Proceedings in Informatics and Information Technologies

Student Research Conference 2019

Michal Kompan and Pavol Návrat (Eds.)

Keynote by Michel Chaudron

2019 iit src

STU FIIT

SPEKTRUM STU

Proceedings in
Informatics and Information Technologies

IIT.SRC 2019
Student Research Conference

Michal Kompan and Pavol Návrat
(Eds.)

IIT.SRC 2019: Student Research Conference

15th Student Research Conference
in Informatics and Information Technologies
Bratislava, April 17, 2019
Proceedings

IEEE Czechoslovakia Section
ACM Slovakia Chapter

STU FIIT
SLOVAK UNIVERSITY OF TECHNOLOGY IN BRATISLAVA
FACULTY OF INFORMATICS AND INFORMATION TECHNOLOGIES

Proceedings in
Informatics and Information Technologies

IIT.SRC 2019
Student Research Conference

Editors

Michal Kompan, Pavol Návrat
Faculty of Informatics and Information Technologies
Slovak University of Technology in Bratislava
Ilkovičova 2
842 16 Bratislava, Slovakia

Visit IIT.SRC on the Web: http://iit-src.stuba.sk/2019/

Executive Editors: Katarína Mršková
Copy Editor: Peter Gašpar
FIIT STU

Cover Designer: Peter Kaminský

FIIT STU
Ilkovičova 2, Bratislava, Slovakia

Published by Slovak University of Technology in Bratislava in SPEKTRUM STU

2019

ISBN 978-80-227-4950-3

Preface

Research has been one of the main priorities of the university education since its very beginning. It is the case also for our university – the Slovak University of Technology in Bratislava and its faculty – the Faculty of Informatics and Information Technologies. Close connection of research and education leads very naturally to an active participation of students in research. This holds not only for students of the doctoral study, where research is a substantial part of their study and one of their principal activities. A participation of students in research is "going down" to students of the master's and even the bachelor study.

Universities of technology have a long tradition of students participating in a research work where they have to apply their theoretical knowledge. Outcomes of such endeavours have usually been presented at various students' competitions or exhibitions. Our university has a long tradition in organizing such events. However, in 2005 we decided to upgrade the framing of the event by transforming into a Student Research Conference covering topics of Informatics and Information Technologies (IIT.SRC). Participants are students of all three levels of studies, i.e. they recruit from both the undergraduate and postgraduate studies. The conference has adopted a process of reviewing as is usual at any other scientific conference. The accepted papers are available to all participants in a form of printed preprints (Proceedings). The aim of reviewing is not so much in achieving a certain acceptance rate, which is currently rather common in our community, but in assisting the student authors to amend their papers in the process of revision by providing helpful reviews. Of course, filtering out papers with a too low quality happens, too.

This is only part of our effort to support students in their (often first) steps in communicating their results to the research community at large. We also encourage students to improve their papers and attempt to publish them in international journals or conferences to make them available to their elder peers. Especially, we support the youngest generation of researchers (bachelor and master's students) by special travel grants that cover partially their travel expenses to conferences.

IIT.SRC 2019 attracted 103 student papers from which 97 were accepted as research papers (28 bachelor, 51 master, 18 doctoral) and 6 as papers to the innovative application and technologies track. The number of papers slightly varies each year. This year we have noticed a increase in bachelor and master categories comparing to IIT.SRC 2018.

The research track of the IIT.SRC 2019 conference was organized in 5 sections presented in course of live discussion in two poster sessions:

- Intelligent Information Processing and Data Analysis (15 papers),
- Machine Learning and Optimization (33 papers),
- Web Science and Engineering (16 papers),
- Software Engineering (13 papers),
- Information Security, Computer Networks and Computer Systems (19 papers).

Papers in Innovative applications and technologies track were presented in a demo session, where student authors presented their applications online.

The conference was opened by Professor Michel Chaudron's keynote titled Empirical Studies into the Effectiveness of Modelling in Software Design. Michel Chaudron is a professor at the Software Engineering division which is part of the joint Department of Computer Science of Chalmers and Gothenburg University in Sweden. His research interests focus software architecture, software design, software modelling with a special focus on UML, software composition.

Besides the 103 papers presented at the conference and included in these Proceedings several accompanying events were organized. This year we organized for the ninth time as part IIT.SRC a showcase of TP-Cup projects. TP-Cup is a competition of master students' teams aimed at excellence in development information technologies solutions within two semesters long team project module. The competition has three stages. 11 teams managed to achieve this stage and presented their projects during the TP-Cup showcase. Extended abstracts of their projects are included in these proceedings.

We continued this year with FIITApixel exhibition. FIITApixel brings together both students and staff of the Faculty as well as its potential students and alumni in an effort to create, share and judge pictures. It is organized as an ongoing event, where anyone can contribute pictures. The IIT.SRC FIITApixel exhibition presented the best pictures of this year contest.

IIT.SRC 2019 was for the seventh time organized in the new FIIT building. We all benefited from well-disposed space, which supports lively discussions. IIT.SRC 2019 is the result of considerable effort by a number of people. It is our pleasure to express our thanks to:

- members of the IIT.SRC 2019 Programme Committee who devoted their effort to reviewing papers and selecting awards,
- members of the IIT.SRC 2019 Organising Committee and accompanying events coordinators (mentioned in particular reports in these proceedings) for a smooth preparation of the event,
- the students – authors of the papers, for contributing good papers reporting their research and their supervisors for bringing the students to research community.

Special thanks go to:

- Katarína Mršková and Peter Gašpar who did an excellent job in the completion of the proceedings,
- Zuzana Marušincová and the whole organizing committee for effective support of all activities and in making the conference happen.

Finally, we highly appreciate the financial support of our sponsors which helped the organizers to provide excellent environment for presentation of the results of student research and valuable awards.

Bratislava, April 2019

Michal Kompan and Pavol Návrat, jointly with Mária Bieliková

Conference Organisation

The 15th Student Research Conference in Informatics and Information Technologies (IIT.SRC), held on April 17, 2019 in Bratislava, was organised by the Slovak University of Technology (and, in particular, its Faculty of Informatics and Information Technologies) in Bratislava.

General Chair

Mária Bieliková (dean, Faculty of Informatics and Information Technologies, Slovak University of Technology in Bratislava)

Programme Chairs

Michal Kompan, Pavol Návrat

Programme Committee

Nadežda Andrejčíková
Michal Barla
Rastislav Bencel
Vanda Benešová
Miroslav Blšták
Anna Bou Ezzeddine
Gabriela Grmanová
Ján Hudec
Ladislav Hudec
Daniela Chudá
Katarína Jelemenská
Peter Kapec
Ondrej Kaššák
Ivan Kotuliak
Tibor Krajčovič
Tomáš Kramár
Štefan Krištofík
Eduard Kuric
Peter Lacko
Ján Lang
Ján Laštinec
Marián Lekavý
Mária Lucká
Dominik Macko
Martin Madaras
Alena Martonová
Róbert Móro
Irina Malkin Ondik
Peter Pištek
Ivan Polášek
Viera Rozinajová
Karol Rástočný
Ivan Srba
Petr Šaloun
Jakub Ševcech
Jakub Šimko
Marián Šimko
Peter Trúchly
Jozef Tvarožek
Valentino Vranić

Organising Committee

Zuzana Marušincová, Chair
Katarína Mršková
Zuzana Horniaková
Zuzana Macková
Zuzana Kozíková
Ľubica Jančaťová
Michal Kompan
Ľubica Palatinusová
Viera Rozinajová
Branislav Steinmüller
Roman Stovíček

all from FIIT STU in Bratislava, Slovakia

Accompanying Events Chairs

- *TP CUP:* Mária Bieliková
- *FIITApixel Exhibition:* Ján Lang

Supporting Professional Societies and Foundations

- ACM Slovakia Chapter
- Czechoslovakia Section of IEEE
- Slovak Society for Computer Science
- Informatics Development Foundation at FIIT STU

Industry sponsors

Studies into the Effectiveness of Modelling in Software Development

Michel CHAUDRON

Joint Department of Computer Science
Chalmers and Gothenburg University in Sweden
Göteborg, Sweden
Michel.Chaudron@cse.gu.se | Chaudron@chalmers.se

Abstract. Modelling is a common part of modern day many engineering practices. However, little evidence exists in Software Engineering about how models are made, how models are used and how they help in producing better software. In this talk, I will present highlights from my last 15+ years of research in the area of software modelling, model-driven development and UML. Topics that will be addressed:

- How are models used in software development? Including how the use and purposes of models evolve over time.
- A discussion of arguments in favour and against software modelling
- Do UML models actually help in creating better software?
- Challenges and future directions

Also, I will overview some of the ongoing projects including our work on creating a dataset of almost 100.000 UML models[1].

1 Introduction

All systems that are engineered go through some process of designing. Such designs may or may not be represented explicitly by the engineering team. A common way to represent a design is through models. Models are a *systematic* way of representing a system. The *systematic* aspect of models comes from a well-defined language, both in terms of syntax and semantics. Another property of models is that they are an abstraction: they represent some aspects of a system, but leave out aspects. Typically aspects are left out because they are considered irrelevant from the perspective of the model. Moreover, the concepts that can be expressed in a modelling language satisfy some *system*: in particular for software modelling, the concept in the semantics are of the same abstraction level. Models may be expressed in textual form or graphical form, or both. To clarify the meaning and encourage precise use of these terms, I will use the following definitions:

Definitions:
Design (verb) = the process of making decisions about something that is to be built or created**.**
Design (noun) = the plans, drawings, etc., that show how something can be made.
Model (noun) = an abstract systematic representation of a thing.
Modelling (verb) = the process of creating a model.

Hence modelling requires that the model-author chooses what to represent. Sometimes it is clear which language is appropriate for representing some model, but sometimes a model-author needs to invent or adapt a language for represent his system. In the area of Software Engineering, this happens for example through the development of domain specific programming- or modelling-languages.

Across engineering disciplines, models may serve many purposes, including:

- Understanding: through creating of a model, it becomes clear which parts of a system are important and how they relate to each other.
- Structuring: especially visual models aid in being able to see an overview, to see which parts of a system belong together, and to see which organization of parts leads to an (aesthetically) pleasing representation.
- Communicating: models are useful in conveying meaning. The systematic use of notation and meaning helps to avoid misunderstanding. Models of software designs are commonly understood as a collection of components that have some types of spatial relation to each other. Especially graphical models are well suited for capturing and conveying such spatial relations between components.

[1] The *Lindholmen* dataset of projects that use UML can be found via http://oss.models-db.com/

- Analysing / Predicting: Models can be used as a basis for analysing and predicting properties of the system to be build. Properties of software systems for which model-based analysis/prediction-techniques exist include o.a.: performance (a system-property; see e.g. [MC04]), maintainability (a development property), and costs (a project management property).
- Coordination: models can serve multiple roles in the project management. These include: a model can be a basis for allocation different parts of the construction to different engineers (division and allocation of work). Also, models can be used sharing information on the progress of the design/construction.
- Guiding: The model can be used to guide the construction work o.a. by showing engineers how different parts build on each other.
- Blueprint for Production: If a model is complete and sufficiently detailed, it can be used to automatically construct the implementation. For software, this can be e.g. through code-generation or low-code platforms where models actually constitute the representation that is executed.

In practice, we see that models play multiple roles, often for different categories of roles in a project. Moreover, the role(s) that models play also change over time. In the early stages of a design project, models play a role in ideation (explore and synthesize ideas). In this stage the design often is changes a lot. For this reason, models in this stage are often represented in fairly informal representations. When a project proceeds to the stage where it aims to define what needs to be built, a more precise model is built that serves as a common reference for all engineers in the team. While designs at this stage are more complete and more precise than those of the inception stage, models are often not totally complete and fully precise. Indeed, economic considerations lead to pragmatic decisions in modelling: engineers choose to focus on parts of the systems that are important. This can be parts that constitute an (engineering) risk to the project or is critical for its functioning. Complementarily, model-authors would leave out information that they believe is 'common understanding' amongst the engineers of the team, and include parts that are relatively uncommon or unique. Indeed we find that level of detail is often not uniform across the model of a design. Some parts may be specified in a high level of detail while other parts are specified in low level of detail.

When models are used for the actual production of a system, then indeed the model must be complete and all necessary detailed included.

The early leaders in developing engineering practices in software engineering were in military and space. Inspired by existing hardware engineering practices they proposed engineering methods that leaned heavily on complete specifications of artefacts at all stages of the engineering cycle. Clearly the military and space domain are very critical. When computers became more universally adopted in the late 1980's and early 1990's, software become more common in all types of business areas, also less critical ones. This has led to a counter-movement in engineering style: iterative, incremental, and adaptive. The most prominent school of approaches goes under the label of 'Agile development'. The Agile community has written down some 'values' that are fundamental to their approach. These values include: 'Working software over comprehensive documentation' (v1), and 'People and processes over methods and tools' (v2). These adagio's are often 'misinterpreted' to mean: 'let's not make *any* documentation (v1) and using fewer tools is better than using more tools, so if we can avoid using modelling tools, then that is a good thing. Indeed this thinking is also justified by efficiency thinking. Both these positions are extremes. Indeed it is clear that modelling is neither necessary nor sufficient to produce good software. But, there are documented projects where individuals and teams do benefit from using modelling in their engineering approaches. To better understand the truth about modelling, and to separate opinions from truths, we need to do empirical scientific studies into modelling.

2 Empirical Studies

The Science of software engineering has accepted that there are 'no silver bullets'; i.e. there is no single technique that automagically makes all defects go away (quality) and/or increases the efficiency of the production of software by some significant factor. As a special case, this also holds for the practice of modelling.

