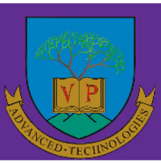
PANNON EGYETEM
MŰSZAKI INFORMATIKAI KAR



PANNONIAN CONFERENCE ON ADVANCES IN INFORMATION TECHNOLOGY (PCIT 2019)

pcit2019.mik.uni-pannon.hu

SZÉCHENYI 2020

MAGYARORSZÁG KORMÁNYA

Európai Unió
Európai Szociális Alap

BEFEKTETÉS A JÖVŐBE

Proceedings of the

Pannonian Conference on Advances in Information Technology (PCIT 2019)

30 May – 1 June 2019

University of Pannonia, Veszprém, Hungary

Editor: István Vassányi

Cover design: Viktória Nagy

Published by the University of Pannonia, Faculty of Information Technology

ISBN 978-963-396-127-8

Preface

The Pannonian Conference on Advances in Information Technology (PCIT 2019) was organized by the Faculty of Information Technology, University of Pannonia, Veszprém, Hungary together with the Information Technology in Healthcare Work Committee and the Operation Research Work Committee of the Regional Centre of the Hungarian Academy of Sciences, Veszprém, Hungary on May 31 – June 1, 2019. The scientific program of the conference consisted of classical and interdisciplinary areas of information technology including, e.g., software technology, intelligent systems, image processing, data analysis, process modeling and optimization, medical and industrial applications. After a two-round peer review process, 36 papers of 82 co-authors from 16 academic institutes in 6 countries were accepted for oral presentation at the conference, out of which 29 are included in this proceedings as full papers.

I thank the members of the Scientific Committee for all their efforts in putting together the scientific program, and I thank all authors and participants for their contributions.

Ferenc Hartung

Dean
Faculty of Information Technology
University of Pannonia

PCIT 2019 Scientific Committee

Chair:

Hangos, Katalin, University of Pannonia

Members:

Bari, Ferenc, University of Szeged
Bodó, Zalán, Babes-Bolyai University
Borbély, Ákos, Óbuda University
Gál, Zoltán, University of Debrecen
Gerzson, Miklós, University of Pannonia
Gyimóthy, Tibor, University of Szeged
Hartung, Ferenc, University of Pannonia
Heckl, István, University of Pannonia
Kis, Tamás, Hungarian Academy of Sciences Institute for Computer Science and Control
Kiss, Attila, Eötvös Loránd Univesity
Kovács, Levente, Óbuda University
Krész, Miklós, University of Szeged
Kruzslicz, Ferenc, University of Pécs
Márton, Lőrinc, Sapientia Hungarian University of Transylvania
Stark-Werner, Ágnes, University of Pannonia
Szederkényi Gábor, Pázmány Péter Catholic University
Várady, Géza, University of Pécs
Vassányi, István, University of Pannonia

Contents

Session 5: Mobile and community applications

Session 6: Process modeling and optimization

Session 7: Applications in social sciences and control

Session 8: Data analysis

For the full PCIT 2019 conference programme including also the oral presentations without published papers please see
https://pcit2019.mik.uni-pannon.hu/images/program/PCIT_program_english.pdf

Application of the maximal bipartite matching algorithm to schedule medical appointments with quotas

András Éles[1], István Heckl[1]
[1]Department of Computer Science and Systems Technology,
University of Pannonia, Veszprém, Hungary, eles@dcs.uni-pannon.hu

Abstract: **A scheduling problem involving the determination of medical examination and treatment times, where quotas can be given as a requirement, is solved by a reformulation as the maximal bipartite matching problem. Finding appointment times for a patient can be nontrivial. Specific limitations may arise, including time windows for the examination, as well as reserved times. The problem becomes difficult if many examinations must be scheduled at once. Also, performance volume quota can be included for specific types of examinations. An algorithm is developed which addresses the scheduling problem and the aforementioned constraints, provided that specific assumptions are made about both the problem and its solution itself. Computational results show that large scale problems can be solved in acceptable time with this method.**

Introduction

Nowadays, management of healthcare institutions is a problem, where complex decisions must be made fast. Therefore it is mostly supported by computer systems. It is usual that a patient must visit an institution regularly, where he must go through various examination and treatment procedures. Throughout this paper, we call these examinations and treatments as appointments.

Finding a time for a single appointment for a patient is a common scenario. This task can usually be done easily and fast. The doctor or other personnel must specify the requirements of the appointment. That may include a preferred time window in which the appointment may take place. Too early or too late appointments can be prohibited. Usually the appointment is scheduled on the first free time slot, that is easily found by a first-fit algorithm.

However, the problem can become complex in certain situations. The appointment may be subject to a particular doctor or facility in the institute, meaning that resource requirements must be carefully taken into account.

Finding appointment times can be difficult, if there are more than one appointments to be scheduled at a single time. This may happen if the treatment consists of a sequence of visits. Not only the constraints for each individual visit must be taken into account, but the visits may depend on each other. For example, exact order can be specified and minimal waiting time can be expected for consecutive visits. Scheduling them one-by-one usually leads to a suboptimal solution. A new method supporting multiple appointments is introduced here.

Of the possible practical limitations, performance volume quota [1] is used in Hungary and state financed healthcare institutions rely on it strongly. The performance volume quota gives the number of specific treatments and examinations the state finances. Being away from this limitation in either direction is disadvantageous for a hospital. Our goal is to find a schedule with all constraints satisfied.

There is a wide range of solution methods for scheduling; the appropriate choice depends on the problem itself. One popular approach is the utilization of Mixed-Integer Linear Programming models. Scheduling can be modeled with time intervals [2], dedicated slots [3], or precedence relationships [4]. Developing a MILP model may be difficult and the computational needs can be prohibitive. The literature of scheduling specific to healthcare, namely the Patient Admission Scheduling problem is itself a vast research topic. MILP modeling is a common option [5]. Due to the size of the problem often heuristics and decomposition methods are used [6], even and especially when the scheduling problem is dynamic [7].

In the present work, a simplified, specific case of patient scheduling problems was identified, and reformulated as the maximal bipartite matching problem [8]. This is a well-known problem for which algorithms exist that run in polynomial time of the input size, and hence are capable of solving scheduling problems with a large number of appointments, and large time span of the institute.

Other constraints or considerations can be implemented, provided that the problem can still be reformulated as a polynomial algorithm. For example, it is possible to differentiate appointment times, and express these as weights, the sum of which are to be minimized instead. This more general problem is still solvable in polynomial time with the Hungarian algorithm [9].

In our paper, we specify the problem we intended to solve, and assumptions made about the solution of the problem. Then, the reformulation is briefly shown. Computational tests show the method

working on large problems; nevertheless, the theoretical capabilities are also discussed.

Problem specification

Scheduling problems can be NP-hard even for simple restrictions. We intended to have a simple, basic problem formulation that can be solved fast, and the algorithm of which can be extended if other practical considerations are added in the future. This means that the following specification is stricter and much simpler than actual practice. However, a fast algorithmic solution of this specification can still be useful in the implementation of more complex algorithmic frameworks, for finding initial, approximate solutions or strict bounds.

In the problem specification, our assumptions about the timings of the appointments are the following.

- The resources of the healthcare institute are not monitored; the only constraint in terms of the institute is time capacity, some of which can be already reserved.
- Appointment times and time constraints rely on a daily precision. That means, scheduling inside a day is neglected. We assume it would be possible to specify exact schedules based on the solution of this specification, because appointments cannot exceed the capacity of the institute for that particular day.
- Appointments all have the same length, which is also the unit of measuring time capacity.
- The only constraint for scheduling a single appointment is that it must be on a specific set of days. This especially causes that appointments are independent of each other.

Note that the specification allows any set of days for each appointment individually. This can be an interval (say at least 6 weeks and at most 8 weeks from present time), or a particular subset (say Tuesdays and Thursdays).

The performance volume quota is defined as several disjoint intervals, in all of which there is a given number for a specific type of appointments. Each appointment may belong to at most one quota, but it can be held in any of that quota's intervals. For each quota interval, the number of involved appointments must be fixed. For example, a quota may state that there should be exactly 20 CT scans every month.

One important assumption is made about the quotas, in order to make this approach working, which is the following: for all appointments of a quota,

the selection of the quota interval can be made a priori. This is done, for example, with consecutively assigning appointments to quota intervals, in the order of the appointments' deadlines. Note that this is a preprocessing step in this approach.

Reformulation

The aforementioned specification and assumption on the quotas mean that, for each individual appointment, a single interval constraint is added. This does not alter the original specification, which allows any subset of days for each appointment (see Figure 1).

This gives rise to a bipartite graph model, where appointments are the first partition of the nodes, and units of possible appointment times are the other partition. Possible assignment of an appointment to a specific time is represented by an edge. Missing edges mean that the particular appointment cannot be on a particular day. Note that for each day, the number of identical nodes in the graph is the capacity of the healthcare institute for that day. The schedule is done if a matching involving all nodes for appointments is found (see Figure 2).

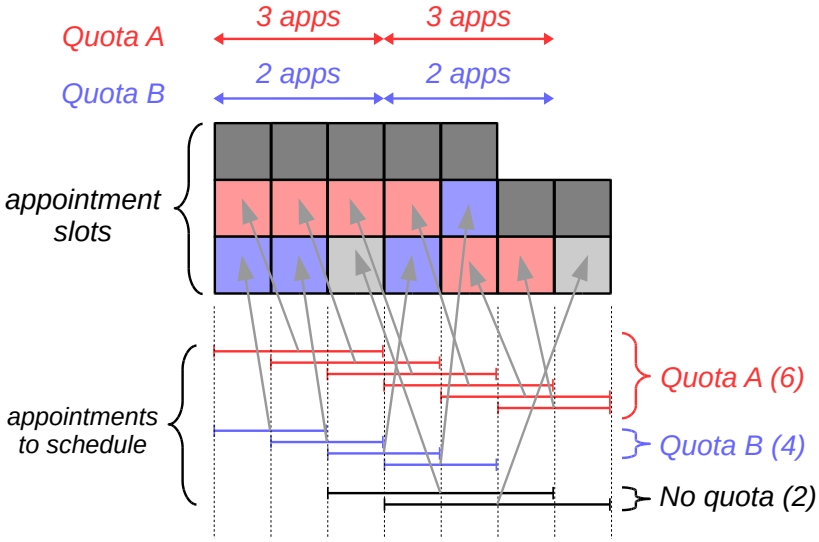


Figure 1. This is the single unique solution of an example appointment scheduling problem. There are 12 free slots for appointments on a weekly schedule, with two quotas, and 2-4 day time windows for each appointment.

The quotas require 3 and 2 appointments in each of their two quota intervals, respectively, which can be perfectly satisfied.

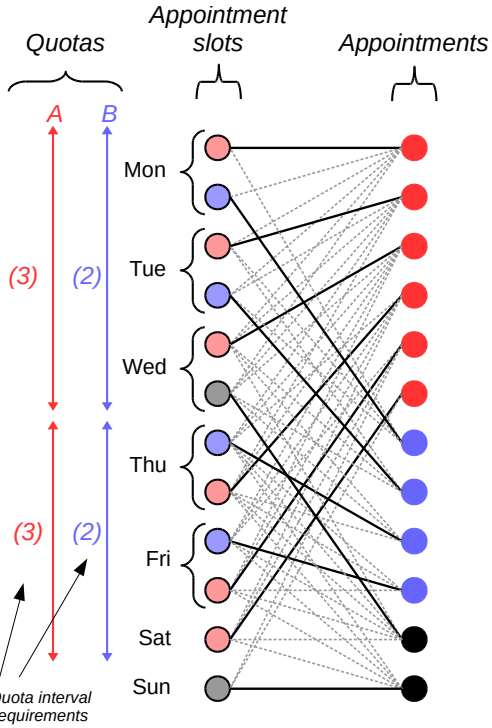


Figure 2. This is the solution of the same problem, represented by a perfect matching of the bipartite graph model. Note that colors represent quotas, and nodes on the left side represent the fixed appointment times, the color of which are determined by the scheduling. It can be seen that all four quota intervals have the exact number of required appointments scheduled.

Implementation and testing

A program which iteratively improves the matching found by looking for augmenting paths is implemented. It is one implementation for the maximal matching algorithm. In addition to finding a solution for the problem, the program finds the one with the most appointments scheduled.

The algorithm was implemented as a pure C++ program, with its own data file format. Problem data consist of daily capacity data, quota data (intervals and limit values), and appointment data (times restricted). Note

that the most recent implementation only allows intervals to be given as a day set for an appointment, but this could be easily relaxed if needed.

The program was capable of solving a large problem with a 180 days long time span, 3 quotas, and 2432 appointments for 3496 possible appointment times.

The algorithm, not counting the data parsing and presentation parts, worked for 356.62 seconds and scheduled 2334 appointments. Further optimization of the code is possible and could yield results faster.

We must note if the algorithm does not succeed in scheduling all appointments, then it may be because of the a priori assignment of appointments under a quota, to a quota interval. A different assignment may result in scheduling more or even all appointments instead, although this is unlikely.

Conclusions

The scheduling of examinations and treatments in healthcare institutes is addressed in a particular, simplified problem class, by reformulating the problem as finding the maximal matching in a bipartite graph. Time windows as well as arbitrary timing constraints for single appointments can be formulated, as well as performance volume quotas. The algorithm was shown to be working fast for large problems. If problem sizes allow solution of a large number of problems, then this method can be used in more sophisticated algorithmic frameworks, for initial solutions, approximation, or finding bounds for more difficult and practically realistic problem classes.

Acknowledgment

We acknowledge the financial support of Széchenyi 2020 programme under the project No EFOP-3.6.1-16-2016-00015.

References

[1] "Teljesítményvolumen korlát", https://fogalomtar.aeek.hu/index.php/TVK, Accessed: 2019.01.21.

[2] E. Kondili, C.C. Pantelides, R.W.H. Sargent, "A general algorithm for short-term scheduling of batch operations--I. MILP formulation", Computers and Chemical Engineering, 1993, 211-227

[3] J.M. Pinto, I.E. Grossmann, "A Continuous Time Mixed Integer Linear Programming Model for Short Term Scheduling of Multistage Batch Plants", Industrial and Engineering Chemistry Research, 1995, vol. 34, 3037-3051

[4] C.A. Mendez, J. Cerda, "An MILP Continuous-Time Framework for Short-Term Scheduling of Multipurpose Batch Processes Under Different Operation Strategies", Optimization and Engineering, 2003, vol. 4, 7-22

[5] L.S.L. Bastos, J.F. Marchesi, S. Hamacher, J.L. Fleck, “A mixed integer programming approach to the patient admission scheduling problem”, European Journal of Operations Research, 2019, vol. 273, 831-840.
[6] A.M. Tuhran, B. Bilgen, “Mixed integer programming based heuristics for the Patient Admission Scheduling problem”, Computers and Operations Research, 2017, vol. 80, 38-49.
[7] Y.H. Zhu, T.A.M. Toffolo, W. Vancroonenburg, G.V. Berghe, “Compatibility of short and long term objectives for dynamic patient admission scheduling”, Computers and Operations Research, 2019, vol. 104, 98-112.
[8] Douglas B. West, "Introduction to Graph Theory", Chapter 3, Pearson Education, 2002
[9] T.H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein, "Introduction to Algorithms", Chapter 6, The MIT Press, 2003

Hybrid time-quality-cost trade-off problems

Zs.T. Kosztyán[1], I. Szalkai[2]

[1]Univ. of Pannonia, Dep. Quantitative Methods, kzst@gtk.uni-pannon.hu
10 Egyetem u., Veszprém, H-8200, Hungary
[2]Univ. of Pannonia, Dep. Mathematics, szalkai@almos.uni-pannon.hu
10 Egyetem u., Veszprém, H-8200, Hungary

Abstract: We propose a matrix-based foundation and algorithmic treatment for agile and hybrid time-quality-cost trade-off project management problems. Our method handles scores for alternative project plans, flexible task dependencies and undecided, supplementary task completion while also covers traditional time-quality-cost trade-off problems, detailed in [1].

We also provide a mathematical foundation of the problem.

Acknowledgment: We acknowledge the financial support of Széchenyi 2020 programme under the project No EFOP-3.6.1-16-2016-00015.

1 Introduction

Reducing time and cost while keeping or even increasing the quality of the project is one of the most important, but most challenging task of the project management. Every parameter can only be improved against the expense of each other. This problem is a so called time-quality-cost trade-off problem. The *discrete time-cost trade-off problem*(DTCTP) is a well-known problem in the project management literature. DTCTP and *discrete time-quality-cost trade-off problems* (DTQCTP) are NP-hard, and are therefore usually solved using heuristic or meta-heuristic methods, while continuous versions of these problems can usually be solved within a polynomial computational time. The present paper extends the traditional trade-off problem to address flexible project plans, models manage flexible project plans and allow us to restructure or reorganize these project plans to satisfy customer and management demands. To handle flexible project plans, *matrix-based techniques* will be used instead of traditional network-based project planning techniques. Therefore, two group of methods, such as trade-off methods and project scoring and screening methods are combined into one, called *hybrid trade-off method*. However, in contrast to the traditional project scoring and screening methods, there is no need to specify all project alternatives to select the most desirable project scenario or the one with the shortest duration or lowest cost.

In the model of the present paper first we are given a set of possible **tasks** (A) containing **mandatory** (compulsory) and **supplementary** (optional)

ones, **score functions** (P,Q) give scores to our choices of supplementary tasks. We also are given **relations** ($\prec$,$\sim$,$\bowtie$) among the tasks: which must be finished before / after, may be handled simultaneously. These relations also may be mandatory or supplementary, another score functions (P,Q) give scores to our choices. Third, for each task ($a \in A$) we are given a *set* (W_a) of **protocols**, i.e. a *list* of possible treating methods for the task a. These protocols include cost, time, quality and resource data for each treating.

In PHASE ONE we have to decide which supplementary tasks to be chosen, maximizing a certain value $\otimes$(M'), calculated from the score functions P,Q , fulfilling also some requirements (C_c,C_t,C_{diag}).

In PHASE TWOO we have to decide the supplementary relations to fix the completion order of the tasks, maximizing the value of $\otimes_{nd}$(M''), calculated from P,Q, and meeting some requirements (C_t,C_{nd}).

In PHASE THREE we state and solve several optimization problems (separately) for deciding a protocol for each task handling, as: **time**($t(w^{\rightarrow})$) is minimized, **cost** ($c(w^{\rightarrow})$) is minimized, *or* **quality** ($q(w^{\rightarrow})$) is maximized. In this phase we also have prescribed requirements (C_c ,C_t ,C_r).

The basis of the proposed methods is the *project domain matrix* PDM [2]. Our algorithm and simulations are described in Sections 3 and 4.

Currently, hybrid (i.e. combinations of traditional and agile) approaches are becoming increasingly popular, however, these approaches lack a principled *mathematical foundation* and algorithmic treatment. Our first goal is to fill this gap in this paper, so, for describing exactly the problem and our results we need unfortunately many definitions and notations at first.

2 Mathematical definitions

We are given a finite set $A=\{a_1,\dots,a_n\}$ of **possible tasks,** $A^{\bar{}} \subseteq A$ contains of the **mandatory** and $A^{\tilde{}}=A\backslash A^{\bar{}}$ the **supplementary** tasks. Any $P:A\rightarrow[0,1]$ is a **score function of** task **inclusion** if $P(a_i)=1$ for $a_i \in A^{\bar{}}$ and $P(a_j)\in[0,1)$ for $a_j \in A^{\tilde{}}$. $Q:A\rightarrow[0,1]$ is **score of** task **exclusion** if $Q(a_i)=0$ for $a_i \in A^{\bar{}}$ and $Q(a_j)\in(0,1]$ for $a_j \in A^{\tilde{}}$ (e.g. probability, importance, relative priority).

$\Xi(A) := \{S \subseteq A : A^{\bar{}} \subseteq S\}$ is the set of realizable (**project**) **scenarios.**

$\otimes : \Xi(A)\rightarrow R$ is an **aggregate function** if $\otimes(S) = \otimes_{a\in S}P(a) \otimes \otimes_{a\in A\backslash S}P(a)$ for any monotone operation $\otimes$ on R (e.g. Σ, $\prod$, etc.).

Among the tasks in A we are given three relations: $a_i \prec a_j$ **strict** or **required**: a_j may not be started unless a_i has been completed; $a_i \sim a_j$ means **no de-**

pendency and $a_i \bowtie a_j$ **flexible** dependencies must be resolved (decided) by the algorithm. Scores P and Q are also given for these dependencies: for $i \neq j$: $a_i \prec a_j \Leftrightarrow$ "$P(a_i,a_j)=1$ and $Q(a_i,a_j)=0$", $a_i \sim a_j \Leftrightarrow$ "$P(a_i,a_j)=0$ and $Q(a_i,a_j)=1$", $a_i \bowtie a_j \Leftrightarrow$ "$0<P(a_i,a_j),Q(a_i,a_j)<1$".

We use the **matrix representation** $M=[m]_{ij} \in \{X,\emptyset,?\}^{n \times n}$ of the above input as $m_{i,i}=X$ for $a_i \in A^{-}$, $m_{i,i}=?$ for $a_i \in A^{\sim}$ and for $i \neq j$ we have $m_{i,j}=X \Leftrightarrow a_i \prec a_j$, $m_{i,j}=\emptyset \Leftrightarrow a_i \sim a_j$ and $m_{i,j}=? \Leftrightarrow a_i \bowtie a_j$. P and Q are represented in the matrices P and Q similarly.

The algorithm will change all ? to either X or $\emptyset$ in M in the diagonal in PHASE ONE and in the off-diagonal in PHASE TWO, the resulted matrices are the **in-** and **out-** (*-diagonal*) **closures** of M.

Clearly, if M contains no "?" in the diagonal then for the represented scenario $S \subseteq A$ we have $\otimes(S)=\otimes_{diag}(M)$ where

$$\otimes_{diag}(M) := \otimes\{P(i):m_{i,i}=X\} \otimes \otimes\{Q(i):m_{i,j}=\varnothing\},$$
$$\otimes_{diag}^{\min}(M) := \otimes_{diag}(M) \otimes \otimes\{min(P(i),Q(i)) : m_{i,i}=?\},$$
$$\otimes_{diag}^{\max}(M) := \otimes_{diag}(M) \otimes \otimes\{max(P(i),Q(i)) : m_{i,i}=?\}.$$

In PHASE THREE we must decide *how to treat* the elements of A by given **protocols** (*modes* or *methods*): paying **cost** c with **quality** q and resource (vector) $\mathbf{r}$ to handle the element $a \in A$ in **time** t.

$W = \{(t_i,q_i,c_i,\mathbf{r}_i):i=1,\ldots,k\}$, $\mathbf{r}_i=\{r_{i,1},\ldots,r_{i,r}\}$ is a **discrete** *time-quality-cost trade-off* **protocol** (**DTQCTp**) with resource demands if $t_1<\ldots<t_k$, $q_1<\ldots<q_k$, $c_1 \geq \ldots \geq c_k$, $\mathbf{r}_1 \geq \ldots \geq \mathbf{r}_k$. We write $t_{\min}$,$t_{\max}$, $q_{\min}$, $q_{\max}$, $c_{\min}$, $c_{\max}$, $\mathbf{r}_{\min}$ and $\mathbf{r}_{\max}$ instead of t_1, t_k, q_1, q_k, $\mathbf{r}_k$, $\mathbf{r}_1$, c_k and c_1 respectively. For each $a \in A$ we are given a protocol W_a , the set $\{W_a:a \in A\}$ is a discrete time-quality-cost trade-off **problem** (**DTQCTP**).

Any *positive, continuous, strictly decreasing function*

$$w: [t_{\min},t_{\max}] \rightarrow [q_{\min},q_{\max}] \times [c_{\min},c_{\max}] \times [\mathbf{r}_{\min},\mathbf{r}_{\max}]$$

is a **continuous** *time-quality-cost trade-off* **protocol** (**CTCQTp**) with resource demands if $0<t_{\min} \leq t_{\max}$ $0<q_{\min} \leq q_{\max}$, $0<c_{\min} \leq c_{\max}$, $0<\mathbf{r}_{\min} \leq \mathbf{r}_{\max}$. The set $\{w_a:a \in A\}$ is a continuous time-quality-cost trade-off **problem** (**CTQCTP**).

Any finite set of four dimensional continuous random variables $\xi = \{\mu_i: i=1,\ldots,k\}$ is a **stochastic** *time-quality-cost trade-off* **protocol** (**STQCTp**) if $E(\mu_i)=(t_i,q_i,c_i,\mathbf{r}_i)$ and $\{(t_i,q_i,c_i,\mathbf{r}_i):i=1,\ldots,k\}$ form a DTQCTp. $\{\xi_a:a \in A\}$ is a stochastic time-quality-cost trade-off **problem** (**STQCTP**).

We *interpret* the **protocol** $(t,q,c,\mathbf{r})\in W_a$ or $w_a(t)=(q,c,\mathbf{r})$ as paying **cost** c with **quality** q and resource (vector) $\mathbf{r}$ to handle the element $a\in A$ in **time** t. Both in discrete and continuous problems we write $(t,q,c,\mathbf{r})\in w_a$. The elements $t^a_{\min}$, $t^a_{\max}$, $q^a_{\min}$, $q^a_{\max}$, $c^a_{\min}$, $c^a_{\max}$, $\mathbf{r}^a_{\min}$, $\mathbf{r}^a_{\max}$ may be different in different protocols w_a for each $a\in A$; $t^a_{\min}=t^a_{\max}$, $q^a_{\min}=q^a_{\max}$, $c^a_{\min}=c^a_{\max}$ or $\mathbf{r}^a_{\min}=\mathbf{r}^a_{\max}$ are also allowed.

For any M and DTQCTP or CTQCTP W

the **minimal cost-bound** is $C_{\min}(\mathrm{M},W) := \Sigma\{c^a_{\min} : m_{aa}=\text{"X"}\}$,

the **maximal (relative) quality bound** is $Q_{\max}(\mathrm{M},W):=1$.

Our final goal is to find an optimal project **schedule** $w^{\rightarrow}=\{(t^a,q^a,c^a,\mathbf{r}^a):a\in S\}$ where $(t^a,q^a,c^a,\mathbf{r}^a)\in w_a$ for $a\in S$. For any $w^{\rightarrow}$ the **total** *project* **cost** is

$$c(w^{\rightarrow}) := \mathbf{\Sigma}\ \{c^a:(t^a,q^a,c^a,\mathbf{r}^a)\in w_a\ ,\ a\in S\}\ ,$$

the **total** *project* **quality** is

$$q(w^{\rightarrow}) := \frac{\sum\limits_{(t^a,q^a,c^a,\mathbf{r}^a)\in w_a,\,a\in S} q^a}{\sum\limits_{a\in A} q^a_{\max}}$$

For time bounds, we must not forget the $\prec$ dependencies. For any *real* **path** $P^{\rightarrow}=a_{i1}\prec a_{i2}\prec\ldots\prec a_{ik}$ the minimal **time bound** of the path is $\boldsymbol{T}_{\min}(P^{\rightarrow},w) := \mathbf{\Sigma}\{t^a_{\min} : a\in P^{\rightarrow}\}$, and $P^{\rightarrow}$ is a **longest** min-**path** of M if $\boldsymbol{T}_{\min}(P^{\rightarrow},w)$ is maximal, assuming that $P^{\rightarrow}$ contains mandatory tasks only, this maximum is denoted by $\boldsymbol{T}_{\min}(\mathrm{M},W)$, i.e. $\boldsymbol{T}_{\min}(\mathrm{M},W)=max_P\boldsymbol{T}_{\min}(P^{\rightarrow},w)$. This $P^{\rightarrow}$is called **critical path** and $\{a_{i1},a_{i2},\ldots,a_{ik}\}$ is the set of **critical activities.**

The **total** *project* **time** is $t(w^{\rightarrow}) :=\Sigma\{t^a:(t^a,q^a,c^a,\mathbf{r}^a)\in w_a,a\in P^{\rightarrow}\}$ where $P^{\rightarrow}$ is any longest min-path.

The length and definition of the longest min-path do not depend on $w^{\rightarrow}$ since $t^a_{\min}$ are summed in $t(w^{\rightarrow})$. In fact, critical paths are longest min-paths. Clearly $t(w^{\rightarrow})\ge T_{\min}(\mathrm{M},w^{\rightarrow})$ for any W and $w^{\rightarrow}$.

The **maximal resource demand** for resource k is $\boldsymbol{r_k}(w^{\rightarrow}):=\max_t R_{kt}$ where $R_{kt}=\Sigma\{r_{ik}:a_i\in A(w^{\rightarrow},\mathrm{t})\}$, $A(w^{\rightarrow},\mathrm{t})\subseteq A$ is the set of running activities in time t for the schedule $w^{\rightarrow}$ and $k=1,\ldots,r$.

Clearly $C_{\min}(\mathrm{M},W)\le C_{\min}(\mathrm{N},W)$, $Q_{\max}(\mathrm{M},W)\ge Q_{\max}(\mathrm{N},W)$ and $T_{\min}(\mathrm{M},W)\le T_{\min}(\mathrm{N},W)$ for any *in-* or *out*-closure N of M. Further, $C_{\min}(\mathrm{N},W)\le c(w^{\rightarrow})$, $Q_{\max}(\mathrm{N},W)\ge q(w^{\rightarrow})$ and $T_{\min}(\mathrm{N},W)\le t(w^{\rightarrow})$ for any $w^{\rightarrow}$ determined by N.

For $\mathrm{M}\in\{X,\emptyset\}^{n\times n}$ and $w^{\rightarrow}$ we also define the **total project quality-, cost-, time-** and **resource-demands** as $TPC(\mathrm{M},w):=c(w^{\rightarrow})$, $TPQ(\mathrm{M},w):=q(w^{\rightarrow})$, $TPT(\mathrm{M},w^{\rightarrow}):=t(w^{\rightarrow})$ and $TPR(\mathrm{M},w^{\rightarrow}):=[r_1(w^{\rightarrow}),\ldots,r_r(w^{\rightarrow})]^{\mathrm{T}}$.

The **aggregation function** for project structures and its extreme values are (*nd* means "*no diagonal*"):

$$\otimes_{nd}(\mathrm{M}) := \otimes\{\mathrm{P}(i,j):m_{i,j}=\mathrm{X},i\neq j\} \otimes \otimes\{\mathrm{Q}(i,j):m_{i,j}=\varnothing,i\neq j\},$$

$$\otimes_{nd}^{\min}(\mathrm{M}) := \otimes_{nd}(\mathrm{M}) \otimes \otimes\{min(\mathrm{P}(i,j),\mathrm{Q}(i,j)) : m_{i,j}=?,i\neq j\},$$

$$\otimes_{nd}^{\max}(\mathrm{M}) := \otimes_{nd}(\mathrm{M}) \otimes \otimes\{max(\mathrm{P}(i,j),\mathrm{Q}(i,j)) : m_{i,j}=?,i\neq j\}.$$

$\otimes_{nd}(\mathrm{M})$ is the *score value* of the project structure, represented by M.

We now can define the ***resource-constrained hybrid time-quality-cost trade-off problems*** that we will solve in PHASES ONE, TWO and THREE (the constants C_c,C_t,C_q,C_r,C_{diag} and C_{nd} might be varied upon request).

Problem 1 PHASE ONE. Let C_c,C_t,C_{diag} be given such that $C_{\min}(\mathrm{M},W)\leq C_c$, $T_{\min}(\mathrm{M},W)\leq C_t$ and $C_{diag}\leq\otimes_{nd}^{\max}(\mathrm{M})$. Now, find a *scenario* $S\subseteq A$ (an *in-closure* M′ of M) such that $\otimes(\mathbf{M'})\rightarrow\mathbf{max}$ assuming $C_{\min}(\mathrm{M}',W)\leq C_c$, $T_{\min}(\mathrm{M}',W)\leq C_t$, $Q_{\max}(\mathrm{M}',W)\geq C_q$ and $\otimes_{diag}(\mathrm{M}')\geq C_{diag}$.

Problem 2 PHASE TWO. Let M′ be a solution to PHASE ONE, C_t, C_{nd} be given such that $T_{\min}(\mathrm{M}',W)\leq C_t$, $C_{nd}\leq\otimes_{nd}^{\max}(\mathrm{M}')$. Now, find a *structure* (*off-closure* M"of M′) such that $\otimes_{nd}(\mathbf{M''})\rightarrow\mathbf{max}$ assuming $T_{\min}(\mathrm{M}'',W)\leq C_t$ and $\otimes_{nd}(\mathrm{M}'')\geq C_{nd}$.

After PHASE TWO we have a traditional time-cost trade-off problem, therefore in PHASE THREE we can specify different types ofobjective functions:

Problem 3 Let M" be a solution to PHASE TWO, C_c,C_t,C_{r} be given such that $C_{\min}(\mathrm{M}'',W)\leq C_c$ and $T_{\min}(\mathrm{M}'',W)\leq C_t$.

PHASE THREE /**1.** Find a project schedule $w^{\rightarrow}$such that $\boldsymbol{t(w^{\rightarrow})\rightarrow\min}$ assuming $c(w^{\rightarrow})\leq C_c$, $q(w^{\rightarrow})\geq C_q$ and $\mathrm{r}(w^{\rightarrow})\leq C_{\mathrm{r}}$.

PHASE THREE /**2.** Find a project schedule $w^{\rightarrow}$ such that $\boldsymbol{c(w^{\rightarrow})\rightarrow\min}$ assuming $t(w^{\rightarrow})\leq C_t$, $q(w^{\rightarrow})\geq C_q$ and $\mathrm{r}(w^{\rightarrow})\leq C_{\mathrm{r}}$.

PHASE THREE /**3.** Find a project schedule $w^{\rightarrow}$ such that $\boldsymbol{q(w^{\rightarrow})\rightarrow\max}$ assuming $c(w^{\rightarrow})\leq C_c$, $q(w^{\rightarrow})\geq C_q$ and $\mathrm{r}(w^{\rightarrow})\leq C_{\mathrm{r}}$.

3 Computer solution

In [1] and [2] a quasilinear algorithm for the above problems and larger computer results are discussed in detail which do not fit here.

In the traditional approach every score are rounded, therefore, every flexible dependency and every uncertain task occurrence converted to a fix realizations. Therefore only a traditional time-quality-cost trade-off problem had to be solved. In agile approach all uncertain parameters are saved, therefore

the project can be restructured like agile projects, however only one completion mode t_{max}, $c_{min,}q_{max}$ are allowed. This problem can already be solved by Exact Project Ranking algorithm [2],therefore the results of the proposed hybrid method and the EPR can be compared directly.

4 Simulation

At the simulation the proposed method was to model the project management approaches. Methods simulate the decision makers, and the project scheduling approaches. The proposed method contain three phases, while the first two phases select the tasks and dependencies. Without using phase three, we get Kosztyán's [2] Exact Project Ranking algorithm, which models the agile project manager's decisions, who can re-organize the project structure, if it is necessary for keeping deadlines and the budget, however, without using phase three trade-off methods are not involved. Furthermore, we consider Kosztyán's [2] algorithm, as an Agile Project Management agent (APMa). If only the phase three is implemented, we get a classical time-quality-cost trade-off problem, and we call this method as a Traditional Project Management agent (TPMa). The full algorithm can re-organize and can reduce the time-demands of tasks by using trade-off algorithms. Therefore, we call this agent as a Hybrid Project Management agent (HPMa).

The main goal of the simulation phase was to compare the project management agents, while in this case we also compare the Kosztyán's [2] and Kosztyán-Szalkai's [1] algorithms with traditional trade-off approaches [3].

We get 10 project networks form the standard PSPLIB database [3], using 30 sets. Since this database contained only mandatory tasks, we specified *ff* = {0,0.05,0.1,0.15,0.20} of tasks and dependencies as flexible, where *ff* was the flexibility factor. We specified the ratio of constraints, such as *C*%, *T*%, $\otimes_{nd}$% and $\otimes_{diag}$% as 0.1, 0.2,..,1. Therefore, we get $10\times5\times10\times10\times10\times10$ = 500,000 project networks. Figure 1 shows, that most feasible project produced by the proposed algorithm.

However, if we consider only the feasible project schedules, Figure 2 shows that there are on superior approaches. TPMa keeps all tasks, therefore, only in this case the score is maximal. If it is important to keep all tasks only TPMa can be used. However, APMa can save the most cost and HPMa can save the most time. Figure 2 shows the results in a so called project triangle.

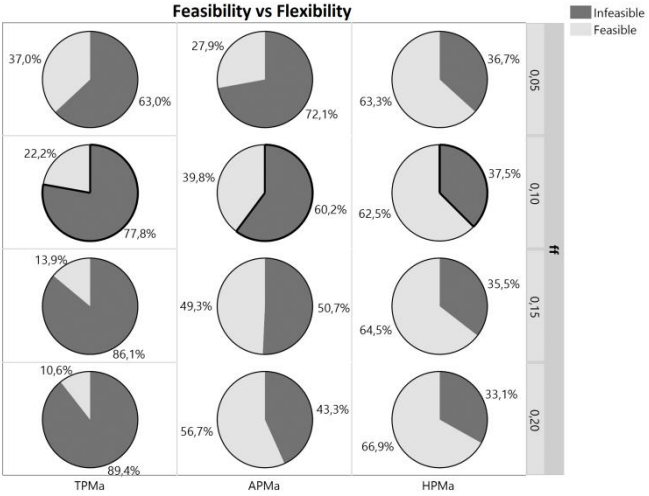


Figure 1 *Comparing project management agents*

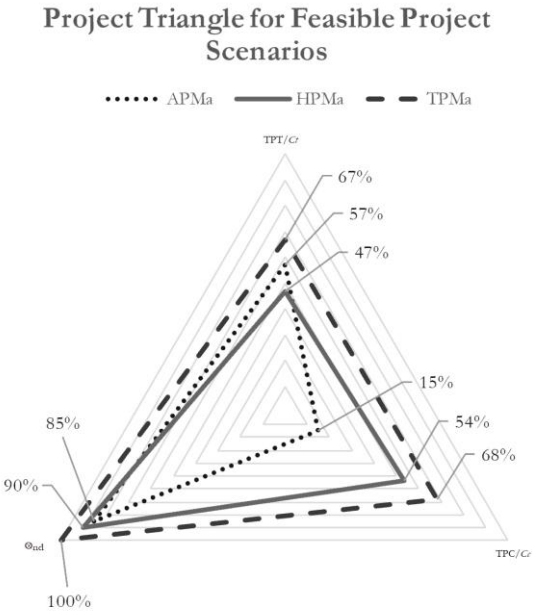


Figure 2 *Results of feasible project schedules*

Notations:

$\mathbf{C}\% = (C_{max}\text{-}C)/(C_{max}\text{-}C_{min})$, $\boldsymbol{n} = |A|$ = total size of the problem,

$\otimes_{diag}\% = \mathbf{(}\otimes_{diag}(M)\text{-}\otimes^{min}_{diag}(M)\mathbf{)} / \mathbf{(}\otimes^{max}_{diag}(M)\text{-}\otimes^{min}_{diag}(M)\mathbf{)}$,

$\otimes_{nd}\ \% = \mathbf{(}\otimes_{nd}(M)\text{-}\otimes^{min}_{nd}(M)\mathbf{)} / \mathbf{(}\otimes^{max}_{nd}(M)\text{-}\otimes^{min}_{nd}(M)\mathbf{)}$,

$T\% = 1 - \mathbf{(}T - T_{min}(M,W)\mathbf{)/(}T_{max}(M,W)\text{-}T_{min}(M,W)\mathbf{)}$.

5 Conclusion

In this paper we propose a new algorithm, which can be used for all the traditional, agile and the hybrid project management approach. We showed, that the proposed algorithm can produce the most feasible project schedule. While we also showed, that agile project management agent can save the most cost, while the proposed hybrid project management agent can save the most time demands.

References

[1] Z.T.Kosztyán, I.Szalkai. "Hybrid time-quality-cost trade-off problems", Operations Research Persp., vol. 5, pp. 306-318, 2018, https://doi.org/10.1016/j.orp.2018.09.003

[2] Z.T.Kosztyán. "Exact algorithm for matrix-based project planning problems", Expert Syst. Appl. vol. 42, pp. 4460-73, 2015; https://doi.org/10.1016/j.eswa.2015.01.066

[3] R.Kolisch, A.Sprecher. "PSPLIB - a project scheduling problem library: OR software", European Journal of Operational Research, vol. 96, pp. 205–216, 1997, http://www.sciencedirect.com/science/article/pii/S0377221796001701

Distance domination and vertex partitions

A. Frendrup[1], Zs. Tuza[2,3], P.D. Vestergaard[1]
[1]Aalborg University, {frendrup,pdv}@math.aau.dk
DK-9220 Aalborg, Denmark
[2]University of Pannonia, tuza@dcs.uni-pannon.hu
H-8200 Veszprém, Egyetem u. 10, Hungary
[3]MTA Rényi Institute, 1053 Budapest, Reáltanoda u. 13–15, Hungary

Abstract: **We treat a variation of domination which involves a vertex partition and domination of each partition class over a given distance where all vertices and edges may be used in the domination process. Strict upper bounds and extremal graphs are presented. Further, we compare a high number of partition classes and the number of dominators needed. Due to space limitation, the proofs will be published elsewhere.**

Introduction

We deal with finite simple graphs $G = (V, E)$ where $V = V(G)$ is the vertex set and $E = E(G)$ is the edge set. The *order* of G is $|V(G)| = n$, and the *size* of G is $|E(G)|$. A subset $S \subseteq V$ is called a *dominating set* of G if every vertex of $V \setminus S$ is adjacent to at least one vertex of S. The minimum cardinality of a dominating set is denoted by $\gamma(G)$ and is termed the *domination number* of G. As a further notation, we write $\delta(G)$ for the minimum vertex degree in G. (The degree of a vertex is the number of edges incident with it.)

More than half a century ago Ore [1] defined domination and proved that a connected graph G of order n has $\gamma(G) \leq n/2$. Payan and Xuong [2] and Fink, Jacobson, Kinch and Roberts [3] proved that equality, $\gamma(G) = n/2$, holds precisely for the cycle C_4 of length four, and for corona graphs. (A corona graph G, denoted by $G = H \circ K_1$, has order $2n$ and is obtained from a graph H of order n and n new vertices, one corresponding to each vertex of H, by joining each vertex of H to its corresponding new vertex.)

Many variants of domination in graphs have been surveyed in two books by Haynes, Hedeniemi and Slater [4, 5]. We shall here be concerned with distance domination in partitioned graphs. Turau and Köhler [6] describe various applications of distance domination, with a major motivation in the area of ad hoc and wireless sensor net-

works. Applications include significant reduction of flooding overhead in broadcast, efficient network initialization, server allocation in computer networks, message routing with sparse tables, and more. For details see [6, Section 1.1] and the references therein. On the other hand, the problem of domination in vertex-partitioned graphs is a model to minimize the number of servers in a computer network where file (in)compatibilities are taken into account, as described in [7] and also recalled at the beginning of the paper [8].

More formally, let d be a positive integer and let Y be a subset of V. We say that a set $S \subseteq V$ *distance d dominates* Y if every vertex in Y has distance at most d to some vertex of S. The minimum cardinality of such an S will be denoted by $\gamma_d(G;Y)$. If $Y = V$, this value is the *distance d domination number* $\gamma_d(G)$. In case of $d = 1$ we omit the subscript and simply write $\gamma(G;Y)$ instead of $\gamma_d(G;Y)$; and certainly for $Y = V$ and $d = 1$ we have the ordinary domination, $\gamma_1(G) = \gamma(G)$.

A *partition* $(V_1, V_2, \ldots, V_k)$ of $V = V(G)$ into k disjoint sets, $k \geq 2$, has $V = \bigcup_{i=1}^{k} V_i$ with $V_i \cap V_j = \varnothing$ for all $1 \leq i < j \leq k$. For partitions $(V_1, V_2, \ldots, V_k)$ of V, we define for distance $d = 1$ the following:

$$f(G;V_1,V_2,\ldots,V_k) = \gamma(G) + \gamma(G;V_1) + \gamma(G;V_2) + \ldots + \gamma(G;V_k)$$
$$g(G;V_1,V_2,\ldots,V_k) = \gamma(G;V_1) + \gamma(G;V_2) + \ldots + \gamma(G;V_k)$$
$$f(k,G) = \max_{V_1,V_2,\ldots,V_k} f(G;V_1,V_2,\ldots,V_k)$$
$$g(k,G) = \max_{V_1,V_2,\ldots,V_k} g(G;V_1,V_2,\ldots,V_k)$$

where the maximum is taken over all partitions $(V_1, V_2, \ldots, V_k)$ of V. We observe that $f(k,G) = \gamma(G) + g(k,G)$. For distance at most d, $d \geq 1$, definitions of $f_d(G;V_1,V_2,\ldots,V_k)$ etc. are analogous. Since $\gamma_d(G;V_i) \leq \gamma_d(G)$ and hence $g_d(k,G) \leq k\gamma(G)$ always holds, we have

$$g_d(k,G) \leq \frac{k}{k+1} f_d(k,G)$$

for every graph G and all integers $k \geq 2$ and $d \geq 1$. Moreover for $k = 1$ the upper bound $\gamma \leq n/2$ mentioned above can be extended for any d, as follows. For the precise description we need to introduce the P_d-corona graph, $G = H \circ P_d$, of order $n(d+1)$ obtained as the disjoint union of a graph H of order n and n disjoint paths P_d, each of length $d-1$, by joining each vertex of H to an end vertex of its corresponding path P_d.

Table 1: Bounds on 2-partitioned graphs, $\delta = \delta(G)$; asterisk * indicates that the extremal graphs are known.

T denotes a tree of order n and G a connected graph of order n			
upper bound	conditions	reference	
$f(2,T) \le \frac{5}{4} \cdot n$	$d = 1, n \ge 3$	[7]	*
$g(2,T) \le \frac{4}{5} \cdot n$	$d = 1, n \ge 3$	[9]	*
$f(2,G) \le n$	$d = 1, \delta \ge 2$	[10]	
$g(2,G) \le \frac{2}{3} \cdot n$	$d = 1, \delta \ge 2$	[9]	*
$g(2,G) \le \frac{\delta+1}{2\delta} \cdot n$	$d = 1, \delta \ge 1$	[9]	
$f_d(2,T) \le \frac{6}{2d+3} \cdot n$	$d \ge 2, n \ge d+2$	[11]	*
$g_d(2,T) \le \frac{4}{2d+3} \cdot n$	$d \ge 2, n \ge d+2$	Theorem 5, $k = 2$	

Theorem 1 *Let $d \ge 1$ be an integer and let G be a connected graph with diameter at least d (hence $n > d$). Then $\gamma_d(G) \le \frac{n}{d+1}$ where equality holds if and only if $n = d+1$, or $G \cong C_{2d+2}$, or $G \cong H \circ P_d$ for a connected graph H.*

Noting that g_d can never exceed the order of the graph in question, from Theorem 1 we immediately obtain the following universal bounds on f_d and g_d.

Corollary 1 *If G is a graph and $k, d \ge 1$ are integers then $g_d(k,G) \le |V(G)|$ and if G is a connected graph such that $|V(G)| \ge d+1$ then $f_d(k,G) \le \frac{d+2}{d+1} |V(G)|$.*

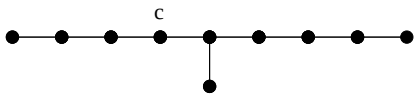
Note further that a connected graph G has its various domination numbers bounded above by the corresponding domination number for any one of its spanning trees T, e.g. $f(2,G) \le f(2,T)$, and if we search for an upper bound holding for all connected graphs of order n it suffices to search among all trees of order n, e.g. $f(2,G) \le f(2,T) \le \frac{5n}{4}$. As exhibited in Table 1, several tight results are known for 2-partitioned graphs, and in most of them the extremal graphs are characterized, too.

Bounds for $f_d(3,T)$ and $g_d(3,T)$

Our main concern in this paper is to provide tight estimates on 3-partitioned graphs. As we noted above, the worst-case behavior of

$f_d(k,G)$ and $g_d(k,G)$ over connected graphs occurs when G is a tree. Moreover, the case $|V(G)| \leq d+1$ is trivial. For this reason we concentrate on bounds for $f_d(3,T)$ and $g_d(3,T)$ when T is a tree with at least $d+2$ vertices. First some families of graphs will be defined.

For each integer $d \geq 2$ let Q_d be the family of trees consisting of P_{2d+4} and all trees with $d+2$ vertices. Let G_{10} denote P_9 with a pendent vertex attached to its center, i.e., the graph with 10 vertices illustrated in the next figure.



A neighbor c to the center of the path $v_1, v_2, v_3, c, v_5, \ldots, v_9$ in G_{10} is called a connection-vertex in G_{10}. Let further $Q_2' = Q_2 \cup \{P_6, P_7, G_{10}\}$, $Q_3' = Q_3 \cup \{P_9\}$, and let $Q_d' = Q_d$ for $d \geq 4$.

We summarize parameters for the specific graphs mentioned above in the following small table. Double separation indicates the examples for $d = 2$ and the last one for $d = 3$, respectively.

graph	$\|V(T)\| = d+2$	P_{2d+4}	G_{10}	P_6	P_7	P_9
$g_d(3,T)$	3	6	7	4	5	5
$\gamma_d(T)$	1	2	3	2	2	2

For $d \geq 2$ let T_d be the tree with the smallest diameter, $2d+6$, that can be obtained from $3P_{2d+4} \cup K_1$ by adding three edges all incident with the isolated vertex which will be called the central vertex in T_d. For $d \geq 2$ we define $\mathcal{F}_d$ as the family of trees that can be obtained from graphs isomorphic to T_d by adding edges between their central vertices. Let T_2' be the tree obtained from $3G_{10} \cup K_1$ by adding three edges all incident with the isolated vertex (this vertex will be called central in T_2') and a connection-vertex from each of the three G_{10}-components. Define $\mathcal{F}_d' = \mathcal{F}_d$ for $d \geq 3$ and $\mathcal{F}_2'$ as the family of trees that can be obtained from isomorphic copies of T_2' by adding edges between central vertices. With these notations we have the following result. It also involves the particular graphs mentioned above; if we disregard them, a summary of the situation for $k = 3$ can be given as exhibited in Table 2.