In this talk I will present some recent studies that increased our knowledge about the effectiveness of modelling. I will not go much into the 'cost'-side of this, but will summarize the findings from various interview studies into the hurdles and challenges of modelling: the main hurdles are not so much the effect that is spent on creating models. Even for complex systems that effort is often small compared to effort spent on all other engineering tasks. Among the more challenging hurdles are: migrating existing documentation, introducing new processes for maintaining models, including practices such as versioning and responsibilities such as quality-assurance of models, integrating tooling for modelling into the tool-ecosystem of organisations.

In this talk I will present some empirical studies that research the following:

- Do defects in UML models affect the understanding of the design?
- Do UML models affect the communication about a design?
- Does the use of UML modelling relate to the quality of the resulting software?

This talk will present an update of a paper that overviews empirical results of UML modelling

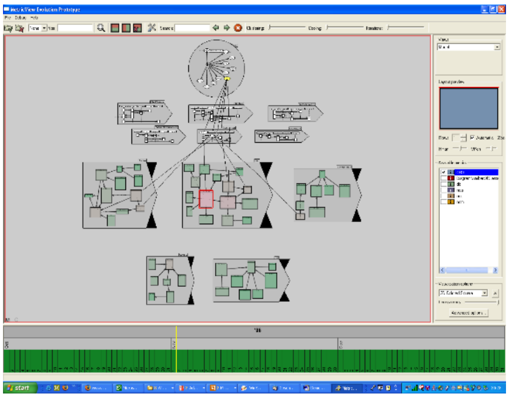

Figure 1 Visualisation of UML diagrams and their quality (from [LC07]).

3 Challenges and Future Directions

Empirical studies have shown that using modelling can contribute to improved communication, improved understanding and improved quality. Also, from case studies in practice, we have learned that each project need to find its own ‘sweet spot’ regarding the rigour and completeness of modelling. We know that such a sweet-spot is related to various context factors such as number of developers, education level of developers, geographic co-location (or distribution) of the team, criticality of the application (often subject to certification standards). Moreover, it is likely that also cultural factors affect the attitude towards the adoption and use of modelling.

From the research we can also identify several directions for advancing the adoption of modelling in software development:

- Better techniques are needed for migrating existing documentation to model-based documentation. Reverse Engineering (RE) can play a role in this. But for RE to be effective, progress needs to be made in several areas, including: i) abstraction from source code/implementation, because models focus on essentials, ii) closely related to the previous: we need techniques for modularisation of reverse engineered diagrams: how to break them up so that we do not get one diagram with 50 classes, but around 3-4 diagrams of 12-16 classes, such that they are cognitively manageable. Together with this we need better algorithms for automatically producing layouts of class diagrams that are compatible with the logic of the design.

For future work, I identify the following promising directions:

- Automated updating of documentation: the most common scenario in modelling is that project teams do make an initial design, but when the project progresses, fail to prioritize updating the design. This scenario where an initial design exists provide good opportunities for automated updating.
- Tools should offer better integration: e.g. editing of models and text should be integrated into a single tool. Also, the usability of tools should be improved. Modern technology enables interaction with touch, gesture, voice which can make the modelling much more intuitive, and moreover may reduce the effort for creating models.
- Other improvements can be obtained, by automatically creating views that provide the right information to developers that suit the task they are working on [LC07]

References

[LC07] Christian F. J. Lange, Michel R. V. Chaudron: Interactive Views to Improve the Comprehension of UML Models - An Experimental Validation. ICPC 2007: 221-230.

[MC04] Johan Muskens, Michel R. V. Chaudron: Prediction of Run-Time Resource Consumption in Multi-task Component-Based Software Systems. CBSE 2004: 162-177

[CHN12] Michel R. V. Chaudron, Werner Heijstek, Ariadi Nugroho: How effective is UML modeling ? - An empirical perspective on costs and benefits. Software and System Modeling 11(4): 571-580 (2012)

Modeling User's Individual Differences during Reading

Matej ČIEF*

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
xcief@stuba.sk

Abstract. In these days, as we are surrounded by huge amount of digital information, there is an increasing need to process overabundant data more effectively. Although visualizations have become a relevant tool for presenting stories from data, they are typically designed following a one size-fits-all approach, not considering the potential needs and preferences of an individual user. Previous research has uncovered which user characteristics play role in how users process textual documents with embedded visualizations, yet, to the best of our knowledge, no work claiming successful inference from such reading tasks has been carried out. Thus, we analyzed impact of individual differences on reading and comprehension of text. We focused on a specific cognitive characteristic (visual working memory) and propose a method of its identification using eye-tracking data. We present our findings as a step toward developing user-adaptive support.

1 Introduction

There are multiple factors influencing the way how a user processes the information contained in the text, such as text complexity, previous domain knowledge of the user, but also user characteristics and cognitive abilities. These traits have impact on user's ability to comprehend the text and amount of effort user must make. Currently most of the user interfaces do not consider the potential needs of an individual user [7]. To help the users (e.g., by highlighting important information or reference between text and part of the visualization), a method for identifying level of these abilities is needed.

We have analysed the impact of these abilities on user's accuracy and speed and based on our findings, we have chosen visual working memory to infer as it showed to have the biggest impact on user's accuracy. We propose a machine learning method for inference of visual working memory using eye-tracking data. Apart from standard eye-tracking metrics, our method also calculates transition probabilities between areas of interest and classifies user based on difference between his or her probabilities and aggregated probabilities of the whole group. We are not aware of other research using such method in domain of combined text and visualizations. We expect transition probabilities to be an important deciding factor for our classifier to successfully infer visual working memory.

2 Related work

There has been extensive research demonstrating how people process textual and graphical information. Conati et al. [7] identifies nine user characteristics (e.g., visual working memory, reading proficiency) that play a significant role in user experience with visualizations or impact user's reading abilities. These findings are backed by multiple studies. According to Dole et al. [2], individuals with high *need for cognition* (extent to which individuals are inclined towards effortful cognitive activities) are more likely to engage with in effortful analysis of ideas, while others will be more likely to process information heuristically. Carenini et al. [1] examined impact of *perceptual speed* (a measure of speed when performing simple perceptual tasks) on time needed to complete a visualization task where significant differences were found with growing complexity of task. Participants with higher perceptual speed performed significantly better when solving more complex tasks. This implies that complexity can considerably influence users' performance depending on their cognitive abilities. Velez et al. presented evidence that spatial abilities directly impact visualization comprehension [8], meaning it is possible to use level of spatial ability as a method for comparison of different errors in comprehension.

There has been some work investigating how to infer cognitive abilities. Steichen et al. [6] classified *verbal working memory* (part of the working memory

* Bachelor study programme in field: Informatics
Supervisor: Dr. Róbert Móro, Institute of Informatics, Information Systems and Software Engineering, Faculty of Informatics and Information Technologies STU in Bratislava

responsible for temporary maintenance and manipulation of verbal information) and *visual working memory* (part of the working memory responsible for temporary storage and manipulation of visual and spatial information) using gaze data. In a study, a sequence of tasks displaying bar or radar graphs was presented. Based on such visualization tasks, they classified level of cognitive ability into two groups – high and low (based on median split). They achieved 63% accuracy for verbal working memory and 58% for visual working memory respectively. The most important metrics were those related to chart labels and areas of interest (AOI). Participants with high level of visual working memory had smaller time to first fixation, meaning they were able to scan through various AOIs more quickly.

Even though there has been conducted research focused on inference of cognitive abilities, it was mostly focused on inference from visualizations and its scope relating reading tasks or combining text with visualizations is very limited. Therefore, our goal is to research combination of text and visualizations where we can find interesting results regarding interaction between these two modalities.

Although we are not aware of other research dealing with adaptive personalized support for processing both text and visualizations, there has been made some research on providing guidance to visualizations. Gotz and Wen [3] achieved significant speed improvement and reduced error rate by recommending appropriate visualization based on pattern detection in sequence of actions performed by users during search and comparison tasks. Carenini et al. [1] gauged individual forms of dynamic highlighting to draw attention to relevant points within bar graphs and reported significant improvement in task performance compared to using no interventions. For reading, Loboda et al. [5] investigated if words relevant for current reader's informational needs could be inferred using eye-tracking data, which proved to be useful to provide personalized content. Adaptive interventions are beyond the scope of this work.

3 Explorative analysis of data

Toker et al. [7] conducted user study with purpose to discover which user characteristics play role in how users process text documents with embedded visualizations. Participants were given a task to read such a document (Magazine Style Narrative Visualizations or MSNV for short) and after finishing they were presented with questionnaire in which they expressed their opinion about article and answered questions testing their comprehension of relevant concepts discussed in article. Each participant read 15 articles.

This experiment was replicated at the Faculty of Informatics and Information Technologies, Slovak University of Technology in Bratislava (the author of this paper assisted in collecting data). The majority of participants consisted of students. Our data consist of 52 participants and 15 articles for each one of them. Data were preprocessed by open-source library EMDAT (*github.com/ATUAV/EMDAT*). Apart from that, the data also contain information about user characteristics and performance. There are 17 independent variables. Nine of them are cognitive abilities (e.g., *visual working memory*, *perceptual speed*) and eight of them are subjectively measured traits collected from user questionnaire (e.g., *bar chart preference*). Regarding dependent variables, the data contain two performance related metrics – *accuracy* and *task time* and two subjectively measured metrics *ease of understanding* and *interest* (both related to the article read by a user). We analysed these data to find user characteristics that effect participants' performance.

First, we examined whether there are any outliers. By defining them using 2 standard deviations, we get 2 – 5 outliers for each independent variable, not enough to significantly skew our data. But we have noticed there are few entries, where *task time* is significantly higher than the rest of entries for an article. Consequently, we defined outliers using 3 standard deviations threshold with task time for each MSNV document. Those are users, who did not pay attention. We detected 7 such cases (out of 726) and these were excluded from further analysis.

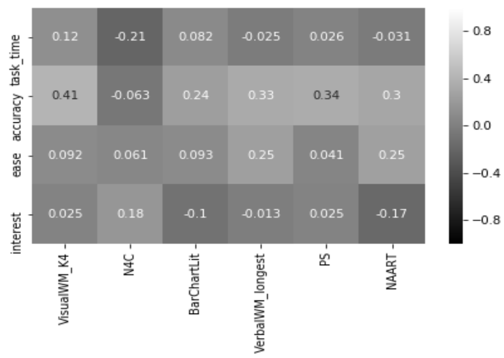

Figure 1. Correlation heatmap of significant user characteristics.

We examined possible correlation between user characteristics and dependent variables – particularly accuracy and task time. Out of 17 examined user characteristics there are 6 cognitive abilities having moderate correlation (above 0.2) with *accuracy* or *task time*. These are (refer to . from left to right) visual working memory, need for cognition, bar chart literacy, verbal working memory, perceptual speed and reading proficiency. Out of these characteristics, only *visual working memory* and *bar chart literacy* were statistically significant (p-value less than 0.05) with both performance-related metrics. *Perceptual speed, verbal working memory* and *reading proficiency* showed statistical significance only with relation to *accuracy*.

We further examined these cognitive abilities. For each, we placed users into one of 2 groups (either high or low) based on the median split. Then we calculated average *accuracy* and *task time* for both groups and compared it. Furthermore, we identified that there is a significant difference in *task time* between groups split by *need for cognition* where participants with higher level finished reading 15% faster in contrary to the original study [7], where people with higher level spent more time looking at the MSNVs.

Other characteristics did not show any significant effect on *task time*, but *verbal working memory, perceptual speed* and *visual working memory* had considerable effect on *accuracy*. Out of these three, *visual working memory* had the biggest impact according to the median split (refer to *Figure 2*).

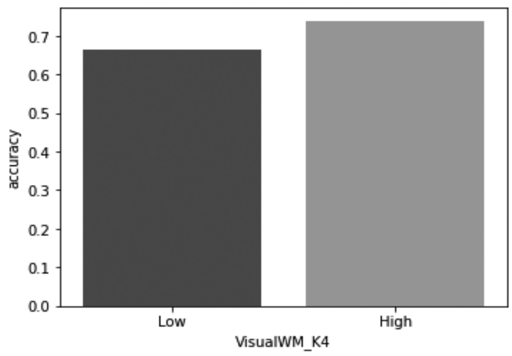

Figure 2. Average accuracy for groups with high and low visual working memory.

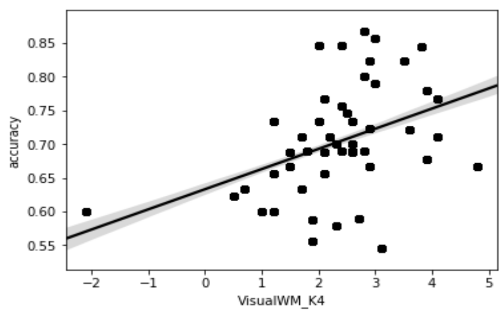

Figure 3. Linear regression to examine impact of visual working memory on accuracy.

We investigated these characteristics further and used linear regression algorithm to confirm our findings. *Figure 3* shows linear trend of visual working memory on accuracy even with one outlier. Without this outlier, the trend would be even stronger. We consider accuracy more important metric, as participants were not time limited and people with worse accuracy are the ones who would profit from adaptive interventions the most. Based on these findings, we chose visual working memory as the user characteristic to infer.

4 Method of automatic visual working memory inference

Our aim is to assess user's level of *visual working memory*. For this purpose, we propose a method of binary classification working with gaze data. Subjects are divided into two groups according to their level of visual working memory. Boundary between these two groups is median. Various metrics are derived from gaze data. These metrics help us determine in which group user belongs.

Because we use a machine-learning method, steps of our method reflect common phases of machine-learning methods. We begin by data preprocessing. This phase includes correction of missing or faulty data, feature extraction and, in few cases, group aggregation.

We are using data mentioned in section 3. The method works on the level of individual user and each article and, in few cases, with aggregated data from all articles for a particular user (e.g., transition probability matrix).

Training of our model begins by calculating transition probabilities between AOIs aggregated for both groups for each article. With respect to our data, four AOIs are chosen to gain general sense of MSNV processing with respect to the two sources of information, text and visualization. These AOIs are defined as:

- *Refs AOI*: The combined areas of all the reference phrases contained in MSNV document.
- *Text AOI*: The rest of MSNV document text.
- *Referenced Bars AOI*: The combined area of all the bars in visualization that are mentioned by any of the references.
- *Viz AOI*: The rest of visualization region.

Input attributes are gaze data about fixations, saccades, pupil size and distance from screen, all in time series. As another input attributes we use output of open-source library EMDAT which extracts multiple useful features from basic gaze attributes.

We take into consideration average, min, max and standard deviations from the following attributes. *Fixation related*: time to first fixation, fixation length, fixation rate, number of fixations, recurrent fixations, fixation-saccade ratio. *Saccade related*: number of saccades, regression saccades, absolute saccadic angles, relative saccadic angles, saccade duration. *Pupil related*: pupil size, pupil speed.

We use Markov model to calculate transition probabilities matrix from gaze data for each group (high and low visual working memory) and store it in our model. Based on these transition probabilities, we calculate their difference with user's transition probabilities. To achieve such result, we aim to use similar approach as Krejtz et al. [4] who used entropy coefficient of the fit Markov model to quantify complexity of individual switching patterns. Our approach is to calculate entropy from a transition probabilities matrix for each group as well as for individuals. The result is a subtraction between entropy of the group and the given individual – another metric used as an input to our model. Lastly, we train our classifier based on all input attributes. We use various algorithms, e.g., *decision tree, random forest, logistic regression.* To increase success rate of trained algorithms without overfitting, we use standard machine-learning approaches such as *cross-validation* and *hyperparameter tuning.* It is possible to use model trained according to this method for inferring new entries.

5 Preliminary results

So far, we have implemented a model inferring user's group of visual working memory using basic gaze attributes without features related to Markov model. We have achieved 66% accuracy with training/testing sample of size 80/20. Majority class represented 50.9% of dataset. As a classifier, we used decision tree and 5-fold cross-validation with training.

6 Discussion and conclusion

We have approached the problem of inference of the level of visual working memory as a binary classification problem, even though the level of visual working memory is a continuous numeric variable. The main reason is accuracy of our inference. We expect better results for a smaller group of categoric values, because size of our training data is limited. Also, our solution can still be used as a solution for adaptive systems thanks to finite number of variations of such system.

The method uses data from MSNV study [7], but our goal is to create universal approach independent from these particular data. Method requires only gaze data and labelled AOIs. Based on these features, our method should achieve similar results with similar problems – where user is looking for connections between text and visualization and where information about transitions between AOIs plays a main role.

During distribution of participants between testing and training sample we make sure that all articles read by one user end up in the same sample. We assume, that participant's performance in one article is similar to his or her performance in other articles. Thus, by isolation into one sample we prevent possible bias of classifier, resulting in more accurate inference.

We have designed method for inferring visual working memory using combination of text and visualization, which can deliver more accurate results than previous studies targeting only stand-alone visualization or text. Furthermore, to achieve such results, our method uses not only standard gaze metrics, but also transition probabilities.

As future work, we plan to implement Markov model, evaluate its significance in decision making of final model and further increase accuracy. We already perform better than inference conducted in [6], where the 58% accuracy was achieved.

Acknowledgement. This work was supported by grants No. 002STU-2-1/2018 supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic.

References

[1] Carenini, G., Conati, C., Hoque, E., Steichen, B., et al.: Highlighting interventions and user differences. In: *Proc. 32nd Annu. ACM Conf. Hum. factors Comput. Syst. - CHI '14*, ACM Press, (2014), pp. 1835–1844.

[2] Dole, J.A., Sinatra, G.M.: Reconceptalizing change in the cognitive construction of knowledge. In: *Educ. Psychol.*, (1998), pp. 109–128.

[3] Gotz, D., Wen, Z.: Behavior-driven visualization recommendation. In: *Proceedingsc 13th Int. Conf. Intell. user interfaces - IUI '09*, ACM Press, (2008), pp. 315.

[4] Krejtz, K., Duchowski, A., Szmidt, T., Krejtz, I., et al.: Gaze Transition Entropy. In: *ACM Trans. Appl. Percept.*, ACM, (2015), pp. 1–20.

[5] Loboda, T.D., Brusilovsky, P., Brunstein, J.: Inferring word relevance from eye-movements of readers. In: *Proc. 15th Int. Conf. Intell. user interfaces - IUI '11*, ACM Press, (2011), pp. 175.

[6] Steichen, B., Carenini, G., Conati, C.: User-adaptive information visualization - Using eye gaze data to infer visualization tasks and user cognitive abilities. In: *Proc. 2013 Int. Conf. Intell. user interfaces - IUI '13*, ACM Press, (2013), pp. 317–328.

[7] Toker, D., Conati, C., Carenini, G.: User-adaptive Support for Processing Magazine Style Narrative Visualizations. In: *Proc. 2018 Conf. Hum. Inf. Interact. - IUI '18*, ACM Press, (2018), pp. 199–204.

[8] Velez, M.C., Silver, D., Tremaine, M.: Understanding Visualization through Spatial Ability Differences. In: *VIS 05. IEEE Vis. 2005.*, IEEE, n.d., pp. 511–518.

Transformation of Full-text Search into Structured Search

Erika REPAŠSKÁ*

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
erika.repasska@hotmail.com

Abstract. In recent years, with a growing amount of data on the Web, search efficiency becomes a challenging problem. While searching using natural language queries and full-text search is more intuitive and natural for users, considering precision, structured search gives us better results. We could solve this issue by transforming full-text search into structured search, so we can get the best properties from both. In this paper, we propose a method of transforming full-text search queries into structured search queries. Our method abstracts from machine learning and enormous query logs and is based on query understanding and fuzzy string matching.

1 Introduction

Nowadays, the search is an inseparable part everywhere, where we are dealing with large volumes of data - which is indeed everywhere. Depending on search engine quality it might take a considerable time to find relevant results when doing an ordinary full-text search. Full-text queries are naturally ambiguous and target thousands of possible candidates, which need to be considered. The advantage is that they are natural for a user to come up with.

On the other hand, structured search using keys and values is more precise and faster, but less intuitive for users. The best solution would be to get the intuitiveness of full-text search and the precision of structured search. The challenge we face is mapping non-structured data from natural language query to structured data stored in the database. The main goal is to identify this kind of mapping without the use of machine learning algorithms. In this paper, we focus on query understanding by using a combination of shingles and Fuzzy string matching. Shingles represent n-grams at the level of tokens.

2 Related work

There are many papers addressing the query understanding problem and proposing different kinds of approaches such as: web query logs [6–8], hierarchical database [3], tries [6, 10], POS tagging [2, 4, 5, 9], knowledge-based [9].

Web query logs are suitable for machine learning algorithms, though they need to contain tens of millions of logs, which becomes impossible for smaller domains. For small query logs, machine learning algorithms are not capable to reliably learn how to identify certain parts of a query. The quality of the approach depends on the amount of data and this plays nicely with big data systems.

Tries approaches are simple to use and their performance does not depend that much on amount of data available. They optimize their performance in terms of time complexity (search/look-up time) by storing n-grams directly in the tries data structure, which is especially useful for fuzzy string matching.

POS tagging of queries deals with some major problems - usually not correct grammar, short query length, lack of capital letters, free word order, etc. However, by using a proper method, we can benefit from POS tagging - e.g., if we identify word *sport* as an adjective in the query *sport watches* and word *watches* as a noun in this query, we can declare that *watches* are a category and *sport* is a type of the watches.

Knowledge-based approaches take a lot of effort in the pre-processing part as they require a definition of relations and connections between data. It makes sense to use it if there are some relations present and the amount of data is large enough for creating these connections. The precision of the approach is high, but the cost is in the complexity of creating such a knowledge base.

---

* Bachelor study programme in field: Informatics
Supervisor: Dr. Michal Barla, Institute of Informatics, Information Systems and Software Engineering, Faculty of Informatics and Information Technologies STU in Bratislava

IIT.SRC 2019, Bratislava, April 17, 2019, pp. 5–8.

3 Transformation of full-text to structured query

Activity diagram in Figure 1 illustrates the algorithm we propose as a method for transforming full-text search into structured search. Overall, the method has two phases. During the first phase, we systematically go through an input query and try to identify, whether a product/manufacturer/category is present. In case one of these items has been found, the second phase tries to find the parameters belonging to this item in the rest of the query. The two nested phases are being repeated until all possible matches are found. Otherwise, the rest of the query, which has not been recognized, is printed as full-text.

In more detail, the flow starts with normalizing a query, i.e. process it through the standard NLP pipeline. After that, we pre-select a subset of items from the database, that are related to the query. Then, we can generate shingles of variable length out of the query. Subsequently, we start to look-up for a good-enough match of a shingle against pre-selected subset. If we were unable to find a match, the process ends here. However, if we were able to find a matching product/category/brand — we consider the result with the highest similarity score and the longest length (considering the number of words) — we try to identify more attributes from the query, which are within the scope of this match. Therefore, we remove tokens used for recognition from the query and re-run the process of shingles generation and match look-up - this time scoped to the product/brand/category retrieved in the previous step.

The whole process boils down to how we are detecting and scoring matches of query terms to product/brand/category parameters. We are evaluating several strategies of similarity scoring to find out which one is the best for the query-attribute mapping.

As we store data and their classification in the database, we can categorize our method as a knowledge based approach. For each item in our database, we can determine, whether it is a product, a category, a brand or a parameter - and what kind of parameter it is (the key of key-value pairs), e.g. color, size, weight, etc. Items - that are stored together with their specification in the database - represent our knowledgebase, which we use in the process of transforming search.

4 Prototype

We performed a basic evaluation of the method by implementing our first prototype in Python. Data we work with are stored in database tables - the *parameters* table and the *product-manufacturer-category* table. The *parameters* table has the following attributes: product ID, category ID, manufacturer ID, key, value.

The whole point of *parameters* tables is to store key-values (product parameters) along with information about parameters scope (which product, category and manufacturer the key-value pair belongs to).

The *product-manufacturer-category* table then contains information about those three primary entities. Its attributes are: item ID, name, URL, parent ID - valid for categories only, item type, search name - the lemmatized form of the name and manufacturer ID and category ID - both valid for the products only.

For the Fuzzy matching and string similarity computation, we use the FuzzyWuzzy [1] Python library, which provides us several different scorers:

1. *Simple ratio* - returns maximal score for two identical strings.
2. *Partial ratio* - returns maximal score if one string is a substring of another, while ordering of individual words does matter.
3. *Token sort ratio* - returns maximal score if both strings contains the same words, while ordering of these words does not matter.
4. *Token set ratio* - returns maximal score if one string contains all words from another, while ordering of words does not matter.

5 Evaluation

For our evaluation, we have a validation dataset consisting of query log of a big Slovak online retailer, along with an information about a clicked result. The information contains a structured version of clicked item - such as a name, a type, attributes, etc. We assume that users know exactly what they are looking for when entering their queries [7]. Therefore, if we consider only the queries which resulted in a click, we can assume that attributes of the clicked item are somehow encoded in the original query.

Dataset for a scorer experiment consists of *213 queries* with an average length of *2.7 words*. We tested how well the prototype works according to used scorer.

Table 1. Scorers evaluation results.

Scorer	*C1*	*C2*	*Precision*	*Recall*
Simple ratio	110	91	**82.73%**	42.72%
Partial ratio	210	31	14.76%	14.55%
Token sort ratio	110	91	**82.73%**	42.72%
Token set ratio	176	34	19.32%	15.96%

Table 1 shows the number of queries, in which the prototype identified something (*column C1*), the number of queries, in which prototype identified something

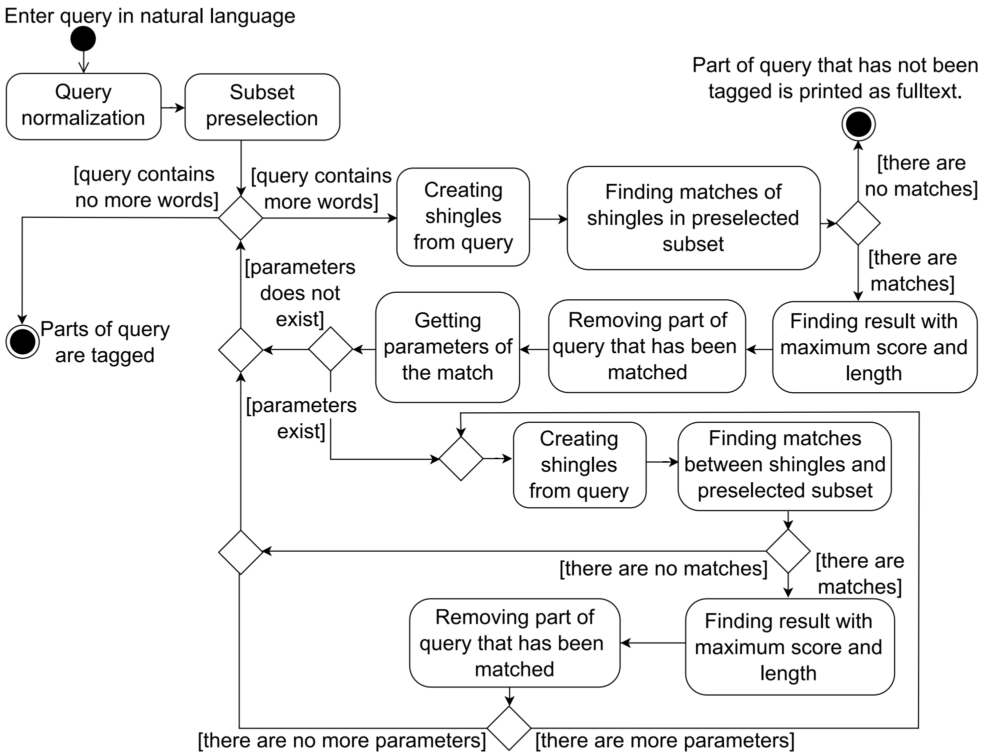

Figure 1. Activity diagram of the proposed prototype.

successfully (*column C2*), the precision and the recall - if we suppose the prototype could identify something in all queries.