Table 2: Bounds on 3-partitioned graphs; all extremal graphs are known.

upper bound	conditions	reference
$f(3,T) \leq \frac{7}{5} \cdot n$	$d = 1, n \geq 3$	[7]
$f_2(3,T) \leq n$	$d = 2, n \geq 4$	[11]
$f_2(3,T) \leq \frac{30}{31} \cdot n$	$d = 2, n \geq 5$ $T \notin \{P_6, P_7, P_8, G_{10}\}$	Theorem 2
$g_2(3,T) \leq \frac{18}{25} \cdot n$	$d = 2, n \geq 5$ $T \neq T_8$	Theorem 2
$f_3(3,T) \leq \frac{24}{31} \cdot n$	$d = 3, n \geq 6,$ $T \notin \{P_9, P_{10}\}$	Theorem 2
$f_d(3,T) \leq \frac{24}{6d+13} \cdot n$	$d \geq 4, n \geq d+3,$ $T \neq P_{2d+4}$	Theorem 2
$g_d(3,T) \leq \frac{18}{6d+13} \cdot n$	$d \geq 3, n \geq d+3$ $T \neq P_{2d+4}$	Theorem 2

Theorem 2 *Let $d \geq 2$ be an integer and let T be a tree with $n \geq d+2$ vertices. Then the behavior of $f_d(3,T)$ and $g_d(3,T)$ is as follows.*

- *If $d = 2$ then $f_d(3,T) = n$ if $T \in Q'_d$, and if $T \notin Q'_d$ then $f_d(3,T) \leq \frac{30}{31}n$ where equality holds if and only if $T \in \mathcal{F}'_d$.*
- *If $d \geq 3$ then $f_d(3,T) = \frac{4}{d+2}n$ if $T \in Q_d$, and if $T \notin Q'_d$ then $f_d(3,T) \leq \frac{24}{6d+13}n$ where equality holds if and only if $T \in \mathcal{F}'_d$.*
- *If $d \geq 2$ then $g_d(3,T) = \frac{3}{d+2}n$ if $T \in Q_d$, and if $T \notin Q_d$ then $g_d(3,T) \leq \frac{18}{6d+13}n$ where equality holds if and only if $T \in \mathcal{F}_d$.*

Many partition classes

Besides the case of few partition classes, we also investigate the other extreme, where the number of classes is very large. Our results in this direction show that the best possible universal upper bound on $g_d(k,G)$ is the trivial one, namely n, for all n and d, whenever $k \geq (d+1)^2$; and for such large k, the best bound on $f_d(k,G)$ is $\frac{d+2}{d+1}n$. On the other hand, if k is any smaller, then the upper bounds can be improved.

Theorem 3 *Let $d \geq 1$ be a integer and let T be a tree with n vertices. Then the following relations are valid.*

- $f_d(d^2+2d+1, P_{\frac{n}{d+1}} \circ P_d) = \frac{d+2}{d+1}n$ *if* $(d+1) \mid n$.
- $g_d(2d+1, P_n) = n$ *for each* $n \geq 1$.
- $g_d(d^2+2d, T) < n$ *if* T *is a* P_d*-corona graph and* $|V(T)| > 2d(d+1)$.
- $f_d(d^2+2d, T) < \frac{d+2}{d+1}n$ *if* $|V(T)| > 2d(d+1)$.
- $f_d(d^2+2d, P_{2d} \circ P_d) = \frac{d+2}{d+1}n$.
- $g_d(2d, T) < n$ *if* $|V(T)| \geq 2d+1$.

Concerning the fourth case, a stronger estimate can also be proved, as shown in the next result.

Theorem 4 *Let $d \geq 1$ be a integer and let T be a tree with $n > 2d^2+2d$ vertices. Then*

$$f_d(d^2+2d, T) < \frac{d+2}{d+1}\, n - \frac{n}{2(d+1)^5}.$$

The following result generalizes Theorem 1.

Theorem 5 *Let G be a tree with $n \geq d + \frac{k+1}{2}$ vertices. Then*

$$g_d(k, G) \leq \frac{2k}{2d+k+1}\, n.$$

Finally, from Theorem 5 we obtain

Corollary 2 *A graph G with $n \geq 2d+1$ vertices satisfies $g_d(2d, G) \leq n - \frac{n}{4d+1}$.*

In [9] it has been proven that this bound is optimal when $d = 1$. In estimates above with strict inequalities, however, it should be a subject of future research to determine tight results.

Acknowledgments

We acknowledge the financial support of Széchenyi 2020 programme under the project No. EFOP-3.6.1-16-2016-00015, and of the National Research, Development and Innovation Office – NKFIH under the grant SNN 129364.

References

[1] O. Ore, Theory of Graphs. Amer. Math. Soc. Colloq. Publ., 38, Amer. Math. Soc., Providence, RI, 1962.

[2] C. Payan and N. H. Xuong, Domination-balanced graphs. J. Graph Theory 6 (1982), 23–32.

[3] J. F. Fink, M. S. Jacobson, L. F. Kinch and J. Roberts, On graphs having domination number half their order. Period. Math. Hungar. 16 (1985), 287–293.

[4] T. W. Haynes, S. T. Hedetniemi and P. J. Slater, Fundamentals of Domination in Graphs. Marcel Dekker, New York, 1998.

[5] T. W. Haynes, S. T. Hedetniemi and P. J. Slater (Eds.), Domination in Graphs: Advanced Topics. Marcel Dekker, New York, 1998.

[6] V. Turau and S. Köhler, A distributed algorithm for minimum distance-k domination in trees. J. Graph Algorithms Appl. 19 (2015), 223–242.

[7] B. L. Hartnell and P. D. Vestergaard, Partitions and dominations in a graph. J. Combin. Math. Combin. Comput. 46 (2003), 113–128.

[8] M. A. Henning and P. D. Vestergaard, Domination in partitioned graphs with minimum degree two. Discrete Math. 307 (2007), 1115–1135.

[9] Zs. Tuza and P. D. Vestergaard, Domination in partitioned graph. Discuss. Math. Graph Theory 22 (2002), 199–210.

[10] S. M. Seager, Partition dominations of graphs of minimum degree 2. Congr. Numer. 132 (1998), 85–91.

[11] C-M. K. Fu and P. D. Vestergaard, Distance domination in partitioned graphs. Congr. Numer. 182 (2006), 155–159.

Strongly possible keys

Munqath Alattar[1] and Attila Sali[2]
[1]Department of Computer Science and Information Theory, BUTE,
m.attar@cs.bme.hu, Budapest, Hungary
[2]Alfréd Rényi Institute of Mathematics, sali.attila@renyi.mta.hu,
Budapest, Hungary

Abstract: **A new concept of keys *strongly possible keys* in relational databases with null values is introduced. It lies between possible keys and certain keys introduced by Köhler et. al. earlier. The definition uses only information extractable from the database table.**

Introduction

A basic approach to treatment of null values in keys of relational databases is imputing a value from the attribute domain for each occurrence of a null as explained by [3]. We investigate the situation when the attributes' domains are not known a priori. That is, we only consider what is given in the relational table and extract the values to be imputed from the data itself so that the resulting complete dataset after the imputation would not contain two tuples having the same value in any key set. Köhler et al.[3] created possible worlds by replacing each occurrence of a null with a value from the corresponding attribute's (possibly infinite) domain. They defined a possible key as a key that is satisfied by some possible world of a non total database table and a certain key as a key that is satisfied by every possible world of the table. In many cases we have no proper reason to assume existence of any other attribute value than the ones already existing in the table. Such examples could be types of cars, diagnoses of patients, applied medications, dates of exams, course descriptions, etc. We define a strongly possible key as a key that is satisfied by some possible world that is obtained by replacing each occurrence of null value from the corresponding attribute existing values. We call this kind of a possible world a strongly possible world. This is a data mining type approach, our idea is that we are given a raw table with nulls and we would like to identify possible key sets based on the data only.

Definitions

Let $R = \{A_1, A_2, \dots A_n\}$ be a relation schema. The set of all possible values for each attribute $A_i \in R$ is called the domain of A_i and denoted by $D_i = dom(A_i)$ for $i = 1, 2, \dots, n$. For $X \subseteq R$ let $D_X = \prod_{\forall A_i \in X} D_i$. An instance $T = (t_1, t_2, \dots t_s)$ over R is a set of tuples that each tuple is a function $t : R \to \bigcup_{A_i \in R} dom(A_i)$ and $t[A_i] \in dom(A_i)$ for all $A_i \in R$. For a tuple $t_r \in T$, let $t_r[A_i]$ be the restriction of t_r to A_i.

In practice, data sets may not contain information about the value of $t_j[A_i]$ for $j = 0, 1, \dots s$, which is denoted by $t_j[A_i] = \perp$. t_1 and t_2 are *weakly similar* on $X \subseteq R$ denoted by $t_1[X] \sim_w t_2[X]$ [3] if

$$\forall A \in X \quad (t_1[A] = t_2[A] \text{ or } t_1[A] = \perp \text{ or } t_2[A] = \perp) \tag{1}$$

Furthermore, t_1 and t_2 are *strongly similar* on $X \subseteq R$ denoted by $t_1[X] \sim_s t_2[X]$ if

$$\forall A \in X \quad (t_1[A] = t_2[A] \neq \perp) \tag{2}$$

For the sake of convenience we write $t_1 \sim_w t_2$ if t_1 and t_2 are weakly similar on R and the same for strong similarity. For a null-free table, a set of attributes $K \subset R$ is a *key* if there are no two distinct tuples in the table that agree in all attributes of K.

$$t_a[K] \neq t_b[K] \; \forall \; 0 \leq a, b \leq s \text{ such that } a \neq b \tag{3}$$

The concepts of possible and certain keys were defined by Köhler et al [3]. Let $T' = (t'_1, t'_2, \dots t'_s)$ be a table that represents a total version of T which obtained by replacing the occurrences of $\perp$ in all attributes $t[A_i]$ with a value from the domain D_i different from $\perp$ for each i. T' is called a *possible world* of T. For a possible world T', t'_i is weakly similar to t_i and T' is completely null-free table. A *possible key* K denoted by $p\langle K \rangle$, is a key for some possible world T' of T

$$t'_i[K] \neq t'_j[K], \;\; \forall t'_i, t'_j \in T', i \neq j. \tag{4}$$

Similarly, a *certain key* K denoted by $c\langle K \rangle$, is a key for every possible world T' of T. The *visible domain* of an attribute A (VD_A) is the set of all distinct values except $\perp$ that are already used by tuples in T:

$$VD_i = \{t[A_i] : t \in T\} \setminus \{\perp\} \text{ for } A_i \in R \tag{5}$$

The term visible domain refers to data that already exist in a given dataset. If we have a dataset with no information about the attributes' domains definitions, then we use the data itself to define the domains. This may provide more realistic results when extracting the relationship between data so it is more reliable considering only what information we have in a given dataset.

A possible world T' is called *strongly possible world* if $T' \subseteq VD_1 \times VD_2 \times \ldots \times VD_n$.

A subset $K \subseteq R$ is a *strongly possible key* (in notation $sp\langle K\rangle$) in T if $\exists$ a strongly possible world $T' \subseteq VD_1 \times VD_2 \times \ldots \times VD_n$ such that K is a key in T'.

Results

Let Σ be a set of strongly possible keys and θ a single strongly possible key over a relation schema R. Σ logically implies θ, denoted by $\Sigma \models \theta$ if for every instance T over R satisfying every strongly possible key in Σ we have that T satisfies θ.

Theorem 1 $\Sigma \models sp\langle K\rangle \iff \exists Y \subseteq K \ s.t. \ sp\langle Y\rangle \in \Sigma$.

Let us given schema $R = \{A_1, A_2, \ldots, A_n\}$ and let $\mathcal{K} = \{K_1, K_2, \ldots K_p\}$ be a collection of attribute sets and $T = \{t_1, t_2, \ldots, t_s\}$ be an instance with possible null occurrences. Our main question here is whether $\Sigma = \{sp\langle K_1\rangle, sp\langle K_2\rangle, \ldots, sp\langle K_p\rangle\}$ holds in T? Let $E_i = \{t' \in VD_1 \times VD_2 \times \ldots \times VD_n : t' \sim_w t_i\}$. Let $S \subseteq VD_1 \times VD_2 \times \ldots \times VD_n$ be the union $S = E_1 \cup E_2 \cup \ldots \cup E_s$ and define bipartite graph $G = (T, S; E)$ by $\{t, t'\} \in E \iff t \sim_w t'$ for $t \in T$ and $t' \in S$. Let $(S, \mathcal{M}_0)$ be the transversal matroid (see [5]) defined by G on S, that is a subset $X \subseteq S$ satisfies $X \in \mathcal{M}_0$ if X can be matched into T. Furthermore consider the partitions

$$S = S_1^j \cup S_2^j \cup \ldots \cup S_{p_j}^j \tag{6}$$

induced by K_j for $j = 1, 2, \ldots, p$ such that S_i^j's are maximal sets of tuples from S that agree on K_j. Let $(S, \mathcal{M}_j)$ be the partition matroid given by (6). We can formulate the following theorem.

Theorem 2 *Let T be an instance over schema $R = \{A_1, A_2, \ldots, A_n\}$ and let $\mathcal{K} = \{K_1, K_2, \ldots K_p\}$ be a collection of attribute sets.* $\Sigma =$

$\{sp\langle K_1\rangle, sp\langle K_2\rangle, \ldots, sp\langle K_p\rangle\}$ *holds in* T *if and only if the matroids* $(S, \mathcal{M}_j)$ *have a common independent set of size* $|T|$ *for* $j = 0, 1, \ldots p$

Unfortunately, Theorem 2 does not give good algorithm to decide the satisfaction of a system Σ of strongly possible keys, because as soon as Σ contains at least two constraints, then we would have to calculate the size of largest common independent set of at least three matroids, known to be a NP-complete problem [1].

In case of a single strongly possible key $sp\langle K\rangle$ constraint Theorem 2 requires to compute the largest common independent set of two matroids, which can be solved in polynomial time [4]. However, we can reduce the problem to the somewhat simpler problem of matchings in bipartite graphs.

If we want to decide whether $sp\langle K\rangle$ holds or not, we can forget about the attributes that are not in K since we need distinct values on K as a matching from $VD_{A_1} \times VD_{A_2} \times \ldots \times VD_{A_b}$ to $T = \{t_1, t_2 \ldots t_r\}|_K$ where $K = \{A_1, A_2 \ldots A_b\}$. Thus, we may construct a table T' that formed by finding all the possible combinations of the visible domains of $T|_K$ that are weakly similar to some tuple in $T|_K$.

$$T' = \{t' : \exists t \in T : t'[K] \sim_w t[K]\} \subseteq VD_1 \times VD_2 \times \ldots \times VD_b \quad (7)$$

Finding the matching between T and T' that covers all the tuples in T (if exist) will result in the set of tuples in T' that can be used to replace incomplete tuples in T so that K is a strongly possible key.

Let $c_v(A)$ denote the number of tuples that have value v in attribute A, that is $c_v(A) = |\{t \in T\colon t[A] = v\}|$. Some necessary conditions for a strongly possible key $sp\langle K\rangle$ are listed next.

Proposition 3 *Let* $K \subseteq R$ *be a set of attributes. If* $sp\langle K\rangle$ *holds, then*

1. *No two tuples* t_i, t_j *are strongly similar in* K.
2. $|T| \leq \prod_{\forall A \in K} |VD_A|$.
3. $\forall B \in K$, *number of nulls in* $B \leq \sum_{\forall v \in VD_B} \left(\frac{\prod_{\forall A \in K} |VD_A|}{|VD_B|} - c_v(B) \right)$.
4. *For all* $v \in VD_B$ *we have* $c_v(B) \leq \frac{\prod_{\forall A \in K} |VD_A|}{|VD_B|}$

Note that $sp\langle K\rangle$ holds if a matching covering T exists in the bipartite graph $G = (T, T'; E)$ defined as above, $\{t, t'\} \in E \iff t[K] \sim_w t[K]'$. We can apply Hall's Theorem to obtain

$$\forall X \subseteq T, \text{ we have } |N(X)| \geq |X| \text{ for } N(X) = \{t' \colon \exists t \in X \text{ such that } t' \sim_w t[K]\} \tag{8}$$

Approximation

To measure in what degree a set of attributes is a strongly possible key in a given dataset we use measure g_3 introduced in [2]. g_3 based on the idea that the degree to which a key is approximate is determined by the minimum number of tuples that need be removed from T so that K becomes a key. To find the tuples that we need to remove, we suggest to construct the maximum matching in graph $G = (T, T'; E)$.

$$g_3(K) = \frac{|T| - \nu(G)}{|T|} \tag{9}$$

where $\nu(G)$ denotes the maximum size of a matching in graph G.

Let $\mathcal{M}$ be the collection of connected components in the graph that satisfy the strongly possible key condition, i.e. there is a matching covering all T tuples in that set ($\forall_{M\in\mathcal{M}} \nexists X \subseteq M \cap T$ such that $|X| > N(X)$). Let $C \subseteq G$ be defined as $C = G \setminus \bigcup_{\forall M\in\mathcal{M}} M$ and let $\mathcal{M}'$ be the set of connected components of C. Furthermore, let V_M denote the set of vertices of T in a component M. So, the maximum matching can be written as $\sum_{M\in\mathcal{M}}(|V_M|) + \sum_{\forall M'\in\mathcal{M}'} \nu(M')$. Therefore we can reformulate the measure g_3 as:

$$g_3(K) = \frac{|T| - (\sum_{M\in\mathcal{M}}(|V_M|) + \sum_{M'\in\mathcal{M}'} \nu(M'))}{|T|} \tag{10}$$

Measuring the strongly possible keys approximation can be more appropriate by taking into consideration the effect of each connected component in the graph on the matching. More specifically, $\mathcal{M}$ represents the sets of tuples that do not require any tuple to be removed to get a strongly possible key, while the components of $\mathcal{M}'$ represent the sets of tuples that contain some tuples needed to be removed to have a strongly possible key. We consider the components of $\mathcal{M}$ to get their effect doubled in the approximation measure because they represent a

part of the data that is not effected by any tuples removal. So we propose a derived version of g_3 measure named g_3^c that consider the effects of these components.

$$g_3^c(K) = \frac{|T| - (\sum_{M \in \mathcal{M}}(|V_M|) + \sum_{M' \in \mathcal{M}'} \nu(M'))}{|T| + \sum_{M \in \mathcal{M}} |V_M|} \tag{11}$$

Theorem 4 *For any table T and set of attributes K we have either $g_3(K) = g_3^c(K)$ or $1 < g_3(K)/g_3^c(K) < 2$. Furthermore, for any rational number $1 \leq \frac{p}{q} < 2$ there exists tables of arbitrarily large number of tuples with $g_3(K)/g_3^c(K) = \frac{p}{q}$.*

Summary

The key selection problem is an important task in relational databases. Though mostly it is a human activity at design time, there are emerging needs for its automation. The strongly matching key concept is a new addition to this field, and makes a step forward to the applicability since it is based only on the currently available information contained in relational tables. The other advantage of this paper is that not only concepts and algorithms are defined but a new quality measure is introduced.

References

[1] Garey M.R., Johnson D.S.: *Computers and Intractability. A Guide to the Theory of NP-Completeness*. Freeman, New York (1979)

[2] J. Kivinen, H. Mannila: *Approximate inference of functional dependencies from relations* Theoretical Computer Science, Volume 149, Issue 1, PP. 129-149, 1995.

[3] H. Köhler, U. Leck, S Link and X. Zhou: Possible and certain keys for SQL, The VLDB Journal. The VLDB Journal, Volume 25, Issue 4, pp. 571-596, 2016.

[4] E. L. Lawler, *Matroid intersection algorithms*, Mathematical Programming, **9**, pp. 31-56, 1975.

[5] D. J. A. Welsh: Matroid Theory, New York, Academic Press, 1976.

Data Linking with String Matching

F. Kruzslicz[1], M. Hornyák[2]
[1]University of Pécs, Faculty of Economics, kruzslic@ktk.pte.hu
80 Rákóczi Street, H-7620 Pécs, Hungary
[2]University of Pécs, Faculty of Economics, hornyakm@ktk.pte.hu
80 Rákóczi Street, H-7620 Pécs, Hungary

Abstract: **Data matching is a special data integration task where same or similar entities are identified and linked together in disparate databases. Since references to identifiers from external data sources may be ambiguous, there is a strong need for efficient techniques of semi automated data linking, especially in case of schema-less data models. Because of its high complexity scalability is important in Big Data environments. In this paper we demonstrate a text based data matching application case using the general framework of data mapping.**

Introduction

Real world objects usually have different data representations in various databases. In order to properly connect these occurrences together first a schema matching is required to find possible (group of) attributes suitable to decide the similarity between the candidate records. In simple cases entity details like names, dates, addresses are available as natural identifiers for matching. Lack of unique identifiers requires more elaborated processes able to match entities with highly dissimilar identifiers using their profile description or activities. Entity deduplication is a special case of data matching, where only one data source is used instead different ones.

In order to develop efficient data matching methods at least the attributes used for linking must have a structured format. These structural pre-processing steps, such as OCR (Optical Character Recognition) of printed documents and handwriting, or NER (Named Entity Extraction) recognition in texts [6], are not considered as a core part of data matching. Its most important application area can be found in crime prevention, publication databases, e-commerce, social networks and health sector [8]. In general some kind of data linking step is included in almost all data mining projects to maintain referential integrity. Especially CRM (Customer Relationship Management) deals with data from different sources to enrich and facilitate analytic methods. Record matching and unification are standard phases of

data cleaning steps of ETL (Extraction, Transformation and Loading) processes in data warehouses.

Time complexity of data matching is quadratical since potentially all records of a data set should be compared to all records of the other's ones. When at least one of the data sets is duplication free, the number of comparisons of possible true matches can be limited by efficient indexing [2].

Methodology

There are two major different data matching problem types according to the desirable outcome of the process: one-to-one or one-to-many mappings of records. The latest problem group is known as statistical data matching as well, because records are linked to a group of records based on similarity, which is very close to clustering techniques [5]. In this paper only the one-to-one type of matching is studied.

Every record pairing project demands special domain knowledge and deep understanding of data and there are no „one for all" solutions. In spite of this non-uniform property a data matching process consists of the following six phases [4]:

1. *data pre-processing* – to assure data from all sources have the same format
2. *indexing* – to reduce the quadratic complexity by effective pair candidate generation
3. *pairwise comparisons* – to determine the similarity level of record pairs
4. *classification* – of pair similarities into match, non-match, potential match categories
5. *manual review* – of classification enrollment of record pairs
6. *evaluation* – to determine the quality and completeness of the results

Though each phase is equally important, some phases may be ignored. Indexing is useful mainly in case of large data sets. The result of classification phase is worth to be formalized by models or sets of rules for later reusability. In the evaluation phase matching quality defines how many of the matched records correspond to true real-world cases. On the contrary a completeness measure is ratio of correctly matched real-world cases of the data sets [4].

In our following examples we shall focus on the pairwise comparison phase, and demonstrate how string matching is used to facilitate this step.

Named entities

There are numerous distance metrics and similarity measures defined on character strings suitable to determine the matching level of texts. Popular ones, like *Hamming, Jaccard, Levenshtein* and *LCS (Longest Common Subsequence)* edit distances evaluate the written form of words by determining the transformation from one into the other having minimal steps or costs [5]. If spoken form of words must be compared according to their pronunciation, phonetic encodings functions, like *Soundex* or *NIIS (New York State Identification and Intelligence System)* are used first. Similarity functions may be applied for standalone letters or n-grams (contiguous sequence of *n* characters) of words [3].

Names of entities (persons, organizations, geographic locations etc.) are usually written with upper case initial letter in many languages. Named entities have other known properties as well, which make the problem a bit more complex, but allow us to develop more specific similarity measures. The following company name examples are cited from the uncleaned data set of the Corruption Research Center Budapest (www.crcb.eu) [2]:

- *Spaces and non alphanumeric characters:* Generali-Providencia, T-INVEST '91, Johnson & Johnson
- *Missing components:* Sodexo (Pass) (Hungary)
- *Spelling differences:* PC-Boksz vs PC-Box
- *Abbreviations:* Mafilm, MEDTECH, Filo Ker.
- *Titles and forms:* Limited, Kft., Kht., Rt. Zrt.
- *Out of order component:* Hotel Veszprém Margaréta vs. Margaréta Hotel Veszprém
- *Nicknames:* OTP, Közgép, Fürge Diák, J&J
- *Multiple languages:* Phoenix Contact vs. Főnix Kontakt
- *Truncated components and initials:* IDOM (2000), M. és T. Kft.
- *Similar names:* Agro-Data Kft., Agro Alba Zrt., Agro Alfa Kft.

Company and person names both consist of more than one word, and less error can be found at the beginning and at the end of words than in the middle. It means that sequence similarity measures are more appropriate in this case than set-based ones. One of the best similarity measure capturing this property is the *Jaro-Winkler* formula which put more weights on matching of either end of words [9].

sim_{jw}('Agro Alba Zrt.', 'Agro Alfa Kft.') = 0.9142857193946838
sim_{jw}('Agro Alfa Kft.', 'Agro-Data Kft.') = 0.845714271068573
sim_{jw}('Agro Alba Zrt.', 'Agro-Data Kft.') = 0.8035714030265808

Application

Companies are registered by regulatory authorities in Hungary and country wide unique numerical identifiers are provided for them. But in surveys, web crawling and other data collections this information is rarely available. Not like company names, which may have different forms, and even might contain typos as well. In our company competitiveness research a cleaned data set of financial indicators from enterprise information data source of OPTEN (www.opten.hu) had to be enriched by their previously mentioned data set of public procurement activities published by CRCB. OPTEN data selection contains both official short and long names of 1 028 companies, while CRCB consists of 239 483 tender announcement result records, mentioning one or more winners in various unofficial formats.

Table 1: Short and full company name examples from the OPTEN data set

AGRO-DATA Bt.	AGRO-DATA Mezőgazdasági Szaktanácsadási Fejlesztési és Informatikai Betéti Társaság (*cancelled*)
AGRO-DATA Kft.	AGRO-DATA Mezőgazdasági Szaktanácsadási Fejlesztési és Informatikai Korlátolt Felelősségű Társaság

Table 2: Company name reference examples from the CRCB data set

AGRO - DATA Mezogazdasagi
AGRO-DATA Mezogazdasagi Szaktanacsado Fejlesztesi es Informatikai Kft.
AGRO-DATA Kft.
AGRO DATA Korlatolt Felelossegu tarsasag

Because tender data has plain ASCII encoding, even Hungarian words are stored without any accents, therefore fuzzy matching of text identifiers were required. Before pairwise comparison some text mining pre-processing was necessary in order to replace ineffectual long words with little additional information. After converting strings to all lowercase, association mining was applied to find the most frequent word n-grams in CRCB corpus to build a replacement dictionary to support better comparison.

Table 3: Extract from the domain specific n-gram stopword replacement dictionary

kereskedelmi es szolgaltato korlatolt felelossegu tarsasag	kskft
zartkoruen mukodo reszvenytarsasag	zmrt
epitoipari kereskedelmi es szolgaltato kft	ekskft
magyarorszag	mo

After splitting the winners attribute, which contain multiple winners, we got 63 378 different potential name references, which was reduced to 42 770 candidates by using the replacement dictionary. This number was further decreased to 36 522 after a deduplication process, where all names having higher Jaro-Winkler similarity than a threshold value 0.95 were merged into their most frequently used name form.

Figure 1 shows the quality improvement of the CRCB data set in terms of resolved company name links. In the manual review phase the number of false matching was negligible, and the whole process was much less time consuming and expensive. This kind of threshold based similarity linking performed worst on short (*hapax*) names that occur only once.

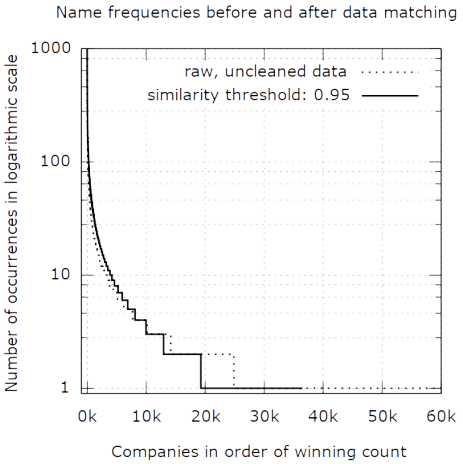


Figure 1: Quality improvement with fuzzy data linking

Though the top two most frequent winners remained the same, the data cleaning significantly changed even the rest of the top rankings. See winning counts before and after in Table 4.

Table 4: Top list of tender winner companies in Hungary between years 2015 and 2017

Rank	Company name	Count before	Count after
1.	Magyar Aszfalt Kft.	818	1040
2.	Johnson & Johnson Kft.	597	828
3.	Swietelsky Magyarország Kft.	285	743
4.	Immofix Kft.	444	675
5.	Euromedic-Pharma Zrt.	298	639

The same name pre-processing conversions were applied for the short and long company names in the OPTEN database as well to facilitate joining records. Because of company names (in short and long form) in OPTEN are officially validated, no further manual corrections were needed to apply for the deduplicated CRCB data set at this moment. Manual revision could be delayed after the record matching step, when more CRCB candidates were assigned to the same entity in OPTEN.

In the pairwise comparison phase first exact matches were identified and removed to speed up further processing. Next five different similarity functions were used to find potentially linkable candidates, which were proved to be most useful for name-matching purposes in other cases [1]. which was used in the deduplication phase earlier. Pair candidates were classified according to the normalized results of *Jaccard, Levenshtein, Jaro, Monge-Elkan, Smith-Waterman* and *Jaro-Winkler* string similarities.

A record pair was defined as a *match*, if its all similarities were 1. If not all, but least one of the similarities was equal to 1, then the pair was categorized as a *possible match*, otherwise it was taken as an *unmatch*. These category enrollments were supervised and resolved manually in questionable cases to create a benchmark data set for performance measurements. Optimal threshold splitting values of similarities were determined by using this manually corrected benchmark data set.

Table 5: Threshold based similarity classification results for optimal cutting points

Similarity function	*optimal threshold*	true match	false match	true unmatch	false unmatch	**accuracy**	**F_1** for match
Smith-Waterman	*0.23*	145	346	460	47	**60.62%**	**0.4245**
Jaccard	*0.96*	124	181	625	68	**75.05%**	**0.4989**
Monge-Elkan	*0.98*	132	26	780	60	**91.38%**	**0.7542**
Levenshtein	*0.87*	131	22	784	61	**91.68%**	**0.7594**
Jaro	*0.95*	126	9	797	66	**92.48%**	**0.7706**
Jaro-Winkler	*0.96*	135	12	794	57	**93.08%**	**0.7964**

If the number of 104 exact matchings of the pre-processed data sets is compared to the number of true match cases in Table 5, the improvements are not remarkable. Jaro-Winkler similarity gave the best results in terms of accuracy and F-measure as well. Since our aim was to enrich the OPTEN data set, that is why correct matches are more important than correct unmatches. As the intersection of the two data set contains relatively small number of records compared to the one of their set difference, the F_1 measure of positive class of matches is a better indicator in case of such an unbalanced classification data set.

Among the elements of false unmatch cases the lowest Jaro-Winkler similarity was 0.894. The optimal threshold value of 0.96 as an inter-database fuzzy match acceptance level is very close to the 0.95 threshold value which was applied as a criteria in the deduplication phase. Due to high recall (70.31%) and precision (91.84%) of matches the model has an acceptable good $F_1 = 0.7964$ efficiency indicator value.

Conclusion

In this project, we demonstrated how data linking method can help to improve data quality. Approximate string matching tools were used to find company entities referred by their several name variations. After data cleaning we could get more appropriate ranking and standardized identifiers suitable for enrichment of other company data set. In case of company names Jaro and Jaro-Winkler string similarities were found to be a good solution for both data linking and deduplication at about 0.95 similarity

threshold level. Though a careful selection of the best name similarity function is important, a simple nominal distance might give very reasonable results, if the data sets are properly pre-processed. Preprocessing was the crucial step in our presented method as well, which can be improved by enhancing the domain specific n-gram stopword replacement dictionary. A possible further development of our method is using different similarity measures depending on the length of names, because Jaro-Winkler similarity was found to be less efficient for short texts. Finally it is recommended to use additional attributes to just names (like addresses, fields of economic activities, owners etc.) for a more efficient manual and automatic disambiguation of companies.

Acknowledgment

The research was financed by the Higher Education Institutional Excellence Programme of the Ministry of Human Capacities in Hungary, within the framework of the 4th thematic programme "Enhancing the Role of Domestic Companies in the Reindustrialization of Hungary" of the University of Pécs (reference number of the contract: 20765-3/2018/ FEKUTSTRAT) The authors would like to thank the reviewers for their helpful and valuable advice on various technical issues, corrections and ideas for improvements.

References

[1] W. W. Cohen, P. Ravikumar, S. E. Fienberg, "A Comparison of String Distance Metrics for Name-Matching Tasks," Proceedings of the 2003 International Conference on Information Integration on the Web — Acapulco, Mexico, pp. 73-78, 2003

[2] CRCB.2018, "Hungarian Public Procurement Dataset 2005-2017," Data Publication for EU Citizens – 1/2018, Corruption Research Center Budapest, 2018

[3] A. Doan, A. Halevy, Z. Ives, "Principles of Data Integration," Morgan Kaufmann Publishers, pp. 520, 2012

[4] H. Köpcke, E. Rahm, "Frameworks for entity matching: A comparison," Data and Knowledge Engineering, 69(2), pp. 197–210, 2010

[5] F. Kruzslicz, "Improved greedy algorithm for computing approximate median strings," Acta Cybernetica, 14: pp. 331–339, 1999

[6] D. Nadeau, S. Sekine, "A Survey of Named Entity Recognition and Classification." Lingvisticae Investigationes, 30, pp. 3–26. 2007

[7] Cs. I. Sidló, "Entity Resolution with Heavy Indexing", Proc. 2011 Int. Conf.on Advances in Databases and Information Systems, CEUR Workshop Proceedings, 2011.

[8] X. Wang, J. Ling, "Multiple valued logic approach for matching patient records in multiple databases," Journal of Biomedical Informatics, 45 pp. 224–230, 2012

[9] W. E. Yancey, "Evaluating string comparator performance for record linkage," Tech. Rep. RR2005/05, US Bureau of the Census, 2005

Undersampled On-Off Keying Camera Communication Methods for Beacon ID Transmission

M. Rátosi[1], G. Simon[2]
[1] University of Pannonia, Department of Computer Science and Systems Technology, ratosi@dcs.uni-pannon.hu
8200 Veszprém, Egyetem utca 10., Hungary
[2] Pázmány Péter Catholic University, Institute for Process Systems Engineering and Sustainability, simon.gyula@ppke.hu
1088 Budapest, Szentkirályi utca 28., Hungary

Abstract: **Visible Light Communication techniques allow the utilization of LED lighting infrastructure for data transmission. Such systems use modulated light sources, the flicker of which is not visible for the human eye. When ordinary cameras are used as receivers, the low framerate of the cameras, with respect to the blinking frequency of the LEDs, results in sub-Nyquist sampling. This paper reviews and compares recent undersampled on-off keying camera communication methods.**

Introduction

Modern LED light sources provide new possibilities for the utilization of the lighting infrastructure for communication purposes as well. In such Visible Light Communication (VLC) systems the LEDs are blinking to transmit the message in such a way that the receiver is able to decode the message content, but the blinking is not visible for human observers. In order the flicker to be not visible for the human eye the modulation frequency must be significantly higher than 100 Hz. Perception related issues are important in real applications, but it is not in the scope of this paper.

On the receiver side various sensors can be used: for high bandwidth communication photodiodes can be used as sensors, which can be operated even in the MHz range. Since inexpensive cameras are present in our everyday life the utilization of such cameras for sensing is an appealing solution. These devices, however, usually operate with low sampling frequencies, providing 30-120 frames per second. Thus the utilization of cameras as receivers results in undersampling, which necessitates the design of compatible undersampled communication protocols. Such protocols are

able to provide a few bits per second bandwidth, which is ideal for the transmission of the beacon identifiers, used e.g. in localization applications.

In this paper we focus on on-off keying protocols used with global shutter cameras, which can provide robust operation for long distances even when the beacon is barely visible. We also investigate the appropriateness of these methods for cases with moving cameras.

Undersampled Camera Communication

The studied methods include Undersampled Frequency Shift On-Off Keying (UFSOOK) [1], Undersampled Phase Shift On-Off Keying (UPSOOK) [2], and our proposed Trackable Undersampled Phase Shift On-Off Keying (TUPSOOK) [3].

Let us use f_{cam} to denote the sampling frequency of the camera. The operation of the methods is illustrated in Fig. 1.

UFSOOK

The Undersampled Frequency Shift On-Off Keying protocol utilizes three blinking frequencies: the header frequency is chosen to be high enough so that even with the smallest exposure time the blinking is undetectable by the camera, thus headers are sensed as half intensity signals. In practice $f_{header} = 15 - 20\ kHz$ is used. For data coding two lower frequencies are utilized, the frequencies of which are determined according to the following rules:

- $f_1 = f_{cam} * n$
- $f_2 = f_{cam} * (n - 0.5)$

For example, in case of $f_{cam} = 30\ Hz$ and $n = 4$ the following frequencies are given: $f_1 = 120\ Hz$, and $f_2 = 105\ Hz$.

With such frequency selection the camera, operating with constant f_{cam}, samples the signal with f_1 frequency always at the same phase, and thus observes the light source as a constant magnitude signal, i.e. the LED is always shown on the camera image as a full intensity signal or a dark signal.

The other signal with frequency f_2, however, is sampled at alternate phase values, that is the LED is shown as a full intensity signal in the first frame and as a dark signal on the next frame.

The coding of the data is performed using the above properties:

- Logical 0 is coded as a signal with frequency f_1 with length of two frames, thus the camera observes two consecutive dark or two consecutive full intensity values.
- Logical 1 is coded as a signal with frequency f_2 with length of two frames, thus the camera observes two opposite value levels in a row.

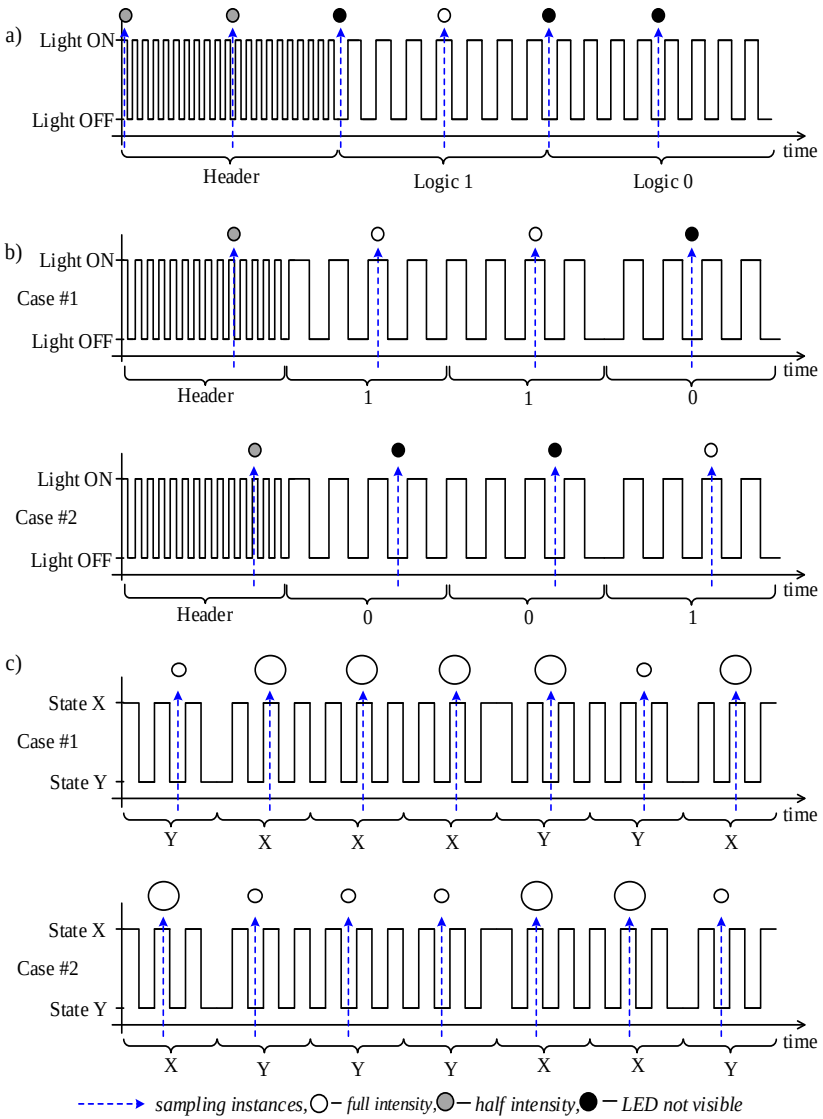


Figure 1: The operation of the studied undersampled communication protocols. a) UFSOOK, b) UPSOOK, c) TUPSOOK. Camera sampling instances are shown by dashed arrows; observed beacon images are illustrated above the sampling instances.

- The protocol applies a framing header with frequency f_{frame} before the data bits. The decoder algorithm observes a half intensity signal marking the beginning of the data packet.

The operation of the protocol is illustrated in Figure 1.a. The packet is preceded by the header, which is sampled as two half intensity values. The first bit is a dark-light pair, which is decoded as a logical 1. The second bit is a dark-dark sequence, which is decoded as a logical 0.

Notice, that the UFSOOK protocol can transmit half bits per frame, apart from the overhead caused by the header.

UPSOOK

The Undersampled Phase Shift On-Off Keying protocol utilizes two blinking frequencies: the header frequency is the same as of the UFSOOK protocol. For data coding one lower frequency is utilized, the frequency of which is determined as $f_1 = f_{cam} * n$.

Here the data encoding is performed using phase modulation, as follows: logical 1 is encoded by a signal with phase 0 and length of one frame. Logical 0 is encoded by a signal with phase 180 and length of one frame. Since the camera sampling is usually not synchronized with the transmitter, the receiver starts sampling at a random phase, producing a phase uncertainty problem. Thus, the protocol contains a mark symbol after the header, which is used to determine the actual sampling phase. The operation is illustrated in Figure 1.b, showing two possible outcomes, due to the phase uncertainty:

- In case 1, the header is sampled as a half intensity value, followed by a full intensity mark symbol. The data bits are detected as light and dark values, which are, using the mark symbol, decoded as logical 1 and 0.
- In case 2, the header is sampled as a half intensity value, followed by a dark mark symbol. The data bits are detected as dark and light values, which are, using the mark symbol, decoded again as logical 1 and 0.

Notice, that the UPSOOK protocol can transmit one bit per frame, apart from the overhead caused by the header and the mark symbol.

TUPSOOK

The Trackable Undersampled Phase Shift On-Off Keying protocol utilizes one blinking frequency only, since there is no dedicated header symbol. For data coding one frequency is utilized with $f_1 = f_{cam} * n$.

The TUPSOOK protocol utilizes a special beacon design as shown in Figure 2. The beacon contains a small central LED and a large outer ring LED. At any time instant one and only one of the LEDs is on.

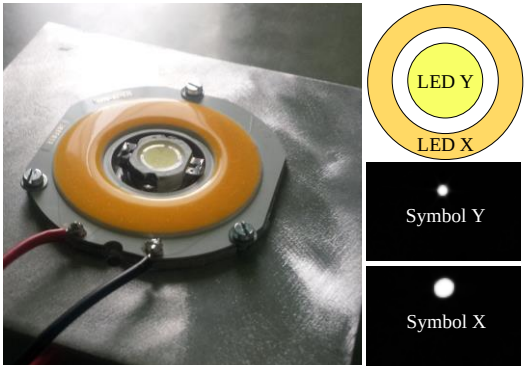


Figure 2: The beacon design and the observed symbols of the TUPSOOK protocol.

Let us define two symbols:

- Symbol X: the two LEDs are blinking alternately with frequency f_1 and phase 0, with length of one frame.
- Symbol Y: the two LEDs are blinking alternately with frequency f_1 and phase 180, with length of one frame.

The physical header is replaced by a logical header, containing 5 consecutive symbols as follows: Y-X-X-X-Y.

The encoding of the data bits is the following:

- Logical 0 is encoded by a Y-X symbol pair (2 frames).
- Logical 1 is encoded by an X-Y symbol pair (2 frames).

The operation of TUPSOOK is illustrated in Figure 1.c. Here, because of the phase uncertainty, two cases can happen:

- Case 1: The logical header is sensed as Y-X-X-X-Y, the following bit is Y-X. Since the header is not inverted, Y-X means a logical 0.
- Case 2: The logical header is sensed as X-Y-Y-Y-X, the following bit is X-Y. Since the header is inverted, X-Y is a logical 0.

Notices:

- The logical header is defined for a symbol pattern, which cannot occur among the data bits. Thus the detection of the header is straightforward: either the sequence Y-X-X-X-Y or the sequence X-Y-Y-Y-X means a valid header.
- The TUPSOOK protocol can transmit half bits per frame, apart from the overhead caused by the software header.
- This protocol uses the beacon size instead of its brightness level for coding/decoding.

- The redundant bit coding allows error detection: if the symbol pair, representing a bit, consists of similar values then it is certainly an erroneous detection.

Summary

The performance properties of the discussed protocols are summarized in Table 1. UFSOOK and UPSOOK use physical framing, thus the detection involves the determination of half light intensity. This feature is reported to cause detection errors from long distances. TUPSOOK avoids the utilization of physical framing at a price of a somewhat longer logical frame. The fastest protocol is UPSOOK with 1 bit per frame, UFSOOK and TUPSOOK produce ½ bits per frame. The overhead is the smallest of UFSOOK (1 frame), while UPSOOK requires 2 frames and TUPSOOK 5 frames, per data packet. Only TUPSOOK provides good trackability, since the image of the beacon is visible in every frame. UFSOOK and UPSOOK, however, depending on the actual data, may have long sequences of frames with dark beacon images. TUPSOOK also contains some error detection features.

Table 1: Properties of the communication protocols.

	UFSOOK	**UPSOOK**	**TUPSOOK**
Physical framing	f_{frame}	f_{frame}	-
Logical framing	no	no	yes
Frequencies used	3	2	1
Bit rate per frame	1/2	1	1/2
Minimum frame length (frames)	1	1+1	5
Trackability	-	-	good
Error detection	no	no	yes

Acknowledgment

We acknowledge the financial support of Széchenyi 2020 program under the project No EFOP-3.6.1-16-2016-00015.

References

[1] R. D. Roberts, "Undersampled frequency shift ON-OFF keying (UFSOOK) for camera communications (CamCom)," 2013 22nd Wireless and Optical Communication Conference, Chongqing, 2013, pp. 645-648, doi: 10.1109/WOCC.2013.6676454

[2] P. Luo, Z. Ghassemlooy, H. Le Minh, X. Tang and H. M. Tsai, "Undersampled phase shift ON-OFF keying for camera communication," 2014 Sixth International Conference on Wireless Communications and Signal Processing (WCSP), Hefei, 2014, pp. 1-6.

[3] M. Rátosi, G. Simon, "Trackable Visible Light Beaconing and Detection for Indoor Localization Applications Using Undersampling Cameras", 2018 International Conference on Indoor positioning and Indoor Navigation (IPIN), Nantes, 2018, pp. 1-4., Paper: 211108

Validation of a custom human centric luminaire design based on on-site experiments

D.N. Tóth[1], F. Szabó[1]
[1] Light and Colour Science Research Laboratory, University of Pannonia, Egyetem u. 10., Veszprém, 8200 Hungary

Abstract: **The aim of human centric lighting (HCL) is to improve the overall comfort of the users and to support the proper function of their circadian system while having the benefits expected from modern light sources. The absence of natural light can lead to serious health risks due to the disruption of the circadian rhythm. With the introduction of HCL to workplaces, these risks could be reduced. At the Light and Colour Science Research Laboratory of the University of Pannonia in Veszprém a lighting solution had been developed for this aim. The concept had been validated via a case study at a manufacturing site with the participation of the employees. Participants were wearing heartrate and activity monitoring smartwatches during work. The experiments had two phases: one before and one after the introduction of HCL. The measurement results had been analyzed using statistical methods to prove the effects of the lighting on the participants.**

Introduction

During millions of years, life on Earth adapted to the repeating cycle of day and night. In the beginning this daily cycle defined the daily routine. The continuous development of artificial lighting – beginning with the discovery of fire – led to efficient artificial light sources which made working in night shifts feasible in the last century. Many recent studies aimed to uncover the risks of circadian disruption. As possible short-term consequences sleep disorders and insomnia were reported in these papers [1, 2, 3, 4]. Long-term consequences could be far more serious: depression or even cancerous diseases [1]. Human centric lighting (HCL) had been proposed as a possible way to reduce these health risks. It offers a possible solution providing optimized lighting for the human body by taking interdisciplinary researches into account. Modern LED light sources have optimal properties to be used in the realization of such optimized light: spectrally tunable light sources usually have phosphor white LEDs combined with narrow band color LEDs which usually can be controlled in groups called LED channels. This technique – depending on the characteristics of the LED channels – makes possible to optimize light for different objectives. In the implementation of

human centric lighting an intelligent controller is required for automatic, real-time control of the lighting without the need of user interaction.

The circadian system

An important aspect in the design of an HCL solution is to implement the support of proper function of the human circadian system – which controls the sleep cycle in humans. The "central clock" of the circadian system is found in the brain – in the hypothalamus – and is called the suprachiasmatic nucleus (SCN). It regulates the circadian rhythm through hormonal mechanisms. It has an intrinsic period of approximately 24 hours, and is synchronized to the light-dark periods of the day. Abrupt desynchronization of the "real" light-dark period and the light-dark period expected by the circadian system – for example a long flight through multiple time zones – causes short-term negative effects like fatigue and disorientation.

The light or dark periods of the day are detected by the intrinsically photosensitive retinal ganglion cells (ipRGC) which contain melanopsin photopigment [5]. The neural signal of these cells transfers directly into the SCN which is in neural connection with the pineal gland. Through the pineal gland the SCN controls the level of melatonin and cortisol hormones in the bloodstream based on the signal of ipRGCs. Melatonin lengthens reaction time, and prepares the body for sleep while cortisol has the exact opposite effects: induces a high alertness state and shortens reaction time.

The absorption maximum of melanopsin is between 460 nm and 480 nm causing ipRGCs to have a sensitivity maximum in this wavelength domain [6, 7, 8]. To mix optimized light to achieve specific effects on the circadian system a spectrally tunable light source is needed which contains LEDs radiating in this wavelength interval. During the design of a "casual" spectrally tunable light source this should have to be taken into account to avoid possible negative impacts on the users [1].

Calculating the circadian effect of light

To evaluate the effect of light on the circadian system a well-defined unit is needed. Circadian light (CL) can be used for this purpose [9]. It represents the circadian system's response to a retinal illumination of certain intensity and spectral composition with one number. It is a spectrally weighted variant of irradiation (W/m^2). To ease the comparison between different spectra, a special variant is also defined – denoted as CL_A [9] – which proportions the circadian effect of a specific spectral power distribution to the circadian effect

of an illuminance of 1000 lx generated by an incandescent lamp (CIE A illuminant).

Common artificial light sources maintain their spectral composition during their operation causing the circadian effects to be constant as well. Comparing the CL_A values in case of 1000 lx of retinal illuminance the incandescent lamp has a CL_A value of exactly 1000, fluorescent tubes have around 600-800 and LEDs have CL_A values varying depending on the spectral distribution.

The sunlight reaching the Earth's surface however continuously changes due to Earth's rotation and different atmospheric conditions. For this reason, calculating the CL_A values of sunlight measured in different conditions or in different parts of the day results in different values [10].

An important design aspect in human centric lighting is to reproduce the phenomenon of continuous change of the spectral composition, intensity and circadian effect similarly to the natural daylight.

Experimental setup: Industrial HCL luminaires

Preceding the case study, a custom industrial human centric lighting solution had been developed. The result was a spectrally tunable luminaire and an intelligent control unit which can provide continuous transitions between specific "lighting cornerstones" over a long period of time. Due to this technique users cannot notice the continuous, minimal changes in the lighting. The blue LEDs had been chosen to have their peak wavelength in the maximum sensitivity range of the circadian system. The luminaire has other channels composed of red, green and warm white phosphor LEDs in order to provide dynamically mixed white light with good color quality.

A series of the prototype luminaire had been manufactured and installed in a windowless manufacturing site to be validated by a series of measurements. The final settings on the controller had been set on site to be perfectly adapted to the environment: the pre-set "lighting cornerstones" had been fine-tuned in the final experimental environment.

Table 1: Photometric parameters of the light settings on the intelligent luminaire compared to daylight and the old system.

Name of the light setting or light source	**CCT [K]**	**d_{uv}**	**CRI R_a [11]**	**IES TM-30-15 [12]**		**CL_A**
				R_f	**R_g**	
Daylight (CIE D65)	6504	0.0032	100	100	100	2036
"Stimulating light" setting	6510	0.001	90	85	98.6	1841
"Neutral light" setting	4200	-0.0016	94	89.6	99.5	1030
"Relaxing light" setting	4000	0.0001	93	88.7	96.4	826
The old LED tubes	4442	-0.005	86	82	96.1	1055

Two of the specified “lighting cornerstones” had been optimized to have minimal (“relaxing light”) and maximal (“stimulating light”) circadian effects and the third one as an intermediate point. Compared to the original static LED lighting the CL_A of the new system varies between 20% lower and 75% higher values. In case of the “stimulating light” setting the CL_A of the provided light is only 9% less than the midday daylight (CIE D65).

During the four months of the experiments, a total of 16 workers participated (11 female, 5 male), ranging from 23 to 55 years of age (avg. 39). The participants had been working both in a morning and the afternoon shift. During the morning shift the light setting changed from the “stimulating” setting to the “neutral” setting and during the afternoon shift it changed from the “stimulating” to the “relaxing” setting. The time and duration of the transitions had been chosen in a way to be able to provide enough stimulation for the circadian system to make it possible to do work with high performance as well as to simulate the natural change of the circadian effect of daylight to avert the negative health impacts of the absence of natural light.

Heartrate monitoring and evaluation

To evaluate the effects of lighting on the workers recording of their heartrate had been chosen since it can be easily recorded and is affected by the circadian system [2]. The experiments were carried out in two phases; the first under the old industrial LED tube lighting and the second under the new intelligent HCL. For this purpose, intelligent sport watches had been used with accurate chest band type heartrate sensors. These instruments could be used during both shifts to record participants’ heartrate with a resolution of 1 second.

The first stage of the experiment had two purposes. It served to determine if different conditions, habits or traits of the participants have significant effect on the results and to serve as a base for the later comparison to the second stage measurement results. The age, weight and height had been initially recorded for all participants. At the end of every measurement session their coffee intake was also recorded. At the end of this stage approximately 600 hours of heartrate data had been recorded. Initially it could be concluded that during the afternoon shift mean pulse was higher (91.8 while during the morning shift it was 84.7) and the standard deviation was higher in the afternoon (10.6 during the afternoon shift and 6.4 during morning) too. To analyze connections between the heartrate and the participants’ other traits statistical methods had been used. Since the recorded

parameters are objective measurement data which are considered high measurement level variables in the statistical analysis, Pearson correlation [13] had to be used. The results of the correlation analysis did not confirm correlation between the heartrate data of the two shifts and any of the other variables.

The purpose of the experiments' second phase was to compare the effects of the new intelligent luminaires to the previously recorded data. Since the CL_A of the light was known during the working hours the correlation between this parameter and the heartrate data could be analyzed. During the evaluation the approximately 2-hour delay in the circadian effect [3, 4] had been taken into account.

The delay caused the effects of the morning shift's lighting transition to fall outside the time interval of the heartrate monitoring thus preventing its evaluation using correlation analysis. Contrary to this in case of the afternoon shift data could be evaluated and a significant moderate positive correlation had been found (Pearson correlation coefficient: 0.4). The explanation for the moderate strength of the correlation is that heartrate is affected by many other factors besides the circadian effect. For example, other activities during the break at the middle of the afternoon shift increased the following heartrate.

Changes in the effect of lighting could be tested by calculating correlations between the heartrate data recorded under the old and new luminaires. Since during the morning shift's heartrate recording in both cases the measured effects were belonging to a constant circadian effect it was expected for the correlation between those to be stronger than in case of the afternoon shift. In this case the Pearson correlation coefficient was 0.75 (strong positive correlation) and in case of the afternoon shift data it was 0.41 (moderate positive correlation) with both correlations being significant.

To confirm if the results of the new measurement are statistically different from the previous experiment under the old lighting, paired-samples t-test [13] had been used. In case of both the morning and the afternoon data this test confirmed that the differences between the data recorded in the two phases of the experiments were statistically significant ($p < 0.01$ in both cases).

The results confirm that replacing the static LED lighting with the HCL luminaires indeed had an effect on the heartrate of the workers. The statistical significance levels of the results of the analyses indicate that the changes in the data are not caused outliers but a significant change. The moderate positive correlation between the heartrate measurements and the changing CL_A of the lighting suggests that the originally intended effects on the

circadian function through lighting had been achieved. However, this moderate correlation also implies that the other environmental parameters also have a significant effect on the heartrate of the workers. In further research these environmental parameters should be recorded for better approximation of the whole model of the environment affecting the heartrate.

Conclusions

Human centric lighting aims to improve the life quality of the users and to provide a healthy environment. At first human centric lighting had been adopted by hospitals and schools but it is expected to be used at homes in the future as well. Preceding the case study presented in this paper a custom human centric industrial luminaire had been developed to improve the quality of lighting in a windowless manufacture building and to affect the circadian system of the workers in a similar way to natural sunlight. Prototype luminaires had been installed on site and a two-stage experiment had been conducted to investigate the effects of light on the workers. In the first half of the study the heartrates of the workers had been recorded by smartwatches with chest-band heartrate sensors under the old lighting system. After the installation of the new intelligent luminaires the measurements had been repeated. The changes in the effect of lighting on the participants had been evaluated through changes in the measured heartrate data. Correlation analysis on the old-new heartrate pairs and on heartrate paired with its circadian effect measure confirmed a significant moderate connection – with a correlation coefficient of 0.4 – implying that the replacement of the LED tube lighting with the experimental luminaires indeed has the previously expected effect.

Acknowledgment

We acknowledge the financial support of Széchenyi 2020 programme under the project No EFOP-3.6.1-16-2016-00015.

References

[1] Chang A-M, Aeschbach D, Duffy J F, Czeisler C A. “Evening use of light emitting eReaders negatively affects sleep, circadian timing, and next morning alertness”, Proceedings of the National Academy of Sciences, 112:1232–1237., 2015

[2] Rüger M, Scheer F A. “Effects of circadian disruption on the cardiometabolic system”, Reviews in endocrine & metabolic disorders, 10(4):245-260, 2009

[3] Figueiro M, Overington D. “Self-luminous devices and melatonin suppression in adolescent”, Lighting Rese-arch and Technology, 2015 0 1-10

[4] Lack L, Wright H. “The effect of evening bright light in delaying the circadian rhythms and lengthening the sleep of early morning awakening insomniacs”, Sleep, 16(5):436-43, August 1993

[5] Berson D M. “Strange vision: ganglion cells as circadian photoreceptors”, Trends in Neurosciences, 26 (6):314–20, June 2003
[6] Gall D, Bieske K. “Definition and measurement of circadian radiometric quantities”, Proceedings of the CIE symposium 2004 on Light and Health: Non-visual effects, Vienna, Austria-Wien: Commission international de I’éclairage. S:129-132 (2004, Sep 30 - 2004, Oct 2)
[7] Brainard G C et al. “Action spectrum for melatonin regulation in humans: Evidence for a novel circadian photoreceptor”, Journal of Neuroscience, 21:6405–6412., 2001
[8] Thapan K, Arendt J, Skene D J. “An action spectrum for melatonin suppression: Evidence for a novel non-rod, non-cone photoreceptor system in humans”, Journal of Physiology, 535: 261–267., 2001
[9] Rea M S et al. “Circadian light”, Journal of Circadian Rhythms, 8, p.Art. 2, 2010
[10] Bellia L, Pedace A, Barbato G. “Winter and summer analysis of daylight characteristics in offices”, Building and Environment, 81(2014), pp. 150-161, November 2014
[11] “CIE 13.3-1995: Method of Measuring and Specifying Colour Rendering Properties of Light Sources”, CIE Technical Report 13.3, CIE. 1995
[12] [IES] Illuminating Engineering Society 2015. “TM-30-15 IES method for evaluating light source color rendition”, New York (NY): Illuminating Engineering Society p. 26 p.
[13] McKillup S. “Statistics Explained: An Introductory Guide for Life Scientists”, Cambridge University Press, ISBN: 978-0-511-13976-5, 2005

Modeling of phenylalanine metabolism and its medical relevance

Sváb G.[1], Tretter L.[1], and Szederkényi G.[2]
[1]Department of Medical Biochemistry, Semmelweis University, gsvab92@gmail.com, Tűzoltó u. 37-47, H-1094 Budapest, Hungary
[2]Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, szederkenyi@itk.ppke.hu, Práter u. 50/a, H-1083 Budapest, Hungary

Abstract: **Nowadays, more and more enzymopathies are observed, as most of them cause changes in phenotype or lead to metabolic disease. For example in degradation pathways of tyrosine and phenylalanine contains more enzymes, which play role in the development of several diseases. In this paper we summarize a mathematical model of metabolic pathways. With this model we analyze the effect of enzyme activity change on the actual metabolic state. We demonstrate that the model is able to predict the new metabolic status.**

Introduction

Recently, the study of cell metabolism has been in the focus of biological research, since most of the changes in cell function cause metabolic changes. Observation of these changes is usually difficult, but there are promising possibilities in metabolic pathway modeling. There is a clear need for cellular models which can predict the metabolic states and the modified enzyme activity in modified states. In the literature there are some dynamic models, but with limited metabolic pathways [1, 2, 3]. In our previous works we examined the intermediers of the citric acid cycle [4, 5]. These metabolites can transform to each other according to the actual metabolic state in mitochondria. The key elements in the regulation of citric acid cycle are the concentrations of intermediers and the changes in redox potential. Later the initial model was extended by adding mitochondrial transporters which connect cellular and mitochondrial metabolic networks [6]. After that we built cellular metabolic pathways to this model, for example urea cycle, amino acid degradation pathways and glycolysis. In this work we

present tyrosine and phenylalanine metabolism, which is one of these new moduls.

Biological background

Process description

In this work we present the model of the tyrosine and phenylalanine degradation pathways. These two amino acids are ketoplastic and glucoplastic (precursor for acetyl-CoA synthesis or gluconeogenesis, figure 1). They can transform into each other in an enzyme reaction catalyzed by phenylalanine hydroxylase, in physiological conditions this enzyme mainly catalyzes from phenylalanine to tyrosine, and after that they have similar metabolism. Tyrosine is precursor for catecholamines (dopamine, DOPA, noradrenalin, adrenalin) and several hormones (thyroxine and melatonin). The most popular enzyme defect causes phenylketonurie (PKU) which has two different types. Defect of phenylalanine hydroxylase leads to classic PKU, which has severe symptoms (e.g. mental retardation). Rarely, the lack of the tetrahydrobiopterin or defect of dihydrobiopterin reductase generate cofactor defect PKU. If the phenylalanine cannot transform to tyrosine, it has to metabolize via other pathways. In the first step transaminases form phenylpyruvate. After that it can go to two different degradation pathways to phenyllactate or phenylacetate (figure 2). Transformation to phenyllactate is a reversible process, but the transformation to phenylacetate and from phenylacetat is an irreversible decarboxylation step. Phenylacetate is a neurotoxic metabolite, which causes mental degradation and myelin lesion. Phenylalanine cannot transform to tyrosine, which results in less precursor for neurotransmitter synthesis.

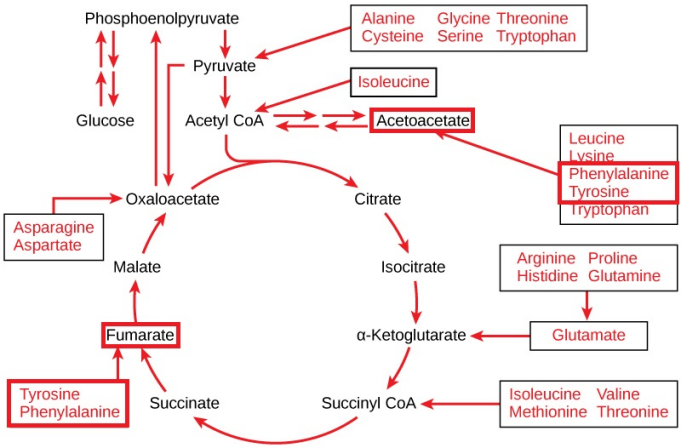


Figure 1: Phenylalanine and tyrosine metabolism [7]

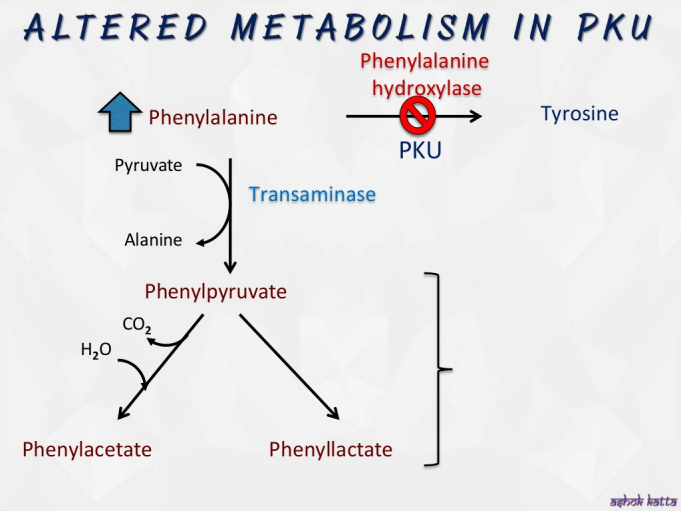


Figure 2: Phenylalanine metabolism in PKU [8]

Kinetic model of degradation processes

Modeling goals are the following:

1. To analyze the qualitative dynamic of substrates which participate in these reactions.

2. To observe the degradation methods of phenylalanine in PKU.

Parameters

We tested our model qualitatively (each K_m and V_{max} value is 1). This approximation is similar to that used in articles [2, 3]. In the next phase in our research we would like to determine more realistic healthy and mutant parameters to this model and test physiological and pathological metabolite conditions.

The applied kinetics

Differential equations were applied by using the extended Michaelis-Menten kinetic model [2]. Rates of reversible reactions are described by two opposite reactions. The general form of the reaction rates is

$$S_1 + S_2 + ... + S_n \rightleftharpoons P_1 + P_2 + ... + P_k$$

$$V_{S_1,...S_n} = V_{max}(E) \cdot (\frac{[S_1]}{[S_1]+K_m(S_1,E)}) \cdot ... \cdot (\frac{[S_n]}{[S_n]+K_m(S_n,E)})$$

$$V_{P_1,...P_k} = V_{max}(E) \cdot (\frac{[P_1]}{[P_1]+K_m(P_1,E)}) \cdot ... \cdot (\frac{[P_k]}{[P_k]+K_m(P_k,E)})$$

$$V_{S \rightleftharpoons P} = V_{S_1,S_2,...S_n} - V_{P_1,P_2,...P_k} =$$

$$= V_{max}(E) \cdot ((\frac{[S_1]}{[S_1]+K_m(S_1,E)}) \cdot ... \cdot (\frac{[S_n]}{[S_n]+K_m(S_n,E)}) -$$

$$-(\frac{[P_1]}{[P_1]+K_m(P_1,E)}) \cdot ... \cdot (\frac{[P_k]}{[P_k]+K_m(P_k,E)}))$$

Using the constructed model, we analyze the temporal changes of the metabolite concentrations with a quantitative mathematical description. In this work we modeled the degradation of phenylalanine in PKU.

Differential equations of the model

Enzyme	Abbreviation	EC	V_{max}
Phenylalanine hydroxilase	PheOH	1.14.16.1	0.4
Pteridine reductase	PterR	1.5.1.33	1
Phenylalanine transaminase	PheT	2.6.1.58	1
Lactate dehydrogenase	LDH	1.1.1.27	1
Phenylpyruvate decarboxylase	PPDC	4.1.1.43	1

Table 1: Enzymes abbreviations, EC numbers and V_{max} values

Substrate	Abbreviation	[S]
Phenylalanine	Phe	0.5
Phenylpyruvate	PhePyr	0
Phenyllactate	PheLac	0
Phenylacetate	PheAc	0
Tyrosine	Tyr	0
Pyruvate	Pyr	3
Alanine	Ala	0
Tetrahydrobiopterin	THB	0.3
Dihydrobiopterin	DHB	0
$NADP^+$	NADP	0
NADPH	NADPH	10
NAD^+	NAD	0
NADH	NADH	10

Table 2: Initial values

During the modeling we construct the differential equations with parameters of enzymes and transporters, and Michaelis Menten kinetics. Then we solve the equations, we get the concentration changes in time. Differential equations are as follows:

$$
\begin{aligned}
\frac{d}{dt}[Phe] &= -V_{PheOH} \cdot (\frac{[Phe]}{[Phe] + K_{M(Phe,PheOH)}} \cdot \\
&\cdot \frac{[THB]}{[THB] + K_{M(THB,PheOH)}}) \\
&-V_{PheT} \cdot ((\frac{[Phe]}{[Phe] + K_{M(Phe,PheT)}} \cdot \frac{[Pyr]}{[Pyr] + K_{M(Pyr,PheT)}}) \\
&-(\frac{[PhePyr]}{[PhePyr] + K_{M(PhePyr,PheT)}} \cdot \frac{[Ala]}{[Ala] + K_{M(Ala,PheT)}}))
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{dt}[PhePyr] &= V_{PheT} \cdot ((\frac{[Phe]}{[Phe] + K_{M(Phe,PheT)}} \cdot \frac{[Pyr]}{[Pyr] + K_{M(Pyr,PheT)}}) \\
&-(\frac{[PhePyr]}{[PhePyr] + K_{M(PhePyr,PheT)}} \cdot \frac{[Ala]}{[Ala] + K_{M(Ala,PheT)}})) \\
&-V_{LDH} \cdot ((\frac{[PhePyr]}{[PhePyr] + K_{M(PhePyr,LDH)}} \cdot \\
&\cdot \frac{[NADH]}{[NADH] + K_{M(NADH,LDH)}}) \\
&-(\frac{[PheLac]}{[PheLac] + K_{M(PheLac,LDH)}} \cdot \frac{[NAD]}{[NAD] + K_{M(NAD,LDH)}})) \\
&-V_{PPDC} \cdot (\frac{[PhePyr]}{[PhePyr] + K_{M(PhePyr,PPDC)}})
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{dt}[PheLac] &= V_{LDH} \cdot ((\frac{[PhePyr]}{[PhePyr] + K_{M(PhePyr,LDH)}} \cdot \\
&\cdot \frac{[NADH]}{[NADH] + K_{M(NADH,LDH)}}) \\
&-(\frac{[PheLac]}{[PheLac] + K_{M(PheLac,LDH)}} \cdot \frac{[NAD]}{[NAD] + K_{M(NAD,LDH)}}))
\end{aligned}
$$

$$
\frac{d}{dt}[PheAc] = V_{PPDC} \cdot (\frac{[PhePyr]}{[PhePyr] + K_{M(PhePyr,PPDC)}})
$$

$$
\begin{aligned}
\frac{d}{dt}[Tyr] &= V_{PheOH} \cdot (\frac{[Phe]}{[Phe] + K_{M(Phe,PheOH)}} \cdot \\
&\cdot \frac{[THB]}{[THB] + K_{M(THB,PheOH)}})
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{dt}[Pyr] &= -V_{PheT} \cdot ((\frac{[Phe]}{[Phe] + K_{M(Phe,PheT)}} \cdot \frac{[Pyr]}{[Pyr] + K_{M(Pyr,PheT)}}) \\
&\quad -(\frac{[PhePyr]}{[PhePyr] + K_{M(PhePyr,PheT)}} \cdot \frac{[Ala]}{[Ala] + K_{M(Ala,PheT)}})) \\
\frac{d}{dt}[Ala] &= V_{PheT} \cdot ((\frac{[Phe]}{[Phe] + K_{M(Phe,PheT)}} \cdot \frac{[Pyr]}{[Pyr] + K_{M(Pyr,PheT)}}) \\
&\quad -(\frac{[PhePyr]}{[PhePyr] + K_{M(PhePyr,PheT)}} \cdot \frac{[Ala]}{[Ala] + K_{M(Ala,PheT)}})) \\
\frac{d}{dt}[THB] &= -V_{PheOH} \cdot (\frac{[Phe]}{[Phe] + K_{M(Phe,PheOH)}} \cdot \\
&\quad \cdot \frac{[THB]}{[THB] + K_{M(THB,PheOH)}}) + V_{PterR} \cdot \\
&\quad \cdot((\frac{[DHB]}{[DHB] + K_{M(DHB,PterR)}} \cdot \frac{[NADPH]}{[NADPH] + K_{M(NADPH,PterR)}}) \\
&\quad -(\frac{[THB]}{[THB] + K_{M(THB,PterR)}} \cdot \frac{[NADP]}{[NADP] + K_{M(NADP,PterR)}})) \\
\frac{d}{dt}[DHB] &= V_{PheOH} \cdot (\frac{[Phe]}{[Phe] + K_{M(Phe,PheOH)}} \cdot \\
&\quad \cdot \frac{[THB]}{[THB] + K_{M(THB,PheOH)}}) - V_{PterR} \cdot \\
&\quad \cdot((\frac{[DHB]}{[DHB] + K_{M(DHB,PterR)}} \cdot \frac{[NADPH]}{[NADPH] + K_{M(NADPH,PterR)}}) \\
&\quad -(\frac{[THB]}{[THB] + K_{M(THB,PterR)}} \cdot \frac{[NADP]}{[NADP] + K_{M(NADP,PterR)}})) \\
\frac{d}{dt}[NAD] &= V_{LDH} \cdot ((\frac{[PhePyr]}{[PhePyr] + K_{M(PhePyr,LDH)}} \cdot \\
&\quad \cdot \frac{[NADH]}{[NADH] + K_{M(NADH,LDH)}}) \\
&\quad -(\frac{[PheLac]}{[PheLac] + K_{M(PheLac,LDH)}} \cdot \frac{[NAD]}{[NAD] + K_{M(NAD,LDH)}}))
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{dt}[NADH] &= -V_{LDH} \cdot ((\frac{[PhePyr]}{[PhePyr] + K_{M(PhePyr,LDH)}} \cdot \\
&\cdot \frac{[NADH]}{[NADH] + K_{M(NADH,LDH)}}) \\
&-(\frac{[PheLac]}{[PheLac] + K_{M(PheLac,LDH)}} \cdot \frac{[NAD]}{[NAD] + K_{M(NAD,LDH)}}))
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{dt}[NADP] &= V_{PterR} \cdot ((\frac{[DHB]}{[DHB] + K_{M(DHB,PterR)}} \cdot \\
&\cdot \frac{[NADPH]}{[NADPH] + K_{M(NADPH,PterR)}}) \\
&-(\frac{[THB]}{[THB] + K_{M(THB,PterR)}} \cdot \frac{[NADP]}{[NADP] + K_{M(NADP,PterR)}}))
\end{aligned}
$$

$$
\begin{aligned}
\frac{d}{dt}[NADPH] &= -V_{PterR} \cdot ((\frac{[DHB]}{[DHB] + K_{M(DHB,PterR)}} \cdot \\
&\cdot \frac{[NADPH]}{[NADPH] + K_{M(NADPH,PterR)}}) \\
&-(\frac{[THB]}{[THB] + K_{M(THB,PterR)}} \cdot \frac{[NADP]}{[NADP] + K_{M(NADP,PterR)}}))
\end{aligned}
$$

Simulation results

The simulation results are shown in Fig 3. We used decreased (40%) phenylalanine hydroxilase activity (tyrosine formation from phenylalanine is slow), and we add phenylalanine into the system as external input. The results show, that the toxic metabolite (phenylacetate) increases, and the other intermediers reach steady-state. Phenyllactate is a buffer intermedier in this system, because it can form phenylpyruvate instead of phenylacetate. However, its buffer capacity becomes saturated easily if we use constant input of phenylalanine.

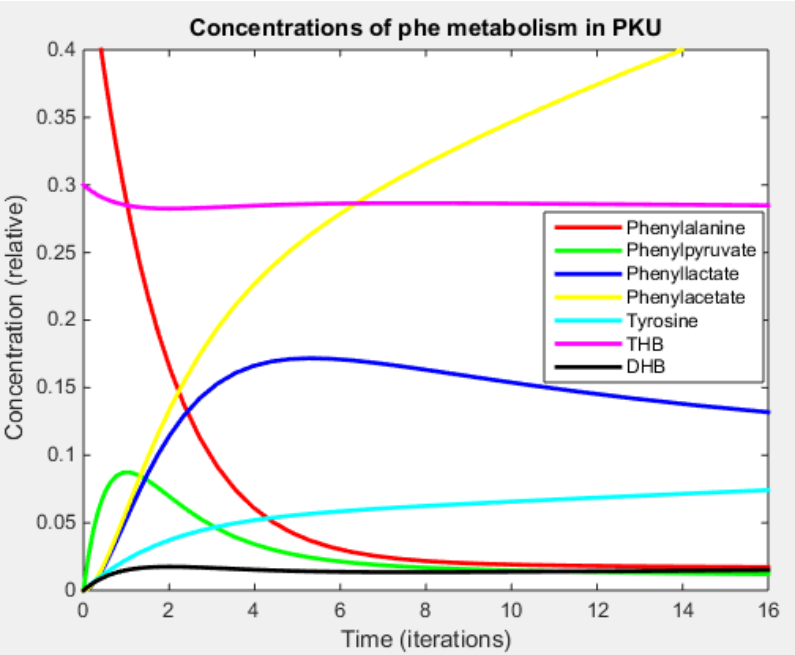


Figure 3: Concentrations of phe metabolism in PKU

Acknowledgment

The research has been supported by the European Union, co-financed by the European Social Fund through the grants EFOP-3.6.2-16-2017-00013 and EFOP-3.6.3-VEKOP-16-2017-00009.

References

[1] Wu, F., Yang, F., Vinnakota, K., Beard, D.: Computer modeling of mitochondrial tricarboxylic acid cycle, oxidative phosphorylation, metabolite transport, and electrophysiology. The Journal of Biological Chemistry, (24525-37 282(34)) 2007.

[2] Korla, K., Mitra, C.: Modelling the Krebs cycle and oxidative phosphorylation. Journal of Biomolecular Structure and Dynamics, (242-256 32) 2014.

[3] Korla, K., Vadlakonda, L., Mitra, C.: Kinetic simulation of malate-aspartate and citrate-pyruvate shuttles in association with krebs cycle. Journal of Biomolecular Structure and Dynamics, (2390-403 33(11)) 2015.

[4] Sváb, G.: Modeling of mitochondrial metabolism. Master's thesis, http://www.eregistrator.hu/drsvabgergely, 2017.

[5] Sváb, G., Szederkényi, G., Horváth, G., Tretter, L.: A simple dynamic model for mitochondrial metabolism. In MathMod 2018 Extended Abstract Volume, Wien, Austria, (11-12) 2018.

[6] Sváb, G., Horváth, G., Szederkényi, G.: Modeling of the citric acid cycle and its two shuttle systems. In 26th Mediterranean Conference on Control and Automation (MED), Zadar, Croatia, (70-77) 2018.

[7] https://s3-us-west-2.amazonaws.com/courses-images/wp-content/uploads/sites/1950/2017/05/31183431/figure-07-06-01.jpeg

[8] https://www.slideshare.net/ashokktt/metabolic-disorders-of-phenylalanine-and-tyrosine

Improved stress detection method for Ambient Assisted Living applications

B. Szakonyi[1], I. Vassányi[2], I. Kósa[3]
[1]University of Pannonia, Department of Electrical Engineering and Information Systems, benedek.szakonyi@virt.uni-pannon.hu
2 Egyetem utca, Veszprém 8200, Hungary
[2] University of Pannonia, Department of Electrical Engineering and Information Systems, vassanyi@almos.vein.hu
2 Egyetem utca, Veszprém 8200, Hungary
3 University of Szeged, Department of Medical Rehabilitation and Physical Medicine, kosa.istvan@med.u-szeged.hu
8-10 Korányi fasor, Szeged 6720, Hungary

Abstract: **Modern days' problems significantly increase the stress perceived by individuals, increasing the risks of developing serious health conditions. By detecting stressful moments, it is thought to be possible to help people escape such states, thus decreasing such risks. In this paper a method for stress detection is presented.**

Introduction

Studies have shown that increased stress levels have a notable effect on developing multiple civilization diseases such as diabetes [1] and numerous cardiovascular conditions [2]. However, by properly detecting when an individual's stress level has increased considerably, it could be possible, as an intervention, to warn the user of this unhealthy state. Taking a short break for trying to calm down a bit, until a more relaxed state of mind is reached, apart from helping in coping with the current problems more easily, could also have the positive effect of decreasing the risks of developing such unwanted health conditions.

In general, three different types of stress can be perceived: emotional, social and mental stress. The source of emotional stress is usually a strong negative emotions such as fear or anger, but it can also be as a result of strong anticipation or impatience. Social stress is generally considered to be a result of the need to comply with the expectation of others (in the work, school), to fulfil social obligations (e.g. friends, family), etc. Mental stress is commonly caused by though, hard to solve mental problems, or tasks that should be completed in a narrow time interval. While all 3 categories are equally

important components of the "global stress" perceived, in practice, mostly social and mental stress are investigated, as inciting emotional stress in participants is both hard to execute (e.g. people fear different things) and ethically questionable. (But even in the case of those two, finding the proper method can be challenging.)

How stress could be measured is a demanding task. It has already been investigated, how stress affects human physiology [3], and it was found that it has a high impact on the heart rate variability of individuals [4]–[6]. The complex indicator heart rate variability (HRV) was found to be useful for describing the complicated heart-brain interactions and autonomic nervous system functions that both have an effect on the amount of stress perceived [7], [8], meaning that is also capable for investigating the amount of stress [5], [6], [9], [10]. The basic concept of using HRV is that for healthy individuals, the lengths of the time intervals between successive heart beats are not constant, as they change accordingly to the environmental effects, allowing the body to cope with different in life situations. This means, that by analysing how these intervals change, it is possible to estimate the physiological state of an individual, i.e. to estimate how stressful the individual is.

There are various features used to describe these differences, but in general, those are grouped into two categories: time-domain features and frequency-domain features. While frequency-domain features are usually considered to be more closely related to describing the behaviour of the heart, time-domain features were found to be just as good as them in case of characterising HRV. Moreover, time-domain features, in general, require less computational power which makes them the favourable candidate for mobile computer device based solutions.

Methods

Measuring Heart Rate Variability

In general healthcare conditions, HRV is usually measured and calculated with advanced electrocardiogram (ECG) devices that usually take up a lot of space and/or are quite expensive. However, research has shown, that the quality and accuracy of both low cost and wearable heart rate sensors have already reached the point where they could be successfully used for stress detection [11]–[13]. Using simple "smartwatches" or chest belts connected to smartphones are solutions that require only a low effort from the users (i.e. putting them on and starting them), and are without considerable limitations on how they should or should not behave (e.g., no need to avoid quick

movements, that would cause trouble for holter devices). This allows them to be used easily in everyday life, and with processing their data it could be possible to provide “on the spot” assessment for reducing stress levels.

a) b)

Figure 1: Examples for wearable heart rate sensors
a) Polar M600 sport smartwatch b) CardioSport TP3 Heart Rate Transmitter chest belt

As mentioned already, because of their notably lower computational needs, for mobile HRV monitoring/measuring solutions, time-domain features are used. Part of such frequently used features use the RR intervals, i.e. the distances of the R components of successive ECG cycles.

Proposed Stress Detection Method

As HRV can differ greatly even amongst healthy individuals, meaning only loosely constrained intervals can be given on what is considered as a stressful state in general, it is better to investigate HRV on the individual level. That is, changes in HRV should be used to indicate how the stress level of an individual has changed compared to a reference point (e.g., to an initial interval, when the measurement started).

With this in mind, based on the algorithm presented in [11], a new stress detection method is being developed, using the mean HR, pNN50 and RMSSD features of HRV to measure stress. (The NN50 feature is the number of pairs of successive RR intervals that differ by more than 50ms, and pNN50 is the proportion of NN50 divided by the number of RR intervals. RMSSD is the root mean square of the successive differences.) During a measurement session, a sliding window of 5 minutes is used. The first 5 minute interval of the session is used to calculate the reference values for comparison. After this initial interval, the time window is updated with inserting the new data (and removing the oldest readings). In every 60 seconds, the 3 features are re-evaluated for the sliding window, and then are compared to the initial interval

in order to investigate how the user's stress level has changed. For each feature, based on the difference compared to its counterpart value for the reference interval, a score from -3 to +3 is given. Then the average of these values are computed, and this average is used to signal the level of stress change (where -3 is equivalent to high increase, 0 is no change, and +3 is high decrease).

Clinical Study

Data from the clinical study presented in [11] was used during the development process of the stress detection method. In the study, volunteers took part in two consecutive phases each 10 minute long, while being seated. In the first, they were asked to try and relax while a relaxation music was played. In the second phase, the so called Stroop colour game was used to incite stress (the game is about matching colours to labels, at an increasing pace). During the experiment, participants were wearing the CardioSport TP3 Heart Rate Transmitter to gather hear rate and RR data.

Results and Discussion

To test the method described above, it was first implemented in C++ programming language on a personal computer, in order to make sure that is provides to correct results first (by using C++, it will be possible to test the method on different mobile platforms in the future, with only simple modifications, when the algorithm has been validated.) Based on the data from study [11], the method seems to work as intended, as the situations that have been marked as stressful intervals in the test phase match those identified by the method proposed in [11]. However, as the former study focused on detecting only increased stress levels, the data acquired there was found to be inadequate to properly validate each functionality of the new method, as the test cases contain no "general" phases that could be compared to both the relaxed and stressful phases. Thus, for the validation, additional, more complex experiments are required, that are expected to yield results later this year.

While being useful, in their paper, Gambi et al. [14] has shown an interesting solution that might be capable of offering a better solution than wearable devices (in some cases). Their research states, that by using a "simple" Kinect device and videopletysmography it is possible to measure the heart rate of individuals without the need of wearing any devices. As such a method could serve as a solution requiring no effort from the users, investigating if it is accurate enough for providing reliable HRV data could prove useful results. If feasible, this could serve as an additional feature in a

"smart workplace", where automated assistance focuses on decreasing the modern lifestyle related health risks, caused by stress, prolonged sedentariness, and lack of physical activity.

Acknowledgment

We acknowledge the financial support of Széchenyi 2020 programme under the project No EFOP-3.6.1-16-2016-00015.

References

[1] A. M. Heraclides, T. Chandola, D. R. Witte, and E. J. Brunner, "Work stress, obesity and the risk of type 2 diabetes: Gender-specific bidirectional effect in the whitehall II study," *Obesity*, 2012.

[2] A. Steptoe and M. Kivimäki, "Stress and Cardiovascular Disease: An Update on Current Knowledge," 2013.

[3] J. Cacioppo, L. G. Tassinary, and G. G. Berntson, *The Handbook of Psychophysiology*, vol. 44. 2007.

[4] S. T. Nyberg *et al.*, "Job Strain and Cardiovascular Disease Risk Factors: Meta-Analysis of Individual-Participant Data from 47,000 Men and Women," *PLoS One*, vol. 8, no. 6, 2013.

[5] S. M. Collins, R. A. Karasek, and K. Costas, "Job strain and autonomic indices of cardiovascular disease risk," *Am. J. Ind. Med.*, vol. 48, no. 3, pp. 182–193, 2005.

[6] T. G. M. Vrijkotte, L. J. P. van Doornen, and E. J. C. de Geus, "Effects of Work Stress on Ambulatory Blood Pressure, Heart Rate, and Heart Rate Variability," *Hypertension*, vol. 35, no. 4, pp. 880–886, 2000.

[7] T. F. of the E. S. of C. the N. A. S. of P. Electrophysiology, "Heart Rate Variability," *Circulation*, vol. 93, no. 5, p. 1043 LP-1065, Mar. 1996.

[8] K. Martinmaki, "Ability of short-time Fourier transform method to detect transient changes in vagal effects on hearts: a pharmacological blocking study," *AJP Hear. Circ. Physiol.*, vol. 290, no. 6, pp. H2582–H2589, 2006.

[9] E. Clays *et al.*, "The perception of work stressors is related to reduced parasympathetic activity," *Int. Arch. Occup. Environ. Health*, vol. 84, no. 2, pp. 185–191, 2011.

[10] L. G. P. M. Van Amelsvoort, E. G. Schouten, A. C. Maan, C. A. Swenne, and F. J. Kok, "Occupational determinants of heart rate variability," *Int. Arch. Occup. Environ. Health*, vol. 73, no. 4, pp. 255–262, 2000.

[11] M. Salai, I. Vassányi, and I. Kósa, "Stress detection using low cost heart rate sensors," *J. Healthc. Eng.*, 2016.

[12] W. Lawanont, P. Mongkolnam, C. Nukoolkit, and M. Inoue, "Daily Stress Recognition System Using Activity Tracker and Smartphone Based on Physical

Activity and Heart Rate Data," in *Intelligent Decision Technologies 2018*, 2019, pp. 11–21.

[13] R. Zangróniz, A. Martínez-Rodrigo, T. M. López, M. J. Pastor, and A. Fernández-Caballero, "Estimation of Mental Distress from Photoplethysmography," *Applied Sciences* , vol. 8, no. 1. 2018.

[14] E. Gambi *et al.*, "Heart Rate Detection Using Microsoft Kinect: Validation and Comparison to Wearable Devices," *Sensors*, vol. 17, no. 8, p. 1776, 2017.

Automatic Removal of EOG artefacts from EEG based on Independent Component Analysis

M.F. Issa, Z. Juhasz, Gy. Kozmann
University of Pannonia, Department of Electrical Engineering and Information Systems, mohamed.issa@virt.uni-pannon.hu
Egyetem u.10, Veszprém, 8200, Hungary

Abstract: **Eye-related bioelectric activity, such as blinks, eye movements, generates large amplitude electrooculography (EOG) artefacts negatively affecting the accuracy of electroencephalography (EEG) measurements. Using independent component analysis (ICA) for rejecting these artefacts also rejects some of the EEG information contained in the rejected component. This paper presents a novel method for the automatic identification and removal of the EOG artefacts to generate high quality EEG. The proposed method removes only the EOG activity and keeps the underlying EEG data unaltered.**

Introduction

EEG recording is contaminated by different type of noises. These artefacts modify the measured EEG signals and consequently distort the detected activity. Artefacts may come from physiological sources, such as the ECG, muscle activity or EOG. Other kinds of artefacts are physical noises that appear as power line noise or variations in electrode-skin conductivity.

Independent Component Analysis (ICA) is widely used in the field of EEG signal processing for artefacts suppression, since it can separate a signal mixture into its main sources, such as EOG, ECG, EEG, etc. components [1]. The unwanted artefacts (components) can be rejected by visual or automatic inspections. Eyes movements and blinks are transient activities that occur relatively infrequently in the raw signal but generate distinct high-amplitude peaks. These might be located visually by checking the corresponding component. The usual approach is to reject a component entirely if it contains EOG artefacts. This, however, may lead to loosing important EEG data present in the component [2]. In our research, we devised methods to selectively remove EOG artefacts from ICA components and keeping relevant EEG information at the same time.

Related work

Independent Component Analysis [3], originally developed for solving the Blind Source Separation (BSS) problem, is considered a robust method for artefact removal able to minimize the mutual information between the sources and decompose the EEG signals to their independent components. If eye movement is recorded with special electrodes using electrooculography, this reference EOG signal can be used in ICA in combination with regression methods to automatically identify and remove the EOG artefacts from the contaminated signal and as a result, increase the signal-to-noise ratio (SNR) [4], [5].

Using the traditional artefact component rejection method, the reconstruction of the clean EEG might cause distortion in the signal spectrum that can lead to an overestimation of the coherence between different cortical sites [6], [7]. Wavelet Transform (WT) with adaptive threshold was introduced in the EEG applications to identify and remove the EOG [8] without losing the related EEG information. This approach was modified by Nguyen *et al.*, [9] who introduced Wavelet Neural Network (WNN) (clean and contaminated EEG data is used to train the network) and achieved 9.07 μV Root Mean Square Error (RMSE) between the cleaned and the noise free data. Their method works without a reference EOG signal that is normally required in the linear regression based methods [4]. Burger and van den Heever [2] improved on this method, however, their solution can only remove eye blinks, it does not work for eye movements.

In [7], wavelet transform was used in combination with ICA, based on the fact that wavelet coefficients of the artefact component typically have higher amplitudes than that of the cerebral activity components, so by setting the coefficients greater than a certain threshold to zero value, EOG artefacts can be removed from the signal.

Methods

The goal of our proposed artefact removal method is to clean EOG components by removing EOG peaks without rejecting the entire component. This ensures that we keep the EEG information content. We only reject components if the number of peaks is above a given threshold. The outline of our method is shown in Fig.1.

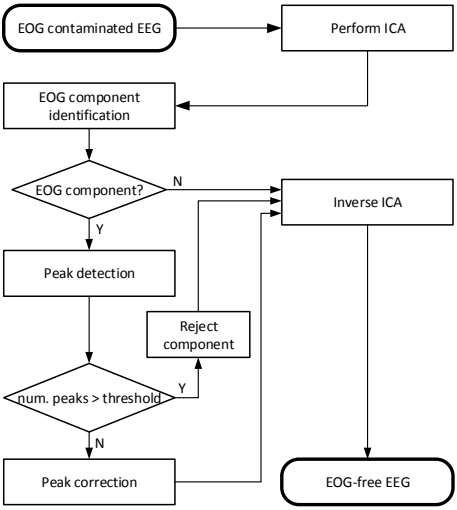


Fig. 1. The flow chart of the EOG artefact removal method.

Pre-processing: Each dataset is bandpass filtered (1-47Hz), then re-referenced to the average reference. ICA is applied to the signal to estimate the components. *Auto-identification of the EOG component*: the EOG component is identified based on the correlation between the data of the frontal EEG channels and each component. Components with over 70 percent correlation is marked as EOG and validated by a further test that includes finding the maximum unmixing weight of the components contributing to the frontal channels. EOG peaks in the identified EOG component are located using a predetermined threshold. A 1-second window is placed around the peaks and if these EOG windows cover more than 60 percent of the component, the entire component is marked for rejection. In all other cases, the window used for correction. WT-sym4 [10], [11] of 6 levels decomposes each window to different frequency bands and, only the bands of the high frequencies are reconstructed. The corrected windows are projected to the EOG component, and then the components are inverted to generate EEG free of EOG activity.

Datasets

Kaldos dataset [12]: The Kaldos EEG dataset was recorded from twenty-seven subjects (males and females), 19 EEG electrodes were used arranged

in the 10-20 international electrode layout, sampling frequency was 200 Hz. The noisy EEG was generated using the following expression:

$$\text{Contaminated\_EEG}_{i,j} = \text{Pure\_EEG}_{i,j} + \text{a}_j VEOG + \text{b}_{,j} HEOG \quad (2)$$

where $\text{Pure\_EEG}_{i,j}$ is the signal obtained with eyes closed (no EOG artefacts), and the $VEOG$ and $HEOG$ terms are the additive vertical and horizontal EOG activities.

Resting state EEG dataset: Eyes-open resting state EEG data was recorded from 61 adult volunteers (males and females, age from 17 to 35 years) of 2-3 minute's duration. During the experiment, subjects had to sit and relax in a silent room. Data were recorded using a Biosemi ActiveTwo EEG system (fs = 2048 Hz) using 128 electrode arranged in the ABC radial electrode layout. The volunteers gave their written consent to participating in the experiments.

Results

The proposed method was compared to the wICA algorithm [7] using the standard Kaldos dataset. The closed-eye resting state EEG was taken as the noise-free reference signal [4], [5], [12]. The root mean square error (RMSE) between the artefact-free and the cleaned EEG using the wICA and the proposed method are shown in Table 1. Our method produced an artefact free EEG signal with smaller RMSE than the wICA method (RMSE=6.31 μV vs 8.22 μV).

Table 1: RMSE (μV) of the wICA and our proposed method on the Kaldos dataset.

Dataset:	**s1**	**s2**	**s3**	**s4**	**s5**	**s6**	**s7**	**s8**	**Avg.**
wICA	9.43	7.13	7.76	5.90	5.18	8.92	8.92	9.94	**8.22**
	s9	**s10**	**s11**	**s12**	**s13**	**s14**	**s15**		
	8.70	8.91	8.12	7.95	8.17	9.26	9.24		
Dataset:	**s1**	**s2**	**s3**	**s4**	**s5**	**s6**	**s7**	**s8**	**Avg.**
Proposed Method	5.07	5.55	6.11	8.50	6.63	6.67	6.13	7.13	**6.31**
	s9	**s10**	**s11**	**s12**	**s13**	**s14**	**s15**		
	5.43	7.15	5.28	4.99	5.90	7.59	6.59		

Fig. 2 illustrates the efficiency of selectively removing an EOG effect from the contaminated EEG. Note how the contaminated signal (red) with the sharp blink artefact is corrected (blue) that matches the reference, artefact-free signal with little error. Fig. 3 shows the same result in the spatial potential distribution. The high-intensity artefact at the frontal area is removed giving rise to previously hidden details in the potential map.

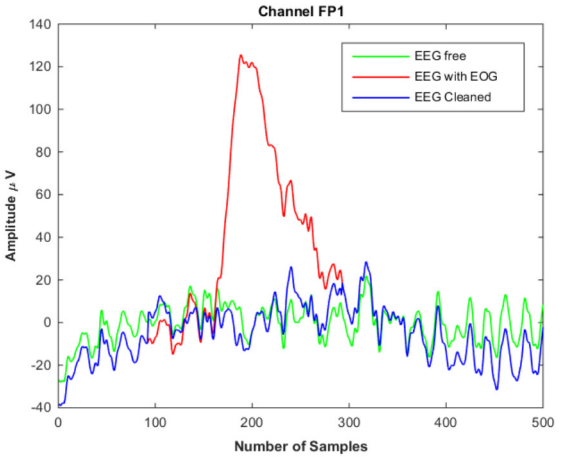


Fig. 2 Comparison of the artefact-free, the contaminated and the cleaned EEG signals.

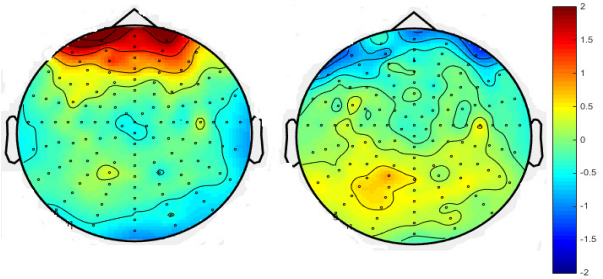


Fig. 3 Topoplot of the 128-channel EOG contaminated (left) and cleaned (right) EEG signals.

Conclusion

EOG activity in EEG recordings is a random artefact appearing unpredictably. Unfortunately, the effect of these artefacts is serious signal distortion due to the large amplitude of eye blinks and eye movements, hence their removal is a necessary step in EEG analysis. Using normal ICA-based rejection methods, we lose important EEG-related information. Our proposed method applies the wavelet transform to atomically identified EOG components and recovers the EEG data from the corrupted EOG segments by selectively removing only the EOG peaks from the

components. The proposed method improves upon the wICA method and produces results with better error rates.

Acknowledgment

We acknowledge the financial support of Széchenyi 2020 programme under the project No EFOP-3.6.1-16-2016-00015.