Taking look at the first two columns, we can see that scorers *Partial ratio* and *Token set ratio* were able to identify something in most cases, but their precision is very small. It is because of the way they work - there is a high chance the part of a query is a substring of some record, so they find a match for almost anything and that is why they are so unsuccessful. On the contrary, *Simple ratio* and *Token sort ratio* are more strict and find a match in much fewer queries, but when a match is recognized, it is likely to be correct.

The reason why *Simple ratio* and *Token sort ratio* give the same results is the way users formulate their queries. Users tend to keep the word order unchanged to the original names (or their substrings). It means no matter which of these scorers we use, results are the same. However, by using *Token sort ratio*, we can give users more freedom to change the word order.

The average length of a successfully identified query is: *2.9 words* for *Simple ratio* and *2.91 words* for *Token sort ratio*. The longer query is, there is the bigger chance of identifying something successfully. When it comes to a one-word query - there are only two possible options: the prototype either identifies the word or does not. When the long query is being analyzed, there is a big chance that some of its parts can be recognized successfully.

Considering that the average length of a query is *2.7 words*, it becomes clear why the recall of the best scorers is only slightly above 42%. As the score threshold is currently set to 90, matches with the smaller scores are being ignored. That is why - for example, serial numbers, which are parts of product names - are not recognized. Comparing them to the whole product name gives us quite a small score.

Further, we focused on efficiency. Comparing a user input with all entities in the database is an enormously expensive task. The improvement we suggest is to do a pre-selection, which tokenizes an input and for every token, it queries a database for retrieving a list of records, which contains that particular token. As a result, we reduce the comparison list to only those entities, which are somehow related to the user input. We tested the impact of this improvement by running prototype 10 times over 50 queries using following variants: *no pre-selection* and *pre-selection w/wo full-text index*. Results (table 2) show that using both - the pre-selection and the full-text index - can make the prototype runs ca. *23,55 times* faster.

Table 2. Effect of preselection and full-text index.

Preselection	*Full-text index*	*Avg. time*
No	No	606.87 sec.
Yes	No	73.65 sec.
Yes	Yes	**25.77 sec.**

Another efficiency issue we faced was long queries, which generate lots of n-grams. The improvement we suggest in this case is to stop generating n-grams for such queries. When a number of words in the query is greater than or equal to 5, we only check for a match for the whole query and do not generate more n-grams. We again tested the effect of the improvement by running 10 times two variants (with restricting generating n-grams and without) over two datasets consisting of 50 queries. The average length of the query in the first dataset was 2.22 words (table 3) and 3.18 words in the second dataset (table 4.).

Table 3. Restricting n-gram generation (1st dataset).

Restriction	*Precision*	*Recall*	*Avg. time*
No	85.71%	60%	29.84 sec.
Yes	85.29%	58%	**25.77 sec.**

Table 4. Restricting n-gram generation (2nd dataset).

Restriction	*Precision*	*Recall*	*Avg. time*
No	89.19%	66%	85.14 sec.
Yes	89.29%	50%	**31.08 sec.**

When it comes to shorter queries, testing shows there is not such a significant difference in execution time. But when analyzing longer queries, the restriction improvement allows the prototype to run ca. *2.74 times* faster, without any especially negative impact on precision or recall.

The conclusion of our evaluation is that the prototype works the best for rather longer queries and with *Token sort ratio* or *Simple ratio* scorer. Talking about the efficiency improvements, we highly recommend to use pre-selection with full-text index and restrict generating n-grams for long queries.

6 Conclusions

In this paper, we proposed the process for transforming full-text search into a structured one. The biggest advantage of our solution is that it is machine learning free and is thus able to work properly in domains, where it is hard to get enough data for ML-based approaches. We showed how the prototype works and what results it gives us when different scorers are used. Finally, we found out which scorers work the best. We also showed how to optimize execution time by using a subset pre-selection with the full-text index and restricting n-grams generation for long queries.

Acknowledgement. This work was supported by grants No. 002STU-2-1/2018 supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic.

References

[1] Cohen, A.: FuzzyWuzzy: Fuzzy String Matching in Python. `https://chairnerd.seatgeek.com/fuzzywuzzy-fuzzy-string-matching-/in-python/`, Last visited: 7/12/2018.

[2] Ganchev, K., Hall, K.B., McDonald, R.T., Petrov, S.: Using Search-Logs to Improve Query Tagging. In: *The 50th Annual Meeting of the Association for Computational Linguistics, Proc. of the Conf., 2012, Korea - Volume 2: Short Papers*, 2012, pp. 238–242.

[3] Ipeirotis, P.G., Gravano, L.: Distributed Search over the Hidden Web: Hierarchical Database Sampling and Selection. In: *VLDB 2002, Proc. of 28th Int. Conf. on Very Large Data Bases, 2002, China*, 2002, pp. 394–405.

[4] Keyaki, A., Miyazaki, J.: Part-of-speech tagging for web search queries using a large-scale web corpus. In: *Proc. of the Symposium on Applied Computing, SAC 2017, Marrakech, Morocco, 2017*, 2017, pp. 931–937.

[5] Liu, J., Pasupat, P., Wang, Y., Cyphers, S., Glass, J.R.: Query understanding enhanced by hierarchical parsing structures. In: *2013 IEEE Workshop on Automatic Speech Recognition and Understanding, Olomouc, Czech Republic, 2013*, 2013, pp. 72–77.

[6] Paparizos, S., Ntoulas, A., Shafer, J.C., Agrawal, R.: Answering web queries using structured data sources. In: *Proc. of the ACM SIGMOD Int. Conf. on Management of Data, SIGMOD 2009, USA, 2009*, 2009, pp. 1127–1130.

[7] Pound, J., Paparizos, S., Tsaparas, P.: Facet discovery for structured web search: a query-log mining approach. In: *Proc. of the ACM SIGMOD Int. Conf. on Management of Data, SIGMOD 2011, Greece, 2011*, 2011, pp. 169–180.

[8] Sarkas, N., Paparizos, S., Tsaparas, P.: Structured annotations of web queries. In: *Proc. of the ACM SIGMOD Int. Conf. on Management of Data, SIGMOD 2010, USA, 2010*, 2010, pp. 771–782.

[9] Wang, Z., Zhao, K., Wang, H., Meng, X., Wen, J.: Query Understanding through Knowledge-Based Conceptualization. In: *Proc. of the Twenty-Fourth Int. Joint Conf. on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, 2015*, 2015, pp. 3264–3270.

[10] Yang, Z., Yu, J., Kitsuregawa, M.: Fast Algorithms for Top-k Approximate String Matching. In: *Proc. of the Twenty-Fourth AAAI Conf. on Artificial Intelligence, AAAI 2010, USA, 2010*, 2010.

Generating Graphs on a Grid

Ivan PETROV*

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
xpetrov@stuba.sk

Abstract. Generating graphs in a grid is a common task of communication systems analysis. There are many variations of it and their corresponding sequences. Typically, such sequences consist of numbers of all possible graphs of a given length. In this paper, we deal with graphs that are restricted to start at the origin point and that each vertex of the generated graph can be expanded in either of two directions, positive horizontal or positive vertical, including both directions at the same time.

The computation of this sequence is a problem with exponential time complexity. The efficiency of our algorithm is based on symmetry, i.e. there is no need to generate absolutely every graph, instead, they can be grouped according to a certain symmetry feature. At this stage, we managed to enumerate all graphs of the length of 0-24 edges in 117 minutes and the number of different 24-edges graphs is 1,246,172,974,048. For 25-edges graphs, the estimated waiting time is 6 hours.

1 Introduction

To study the properties of communication and other networks, it is necessary to create their model, which can often be represented by a graph. In the case of sensor networks, the graph can be represented as a grid. Such networks usually contain many nodes that have the same data transmission radius. Each node can transmit information to all nodes, the distance to which is not greater than the radius of its transmission. For example, to determine the probability of delivery of a message at a given probability of failure of each line means to find all valid graphs in the grid. Generally, this is a challenging problem. Therefore, the question of developing an effective generator of such graphs is relevant.

First, we explain which graphs in the grid are valid. In the case of sensor networks, each topology point can only be connected to those points that are within the data transmission radius. As part of our problem, we assume that the cell size in the grid is 1×1, and the transmission radius is 1. Thus, points can be interconnected only along grid lines. To avoid cyclicity in graphs when generating, we will add a condition, that each vertex of the generated graph can be expanded in either of two directions, positive horizontal or positive vertical, including both directions at the same time and none of them as well. So, there are four possible cases for each vertex how it could be expanded. If one of these four transmission methods is not yet selected for a vertex, this vertex will be called active. In Figure 1 these vertices have a white core.

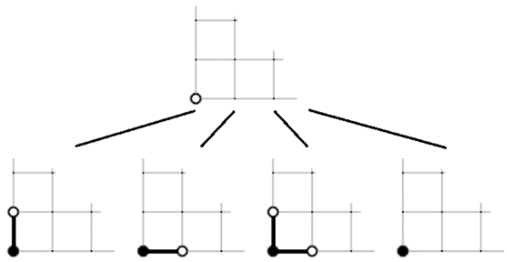

Figure 1. The deployment of the active vertex.

Second, let us explain what we mean by the length of the graph. Since there is no such term as "the length of the graph" in graph theory, but there is "the length of the walk", it is important not to confuse them. According to graph theory [2], a walk in the graph $G = (V, E)$ is a finite sequence of the form

$$v_{i_0}, e_{j_1}, v_{i_1}, e_{j_2}, \dots, e_{j_k}, v_{i_k},$$

which consists of alternating vertices and edges of G. The walk starts at a vertex. Vertices $v_{i_{t-1}}$ and v_{i_t} are end vertices of e_{j_t} $(t = 1, \dots, k)$. v_{i_0} is the initial vertex and v_{i_k} is the terminal vertex. k is the length of the walk. A zero length walk is just a single vertex v_{i_0}. Speaking about the length of the graph, we simply mean the number of edges used (or the size of the set E).

* Bachelor degree study programme in field: Informatics
Supervisor: Dr. Dušan Bernát, Institute of Computer Engineering and Applied Informatics, Faculty of Informatics and Information Technologies STU in Bratislava

IIT.SRC 2019, Bratislava, April 17, 2019, pp. 9–12.

2 Depth-first search

The obvious solution to generating all graphs is a search tree, where the simplest graph, namely a lone vertex, will be a root node. Its successors will be four nodes, corresponding to the four possible ways to expand the active node.

Note also that the number of successors of a node is determined by the formula:

$$4^{a} - 1 \tag{1}$$

where a is a number of active vertices. Subtraction of one at the end means that we do not consider a successor identical to the parent node. The presence of such a successor for each node is due to the fact that there is always such an option, in which each of the active vertices will not be expanded.

Thus, the deeper we traverse along the search tree, the wider it becomes, and significantly. For example, there is only one node at the depth of 0. Further, we have 3 nodes and 21 at the next two depth levels. Traversing the search tree implies some data structure where we can store the generated but not yet processed successors. Therefore, to use memory rationally, it is reasonable to use the method of traversing the tree called Depth-First Search. By storing the vertices in a LIFO stack, we explore the vertices by lurching along a path, visiting a new neighbour if one is available, and backing up only when we are surrounded by previously discovered vertices [3]. So, when we move to a new branch, we can free those memory blocks where the previous one was stored. If we were used Breadth-First Search, we would have to store the entire tree in memory, which is unacceptable because of the exponential increase in the number of nodes with depth increasing.

3 Graph representation

Let the graph increment be the set of edges and vertices of the graph, removing which we get the predecessor (in the search tree) of this graph. For example, in Figure 1, a graph with two active vertices has an increment of the set of these two vertices and their adjacent edges. In other words, a graph increment is the difference between that graph and its parent (predecessor). Thus, a graph G can be defined as a set of all the increments obtained by traversing from the node corresponding to this graph G up to the root node. So, it is enough to store only the corresponding increment in nodes of the tree instead of the whole graph.

Since the increment consists of a set of vertices and edges, it will be stored in two variables: $edges$ and $vertices$. Each of these variables is a sequence of bits, where 0 corresponds to the unused edge/vertex of the grid, 1 corresponds to the used edge/vertex of the grid. For example, in Figure 2 below, when the grid is rotated, the i_3 increment corresponds to $edges = 011100b$ and $vertices = 0110b$. However, based on the i_1 increment, it becomes apparent that all further increments have the last two bits of edges and the last one of vertices being zero. Therefore, we assume that for the i_3 increment $edges = 0111b$ and $vertices = 011b$. Note also that the i_0 increment always looks like this: $edges = \emptyset$ and $vertices = 1b$.