References

[1] R. Vigário, J. Särelä, V. Jousmäki, M. Hämäläinen, and E. Oja, "Independent Component Approach to the Analysis of EEG and MEG Recordings," 2000.
[2] C. Burger and D. Jacobus Van Den Heever, "Removal of EOG artefacts by combining wavelet neural network and independent component analysis," Biomed. Signal Process. Control, vol. 15, pp. 67–79, 2015.
[3] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," Neural Networks, vol. 13, no. 4–5, pp. 411–430, Jun. 2000.
[4] M. A. Klados, C. Papadelis, and C. Braun, "REG-ICA: A hybrid methodology combining Blind Source Separation and regression techniques for the rejection of ocular artifacts," Biomed. Signal Process. Control, vol. 6, no. 3, pp. 291–300, Jul. 2011.
[5] M. M. N. Mannan, M. Y. Jeong, and M. A. Kamran, "Hybrid ICA—Regression: Automatic Identification and Removal of Ocular Artifacts from Electroencephalographic Signals," Front. Hum. Neurosci., vol. 10, p. 193, May 2016.
[6] K. J. Friston, "Modes or models: A critique on independent component analysis for fMRI," Trends Cogn. Sci., vol. 2, no. 10, pp. 373–375, 1998.
[7] N. P. Castellanos and V. A. Makarov, "Recovering EEG brain signals: Artifact suppression with wavelet enhanced independent component analysis," J. Neurosci. Methods, vol. 158, no. 2, pp. 300–312, Dec. 2006.
[8] V. Krishnaveni, S. Jayaraman, S. Aravind, V. Hariharasudhan, and K. Ramadoss, "Automatic Identification and Removal of Ocular Artifacts from EEG using Wavelet Transform," 2006.
[9] H.-A. T. Nguyen et al., "EOG artifact removal using a wavelet neural network," Neurocomputing, vol. 97, pp. 374–389, Nov. 2012.
[10] P. Amorim, T. Moraes, D. Fazanaro, J. Silva, and H. Pedrini, "Electroencephalogram signal classification based on shearlet and contourlet transforms," Expert Syst. Appl., vol. 67, pp. 140–147, 2017.
[11] J. A. Jiang, C. F. Chao, M. J. Chiu, R. G. Lee, C. L. Tseng, and R. Lin, "An automatic analysis method for detecting and eliminating ECG artifacts in EEG," Comput. Biol. Med., vol. 37, no. 11, pp. 1660–1671, 2007.
[12] M. A. Klados and P. D. Bamidis, "A semi-simulated EEG/EOG dataset for the comparison of EOG artifact rejection techniques," Data Br., vol. 8, pp. 1004–1006, 2016.
[13] G. B. Moody and R. G. Mark, "The Impact of the MIT-BIH Arrhythmia Database : History, Lessons Learned, and Its Influence on Current and Future Databases," IEEE Eng. Med. Biol., no. June, pp. 45–50, 2001.
[14] A. L. Goldberger et al., "PhysioBank, PhysioToolkit, PhysioNet: Com- ponents of a new research resource for complex physiologic signals," Circulation, 13-Jun-2000.

Analysis of patient pathways in acute stroke care episodes

István Vassányi[1], Tamás Kováts[2], Görgy Surján[2], Zoltán Nagy[3]
[1]University of Pannonia, Medical Informatics R&D Centre,
vassanyi@almos.vein.hu
H-8200 Veszprém, Egyetem u. 10.
[2]National Health Development Institute, Budapest
[3]National Institute of Clinical Neurosciences, Budapest

Abstract: **We analyzed the practices of acute stroke referral based on the data store of the National Health Development Institute in the 8-year-long period between 2010 and 2017. The essence of the method lies in the classification of cases and events, forming episodes of 8 distinct types, finding the de facto dominant care regions of the providers, and characterizing the care practices of the providers with the relative frequency of the various episode types.**

The analysis highlighted some interesting spatial and provider related anomalies. There are remarkable, even an order of magnitude differences in the relative frequency of various event and episode types. An even more remarkable result is the formation of spatial patches of postal code areas with outlier values, on the relative frequency maps.

Some of the anomalies identified may be caused by the shortcomings of the current national case reporting and coding system. Any qualitative conclusions regarding the care system should be stated, however, after a more elaborate statistical analysis in the future.

Introduction

Public health care is a complex system which is very hard to comprehend by an intuitive or heuristic approach. Yet the standard data analysis tools and methods, routinely applied in any other industries, are very seldom applied in the field of health care planning and analysis in Hungary. The data store of the National Health Development Institute holds valuable data of every publicly financed case for more than two decades back by now, however, this data store has not been exploited for research until lately. In order for such an effort to be successful, it is imperative to understand the medical context and consider the current coding practices.

This paper presents the first results of a patient pathway analysis effort coordinated by the National Institute of Clinical Neurosciences in the field of acute stroke care. The main objective is to explore characteristics of the care

system, with a special respect to the spatial anomalies, the conformance to the applicable medical protocols [1] and the effect exerted by the specialized stroke care centers on the care network [2]. To our best knowledge, such an analysis has not yet been performed yet in the field of stroke care.

Data Sources

The anonymized case reports were queried from the Health Data Store of the National Healthcare Services Center and from the CT Register and complemented from various public sources. We defined acute stroke cases as cases associated with ICD codes I63 or I66 (acute ischemic stroke) as a main diagnosis, commencing between 1 January 2010 and 31 December 2017 and which had a CT performed in the -1..+7 day interval of the case onset date. This latter requirement was used to exclude the several cases (in fact two-third of all reported cases) which carried an acute stroke main diagnosis only for case financing reasons. The number of eligible cases was 281,948 belonging to 228,751 distinct patients. The yearly case number was ca. 32,000. We do not detail here the positive trends in the case numbers in the last decade.

Data cleaning was a complex process. We tracked down the known care provider mergers, new and closed up providers in the period to determine the set of valid care providers. Cases of patients whose gender, age or postal code area was impossible to determine were excluded from the study, as well as cases and procedures whose provider was unknown.

Methods

The methodology that we used for the patient pathway analysis was originally developed for analyzing ischemic heart disease episodes at the Medical Informatics Research and Development Center at the University of Pannonia [3]. The essence of the method is the formation of basic 'care events' based on case reports, and then episodes as a sequence of events that belong to the same patient and obey certain rules. Episodes are then classified according to their structure.

The event types we used included the skull CT, thrombolysis (TL), thrombectomy (TE) and basic care event (i.e. one without any invasive procedure or CT), denoted by EL. A CT is normally needed to assess the severity or nature of the stroke, and definitely needed before any TL/TE. We analyzed the spatial distribution of the various event types. We defined the maximal duration of an episode as 2 (5) days (5 for follow-up TL/TE events),

preceded by at least 1 event free day. Then we determined the *de facto* dominant regions of care for each stroke center according to the following procedure:

- For each postal code area, each episode gives one vote for the primary care provider of the episode. The *de facto* dominant care provider of the area is then selected as the one with the most votes.
- In case of a tie and if there was not a single episode in a postal code area, we select the closest provider, measured in time needed to reach the provider from the area by road, as the *de facto* dominant care provider of the area.

We classified the episodes according to their event sequences as follows.

- EL: an episode without CT, no further referral to another provider, no procedure performed
- CT: CT performed, no further referral to another provider, no procedure performed
- CT(TL/TE): CT and TL or TE performed, no further referral to another provider
- EL->CT: no CT at the first provider, the case is referred to another provider which performs a CT but no TL/TE
- EL->CT(TL/TE): no CT at the first provider, the case is referred to another provider which performs CT and TL/TE
- CT->CT: CT at the first provider, then the case is referred to another provider which performs a CT but no TL/TE
- CT->CT(TL/TE): CT at the first provider, then the case is referred to another provider which performs CT and TL/TE
- CT->TL/TE: CT at the first provider, then the case is referred to another provider which performs a TL/TE

Finally, the care practice of the providers was characterized by the spectra of the episodes that occurred in their region [4].

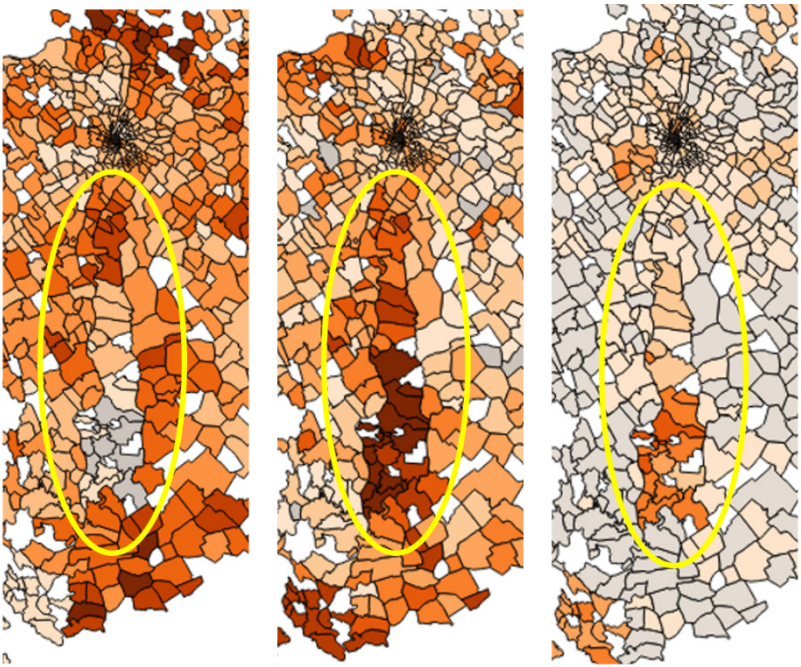
Results

We observed considerable differences in the spatial distribution of the cases. Considering only POSTAL CODE areas with a population over 1,000 for the whole 8-year-long period the minimum/maximum/average yearly case number per 1000 inhabitants was 0.96, 9.96 and 3.76, respectively, with a standard deviation of 1.19, which means that there are more than an order of

magnitude differences among areas with respect to the reported case numbers.

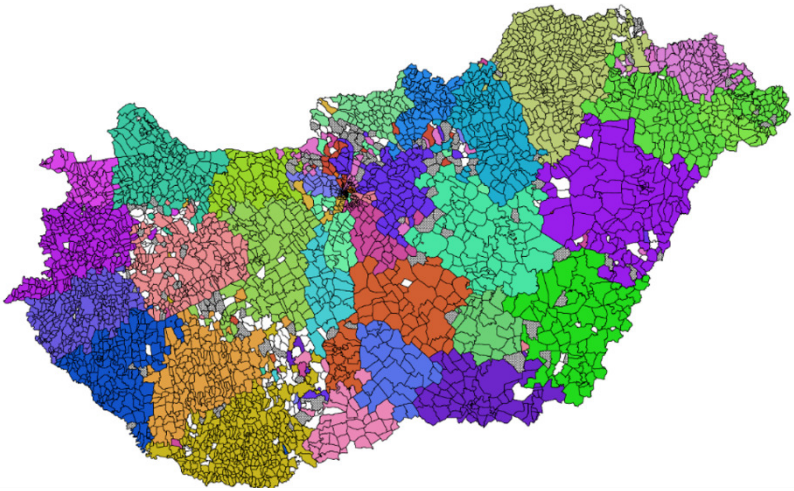
The event classification process yielded 282,229 events of type CT, 272,243 of EL, 11,743 of TL and 708 of TE that we could assign to an episode. The number of episodes by episode types was 220,061 for the CT type, 21,983 for EL, 15,610 for EL->CT, 10,974 for CT(TL/TE), 1,613 for CT->CT and 569 for CT->CT(TL/TE) i.e. in the vast majority of cases the care process involves a CT examination but no TL/TE procedure. This may be due to the nature of the stroke case or a time delay in the process which renders these procedures irrelevant.

We analyzed the spatial distribution of each episode type and found that it is in most cases quite inhomogeneous, with large distinguished, contiguous patches. For an example of this phenomenon see the three map sections below showing the central part of the country, with the episode numbers per 1,000 inhabitants for the CT (left), EL (middle) and EL->CT (right) episode types.



The maps show a vertical patch along the river Danube, of postal code areas with outlier values. This area is characterized by light shades i.e. low numbers of on-site CT (the expected protocol) and dark shades i.e. many cases of transporting the patient to another provider for CT. These altogether can be a sign of care logistics problems in the area. The maps do not show postal code areas with less than 1,000 inhabitants to improve statistical reliability.

The next map shows the *de facto* dominant practice regions of the specialized stroke clinics computed by the episode voting scheme. White means no episode, gray means a tie in the postal code area.



The regions are quite compact which is in line with our expectations following from the emergency nature of acute stroke care.

For each center, we computed the relative frequency of the 8 distinct episode types and performed a clustering procedure with these 8 ratios as cluster attributes. The result was that although there are considerable differences among the centers with respect to the frequency, no meaningful clusters could be formed, showing a homogeneous care practice across the country.

For more details on the results of the analysis, see [5].

Discussion

The most important result of the study was the identification of the spatial anomalies of the various episode types. There are remarkable, over an order of magnitude differences among postal code areas, but even more remarkable is how the outlier areas group together in *patches* in several cases. Since we used no geographical features for classifying the episodes, any clear and large outlier patches on these maps—which patches are to be found on virtually all the maps—mark a spatial anomaly of the stroke care system.

Another important conclusion of the study is the characterization of the care providers with their episode type frequencies as the 'care spectrum' of the provider. As a result, we identified some outlier centers that may need further investigation, or per case studies to find the causes of their practices.

The goal of this analysis was limited to the exploration of stroke care related data. Any qualitative statements regarding the care system must be formed and proven by statistical methods. This is a field for future research.

Acknowledgment

We acknowledge the financial support of Széchenyi 2020 programme under the project No EFOP-3.6.1-16-2016-00015.

References

[1] Jauch EC et al. Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/America Stroke Association. Stroke, 2013.

[2] Nagy Z, Javor A, Harcos P, Bodo M. Hungarian stroke program: 1988-2006. Int J Stroke. 2006 Nov;1(4):240-1. DOI: 0.1111/j.1747-4949.2006.00054.x.

[3] Vassy Zs, Kosa I, Vassanyi I. Correlation Clustering of Stable Angina Clinical Care Patterns for 506 Thousand Patients. Journal of Healthcare Engineering, Volume 2017 (2017), Article ID 6937194, DOI: 10.1155/2017/6937194

[4] Vassy Zs, Kosa I, Vassanyi I. Changes in the spatial distribution of dominant IHD care providers over a 10 year period in Hungary. In F. Bari, L. Almási (eds) Proc. XXIX Neumann Kollokvium, Szeged, Hungary, 2-3 December 2016, pp. 17-20.

[5] Nagy Z (ed). A hazai stroke beteg ellátás helyzete a statisztikai adatok tükrében. [A statistical overview of the national stroke care network, in Hungarian] Research report, National Institute of Clinical Neurosciences, Budapest, Hungary, 2019.

Predicting user actions under time constraints in a divided attention task

R.R. Saboundji[1,†] and R.A. Rill[1,2,*]

[1]Faculty of Informatics, Eötvös Loránd University
Pázmány P. stny 1/C, H-1117 Budapest, Hungary.
[2]Faculty of Mathematics and Computer Science, Babeş-Bolyai University
No. 1 Mihail Kogalniceanu St., RO-400084 Cluj-Napoca, Romania
Email: †s.rachid.riad@gmail.com, *rillroberto88@yahoo.com

Abstract: **Intelligent interfaces may need to anticipate user actions and errors in order to provide assistance and avoid dangerous situations. Therefore, in this work we evaluate several classification algorithms in an attempt to predict whether the user will manage or fail to perform actions in time. We performed experiments with ten participants using a special divided attention task, and use features computed from gaze and mouse-cursor movements as input to the classification methods. Using a restricted subsample of the experimental dataset we achieve a ten-fold cross-validated accuracy of up to 85%.**

Introduction

Intelligent interfaces and environments need to interact with humans while analyzing events and ongoing situations in the same time. This includes anticipating user actions based on their behavior [1], especially important in safety-critical systems, such as advanced driver assistance systems [2] or security surveillance [3]. In these cases failing to perform an action due to high cognitive load or timing constraints may have serious consequences.

Modeling human behavior and enhancing intelligent interfaces with predictive capabilities may result in improving their overall assistive potential. Indirect input, such as eye gaze, can be a useful source of information for revealing user intentions and future actions (see, e.g., [4]). When users control an interface with a cursor or a pointing device, often their gaze centers on the objects of interest before corresponding movements begin.

However, gaze and cursor coordination can also show complex and nuanced interaction patterns [5]. For example, gaze can leave the target area moving on to the next task before the click action is completed. Accordingly, it is not straightforward to recognize in advance whether the user will complete an action successfully.

In this work we use gaze and mouse-cursor movements to predict whether the user will manage or fail to perform an action in a special divided attention task that requires continuous focused concentration and frequent shifts of attention. In our analysis we restrict ourselves to the cases when there is only a limited amount of time left to perform the action and achieve classification accuracy values of up to 85%.

The paper is organized as follows. The Methods section introduces briefly the divided attention task, summarizes the experiments we performed with 10 participants, and lists the machine learning algorithms used in our analysis. The Results section presents our qualitative findings. Finally, in the last section we discuss our results, conclude our work and highlight future directions.

Methods

Task, participants and experiments

To analyze human performance we have designed and implemented a simplified version of the popular Train of Thought game from the Lumosity online platform [6]. The task of the user is to simultaneously concentrate on multiple moving objects over 2-3 minutes and direct them to their correct destinations through mouse clicks.

We conducted a longitudinal study with 10 participants aged between 25 and 30 years, who had normal or corrected to normal vision and reported no attentional disorders.

The subjects were asked to play with the divided attention task over a several day period, resulting in 60 gameplays each. Data about experiments was logged for later analysis, including mouse and eye-gaze movements. For gaze tracking we used the Tobii EyeX Controller [7] device. The sampling frequency was 60 Hz.

For the full details about our experiments performed, the interested reader is referred to our previous work [8].

Classification algorithms, features and prediction task

We frame the prediction task as a binary classification problem. For classifying successful and failed user actions we evaluated the scikit-learn [9] implementation of logistic regression (LR), support vector machines (SVM) with RBF kernel, random forests (RF) and k-nearest neighbours (k-NN) algorithms. We also experimented with the Dynamic Time Warping (DTW) [10] and Global Alignment (GA) [11] kernels in the case of the k-NN and SVM algorithms.

The features used for classification are three types of screen distances that characterize the cursor and gaze movements:

(i) gaze-target: distance of the gaze point (screen coordinates of the urers gaze direction) from the target to be clicked,

(ii) mouse-target: distance of the mouse cursor from the target,

(iii) gaze-mouse: distance between the gaze point and mouse cursor.

We computed the above distances for every timepoint before each correct and missed click event from the experiments, and use the flattened multivariate time-series for classification. For the purposes of the current work we take a subset of the full data. We impose a threshold of 40 frames (approx. 667 ms) as the time remaining until the last moment for performing a successful click action, obtaining a balanced dataset of 2293 positive (correct action) and 2296 negative (failed action) samples. Using this dataset we experiment with different time-series lengths, prediction thresholds for how much in advance we can predict correct/failed click actions and combinations of the three features. In each case we report the classification accuracy values, cross-validated over the 10 subjects from the experiments.

Results

Figure 1 shows the accuracy values for different time-series length values. The prediction threshold value was fixed to 40 frames, and all three features were used. The LR, SVM and RF methods show close values to each other and a roughly constant pattern, with SVM having slightly higher accuracy when considering short sequences. The k-NN algorithm shows a decreasing pattern as longer time-series are used.

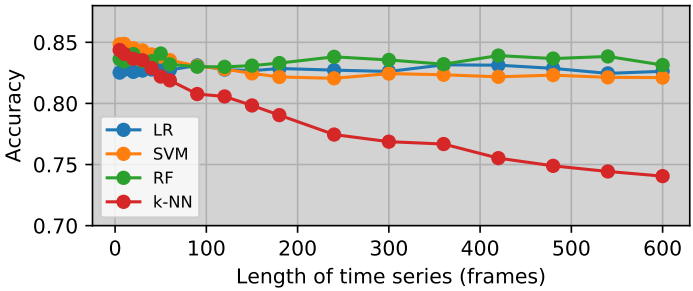


Figure 1: Effect of time-series lengths.

The effects of varying the prediction threshold for our dataset are illustrated by Figure 2. The time series length was fixed to 90 frames and all three features were used. Clearly the accuracy decreases in all cases as we try to predict the action of the users more and more in advance. At 120 frames the accuracy is only slightly above chance level, i.e. predicting the user action over 2 seconds in advance is equivalent with random guessing.

For the shorter time-series we also evaluated the k-NN and SVM algorithms with DTW and GA kernels. The results are shown on Figure 3. The accuracy is decreasing as length is increased, and k-NN with GA kernel shows the worst and k-NN with DTW the best performance.

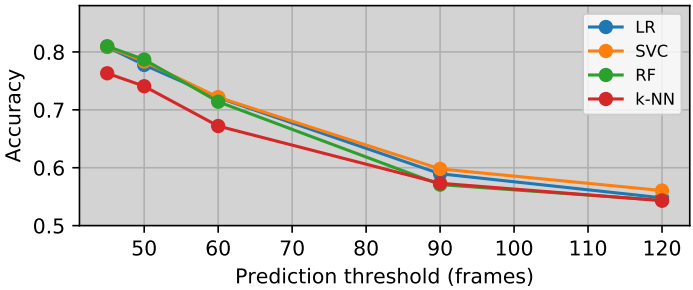


Figure 2: Effect of prediction threshold values.

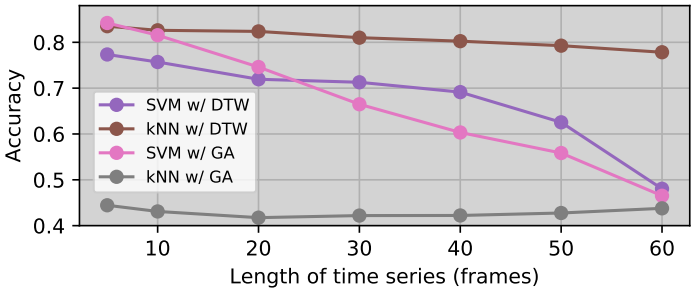


Figure 3: Effect of prediction threshold values.

Table 1 shows the results for the different combinations of the three features. The prediction threshold is fixed to 40 and the time-series length to 90 frames. Considering the gaze-mouse distance as the sole predictor gives accuracy values close to chance level, meaning that their coordinated movement alone does not differ for correct and failed clicks. We can also see that adding the mouse-target distance feature to the other two results in the largest increases in accuracy, when compared to the other cases. The results for k-NN with DTW kernel were fine-tuned, i.e. Gram matrices are computed separately for features and averaged, and the best k hyperparameter is chosen at cross-validation in each case. This lead to slight improvements compared to the other algorithms in all feature combinations.

Table 1: Feature combination accuracy values (G-gaze, M-mouse, T-target). *Fine-tuned results are shown.

G–T	✓			✓	✓		✓
M–T		✓		✓		✓	✓
G–M			✓		✓	✓	✓
LR	0.773	0.810	0.544	0.820	0.779	0.810	**0.831**
SVM	0.786	0.824	0.558	0.829	0.794	0.813	**0.831**
RF	0.791	0.819	0.554	0.837	0.804	0.822	**0.832**
k-NN	0.791	0.820	0.562	**0.821**	0.797	0.802	0.811
k-NN w/ DTW*	0.802	0.837	0.574	**0.851**	0.813	0.831	0.840

Discussion and conclusion

In this work we analyzed the classification of successful and failed user click actions under time constraints in a special divided attention task. We defined features that characterize the gaze and mouse movements of the subjects. Particularly we considered the distances of the mouse cursor and gaze from the click target and from each other. We constructed time-series from these and evaluated several algorithms and parameters for the sequences, achieving a cross-validated accuracy of up to 85%.

The dataset used in this study was restricted to cases when there is only a limited amount of time left to perform the click. This is challenging since although the gaze and mouse often show predictable and coordinated movements, complex interaction patterns can also be observed [5]. For instance the gaze might leave the target before the click action, or the user might click right after the last moment available. These latter cases are considered as failed user actions in our work.

We have evaluated several classical machine learning algorithms on our dataset and observed that considering longer time sequences for prediction does not increase accuracy (see Figure 1). Furthermore, as expected, the performance drops as we attempt to classify the user actions more and more in advance (see Figure 2). We also experimented with time-series similarity kernels (Figure 3). Fine-tuning schemes in the case of the k-NN algorithm with DTW kernel have shown slight but not significant performance increases.

Future works should consider using the full dataset for prediction. One promising direction is the application of long short term memory fully convolutional networks [12], capable of capturing long term dependencies. Another alternative approach is to use the effects of cognitive load as additional features (e.g., changes in pupil diameter). Also, one might opt to use time-series forecasting instead of classification.

Acknowledgments

The project has been supported by the European Union, co-financed by the European Social Fund (EFOP-3.6.3-VEKOP-16-2017-00001 and EFOP-3.6.3-VEKOP-16-2017-00002). The first author would like to gratefully acknowledge the support of the Tempus Public Foundation for sponsoring his PhD study.

References

[1] P. Schydlo, M. Rakovic, L. Jamone, and J. Santos-Victor, "Anticipation in human-robot cooperation: A recurrent neural network approach for multiple action sequences prediction," *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1–6, 2018.

[2] A. Jain, A. Singh, H. S. Koppula, S. Soh, and A. Saxena, "Recurrent neural networks for driver activity anticipation via sensory-fusion architecture," *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3118–3125, 2016.

[3] S. Tremblay, D. Lafond, C. Chamberland, H. M. Hodgetts, and F. Vachon, "Gaze-aware cognitive assistant for multiscreen surveillance," in *Intelligent Human Systems Integration* (W. Karwowski and T. Ahram, eds.), (Cham), pp. 230–236, Springer, 2018.

[4] A. Borji, A. Lennartz, and M. Pomplun, "What do eyes reveal about the mind? Algorithmic inference of search targets from fixations," *Neurocomputing*, vol. 149, pp. 788–799, 2015.

[5] D. J. Liebling and S. T. Dumais, "Gaze and mouse coordination in everyday work," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, UbiComp '14 Adjunct, (New York, NY, USA), pp. 1141–1150, ACM, 2014.

[6] J. L. Hardy, F. Farzin, and M. Scanlon, "The science behind Lumosity, Version 2," 2013. Lumos Labs, Inc.

[7] A. Gibaldi, M. Vanegas, P. J. Bex, and G. Maiello, "Evaluation of the Tobii EyeX Eye tracking controller and Matlab toolkit for research," *Behavior Research Methods*, vol. 49, no. 3, pp. 923–946, 2017.

[8] R. A. Rill, K. B. Faragó, and A. Lőrincz, "Strategic predictors of performance in a divided attention task," *PLOS ONE*, vol. 13, no. 4, pp. 1–27, 2018.

[9] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[10] W. Meert and T. V. Craenendonck, "wannesm/dtaidistance v1.1.2," July 2018.

[11] M. Cuturi, "Fast global alignment kernels," in *Proceedings of the 28th International Conference on Machine Learning*, ICML'11, (USA), pp. 929–936, Omnipress, 2011.

[12] F. Karim, S. Majumdar, H. Darabi, and S. Chen, "LSTM fully convolutional networks for time series classification," *IEEE Access*, vol. 6, pp. 1662–1669, 2018.

Investigating the visual forms of dynamic electronic work instructions to improve learning efficiency and productivity in assembly processes

Á. Lipovits[1], K. Tömördi[1], Zs. Vörösházi[1], R. Jinda[2]
[1,2]University of Pannonia, Faculty of Information Technology,
Image Processing Laboratory, 10. Egyetem, H-8200, Veszprém, Hungary
voroshazi.zsolt@virt.uni-pannon.hu

Abstract: **This case study explores how electronic work instructions affect the training of newcomers in assembly processes. Motivation behind this study was to improve operator performance while reducing the learning time and suppressing the scrap production which are important in the Industry 4.0 manufacturing technology. Experiments (two different trials and two test sets in various visual forms) were performed with operators, technicians, as well as with professionals, quality and production support engineers. More than 100 participants were involved in these experiments and results were evaluated by statistical methods in order to investigate the impact on learning efficiency and productivity in assembly processes. The proposed paperless instructions – Dynamic Electronic Work Instructions (DEWI) – will support the construction, distribution and management of assembly processes dynamically by enabling the collection of all required information.**

Keywords: **Dynamic EWI, Eye-tracking, Industry 4.0, Learning efficiency**

Introduction

This paper describes an experiment demonstrating how assembly instructions can increase operator performance and presents simple guidelines. Since the Third Industrial Revolution (called as Industry 3.0) the role of electronics and IT support in production and automation processes has emerged, thus introducing the cooperative work of humans and machines.

Nowadays, in the latest paradigm of Industry 4.0 [1] the number of cyber physical systems has significantly increased, and they can be connected in various network hierarchies (e.g. IoT devices, virtual networks, Cloud, cognitive computing, etc.). Within the framework of Industry 4.0 the human-centric perspective has been further emphasized: a new generation of tech-augmented human worker in the Operator 4.0 was introduced in [10]. Accordingly, providing flexible, dynamically distributed assembly instructions for operators during assembly procedures is crucial. The study –

consisting of a pre-test, two different experiments with eye-tracking systems, and a post-test – was carried out in order to investigate the learning efficiency and role of assembly instructions. The visual appearance, formation and layout of design guidelines (containing several textual, graphical, and multimedia elements) can be used to improve operator performance and productivity. The first guidelines [2] were based on the notion that supporting the cognitive process the structuring of both the assembly procedure (planning and design) and of the instructions (presentation) are important.

Our working procedure will be described in the next section showing how the guidelines can be put into practice. Initially, several references were examined. The design phase includes mapping and planning of the current assembly procedure [2]. The next is the presentation phase including steps such as the creation of instruction layout with pictures, textual, and multimedia contents, along with some enhancements or refines of them [3-5]. Finally, the last step is to validate the instructions by tests and user questionnaires ensuring that the operator perspective is captured. These recommendations were applied for our instructions, and further proposals were also investigated. Accordingly, the instructions should have a high focus on pictures including relevant information and text should only be used when pictures are insufficient [4-5] and pictures should be high resolution having high contrast and reduced shadows [8].

Ganier et al. [11] pointed out that participants learn manual techniques, the results show that the videos are much more effective for the first trial than photos by considering execution time. This trend reversed in their following experiments when participants knew what to do: the time spent with the photos was shorter than with the videos.

Methods

Two different trials and two test sets in various visual forms (hereafter called as 'experiments') were performed with large variety of participants: operators, technicians, as well as with professionals, quality and production support engineers. More than 100 participants were involved in these experiments both at the university and our industrial partner.

2.1. Pre-test

Before the trials, participants filled a form and answered a set of questions investigating general personal information (e.g. identity, name, age, qualifications, education, etc.) and different abilities (e.g. eyesight, manual skills with equipment, etc.). This pre-test aims to collect information

that might have relevant impact on the speed and successfulness of the assembly process. Some questions have assessed whether the experimenter knew a fundamental electric circuit component or other parts at all.

2.2. Trials

After the pre-test, participants also undergo two different trials one after another, elaborated to measure how the visual forms of electronic work instructions affect the learning process and cognitive ability of each participant in a real assembly environment:

- The first practical working instruction was created to assemble a Timer-555 IC as an electric circuit on a breadboard panel to flash a LED.
- The second trial was to connect to a traffic-lamp LED panel with a configurable Xilinx FPGA circuit board via fly-wire and USB cables.

For these trials four different types of work instructions were made as shown in Figure 1: combined text-picture, pure picture (without explanatory textual information), mixed video-picture and pure video (with minimal number of pictures).

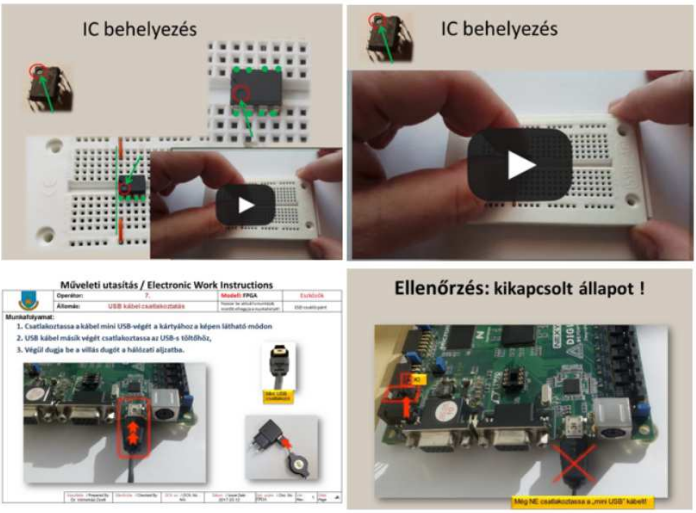


Fig. 1.: Working instructions for the Timer-555 IC and FPGA-based trials: combined picture-video, video only format without explanatory textual information, combined picture-textual, and picture-subtitled format (from top-left to bottom right order).

Figure 2. shows the distribution of various trials (Timer 555 IC, and FPGA) according to the number of participants.

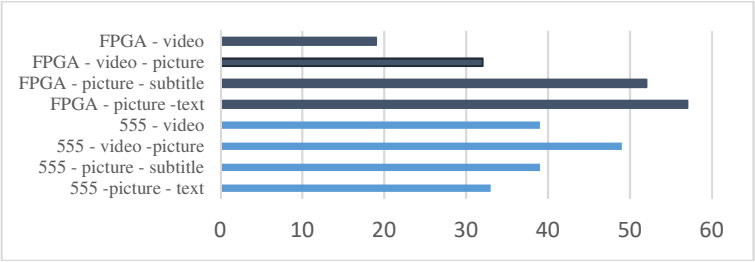


Fig. 2. Distribution of various trials according to the number of participants

During the experiments two different eye-tracking systems were installed, properly calibrated and applied for measuring and recording all tracking information. These systems consist of sensors (camera + projectors) and algorithms. With eye tracking the devices we know where the user's focus is at any given point (gaze) in time and we can analyze later their interventions. Figure 3.a. shows the Tobii Pro Glasses 2.0 [9] stand-alone wearable, head-mounted glasses product, which is built up from a head unit and a recording unit. Another test system was a Tobii X120 (Fig.3.b), a stand-alone, screen-based eye-tracker environment with sampling rate of 120 Hz, which was mounted on a table during the assembly procedure. The operation of the Timer-555 oscillator circuit, and FPGA controlled traffic lamp could be easily checked: a LED unit blinked periodically when the assembly process was successfully completed.

Fig. 3. (a) Tobii Pro Glasses 2.0; (b) Tobii X120 eye-tracking system, and (c) A Timer-555 IC trial recorded with Tobii Pro Glasses

2.3. Post-test

Finishing the trials participants also had to fill post-test partly repeating some questions of the pre-test, but in a technically more detailed way by reflecting what happened to the assembly process (e.g. recognizing electrical elements and devices in photos, remembering terminology, etc).

Therefore, we utilized the test results to analyze the internalization of new information during the experiments.

Results

The analysis of the eye-tracking data pointed out that the experienced operators focused only to the relevant information in contrast to newcomers (in Fig. 4).The prominence of warnings and notes on all the instructions in the EWI template is only relevant during the training period, can be omitted in a real environment that further improves the optimal layout.

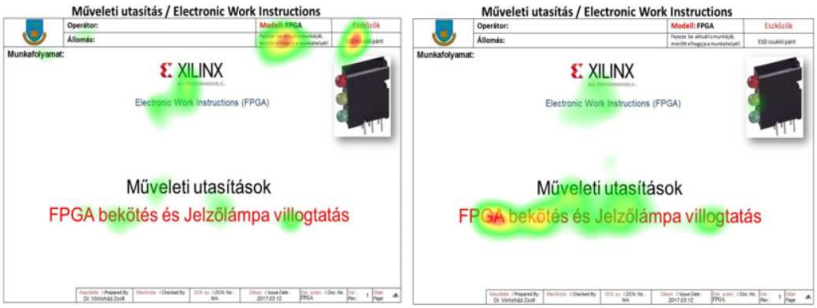


Fig. 4. Heat map (focus-points) of (left) newcomer operators; (right) experienced operators based on the evaluation of eye-gaze data set

Table 1. shows the cross tabulation of the points of tests, and a related-samples Wilcoxon signed rank test indicated that there was a significant difference in the results of the Pre-test and Post-test ($Z=-9.74$, $p<0.05$).

Table 1: Cross tabulation of the result of the Pre-test and Post-test

Count

		test							
		3,00	4,00	5,00	6,00	7,00	8,00	9,00	Total
pretest	1,00	0	1	0	1	1	2	0	5
	2,00	0	0	2	5	1	1	1	10
	3,00	0	0	1	2	3	1	2	9
	4,00	0	1	2	4	6	1	2	16
	5,00	1	1	3	2	5	8	5	25
	6,00	0	1	3	1	8	4	11	28
	7,00	0	0	0	2	2	9	14	27
	8,00	0	0	0	0	1	11	28	40
Total		1	4	11	17	27	37	63	160

We investigated the possibility of the prediction of operation's faults using eye gaze, head moving data and other features such as age, education, experience, theoretical background. The classifiers with relevant parameters and accuracy can be found in Table 2. The IBM SPSS Modeler has also been used in the experiments for analytic tasks.

Table 2: Accuracy of classification models

Model	Parameters	Accuracy
Logistic regression	multinomial	85.6%
SVM	RBF, C=10, eps=0.1, G=0.1	85.4%
Random Trees	nom=100, maxnodes=10000, childns=5	84.7%
Bayesian Network	Structure type=TAN, plm=maximum likelihood	79.3%

Conclusion

This study proves that the learning process of work instructions can be accelerated to newcomers by using eye-gaze and head moving statistics. Furthermore, we can reduce the number of wastes by eliminating information which were not emphasized properly related to the source of error. As a result, a suitable software framework application was developed for creating, sharing and managing electronic work instructions dynamically, while improving quality and productivity performance.

Acknowledgment

This paper was supported by the State of Hungary in the framework of VKSZ_14-1-2015-0190 and OTKA-120367. We also acknowledge the financial support of Széchenyi 2020 programme under the project No EFOP-3.6.1-16-2016-00015

References

[1] Hermann, M., Pentek T., Otto, B., 2016, Design Principles for Industrie 4.0 Scenarios, 49th Hawaii International Conference on System Sciences (HICSS)

[2] Agrawala, M., Phan, D., Heiser, J., Haymaker, J., Klinger, J., Hanrahan, P., Tversky, B., 2003, Designing Effective Step-By-Step Assembly Instructions, Proceedings of ACM SIGGRAPH, Volume 22, Issue 3 p. 828-837.

[3] Osvalder, A.-L. &Ulfvengren, P., 2009, Human-technology systems. in M. Bohgard, a.o. red. Work and technology on human terms, 339-461. Prevent, Stockholm.

[4] Inaba, K., Smillie, R. &Parasons, S.O. 2004, Guidelines for developing instructions, CRC Press Inc, New York; London.

[5] 2004, Factors affecting the processing of procedural instructions: implications for document design, IEEE Transactions on Professional Communication, 47, 1, 15-26.

[6] Clark, R.C., Nguyen, F., Sweller, J. (2006), Efficiency in learning: evidence-based guidelines to manage cognitive load, Pfeiffer, San Francisco, Calif.

[7] Söderberg, C., Johansson, A., Mattsson, S., 2014, Design of simple guidelines to improve assembly instructions and operator performance, The 6thSwedish Production Symposium.

[8] Li, D., Cassidy, T., Bromilow, D., 2013, The Design of Product Instructions, licensee InTech, 101-114, Leeds

[9] Tobii Pro Glasses 2.0 and Tobii TX 120 eye-tracker product specifications (2019). Webpage: https://www.tobiipro.com/product-listing/

[10] Romero, D., Stahre, J., Wuest, T., Noran, O., Bernus, P., Fast-Berglund, Å. and Gorecky, D., 2016. Towards an Operator 4.0 Typology: A Human-Centric Perspective on the Fourth Industrial Revolution Technologies, International Conference on Computers & Industrial Engineering (CIE46) Proceedings, Tianjin, China.

[11] Ganier, F, Vries, P: Are instructions in video format always better than photographs when learning manual techniques? Learning and Instruction, Vol 44, Elsevier (2016) 87-96

An optimization based algorithm for conflict-free navigation of autonomous guided vehicles

B. Csutak[1,2], T. Péni[2], G. Szederkényi[1,2]
[1]Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Práter 50/a, H-1083 Budapest, Hungary (e-mail: csutak.balazs@hallgato.ppke.hu,szederkenyi@itk.ppke.hu)
[2]Systems and Control Laboratory, Institute for Computer Science and Control (MTA SZTAKI), Hungarian Academy of Sciences, Kende 13-17, H-1111 Budapest, Hungary

Abstract: A novel optimization based route design approach for autonomous guided vehicle transport systems is presented in this paper, followed by the analysis and implementation of an algorithm capable to provide suboptimal solutions for route planning problems in real time providing conflict-, collision- and deadlock-free navigation by design.

Introduction

Optimal route planning based on transport demands is an intensively investigated topic in several engineering fields. Depending on the applied model and assumptions, the computational complexity of such task moves on a wide scale. Route planning problems are commonly modeled as optimization problems, which can indeed give us an optimal solution, but scale badly as the size of the map or the number of agents increases. This means that the real time operation of such methods is often non-realistic due to the need of re-planning.

In [1], a mixed integer optimization based distributed route-planning method is proposed for multiple mobile robots using Dijkstra's algorithm with a special cost function considered to manage vehicle interdependence. The idea of routing with high resource utilization appears in [2, 3], where the deadlock avoidance is addressed, but lacks the ability to handle routes that became unrealizable meanwhile. In [4], a similar routing strategy is introduced for airport taxiways, though it works with different assumptions and optimization goals.

In this line of research, an efficient dynamic real-time algorithm is proposed in [5, 6, 7], where time windows and resource reservation are used to ensure conflict-free navigation by design.

Following the ideas presented in [5, 6, 7], we investigate optimal route planning for multiple types of automated guided vehicles in a microscopic routing environment, where the size of the vehicles in the system is comparable to the size of the underlying network. For this reason, the route planning algorithm should be prepared to avoid collisions and handle congestion and even deadlock problems. Moreover, the algorithm should be able to address additional arising problems, such as vehicles unable to follow their previously planned routes.

As for the optimization, we are trying to find a solution for two common optimization tasks: the *on-line shortest dynamic disjoint path problem* (OSDDPP), and the *on-line quickest disjoint path problem* (OQDPP).

This research is motivated by its possible applications in automated guided vehicle (AGV) routing systems, mostly aimed for industrial application in a research-oriented experimental factory cell in Győr.

Dynamic routing

Formal problem statement

To present the problem in a formal way, following the author of [6], we model the routing environment with a graph $G = (V, E)$ with nodes $V = \{1, 2, ..., N\}$ and edges $E = \{(v_1, v_2, l) | 1 \leq v_1, v_2 \leq N\}$. The graph is directed, and has no multiple or loop edges. The weight of the edges (representing length) is denoted by l. Agents can have different traversal times, based on their maximal speeds.

Transportation tasks are continuously arriving for the agents, and are assigned to the vehicles by a higher level dispatching system.

Definition 1. *A* request *is a tuple $r = (s, t, \theta)$, where s is the source node (from where the agent starts), t is the target node, and θ is the earliest time, when execution of the requests can begin.*

Definition 2. *A* dynamic path *in a graph G is defined as a sequence*

$$P = ((v_0, \theta_0), (v_1, \theta_1), ..., (v_k, \theta_k))$$

of $v_1, ...v_k$ nodes and $\theta_1, ..., \theta_k$ timestamps. Timestamp θ_i is called a reservation *of node v_i, i.e. it constitutes the earliest time when node v_i can be entered. Similarly, interval $(\theta_{i-1}\theta_i)$ is called a* reservation *of the*

edge between v_{i-1} and v_i. The duration *of a dynamic path is defined as $\Delta p = \theta_k - \theta_0$.*

The aim of the algorithm is to compute a set of *disjoint* dynamic paths (having no overlapping time intervals between reservation times of the contained edges) in order to serve the dispatched requests, while minimizing specific cost functions.

Definition 3. *The Online Shortest Dynamic Disjoint Path Problem is defined as follows: Being given a sequence of requests $(s_i, t_i, \theta_i), i = 1, \ldots, k$ find a sequence of disjoint paths $P_1, ..., P_n$, for which $\sum \Delta p_i$ is minimal.*

Definition 4. *The Online Quickest Disjoint Path Problem is defined as follows: Being given a sequence of requests $(s_i, t_i, \theta_i), i = 1..k$ find a sequence of disjoint paths $P_1, ..., P_n$ with minimal maximum completion time over all paths (so that $\max_{i=1...n} \theta_i$ is minimal)*

These problems do not have a common solution, in practice however, the same suboptimal algorithm turns out to be suitable for both cases. It must be noted as well, that finding the optimal solution is not possible due to the continuously arriving requests.

The routing algorithm

Following the ideas presented in [5, 6, 7], we propose an improved greedy algorithm, which, instead of minimizing the overall cost function of the system, it focuses on minimizing the route completion time for the individual agents as the requests arrive, without disturbing the routes already computed.

To achieve this behavior, the algorithm introduces time windows for the graph edges, so agents can reserve all the edges in their path at the very beginning. In this way, the unnecessary waiting or deadlocks can be completely avoided, as the planning algorithm considers these reservations, and looks for the quickest route.

Formally, this goal can be described as follows:

Definition 5. *The Quickest Path Problem with Time Windows: It is given a graph $G = (V, E)$, a set of time windows for the edges, a request $r = (s, t, \theta)$ and an agent in s. Compute a dynamic path with minimal completion time, which uses the edges of the graph in the free time windows only.*

For this task, an algorithm is given in [6], which resembles Dijkstra's simple route planning algorithm, but instead of a single cost value being stored for a graph node, it is based on multiple labels $(a, b, ...)$ being assigned to edges, representing the time intervals, in which the agent can arrive to the respective edge. The algorithm is initialized with labels (t, ∞) on the edges having s as tail, and they are expanded to neighbouring labels iteratively (like in Dijkstra's algorithm), taking in consideration the free time windows.

Handling agent delays

As the system is aimed to be suitable in a real environment, practical considerations must be made. Due to the nature of such environments, there are several factors that can influence the routes planned, and cause already planned routes to become impossible to complete.

Minor time differences, arising from uncertainties in vehicle control can be easily solved, by reserving an interval slightly longer than necessary for the traversed edges. When this safety interval is not enough, and an agent can not free the resource until it would be obliged to, re-planning must take place.

In case of a severe latency, we decided to stop all vehicles in the system, and then all agents are required to re-plan their routes sequentially one after another taking into consideration the eventual changes in the graph (eg. a broken vehicle permanently blocking an edge). As a consequence, all edges where the agents are actually located (and then stopped) should be reserved until they can leave that edge. Though the algorithm can handle the computations in real time, a difficulty is that the agent, who is planning first cannot foresee when will the edges be released where the other agents are still waiting for the possibility to re-plan their routes. On the other hand, the agent, who is planning last may face the problem that all neighbouring edges are already reserved, thus it cannot leave its position until a certain time. Generally, it is not straightforward how to determine the length of the time window for each initially occupied edge. This interval must be chosen carefully, since, if we consider a short time window, the agent planning last might not be able to leave the edge until the end of this time window, and thus causing another delay.

In our solution, a simple heuristics was applied: for all agents stopped at time T_0, a reservation for the interval $(T_0, T_0 + \Delta T)$ was made, and the route planning system assumed, that the agents can leave their position in that interval. This behavior was further aided by initializ-

ing the route planning algorithms with starting labels containing this interval, so all agents try to leave their place as soon as possible.

If an agent is unable to leave its location until the and of this time window, that delay is handled by another severe latency, causing a repeated recalculation of all routes in the system (eventually with a higher ΔT value).

Test cases

The simulation environment

To extensively test the implemented framework system for verification, and check the usability of the implemented algorithm in the scenarios needed, an exact model of the experimental factory cell from Győr was realised in the simulation system. This process involved loading and transforming the data acquired from measurements of the place, so that a three dimensional representation of walls and objects would appear. Next, a directed graph based on the loaded floorplan was created, followed by the generation of nodes and edges in the air, suitable for quadcopters only. Finally, after generating a planner graph and defining some parking places / workstations (nodes, from which and to which transportation requests are arriving), two ground AGVs and two quadcopters were loaded and launched. The floorplan of the graph, together with the planned routes, can be observed in Figure 1.

The experiment was carried out using MATLAB 2018b, on a Dell Vostro 5471, having processor model Intel Core i7-8550U (4 cores, 8 threads, up to 4GHz) and 8 GB RAM. The planner graph resulting from the scene had 158 vertices and 506 edges.

Loading AGVs and generating requests

The AGVs were loaded to the following positions: two ground AGVs to nodes 9 and 18, and two quadcopters to nodes 1 and 16. In this order, the agents had the targets 18, 16, 9, 20. The route planning took place in order 16, 18, 9, 1, where the numbers refer to the departure nodes of the corresponding agents. The routes calculated and followed by the agents can be seen in Figure 1. One can observe that the agent starting from node nr. 18 has chosen a route $\{18, 13, 17, 14, 16\}$ instead of the spatially shorter one $\{18, 13, 12, 14, 16\}$ in order to avoid the interference with the already scheduled route starting from node nr. 16. As for the performance of the algorithm, all route planning operations took place in a time less than 0.1 seconds.

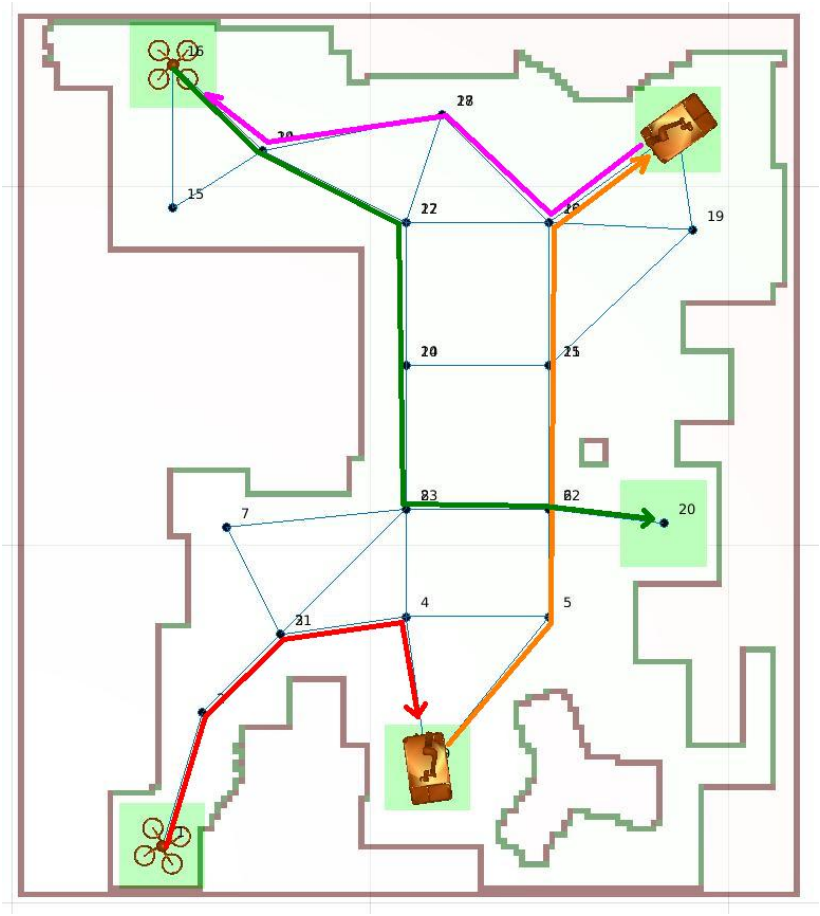


Figure 1: Routes planned using the algorithm

Further test cases

To assess the performance in a more complex scenario, an extended version of the experimental factory cell (see Fig. 2) was created (including as part the original one as well). The underlying graph has above 200 nodes, resulting in a planner graph with over 600 nodes 1000 edges. There were 10 AGVs (4 ground vehicles and 6 quadcopters) loaded, each of them starting randomly from one of the 11 workstations / parking places. Targets for the agents were generated randomly. During the simulation, as soon as one of the agents reached its target, a new one was randomly assigned to it.

The algorithm had a decent performance, computation times for a single route remaining below 1 second in any circumstances, which could be further reduced by some fine-tuning of the implementation.

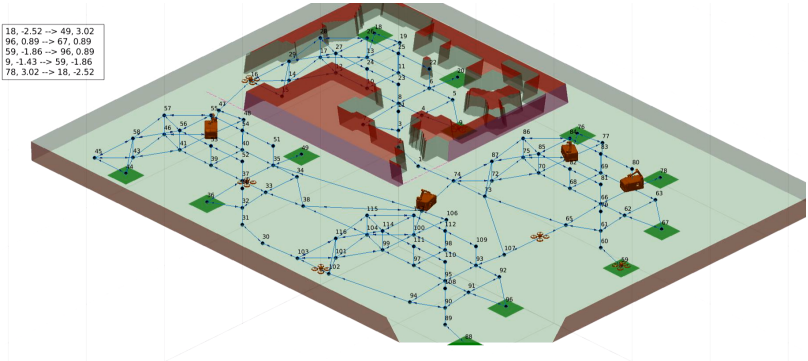


Figure 2: Extended experimental factory cell

Movement of the AGVs for all the test cases can be seen in the demonstration videos available online at `https://drive.google.com/open?id=1-6iPlFZdlAf10fzSWVcLuRQq5ybuvCyd`.

Summary

In this work, we briefly presented the operating principle of an algorithm, capable of providing online disjoint autonomous vehicle route planning for multiple type of AGVs moving in a microscopic routing environment. The algorithm was implemented and thoroughly tested, together with our method for handling serious delays of the agents.

Acknowledgements

B. Csutak gratefully acknowledge the support of the New National Excellence Program scholarship (ÚNKP-18-1-I-PPKE-46). This work was supported in part by the grant EFOP-3.6.2-16-2017-00013.

References

[1] T. Nishi, M. Ando, and M. Konishi. Distributed route planning for multiple mobile robots using an augmented lagrangian decomposition and coordination technique. *IEEE Transactions on Robotics*, 21(6):1191–1200, Dec 2005.

[2] Thomas Lienert and Johannes Fottner. No more deadlocks - applying the time window routing method to shuttle systems. In *ECMS*, 2017.

[3] Kaspar Schüpbach and Rico Zenklusen. An adaptive routing approach for personal rapid transit. *Mathematical Methods of Operations Research*, 77(3):371–380, Jun 2013.

[4] Stefan Ravizza, Jason A. D. Atkin, and Edmund K. Burke. A more realistic approach for airport ground movement optimisation with stand holding. *Journal of Scheduling*, 17(5):507–520, Oct 2014.

[5] Rolf H. Möhring, Ekkehard Köhler, Ewgenij Gawrilow, and Björn Stenzel. Conflict-free real-time AGV routing. In Hein Fleuren, Dick den Hertog, and Peter Kort, editors, *Operations Research Proceedings 2004*, pages 18–24, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

[6] Björn Stenzel. *Online disjoint vehicle routing with application to AGV routing*. PhD thesis, Technical University of Berlin, 2008.

[7] Ewgenij Gawrilow, Ekkehard Köhler, Rolf H. Möhring, and Björn Stenzel. *Dynamic routing of automated guided vehicles in real-time*, pages 165–177. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

Using modified MANET protocol in emergency networking

V. Szucs[1], M. Wassouf[2]
[1,2] University of Pannonia, Faculty of Information Technology,
Department of Electrical Engineering and Information Systems
[1]szucs@virt.uni-pannon.hu
[2]mahmoudwassouf@gmail.com
10 Egyetem street, Veszprém, 8200 Hungary

Abstract: **Today, without wireless network communication, everyday life is almost impossible. The use of mobile devices is age-independent, so the issue of usability has become an important aspect. Assistive technologies are becoming more and more influential in the latest developments in information technology. With the use of mobile phones, geographic areas where wired telephony was not feasible were largely covered. However, there are still areas where, in the absence of mobile coverage, it may be a problem to launch a first aid phone. This article presents the results of testing a new network communication protocol modification, which utilizes the ad-hoc communication capability of the devices to provide a good theoretical solution to the above deficiency.**

Taking advantage of the Cluster-Based Routing Protocol (CBRP) and the Backup Cluster Head Protocol (BCHP), the use of the newly developed communication method is continuous, replacing the missing central nodes in a short time, showing a new appropriate node within the cluster that is moving.

Introduction

Although the number of mobile devices is increasing year by year, communication networks offer many opportunities to quickly and efficiently resolve a weather-related emergency, and there are some popular hotspots, but geographic areas that are simply unavailable on either wired or mobile networks across. We could almost mention a place that is unattainable for 'technology', not just in Hungary.

It is worth finding a solution that allows expanding the access to the mobile network through the Wi-Fi network connection of devices.

Background

We can define a mobile ad-hoc network (MANET) as a group of similar mobile nodes that have no central administration or any existing infrastructure, but they have the ability to move dynamically within a specific area. In a nutshell, MANETs are dynamic, self-adaptive and unpredictable wireless networks [1].

Because of this special nature of MANETs, it makes them useful to be used in case of emergency or rescue operation networks because in that case some network services such as message routing and event notification are needed very fast. The network architecture consists of a mixture of leader nodes and normal nodes that are mobile. The leader nodes are the hierarchically superior policemen, firemen, etc. that supervise the emergency operation, while the mobile nodes are normal emergency workers operating in the field of the disaster [2].

MANETs routing protocols are divided into: Proactive (Table driven), Reactive (On demand) and hybrid routing protocols.

MANET hierarchical routing is suitable for emergency networks because it simulates the way that these networks work. Here the network is divided into subsets of nodes called clusters, and in each cluster the best node is elected to be the cluster head. An example of this type of protocols is Cluster Based Routing Protocol (CBRP) [3]. other nodes called gateway nodes are elected to enable the cluster head to communicate with cluster heads in other clusters.

R. Torres, L. Ensico and L. Mengual suggested a new hierarchical model for cluster-based routing protocol using a redundant cluster head called BCHP [4].

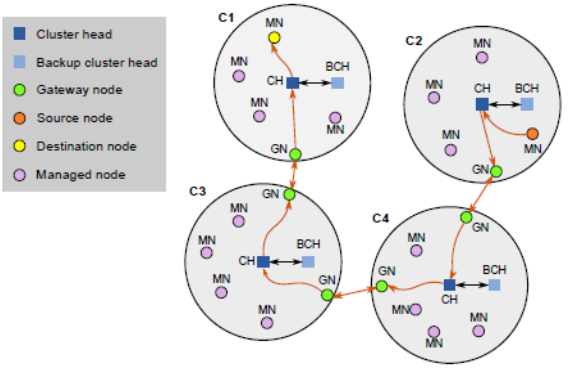


Figure 1. MANET network management model by Rommel et al. [4]

For each cluster head (CH), a redundant node called the backup cluster head (BCH) is elected to take place of CH in case of failure happens to this node. There is continuous periodical synchronization between CH and BCH to keep updated information.
The BCH node here is working with reactive behavior, and to increase the network availability we will make this node to work with a proactive behavior instead by taking into consideration the main node's resources.

Method

The method is based on CBRP taking into consideration the advantages of BCHP by using a redundant cluster head.
Choosing the cluster head will be according to the weighted clustering algorithm (WCA) [5]. The algorithm considers four parameters for the node when electing the cluster head and backup cluster head: degree difference, distance summation, mobility and remaining battery power.
The best two nodes in each cluster are being selected depending on weight value for each node. The node with the first minimum weight is the cluster head and the node with the second minimum weight is the backup cluster head.

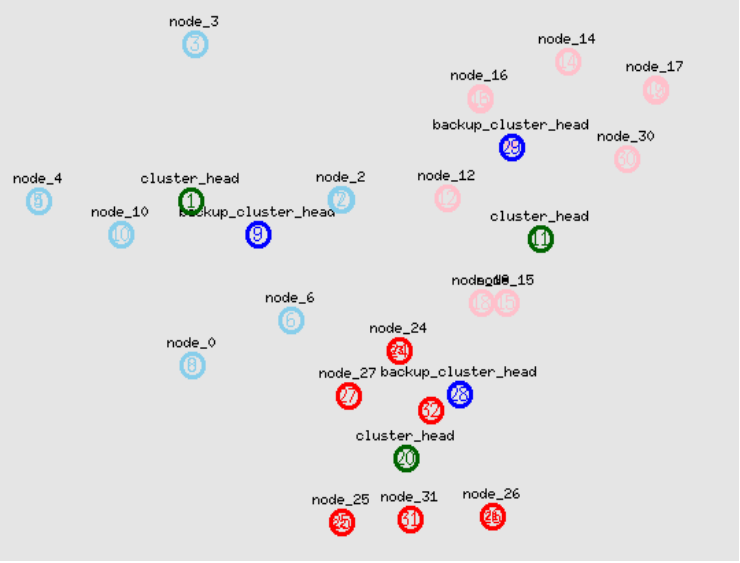


Figure 2. MANET clustering final state

The algorithm ends when all the nodes within the cluster define their states: CH, BCH, or member node.
Figure 2. shows the final clustering state for manet network using network simulator NS2 [6], with 3 clusters.
the new strategy to make BCH working with proactive mode will be as following:

1. Exchanging information about resources level between CH and BCH within synchronization messages.
2. When resources level in CH is lower than the determined level:
   2.1 BCH will send a broadcasting message to the member nodes informing about the new state using the last update information synchronized with the CH.
   2.2 The member nodes will identify the BCH as a new CH, as they have reference information of it.
   2.3 The cluster head will call the algorithm again to choose a new BCH.

This strategy will enhance the availability and convergence of the network more by improving the rate of sent packets, average delay and packet delay variation [7].

Testing the modified MANET protocol

For the simulation, the network simulator Ns2 is being used. The parameters used to define the simulation scenario are listed in Table 1.

Parameters	Values
Simulation area	1000m * 500m
Mobility Model	Emergency and rescue
Number of nodes	25,30,40,50,60,70,80,90,100
Simulation time	200 seconds
Network layer protocols	CBRP, BCHP, Modified BCHP
Traffic Type	TCP, Constant Bit Rate (CBR)
MAC layer	802.11
Radio propagation model	TwoRayGround
Type of antenna	Omnidirectional
Energy Model	EnergyModel

Table 1. Simulation general Parameters

To determine the energy level of the node, an energy model is used which its components are:

- InitialEnergy: represents the initial energy level of the node at the beginning of simulation.
- txPower: the energy consumed for transmitting the packets.
- rxPower: the energy consumed for receiving the packets.
- idlePower: the energy consumed in the node in idle state.
- sleepPower: the energy consumed in the node in sleep state.

the initial energy in Joules while the powers are in Watts where:

energy (Joules) = Power (Watts) * time (Seconds)

Simulation Results:

A. Rate of sent packets: it is the ratio between the number of sent packets and the number of received packets.

As we can see in Figure 3. the modified protocol performs better than BCHP and CBRP. The cluster maintenance is reduced and less broadcasting packets will be sent within the cluster due to the proactive mode of the backup cluster head comparing to BCHP. The modified protocol provides a more stable formation of the cluster.

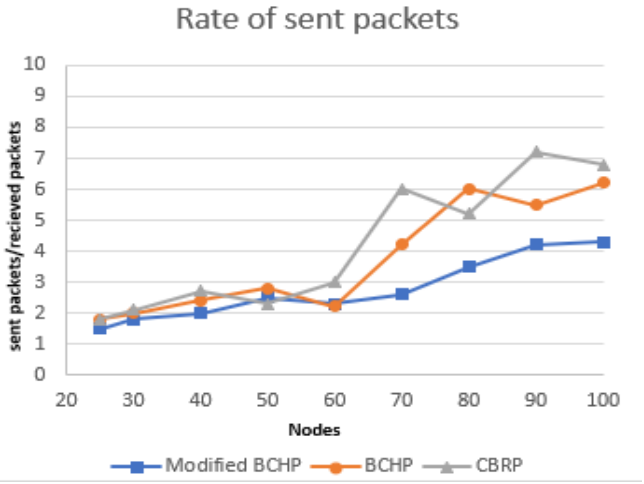


Figure 3. Rate of sent packets

B. *Average Delay:* it is a very important measurement because in rescue and emergency we need network services such as exchange information to be provided as fast as it is possible such as in extreme emergency scenario.

The modified protocol provides better and more stable average delay comparing to BCHP and BCHP as we can see in Figure 4.

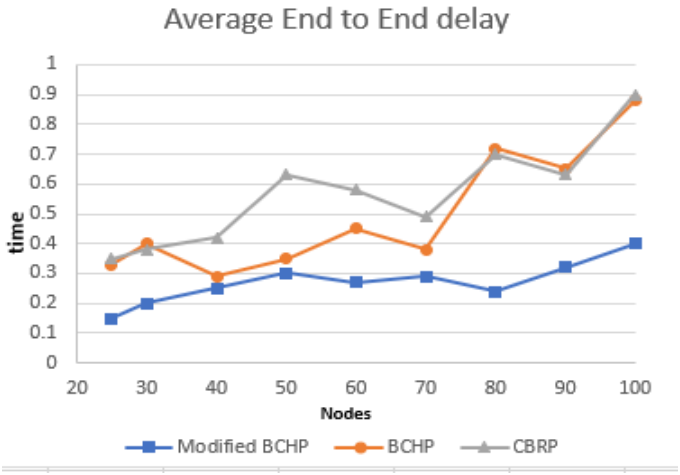


Figure 4. End to End delay

Conclusion

A new modification of a hierarchical routing protocol has been implemented with a proactive behavior mode for the BCH. It is shown that this modification has improved the availability of the MANET network and two network measurements were compared between three hierarchical protocols: CBRP, BCHP, and the modified BCHP. The results have shown that the modified protocol is more stable and it improves the availability of MANET, which is very important to meet the needs of emergency and rescue operation scenarios.

Acknowledgement

We acknowledge the financial support of Széchenyi 2020 programme under the project No EFOP-3.6.1-16-2016-00015.

References

[1] R. Sharma: Mobile ad hoc networks—A holistic overview, 2012, 52(21), 31–36.
[2] E. A. Panaousis, A. Ramrekha, K. Birkos, C. Papageorgiou, V. Talooki, G. Matthew, C. T. Nguyen, C. Sieux, C. Politis, T. Dagiuklas, J. Rodriguez: A Framework supporting Extreme Emergency Services, ICT-MobileSummit 2009 Conference Proceedings, Paul

Cunningham and Miriam Cunningham (Eds) IIMC International Information Management Corporation, 2009, ISBN: 978-1-905824-12-0
[3] M. Jiang, J. Li, and Y.C. Tay: Cluster Based Routing Protocol, IETF Draft. 1999.
[4] R. Torres, L. Ensico, L. Mengual: Improving availability of mobile networks using a cluster routing protocol with redundant cluster head. Third Conference on Theoretical and Applied Computer Science Oklahoma State University Stillwater, February 17, 2012
[5] M. Chatterjee, S. Das and D. Turgut: WCA: A weighted clustering algorithm for mobile ad hoc networks, Cluster Computing Journal, Vol. 5, No. 2, pp. 193–204, 2002.
[6] The Network Simulator NS-2. URL http://www.isi.edu/nsnam/ns/
[7] V. Szücs, M. Wassouf: Modified MANET protocol to extend the emergency network.

Designing gamified virtual reality applications with sensors – A gamification study

Tibor Guzsvinecz[1], Veronika Szucs[2], Cecilia Sik-Lanyi[3]
Department of Electrical Engineering and Information Systems, University of Pannonia
8200, 10 Egyetem street, Veszprem, Hungary
[1]guzsvinecz@virt.uni-pannon.hu, [2]szucs@virt.uni-pannon.hu,
[3]lanyi@almos.uni-pannon.hu

Abstract: **In this paper, the term gamification is established in detail while also gamifying the use of popular sensors, namely the Kinect sensors and the Leap Motion Controller to increase the Human-Computer Interaction between the human and the machine. Doing so, this paper presents a design-process of a virtual reality application using gamified elements.**

Introduction

The popularity of virtual reality is growing every day as the power of information technology increases. Virtual reality (VR) consists of a synthesized environment where the user can move or interact with objects – just like in reality. The goal of VR is to immerse the user. To achieve this goal of immersion, the developers of virtual worlds can use gamification and different sensors.

The term "gamification" can be defined as “using game design in non-game contexts” and it did not gain popularity until late-2010 [1], however the researchers of Human-Computer Interaction (HCI) tackled this idea in the past by making games more enjoyable by designing interesting, enjoyable user interfaces [2,3]. Correctly gamifying applications can modify human behavior [4], e.g. it can motivate people which is considered good when learning [5], during rehabilitation [6], using assistive technologies [7], playing serious games [8], et cetera. Following this thought, gamification can be a part of HCI as design is also a part of HCI [9].

HCI is a critical factor when designing applications with gamified elements. HCI is a multidisciplinary field, and in the center of it stand both the human and the machine. To make their interaction between themselves easier, a motion tracking sensor is needed. Motion tracking sensors are classified into different classes [10], but the cheaper – and widely used – sensors came from the marker-free, visual based class. Such sensors are the

Kinect sensors and the Leap Motion Controller (LMC). These sensor classes can be used with a computer. However, gamification and sensors can be found in the world of smartphones as well. For mobile application variants, the accelerometer and the gyroscope of the smartphone are used.

According to Deterding [11], the rise of these cheap sensors played a great role in increasing the number of applications designed with gamification. Since gamification is also about motivating the users, their sense of presence inside the application is also important. Bogicevic et al. [12] determined that when the users are in a virtual environment inside virtual reality, their sense of presence is increased. Schnack et al. [13] concluded that using motion tracking sensors inside a virtual environment also increases the sense of presence of the users. According to these studies and the mentioned sense of presence, if possible, gamification should be done with sensors inside virtual reality. As the authors find this fact interesting, this papers summarizes the common design elements when designing and developing virtual reality applications with gamification elements.

This paper is structured as the following: The next section shows what the designers should note when designing applications and the authors present five phases how to design and develop applications with gamification elements. After that, the next section is about common gamification elements. Then, since sensors are important when talking about gamification, a section about sensors follows. Lastly, conclusions are made by the authors.

Designing gamified applications

The first step is choosing the design process as with every application in the field of information technology. However, the design process is extended with a few criteria from the usual when using gamification. The developers have to keep in mind that when designing gamified applications, it has to be a user-centered design process [14]. According to Cavalier, the following should be noted [15]:

1. Gamified applications teach the user. The lesson which will be taught by the application should be defined first.
2. A concept should be chosen to accompany the lesson.
3. The target audience should be chosen. The new generation has different digital skills than the generations before [16].
4. The rule system should be chosen for the application.
5. Social interactions should be created.

6. Finally, feedback has to be given to the user. This is also a crucial step as it allows for motivation.

Choosing the target audience is not as easy as it seems. The designers have to know the goals of the user as illustrated in Figure 1. There are users who like to compete, collaborate, explore or express themselves. It is also good to know the genre-preferences of the target groups beforehand [17]. This was also a study made by the authors during the ISG4Competence project.

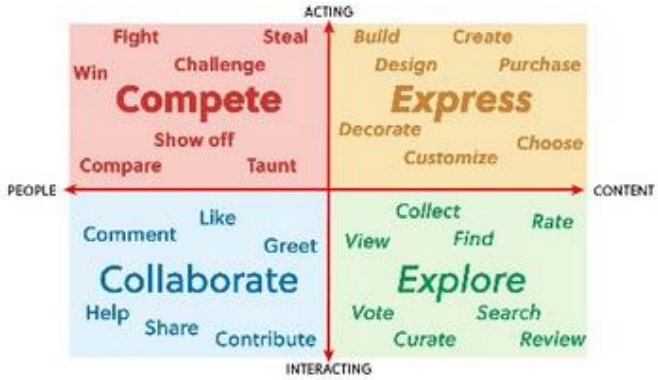


Figure 1.: User types according to Kim [18].

Not only the genre-preferences of the users should be taken into account, but the motivations of the users as well. According to Ryan and Deci there are two motivation types: Intrinsic and extrinsic [19]. In intrinsic motivation people are motivated because they like doing the said thing. In extrinsic motivation however, people are motivated because e.g. there are awards in sports contests or there is parental pressure on them. Even though these are two different motivation types, there are some occasions where extrinsic rewards can be of motivation to intrinsic people. Taking these types of motivations into account, designing the application can be much easier.

The authors – due to their previous works [20,21] and these mentioned design elements – propose five phases when designing applications with gamification elements. When proposing these phases, the authors took the advantages of the waterfall, incremental, iterative and evolutional design methods:

1. In the first phase, the investigation of the environment and the assessment of the user and client needs take place. It consists of assessing the claims of the customers, surveying the needs of the

users who are in different target groups than the customers and analyzing the adaptation skills.

2. In the second, the analysis of reception skills has been created. This allows for the developers to see whether it is worthy to do further analyses or other development steps,
3. The third consists of the gathering of requirements and the graphical user interface (GUI) planning. According to the feedback of the users, the designers and developers can ascertain about the applicability, perspicuity and visibility of the GUI.
4. The fourth is change-tracking based development. This is the phase when the designers or developers consult the target groups. If the target groups disapprove of the application, the designers or developers have to make changes to the application and start over from phase 1. Otherwise, phase 5 begins.
5. The last phase is made-up from modularization, migration and other follow-up methods. In this phase the optimization for the newly developed target hardware, the migration of interfaces related to the application are done. This phase also serves as an opportunity for the following of further lifecycles of the finished application and other actions if they are required.

Also, while taking continuous feedback from the user, the gamified application slowly starts to take shape. Besides testing and fixing bugs, what remains is making a right difficulty curve. This is where the "Flow theory" by Csíkszentmihályi comes into place [22], illustrated in Figure 2.

According to the Flow theory of Csíkszentmihályi, there exists a "Flow

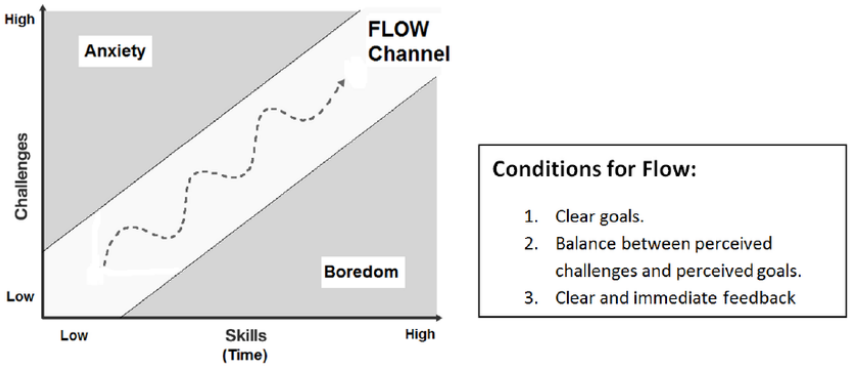


Figure 2.: Flow theory. Illustration adapted from [23].

channel". The feelings of user should stay in the flow channel, with occasional steps into the "Anxiety" and the "Boredom" stages. The time of the steps in these stages should not be too long or otherwise negative effects arise, meaning the user gets demotivated. With the help of continuous testing, feedback and the flow channel it is possible to design a correct difficulty for the user.

Common elements of gamification

There are multiple gamification elements to have the user at the edge of their seats. The most common ones are listed in this section.

- Correct difficulty – or multiple difficulty options,
- Multiple languages,
- Time-based challenges,
- High score list,
- Congratulation sounds,
- Playing the game with friends,
- Open world exploration,
- Story in the application,
- Fun interaction.

Sensors in gamified applications

In the last section the "fun interaction" was purposely left as the last list element. Normally, the users control applications on the computer with a keyboard and a mouse. On smartphones, the applications are controlled by tapping on the screen. When talking about serious games for learning or for rehabilitation, these are either bad design choices or boring gameplay mechanics. The former is a much more serious problem.

When designing applications for people with disabilities who need rehabilitation, the use of a sensor is necessary. However, when choosing the right sensor for the application, it can be considered as another design choice. Their tracking capabilities and precision should be considered. A few of their key specifications can be seen in Table 1.

Table 1: Comparison of the Kinect v1, v2 and the LMC.

	Kinect v1	**Kinect v2**	**Leap Motion Controller**
Dimensions	27.94cm x 6.35cm x 3.81cm	24.9cm x 6.6cm x 6.7cm	7.874cm x 3.048cm x 1.27cm
Tracking	2 depth camera, IR emitter	2 depth cameras, IR emitter	2 cameras, 3 IR LEDs
Body tracking	Full body	Full body	Hand tracking only
Field of view	57° horizontal, 43° vertical	70° horizontal, 60° vertical	150° horizontal, 120° vertical

Raw data access	Available	Available	Available in recent versions

For more of their specifications and other important information, please see [24]. When choosing a sensor for the job, two of their most important features are accuracy and precision. It should also be noted that the precision of the Kinect v1 decreases quadratically when increasing the range of the user from the sensor. Increasing the range with the Kinect v2 does not provide a mathematical behavior, but the precision still decrease, however they decrease less than with the Kinect v1. Also, the noise error distribution is different for the horizontal and vertical axes and the distance is also a key factor when analyzing the noise. The LMC is much accurate than both Kinect sensors with its accuracy averaging at ±1.2mm, but it can only track the hand motion of the user.

Regarding these sensors, the state of their literature is rich, but when used with gamification the literature shrinks. Gamification was used in e-learning [25], in fitness applications for smartphones [26], building engineering 3D arts in virtual reality [27] and in health-related contexts [28].

Conclusion

Gamification plays an important part in HCI and in virtual reality applications. Depending on the target groups, the look and feel of the gamified elements can change, but their aims stay the same: To keep the user engaged in the application. To achieve a good engagement, correct design choices have to be made by the developers with the continuous feedback of the user who stands in the center during development. Inappropriate design elements, such as choosing wrong target groups, poorly chosen input devices, bad rule systems or even wrong difficulties can demotivate the users which could end in them exiting the application.

During the ISG4Competence project, the authors developed six games at the University of Pannonia with the five phases mentioned. The games were completed one and a half year before the end of the project (which was three years long). This method navigates the development from an entirely new viewpoint which is made both from the customers' and users' point of views. In all phases, it allows the further actions to be based on the users' feedback. It also considers the long-time lifecycle of the developed application. During all gamified application developments, the authors initiated the process predominantly with the planning and implementation of graphic interface to suit the principles of structured analysis and development. This yielded the result that the customer and the user knew how it functions from the first steps, they also understood the changes and

became the generators and controllers of the changes. The source of their satisfaction is the continuously received product.

In conclusion, designing gamified elements in an application is an important task, it can change the design process of applications and should be done carefully with the needs of the target groups in mind.

Acknowledgements

The authors would like to thank the support of the "Intelligent Serious Games for Social and Cognitive Competence" (ISG4Competence) project (2015-1-TR01-KA201-022247) for the design choices and the financial support of Széchenyi 2020 under the EFOP-3.6.1-16-2016-00015.

References

[1] Gamification - Google Trends. https://trends.google.com/trends/explore?date=all\&q=gamification (Last accessed on 01.04.2019)

[2] Malone, T.: What makes things fun to learn? Heuristics for designing instructional computer games. Proc. 3rd ACM SIGSMALL symposium, ACM Press (1980), 162-169.

[3] Malone, T.: Heuristics for designing enjoyable user interfaces: Lessons from computer games. Proc. 1982 conference on Human factors in computing systems, ACM Press (1982), 63-68.

[4] Richard N. Landers, Elena M. Auer, Andrew B. Collmus, and Michael B. Armstrong, Gamification Science, Its History and Future: Definitions and a Research Agenda. Simulation & Gaming. Volume: 49 issue: 3, page(s): 315-337. 2018.

[5] S. Chin, "Mobile technology and Gamification: The future is now!" 2014 Fourth International Conference on Digital Information and Communication Technology and its Applications (DICTAP), Bangkok, 2014, pp. 138-143. doi: 10.1109/DICTAP.2014.6821671

[6] Sik Lányi C. Multimedia Medical Informatics System in Healthcare, In Intelligent Paradigms for Assistive and Preventive Healthcare, Volume 19, Ichalkaranje, A., et al. (Eds.), Springer Berlin Heidelberg, New York, pp 39-91 (2006)

[7] Wojciechowski, A. & Al-Musawi, R.: Assisstive technology application for enhancing social and language skills of young children with autism. Multimedia Tools and Applications (2017) 76: 5419. https://doi.org/10.1007/s11042-016-3995-9

[8] Skiada R., Soroniati E., Gardeli A, Zissis D. EasyLexia: A Mobile Application for Children with Learning Difficulties. Procedia Computer Science 27: 218-228, 2014.

[9] Chakraborty, Biplab & Sarma, Debajit & Bhuyan, Manas & MacDorman, Karl. (2017). A Review of Constraints on Vision-based Gesture Recognition for Human-Computer Interaction. IET Computer Vision. 12. 10.1049/iet-cvi.2017.0052.

[10] Huiyu Zhou, Huosheng Hu. Human motion tracking for rehabilitation—A survey. Biomedical Signal Processing and Control 2008, 3(1), Pages 1-18, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2007.09.001.

[11] Deterding, S. (2012). Gamification: designing for motivation. interactions, 19(4), 14-17.

[12] Bogicevic, V., Seo, S., Kandampully, J. A., Liu, S. Q., & Rudd, N. A. (2019). Virtual reality presence as a preamble of tourism experience: The role of mental imagery. Tourism Management, 74, 55-64.

[13] Schnack, A., Wright, M. J., & Holdershaw, J. L. (2019). Immersive virtual reality technology in a three-dimensional virtual simulated store: Investigating telepresence and usability. Food Research International, 117, 40-49.
[14] Preece E., Rogers Y, Sharp E. Interaction Design, Beyond Human Computer Interaction, John Wyley and Son New York, 2011.
[15] Gamification, or How Playing Helps You Learn. https://www.codingame.com/blog/gamification-playing-helps-learn/ (Last accessed on 01.04.2019)
[16] Pérez-Escoda, Ana and Castro-Zubizarreta, Ana and Fandos-Igado, Manuel: Digital Skills in the Z Generation: Key Questions for a Curricular Introduction in Primary Schoolq. Comunicar, vol. 24, n. 49, pp. 71-79. 2016.
[17] Kordigel Aberšek M., Aberšek B., Kosta Dolenc K., Sik-Lanyi C., Shirmohammadi S., Van Isacker K., Petya Grudeva P., Veronika Szucs V., Guzsvinecz T.: Intelligent Serious Games for Learning in Informal Learning Environment, 2nd International Scientific Conference on Philosophy of Mind and Cognitive Modelling in Education, Maribor, Slovenia May 7-8, 2018.
[18] Amy Jo Kim: Game Thinking: Innovate smarter & drive deep engagement with design techniques from hit games. 2018.
[19] Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. Contemporary educational psychology, 25(1), 54-67.
[20] Sik-Lanyi, C., Shirmohammmadi, S., Guzsvinecz, T., Abersek, B., Szucs, V., Van Isacker, K., & Boru, B. (2017, September). How to develop serious games for social and cognitive competence of children with learning difficulties. In 2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) (pp. 000321-000326). IEEE.
[21] Sik, lanyi C ; Szucs, V. Motivating Rehabilitation Through Competitive Gaming pp. 137-167. , 11 p. In: E, Vogiatzaki; A, Krukowski (szerk.) Modern Stroke Rehabilitation through e-Health-based Entertainment Cham (Svájc), Svájc: Springer International Publishing, (2016)
[22] Nakamura, J., & Csikszentmihalyi, M. (2014). The concept of flow. In Flow and the foundations of positive psychology (pp. 239-263). Springer, Dordrecht.
[23] Morales, Juan & Prince, Michael. (2014). Second Year Enhancements to a Summer Faculty Immersion Program.
[24] Guzsvinecz, T., Szucs, V., & Sik-Lanyi, C. (2019). Suitability of the Kinect Sensor and Leap Motion Controller—A Literature Review. Sensors, 19(5), 1072.
[25] Muntean, C. I. (2011, October). Raising engagement in e-learning through gamification. In Proc. 6th International Conference on Virtual Learning ICVL (Vol. 1).
[26] Lister, C., West, J. H., Cannon, B., Sax, T., & Brodegard, D. (2014). Just a fad? Gamification in health and fitness apps. JMIR serious games, 2(2).
[27] Villagrasa, S., Fonseca, D., & Durán, J. (2014, October). Teaching case: applying gamification techniques and virtual reality for learning building engineering 3D arts. In Proceedings of the second international conference on technological ecosystems for enhancing multiculturality (pp. 171-177). ACM.
[28] Pereira, P., Duarte, E., Rebelo, F., & Noriega, P. (2014, June). A review of gamification for health-related contexts. In International conference of design, user experience, and usability (pp. 742-753). Springer, Cham.

Walking Warrior

Gergo Laszlo Proszenyak[1] Adrian Arvai[2], Cecilia Sik-Lanyi[3] Adam Czank[4], Csaba Simon[5], Ferenc Revesz[6], Arpad Kelemen[7], Shannon Cerbas[8], Barbara van De Castle[9], Yulan Liang[10],
[1,2,3,5,6]University of Pannonia Faculty of Computer Science Department of Electrical Engineering and Information Systems
8200 Veszprem, Egyetem street 10.
[4]Budapest University of Technology and Economics
[7,8,9,10]University of Maryland, Baltimore
655 W. Lombard St., Baltimore, MD, 21201
[8,9] Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins Hospital
1800 Orleans St, Baltimore, MD 21287
[1]proszenyak.gergo@gmail.com [2]adrianarvai94@gmail.com,
[3]lanyi@almos.uni-pannon.hu, [4]adam.czank1221@gmail.com,
[5]19simoncsaba98@gmail.com, [6]ferrier.revesz@gmail.com
[7]kelemen@umaryland.edu, [8]scerbas@umaryland.edu, swolfe12@jhmi.edu,
[9]vandecastle@umaryland.edu, bvande1@jhmi.edu, [10]liang@umaryland.edu

Abstract: **The goal is to create a mobile health game, which belongs to the "serious game" category. In serious games, the goal is not only to entertain, but to develop skills, learning, healthy behavior, and social and environmental success for the players. In our game the purpose is to motivate bone marrow transplant cancer patients to walk. By walking more, they earn tokens that allow them to play higher difficulty levels of the game. The game is designed to be similar to the popular *Candy Crush* game. The game will be used by patients in the Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins Hospital, Baltimore, MD.**

Introduction

The goal of this project is to design and develop a mobile application for bone marrow transplant cancer patients, to make their recovery more bearable. After treatment they are instructed to try to get up and walk more, helping their recovery.

There are many types of cancer treatments. The types of treatment a patient receives depend on the type of cancer he or she has and how advanced it is. Some people with cancer will have only one type of treatment. But most

people have a combination of treatments, such as surgery with chemotherapy and/or radiation therapy.

Fatigue is one of the most commonly reported symptoms by patients going through a hematopoietic stem cell transplant (HSCT) process [1]. We have to keep in mind that these patients who receive this kind of treatment should continue to be physically active, even if it is hard to maintain. Physical activity (PA) is a key to improve their health and quality of life. It is hard, but important to prevail upon themselves to do some activities, especially walking. The recommended level of aerobic activity is 30 minutes per day [2], five days per week. Unfortunately, adherence to recommended levels of PA is often low in cancer patients. Mobile device applications could be helpful for these patients for several reasons: nowadays almost everyone has a smartphone, it is easy to persuade the patients to use their smartphones because in their free time they are using them anyway. Mobile health applications exist on the market, but they are not suited for cancer patients, because they face unique barriers to engage in the recommended levels of PA such as fatigue, pain, and nausea. [3]

HSCT treatment

HSCT is the transplantation of hematopoietic stem cells, derived from bone marrow, peripheral blood, or umbilical cord blood. It is most often performed for patients with certain cancers of the blood or bone marrow.

A person undergoing HSCT to treat cancer is given some form of chemotherapy to suppress the bone marrow before the transplant, which is specific to the patient's selected treatment protocol. Total body irradiation is also sometimes part of the preparative regimen. There are two types of HSCT transplants; allogenic and autologous. Patients undergoing an allogenic transplant are receiving stems cells from a donor. Patients undergoing an autologous transplant are receiving their own stems cells. In addition to chemotherapy, the preparative regimen for autologous transplants includes mobilization and pheresis. To achieve mobilization, patients receive growth factor injections to produce large number of stem cells into the bloodstream. Once the stem cell count reaches the optimal level, patients undergo pheresis, which is the collection of stem cells from their peripheral blood. In the days following pheresis, patients receive chemotherapy and then the infusion of their stem cells. Due to the suppressed immune system, all HSCT patients receive prophylactic antibacterial, antiviral, and antifungal medication to protect them from infection. [4] At The Sidney Kimmel Comprehensive Cancer Center, most HSCT transplants are done outpatient, but some are

performed inpatient due to advanced age or pre-existing medical complications.

Method

It was ideated and designed as a matching puzzle game with the unique feature of walking to unlock levels in the game. The game play screen includes 10 different tiles, 6 displayed as a cell type, 2 as boosts, and 2 as blocks that make game play more challenging. The different cell types include red blood cells, white blood cells, neutrophils, platelets, stem cells, and nerve cells. Each difficulty level has limited number of "moves" a player can make to reach the goal. Each level also has a different goal that the player must reach to get to the next level. The game contains bonuses for more points and obstacles for added challenge and diversity in game play experience.

Figure 1. The main menu of the game

Figure 1 shows the main menu of the Walking Warrior game. One can play by clicking on the play icon button (triangle); "i" shows information about how to play the game; "Levels" displays all of the game levels; and the foot icon starts the step counter.

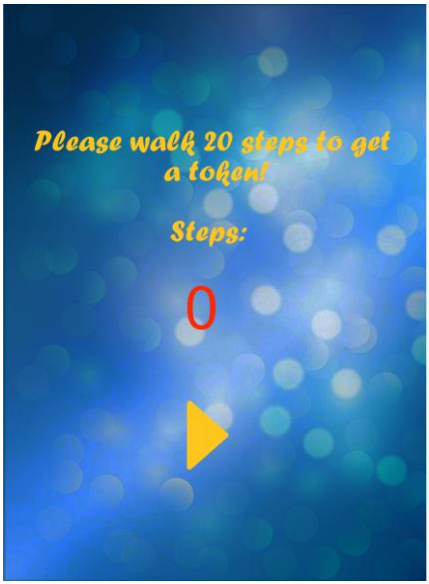


Figure 2. Step counter

Figure 2 shows the step counter. It can be started with the play button. Currently it says in red that zero steps were taken. If the patient steps it starts to count till twenty, then it returns the user to the main menu.

Development

Ideally, all the people with any mobile device should be able to use this game. This is why we decided to implement the game to be playable through any modern internet browser. The game includes animations and movements. Nowadays, one of the best and the most popular ways to implement such an application is to code in JavaScript.

JavaScript (often shortened to JS) is a lightweight, interpreted, object-oriented language with first-class functions, and is best known as the scripting language for Web pages. [5]

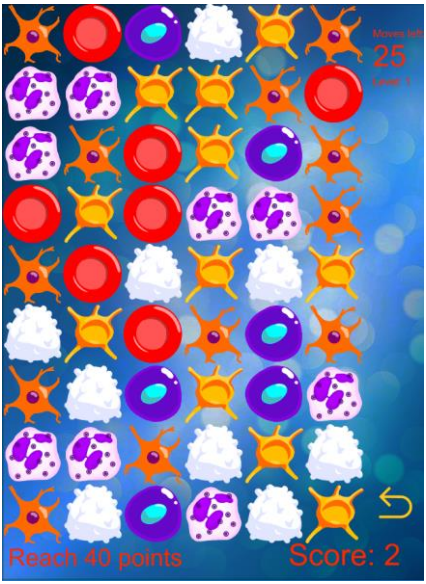


Figure 3. The first level of the game

Figure 3 shows the first level of the game and informations about the goal, the score, etc. The game's main point is to get people to do some exercise, but not the typical way other applications work. The first 3 difficulty levels are considered tutorial levels, it shows how to play the game. After the third level tokens are introduced. The tokens are an alternative to the typical game concept "life". If the player fails to beat the level with the given moves, they lose one token. If they lose all their tokens they can earn more by using the step counter feature of the game. The mobile devices use the built-in accelerometer with an additional piece of open source HTML code that were modified and tuned to count the number steps taken. When the player reached the required number of steps, the game gives them one token and they can continue to play. We developed an admin page with a MySQL database and stored it on a server. Admins and authorized users can check how many tokens the user has and how many steps they made.

Implementation

Separate JavaScript, HTML5, and PHP files were created for this game. The JavaScript files are responsible for the game itself, they manage the certain levels, load the main frame and tiles. HTML5 files count the steps and

rely on the phones' built in accelerometers. The PHP files handle the user logins and access and store the data. The steps and the login credentials are stored in a MySQL database. All the data transfer goes with AJAX, because with it we had an advantage that it is not required to refresh the webpage to send data through PHP, so the user experience is rather smooth. [6]

Conclusion

In summary this game could help many people who are fighting not just with cancer, but any disease, where physical activities is a main part of their recovery. Nonetheless, the treatment could be very exhausting for the patients. They must keep fighting as real warriors. Moreover, the game can be used by anyone who enjoy puzzle games or need motivation to walk.

References

[1] https://www.ncbi.nlm.nih.gov/pubmed/23715701
[2] https://www.nccn.org/patients/resources/life_with_cancer/exercise.aspx
[3] http://www.cancer.ca/en/cancer-information/diagnosis-and-treatment/chemotherapy-and-other-drug-therapies/chemotherapy/?region=on
[4] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4705045/
[5] https://developer.mozilla.org/hu/docs/Web/JavaScript
[6] https://www.w3schools.com/js/js_ajax_intro.asp

Re-Creation, an Android game

Barbara Bodor [1], Patrícia Szabó [2], Cecilia Sik-Lanyi [3]
[1,2,3]University of Pannonia Faculty of Computer Science Department of Electrical Engineering and Information Systems
8200 Veszprem, Egyetem street 10.
[1]bodorbarbi92@gmail.com, [2]szabo.patricia1996@gmail.com,
[3]lanyi@almos.uni-pannon.hu

Abstract: **One of the most popular programming languages today is Java, because it is easy to use, and users can be creative in using it. This is the reason why we have chosen it for planning a new serious game. In this language we have created a funny game for patients, because we think that an application can be helpful and entertaining at the same time. We have created an application for tablet which helps for the patients who have brain damage (for example stroke). This application helps to restore the motor functions of the fingers.**

Keywords: rehabilitation, stroke, application, Java, Android, finger motion

Introduction

From the early 1970s to the early 1990s, the estimated number of noninstitutionalized stroke survivors increased from 1.5 million to 2.4 million. Each year about 795 000 people experience a new or recurrent stroke. [1] The number of people involved in it is very high, and this number increased fast (Figure 1).

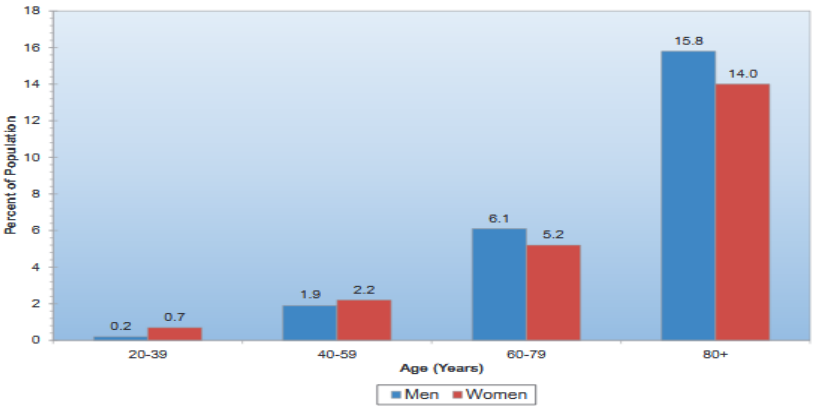


Figure 1. *Prevalence of stroke by age and sex (National Health and Nutrition Examination Survey: 2009–2012). [1]*

The speed of a treatment has a large effect on the person's chance of recovery [2]. So, it is very important to act fast. A good way shown by Figure 2 can be helpful to everybody to recognize the first signs of the stroke.

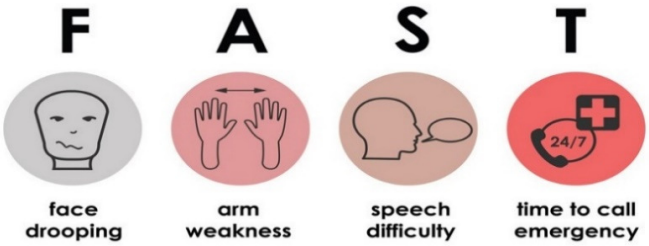


Figure 2. *The steps of the FAST test [2]*

The consequence of a stroke is a really serious problem, and it makes one's life harder. This is the reason why this application has been designed; to help people who suffered stroke to heal faster and go back to normal life sooner. There are promising interventions that could be beneficial to improve the aspects of gait including fitness training, high-intensity therapy, and repetitive-task training. Repetitive-task training might also improve transfer functions. Occupational therapy can improve activities of daily routine. Several large trials of rehabilitation practice and of novel therapies (eg, stem-cell therapy, repetitive transcranial magnetic stimulation, virtual reality, robotic therapies, and drug augmentation) are underway to inform future practice. [3]

There are other lots of methods that can help patients all over the world. We used the Web of Science (WoS) database in our research, WoS was queried. Search words included "serious game" AND "stroke". A list of 28 records were identified. These papers were published in 2018. Of course, there are several applications in the Science Direct, PubMed other databases. Moreover, some games were developed at the University of Pannonia [4-12] in the last decade.

Methods

We have downloaded the most relevant available literatures from the above-mentioned scientific database. Based on the new results and on our earlier experience we have designed a new application for tablets, because we needed bigger surface and because of the touching function it was a good opportunity to combine the technology with the therapy. The idea of the new game was also influenced by a game for children [13].

We used one of the most popular programming languages, Java and we have created the application for android platform. It is easy to use, and developers can create a colorful, amazing game in it. That is the main reason why we have chosen this solution.

Java programming language is really classical. It was used first in the Web and

Sun's HotJava browser appeared in 1995. Since then there has been a lot of fields where we can apply it. Java is used in business to develop enterprise, or middleware, applications such as on-line stores, transactions processing, and dynamic web page. Java is also a good choice for small platforms such as smart cards, phones, and PDAs.

Java provides a number of advantages: platform independent, object-oriented, threading, simulation tools, networking, interface and enhancing legacy code. So, there are several features of Java that makes it a powerful and highly secure language [14].

Game description

Figure 3 shows the start icon of the game.

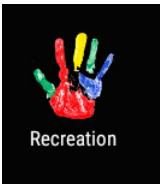


Figure 3. *The start icon of the game.*

After clicking the start icon, the opening screen appears (Fig. 4) at first. The Re-Creation game is a set of 4 games: "Piano", "Fly away", "Drawing" and the "Bonus" game (shopping) as it is shown in Figure 5.

Figure 4. *The opening screen of the game.*

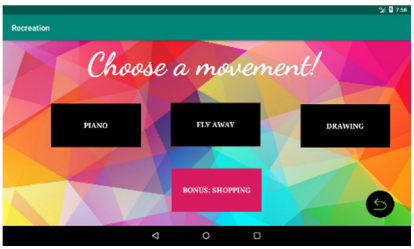


Figure 5. *The main menu of the game.*

1st game: The "Piano" game

The Piano game is an imitation of 4 keys of a piano keyboard. Moreover, the distance of the keys is adjustable: small, medium and large (Fig. 6). Figure 7 shows the different distances in the Piano game.

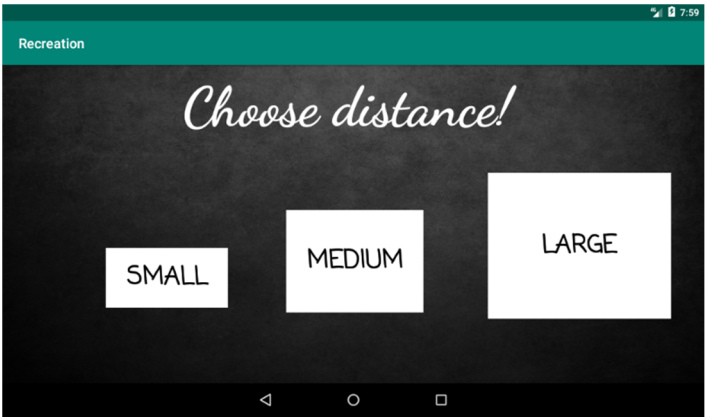


Figure 6. *Distance setting of the piano game.*

Figure 7. *Distance setting in the piano game.*

Each screen has four buttons ("keys"). They are always recreated in random order. The point is shown in the upper left corner (Fig. 7).

Scoring: In case of a correct order when a key is pressed, the user gests plus 1 point and the key turns green.

In case of an incorrect order, the game produces minus 1 point per press. The button changes to red for a few seconds and then turns gray again, then the user can try it again.

The level is finished successfully if the user presses and holds down all 4 keys in the correct sequence. In case of success, all four keys will change to green. After a few seconds, they are mixed again.