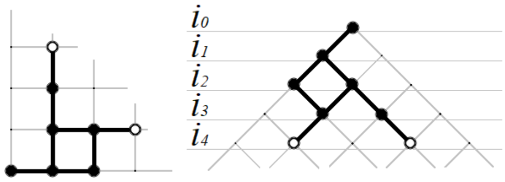

Figure 2. Graph increments.

3.1 Pros and cons

The benefits of this graph representation are, first, the economical use of memory, since each node of the tree does not contain the whole graph, but only the increment. Second, the use of a sequence of bits (as the *unsigned long long* data type) gives us an advantage in time over other ways of representation since, generating a new node, we do not have to regularly allocate memory to edges and vertices variables. The disadvantages are limits of data type (at least in C language). That is, if some increment contains more edges than the size of the data type in bits, it is likely that the program will fail.

4 Symmetry

The naive solution algorithm is to generate absolutely every graph. However, among the generated graphs there are many groups of graphs, there is no practical difference between them. In this case, it will be enough to generate only one graph from the group and count it exactly as many times as there are graphs in this group.

Let us imagine that at a certain stage for some node among the potential successors there are two increments with sequences of edges 101001b and 100101b and respectively vertices 1101b and 1011b (see Figure 3). That is, one sequence is the reverse of the other. We can guess that the successors of such increments will be the same (to be more precise, it is mutually reversible). So, it makes no sense to generate the same thing twice. Therefore, we will consider such increments as *equal*.

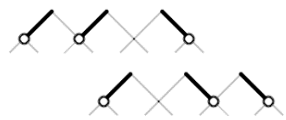

Figure 3. Two equal increments.

Consider an example:

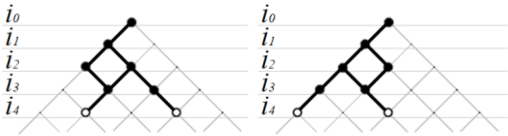

Figure 4. Two graphs with equal increments.

The difference between graphs in Figure 4 is only that for the i_3 increment in one case edges sequence is equal to 0111b, and in the other, this sequence is equal to 1110b. Otherwise, they have the same number of used edges and vertices and unused edges. That is, having only the left graph, it is safe to assume that there is also a graph shown on the right. Let's pay attention to the i_1 increment: edges sequence, in this case, is 10b. So, we can also assume the existence of two more graphs of this configuration (if the edges sequence at i_1 were 01b, see Figure 5). Thus, this group consists of 4 graphs.

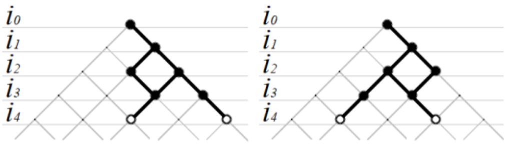

Figure 5. Two graphs with the same increments as in Figure 4.

So, having only one graph, we can conclude that there are three more with the same parameters. Any other groups can contain from 1 to 2^n (n – a given maximum length of generated graphs). That is, the best case would be when for some group, instead of executing a cycle 2^n times, the program will execute only one.

Let us say a few words about how it looks in the code. Each time we pop an increment from the stack, we first generate a set of all possible successors. Among this set we need to identify pairs of *equal* (in other words, mutually reversible) increments. For each such pair, we push only one increment onto the stack and remove the other. At the same time, the pushed increment stores information that it represents not one graph, but two. This information will also be shared with all successors of this increment. Finally, we push all other successors of the set (for which there was no pair) onto the stack.

5 Testing and results

The test results are given for graphs with a maximum length of 24. The desired sequence X_n is presented in Table 1 in the second column. The third column contains data on the number of graphs generated by the program (X'_n). Based on the difference between this and the second column (Figure 6), we can draw conclusions about the efficiency of our algorithm in comparison with the naive one. The ratio of the number of graphs generated by the program to their actual number is presented in column 4.

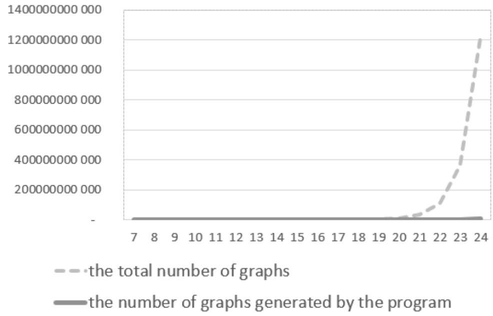

Figure 6. Comparison of the number of graphs generated by the program to their actual number.

Figure 6 clearly shows how much unnecessary work we have avoided. Its size is impressive. Even for $n = 24$, it seems that the bottom curve still looks like a straight line. But it is not: the bend is there, but it is so small that we do not notice it.

Table 1. Number of graphs and other dependencies.

n	X_n	X'_n	ratio	time[s]
0	1	1	1.0	0.000
1	2	1	2.0	0.000
2	5	2	2.5	0.000
3	14	4	3.5	0.000
4	42	10	4.2	0.000
5	130	23	5.7	0.000
6	412	57	7.2	0.000
7	1,326	142	9.3	0.001
8	4,318	364	11.9	0.002
9	14,188	944	15.0	0.003
10	46,950	2,492	18.8	0.005
11	156,258	6,656	23.5	0.011
12	522,523	18,008	29.0	0.024
13	1,754,254	49,263	35.6	0.063
14	5,909,419	136,208	43.4	0.153
15	19,964,450	380,166	52.5	0.409
16	67,618,388	1,070,692	63.2	1.157
17	229,526,054	3,040,728	75.5	3.343
18	780,633,253	8,703,323	89.7	9.717
19	2,659,600,616	25,093,487	106.0	28.551
20	9,075,301,990	72,851,080	124.6	84.525
21	31,010,850,632	212,881,770	145.7	252.712
22	106,100,239,080	625,918,631	169.5	757.056
23	363,428,599,306	1,851,124,928	196.4	2318.450
24	1,246,172,974,048	5,505,132,809	226.4	6927.653

Comparing the time spent on calculations with the results achieved by other students of our faculty last year [1], we can see a significant improvement. Last time, to calculate the value of X_{23} and X_{24}, the calculation time was 25409 and 76911 seconds. In other words, our program calculates these values 11 times faster. Our algorithm is different in that the symmetry is considered not only for the root node of the search tree but for each node. A visual comparison of the calculation time is shown in Figure 7.

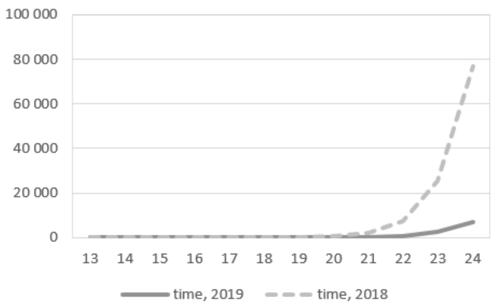

Figure 7. Comparison of calculation time (s).

6 Conclusions

Although we can be satisfied with the work done, this algorithm can still be improved. At least the program can be implemented using multithreading. Splitting the current program into n threads should increase efficiency by n times. It would be best if we could see some pattern based on which it would be possible to detect the hidden formula.

Acknowledgement. This work was supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic within the Research and Development Operational Programme for the project “University Science Park of STU Bratislava”, ITMS 26240220084, co-funded by the European Regional Development Fund. The work was also partially supported by the Slovak Research and Development Agency (APVV-15-0789).

References

[1] HROŠ, Roman. Generating subgraphs of finite grids. *IIT.SRC 2018*. Bratislava, 2018, pp. 1–4.

[2] RUOHONEN, Kejho. Graph Theory [online]. Tampere, 2013 [cit. 2018-12-10]. Available at: http://math.tut.fi/~ruohonen/GT_English.pdf. Lecture nodes. Tampere University of Technologies.

[3] SKIENA, Steven S. *The Algorithm Design Manual*. 2. London: Springer-Verlag, 2008. ISBN 978-1-84800-069-8.

Evolution of Image Filters for Clustering Based Image Segmentation

Tomáš ANDA*

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
tomas.anda384@gmail.com

Abstract. Segmentation of brains structures is important in quantitative analysis, due to the detection of diseases such as Alzheimer's disease causing brain atrophy. In this paper, we propose a new method of brain structures segmentation, i.e. white matter (WM), grey matter (GM) and cerebrospinal fluid (CSF) based on clustering. In our method, the input data are transformed by a filter, which is created using cartesian genetic programming (CGP). We evaluate our method on a publicly available annotated dataset obtained from the evaluation framework MRBrainS.

1 Introduction

One of the subsets of evolutionary algorithms is genetic programming (GP), which is used to evolve executable programs and is often used in data analytics. The individuals in GP represents computer programs or functions which are represented as tree structures. Cartesian Genetic Programming (CGP) is a derivation of GP, which does not represent the program using a tree structure, but rather as a directed acyclic graph [1].

One of the major problems of GP is bloat. This phenomenon caused that after several generations, the size of programs become too large without changing fitness [2]. This type of problem solves CGP where the representation of a program or function is using a directed acyclic graph, and the individuals in the population have fixed size of genotype. This makes CGP calculation much more efficient.

This paper focuses on application of CGP to quantitative brain analysis, which is area where CGP have not been used before.

2 Alzheimer's detection based on segmentation of MRI images

The brain is one of the most complex part of the human body and together with the spinal cord forms the central nervous system. The brain has many important functions and is responsible for maintaining balance, thinking, planning, movement control, and other functions. Since the brain forms the center of the nervous system and has many important functions, improper brain function can lead to failure of these functions, which can have catastrophic consequences. Pathological changes in the brain can lead to disease such as Alzheimer's disease.

Alzheimer's disease is irreversible and is currently incurable, only certain drugs are known to suppress symptoms. For this reason, it is important to recognize the disease in the early stages, and then deploy appropriate drugs to suppress symptoms and slow down the disease progression. That is why the emphasis is put on the development of tools that can detect Alzheimer's disease in the early stages from MRI images. Studies show that, based on the ratio of WM and GM in the brain, it is possible to detect certain anomalies and evidence of Alzheimer's disease [3, 4].

3 Dataset

In our work, we are using dataset obtained from the evaluation framework MRBrainS [5]. This framework provides a publicly available tagged dataset with 5 patients to train algorithms for segmentation of brain structures. Each of the patients had 48 brain images. These data represent volumetric brain imaging data collected by MRI from patients between 65 and 80 years of age. Some of the patients suffered from brain atrophy and white matter lesions.

Datasets contain raw brain MRI scans and manually segmented brain scans, where image segmentation was manually performed by experts.

Scanning was performed with a 3T MRI scanner, including the following MRI modalities: T1-weighted, T1-weighted inversion recovery and T2-weighted fluid-attenuated inversion recovery (FLAIR).

* Master degree study programme in field: Intelligent Software Systems
Supervisor: Jaroslav Loebl, Institute of Informatics, Information Systems and Software Engineering, Faculty of Informatics and Information Technologies STU in Bratislava

IIT.SRC 2019, Bratislava, April 17, 2019, pp. 13–16.

4 Related work

In the field of segmentation of brain structures, deep learning methods prevail, which can overcome classic machine learning algorithms and bring better results [6]. Deep learning describes neural networks (NN) with a larger amount of layers. These neural networks can perform hierarchical feature extraction from the input image. Such networks can deliver very effective results when training on a large amount of data.

There have been developed a several approaches based on deep learning for various purposes, such as object detection, image segmentation, voice recognition, or disease classification. In the image segmentation and classification, the most commonly applied type of NN is the convolutional neural network (CNN).

We point out several works below, that focus on brain structure segmentation using CNN. Authors in these works use for quantitative measuring of brain segmentation quality one of the most popular statistical methods: Dice Similarity Coefficient (DSC) [6]. DSC is expressed in percentage and serves to determine the spatial overlap of the binary segmentation, where A represents the result of the binary segmentation and G represents the reference standard [5]:

$$DSC = \frac{2|A \cap G|}{|A|+|G|} \cdot 100 \quad (1)$$

Zhang et al. [7] describes the CNN-based 2D approach using patches of size 13x13 pixels. In their work, they segmented WM, GM and CSF and pointed out that their CNN-based approach was able to overcome classical machine learning algorithms such as support vector machines (SVM) and random forest (RF). With SVM they reached DSC with 76.95% ± 3.55%, with RF they achieved DSC with 83.15% ± 2.52% and with CNN they achieved the best DSC with 85.03% ± 2.27%.

Nie et al. [8] in work describes a fully convoluted network (FCN). They performed a brain segmentation of WM, GM and CSF where they worked with the same dataset as Zhang et al. [7]. Compared to work [7] they reached better results with overall DSC 88.7% for WM, 87.3% for GM and 85.5% for CSF.There are also other works in this area [9, 10] that has produced similar results in brain segmentation with CNN. We can see that deep learning approaches are very effective in the segmentation of brain structures.

In the field of image processing, it is important to mention one of the image segmentation methods: clustering. In the cases of brain structure segmentations, such as WM, GM or CSF clustering algorithms were the most used methods [11].

In the field of image processing CGP achieved good results. Harding [12] in work among the first described the evolution of image filters using CGP, which was able to reduce noise in the image more effectively than conventional noise reduction filters.

The use of CGP has also been found in the field of medicine. Harding et al. [13] in the work described a method using CGP on the detection of cancer cells in the diagnosis of breast cancer.

We see that CGP achieved good results in the field of image processing. Based on our previous knowledge, there is no work that presented brain segmentation using CGP. For this reason, we think it is advisable to explore this area and compare the deep learning approach with CGP.

5 Proposed methods

We propose two methods for segmentation of brain structures using the CGP and clustering. In our methods, CGP modifies the input image by using patches of a certain size. This creates a specific image filter. Then we use clustering algorithm on filtered image to perform image segmentation for specified brain structures (WM, GM, CSF).