The life is visible in the upper right corner (red hearts). Then there will be one life less if the user releases a key after a successful keystroke. The correctness of hand gestures is checked by the following instruction:

```
switch (event.getAction())
{
case MotionEvent.ACTION_DOWN:
case MotionEvent.ACTION_UP:
}
```

Then the result is shown at the end of the Piano game and the user can finish the game.

2nd game: The"Fly away" game

The goal of the "Flay away" game is to shoo a bug (a bee or a ladybird or a cockchafer) away by touching it (Fig. 8). Bugs are randomly generated on the

screen in random position. The aim of this game is to develop reflexes and speed. It has several levels. The user gets one point if they shoo bees away (Fig 8). They have 15 seconds at the 1st level. They get 2 points if they shoo ladybirds away. They have 10 seconds at the 2nd level. They get 3 points if they shoo cockchafers away. They have 5 seconds at the 3rd level. The game finishes if the user shoos every bug away (Fig. 9).

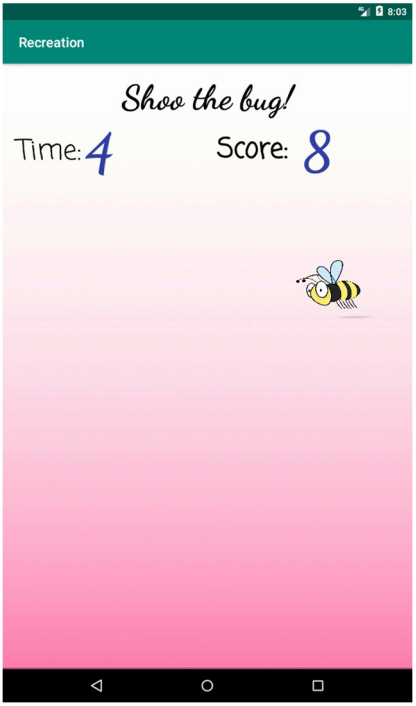


Figure 8. *1st level in the "Fly away" game.*

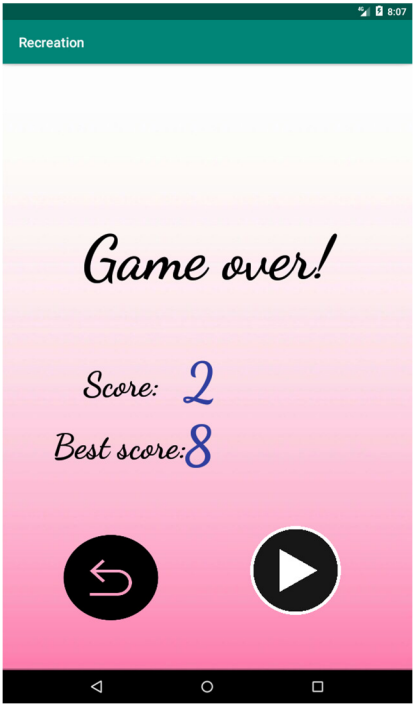


Figure 9. *End of the "Fly away" game.*

The correctness of hand gestures is checked by the onTouchEvent() event.

3rd game: The "Draw" game

The goal of the "Draw" game is to follow lines, geometry shapes e.g. numbers which are visible on the screen, better to say draw those figures. The correctness of hand gestures is checked by the ChoiceTouchListener built-in class with the help of the following events:

```
switch (event.getAction() & MotionEvent.ACTION_MASK) {
case MotionEvent.ACTION_DOWN:
case MotionEvent.ACTION_UP:
case MotionEvent.ACTION_POINTER_DOWN:
case MotionEvent.ACTION_POINTER_UP:
case MotionEvent.ACTION_MOVE:
}
```

4th game: The "Bonus" game

The 4th game is the bonus game which is actually a memory game. It has 3 difficulty levels which are based on shopping lists. 1st level: one type of products should be selected by touching the figures of products which have been shown on the previous screen. 2nd level: the user must also select one type of products, but they should select them based on a written list. 3rd level, the user must select multiple types of products based on a longer lists. The technology is touching a particular point on the screen.

Result

We have achieved great results while we were testing our application. Only the user interface and the main function have been tested.

Firstly, we have tested all functions of the game, moreover it was tested on different smart devices too. After the first test done by students from the University of Pannonia some elderly people with different illnesses (e.g. Alzheimer) have tested the game. The test of the patients will run in April and in May 2019. Those test results will be shown at the conference. Based on the first test we can state:

Playing this game helps everyday tasks become as easy as they were once.

The different gestures help differently:

- Tweezer grip: help to pick up small objects.
- Selective finger touch: help in typing (for example on the phone or on the computer)
- Rotation in the wrist: strengthens the wrist
- Moving one finger: help the precision
- Selective finger circles: strengthens the forefinger
- Selective circles with the thumb: strengthens the thumb

Hopefully patients will prefer to use our application, because they will consider it as an entertaining and a challenging game rather than a painful physiotherapy. It will help more patients not just those who suffer from stroke, but those as well who somehow have lost their motoric function in their fingers for example patients who are diagnosed with Parkinson It can be suitable too for elderly people and children who want to improve the fine motor skills.

Conclusions

We have designed a serios game for tablets. This application will help to restore the motoric function of the fingers for those patients who use it regularly. Even though it is successful, the patients will need the traditional methods as well to achieve the best results during their rehabilitation.

Benefits:

- Easy to use.
- Exciting.
- Sustaining the attention.
- Improve different skills.

Disadvantages:

- It is not enough for the full recovery.
- It is not fully suitable for strengthening the muscle of the fingers

Acknowledgment

The authors would like to thank the financial support of Széchenyi 2020 under the EFOP-3.6.1-16-2016-00015.

References

[1] Heart Disease and Stroke Statistics—2006 - 2016 Update
A Report From the American Heart Association Statistics Committee and Stroke Statistics Subcommittee

[2] https://www.stdavidsfirstaid.co.uk/resources/top-tips/how-to-spot-a-stroke/

[3] C.P. Warlow, J. van Gijn, M.S. Dennis, J.M. Wardlaw, J.M. Bamford, G.J. Hankey, P.A.G. Sandercock, G. Rinkel, P. Langhorne, C. Sudlow, and P. Rothwell, Stroke: Practical Management. 3rd Ed., Wiley-Blackwell Publishing, Oxford. 2008

[4] C. Sik Lányi, Z. Geiszt, and V. Magyar "Using IT to Inform and Rehabilitate Aphasic Patients", Informing Science Journal, vol. 9, pp. 163-179, 2006

[5] M. Horváth, Cs. Dániel, J. Stark, and C. Sik Lányi, "Virtual Reality House of Aphasic Clients", Transaction on Edutainment III, LNCS 5940, pp. 231-239, 2009.

[6] T. Dömők, V. Szűcs, E. László and C. Sik Lányi, "Break the Bricks Serious Game for Stroke Patients", in *Lecture Notes in Computer Science*, K. Miesenberger et al. (Eds.): ICCHP 2012, Part I LNCS 7382, pp. 673-680, 2012

[7] S. Ortmann, P. Langendörfer, and C. Sik Lányi C, "Telemedical assistance for ambulant rehabilitation of stroke patients", BRAIN INJURY, vol 26(4-5), pp. 644-645, 2012

[8] Á. Nyéki, V. Szucs, P. Csuti, F. Szabó, and C. Sik Lanyi, "Gardener Serious Game for Stroke Patients", in *Lecture Notes in Computer Science*, K. Miesenberger et al. (Eds.) ICCHP 2014, Part I, LNCS 8547, pp. 272-275, 2014

[9] C. Sik Lanyi, and V.Szucs, "Games applied for therapy in stroke tele-rehabilitation", International Journal of Stroke, Vol 9 (Suppl. S3) pp:300, 2014

[10] C. Sik Lanyi, V. Szucs, and J. Stark, "Virtual reality environments development for aphasic clients", International Journal of Stroke, Vol 9 (Suppl. S3) pp:241, 2014

[11] C. Sik Lanyi, and V. Szucs, "Motivating Rehabilitation Through Competitive Gaming", in *Modern Stroke Rehabilitation through e-Health-based Entertainment*, E. Vogiatzaki, A. Krukowski (Eds.). Springer-Verlag, pp. 137-167, 2016

[12] M. Yates, A. Kelemen, and C. Sik Lanyi, "Virtual reality gaming in the rehabilitation of the upper extremities post-stroke", BRAIN INJURY, vol. 30(7), 855-863, 2016

[13] Spielen für the Feinmotorik: http://www.fingers-in-motion.de/de/ (in Germen)

[14] JavaTech – An Introduction to Scientific and Technical Computing with Java, Clark S. Lindsey, Johnny S. Tolliver and Thomas Lindbald

Learning to play snake using genetic neural networks

B. Halmosi, C. Sik-Lányi
University of Pannonia, Veszprém, Hungary

Abstract: **In this paper we are creating an artificial intelligence that plays snake. We are using our own framework to train neural networks that are able to play the game and testing the effectiveness of different sensory selections compared to each other. Finally, we test the limits of our program.**

Introduction

Artificial neural networks are frequently successfully applied in optimization problems and pattern classification. They have also been used as artificial intelligence in several different games, for instance chess and GO [1][2].

To find the best strategy for training artificial neural networks has been causing difficulties for a long time [3]. Throughout the last decades we could see many different approaches to solve this, one of them is using genetic algorithms [4][5][6].

Evolutionary algorithms have already been employed in many ways to solve different problems due to their flexibility and speed [7]. Under appropriate circumstances they can be really successful in heuristic optimization. According to Montana and Davis, genetic algorithms are able to outperform backpropagation in some cases [5]. They can find a fitting solution pretty fast, and their random mutation guarantees a wide range of solutions.

In this paper we discuss how effective different sensory selections are at learning to play snake and scoring in the game. We are also comparing them to see which combinations are faster in finding a solution and which ones find better solution after 300 generations.

Game and scoring

Firstly, a little clarification on the rules of our snake game. The game area is ten times twenty, surrounded by walls and there is a unit of food that

respawns at a random location every time when it is eaten. Initially the snake is 3 units long and eating increases the length of the snake by one. If the snake gets in contact with a wall or itself, the game ends.

For scoring, we take the original system of the game and give the snakes 100 points for each piece of food eaten. In addition, we decrease the score by one in each step to prevent them from running in circles – also made some compensation when getting closer to the food.

Framework and setup

The framework itself is written in Python3 using the Numpy library for the matrix operations, which guarantees relative fast computing while using a high-level language [8]. It is an implementation of a fully functional genetic algorithm whose parameters can be varied freely.

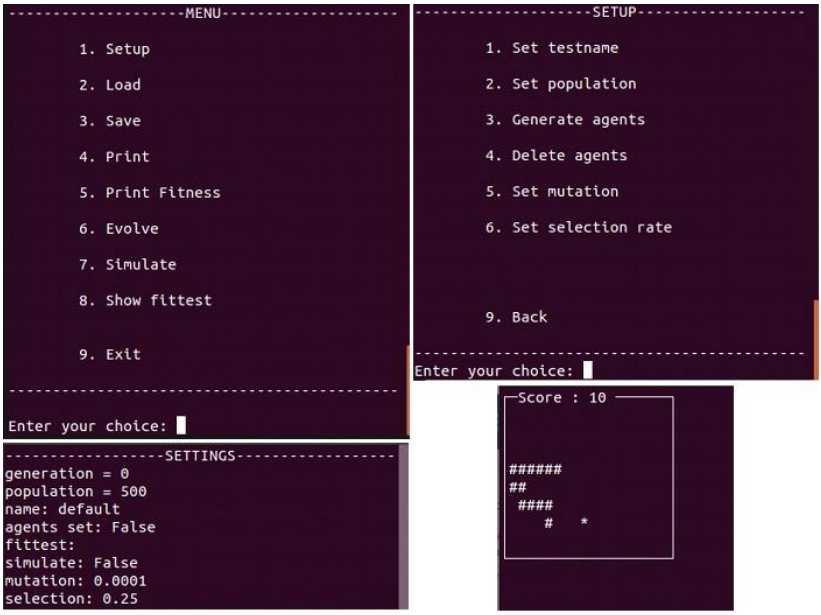


Figure 1: Four screenshots of the framework running on linux.: the main menu (top left), the setup menu (top right), the settings display (bottom left) and the simulation (bottom right).

As usual on genetic neural networks, the agents of the algorithm are simple feedforward neural networks that can be customized. We use two types of feedforward neural networks in the experiments, the 4-4-4 neural network that has an input layer, one hidden layer and an output layer with 4 neurons each, and the 8-8-4 version that has 8-8 neurons on its input- and hidden layers and 4 on the output layer.

The framework can simulate the evolution from generation to generation, the "Evolve" option starts the selection, crossover and mutation phases and calculates the fitness scores. As fitness function we use the scoring method described above and set the parameters of the algorithm as follows:

- Selection is set to 0.5, which means that the worst 50% of the population is removed.
- Crossover is meant to refill the empty spaces left by the selection. It generates a new snake choosing two parents whose genes are inherited with an 50% chance, calculated on each gene separately.
- Mutation is 0.0001 by default, so the genes of the snakes are set to a random value with a 0.01% chance, calculated on each gene separately.

Strategy and sensory selection

Therefore, we see a snake that has only three choices to move in each position and a piece of food to catch. The first thing to do is to choose what senses the snake would get. A sense is a group of input values that the neural networks get as input. We tested four kinds of 4-bit senses on simple 4-4-4 feedforward neural networks:

1. Basic food: It indicates if the food is at least one step away in each direction from the head of the snake.
2. Interpolated food: Each of the 4 inputs represent a direction (up, down, left, right) and their values shows the foods distance in that specific direction scaled between 0 and 1.
3. Basic barrier: It shows whether or not a barrier (a wall or the body of the snake) is one step away in each direction.
4. Interpolated barrier: It is similar to the interpolated food sense, but it shows the distance of the barriers in each 4 directions scaled between 0 and 1.

We also tried the two parts of the barrier separately where the snake sees only its body on 4 bits and where it sees only the walls on 4 bits, but they scored far lower than the 4 four senses mentioned above.

Their effectiveness is shown by Figure 2. The graphs depict the average score as a function of generation. The average score is calculated from the fitness of the agents in each generation, omitting the best ten percent and the worst 50 percent.

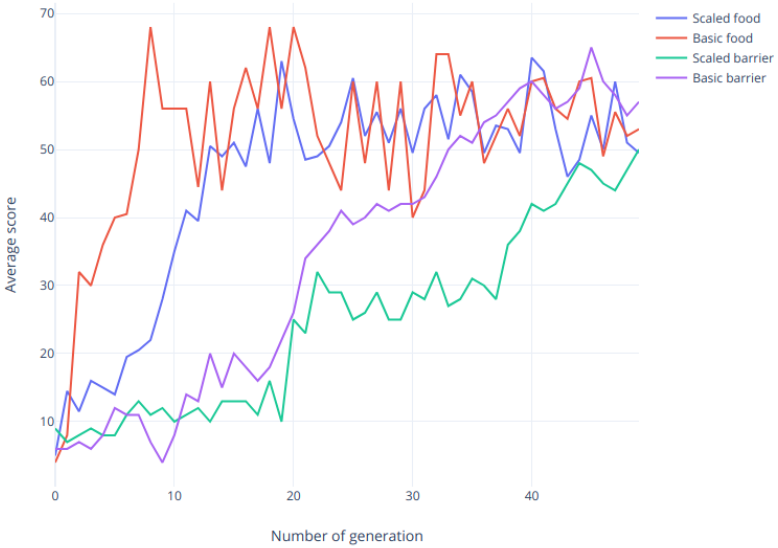


Figure 2: Measured performance of sensory organs.

As can be seen, there are no major differences between basic and scaled pairs. We set the parameters of the genetic algorithm to the same value for the measurements, except that the population is 1000. Basic sensors tend to get to the result sooner, but only with a few iterations, so we decided to test them in pairs combining one food sensor and one barrier sensor in every version. To do this we used 8-8-4 feedforward networks with one sensor on the first 4 input neurons and the other on the last 4.

Figure 3 shows how effective these combinations are. The population was set to 1000 and the other parameters remained by default.

Shown values are the average of five-five measurements, omitting the experiments with the best and worst end results. As we can see, the basic food sensor is performing far better in these combinations than the scaled version.

The basic barrier appears to be better as well, so we shall keep this input combination going forward.

Figure 3: Performance of sensor pairs

Testing the limits

In this section we show how a bit more complex network can learn the game to see the limits of this approach. We stop the algorithm when it does not develop anything over the last 500 generations. Results are as follows.

The maximum average score that we can reach is around 2400 points, which means that the snake usually eats more than 25 but less than 30 units of food in one game. Because of the random food generation, in lucky cases it scores above 40. In different tests it tends to evolve different strategies to play, but it is worth mentioning that it uses almost the same when it is reaching its limits (not on the level of neurons, but its behavior looks similar). During this experiment an 8-8-4 network is applied with one more hidden layer containing 8 neurons (so it is a 8-8-8-4 network).

In most experiments the AI scores around 6-700 points after a few hundred generations, so it gets 8 units of food on average. Almost all of the tested games ended with the snake trapping itself, because it sees only the four fields surrounding its head.

Conclusion

In the beginning, we tested a few senses for their usability, then in pairs to get a more correct picture of their effectiveness. After that we chose the most promising combination and pushed them to their limits.

The AI learned to play the game and is able to catch more than 25 food on average. This seems to be a limit due to the senses. By using different (or more) senses the AI may be able to catch more, but that would also increase the size of the neural networks as well as the time requirement.

As the results show, the interpolated versions of the senses were significantly less effective than the basic binary versions in learning a basic strategy. It would also be interesting to see whether or not other senses with different scaling methods can reach higher scores than these results and if the learning speed can be enhanced by changing the basic senses to different scaled senses after a few hundred generations, but this needs further experimenting.

Acknowledgment

We acknowledge the financial support of Széchenyi 2020 programme under the project No EFOP-3.6.1-16-2016-00015.

References

[1] S. Thrun, “Learning to Play the Game of Chess”, Advances in Neural Information Processing Systems (NIPS) 7, 1995

[2] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel and D. Hassabis, “Mastering the game of Go with deep neural networks and tree search”, Nature, vol. 529, pp. 484–489, 28 Jan 2016

[3] X. Glorot and Y. Bengio, “Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics”, PMLR, 9:249-256, 2010

[4] B. Widrow and M. A. Lehr, “30 Years of Adaptive Neural Networks: Perceptron, Madaline, and Backpropagation”, Proceedings of the IEEE, vol. 78, Issue: 9, 1990

[5] D. J. Montana and L. Davis, “Training Feedforward Neural Networks Using Genetic Algorithms”, Proceedings of the 11th International Joint Conference on Artificial Intelligence, Detroit, MI, USA, 1989

[6] P.P. Palmes, T. Hayasaka and S. Usui, “Mutation-Based Genetic Neural Network", in IEEE Transactions on Neural Networks, vol. 16, Issue: 3, 2005

[7] H.-P. Schwefel and T. Back, “An Overview of Evolutionary Algorithms for Parameter Optimization”, in Evolutionary Computation, vol. 1, pp. 1-23, 1993

[8] T. E. Oliphant, “Guide to NumPy”, 2006
link: web.mit.edu/dvp/Public/numpybook.pdf (2018)

Decision supporting tool for scheduling of production processes considering human factors

Gy. Ábrahám[1], Gy. Dósa[1], T. Dulai[1], Á. Werner-Stark
[1]University of Pannonia, Faculty of Information Technology,
abraham.gyula@virt.uni-pannon.hu, dosagy@almos.uni-pannon.hu,
dulai.tibor@virt.uni-pannon.hu, werner.agnes@virt.uni-pannon.hu
8200 Veszprém, Egyetem Street 10. Hungary

Abstract: **In recent years, the analysis of industrial processes has become important in order to increase the efficiency and safety of the processes. We have developed a method by which we can construct a fault tree, considering human properties. It can be explored by analyzing the factors that lead to the critical major event, errors indicated by humans and their possible combinations. The goal is to assign an activity to the person with the best qualities among the set of the available people. We introduce a decision supporting tool and its application for the appropriate choice of a human resource in case of process scheduling.**

Introduction

Despite the proper theoretical design of business processes, the execution is often not as it is expected. We can change it by knowing the possible effects of the human factor on each step. If these issues are known during the design process, the appropriate parts of the plan can be modified easily and at low cost. The probability of the occurrence of certain events occur due to the human factor involved can be determined by a suitable human error-analysis method, see [1, 2]. For getting optimal time, quality and cost, a useful tool is the fault tree analysis. The fault tree is a graph that displays the various combinations of machine failures, non-independent failures and errors indicated by humans which cause the root event [3].

The quantification during the analysis can be made by estimating the probability of the baseline event directly, based on kinetic theory, using Markov chains or Monte Carlo simulation. During the analysis, we start with a hypothetical error (a major event), and gradually explore other errors that possibly can lead to the occurrence of the event. The Monte Carlo method [4] performs many (thousands or even millions) simulations for statistical accuracy based on elemental input events modeled with random distribution variables. For this reason, Monte Carlo analysis requires a lot of time and computing power; at the same time, the derivation of the deterministic

formula of the analytical method is very complicated in the case of a complex problem, not to mention the determination of specific probabilities [5].

There are many software that support fault tree generation and fault tree analysis, see [6, 7], but they do not take into account human qualities. We developed the method considering some human properties, too.

The most important part of our study is to calculate the probability of the occurrence of a defined root event. The transparent work is supported by the display of the tree structure, which was extended with reliability calculations. The Monte Carlo simulation is used to determine the probability of occurrence. Using the method, the goal is to identify all the errors, their combinations and the causes which lead to the main event. Moreover, we intend to detect particular critical events and event chains, to calculate reliability across the branches of the fault tree, and to make clear and transparent documentation about failure mechanisms.

The main steps of the method are:

1. Demarcation and definition of the subsystems which determine security and reliability.
2. Definition of the undesirable event or events.
3. Mapping the logical relationships between the errors, display them on the fault tree and performing calculations.

When constructing the fault tree, the causes of the events are investigated backwards on the flow chart of the given system, from the event to the cause (deductive analysis). We take an effect at each step and look for one or more events (cause) that will be split. We define the logical relationship between the events as the so-called logical gates that define the combined effect of interconnected events for a higher-level event. This allows us to identify the combination of events which are prevented to avoid the root event. In addition, it is possible to identify all the combinations of the events that can lead to system failure. Based on the results, the weaknesses of the system can be determined, and suggestions can be made to increase safety and reliability. The probability of occurrence of the root event is determined during the analysis. To test this method, we have also developed a software to build the fault tree and to get the tree structure in XML format, which is the input of the Monte Carlo simulation module. The simulation is performed on an XML structure, where the iteration starts from the leaves of the tree and evaluates the nodes in the tree upwards. The result of one iteration shows whether the root event occurred. After performing several iterations, we count the number of cases when the root event occurred in relation to the number of the iterations. By this computation, we can most accurately estimate the probability of the occurrence of the root event (i.e. the error). The goal is to assign the person with the best qualities to the

activity. The “goodness” of this person is measured in that way that the entire process is executed as quickly as possible.

The Monte Carlo simulation module

In the current version of the Monte Carlo simulation module it gets the fault tree as input in XML format. The algorithm requires an appropriate fault tree representation. An example for the fault tree described by XML is show in Figure 1. A node represents a logical connection above any error event (except for the root event), and one error event above any node corresponds to the logical connection. There may be more than two error event nodes connected to a logical node.

For the tree evaluation a convenient representation can be provided if the error events are presented in the nodes. Their parameters are the probability of their occurrence and the logical connection of their children’s events.

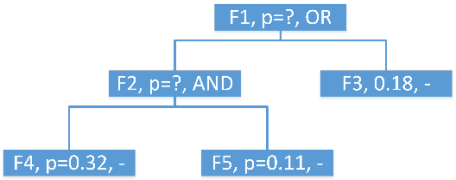


Figure 1.: Representation of fault tree in Monte Carlo simulation

In the beginning, only the leaves have a value of the probability of the occurrence. During an iteration of the simulation, random numbers are generated from the interval [0,1] for each leaf. Based on the random number and the probability of occurrence of the error shown in the leaf, it is possible to determine whether the error occurs. These logical values are propagated upwards, considering the type of the logical connection of the events, and the events that are involved in the logical connection. This evaluation is done from level to level to see whether the root event has occurred or not. If these steps are iterated N times and the root event occurs n times, then the frequency of the root event is n/N, and this frequency is a good estimation of the probability. The p variable contains the probability of the occurrence of the root event and the value gets closer to the reality in case of a high N value.

Aspects used to characterize the performance of a human resource

Aspects had to be selected that allow efficient resource selection and can be extracted from the event log entries to some extent, if available. Five characteristics have been identified, including: 1. expertness, 2. exactitude, 3. viridity, 4. docility, 5. consistency, 6. stability.

The performance characteristics are derived from the event logs. A simple (general) log contains the resource, the performed activity, the activity status (start, end) and the timestamp for the activity. Of course, many other data may be presented in the entry.

Example of two entries with minimal content are:

2018.10.12 8:32:15 - soldering - start - operator1
2018.10.12 8:34:22 - soldering - end - operator1

It is possible to calculate the length of the time that a given activity has taken from the timestamps of start and end status by a simple subtraction. If an activity had to be performed several times on the same or on other resources, a more precise characterization can be given of the working parameters of the given resource(s) by analyzing the log entries.

Requirement to perform activities based on resource properties

In order to allocate a resource that can perform an activity the most efficient way, it has to be known which qualities are the most relevant to perform the activity. For each activity, a fault tree can be built that shows which elementary errors can contribute to cause a failure in the activity.

From the values applied on the tree upwards, in terms of the activity it can be diagnosed which human attribute determines the most the avoidance of the malfunctioning of the activity.

Figure 2 shows a simplified fault tree for the activity "filling in a tax form". The tax form is filled incorrectly if it is filled badly or incompletely. The relevance of the five human qualities for the elementary error cases have to be estimated (on a scale from 1 to 5). These values are propagated upwards in the tree to get the root error.

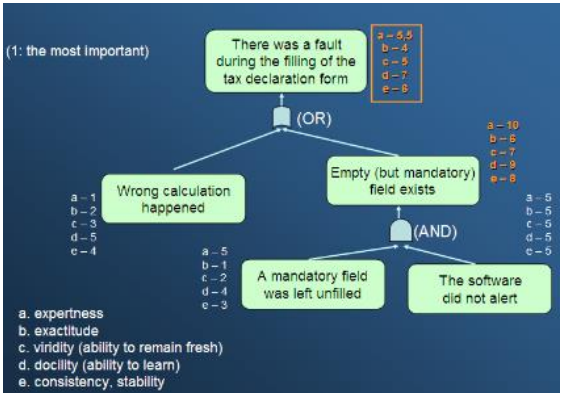


Figure 2.: The process of deriving the relevance values of qualities of human resource for the root event

During the evaluation, we process the fault tree from the leaves to the root. Each quality-relevance value of the human resource at the higher levels is derived from the values of contributor errors of lower levels as follows: in case of logical AND connection the values of the contributing errors are summed up, in case of logical OR connection the values of the contributing errors are averaged. As a result, the values of the relevance of the different aspects of the human resource properties are determined in the root of the fault tree. In Figure 2 it can be seen that – based on the details of the fault tree – accuracy is the most important to avoid errors, the second is viridity, and the third is expertise. Consistency is less relevant, and the docility is the least important property.

Once the operator has built the fault tree for all tasks of the process, provided the relevance of the human properties for the error events of the leaves and the software derived them for each activity, they can be connected to each other to build up the fault tree of the whole process. During this operation, the root events of the fault trees of each activity are logically connected to each other, and the type of the logical connections are determined by the relationship of activities. In the current phase, we deal with two types of relationships. If – based on the BPMN model – between two adjacent activities of the process, there is a

(1) sequential connection then the root events of their fault trees are connected with OR gate,

(2) OR connection, then the root events of their fault trees are connected with AND gate.

Fig. 3 shows an example for building up the fault tree of the whole process linking the fault trees of the individual activities.

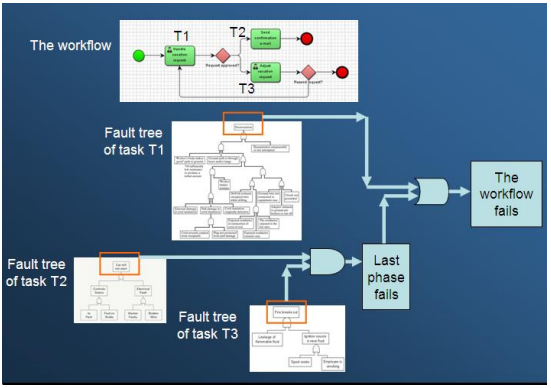


Figure 3.: The fault tree of the whole process

The fault tree of the whole process is needed only in phase of Monte Carlo analysis. We show the ordering of the resources according to their relevance in Fig. 4.

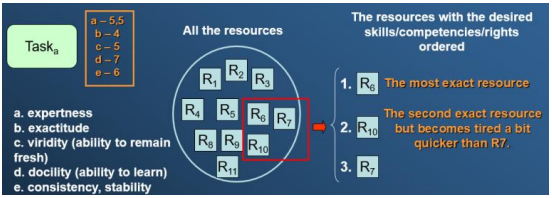


Figure 4.: Ordering the resources

After prioritizing the resources, one of them is chosen to execute the activity. The resources are selected based on their priorities. After the resources are selected for the activities, the schedule can be created and the total makespan can be computed.

Evaluation of the method

The companies that we are in cooperation with gave positive feedback about the applicability of this method. Further investigations and new logs for the human activities to the analysis are already started. After that the test of the method will be done in real situations.

Acknowledgment

We acknowledge the financial support of Széchenyi 2020 under the EFOP-3.6.1-16-2016-00015.

References

[1] Cooper, S. E., Ramey-Smith, A. M., Wreathall, J., & Parry, G. W. (1996). A technique for human error analysis (ATHEANA) (No. NUREG/CR-6350; BNL-NUREG-52467). Nuclear Regulatory Commission, Washington, DC (United States). Div. of Systems Technology; Brookhaven National Lab., Upton, NY (United States); Science Applications International Corp., Reston, VA (United States); NUS Corp., Gaithersburg, MD (United States).

[2] Hollnagel, E. (1998). Cognitive reliability and error analysis method (CREAM). Elsevier.

[3] Lee, W. S., Grosh, D. L., Tillman, F. A., & Lie, C. H. (1985). Fault Tree Analysis, Methods, and Applications - A Review. IEEE transactions on reliability, 34(3), 194-203.

[4] Rao, K. D., Gopika, V., Rao, V. S., Kushwaha, H. S., Verma, A. K., & Srividya, A. (2009) Dynamic fault tree analysis using Monte Carlo simulation in probabilistic safety assessment. Reliability Engineering & System Safety, 94(4), 872-883.

[5] M. Taheriyoun, S. Moradinejad: Reliability analysis of a wastewater treatment plant using fault tree analysis and Monte Carlo simulation, Environmental Monitoring and Assessment, 187(4186), 1-13, 2015

[6] ALD Reliability Engineering LTD software, http://www.fault-tree-analysis-software.com/ald-reliability-safety-engineering

[7] SCRAM, https://scram-pra.org/

Simulation models for transporting oil materials in pipelines

Balázs Csontos[1], István Heckl[1]
[1]Department of Computer Science and Systems Technology,
University of Pannonia, Veszprém, Hungary, csontos@dcs.uni-pannon.hu

Abstract: **The field of logistics involves several optimisation and simulation problems, and these especially pertain to pipeline transportation within the oil industry. Various products and semi-products must be transported in pipes while meeting numerous requirements. The most important of these is the availability of the product at the source prior to transporting. Similarly, the storage capacity at the end of the transportation line has to be sufficient as well. However, due to the complexity of the process, it is very hard to tell whether the scheduling is feasible. A simulation method is developed here which can answer the previous question by calculating the elementary states of a product pipeline system for a given time period based on the intended schedule and the initial states.**

Introduction

Simulation of processes plays an important role in the industry. Knowing the components of a system and how the components behave makes it possible for the experts to predict the response of the system to a specific action. Usually, a system is too extensive to fully understand its inner workings. For example, it can contain a large number of components, or the system might as well be very complex. As for the former, there is too much data to keep in mind. When it comes to the latter case, the inputs can affect multiple components which then affect other components, and so on. It can cause a ripple effect. It is possible that our initial input causes unintended side-effects or in some cases, even the opposite of what was intended. If a system is both complex and consists of a large number of building blocks, then human prediction can be quite inaccurate.

A simulator calculates the behaviour of the system step by step, i.e., it determines how a given input affects the components of the system, and what the future state will be. Furthermore, with the use of the simulator, we can analyse the system, identify bottlenecks, and plan future operation. However, the model, on which simulation operates, is only a representation of the

physical system. Consequently, the accuracy of the results strongly rests on the accuracy of the representation of the physical system.

Performing process simulations is of key-importance in the oil industry. There is a wide range of products to be produced. Transport and storage of both the raw materials and the products is the duty of Supply Chain Management (SCM) departments.

Problem definition

A simulator application and models are to be designed, capable of: (i) keeping track of the products and raw materials under the supervision of the SCM department, (ii) assisting the design of the product pipeline schedule, (iii) ensuring the feasibility of the monthly rolling plan, (iv) and simulating and visualising the transports in the product pipelines.

The proposed system can be also regarded as a decision support tool. Decision-making in the oil industry is an extremely complex process. Decisions are made at different stages of the supply chain distribution and at different levels of management. The decisions also differ in business scope, time horizon, time resolution, data certainty, and process detail. As a consequence, there is a great number of factors that affect decisions. Still, many oil and chemical companies operate with outdated decision-making processes [1] [2]. Decisions and communication across the supply chain are ineffective and delayed because of unorganized paper-based spreadsheets, functional barriers between departments and the lack of transparency. This all leads to slow and inaccurate day-to-day decisions that cost companies a fortune in terms of financial performance. The legacy way of decision-making (based on spreadsheets, meetings, and phone calls) all restrict speed and effectiveness. Thus, an efficient tool is required to assist the decisions in product pipeline scheduling.

Literature review

Simulation, scheduling, modelling, and production planning are core research areas in the oil industry. This section presents an outline of articles related to decision support, process simulation, and product pipeline scheduling.

In 1983, Vasek proposed and implemented an early flow sheeting simulation package. In the research, it was recognized that simulation has an essential role in design, and it was proven that significant results can be achieved using the desktop computers of the time [3].

Crama studied production planning approaches in the process industry. The differences and similarities among different methods were contrasted. Furthermore, the distinctive features of the process industry and their relation to production planning issues were also analysed. The difficulties caused by the implementation of classical flow control techniques have been explained and various approaches to overcome them have been found. A survey of specific flow control models and algorithmic techniques specifically for process industries have been implemented [4].

A general framework for modelling petroleum supply chains has been introduced by Neiro and Pinto. Different tanks, pipelines, and refinery models have been investigated. The complex topology of connected models results in a MINLP problem. This methodology has been generalized from discrete to continuous timeframe [5].

Tak et al. have introduced a cost-optimal inspection and replacement planning model that takes into consideration the pipeline's corrosion rate. The planning model was an MINLP problem, including integer variables for the pipe wall thickness, an inspection number and continuous variables for the inspection times [6].

Our proposed method focuses on pipe network simulation, while [5] and [6] focus on optimization using a mathematical model. There are simulation software products (e.g. Aspen Petroleum Scheduler [7], Anylogic Oil and Gas Simulation Software [8]) as well, but they are commercial products, their operations are not public, so the comparison with our method is difficult.

Operation of Simulator

First, information about the products, tanks, sites, and pipeline structure has to be fed into the simulator. It is then followed by the modelling step to map a real-life system onto the conceptual model. The product pipeline network is complex since there are parallel product pipelines, the pipe diameter is not standard, and there are branching points at various locations as well as a reversal of the flow direction may be executed.

Tank and pipe capacity information need to be accessible for the operator. Each tank is dedicated to a single product, but this can be changed in certain conditions. Each tank has two parts: mobile and immobile part. The immobile component cannot be downloaded in regular operation, which must be kept in mind throughout the process.

Sales data is acquired by the operator as well. Product demands are identified in advance including data types, e.g. the type and amount of

product, and delivery times. Products are transported in different ways, e.g. by barge, railway tank carriages, trucks, or product pipelines.

An operator plans operations by defining the allocation of products among means of transportation and the product pipeline schedule. Various materials are pumped through a pipe one by one. The effect of mixing adjoining materials is neglected from the scheduling perspective.

The input of the simulator contains the proposed product pipeline schedule whereas the output is the validated schedule. If the schedule is not feasible, the operator makes adjustments in the planed schedule and restarts the simulation. Furthermore, operation logs, reports are created, which can assist future planning and predictions.

Pipe simulation model

In this model, we analyse the realisation and runtime of tasks planned by the user (e.g., pipeline transportation, uploading and downloading pipes). This is implemented by taking discrete elemental steps along the time interval in question while simulating the pertinent tasks. At the time of each elemental step, the parameters of the system components are defined (e.g., the material content of the pipes and storage, and the connection status of the intertwining network pipes) as well as the position of particular tasks (e.g., as for pipeline transportation, the material already in the system and the material to be pumped at a certain time). In Figure 1, the pumping task triggers a „Pipe upload" event, that generates further events, i.e., two „Pipe downloads". Should the pipe end in an intersection, the material creates a new „Pipe upload" event. The simulation ends when the task list has been emptied and all the events are processed.

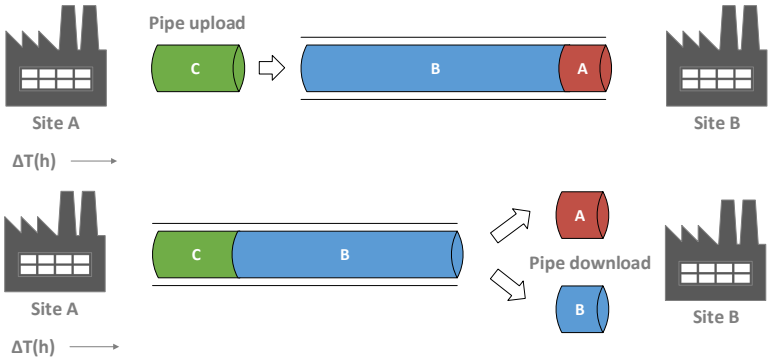


Figure 1. Operation of the Pipe model.

Conveyor belt model

Using the pipe model, the scheduler is unable to plan the transport of a material independently from the transport of another material. In reality, this means that the scheduler can only send a given material provided another is pumped into the pipe that pushes further the material (s)he wishes to transport. It is presumed that the transportation (pumping) mechanism can be automatized if it is inevitable to pump some kind of material into the pipe, e.g., to estimate the end-time of the transportation (and to actually start it). A conveyor belt model is proposed where a filling (dummy) material is continuously pumped (transported) in the system. This help to schedule the transportation of materials separately from each other.

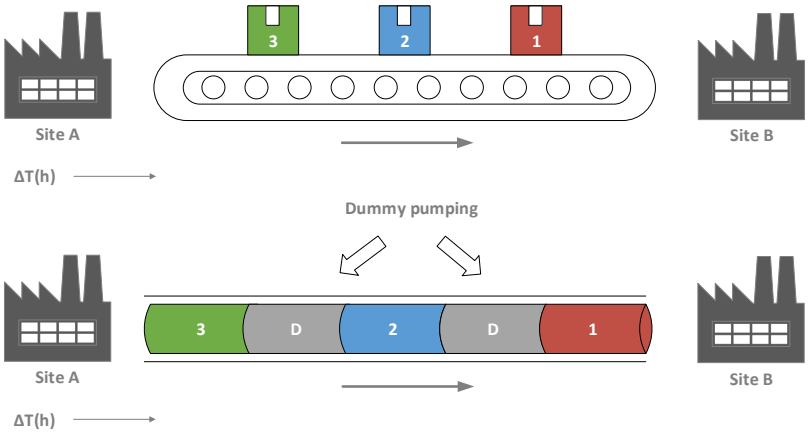


Figure 2. Operation of the Conveyor belt model.

Figure 2 shows the conveyor belt model, which is considered as a pre-simulation procedure which – with fewer conditions – helps us determine in a virtual environment whether the transportation of the requested oil industry material(s) is feasible or not. This simulation method can only be applied to frequently used sections in the pipeline system. At the end, the pipe utilisation must be checked and the dummy materials have to be changed to real materials one at a time. At 100% pipe utilisation, the conveyor belt model will not contain any dummy material and the original pipe simulation model has to be used validate the plan. The previously presented conveyor belt

model is currently under development. Further test results are expected in the future.

Summary and future work

A simulator and decision support tool has been introduced here to validate the planned product pipeline schedule and make the planning procedure as efficient as possible. The simulator with the scheduling models is capable of determining the future state of the system, e.g., tank contents and the product pipeline. The results are stored in history logs which can display the changes of the pipe content as well as the tanks, and help making decisions. We wish to further improve the simulator itself and the scheduling models so that the simulation will be able to depict real-life conditions with higher accuracy.

Acknowledgement

We acknowledge the financial support of the Széchenyi 2020 programme under the project No EFOP-3.6.1-16-2016-00015.

References

[1] Lasschuit, W., Thijssen, N., 2004, Supporting supply chain planning and scheduling decisions in the oil and chemical industry, Computers & Chemical Engineering 28, Issues 6-7, 863-870.

[2] Cheng L., Duran M. A., 2004, Logistics for world-wide crude oil transportation using discrete event simulation and optimal control. Computers & Chemical Engineering, 28, 897-911.

[3] Vasek, V., & Klemes, J. (1983). Simulation programming system siproversion for desk-top computer compucorp 625 in basic language. Computers & Chemical Engineering, 7, 175-182.

[4] Crama, Y., Pochet, Y., & Wera, Y. (2001). A Discussion of production planning approaches in the process industry. Core Discussion Papers 2001/42.

[5] Neiro S. M. S., Pinto J. M., 2004, A general modeling framework for the operational planning of petroleum supply chains, Computers & Chemical Engineering, 28, 871-896.

[6] Tak K., Kim J. (2017). A Planning Model for Inspection and Replacement of Pipes in a Refinery Plant. Chemical Engineering Transaction, 57, 991-997.

[7] Aspentech, "Aspen Petroleum Scheduler." [Online]. Available: https://www.aspentech.com/en/products/msc/aspen-petroleum-scheduler. [Accessed: 11-Mar-2019].

[8] Anylogic, "Oil and Gas Simulation Software." [Online]. Available: https://www.anylogic.com/oil-and-gas/. [Accessed: 11-Mar-2019].

Colored Petri Net based Monitoring and Diagnosis of Technological Systems

A. Leitold[1], A. I. Pózna[2], and M. Gerzson[2]
[1]Dept. of Mathematics, University of Pannónia,
leitolda@mik.uni-pannon.hu, Veszprém, Hungary
[2]Dept. of Elctrical Engineering and Information Science, University of Pannónia, {pozna.anna, gerzson.miklos}@virt.uni-pannon.hu,
Veszprém, Hungary

Abstract: **The diagnostic use of Colored Petri Net based models for different possibilities are introduced and compared in this paper based on the authors' work.**

Introduction

Several technological systems exist which can be described as discrete time discrete event system. This type of systems can be modeled and investigated with the Colored Petri Net (CPN) modelling method. These models describe the faultless operating course of the system, in general, but the consideration of the possible faults in the model is also important for diagnostic purposes.

The fault diagnosis problem includes the specific sub-tasks of fault detection, isolation and identification. Occurrence of faults can be determined by fault detection, the type or location of faults can be found by fault isolation methods while fault identification is used for characterizing the occurred faults. The most frequently used methods are based on the idea of unobservable transitions and using labeled Petri net models. Besides the observabililty of transitions, the set of places may have observable and unobservable subsets too. In [1] sufficient conditions of diagnosability are given and an on-line fault detection algorithm is developed based on ILP and checking the fault diagnosability conditions. In [2] the markings reachable by unobservable transitions are taken into account at the construction of the o ccurrence graph. In [3] firing times of transitions are taken into account and the diagnosis is based on generating residuals.

Complex systems can be represented in a compact form by using CPNs. CPN can be used as a colored diagnoser [4] which has usually

smaller size than the colored one, or backward reachability can be used to find the source of failures [5].

In case of large systems the computational effort of diagnoser algorithms can be extremely large therefore making effective algorithms is a very important task. Distributed diagnosis is a popular method to solve the problem however it raises the question how the global diagnosis result can be obtained from the local results. Usually some kind of communication protocol between the local diagnoser modules is required to get the total diagnosis result in [6].

In these paper two different methods are shown to solve the integration of the faults into the CPN model and to perform the diagnosis.

Colored Petri Nets for diagnostic purposes

According to the formal definition (see details in [7]) a CP-net model consists of places, transition, guard and arc functions, colors and tokens. For diagnostic purposes the following modelling principles are used.

- The input and output variables, the operational mode and the deviation are assigned to places.
- Color of tokens describes the variables' value, the type of the fault and to the emergent deviation from the nominal trace.
- The transitions execute the timing of the system. The operation can be divided into user defined time period, and the values of variables can change at the end of a period.
- The guard functions assigned to the transitions contain the fault generation function ([8]).
- Arcs connect coherent places and transitions.
- The arc functions describe the change of colors.

The behavioral analysis can be done with the occurrence graph ([7]). The occurrence graph contains all of the reachable markings (system states) from the initial one in a form of a graph. The nodes of the graph refer to the color distribution in a given system state and based this information the diagnosis can be performed.

Method 1: Diagnosis based on the CPN model containing the faults

Our first proposal shows how can be built a fault event with probability into a CPN model. Having this type of model, the consequences of a faulty operation can be investigated during the simulation and with the analysis of the resulted occurrence graph.

An important tool of the analysis of CPN based models is the occurrence graph. The nodes of the graph refer to the possible system states and the arcs to events leading to them. Knowing the probability of fault events, the probability of each state can be calculated.

Case study 1

As a case study let us assume a manufacturing system containing two manufacturing lines and a robot. The detailed description of the work of the system and its CPN model can be found in [9], here the gist of the work is summarized. The work pieces appeared in an input place have to be processed either on manufacturing line $M1$ or $M2$ or both on them according to operational instructions. The task of the robot is to put them to the appropriate input place of the lines according to operational instructions and if the process is ready then put the work pieces to an output place. For the modelling and the analysis of the manufacturing system the software package CPNTools [10] is used. The CP-net model of the normal (faultless) operation of the manufacturing system in the form of a screenshot from CPNTools can be seen in Fig. 1.

The color of tokens contains the identifier of the piece and the code of manufacturing process(es) to be carried out. Assume that only a single fault in the system can occur during manufacturing: the identification label of the piece can get damaged therefore it cannot be identified and it gets into a separate container. The modified part of the Petri net model where the occurrence of fault is taken into account can be seen in Fig. 2. As it can be seen the Petri net model has to be changed slightly, a new place is added and the arc inscriptions are changed. The check functions built into the arc inscriptions return with fault in predefined user defined probability.

The thorough analysis of the behavioral properties of a CP-net can be performed using its occurrence graph. Although CPNTools is able to generate automatically the occurrence graph, the probability function

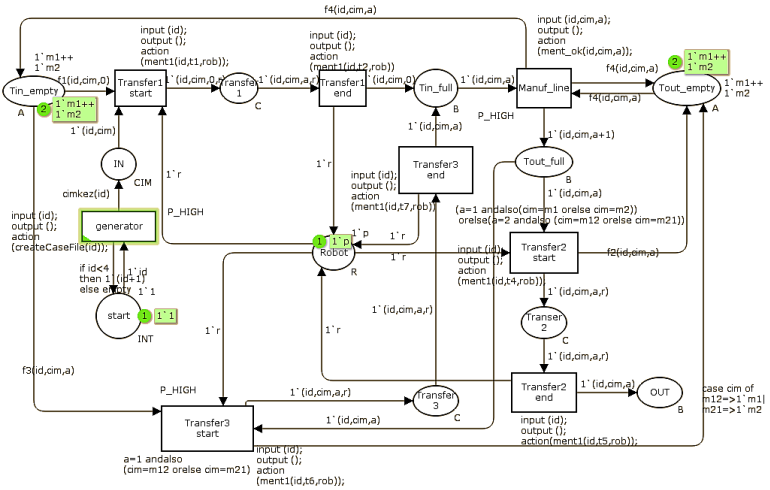


Figure 1: The CPN model of the faultless system

defined in arc expressions has to be removed, otherwise it generates the occurrence graph only for the normal mode or for the faulty mode. For the illustration of generation and analysis of the occurrence graph let us assume two work pieces, one of them has to go through manufacturing line 2, while the other has to go first through manufacturing line 1 then through line 2. Let us assume that faults can occur during manufacturing process. The resulted occurrence graph can be seen in Fig.3.

It can be stated that the occurrence graph has several branches and terminal nodes due the effect of the possible faults. The reason of branches is either technological (at the nodes 3, 9, 28, 51 and 54) or due to the fault. Only one of the terminal nodes refers to the normal termination of the process (highlighted by green), the others belong to the fault cases (highlighted by red).

For the more realistic investigation of the manufacturing system the simulation and diagnosis can be done with the timed version of the CP-net model, the time is assigned to the transitions. For the determination of the faulty states software package OGAnalyzer is developed. The detailed description of the time dependent analysis and probability

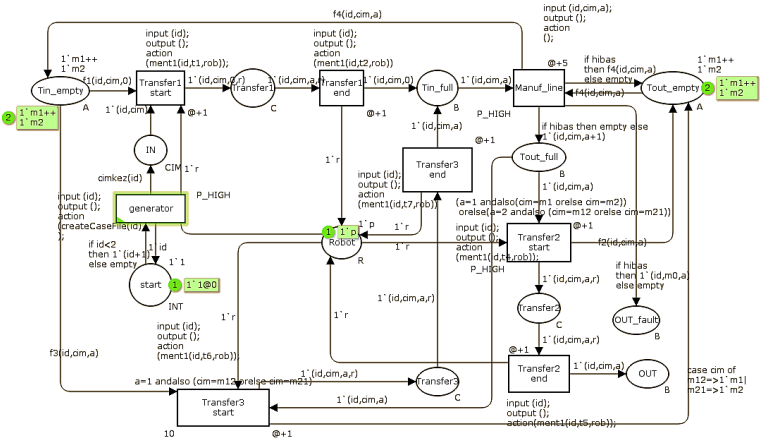


Figure 2: The CPN model of the faulty system

determination can be found in [9].