In the first method, a raw training MRI image is taken at the input, from which NxN patches are created around each pixel. These patches are transformed by a CGP individual, which creates one pixel from each patch. The input for CGP represents patch size with NxN pixels and at the output of CGP is one transformed pixel. These pixels represent the resulting transformed input image after CGP application. This transformed image is then sent to the cost function where we perform patch extraction around each pixel with size NxN. After extraction of patches, there is performed a binary patch clustering with K-means. We used DSC to evaluate the clustering result where this value represents the fitness value for CGP.

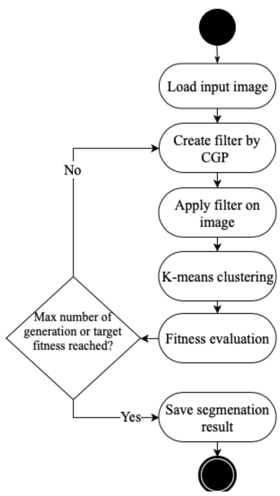

Figure 1. Activity diagram for the first method.

Figure 2. Segmentation of the white matter using the second method with 3x3 patch and extended function set. The first image from left represents the input MRI image, the second expected output, and the third represents the final segmentation of the white matter using CGP where we achieved DSC with 87.62%.

Table 1. Comparison of segmentation performance of our proposed methods with the state of the art methods and with Naive K-means.

			Dice coefficient		
Method	Function set	Patch size	WM	GM	CSF
1	basic	3x3	64.85% ± 16.95%	54.73% ± 0.48%	18.20% ± 6.60%
		5x5	73.1% ± 4.57%	47.67% ± 11.67%	12.00% ± 6.38%
		7x7	69.61% ± 5.03%	51.74% ± 0.9%	19.40% ± 6.06%
		9x9	70.8% ± 1.8%	50.54% ± 1.25%	22.72% ± 5.59%
	extended	3x3	72.0% ± 3.45%	54.74% ± 0.58%	19.15% ± 6.9%
		5x5	71.47% ± 5.09%	53.8% ± 0.56%	17.05% ± 6.05%
		7x7	67.89% ± 7.36%	51.74% ± 1.05%	21.09% ± 4.08%
		9x9	65.42% ± 6.7%	50.37% ± 1.0%	24.00% ± 6.28%
2	basic	3x3	51.75% ± 27.49%	36.34% ± 18.64%	1.58% ± 2.36%
		5x5	68.4% ± 16.26%	15.08% ± 15.78%	1.1% ± 1.35%
		7x7	63.45% ± 17.97%	20.93% ± 15.81%	1.67% ± 1.57%
		9x9	70.53% ± 12.2%	16.3% ± 8.5%	2.04% ± 2.19%
	extended	3x3	**75.73% ± 1.18%**	**55.27% ± 0.26%**	**24.9% ± 9.8%**
		5x5	75.25% ± 2.05%	54.73% ± 0.95%	23.22% ± 7.21%
		7x7	75.69% ± 0.98%	54.18% ± 1.77%	20.88% ± 3.65%
		9x9	72.58% ± 2.87%	52.31% ± 4.90%	20.54% ± 10.71%
State of the art (neural network) [5]			89.87%	86.58%	85.16%
Naive K-means			51.59%	52.48%	8.23%
Our fine-tuned models			71.1% ± 9.58%	52.15% ± 0.29%	24.9% ± 10.53%

The second method is similar to the first method where the input is a raw training MRI image from which are extracted NxN patches around each pixel. Each pixel from patch represent the inputs to CGP (the input amount is NxN) and the CGP output represents the same amount of pixels (the output amount is NxN), which represents the transformed patches. These patches are sent to the cost function and there is performed a binary clustering using K-means. As in the first method, we use DSC to evaluate the clustering result where this value represents the fitness value for CGP.

This method is a modification of the first one, where we get rid of the step of creating patches around individual pixels from a transformed image. By removing an additional extraction of patches from a transformed image in this second method, we expect faster run of the algorithm as in the first method.

6 Experiments and results

We created a prototype, to point out the possibility of using CGP in brain structure segmentation. For the prototype, we decided to use the publicly available CGP library called TenGP[1]. It uses the evolution strategy $1 + \lambda$ and it can define parameters such as the type of mutation, number of inputs and outputs, number of rows and columns, fitness threshold or mathematical functions.

We performed several experiments using T1 weighted MRI scans of the brain in a transverse plane. For training set we have randomly selected 8 MRI images from all images. Validation set contained 2 randomly selected MRI images and the test set contained also 2 randomly selected MRI images. Because of the stochasticity of the CGP algorithm, we execute every experiment repeatedly ten times and compute the average from runs. Our fine-tuned models represents the best performing models on validation set. After we find these models, we performed testing on test set.

In the first method, the number of rows and columns in CGP parameters were set to 25. The number of inputs were set to size of the patch and the number of outputs were set to 1. The population size

[1] https://github.com/Jarino/tengp

were set to 5 individuals and the number of generations for one image were set to 300 generations.

In the second method, the CGP parameters were set identical except to the number of outputs, where this value was set to the patch size (NxN). We have created two types of function sets, a basic function set and an extended function set for experimentation. The basic function set contains functions +, -, *, /, and the extended function set contains functions +, -, *, /, *min*, *max*, **0.5*, **1.5*, **2* and **3*.

Comparison of our proposed methods with different patch sizes and function sets for segmentation of WM, GM, and CSF with other methods are described in Table 1. Values of the DSC in the table represent the average of ten runs of CGP algorithm on validation set. The values with bold font represent the best DSC we achieved from brain structures segmentation. The state of the art results represents the results of best algorithms on same dataset from MRBrainS. We also performed experiments for brain image segmentation using only naive K-means without CGP. In naive K-means we use smart cluster centers initialisation, which has always led to the same clustering result.

We see that we achieved the best DSC with second method and use of extended function set and patch size set to 3x3. With extended function set has CGP more options to create more different filters. It follows that better results could be achieved by adding new functions to the function set. In our opinion, compared to the results, we achieve interesting results and bring interesting research in the area of brain segmentation using CGP.

7 Conclusion and future work

We proposed a new method of brain structures segmentation based on clustering and cartesian genetic programming. Our method uses a filter to transform the input image, which is created using CGP. Then we use clustering algorithm on transformed image to perform image segmentation for specified brain structures, i.e. WM, GM and CSF. We point out the possibility of using CGP in brain structure segmentation, where this alternative has not been researched so far. With our experiments and prototype, we have shown that the use of CGP in brain segmentation can bring potentially interesting results. In further experiments, we focus on extending the function set for CGP and experimenting with CGP parameters such as the number of generations, population size and with the number of rows and columns. We assume that adding new functions to function set will improve the results of binary image segmentation.

Acknowledgement. This work was supported by grant No. 002STU-2-1/2018 supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic.

References

[1] Miller, J.F., Harding, S.L.: Cartesian genetic programming. In: Proc. of the 10th conf. companion on Genetic and evolutionary computation, ACM, 2008, pp. 2701–2726.

[2] Poli, R., Langdon, W.B., McPhee, N.F., Koza, J.R.: A field guide to genetic programming. Lulu. com, 2008.

[3] Biju, K., Alfa, S., Lal, K., Antony, A., Akhil, M.K.: Alzheimer's Detection Based on Segmentation of MRI Image. Procedia computer science, 2017, vol. 115, pp. 474–481.

[4] Alzheimer's, A.: 2015 Alzheimer's disease facts and figures. Alzheimer's & dementia: the journal of the Alzheimer's Association, 2015, vol. 11, no. 3, p. 332.

[5] Mendrik, A.M., Vincken, K.L., Kuijf, H.J., Breeuwer, M., Bouvy, W.H., De Bresser, J., Alansary, A., De Bruijne, M., Carass, A., El-Baz, A., et al.: MRBrainS challenge: online evaluation framework for brain image segmentation in 3T MRI scans. Computational intelligence and neuroscience, 2015, vol. 2015, p. 1.

[6] Akkus, Z., Galimzianova, A., Hoogi, A., Rubin, D.L., Erickson, B.J.: Deep learning for brain MRI segmentation: state of the art and future directions. J. of digital imaging, 2017, vol. 30, no. 4, pp. 449-459.

[7] Zhang, W., Li, R., Deng, H., Wang, L., Lin, W., Ji, S., Shen, D.: Deep convolu. neural networks for multi-modality isointense infant brain image segmentation. NeuroImage, 2015, vol. 108, pp. 214–224.

[8] Nie, D., Wang, L., Gao, Y., Sken, D.: Fully convolutional networks for multi-modality isointense infant brain image segmentation. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), IEEE, 2016, pp. 1342-1345.

[9] Moeskops, P., Viergever, M.A., Mendrik, A.M., de Vries, L.S., Benders, M.J., Išgum, I.: Automatic segmentation of MR brain images with a convolutional neural network. IEEE transactions on medical imaging, 2016, vol. 35, no. 5, pp. 1252-1261.

[10] Bao, S., Chung, A.C.: Multi-scale structured CNN with label consistency for brain MR image segmentation. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 2018, vol. 6, no. 1, pp. 113–117.

[11] Mirzaei, G., Adeli, H.: Segmentation and clustering in brain MRI imaging. Reviews in the Neurosciences, 2018, vol. 30, no. 1, pp. 31–44.

[12] Harding, S.: Evolution of image filters on graphics processor units using cartesian genetic programming. In: 2008 IEEE Congress on Evolutionary Computation IEEE, 2008, pp. 1921–1928.

[13] Harding, S., Leitner, J., Schmidhuber, J.: Cartesian genetic programming for image processing. In: Genetic prog. theory and practice. Springer, 2013.

Fact or Fiction?

Igor STUPAVSKÝ*

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
igor.stupavsky@fiit.stuba.sk

Abstract. The topic is very recent in recent times, as the need for message filtering is becoming increasingly widespread. Not every piece of information comes from a relevant source, and even media-makers get fake news information. As part of the political struggle, there is nowadays more and more frequent misrepresentation of misleading information about individual candidates as part of a political struggle, in order to destroy their political reputation, thus leaning public opinion in their favour. In our research, we have focused on binary classification. Splitting real and misleading messages into election campaigns for US President in 2016. How do you know the fake message?

1 Introduction

Hoax, trolling, or distorted reports are believed today by the younger generation, as well as those born earlier. However, the dangers are most at risk for the younger generation, as they do not yet have enough life experience and can not match the information to their own experience. Therefore, it is relatively easy to influence these people, to give them their vision of reality.

This property of the younger generation abuses various organizations, often of an extremist nature. In the family life and pedagogical process, there is a need to focus on this new reality that has come with the massive expansion of information and communication technologies into our everyday life.

Let us therefore bear in mind the words of the founder of behaviourism:

"Give me a dozen healthy infants, well-formed, and my own specified world to bring them up in and I'll guarantee to take any one at random and train him to become any type of specialist I might select—doctor, lawyer, artist, merchant-chief and, yes, even beggar-man and thief, regardless of his talents, penchants, tendencies, abilities, vocations, and race of his ancestors." [1] and we want to make sure that online users have all the relevant information on which they can make a separate decision. We limit the reach of groups that want to influence the thinking of us all and gain an advantage.

2 Related work

The report is generally defined in the glossaries as information about the recently changed situation or a recent event shared by means of communication with others. As we can see, the definition does not say anything about whether this message is real or invented.

Virtually every person thinks he can figure out what kind of message these two are. However, as we saw in our first attempt, where we followed this questionnaire method, this expectation is not always real.

In the area we are doing in our research, it is necessary to distinguish fake news from real news and, after examining available research work in this area, we can distinguish three basic approaches that we can further divide. These approaches are:

1. *Natural language processing (NLP)* - where fake news is revealed in the form of word analysis.
2. *Facts* - Comparing messages with proven resources as the most common way.
3. *User Model* - The approach focuses on observing the metadata of the post.

2.1 Natural language processing

In the field of NLP detection fake news can be seen:

1. detection of emotionally coloured words
2. analysis of sentence structure

When detecting emotionally coloured words, it is based on the assumption that fake news as a post will attack the recipient's feelings, making this person more prone to take the false message as real. Such work was the work of Mykhailo Granik and Volodymyr Mesyura 2017 [1] on Facebook status analyses, where they reached a 75.40% success rate in revealing the division between administrations. It should be noted that the better classification success

* Doctoral degree study programme in field: Intelligent Information Systems
Supervisor: Assoc. Professor Daniela Chudá, Institute of Informatics, Information Systems and Software Engineering, Faculty of Informatics and Information Technologies STU in Bratislava

IIT.SRC 2019, Bratislava, April 17, 2019, pp. 17–20.

achieved on real news. It is also clear from this that the assumption of better detection is the area of this approach.

By detecting the structure of words and its linguistic styles, which can be detected as being different, the work of Biyani et al. from 2016 [1]. Where did the authors discuss individual techniques, how to crack the report to maximize the number of ladies they can get for their contribution. This structure on fraudulent documents within the Wikipedia website has been dealt with in Kumar et al. also in 2016 [1]. Recent work has addressed Wang 2017 [1] on the PolitiFact domain.

2.2 Fact checking

The theme of the presidential campsite in the US is in the form of facts such as H. Allcott and M. Gentzkow, 2017 [1], on "Social Media and Fake News in the 2016 Election", where they discuss the facts about these voices.

2.3 User modelling

The user model is used to detect fake news by comparing the metadata of individual posts. These metadata are, for example, the names of the individuals who posted or disseminated the text. Other metadata include posting date and other additional data. Such work is, for example, the work of Zhou et al. 2015 [1], which deals with the issuance of real-time certificates to authors. In this way, the recipient of the message can be sure that this is not fake news.

More work that dealt with the issue: J. Nelson [1] conducted an analysis of the US presidential election in 2016. He found that Internet users who believe in false or distorted reports are only very likely to verify the information. At the same time, he noticed a sharp increase in false reports between January 2016 and January 2017. The results of measuring the location of individual reports have shown the belief that these distorted reports are also targeted locally. Within the media that is the source of information in the US, it was The New York Times - 40%, ABC News - 50%. However, within the other countries' media it was less BBC - 1.5% and Aljazeera 0.7%. In our opinion, however, this article has also pointed to a number of false messages that come to the genuinely creating media.

This is a dangerous condition that is easy to use to discriminate against the individual or the entire group of companies in the real world.

As an example of US manipulation, it can also be seen that manipulation is not just a matter for an individual with access to the internet and the necessary knowledge. But it is still more a part of a strategy of political struggle, with which funding is already worth budgeting for a given political campaign. Our view is that the production of counterfeit messages in this case is not motivated by the real intention of physically harming someone, but in order to gain predominantly preferential results. Therefore, the available means are usually used for this fight. If the goal of this work is met and the amount of such reports in the electronic world is reduced, the campaign will also be moved to places more likely to interfere with the views of the marginal group of potential voters.

According to a report published in The New York Times as an example of well-known makers of fake political reports, their creation and dissemination preferentially only result in a headline change. The text itself is not so important, and this text is often copied from previous texts. In this way of spreading news, informatics can find such articles by searching. According to the time of addition, we know how to spread the message and its source. According to Professor N. Kshetri of the University of North Carolina [9], it is very difficult for the disseminators of such reports to be punished because of their frequent out-of-state activities and the weak legal possibilities of punishing them. Therefore, we believe that the solution of the disseminators is in the hands of the competent authorities and we are focusing on the detection of these messages and their labelling, or the prediction of suspicious behaviour.

Another way of gaining voters' voices is based on how to manipulate discussion groups on individual topics.

Manipulating debates is to create false online identities (often in a larger number) to influence the direction of the discussion and the views of other users by wishing to.

One of the known cases was September 11, 2014, where a large number of accounts on networks like Twitter or YouTube reported a blast at a chemical plant in Lousian. This report was artificially created for the direct manipulation of people's opinions. Beginning was a hoax about an explode spread through SMS by picking individuals and gradually using social networks. Although the report was shortly demoted to company executives, it did not get to the same number of people, creating a hostile mood against this producer.

The same approach also uses groups of people to turn the opinion of people who are discussing opinion-forming articles on their side. This is the direction of the discussion based on apparently greater number of contributions on the topic. At the time of easy access to e-mail address creation and various other sign-on methods, it is not difficult for an individual to create dozens of hundreds of different profiles. The ultimate goal is to create an impression of the majority of the discussion and to suppress the plurality of screens.

3 Dataset

The dataset was in JSON format with the following structure shown in the picture paper (see Figure 1). However, not every record has provided all the data, so some may have some variables outage. However, it does not have a significant impact on our experiment.

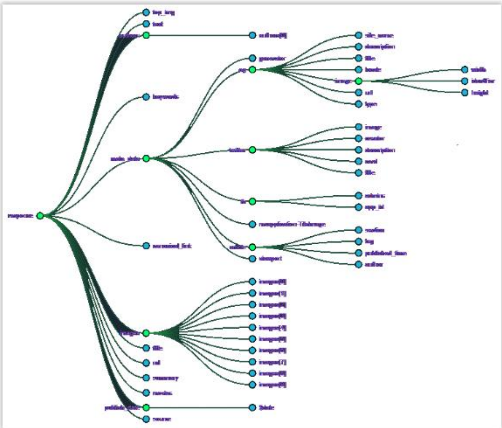

Figure 1. JSON dataset format.

The dataset was split into fake news and real news and each section had 75 records. As a result, 150 records were available.

When examining a dataset, we compared the amount of characters in each type of post. However, we did not find any significant difference in their length indicating the possibility of such a differentiation.

4 Analysis of specific features for fake news

Our first concept was that on average we can recognize real news from fake news using the amount of characters used in each text status. By comparison, however, we found that the average number of characters in fake news is 2271,417 and in real news it is 3880,042 characters. On a more detailed assessment of the breakdown of the individual points associated with individual records, we have come to the conclusion that there is no number of characters in which we could assume with certain validity that the text message will be fake or real. This fact is evidenced by the accompanying graph (see Chart 1).

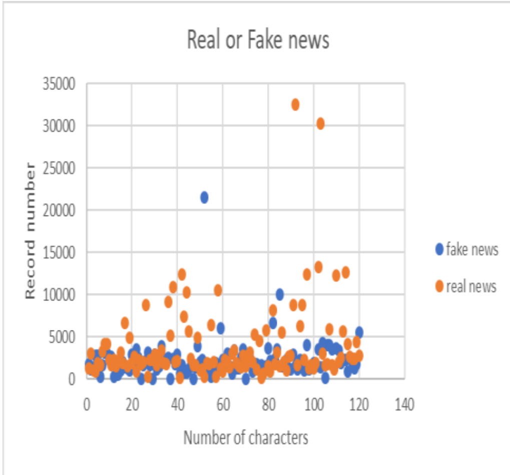

Chart 1. Sample output of a fractal tree drawing algorithm.

Our second concept was that according to the number of expressions in the textual contribution it is possible to judge with certainty whether it is a false message. Based on these beliefs, dictionaries of the most commonly used emotionally coloured words have been created. The dictionary I use in my work is created by the authors of the most frequently used meanings of each word. I created a 10 subcategory of the meaning of emotionally coloured words. From the most negative to the most positive. However, the final identification of emotionally coloured words consisted only of a positive or negative emotion evaluation.

The experiment with our dataset we had available was programmed in the Python programming environment. We analyse 150 texts annotated by dataset authors on real news and fake news. We also used this information for cross-validation to check the correct report. Upon completion of validation, we realized that in our case of political news it is not entirely correct to use the method of searching for emotionally coloured words. As the information is generally distorted in this way of identification. Ruthlessly because of citing the authors of most contributions from official media.

The results of our experiment show that when determining whether the message fake news has reached a success rate of 76.32% and a real news identification of 68.00%.

However, we realized that it is much better to divide into much more categories. These more correspond to the individual's feelings. Unfortunately, we have not noticed these intermediate results and therefore we will have to come back to the experiment in the future.

5 Conclusion and work to future

Obligation that this analysis will be further reviewed with more individual moods. However, good results are seen.

After analysing the work in the area of our interest, we will deeper into the various datasets that we have already prepared for testing. We will add other methods of investigating and recognizing fake news. However, the method of examining emotionally coloured words is still to be explored, since the writing style of the author of the work is quite characteristic. From other methods, the most promising User Model method appears to us based on various parameters.

A user model that could identify suspicious messages not just offline. This method provides important information about authors for different types of metadata. This tool is mentioned in his article.

Acknowledgement. This work was supported by grants No. 002STU-2-1/2018 supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic and the Science and Technology Assistance Agency under the contract No. APVV-17-0267.

References

[1] https://www.researchgate.net/publication/282850983 _Bahaviorism, 14. 2 2019

[2] Granik, M., Mesyura, V., (2017), "Fake news detection using naïve Bayes classifier", 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)

[3] Biyani, P., Tsioutsiouliklis, K., Blackmer, J., (2016), "8 amazing secrets for getting more clicks": Detecting clickbaits in news streams using article informality. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. AAAI Press, pages 94–100

[4] Kumar, S., West, R., Leskovec, J., (2016), Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes. In Proceedings of the 25th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, pages 591–602

[5] Wang, W.Y., (2017), "Liar, liar pants on fire": A new benchmark dataset for fake news detection. In Proceedings of the Association for Computational Linguistics Short Papers. Association for Computational Linguistics

[6] Allcott, H., Gentzkow, M., (2017), Social Media and Fake News in the 2016 Election, Journal of Economic Perspectives, num. 2, Volume 31, pag. 211-236

[7] Zhou, X., Cao, J., Jin, Z., Xie, F., Su, Y., Zhang, J., … Cao, X. (2015). Real-Time News Certification System on Sina Weibo, 983–988

[8] Nelson, L. J., (2017). Fake News, Fake Problem? An Analysis of the Fake News Audience in the Lead Up to the 2016 Presidential Election. *Conference: TPRCAt: Arlington, VA*

[9] Kshetri, N., & Voas, J. (2017). The Economics of "Fake News." *IT Professional*, *19*(6), 8–12

[10] Návrat, P., Chudá, D., Bieliková, M., Šimko, M., Šimko, J., Martonová, A., … Bednár, P. (2018). Rebelion – odhaľovan ie antisociálneho správania na Webe, 2–5

[11] Lessenski, M. (2018). *Resilience To ' Post-Truth ' and Its Predictors in the New Media Literacy Index 2018*. Sofia

[12] Vajdová, D., Masaryk, R,, Kostovičová, L. (2018). Intervencia zameraná na odlišovanie dôveryhodných a nedôveryhodných správ u stredoškolských študentov, Intervention Focused on Discerning Trustworthy and Untrustworthy News in Seco ... INTERVENTION FOCUSED ON DISCERNING TRUSTWORTHY AND. *Nekonečno v Psychologii*, (March), 87–96.

[13] Informa, K. (2013). Sdružení Bezcitných Cynických a Zlých Individuí: Trolling na sociální síti Facebook v českém prostředí, Bakalárska práca. Retrieved from

[14] Siersdorfer, S., Chelaru, S., Nejdl, W., Pedro, J. S. (2010), „HowUseful are Your Comments? Analyzing and Predicting YouTube Comments and Comment Ratings", Conference: WWW 2010, NC, USA, (April), 26-30Owicki, S., Gries, D.: Verifying properties of Parallel Programs: an Axiomatic Approach. Communications of the ACM, (1976), vol. 19, no. 5, pp. 279–285.

[15] Vardhan, A.: Distributed Garbage Collection: A Transformation and its Applications to Java Programming. Master's thesis, (1998).

[16] Widom, J., Panangaden, P.: Expressiveness Bounds for Completeness in Trace-based Network Proof Systems. In: CAAP'88: Proc. of the 13th Colloquium on Trees in Algebra, LNCS 209. Springer, Berlin, (1988), pp. 200–214.

Synthetic Gradient in Residual Networks

Lukáš HANINČÍK*

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
lukas.hanincik@gmail.com

Abstract. Phases of forward-propagation and backpropagation in training of neural networks are processed by synchronous approach. All these phases consist of sequential computation steps of forward transformation and backward propagation of the error gradient. It causes all layers to be dependent on the neighboring layers. We show a method, how to remove the phase of backpropagation from the training process. This idea is based on the synthetic gradient which is an approximated version of the real gradient and is provided by modules of synthetic gradient. This method replaces the sequential approach of the training process by an parallel approach. Based on performed experiments, synthetic gradient increase a speed of training process of residual networks and it reaches as similary results as an original method, backpropagation. We implemented this method in a residual neural network in the DeepCYTOF algorithm which is trained for classification of cytometric data.

1 Introduction

The training process in ordinary neural networks consists of the forward-propagation and backpropagation phases. In forward-propagation, input data is propagated through the layers along the whole neural network. Each layer performs some transformation of received data and sends it to the following layer. When data reaches the output of the neural network, the error value is calculated (basically as difference of output of neural network and expected value). Forward-propagation ends by quantifying the error value [2].

After forward-propagation follows backpropagation. During backpropagation, the error value propagates backward through the whole network from the output to the input. Each layer uses the propagated error for updating its parameters (as known as weights) [2, 4]. A layer which obtains an error value calculates the gradient $\frac{\partial L}{\partial h_i}$ where L is an error value and h_i are transformed data of layer i [2].

Each layer which updated its parameters is going to propagate a gradient to the previous layer, which uses it for calculating its own gradient. The backpropagation algorithm does not allow to update a layer's parameters parallel with the updating parameters of the other layers [4]. Each layer which is ready to update its parameters must wait for the gradient propagated from the following layer and it means that each layer is dependent on the neighboring layers. Based on this limitation, a neural network is able to train just by synchronous approach, layer by layer [1, 4].

The synchronous approach of training neural networks causes three types of network lockings to occur during the training process [1, 4]:

1. *Forward locking* - no module can process its incoming data before the previous nodes in the directed forward graph have executed.
2. *Update Locking* - no module can be updated before all dependent modules have executed in forwards mode.
3. *Backwards Locking* - no module can be updated before all dependent modules have executed in both forwards mode and backwards mode.

In this paper we show a method called *Decoupled neural interface*, which removes *Update locking* and *Backwards locking* from the training process of neural networks. Decoupled neural interface (DNI) is used as provider of a synthetic gradient and allows to update a layer's parameters in parallel. We implemented DNI to a residual network (ResNet) which is trained for the blood cells calibration. Given Resnet is one of the components of the DeepCyTOF algorithm. Calibrated blood cells are subsequently classified to discrete groups which allows to predict a wide spectrum of diseases.

---

* Bachelor study programme in field: Informatics
Supervisor: Tomáš Jarábek, Institute of Informatics, Information Systems and Software Engineering, Faculty of Informatics and Information Technologies STU in Bratislava

2 Decoupled neural interface and synthetic gradient

As we introduced above, common backpropagation of error value limits neural network on synchronous approach. Each layer, which transformed input data is ready to update its parameters but it must wait for gradients propagated from following layer. Based on this fact, all layers are trained sequentially (as shown on Fig. 1 a).