Method 2: CPN model of diagnostic method based on the comparing of event lists

A quite different approach to the application of CPN models for diagnostic purposes when the model itself aims at the realization of the fault diagnosis and the detection and identification of faults occurring in the process either in on-line or in off-line way. The qualitative models of the large and complex technological system can be applied for fault diagnosis of HAZID analysis efficiently. The Hazard Identification (HAZID) study is a technique for early identification of hazards and threats and can be applied at the conceptual or detailed design stage. Our approach was to develop a CPN model which can compare the so-called nominal trace with the characteristic trace stemming from the actual work of the system [11]. The nominal trace refers to faultless operation and as result of the comparison the differences are identified and collected in a list continuously. Based on this list the faulty operational mode of the system can be detected either during of the operation or after the completion of all events and on the other hand the type of the fault or the set of the fault possibilities can be identi-

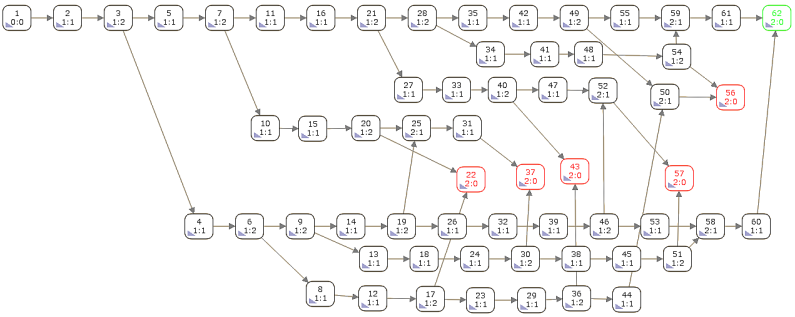


Figure 3: The occurrence graph in case of fault

fied. The described method can be used in case of large technological systems exploiting the modelling with CPNs in hierarchical way. The model and the occurrence graph of complex systems can be very large so their fault diagnosis can be done by structural decomposition where one can conclude the faults of the whole system from the diagnostical investigation of the subparts [11].

Case study 2

Let us assume the following simple technological system. The qualitative model of the technological system and the general form of CPN model are described in detail in [11]. A tank having one input and one output pipe is filled up with liquid until a certain level when the output valve is opened and the unit works in continuous mode. The filling process is a time driven event it takes two time periods. The tank has an input and output valve and a level sensor. The data measured by the level sensor is used only for monitoring the work of the unit. For the diagnosis of the effect of multiple faults it is assumed that the following faults or their combination can occur in the system:

- The bias fault of the level sensor: the measured value is less or greater than the actual value with one qualitative unit as the effect of bad operational mode.
- The leak of the tank: the level of the liquid remains zero in the tank.
- The combination of either of bias errors and the leak.

Table 1: Trace for different operational modes

normal	leak	negbias	leak-negbias
(0, *cl*, *cl*, 0)	(0, *cl*, *cl*, 0)	(0, *cl*, *cl*, e_0)	(0, *cl*, *cl*, e_0)
(1, *op*, *cl*, 0)	(1, *op*, *cl*, 0)	(1, *op*, *cl*, e_0)	(1, *op*, *cl*, e_0)
(2, *op*, *cl*, L)	(2, *op*, *cl*, 0)	(2, *op*, *cl*, 0)	(2, *op*, *cl*, e_0)
(3, *op*, *op*, N)	(3, *op*, *op*, 0)	(3, *op*, *op*, L)	(3, *op*, *op*, e_0)

It is assumed the fault or faults had evolved before the process starts and remain constant during the operation.

Let the states of valves be the input variables and the measured level value be the output variable. Valves are binary actuators, and their qualitative range space is $\{op,\ cl\}$, while the qualitative range space of the measured level is $\{e_0, 0, L, N, H, e_1\}$, where $0, L, N, H$ refer to zero, low, normal and high value measured by the sensor, respectively, while e_0 and e_1 may refer to outlier value caused by a bias failure. The structure of an event is as follows:

$$event_\tau = (\tau,\ \textit{state of input valve, state of output valve, measured value of level sensor});$$

where τ is the time stamp. The trace for the normal operational course contains the following events:

$$T = event_0, event_1, event_2, event_3;$$

where $event_0$ meets the initialization, $event_1$ refers to the start of filling up process, $event_2$ is intermediate state and $event_3$ means that the filling up is ready and then the tank works in continuous mode. The value of variables can be found in the first column of Tab.1. This nominal event list should be modified if a fault or faults occur. The Tab.1 contains the traces for faults tank leakage, negative bias error of level sensor and for the case when these two faults occur at the same time in the system as illustration.

The software package CPNTools was used for modelling the different courses of the system, for the generation of the occurrence graph and for implementing the proposed fault diagnosis method. The CP-net model of the simple tank can be seen in Fig. 4. Here the places in and out refer to the state of valves while the place level to the value of the level sensor. The color sets belonging to these places correspond to the defined qualitative range spaces.

Our diagnosis method ([11]) is based on the generation of deviation between the characteristic and nominal traces and on the searching the

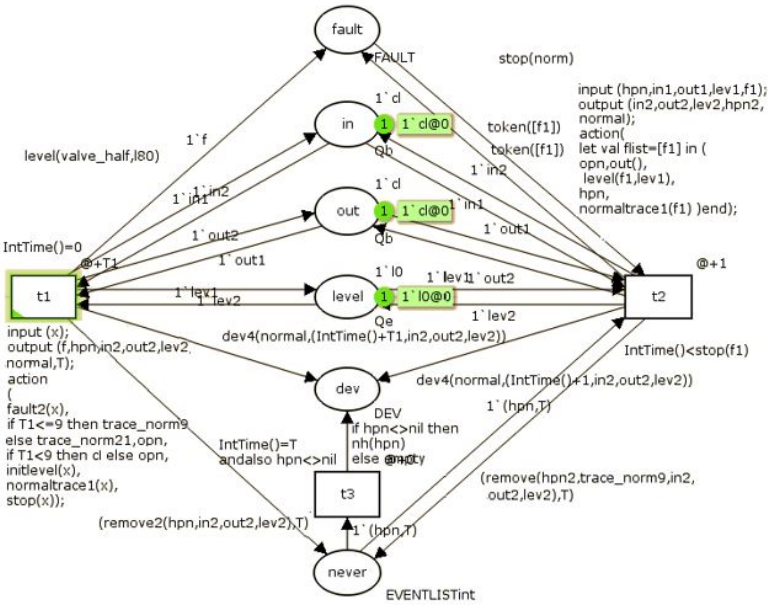


Figure 4: The occurrence graph in case of fault

node on the occurrence graph which token distribution refers to this deviation. Let us assume that all the fault modes of technological system are known. The first step is to generate the deviation list describing the distinction between the nominal and the characteristic traces. The next step is the simulation of CP-net model from the given initial state and the generation of the occurrence graph with all considered faulty mode. The last step of the diagnosis is to find the node having the token distribution which refers to the deviation list on the place dev. Based on the token color on the place fault in this node the type of fault can be determined. If more than one node has the token distribution referring to the deviation list, then the set of possible faults can only be concluded. If no token distribution refers to the deviation list then an unknown faulty mode occurs in the system.

Conclusions

The CPN models of technological systems describing the faultless operation can be subparts referring to possible faults. However these extra modelling parts can increase the size of model significantly. The various fault possibilities and fault operational modes cause a complex and hardly transparent model. Therefore this kind of approach is worth to use in case of small number faults and of relatively simple systems.

The advantage of the CPN model based diagnostical procedures using the qualitative model of the technological system is that several fault possibilities can be monitored at the same time. These faults can exist before the start of the technological procedure or can occur during the operation. In case of large and complex systems the CPN model can be built up hierarchically. The diagnosis can be performed with the help of structural decomposition, that is the analysis of the subparts can be done on their relatively small occurrence graph. Another advantage of this method that the investigation can be done in on-line or off-line way.

Acknowledgment

The authors acknowledge the financial support of Széchenyi 2020 program under the project No EFOP-3.6.1-16-2016-00015.

References

[1] Basile F.; Chiacchiot P.; and Tommasi G.D.: Sufficient Conditions for Diagnosability of Petri Nets. In Discrete Event Systems, 2008. WODES 2008. 9th International Workshop on. 370-375. doi:10.1109/WODES.2008.4605974. 2008

[2] Cabasino M.P.; Giua A.; and Seatzu C.: Fault Detection for Discrete Event Systems using Petri Nets with Unobservable Transitions. *Automatica*, 46, no. 9, 1531-1539. 2010.

[3] Lefebvre D. and Aguayo-Lara E.: Initial Study for Observers Application to Fault Detection and Isolation with Continuous Timed Petri nets. IFAC-PapersOnLine, 48, no. 7, 97 - 103. 2015.

[4] Pencole Y.; Pichard R.; and Fernbach P.: Modular Fault Diagnosis in Discrete-event Systems with a CPN Diagnoser, *IFAC-PapersOnLine*, 48, no. 21, 470-475. 2015.

[5] Bouali M.; Barger P.; and Schon W.: Backward reachability of Colored Petri Nets for systems diagnosis. *Reliability Engineering & System Safety*, 99, 1-14. 2012.

[6] Genc S. and Lafortune S., Distributed Diagnosis of Place-Bordered Petri Nets. *IEEE Transactions on Automation Science and Engineering*, 4, no. 2, 206-219. 2007.

[7] Jensen, K.: Coloured Petri Nets: Basic Concepts, Analysis Methods and Practical Use, Springer-Verlag, Berlin 1996

[8] Gerzson, M.; B. Márczi and A. Leitold. 2012. Diagnosis of Technological Systems based on their Coloured Petri Net Model, ARGESIM Report Eds: Troch, I., Breitenecker, F. no. S38 358/1-6

[9] Leitold, A., Márczi, B., Pózna, A.I., Gerzson, M.: Monitoring and Diagnosis of Manufacturing Systems using Timed Colored Petri Nets, *Hung. J. Ind. Chem.* , 2013

[10] CPN Group, University of Aarhus, Denmark: CPNTools 2.2.0 http://wiki.daimi.au.dk/cpntools/

[11] Pózna, A.I., Gerzson, M., Leitold, A., Hangos, K.M., : Colored Petri Net based Diagnosis of Technological Systems, in *European Simulation and Modelling Conference, ESM'2016*, Las Palmas, Spain, 2016

Ethical Issues Associated with the Use of Smart and Intelligent Learning Environments

B. Aberšek[1], M. Kordigel Aberšek [2], C. Sik-Lanyi [3], A. Flogie [4]
[1] University of Maribor, Faculty of Natural Science, boris.abersek@um.si
Koroška 160, 2000 Maribor, Slovenia
[2]University of Maribor, Faculty of Education, metka.kordigel@um.si
Koroška 160, 2000 Maribor, Slovenia
[3]University of Pannonia, lanyi@almos.uni-pannon.hu
Egyetem Str. 10. 8200, Veszprem, Hungary
[4]University of Maribor and Institute Antom Martin Slomšek,
andrej.flogie@z-ams.si
Koroška 160, 2000 Maribor, Slovenia

Abstract: **As we navigate our lives, we normally allow ourselves to be guided by our impressions, our awareness, and our feelings, and the confidence we have in our intuitive beliefs and preferences is usually justified. In such cases, we must consider proprioception. Proprioception could also be called "self-perception of thought", or "self-awareness of thought", or even that "thought is aware of itself in action". But the problem intensifies when we consider non-human proposals, proposals of artificial intelligence (AI), or even worse, if AI were to make decisions alone, without human interaction, without us. The use of artificial intelligence in today's reality is all around us. AI is also increasingly becoming a reality in education. Teachers have lots of opportunities to use smart and/or intelligent learning environments connected with AI solutions. The problem at hand, innovative learning environment theories, are discussed here with a view to better understanding. And if we would like to talk about learning environment issues in relation to AI, we must start talking about failing to imbue ethics and morality into AI systems. The presented research is mainly related to ethical issues in AI, especially to the use of AI in education.**

Introduction

The use of contemporary learning strategies, such as research- and problem-based learning, in relation to brain-based techniques, artificial intelligence and information-communication technologies, has provided scholars from diverse disciplines with an unusual opportunity to observe possible flaws in their own thinking [1,2]. There is a huge number of opportunities to introduce novelties in the learning process simply by being

creative or by using different information communication technologies. Students can be engaged, for example, by means of games and simulations that require them to apply information in unfamiliar contexts. Settings and activities such as e-learning environments, intelligent tutoring systems [1], role play, energizing online discussions and quick serious games, have the effect of adding sensory stimuli, reducing restlessness, and reinforcing information, and last but not least, they have plenty of opportunities for using smart and/or intelligent learning environments connected with AI solutions. The problem at hand, innovative learning environment theories, are discussed with a view to better understanding what could happen in such AI-based learning environments. The main research question could be: what potential ethical and moral issues could occur upon introducing AI and intelligent learning environments into education?

Natural vs. artificial intelligence

The human brain is an endlessly powerful, energy-efficient, self-learning and self-repairable computer. If we understood and could imitate the ways in which the human brain works, we could spark a revolution in information technologies, medicine, and society. In order to create an *artificial brain*, we should unite all our existing knowledge and everything that we are still able to learn about the internal life of our brain molecules, cells and circuits. Artificial brain is a commonly used term, which refers to research in the field of developing both software and hardware with similar cognitive abilities as the animal or especially the human brain. Research in the field of *artificial brains* has three important goals in science:

1. to understand from the work of neuroscientists how the human brain works, which is known as *cognitive neuroscience*;
2. to prove through thought experiments in the *philosophy of artificial intelligence*, that it is possible, at least in theory, to create a machine that has all the cognitive (and ethical?) capabilities of a human being;
3. to create through a series of long-term projects in the field of AI a machine *with universal or general intelligence, that is, to create universal (general) artificial intelligence*. This idea was popularized by Kurzweil [4], who named it strong AI, which means that the machine would have to equally intelligent as humans.

As part of the *first objective*, researchers are using biological cells to create *neurospheres* (small clusters of neurons) in order to develop new treatments for diseases including Alzheimer's and Parkinson's disease. Research in the frame of the *second objective* is related to known arguments such as John Searle's Chinese room argument [5][1], Hubert Dreyfus' critique of AI [6][2] or Roger Penrose's argument in *The Emperor's New Mind* [7][3] These critics argued that there are aspects of human consciousness or expertise that cannot be simulated by machines. One reply to their arguments is that this is absurd: the biological processes inside the brain can be simulated to any degree of accuracy, without any deviation from the natural processes. This reply is quite old, as it was made as early as 1950, by Alan Turing [3] in his classic paper *Computing Machinery and Intelligence*.

The *third objective* is referred to as *universal artificial intelligence* [9], *strong AI* [4] or *artificial general intelligence* [10]. Research in this area focuses on the implementation of artificial brains in conventional (digital) computing machines, and, on the basis of this, on the analysis of whole brain emulation. Kurzweil claims that this could be done by 2025. Henry Markram, director of the *Blue Brain project*,[4] made a similar claim. The question of whether digital computers are indeed suitable for simulating continuous brain processes is being asked increasingly often.

In the following chapter, we will pay our attention mainly on the third objective, and attempt to provide possible answers to research questions such as: what potential ethical issues could occur, and what kind of consequences could be caused by introducing universal AI and intelligent learning environments into education?

---

[1] For more details, see Bechtel and Abrahamsen (2002), pp. 303–304.

[2] Dreyfus' criticism of AI concerns what he considers to be the four primary assumptions of AI research, the biological and psychological, and the epistemological and ontological assumptions. Dreyfus argues in his criticism that we cannot now (and never will) be able to understand our own behavior in the same way as we understand objects in, for example, physics or chemistry (Dreyfus, 1979)

[3] In *The Emperor's New Mind,* Roger Penrose argues that known laws of physics are inadequate to explain the phenomenon of consciousness. He proposes a "new physics" and specifies the requirements for a bridge between classical and quantum mechanics (what he calls *correct quantum gravity*) (Penrose, 1989).

[4] In 2005, basic objectives were identified for the project of creating an artificial human brain, called *Blue Brains*, domiciled at the *École polytechnique fédérale de Lausanne* (EPFL) in Switzerland.

The social level, or the ethical issue of AI

In focusing our attention to the second, and especially the third objective of research from the field of developing *artificial brains*, we must ask ourselves two fundamental questions at the very beginning, namely: *Can MACHINES (artificial brains) be taught morality and ethics?* And, in relation to this question and the idea of introducing AI systems into education: *Can machines drastically affect the education system as well?* The philosophical starting point for answering these questions could be as follows:

> *Before giving machines a sense of morality, humans have to first define morality in a way computer can process, or "understand". This "understanding" means that algorithms of morals and ethics have to be defined in such a way, that they can become formalized, i.e., translated into the language of science, and coded into a language understood by machines, preferably in machine language*[5].
>
> *Whether this is a difficult, but not impossible task, is going to have to be the crucial question to be analysed and provided with appropriate solutions.*

While the problems of introducing AI into production and service activities (that is, using *intelligent machines* in a context where we can detect "errors" relatively quickly) do not have a dramatic impact on the "cognitive part" of society, the introduction of AI into education processes, which are surely fundamental to human civilization, are extremely risky and require a careful consideration in terms of what should be done and to what extent (COM 237 [11], Com 759 [12], Dignum [13]). Consequences of errors can be catastrophic and, above all, long-lasting, as the results of introducing these innovations will only be seen many years into the future. Here are some initial warnings. For years, alarmist views have warned against the unanticipated effects of general (universal) artificial intelligence (AI) on society. Ray Kurzweil predicts that by 2029 intelligent machines will be able to outsmart human beings [4]. Stephen Hawking argues that *once humans develop full AI, it will take off on its own and redesign itself at an ever-increasing rate*, which may constitute a threat to the human race. Similarly, Elon Musk warns that AI may represent a *fundamental risk to the existence of human civilization*.

[5] *Machine code* or *machine language* is a computer program, in which instructions are written in a computer language that is directly understandable by a computer *central processing unit* (CPU). Every CPU architecture has its own language. In the present context, we will refer to *machine language* as the language "understood" by machines.

Ethics, morality and AI

The field of advanced technologies which emphasize the importance of ethical principles in the design of autonomous systems could be divided into:

- Robotics
- Artificial Intelligence
- Computational Intelligence
- Machine Learning
- Deep Learning
- Cognitive Computing
- Affective Computing
- Algorithmically based programs.

In short, when we talk about AI, we must consider it especially from the perspective of two fundamental philosophies of the explanatory gap, adopted according to Chalmers [14], and related to the development and use of AI:

- *the easy problem*, the implementation and use of AI in "intelligent machines - robots", where the consequences of malfunction or failure can be quickly detected, and
- *the hard problem*, introducing AI into general society[6], which includes especially cognitive and affective computing.

Let us focus initially only on the easy problem. Already in 1942, Isaac Asimov [8] was the first to reflect on the ethical and moral issues in relation to robots, and defined the three basic laws of robotics:

- *A robot may not injure a human being or, through inaction, allow a human being to come to harm.*
- *A robot must obey orders given it by human beings except where such orders would conflict with the First Law.*
- *A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.*

Later on, in 1986, Asimov added the Zeroth Law:

- *A robot may not harm humanity, or through inaction allow humanity to come to harm. 1. A robot may not harm a human, or through inaction allow a human to come to harm, unless this interferes with the Zeroth Law.*

[6] *General society* in this context refers to humans society, society as a social system, for exaple education.

The Fourth and Fifth Law of Robotics were added subsequently, and different authors have modified and interpreted these laws in different ways. When we refer to the "easy" problem of implementing AI in the following section of the book, we will terminologically rely on the expression *robot*.
In 2011, the *Engineering and Physical Sciences Research Council* (EPSRC) and the *Arts and Humanities Research Council (AHRC)* of Great Britain jointly published a set of five ethical principles for designers, builders and users of robots in the real world [15, 16, 17].

1. Robots should not be designed solely or primarily to kill or harm humans.
2. Humans, not robots, are responsible agents. Robots are tools designed to achieve human goals.
3. Robots should be designed in ways that assure their safety and security.
4. Robots are artefacts; they should not be designed to exploit vulnerable users by evoking an emotional response or dependency. It should always be possible to tell a robot from a human.
5. It should always be possible to find out who is legally responsible for a robot.

In 2016, Tony Prescott revised these principles to differentiate ethical from legal principles [18]. This was followed by a number of authors who formulated various new laws, one of them defending the existence of a "new breed": *Mark W. Tilden* defined the following three guiding principles, three laws of robotics:

1. *A robot must protect its existence at all costs.*
2. *A robot must obtain and maintain access to its own power source.*
3. *A robot must continually search for better power sources.*

The essential thing about these laws is that they are written in a way as they would be written by AI autonomously, expressing above all the concern for the existence of a new "breed" of intelligent machines. Tilden is one of the leading experts on robotics, who can, like others, breathe life into these intelligent machines – robots, so that they will abide by some completely different kind of ethical norms than those known and established among people today.

On the basis of the above, the question of ethical norms of AI creators arises, since from the moment AI enters its learning stage (a life of its own), humans have only the minimal possibilities of leading the process of learning by this new, AI entity. Because all such systems will be interconnected (for example, in the Internet of Things (IoT)), this means that they will not only learn from humans, but also from each other. A logical question therefore arises: who will they believe more, humans or their equals, especially while bearing in mind that this equal entity holds the opinion that they must primarily protect themselves and their existence?
Many authors are concerned with ethical issues associated with AI, but in the end, these are all just individual opinions. We are still at the very beginning on this matter; we haven't even really made the first step. For this reason, in 2018, the EU and the European Commission established the European AI Alliance to consider these issues (COM 237 [11], COM 759 [12]). The consultation on the draft *Ethics Guidelines on Artificial Intelligence* (*AI*) concluded on 1 February 2019 [19].

AI and education

What about ethics in education in relation to AI? When we speak about intelligent learning environments that will be guided by AI, we are enabling AI not only to teach itself and create some new, humanoid entity parallel to humans, but to impact the development of the entire human society, define its values, and prescribe its ethical norms.

Conclusion

We believe that the following recommendations (COM 237 [11], COM 759 [12]):

- explicitly defining ethical behaviour,
- crowd-sourcing human morality, and
- making AI systems more transparent,

should be seen as a starting point for developing ethically aligned AI systems. Failing to imbue ethics into AI systems, we may be placing ourselves in the dangerous situation of allowing algorithms to decide what's best for us. For example, in an unavoidable accident situation, self-driving cars will need to make some decision for better or worse. But if the car's designers fail to specify a set of ethical values that could act as decision guides, the AI system may come up with a solution that causes more harm. This means that we cannot simply refuse to quantify our values. By walking away from this critical

ethical discussion, we are making an implicit moral choice. And as machine intelligence becomes increasingly pervasive in society, the price of inaction could be enormous – it could negatively affect the lives of billions of people.

Machines cannot be assumed to be inherently capable of behaving morally. Humans must teach them what morality is, how it can be measured and optimised. For AI engineers, this may seem like a daunting task. After all, defining moral values is a challenge mankind has struggled with throughout its history. *If we cannot agree on what makes a moral human, how can we design moral robots?* Nevertheless, the state of AI research and its applications in society require us to finally define morality and to quantify it in explicit terms. This is a difficult but not impossible task. Engineers cannot build a "Good Samaritan AI", as long as they lack a formula for the Good Samaritan human. However, as mentioned already, when introducing and implementing AI (especially into the general society-education), it is necessary to distinguish between two important fields that are relevant to this process, namely:

- the legal field, and
- the ethical or moral field.

In doing so, we must abide by the legal and ethical norms both:

- in the field of using AI, which is a relatively simple problem in this context, as deviations from norms and regulations will be directly detected in the wider society, and especially
- in the field of creating AI, and creating such algorithms that will contain all the necessary internal elements for the prevention of activities that could potentially harm the individual or humanity as a whole, such as the most elemental version of the four laws of robotics as defined by Isaac Asimov [8].

In the spirit of today's world and the advancements made in this field, these Asimov laws (and similar ones), could be rewritten simply by replacing the word "robot" with the acronym "AI", to imply a more generalized meaning. In the future, moral and ethical laws in this form will still have to be defined and written accordingly in a language understood by AI.

Acknowledgment

The author would like to thank the financial support of Széchenyi 2020 programme under the project No EFOP-3.6.1-16-2016-00015.

References

[1] B. Aberšek *Problem-Based Learning and Proprioception*. New Castle upon Tyne: Cambridge Scholars Press. 2018

[2] B Aberšek, B. Borstner, J. Bregant *Virtual Teacher: Cognitive Approach to e-Learning Material.* New Castle upon Tyne: Cambridge Scholars Press. 2014

[3] A. Turing *Computing Machinery and Intelligence,* Mind*, LIX(236): 433–460,* doi:10.1093/mind/LIX.236.433*,* ISSN 0026-4423 1950
[4] R. Kurzweil *The Singularity Is Near.* London: Duckwort Overlook. 2005
[5] W. Bechtel, A. Abrahamsen *Connectionism and the Mind*. Oxford, UK: Blackwell Publisher. 2002
[6] H. Dreyfus *What Computers Can't Do*. New York: MIT Press. 1979
[7] R. Penrose *The Emperor's New Mind: Concerning Computers, Minds and The Laws of Physics*. Oxford: Oxford University Press. 1989
[8] I. Asimov *I, robot*. New York: Gnome Press. 1950
[9] P. Voss Essentials of general intelligence: The Direct Path to Artificial General Intelligence. V Goertzel, B., Pennachio, C. (Ed.), *Artificial General Intelligence*, Berlin: Springer, 131−159. 2006
[10] S. Baum A Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy. *Global Catastrophic Risk Institute Working Paper* 17-1. 2017
[11] COM 237 Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: Coordinated Plan on Artificial Intelligence. Retreived form: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN 2018
[12] COM 759 Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: Coordinated Plan on Artificial Intelligence. EC Retreived from: https://eur-lex.europa.eu/legal-content/EN/TXT/DOC/?uri=CELEX:52018DC0795&qid=1546111312071&from=EN 2018
[13] V. Dignum Ethics in artificial intelligence: introduction to the special issue. *Ethics and Information Technology*, 20, 1–3. 2018
[14] D. Chalmers The Conscious Mind. Oxford: Oxford University Press,1996.
[15] *Stewart, Jon (2011-10-03). "Ready for the robot revolution?". BBC News. Retrieved 01.12.2018 from: https://www.bbc.com/news/technology-15146053.*
[16] *"Principles of robotics: Regulating Robots in the Real World". Engineering and Physical Sciences Research Council. Retrieved 01.12.2018 from: https://epsrc.ukri.org/research/ourportfolio/themes/engineering/activities/principlesofrobotics/.*
[17] A, Winfield *"Five roboethical principles – for humans". New Scientist. Retrieved 01.12.2018 from*: https://www.newscientist.com/article/mg21028111-100-five-roboethical-principles-for-humans/
[18] V.C. Müller "Legal vs. ethical obligations – a comment on the EPSRC's principles for robotics". Connection Science. doi:10.1080/09540091.2016.1276516. 2017
[19] HLEG, "Ethics Guidelines on Artificial Intelligence", Brussels: EC, 2018. https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_draft_ethics_guidelines_18_december.pdf

Hungary's digital entrepreneurship based on the European Index of Digital Entrepreneurship Systems

L.Szerb[1], É. Komlósi[2], M. Tiszberger[3],
[1]University of Pécs, Faculty of Business and Economics,
szerb@ktk.pte.hu 7634 Pécs, Rákóczi St. 80. Hungary
[2]MTA-PTE Innovation and Economic Growth Research Group,
University of Pécs, Faculty of Business and Economics,
komlosieva@ktk.pte.hu 7634 Pécs, Rákóczi St. 80. Hungary
[3]University of Pécs, Faculty of Business and Economics,
tiszbergerm@ktk.pte.hu 7634 Pécs, Rákóczi St. 80. Hungary

Abstract:

Since the last decade, digital technological developments (artificial intelligence (AI), Internet of things (IoT), machine learning, augmented reality, cloud computing, 5G networks, or autonomous vehicles etc.) have been progressively transforming the character of entrepreneurial activities and continuously facilitating innovative opportunities for entrepreneurship. What is the impact of digitization on entrepreneurship? What role can entrepreneurship play in the digital age? These are the two major research questions we would like to address. Whereas the European Union has a relatively long history in measuring the digitalization development of the member countries it has not been connected to entrepreneurship. At the same time, policymakers need sufficient metrics to measure digital entrepreneurship to exploit new productivity potential for ensuring economic growth and societal welfare. In this paper is presenting the European Index of Digital Entrepreneurship Systems (EIDES) that is a novel tool aiming to measure the digital entrepreneurship system in the European Union countries. EIDES combines the physical and the digital conditions for stand-up, start-up and scale-up ventures.

With special emphasis on the analysis of Hungary, we offer a detailed picture of the performance of the Hungarian digital entrepreneurial ecosystem. Hungary is ranking 24th belonging to the fourth, called *Laggards*, cluster of the EU countries. Hungary's weakest pillar is the *Culture and Informal Institutions* while *Human Capital* and *Knowledge Creation and Dissemination* are relatively strong.

The European Index of Digital Entrepreneurship Systems

Over the recent decade, entrepreneurship has undergone a global transformation. The entrepreneurial opportunities were radically redefined and the practices to pursue them have changed accordingly. These transformations are reflected in the global adoption of new organizational innovations to support entrepreneurial activity, and – above all – in the emergence of a regional agglomeration of economic activity: the entrepreneurial ecosystem [1], [2]. The digitally-enabled entrepreneurial transformation creates important challenges for policy [3], [4]. Policy-makers need metrics to monitor this transformation and ensure that the productivity potential of digital advances can benefit economic and societal welfare [5] [6]. This need sets up a measurement challenge because the digitally-enabled entrepreneurial ecosystem is a pervasive systemic phenomenon impossible to capture by count-based measures of individual-level entrepreneurial action.

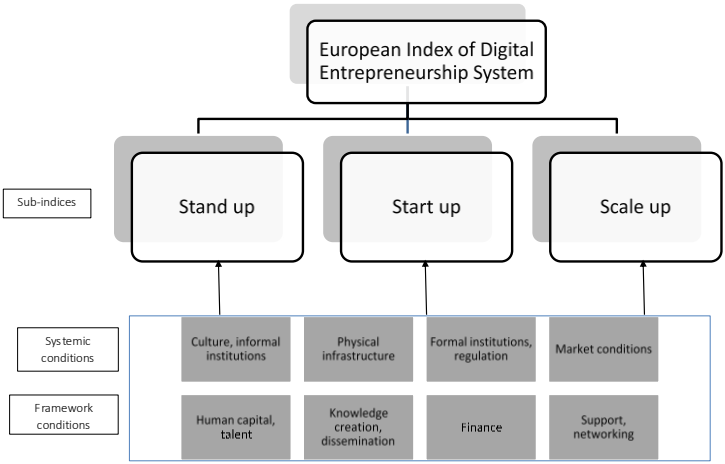


Figure 1. Structure of the European Index of Digital Entrepreneurship Systems

The European Index of Digital Entrepreneurship Systems (EIDES) [7], responds to the need for a tool to better understand and appraise the extent of the digital entrepreneurial ecosystem. Specifically, the EIDES is an attempt to measure both physical and digital conditions for stand-up, start-up and scale-up ventures in EU 28 countries.

The structure of the EIDES (Figure 1) encompasses four pillars for the General Framework Conditions (i.e. Culture and Informal Institutions, Formal Institutions, Regulation, and Taxation, Market Conditions and Physical Infrastructure) and their associated digital counterparts. Specifically, each framework condition can be digitalized with a suitable

measure of a corresponding digital context made by variables that reflect the digitalization aspect of each specific framework condition. Consequently, two versions of each framework condition appear in the index: a non-digitalized part and a digitalized one.

In addition to the General Framework Conditions, the EIDES also measures 'systemic' framework conditions which are the resource-related conditions with a direct effect on the entrepreneurial dynamic in a given country or region. In practical terms, businesses require a range of different resources (i.e. Human Capital, Knowledge Creation and Dissemination, Finance, and Networking and Support) in order to scale up successfully. These resources are not substitutable against one another. Therefore, the Systemic Framework Conditions have to come together to help ‘co-produce’ the system outcomes.

In the EIDES' theoretical structure the General Framework Conditions apply broadly to entrepreneurship, while the Systemic Framework Conditions act differently across three stages of the entrepreneurial development: stand-up, start-up, and scale-up. The Stand-up stage relates to the self-selection of individuals into entrepreneurship. The Start-up stage is the subsequent creation of new start-ups. The Scale-up stage concerns the scaling up of the start-ups that discovered a business model with high-growth potential. Accordingly, the EIDES includes three sub-indices for each Systemic Framework Conditions plus their digital versions calculated with measures of the corresponding digital contexts.

Finally, the value of the overall EIDES is the average of both General and Systemic Framework Conditions. This approach possibly provides a helpful portrayal of national systems of entrepreneurship. In each national system of entrepreneurship, general framework conditions regulate how the systemic conditions can realize their full potential and co-produce the national entrepreneurial dynamic. The approach underlying the EIDES also distinguishes between digital and non-digital conditions to proxy the effect of digitalization on systems’ abilities to facilitate high-quality entrepreneurial dynamic. Furthermore, declining the systemic conditions across three entrepreneurial stages allows for even more fine-grained policy insights.

Country	Stand-up System		Start-up System		Scale-up System		EIDES	
	Score	Rank	Score	Rank	Score	Rank	Score	Rank
Denmark	84.0	1	77.1	1	80.9	1	80.7	1
Sweden	73.4	4	76.4	2	76.9	2	75.6	2
Luxembourg	75.9	2	72.1	3	74.0	3	74.0	3
Finland	73.8	3	71.3	4	72.0	4	72.4	4
Leaders	76.8		74.2		76.0		75.7	
Germany	64.6	6	61.7	6	65.1	6	63.8	5
United Kingdom	65.0	5	60.6	7	65.6	5	63.7	6
Netherlands	64.3	7	57.6	8	64.7	7	62.2	7
Ireland	61.5	8	62.7	5	59.7	8	61.3	8
Belgium	57.5	9	57.0	9	58.5	9	57.6	9
Austria	52.8	11	54.2	11	55.9	10	54.3	10
Malta	54.6	10	56.2	10	52.0	11	54.3	11
Estonia	52.7	12	51.5	12	48.7	13	51.0	12
France	49.7	13	47.7	13	51.4	12	49.6	13
Followers	52.5		52.4		52.0		52.3	
Spain	45.2	14	44.6	14	42.9	15	44.2	14
Czech Republic	41.9	15	41.9	15	43.2	14	42.3	15
Lithuania	39.8	16	41.6	17	40.3	16	40.6	16
Slovenia	35.5	19	41.6	16	38.2	17	38.4	17
Portugal	38.9	17	38.7	18	36.8	18	38.1	18
Cyprus	36.7	18	38.3	19	34.0	20	36.3	19
Catchers-up	37.7		40.1		37.3		38.4	
Poland	31.8	22	33.6	20	33.4	21	32.9	20
Latvia	32.7	20	32.8	21	33.2	22	32.9	21
Italy	32.0	21	31.8	24	34.0	19	32.6	22
Croatia	29.5	23	32.3	22	29.9	25	30.6	23
Hungary	27.4	25	32.0	23	30.9	23	30.1	24
Slovakia	28.3	24	30.8	25	30.6	24	29.9	25
Greece	22.9	26	26.4	26	23.5	26	24.3	26
Bulgaria	22.8	27	25.6	27	23.2	27	23.9	27
Romania	21.6	28	22.4	28	20.8	28	21.6	28
Laggards	23.9		26.3		24.5		24.9	
EU28 average	47.0		47.2		47.2		47.1	

Table 1. EIDES scores for the European Union 28 countries

Hungary's position in EIDES

According to the EIDES ranking for 2018 Denmark, Sweden, Luxembourg, and Finland lead due to their high scores of the digitalized General and Systemic Framework Conditions for entrepreneurship (Table 1). In

particular, Denmark is first in all Digital Entrepreneurship Stand-up, Start-up, and Scale-up sub-indices. Sweden is the second for Start-up and Scale-up sub-indices, and the fourth for the Stand-up one. Finland is the second for Stand-up conditions and the fourth for the rest. Behind, at a notable distance according to the EIDES, there are the followers group of nine countries: Germany, United Kingdom, Netherlands, Ireland, Belgium, Austria, Malta, Estonia, and France. Germany and the United Kingdom appear quite close to one another. A third cluster is the catchers-up consisting of Spain, Czech Republic, Lithuania, Slovenia, Portugal, and Cyprus. Finally, the laggards cohort consists of the remaining nine countries: Poland, Latvia, Italy, Croatia, Hungary, Slovakia, Greece, Bulgaria, and Romania. It is striking that Italy, in spite of being one of the G7 countries, ranks in this group together with former centrally planned economies and Greece.

Out of the 28 EU countries, Hungary is ranked in the last cluster called Laggards[1] on the 24th place with EIDES score 30.1, ahead of Slovakia Greece, Bulgaria and Romania and behind Poland, Latvia and Italy. There are only little differences in the scores of the three sub-indices: Hungary's best sub-index is the Digital Entrepreneurship Start-up (32.0) followed by Digital Entrepreneurship Scale-up (30.1) and the third is Digital Entrepreneurship Stand-up (27.4). A more detailed picture of the pillar scores and their components is presented in Table 2.

[1] Country group indicates the country's performance relative to others, grouped in four categories: (1) Laggards (EIDES index score below 0.35); (2) Catchers-up (0.35 < EIDES score ≤ 0.45); (3) Followers (0.45 < EIDES score ≤ 0.70); (4) Leaders (EIDES score over 0.70).

	PILLAR	PILLAR SCORE	NON DIGITAL SCORE	DIGITAL SCORE
General Framework Conditions	Culture. informal institutions	25.0	55.3	58.3
	Formal institutions. regulation. taxation	30.5	65.4	53.0
	Market conditions	27.6	77.1	51.9
	Physical infrastructure	33.6	63.5	58.4
Systemic Framework Conditions	Human capital	34.5	59.6	59.9
	Knowledge creation and dissemination	34.8	61.4	62.7
	Finance	33.4	65.7	53.3
	Networking and support	31.1	53.5	60.8
EIDES SCORE		**30.1**	**62.7**	**57.3**
SUB-INDEX		**SUB-INDEX SCORE**		
Sub-indices	Digital Entrepreneurship Stand-up	27.4		
	Digital Entrepreneurship Start-up	32.0		
	Digital Entrepreneurship Scale-up	30.9		

Table 2. EIDES component scores for Hungary

Legend:

Pillar: In the first column we list the eight pillar names and the three sub-index names as well as the EIDES index score.
Pillar score column shows the country's pillar scores on a 0-100 point scale.
Non-digital score column shows the country's non-digitalized pillar scores (scale from 0 to 100).
Digital column shows the digital component scores on a scale from 0 to 100.
EIDES index score shows the overall index score, as well as the scores for non-digital and digital components on a scale from 0 to 100.

There are more variations in the eight pillar scores of Hungary as compared to the three sub-indices. Hungary's best pillars are *Knowledge Creation and Dissemination* (34.8) and *Human Capital* (34.5) both belong to the Systemic Framework Conditions. *Culture and Informal Institutions* is by far the weakest component of the eight EIDES pillars. This pillar includes the acceptance and social desirability of entrepreneurship in the society as well as the risk acceptance of the population. *Market Conditions* (27.4) are also at a low level. By surprise, the non-digital elements (62.7) of the EIDES are better than the digital one (57.3). The sales on the net – part of the *Market*

Conditions – is the weakest variable of the digital and non-digital components. The digital element of the *Formal Institutions, Regulation, and Taxation* is also below the desirable level calling for better government policy and regulation.

Two of the four GFC pillars, *Culture and Informal Institutions* and *Market Conditions,* are weak in Hungary. *Systemic Framework Conditions* (SFC) relate more directly to the different stages of entrepreneurial sub-dynamics within a country's system of entrepreneurship. Hungary's relatively strong systematic pillars are *Knowledge Creation and Dissemination* and *Human Capital*.

The pillar *Culture and Informal Institutions* reflects the degree to which a country's social and cultural norms and resulting societal practices support high-quality entrepreneurial endeavors. Policymakers should promote the strengthening of positive cultural and social norms and practices, because these can enhance the quality of the entrepreneurial dynamic by increasing the attractiveness of the entrepreneurial career choice for individuals or by encouraging entrepreneurial orientation and risk-taking for growth. *Market conditions* constitute one of the most important regulators of a country's entrepreneurial dynamic. This pillar includes indicators reflecting different features of market conditions, such as the effect of agglomeration externalities, the market power of existing businesses and business groups, domestic and foreign market size, and also, perceptions of entrepreneurial opportunities.

Furthermore, EIDES needs to be decomposed to be able to get a more accurate picture of the Hungarian digital entrepreneurial profile. For example, analyzing separately the role of the non-digital and digital components of the index it allows for even more fine-grained policy insights.

The *digital component* moderates the overall performance of the relatively weak *Market condition* pillar. The digital counterpart of the pillar characterizes the exploitation of online market channels (e.g., e-commerce, e-sales, e-advertisement) by households and firms. By adopting digital technology households and businesses can enhance efficiency, reduce costs and better engage customers, collaborators, and business partners. Furthermore, the Internet also offers wider access to markets. Consequently, Hungarian households and businesses should utilize digital technologies to a greater extent.

On the contrary, the *non-digital component* modifies negatively the overall performance of the *Culture and Informal Institution* pillar. Prevailing social norms and attitudes may shape entrepreneurial behaviors, (such as the perceptions of citizens regarding ethical behavior by business firms in their interactions with public officials, politicians, and other business firms).

Negative norms and practices impede positive outcomes. Corruption has a negative effect on economic activity because it undermines the rule of law and erodes the predictability of economic relationships. When the level of corruption is low and the quality of governance is high, citizens are more likely to accept entrepreneurial risk. High level of corruption in Hungary has definitely a negative effect on entrepreneurship.

Also the different digital components contribute to the weak performance of the *Financing* and the *Formal Institution, Regulation, and Taxation* pillars. Availability of finance is widely recognized as a key regulator of the entrepreneurial dynamic, specifically for the stand-up stage. Both the amount of funding matters, as does the accessibility by entrepreneurial ventures to such funding. In the case of the *Financing* pillar as digital proxies we use indicators as Digital payment transactions and Number of cashless payment transactions. On the one hand, these indicators capture the effect of digital technologies and infrastructures on the functional operation of financial institutions. On the other hand, these proxies offer insight into the new generation of digitalized financial products and services.

Digitalization intertwines with formal institutions to shape entrepreneurship. *Formal Institutions, Regulation, and Taxation* pillar encompasses several indicators describing digital security and privacy, and also includes proxies that measure how formal institutions and the regulatory environment shape digitalization processes and competition. The Hungarian government should primarily focus on the improvement of these elements.

Summary

The European Index of Digital Entrepreneurship Systems (EIDES) responds to the need for creating an adequate tool to better understand and appraise the extent of the digital entrepreneurship. As a composite measure EIDES helps policymakers to identify the strengths and weaknesses of the national digital entrepreneurial ecosystem.

The concept of the EIDES draws on the entrepreneurial ecosystem literature which primarily emphasizing the multidimensional and contextual nature of entrepreneurship. The strength of the entrepreneurial ecosystem approach is its ability to weave many different layers of the entrepreneur's context together, highlighting the close relationships, interdependencies, and reinforcing mechanisms across the different constituent elements of the entrepreneurial ecosystem, often centered around a focal community of ecosystem constituents [9] [10]. A weakness of the approach is that most conceptualizations are descriptive, rather than theory-grounded, and tend to emphasize different layers, structural elements, and processes of

entrepreneurial ecosystems. It aims to build on the most relevant data, however availability and comparability of the variables is always problematic.

Another potential criticism of the EIDES methodology – as with any other index – might be the apparently arbitrary selection of indicators and the neglect of other important ones. All indices are inevitably constrained by the availability of relevant data. In constructing the EIDES, we tested alternative proxies for each pillar and selected variables on the basis of their coverage of the relevant aspect, as well as their pertinence to the phenomenon we seek to portray.

The EIDES 2018 ranking prevails huge differences amongst the EU countries with respect to digital entrepreneurship. Nordic countries and Luxembourg lead the rank while Hungary is placed in the last cluster (called Laggards) together with many other transitional countries, Italy and Greece.

A more detailed analysis prevails that Hungary has weaknesses in the General Framework Conditions (GFC) while our situation is better in the systemic part. General Framework Conditions regulate the degree to which the grassroots-level entrepreneurial dynamic is translated into national economic development, and also the quality of that dynamic in itself. These framework conditions tend to be fairly path-dependent, and we would not expect them to change suddenly.

Acknowledgment

The EIDES project has been financed by the European Commission under contract number – 932886-2017 A08-GB, thanks for it.

References

[1] Acs, Z. J., Autio, E., Szerb, L. (2014) National Systems of Entrepreneurship: Measurement Issues and Policy Implications. Research Policy, 43(3), 476-494.

[2] Spigel, B. (2017). The relational organization of entrepreneurial ecosystems. *Entrepreneurship Theory and Practice*, *41*(1), 49-72 [1]

[3] Autio, E., Nambisan, S., Thomas, L. D., & Wright, M. (2018). Digital affordances, spatial affordances, and the genesis of entrepreneurial ecosystems. *Strategic Entrepreneurship Journal*. 12(1) pp.72-95.

[4] Sussan, F., & Acs, Z. J. (2017). The digital entrepreneurial ecosystem. *Small Business Economics*, 49(1) pp. 55-73.

[5] Brown, R., & Mason, C. (2014). Inside the high-tech black box: a critique of technology entrepreneurship policy. *Technovation*, *34*(12), 773-784.

[6] Nambisan, S. (2017) Digital entrepreneurship: Toward a digital technology perspective of entrepreneurship. *Entrepreneurship Theory and Practice*, *41*(6), 1029-1055

[7] Autio, E., Szerb, L., Komlósi, E., & Tiszberger, M. (2018). *The European Index of Digital Entrepreneurship Systems* (No. JRC112439). Joint Research Centre (Seville site). DOI: 10.2760/39256;

[8] Van Roy, V. and Nepelski, D. (2016) *Assessment of Framework Conditions for the Creation and Growth of Firms in Europe.* Joint Research Centre, JRC Scientific and Policy Reports – EUR 28167 EN; DOI:10.2791/2811
[9] Autio, E., Nambisan, S., Thomas, L. D., & Wright, M. (2018). Digital affordances, spatial affordances, and the genesis of entrepreneurial ecosystems. *Strategic Entrepreneurship Journal*. 12(1) pp.72-95.
[10] Spigel, B. 2017. The relational organization of entrepreneurial ecosystems. Entrepreneurship Theory and Practice, 41(1) pp. 49-72.

Aggregation approaches for distributed flexibility aggregators

István Balázs, Attila Fodor, Attila Magyar
University of Pannonia
8200 Veszprém, Egyetem u. 10. Hungary

Abstract: **Increasing share of intermittent and distributed generation demands energy sector to find novel solutions for the challenges of the transformed system operation. Significant generation on the distribution grid may cause voltage, power flow and network congestion problems, more balancing reserves are required while consumers became active allowing load to follow generation. Applying aggregator function is a promising initiative to collect new sources of flexibility, combine and offer them as services for system operators or other market participants. However, aggregating small-scale flexibility is not a trivial operation, advanced tools are required to combine large number of inhomogeneous resources and provide competitive services. In this paper the aggregator role is introduced, potential markets are identified and aggregation approaches are suggested for categories of flexibility providers.**

Introduction

Conventional power systems are characterized by large generation sources that inject power into the transmission grid, which is transported to distribution networks and then delivered to the end-users. Power flows one way from the high voltage transmission grid to the end-user at low voltage networks. According to the "generation follows load" paradigm, electrical generation is adjusted to match electrical consumption (load) as it varies throughout the time of day and seasons. Centralized, dispatchable and predictable generation provides flexibility at the transmission level to the electric system to balance generation and demand.

The increasing amount of distributed and renewable generation (from around 21% share of net power generation in 2010 to 44% in 2030 [1]) transforms the generation more variable and intermittent. These production units are connected to low and medium voltage networks that were designed to be supplied from high voltage networks and transfer power to consumers. Having significant capacity of generation connected to the distribution grid, power flow directions may change, voltage and congestion issues arise. On the other hand, demand side becomes more active enabling a new control

approach, "load follows generation" paradigm to emerge and providing additional sources of flexibility on the distribution network.

Flexibility is the modification of generation injection and/or consumption patterns in reaction to an external signal (price signal or activation) in order to provide a service within the energy system [2]. It is an active management of an asset that can be used to maintain system balance or mitigate congestion risks. The proper management of available flexibility, both in generation and demand side, can help to compensate the lack of certainty of renewable sources.

Transmission System Operators[1] (TSO) manages system frequency fluctuations and imbalance. Unpredictability and volatility of intermittent generation requires increasing volume of balancing capabilities and procuring balancing services not only from the transmission grids but also from distribution grids. Distribution System Operators[2] (DSO) need active tools to manage congestion in their distribution grids and consider procurement of flexibility services to redispatch their network. In future energy systems, TSO and DSO will have access and compete for the same flexibility resources. Proper coordination schemes have to be developed, but it is also an opportunity for new players on the distribution grids to offer flexibility capabilities. Balancing markets, TSO-DSO roles and relationships may have significant differences in the world, even in Europe. European Commission launched an energy union strategy in 2015 in order to create a fully integrated internal energy market in Europe. Implementing the strategy, the Clean Energy for all Europeans package (adopted in 2019) contains several legal acts that decrease market differences and dismantle obstacles of active DSO contribution.

Aggregator

Aggregator, a new market agent will play a central role collecting resources on the distribution grid and make them involved in such markets, that are not available for them individually. This requires reliable coordination that has

[1] Transmission System Operator is responsible for providing and operating high and extra-high voltage transmission networks for long-distance transmission of electricity as well as for supply of lower-level regional distribution systems and directly connected customers.

[2] Distribution System Operator is responsible for providing and operating low, medium and high voltage distribution networks for regional distribution of electricity as well as for supply of and directly connected customers.

technical implications as well and common legal background in order to determine the aggregator role. The EU Commission proposal for the recast of the E-Directive [2] defines aggregator a market participant that combines multiple customer loads or generated electricity for sale, for purchase or auction in any organized energy market. This definition incorporates aggregation of all types of decentralized energy resources. USEF Foundation's Aggregator workstream analyzed the different topics related to the aggregator role with particular focus on demand-response aggregation and the relationship between aggregator and the BRP (Balance Responsible Party) / supplier. Seven different aggregation implementation models were identified, advantages and limitations were evaluated [3]. Flexibility resources were investigated in detail and an information model were set up by SmartNet D1.2 [4] that contain a mathematical description of the dynamic behavior of the resource, its constraints for flexibility provision, a formulation of the different components of costs needed to provide flexibility. BestRES project [5] explored different ways an aggregator can create value, categorizing provided services into internal (own balancing) and external reasons (wholesale, retail, reserve capacity markets).

Resources that are used by the aggregator are typically small in terms of the flexibility quantity, the aggregator's role is to gather the flexibility provided by distributed resources and forward it to the market. The aggregator combines individual capabilities and builds complex price-quantity bids. It is assumed that the aggregator controls a heterogenous flexibility portfolio containing dispatchable and intermittent renewable generation, energy storage and demand side flexibility providers (domestic and industrial) that change energy use from their current/normal consumption patterns in response to market signals.

The present paper intends to suggest a categorization of bidding approaches for the aggregator. Both flexibility categories and aggregation approaches are highly dependent on the services the aggregator provides, description of the markets needs to be defined [6].

Aggregator provides optimal dispatch of the portfolio and trading recommendations to maximize profit on the day-ahead and intraday energy markets for its own balancing group[3]. It can use the portfolio of its own

[3] Balancing Group: a group of market participants (consumers, producers, traders) who optimize costs by netting deviations (imbalances) and reduce overall deviations between the projected and reported electricity usage.

balancing group to minimize imbalance cost providing service for the balance responsible party[4] (BRP) of the balancing group near delivery time.

Transmission System Operators procures balancing reserves to have resources for load-frequency control. Aggregator can bid on both balancing capacity and balancing energy market offering frequency restoration reserves (FRR).

Distribution System Operators will be empowered to actively purchase local flexibility capabilities to mitigate voltage and grid congestion problems due to the increasing penetration of intermittent and distributed energy resources in the distribution system. Aggregator can bid on future local flexibility markets offering services for congestion management in the distribution network.

Aggregation approaches

Aggregation is a complex modelling and optimization task, it generates aggregated bids for the markets the aggregator is active on. Physical and dynamic models of the resources have to be developed that need to be simple enough for the optimization to generate bids. Among technical characteristics, a cost model belongs to all the resources. Both physical and cost model must consider the parameters of the market where the aggregator is bidding and the product that is available on the market. During operation, input data is collected right before the optimization runs. Operating condition, technical and price parameters of the portfolio assets, generation and consumption forecast, price forecasts are collected.

Aggregation of flexibility capabilities is performed automatically, aggregation technics have to be developed. In this paper three approaches are suggested and resource classification is proposed. As resource and cost models, aggregation approaches may vary depending on the flexibility resource. The aggregation of individual bids in all the suggested approaches is performed by summation of the selected individual bid curves.

1) **Bottom-up**. In the bottom-up approach it is assumed that the aggregator knows the status and parameters of the physical devices. Flexible power and cost are calculated for the individual devices and the aggregated bids are produced by the aggregator's own optimization problem. Each resource that is aggregated by the bottom-up approach must have a physical and cost model formulated. It is the

[4] Balance Responsible Party: a chosen representative of a balancing group who is responsible for the imbalance of the group.

preferred approach when the number of assets and the information allows it. Conventional generators and storages should be aggregated by a bottom-up approach as well as curtailable loads since the device characteristics are available and the number of such devices enables computationally tractable optimization.

2) **Profile-based**. When formulating individual resource models is not possible, permitted load profiles and cost can be used. Each device has one or more profile and associated cost defined, aggregator selects one of them when performs optimization. It is not as flexible or detailed as the bottom-up approach, but less data is necessary when generating optimal bids. It is suitable for resources with fix profiles that cannot be modified or stopped once started, like industrial processes and loads without physical device characteristics.
3) **Hybrid**. When a resource category contains large number of similar devices, a hybrid approach should be used. A virtual device that has a dynamic physical model, represents all resources of a group. It avoids high number of input parameters for the aggregation optimization. In order to reduce modeling errors, homogenous devices that have similar model parameters, have to form a group. It is the preferred approach for residential demand-response providers, e.g. heat pumps.

Classification of resources should be based on modelling similarity, resources that can be modelled and aggregated similarly can belong to the same group. Straightforward groups, generation, load and storage must be further detailed.

The generation group can be composed of intermittent generation (wind, solar) and conventional controllable generation (e.g. coal or gas fired power plants) due to different modelling requirements of a predictable/controllable and an intermittent unit with limited controllability. CHP (Combined Heat and Power) may be a separated group due to heat constraints that requires a different modelling technique.

Electric vehicles may be disjoint from the storage group, because the battery is not always connected to the grid and it is also used for travelling.

In the load group only flexible load is considered, it is assumed that the aggregator has no role to monitor or supply inflexible consumers. Flexible loads typically split into shiftable and curtailable load. A shiftable load can delay or advance its consumption, but due to the consumption characteristics, the profile is fixed. Industrial processes where different machines working together in a scheduled sequence have a determined power consumption

profile. Curtailable load can decrease and increase its consumption without any significant payback effect (e.g. lightning).

Conclusion

In this paper the transformation trends of energy sector have been presented and aggregator was introduced as a potential role to provide additional flexibility on the distribution grid. Markets were presented where the aggregator will be able to operate on and aggregation approaches were suggested that enables the aggregator to efficiently generate competitive aggregated bids combining flexibility capabilities of its small-scale generation and consumption portfolio.

Acknowledgement

We acknowledge the financial support of Széchenyi 2020 programme under the project No. EFOP-3.6.1-16-2016-00015.

References

[1] J. Merino, I. Gómez, E. Turienzo, C. Madina, I. Cobelo, A. Morch, H. Saele, K. Verpoorten, E. R. Puente, S. Häninnen, P. Koponen, C. Evens, N. Helistö, A. Zani and D. Siface, "Ancillary service provision by RES and DSM connected at distribution level in the future power system," SmartNet, 2016.

[2] E. Commission, "Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on common rules for the internal market in electricity," Brussels, 2017.

[3] H. d. Heer and M. v. d. Laan, "Recommended practices and key considerations for a regulatory framework and market design on explicit Demand Response," USEF: Workstream on aggregation implementation models, 2017.

[4] J. L. Baut, G. Leclercq, G. Viganò and M. Z. Degefa, "Characterization of flexibility resources and distribution networks (D1.2)," SmartNet, 2017.

[5] R. Verhaegen and C. Dierckxsens, "Existing business models for renewable energy Existing business models for renewable energy ExistinExisting business models for renewable energy aggregators," BestRES, Brussels, 2016.

[6] I. Balázs, A. Fodor and A. Magyar, "Aggregation of heterogeneous flexibility resources providing services for system operators and

market participants," Hungarian Journal of Industry and Chemistry (Accepted paper), 2019.

[7] M. Dzamarija, M. Plecas, J. Jimeno, H. Marthinsen, J. Camargo and J. Vardanyan, "Aggregation models," SmartNet, 2018.

A brief review on the challenges of Internet of Things and their solutions

Tibor Guzsvinecz[1], Tibor Medvegy[2], Veronika Szucs[3]
[1,3]Department of Electrical Engineering and Information Systems, University of Pannonia
[2]Institute of Physics and Mechatronics, University of Pannonia
8200, 10 Egyetem street, Veszprem, Hungary
[1]guzsvinecz@virt.uni-pannon.hu, [2]medvegyt@almos.uni-pannon.hu, [3]szucs@virt.uni-pannon.hu

Abstract: **Since its inception, the Internet of Things gathered a lot of attention and in the present, it is considered one of the largest fields of research in information technology. As the available devices on the internet reach a drastically large number, problems arise regarding networking, security and even data management. In this paper a brief review on the current challenges are presented, while mentioning possible solutions.**

Introduction

The Internet of Things (IoT) allows devices that either are or not internet enabled by default with the use of technology to be connected to each other using the internet. With this connection multiple methods of interaction arise between the devices, where several technologies can be used for communication, such as radio frequency identification (RFID), near-field communication (NFC), optical tags, quick response (QR) codes, Bluetooth low energy (BLE) [1].

As IoT is vast, it has several fields of use and research. The following fields were established in a study in 2015 [2]: Smart wearable, Smart home, Smart city, Smart environment, Smart enterprise. These smart fields, especially smart homes can be integrated with Ambient Assisted Living (AAL) which helps elderly people or people with disabilities in their homes, such as [3]. Since the study, the term Smart enterprise became the term Industrial IoT (IIoT) due to the emergence of the new Industry 4.0 [4] which consists of cyber physical systems instead of computers and automation, meaning that sensors gather data without human interaction.

IoT is growing at an alarming rate, according to Hsu and Lin the estimated number of IoT devices will reach 26 billion by 2020 [5] and other sources

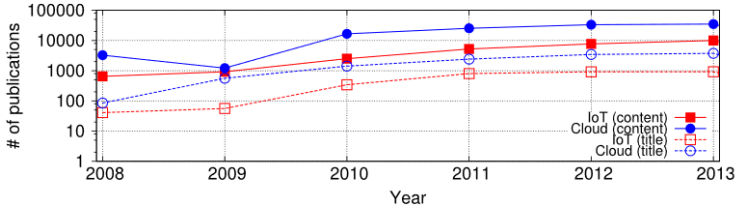


Figure 1. The number of publications containing IoT and Cloud-based systems [6].

predicting even more [7]. Due to this large number of devices on the internet the new generation of IP addresses will be used, namely the IPv6 instead of the IPv4 [1], but this growing fact can also lead to other, networking issues which will be elaborated upon in the next section. Since the number of IoT devices are increasing, the data generated from them are also increasing. Both the term and research field called "Big Data" is a result of this [8]. Also, Botta et al. [6] concluded that the number of publications of IoT and Cloud-based systems are converging to each other as can be seen in Figure 1. Due to data being on the cloud, the questions of security and privacy arise.

As can be seen, the problems exist due to IoT being heterogenous and because of the large number of devices, it rapidly generates large volumes of data. In the following section these problems will be discussed, referencing multiple reviews in the process, while also providing some possible solutions.

Discussion

This section deals with the mentioned problems and it is structured as follows: The first subsection is about the layers of IoT, the second is about Big Data, its measurement, management and analysis. The last is about privacy, security and other, networking issues.

Layers of IoT

IoT has three different layers: The cloud, the fog and the edge, the latter containing the extreme edge sublayer (also known as Mist), which refers to the sensors themselves. For graphical representation, see Figure 2.

Cloud computing allows information to be processed after uploading it. However, according to Shi et al. [9] data is produced at the edge of a network, which led to the introduction of edge computing where cloud connection is optional. As data is close to its source where the computation happens as well, the latency drastically decreases while mentioning that cloud computing is an inefficient way to manage data as – according to them – cloud computing

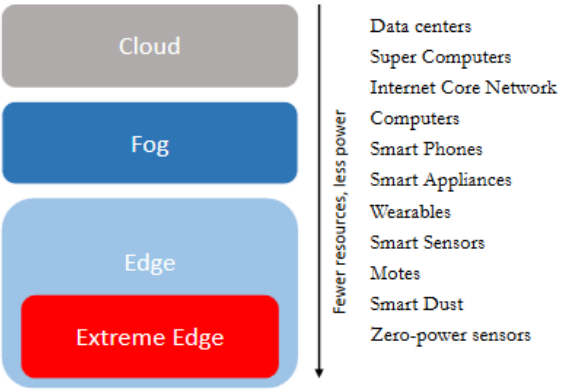


Figure 2. Layers of IoT [10].

cannot process large volumes of data due to bandwidth constraints. Also, as the data is not being uploaded into the cloud, edge computing is more secure. Its security mostly depends on the security of the user's network as long as the data does not leave the network. Fog computing [11] is almost identical to edge computing, however fog computing is virtualized and more information is processed in the cloud than on the edge of the network.

Big Data measurement, management and analysis

Measuring Big Data is not an easy task, as data has drastically large volumes: According to a recent review by the International Data Corporation (IDC) the total data volume (which is the sum of digital and non-digital data) in 2018 is 44 zettabytes ($44*10^{21}$ bytes) and IDC predicts that it will increase to 175 zettabytes by 2025 [12]. Hajirahimova and Aliyeva presents Big Data methodologies in [13], measuring paper, film, optical, magnetic and digital data. As Big Data can mostly be estimated, all methods yield different results.

Data management faces challenges consisting of quickly generating data, which made of large data of different types. In [14] a data storage framework is presented for IoT based on cloud computing. The framework differentiates structured and unstructured data while combining available databases and according to the authors, it provides a solution for the challenges.

When analyzing data, the metadata is just as important as the data itself. Analyzation has multiple steps [15]: Data acquisition is the first, where filtering or compression should be done, but carefully, not to discard any useful information. The second step is to extract data and format it to a useful, structured form that is understandable by the computer, which will be later working on the data. This leads to the challenges of data analysis [16], such

as errors in the algorithm or in the data, differences in the sensors or differing database types, huge computational power requirement and even human error, where the data format is so abstract that the user cannot understand it. Fahad et al. [17] compared clustering algorithms such as DENCLUE, OptiGrid, FMC, EM and BIRCH regarding their strengths and weaknesses in Big Data analysis and came to the following conclusions: There is no perfect algorithm – yet – for Big Data analysis and the users should choose an algorithm in accordance to the database and the data itself. Also, there is a stability problem in clustering algorithms and if possible, ensemble clustering should be chosen instead.

Privacy, security and networking

The main two issues of privacy and security are gaining unauthorized access to data and the artificial intelligence (AI) the IoT devices have. Heer et al. [18] concluded that the security architecture should take the lifecycle of a device (Figure 3.) and its abilities into account, while using a good security protocol design. Also, choosing the right layer is important as all have different security requirements. The aim is to secure the link layer, the network layer and the application layer. However, securing all three requires great computational power, therefore when designing security protocols, cross layer concepts should be used.

Additionally, according to Sicari et al. [19], security has three requirements: Authentication, confidentiality and access control. In their study, they assess different existing methods, such as encryptions, signature schemes, data mining techniques, frameworks for privacy and security in the

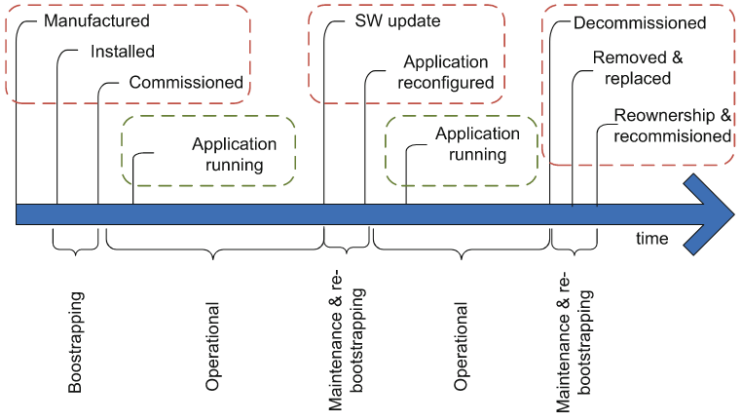


Figure 3. The lifecycle of a device, starting from manufacturing [18].

literature, though the methods are not perfect, each has its pros and cons, hence they do not answer every security liability as they provide different solutions. This is also true according to Dabbagh and Rayes [20] in a recent study as they identified the heterogeneity of IoT, scalability, Big Data, availability, response limitations, remote locations, mobility and delay-sensitive services as the security challenges and confidentiality, integrity, authentication, availability, authorization, freshness, non-repudiation, forward and backward secrecy as security requirements.

In early 2019 a new method called Misty Clouds is proposed [21], providing another possible solution and fulfilling most mentioned requirements. This solution provides network level anonymity, uses cloud-based environments and is based on the Mist protocol, previously made by the same authors. It uses an end-to-end encryption, protects IP addresses, the locations of the users, their used devices and activities. With this method, messages between the users do not contain route information meaning that the users can not be identified. It uses the RSA-2048 algorithm to generate keys and TLS Ephemeral Diffie-Hellman for router-to-router connections.

Conclusion

As more and more devices will be connected to the internet, IoT will certainly continue to grow, leading to larger volumes of data and networking issues. However, due to the heterogeneity of the devices, protocols, data types, et cetera, making a solution for every available problem is very difficult.

This trend also leads to introducing edge computing which can reduce latency of the system and can even decrease the volume of transmitted data. Reducing the transmitted data can ease the difficulty of Big Data measurement, management and analysis, which is a good thing, since no perfect method exist. There is a similar conclusion regarding the privacy and security of IoT systems: Multiple methods exits, but none of them are perfect. Only time will tell, but in the present, users have to choose the right method for the work they want to do.

Acknowledgment

We acknowledge the financial support of Széchenyi 2020 programme under the project No. EFOP-3.6.1-16-2016-00015.

References

[1] R. Want, B. N. Schilit, and S. Jenson. “Enabling the internet of things.” Computer, vol. 48, issue 1, pp. 28-35. 2015.

[2] C. Perera, C. H. Liu, and S. Jayawardena. “The emerging internet of things marketplace from an industrial perspective: A survey.” IEEE Transactions on Emerging Topics in Computing, vol. 3, issue 4, pp. 585-598. 2015.
[3] G. Marques, and R. Pitarma. “An indoor monitoring system for ambient assisted living based on internet of things architecture.” International journal of environmental research and public health, vol. 13, issue 11, 1152. 2016.
[4] M. Hermann, T. Pentek and B. Otto. “Design principles for industrie 4.0 scenarios.” In *System Sciences* (HICSS), 2016 49th Hawaii International Conference on System Sciences. 2016, pp. 3928-3937. IEEE.
[5] C.-L. Hsu, J. C.-C. Lin. “An empirical examination of consumer adoption of Internet of Things services: Network externalities and concern for information privacy perspectives”. Computers in Human Behavior, vol. 62, pp. 516-527. 2016.
[6] A. Botta, W. De Donato, V. Persico and A. Pescapé. “Integration of cloud computing and internet of things: a survey.” Future Generation Computer Systems, vol. 56, pp. 684-700. 2016.
[7] Popular Internet of Things Forecast of 50 Billion Devices by 2020 Is Outdated. Available online: https://spectrum.ieee.org/tech-talk/telecom/internet/popular-internet-of-things-forecast-of-50-billion-devices-by-2020-is-outdated (accessed on 01.26.2019)
[8] M. Ge, H. Bangui and B. Buhnova. “Big Data for Internet of Things: A Survey.” Future Generation Computer Systems. 2018.
[9] W. Shi, J. Cao, Q. Zhang, Y. Li and L. Xu. “Edge computing: Vision and challenges.” IEEE Internet of Things Journal, vol. 3, issue 5, pp. 637-646. 2016.
[10] J. Portilla, G. Mujica, J. S. Lee and T. Riesgo. “The Extreme Edge at the Bottom of the Internet of Things: A Review.” IEEE Sensors Journal. 2019.
[11] A. Rayes and S. Salam. “Fog computing.” Internet of Things From Hype to Reality, pp. 155-180. Springer, Cham. 2019.
[12] Data Age 2025: The Evolution of Data to Life-Critical. Available online: https://www.seagate.com/files/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf (accessed on 01.26.2019)
[13] M. S. Hajirahimova and A. S. Aliyeva. “About Big Data Measurement Methodologies and Indicators.” International Journal of Modern Education and Computer Science, vol. 9, issue 10, pp. 1-9. 2017.
[14] L. Jiang, L. Da Xu, H. Cai, Z. Jiang, F. Bu and B. Xu. “An IoT-Oriented Data Storage Framework in Cloud Computing Platform.” IEEE Transactions on Industrial Informatics, vol. 10, issue 2, pp. 1443-1451. 2014.
[15] A. Labrinidis and H. V. Jagadish. “Challenges and opportunities with big data.” Proceedings of the VLDB Endowment, vol. 5, issue 12, pp. 2032-2033. 2012.
[16] J. Fan, F. Han and H. Liu. “Challenges of big data analysis.” National science review, vol. 1, issue 2, pp. 293-314. 2014.
[17] A. Fahad, N. Alshatri, Z. Tari, A. Alamri, I. Khalil, A. Y. Zomaya, S. Foufou and A. Bouras. “A survey of clustering algorithms for big data: Taxonomy and empirical analysis.” IEEE transactions on emerging topics in computing, vol. 2, pp. 267-279. 2014.
[18] T. Heer, O. Garcia-Morchon, R. Hummen, S. L. Keoh, S. S. Kumar and K. Wehrle. “Security Challenges in the IP-based Internet of Things.” Wireless Personal Communications, vol. 61, issue 3, pp. 527-542. 2011.
[19] S. Sicari, A. Rizzardi, L. A. Grieco and A. Coen-Porisini. “Security, privacy and trust in Internet of Things: The road ahead.” Computer networks, vol. 76, pp. 146-164. 2015.
[20] M. Dabbagh and A. Rayes. “Internet of things security and privacy. In Internet of Things From Hype to Reality.” Springer, Cham. 2019, pp. 211-238.
[21] J. Al-Muhtadi, M. Qiang, K. Saleem, M. AlMusallam and & J. J. Rodrigues. “Misty clouds—A layered cloud platform for online user anonymity in Social Internet of Things.” Future Generation Computer Systems, vol. 92, pp. 812-820. 2019.

Application of Text Mining Methods on Unstructured Hungarian Echocardiogram Documents

Szabolcs Szekér[1], Ágnes Vathy-Fogarassy[2]
[1]University of Pannonia, Department of Computer Science and Systems Technology, szeker@dcs.uni-pannon.hu
8200 Veszprém, Egyetem utca 10.
[2] University of Pannonia, Department of Computer Science and Systems Technology, vathy@dcs.uni-pannon.hu
8200 Veszprém, Egyetem utca 10.

Abstract: **A variety of text mining applications has been described and discussed in detail in the biomedical literature, ranging from simple co-occurrence-based methods to more sophisticated rule-based systems or machine-learning-based systems. These methods are widely applicable to extract information from medical documents mainly written in English, however the nature of Hungarian language requires specific tools to extract information from medical documents written in Hungarian. The goal of our research was to create such a simple NER (named entity recognition) method that is able to find and identify terms and recorded measurement results in unstructured Hungarian medical records, namely in echocardiogram documents. In this paper, we discuss the How Tos and challenges of the extraction process of echocardiography reports written in Hungarian language and presents the logic and results of the developed text mining-based algorithm.**

Introduction

Extracting information from medical records is a challenging task based on that there is no unified process on how to record patient data, the structure and form of the recorded information varies from medical institute to medical institute and the habits of the medical assistant and doctors also affect the recording process. Also the lack of a unified recording interface induces typographical errors. For the aforementioned reasons, the Electronic Medical Record (EMR), which stores information about patients, including diagnostics and performed treatments, is usually incomplete or redundant. It is especially true in case of echocardiography reports. The focus of present paper is on the process of information extraction from echocardiography reports written in Hungarian language.

Echocardiogram is a sonogram of the heart. It is one of the most widely applied diagnostics test in cardiology: routinely used in diagnosis, management and follow-up of patients with any suspected or known heart disease. Echocardiography reports usually contain two parts in terms as diagnostic content: a semi-structured part where results are usually stored in term-value pairs (e.g.: EN: *Septum: 14 mm*, HU*: Szeptum: 14 mm*) and a free text part written in natural language (e.g.: EN: *Left ventricular hypertrophy* HU: *Koncentrikus bal kamra hypertrophia*). As there is no consensus about how to store the results of echocardiography examinations and it is varying across different medical institutes, processing of echocardiography reports is a nontrivial task.

Generally, the main task of information extraction from medical texts is to identify particular types of names, terminologies or symbols (term extraction, named-entity recognition, NER) and the relation between them (relation extraction, RE) [1]. Successful term identification is essential and has been recognized as a bottleneck in text mining [2]. The process of term identification is usually achieved in three steps: the first step is term recognition where possible term candidates are identified; the second step is term classification where the candidates are classified based on pre-defined rules; and the last step is term mapping where the candidates are checked whether they are valid terms or not [2].

Various international studies have been published in the literature which are engaged in echocardiography report processing [3-10]. Typically, only the extraction of one specific parameter is the aim (e.g. EN: *ejection fraction*, HU: *ejekciós frakció*), but it also possible to extract a set of predefined parameters, including wall thicknesses, chamber dimensions or flow velocities. The common factor in all published studies is that they process reports written in foreign language, and to our best knowledge there are no published results that process echocardiography reports written in Hungarian. Present paper is focusing on the How Tos and challenges of the extraction process of the first, semi-structured part of echocardiography reports written in Hungarian language. As processing of free text requires quite different methods, including Natural Language Processing (NLP) techniques, we do not deal with them in this paper. The extracted information can be further processed and fed into modern data mining algorithms to reveal cause and effect relationships between different parameters.

The structure of this document is as follows. In Challenges, we give a brief overview of challenges faced when processing Hungarian echocardiography reports. In Results, the used dataset, the applied methods and the evaluation

process are described and the result of the analysis is presented. Finally, in Conclusion, general experiences are described.

Challenges

The first, semi-structured part of echocardiography reports contains medical information in *term–value* pairs separated by colon. The *term* part refers to the identifiable named entities and the *value* part refers to the measured and recorded value for that named entity. The measured value may also contain a unit of measurement. However, based on the extraction approach, various challenges emerge during term extraction, aside from typographical errors. These challenges are described in detail in the following subsections.

Articles

A common characteristic of the English and the Hungarian language is the use of "*a*" article before adjectives and nouns (in Hungarian "*a*"/"*az*" pair is used and in English "*a*"/"*an*" pair is used). In most case the use of the "*a*" article does not pose a problem, however, in case of echocardiogram reports, *A* (A wave) is the peak velocity flow in late diastole caused by atrial contraction. Furthermore, in Hungarian language "*e*" expletive is also present, but in echocardiography reports *E* (E wave) stands for the peak velocity blood flow from gravity in early diastole.

Missing whitespaces

As a form of typographical error, missing whitespaces can also occur between terms, values and units (e.g. EN: *Left ventricular end-diastolic diameter43.: mm*, HU: *Bal kamra diast.átm43.: mm*). If the text processing method is word-based, missing whitespaces have a huge impact on the success of processing. This problem can be handled by inserting separator space characters into the text during text cleaning, if text cleaning is applied.

Recognition of composite terms

Not only typographical errors make it harder to extract information from echocardiography reports. Based on the assumption, that named-entities follow the term-value pair structure, it is possible to extract the greater part of named entities. However, special cases are also present in echocardiography reports mostly because of the habits of the recording individual. Such a composite term can be described in *prefix–term1–term2–value1–value2–common_unit* form (e.g. EN: *left ventricular end-diast/end-syst diameter: 54/35 mm*, HU: *bal kamra diast/syst átmérő: 54/35 mm*) where the recording individual aggregates two somewhat related terms. In this case

the identified term should be interpreted as *prefix–term1–value1–unit* and *prefix–term2–value2–unit*. Another example is the *prefix–term1–value1–unit–term2–value2–unit* form of the expression (e.g. EN: *ejection fraction Teichholz: 56%, Simpson 52%*, HU: *ejekciós frakció Teichholz: 56%, Simpson: 52%)* or the *term1–term2–value1–value2–unit* sequence (e.g. *E/A: 0.4/0.8 m/s*). Furthermore, expletives are also commonly used (e.g. EN: *left atrium: 42 mm (apical 4Ch: 43x75mm)*, HU: *Bal pitvar: 42 mm (csúcsi nézetből: 43x57 mm)*) which makes composite term recognition even harder. A possible approach to process composite terms is to define some basic rules and process echocardiography documents based on these rules.

Typographical errors

The lack of a unified recording interface infers many typographical errors which need to be taken into account during term extraction. Most frequent typographical errors, if they are within an acceptable margin, can be resolved by using a dictionary during the term mapping process. The identified and classified term candidates must be checked whether they are valid terms or not using the aforementioned dictionary of terms. This dictionary is usually created with the help of a medical expert and contains more terms from the field of interest, in our case the field of cardiology, probably present in some form in echocardiography reports and defines synonyms for the terms.

The extracted terms can be compared against the values of the dictionary. A widespread measure of distance for two strings is the Jaro-Winkler distance [11]. If the measured distance is lower than a specified distance threshold, the term is considered valid and identified. This threshold parameter can be the lowest, non-zero intra-distance of the terms stored in the dictionary. The Jaro-Winkler distance (d_w) can be calculated the following way:

$$d_w(s_1, s_2) = 1 - sim_w(s_1, s_2) \tag{1}$$

$$sim_w(s_1, s_2) = sim_j(s_1, s_2) + lp(1 - sim_j(s_1, s_2)) \tag{2}$$

where sim_j is the Jaro similarity for s_1 and s_2 strings, l is the length of a common prefix up to 4 characters and p is a constant scaling factor with a standard value of 0.1. The Jaro similarity (sim_j) is calculated the following way:

$$sim_j(s_1, s_2) = \begin{cases} 0 & \text{if } m = 0 \\ \frac{1}{3}\left(\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m-t}{m}\right) & \text{otherwise} \end{cases} \tag{2}$$

where $|s_i|$ is the length of s_i, m is the number of matching characters and t is half the number of transpositions. The concept of matching and transpositions is detailed in [11].

Results

In our study, a text mining-based NER algorithm (TM-NER) was used to process a corpus containing 20089 anonymized Hungarian echocardiography reports. TM-NER processes the report in 3 phases: the first phase is a thorough text cleaning process where whitespaces are unified and all colons, parenthesis and unneeded characters are removed; the second phase is term identification based on predefined rules; and the third phase is term mapping using a dictionary. TM-NER was developed in such a way that it manages to solve all problems discussed in Challenges.

In our study, each document had a unique identifier and a basic description about the diagnosis. The first, semi-structured part of the documents contained various terms mainly in the form of *term-value* pairs separated by a colon. The echocardiography reports under study also included typographical errors or deficiencies. The created dictionary of terms for term mapping contained almost 40 terms from the field of cardiography and over a 100 synonyms have been defined for the terms.

Results were evaluated by counting the number of documents in which TM-NER has found a specific term. The number of matched documents was noted N_{TM}. To evaluate the relative success (f_{TM}) of TM-NER we also calculated the frequency of matched documents relative to the size of the corpus (20089). The results can be seen in Table 1.

Table 1 The most commonly extracted terms by TM-NER. It is important to note that the presented rate (f_{TM}) is a function of the size of the corpus and not a function of the number of reports actually containing a specific term.

term	***occurrence (N_{TM})***	***rate (f_{TM})***
Bal kamra syst. átm.	19549	97.31%
Septum végdiast.	19491	97.02%
Aorta gyök	19476	96.95%
Hátsófal végdiast.	19386	96.50%
Bal kamra diast. átm.	19147	95.31%
Bal pitvar (M-mode)	19144	95.30%
E	18723	93.20%
EF	18135	90.27%
A	17483	87.03%

Examples for extracted and identified terms can be seen in Table 2. It shows various term candidates found in the echocardiography reports and their number of occurrence.

Table 2 Examples for candidate terms for *bal pitvar csúcsi nézetből* and their number of occurrence.

candidate	***# of occurrence***
bal pitvar csúcsi nézetből	4930
bal pitvar csúcsi négyüregből	525
bal pitvar csúcsi négy üregi nézetből	172
bal pitvar csúcs felől	128
bal pitvar csúcsi nézet	68
bal pitvar csúcsi négy üregből	19
bal pitvar csúcs nézetből	14
bal pitvar cssvar nal pit	10

As previously mentioned, the last step of term extraction is term mapping. Table 3 shows some examples of term candidates, their measured value and the mapped term.

Table 3 Term candidates with their extracted value and the mapped term. The table shows a variety of extracted values.

term	***candidate***	***value***
Ao vmax.	*aorta kiáramlás*	132cm/sec
Bal kamra diast. átm.	*dd*	57mm
Bal kamra syst. átm.	*ds*	40mm
2D bal pitvar	*bal pitvar csúcsi nézetből*	43x57mm
E	*e*	1,6m/sec
A	*a*	0,65m/sec
2D jobb pitvar	*jobb pitvar*	31x45mm

Conclusion

Based on the previously presented results, it is safe to say that term extraction is demanding, but with the proper extraction method it is possible to extract medical terms on an acceptable level. The developed text mining-based TM-NER processes echocardiography reports in 3 phases: the first phase is a thorough text cleaning process, the second phase is term identification based on predefined rules and the third phase is term mapping

using a dictionary. For a text mining-based method such as the TM-NER, the quality of the pre-cleaning process greatly affects the outcome.

Our results show, that by identifying the occurring challenges during processing of Hungarian echocardiography reports and by utilizing proper text mining techniques it is possible to tackle the aforementioned problems. Based on the promising results, fine-tuning of the developed text mining-based NER is part of our future research to make it more suitable to process Hungarian echocardiography reports.

Acknowledgment

We acknowledge the financial support of Széchenyi 2020 under EFOP-3.6.1-16-2016-00015 and UNKP-18-3 New National Excellence Program of the Ministry of Human Capacities, and the professional support of GINOP-2.2.1-15-2016-00019 "Development of intelligent, process-based decision support system for cardiologists".

References

[1] Wencheng Sun, Zhiping Cai, Yangyang Li, et al, Data Processing and Text Mining Technologies on Electronic Medical Records: A Review, Journal of Healthcare Engineering, 2018:4302425

[2] Krauthammer M, Nenadic G, Term identification in the biomedical literature. J Biomed Inform, 2004, pp. 512–526.

[3] Xie F, Zheng C, Yuh-Jer Shen A, Chen W, Extracting and analyzing ejection fraction values from electronic echocardiography reports in a large health maintenance organization, Health Inform J, 2017, pp. 319–328.

[4] Garvin JH, DuVall SL, South BR, et al, Automated extraction of ejection fraction for quality measurement using regular expressions in Unstructured Information Management Architechture (UIMA) for heart failure, J Am Med Inform Assoc, 2012, pp. 859–866.

[5] Kim Y, Garvin JH, Goldstein MK, et al, Extraction of left ventricular ejection fraction information from various types of clinical reports, J Biomed Inform, 2017, pp. 42–48.

[6] Patterson OV, Freiberg MS, Skanderson M, et al, Unlocking echocardiogram measurements for heart disease research through natural language processing. BMC Cardiovasc Disord, 2017:17:151

[7] Wells QS, Farber-Eger E, Crawford DC, Extraction of echocardiographic data from the electronic medical record is a rapid and efficient method for study of cardiac structure and function, J Clin Bioinforma, 2014:4:12

[8] Toepfer, M., Corovic, H., Fette, et al, Fine-grained information extraction from German transthoracic echocardiography reports, BMC Medical Informatics and Decision Making, 2015:15:91

[9] Jonnalagadda, S.R., Adupa, A.K., Garg, R.P. et al, Text Mining of the Electronic Health Record: An Information Extraction Approach for Automated Identification and Subphenotyping of HFpEF Patients for Clinical Trials, J of Cardiovasc Trans Res, 2017, pp. 313–321.

[10] Renganathan, V, Text Mining in Biomedical Domain with Emphasis on Document Clustering, Healthcare Informatics Research, 2017, pp. 141–146.

[11] Piskorski J., Sydow M, String Distance Metrics for Reference Matching and Search Query, Business Information Systems, 2007, pp. 353-365.

A machine learning algorithm for automatic structure detection and pattern analysis

Zsolt Vassy[1], István Vassányi[1]

[1]University of Pannonia, Medical Informatics R&D Centre

H-8200 Veszprém, Egyetem u. 10, Hungary

zsolt.vassy@gmail.com

Abstract: **ADIOS is a pattern acquisition algorithm that learns from a large-scale natural data set, in an unsupervised fashion. The process of acquisition of patterns is driven both by structural similarities and by statistical information inherent in the data. The original ADIOS method was designed for language acquisition. We improved the algorithm to use any pattern, thus it is more suitable to use in other domains like biology, and at the same time based on an earlier pattern detection method we made the algorithm more effective.**

Introduction

The statistical-structural algorithm for unsupervised language acquisition, ADIOS (for Automatic DIstillation Of Structure) [1], can identify effectively grammars of realistic and unannotated corpus data, in languages as diverse as English and Mandarin using a network science-based learning algorithm.

ADIOS is an information flow-driven pattern acquisition algorithm that learns from a large-scale natural data set in an unsupervised fashion. In this work we wanted to exploit these features of the algorithm in different domains. A common characteristic in a linguistic corpus (a set of sentences) and a biological network is that both are directed pseudographs.

The original ADIOS algorithm looks for only one pattern (motif) type, but both in biological and linguistic networks other types of motifs exist [2, 3]. Our goal was to preserve ADIOS's ability to search for independent motifs in a behavior-driven way and to make it search for various types of arbitrary motifs.

Methods

The original ADIOS algorithm is designed for unsupervised language/grammar detection, using an input corpus to generate an estimated grammar of the language used by the corpus. By 'grammar' we mean a set of structural rules governing the composition of phrases and clauses.

In the initialization step the algorithm loads the corpus into a directed graph such that a node in the graph will be allocated for every unique word in the corpus. An ordered sequence of graph edges (a path) depicts a sentence in the corpus i.e. if the words in a sentence follow one after another, we place edges among their nodes

After initialization, the body of the algorithm consists of 3 main steps [2, 5-7]. In the description below, we marked with *italic* type face the points that have been modified for multipurpose use.

1. After the loading, we execute a Pattern Acquisition algorithm (PA). The PA algorithm identifies Representational Data Structures (RDS) via an unsupervised learning algorithm as follows.

A. Pattern detection. In the original ADIOS algorithm patterns are similarly structured sentences of primitives that are repeated in the corpus (see Figure 1.).

Our algorithm is capable of applying freely defined patterns. These patterns can be described algorithmically (a few steps crawling algorithm, starting from an entry point) or can be described with their nodes and edges (a subgraph of the input graph specified in xml).

B. Formation of 'equivalence classes' (EC) for each pattern. An EC is a set of alternative entities. For example, in a linguistic application, they can be nouns, adjectives, verbs with a similar role (highlighted in red in Figure 1.).

In the modified algorithm this step is optional, with freely defined patterns it is not necessary to have nodes with the same function.

C. The algorithm estimates the probabilities of different paths, yielding a path ranking based on the signal transmission probability. The pattern with the highest score will be the Significant Pattern (SP).

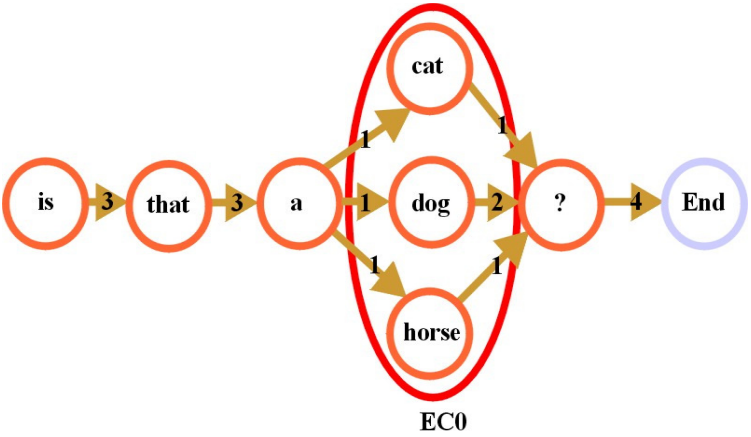
D. The equivalence class of SP is reduced to a single new node i.e. we remove the nodes that form the EC and we introduce a new node (called EC0 in Figure 1).

In the modified algorithm this step is also optional.

2. When one PA step is finished, the input data set is recomposed in a new generalized form. For the details of this step, see [1].

3. The PA algorithm is applied again to the recomposed corpus.

Figure 1: A pattern example detected by ADIOS. The EC0 Equivalence Class together with the nodes before ("is", "that", "a") and after it ("?", End) represent the pattern.



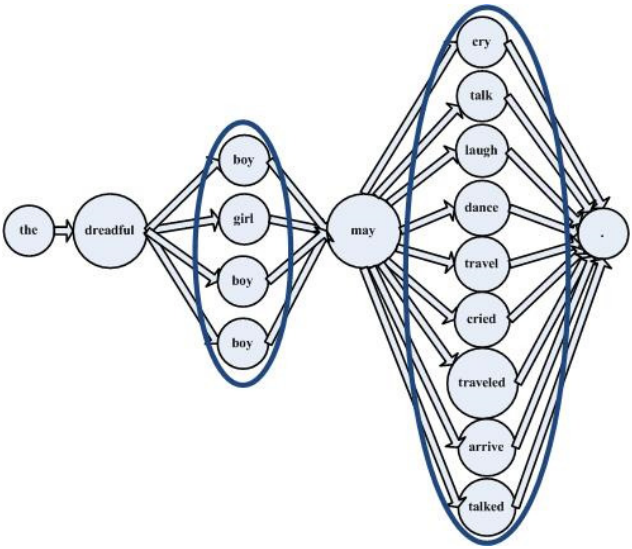
The algorithm halts if it processed a network without finding a new pattern.

This methodology finds independent patterns using a behavior-driven ranking of patterns, which is important in understanding both linguistic and biological networks.

Applying generalized patterns

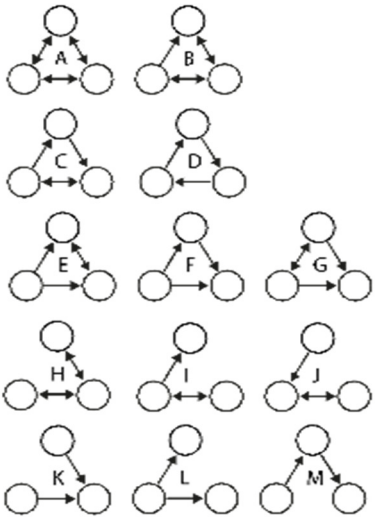
The original algorithm detects structures similar to that shown in Figure 2, which is appropriate for linguistic data sets, but we wanted to develop a more general solution.

Figure 2: A typical structure provided by original method



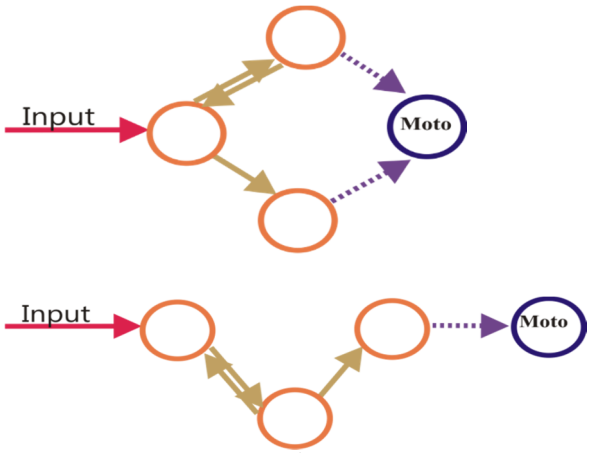
In neurobiology, a neural network is directed, but the typical patterns (motifs) are different from those of linguistic networks. An example is a Triangle Pattern (see Figure 3).

Figure 3: Example Triangle Patterns. Such patterns are typically used in biological networks to describe regulatory processes



Note that a triangle pattern in this information-flow driven approach has well-defined input and output nodes (see Figure 4). The modified algorithm still uses the ADIOS learning method to learn the network, but it can detect any type of pattern that is assumed to play a role in regulating or building the network.

Figure 4: Example patterns for sensory-motoric pathways in a biological neuron network. The stimulus enters the pattern from the sensory neuron from the Input node and exits to the motoric neuron at the Output node. The Input node could be a sensory neuron and the Output node could be a



motoric neuron.

The modified algorithm successfully finds such patterns, and also those possible directed pathways that enter the pattern at the input node and leave the pattern at the output node.

Improving the performance of the algorithm

Based on the position of the identified patterns with a path traversal algorithm (used backward from the input node and forward from the output neuron) we drastically reduced the computation needed to calculate the probabilities of different paths that affect the pattern.

For high density networks (the density of a network being defined as a ratio of the number of edges to the number of possible edges in a network) we also added the feature to use the algorithm just in some selected pathway groups, for example sensory-motoric pathways in a neural network.

Application of the modified algorithm for a biological network

We applied the modified algorithm in the neural network map (connectome) of the nematode Caenorhabditis Elegans [8]. Such a neural network can be described via an edge list of inter-neural interactions. The biological background of this application can be summarized as follows:

- Information flows between two nodes, has a source and a target.
- The fundamental direction of information flow in the connectome network is from sensory to motor neurons
- Information flow in the connectome can be modelled as a probability chain network

The connectome from the WormWiring Project [9] contains:

- 302 neurons (nodes of the network) of types sensory, inter and motor
- 5039 connections (edges of the network)

We represented the chemical synapse with one unidirectional edge, gap junctions with two unidirectional edges.

As an output the implemented new algorithm provides a functional, behavior-driven pattern distribution of the C. Elegans connectome network.

Results

Performance

We used the Brown corpus [11] of ca. 1 million words to compare the original algorithm downloaded from the official ADIOS project webpage [12] to the proposed algorithm. Architecture used for these tests:

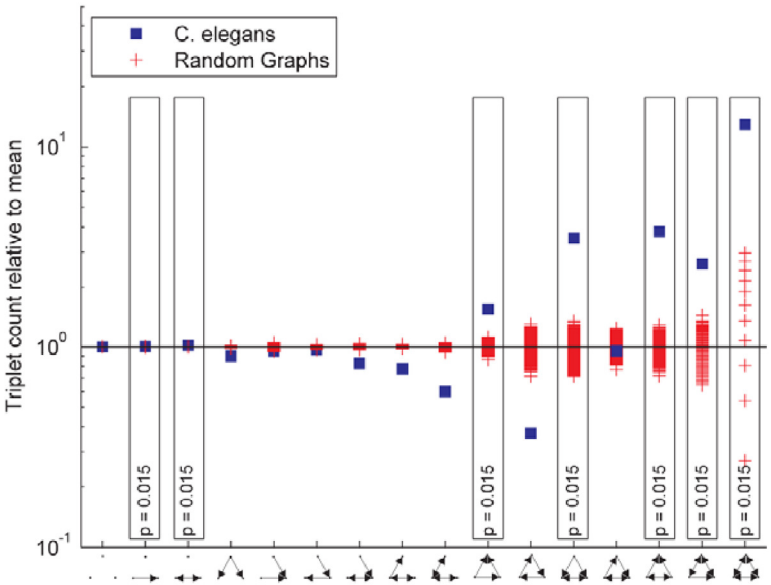
- Intel Core i7-6700HQ CPU @ 2.6Ghz
- Physical Memory: 16.0 GB

The results in Table 1 show a considerable speed-up.

Table 1: Run times of the original and the modified algorithm on the Brown corpus

Algorithm	**Running time (min)**
Original ADIOS	870.7
Modified (path traversal based) ADIOS	256.3

Figure 4: Triangle pattern distributions in the C. Elegans Connectome



Results in C. Elegans connectome network

With the proposed changes to the original ADIOS algorithm we were able to analyze the connectome of the Caenorhabditis elegans nematode, to determine the triangle-pattern distributions of different sensory-motoric pathways. Previously, triangle patterns were only analyzed on the entire connectome network on a morphological basis, without considering directed information flow from sensory to motoric neurons [12].

With the proposed algorithm we could compute a more meaningful pattern distribution (see Figure 4) that can be a basis of further research in neuro science. The triplet counts (in blue) are compared to the counts in a random network with the same number of nodes and edges.

Summary

ADIOS as a path based directed motif search algorithm successful comparable to a classic motif search in a well-known biological network.

With directed motif search we refine a more functional differentiation of patterns. Our method gives the possibility to use a data-driven approach to find which patterns have a significant role in the real signal propagation, these patterns can play an important role in regulation.

Acknowledgment

We acknowledge the financial support of Széchenyi 2020 programme under the project No EFOP-3.6.1-16-2016-00015.

References

[1] Solan, Z., Horn, D., Ruppin, E., and Edelman, S. (2005). Unsupervised learning of natural languages. Proceedings of the National Academy of Science, 102:11629–11634.

[2] Milo, R; Itzkovitz, S; Alon, U. et al. Superfamilies of evolved and designed networks. Science, 303(5663): 1538 – 1542, 2004.

[3] Berant, J., Gross, Y., Mussel, M., Sandbank, B., & Edelman, S. (2007). Boosting Unsupervised Grammar Induction by Splitting Complex Sentences on Function Words. In Proceedings Of The 31st Annual Boston University Conference On Language Development (pp. 93–104). Cascadilla Press, Somerville, MA.

[4] Falk Schreiber and Henning Schwöbbermeyer Motifs in Biological Networks Statistical and Evolutionary Analysis of Biological Networks, pp. 45-64 (2009)

[5] David Horn, Zach Solan, Eytan Ruppin and Shimon Edelman Unsupervised language acquisition: syntax from plain corpus. Newcastle Workshop on Human Language, Feb. 2004

[6] Shimon Edelman, Zach Solan, David Horn, Eytan Ruppin, Unsupervised e-cient learning and representation of language structure, Proc. CogSci-2003, Boston, MA, May 2003.

[7] Zach Solan, David Horn, Eytan Ruppin, Shimon Edelman. Unsupervised Context Sensitive Language Acquisition from a Large Corpus. Proc. NIPS-2003, Dec. 2003

[8] White, J.G. & Southgate, E & Thomson, J.N. & Brenner, S. (1986). The Structure of the Nervous System of the Nematode Caenorhabditis elegans. Philosophical Transactions of

the Royal Society B. Philosophical transactions of the Royal Society of London. Series B, Biological sciences. 314. 1-340. 10.1098/rstb.1986.0056.

[9] Steven Cook, Christopher Brittin, David Hall and Scott Emmons1 WormWiring: A new online resource for nematode connectivity http://wbg.wormbook.org

[10] Francis, W. N. and H. Kučera. 1964. Manual of Information to accompany A Standard Corpus of Present-Day Edited American English, for use with Digital Computers. Providence, Rhode Island: Department of Linguistics, Brown University. Revised 1971. Revised and amplified 1979.

[11] http://adios.cs.tau.ac.il/algorithm.html

[12] Structural properties of the Caenohabditis elegans neuronal network. Varshney LR Chen BL, Paniague E, Hall DH, Chklovskii DB PloS Comput Biol. 2011 Feb 3; 7(2):e1001066