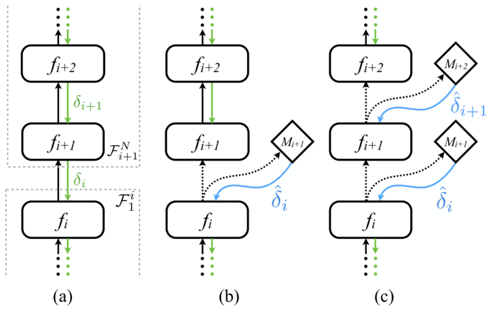

Figure 1. (a) A section of a vanilla feed-forward neural network F_1^N. (b) Incorporating one synthetic gradient model for the output of layer i. This results in two sub-networks F_1^i and F_{i+1}^N which can be updated independently. (c) Incorporating multiple synthetic gradient models after every layer results in N independently [1, 4].

Decoupled neural interface presents an interface for communication with the *module of synthetic gradient*. Module of synthetic gradient is a plain neural network with the one dense layer which is trained for asymptotic approximation of real gradient. Generally, the module of synthetic gradient predicts a real gradient backpropagated from the output of neural network [1, 4].

The module of synthetic gradient receives the transformed data of accompanied layer (output of accompanied layer). As shown on Fig. 1 b, layer f_i sends transformed data to layer f_{i+1} and so on to the module of synthetic gradient M_{i+1}. Module of synthetic gradient M_{i+1} receives transformed data of layer f_i and uses it for predicting a gradient. The predicted gradient is sent back to layer f_i which shall update its parameters immediately [4].

Since layer f_i is able to update its parameters immediately without the need of additional informations from other layers, it breaks a dependence on gradient propagated from layer f_{i+1}. Each layer which directly communicates with the module of synthetic gradient is not dependent on the other layers in phase of backpropagation anymore (as shown on Fig. 1 c). Each layer which is ready to update its parameters, doesn't have to wait for the gradient backpropagated from following layer because it can updates its layers using the synthetic gradient provided by accompanied modules of synthetic gradient. It completely removes *Update locking* and *Backward locking* from the neural network training process [4].

2.1 Training of modules of synthetic gradient

As we introduced above, modules of synthetic gradient are plain neural networks trained for asymptotic approximation of the real gradient. Performance of modules of synthetic gradient is increased by training the model of synthetic gradient. As shown on Fig 2, the module of synthetic gradient M_B receives a gradient by layer f_B. That is the same gradient as which was used for updating parameters of layer f_B. Module of synthetic gradient M_B considers the obtained gradient δ_A as ground-truth gradient and updates its parameter in order to minimize the L_2 distance $||\delta_A - \hat{\delta}_A||$, where δ_A is real gradient propagated by layer f_B and $\hat{\delta}_A$ is synthetic gradient used for updating parameters of layer f_A [1,4]. The feedback model M_B can also be conditioned on any privileged information or context, c, available during training such as a label [4].

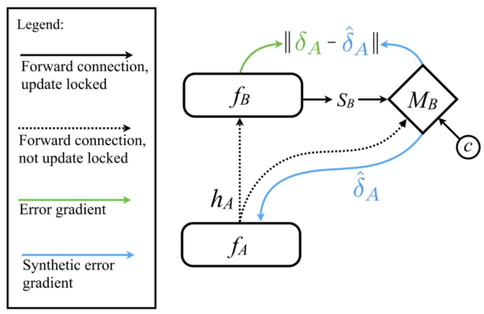

Figure 2. General communication protocol between f_A and f_B [4].

The key idea of DNI is to approximate the gradient of the error $\frac{\partial L}{\partial h_i}$, using the synthetic gradient $\hat{\delta}_i$ as $\hat{\delta}_i \approx \frac{\partial L}{\partial h_i}$, where L is the prediction error and h_i is the output of layer i (data transformed by layer i). Synthetic gradient is the output of activation function $\hat{\delta}_i = \hat{\delta}(h_i, \phi_i)$, where ϕ_i are parameters of the module of synthetic gradient i [4].

As we mentioned, the synthetic gradient must be trained to approximate the true gradient. For updating the parameters of the module of synthetic gradient i, we use the synthetic gradient provided by module of synthetic gradient $i + 1$, $\hat{\delta}_{i+1}$. We "unroll" our synthetic gradient just one step,

$$\hat{\delta}_i \approx \frac{\partial L}{\partial h_i} = \frac{\partial L}{\partial h_{i+1}} \frac{\partial h_{i+1}}{\partial h_i} \approx \hat{\delta}_{i+1} \frac{\partial h_{i+1}}{\partial h_i}$$

Unrolled synthetic gradient $z_i = \hat{\delta}_{i+1}\frac{\partial h_{i+1}}{\partial h_i}$ is considered as constant training target for the synthetic gradient $\hat{\delta}_i$ [4].

We update parameters ϕ_i of module of synthetic gradient i so as to minimize the mean-squared error of these one-step unrolled training targets z_i. For updating parameters ϕ_i stochastic gradient descent [2] is used [4].

3 Residual network

Very deep neural networks have a problem called *vanishing gradient* [3]. A way to solve the problem of vanishing gradient is by replacing the original nonlinear layers [2] by residual blocks. A residual block consists of a few nonlinear layers (as known as *stacked layers*) and a identity shortcut. The identity shortcut is a specific interface which provides identity mapping, received on the input of the residual block, to the output of the residual block (as shown on Fig. 3)

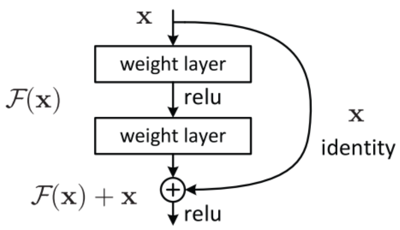

Figure 3. An illustration of a residual block. [3].

As introduced [3, 7], it is easier to optimize the residual mapping than to optimize the original, unreferenced mapping. Even thought the identity shortcut, a residual network can still be trained end-to-end by stochastic gradient descent with backpropagation [3]. [4] introduced, there is nothing to restrict the use of DNI for arbitrary network graph. We think, an implementation of DNI to residual network should not cause a significant degradation of network performace. Deeper functionality of residual networks are explained in [3, 7].

4 DeepCyTOF algorithm

DeepCyTOF algorithm is a algorithm designed for cytometric analysis. Output of DeepCyTOF algorithm are gated blood samples. In cytometry science, gating represents a classification of blood cells to discrete groups. It helps to know and understand different human diseases, predict occurrence of disease and diagnose a disease in early stage [5, 6].

The DeepCyTOF algorithm consist of three separate neural networks. First neural network is denoising autoencoder (DAE). DAE is the neural network trained for replacing zero values for real non-zero values. This process is called denoising. Denoised data follows to MMD-ResNet. MMD-ResNet is specific residual network trained for calibration input samples to a reference target sample. This residual network uses maximum mean discrepancy (MMD) as loss function [2]. Training of MMD-ResNet leads to decreasing of gap between two distributions. Denoised and calibrated data are finally gated using feed-forward neural network [2]. Deeper functionality of DeepCyTOF algorithm is explaned in [6].

5 Experiment and results

As we introduce DeepCyTOF algorithm which is constructed by three neural networks. Subject of our research is observation of influence of synthetic gradient on the accuracy of residual network.

We implemented DNI to MMD-ResNet which is trained for calibration of input samples. We used DNI as provider of synthetic gradient for each layer (except of input and output layers). Each stacked layer of MMD-ResNet includes a batch normalization and ReLU non-linear activation function [2, 8]. Residual blocks consist of two stacked layers. Input stacked layer of residual block has weight matrix of size 25×8 and output stacked layer has weight matrix of size 8×25. Each layer updates its parameter using *RMS Prop* optimizer [2]. Mini-batch size for the MMD-ResNet is 1000. In the MMD-ResNet, a penalty of 10^{-2} on the l_2 norm of the network weights is added to the loss for regularisation. We used the same data as [6] which was available on its GitHub repository.

The kernel used for MMD-ResNet is a sum of three Gaussian kernels

$$k(x, y) = \sum_{i=1}^{3} exp\left(- \frac{||x - y||^2}{\sigma_i} \right)$$

we set the $\sigma_i \in \{\frac{m}{2}, m, 2m\}$, where m is the median of the average distance between a point in the target sample to its 20 nearest neighbours.

First runs give a promising results. We observed a progress of MMD between calibrated source sample and reference target sample. First results indicated almost double increase of time spent for training MMD-ResNet with using an original backpropagation against a MMD-ResNet trained using synthetic gradient. This measurement was performed on 400 iterations.

As shown on Fig. 4, during the training of the MMD-ResNet, the MMD between source sample and target sample is decreasing. Decreasing the MMD means, that the distributions of values in source sample is getting similar to distribution of values in target sample, but not equal. Based on this experiment, we can say, that implementation of synthetic gradient in residual network do not have a significantly negative influence to performance of given residual network. As we expected, MMD of ResNet trained by synthetical gradient decresed slowly but in last iterations, it reached the MMD of ResNet trained by original backpropagation.

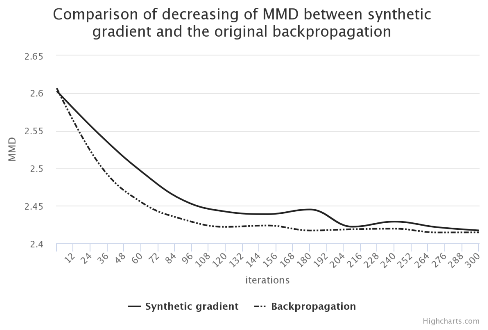

Figure 4. Progress of decreasing of MMD between source sample and target sample. On the plot above we can see, that ResNet trained by backpropagation stops learning after a hundred iterations of learning. ResNet trained by synthetic gradient is able to reach a similar results after another hundred iterations. Afterwards, both of them begin to overfit.

MMD metrics of evaluation of ResNet is based on stochastical approach. A MMD value is computed between a random batch from source sample and random batch from target sample. Size of these batches is 1000. Stochastic selection of batch from target sample caused a variability of error value (MMD). With respect to this fact, we smoothed a MMD progress line in plot.

Table 1. Table displays a basic comparison of MMD ResNet trained by synthetic gradient and original backpropagation. All values are measured for models to reach an optimal value of MMD, 2.419. $Iterations$ *shows a number of iterations needed for a reaching an optimal values.* $Time$ *row shows a time needed for a reaching an optimal value of MMD.* $F1score$ *presents a* $F1$ *score of final classifications of data calibrated by MMD-ResNet.*

MMD = 2.419	**Backprop.**	**Synth. grad.**
Iterations	150	240
Time	107s	90s
$F1$ score	81.71	76.29

As shown in table 1, MMD-ResNet trained by synthetic gradient reachs a bit lower $F1$ score in final classification of calibrated data. Difference is prominent, but it can be resolved by upgrading an implementation of modules of synthetic gradient. Training of MMD-ResNet by backpropagation spent more time for reaching an optimal value of MMD. With respect to this fact, we believe, that synthetic gradient has significantly better time results in models, which requires more iterations of learning for reaching a better performance.

6 Conclusion

We mentioned above, that output layer of our MMD-ResNet does not have a accompanied module of synthetic gradient. This is not needed because the output layer receives ground-truth gradient calculated on the real error value of MMD function.

We used residual network for implementation of synthetic gradient because of its depth. Deeper networks which use synthetic gradients in the training process, usually reach a higher performance than the shallower networks [4]. From that point of view, residual networks are a great choice for observing the influence of synthetic gradients, because we can better understand, how the neural network behaves by using synthetic gradient to update its layer parameters. For the evaluation of MMD-ResNet, we used the $F1$ score statistical test. $F1$ score was applied on the final classification of data, which was calibrated by MMD-ResNet trained by synthetic gradient.

Another reason, why we used residual networks for this experiment is, that we haven't noticed any documented experiment where a residual network with implemented DNI is used.

References

[1] Czarnecki, W.M., et al.: Understanding Synthetic Gradients and Decoupled Neural Interfaces, 2017.

[2] Goodfellow, I., et al.: *Deep learning*. Volume 1. MIT press Cambridge, 2016.

[3] He, K., et al.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[4] Jaderberg, M., et al.: Decoupled Neural Interfaces using Synthetic Gradients, 2016, vol. 1.

[5] Li, H., et al.: DeepCyTOF: Automated Cell Classification of Mass Cytometry Data by Deep Learning and Domain Adaptation. *bioRxiv*, 2016, p. 054411.

[6] Li, H., et al.: Gating mass cytometry data by deep learning. *Bioinformatics (Oxford, England)*, 2017, vol. 33, no. 21, pp. 3423–3430.

[7] Targ, S., et al.: Resnet in Resnet: Generalizing Residual Architectures, 2016, pp. 1–7.

[8] Xu, B., et al.: Empirical Evaluation of Rectified Activations in Convolutional Network, 2015.

Conversion of the Spoken Word into Text using Neural Networks

Denis MITANA*

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
xmitana@stuba.sk

Abstract. The problem of converting spoken word into text exists in many languages. There are several solutions in popular languages like English and therefore we have focused on a less known Slovak language. Speech recognition can be used in multiple fields such as automated identification, automated transcription, education or everyday life. Our goal is to develop artificial neural network that can convert spoken words into a text. For this purpose, we decided to use and modify the existing WaveNet model, whose architecture is based on the use of dilated convolutions. Although WaveNet operates with raw waveform, we experimented with both raw waveform and *Mel frequency cepstral coefficients* (MFCCs). Best achieved *word error rate* (WER) is 54.62% by using MFCCs.

1 Introduction

The problem of speech recognition is difficult and involves a lot of sub-problems such as dealing with noise in the environment, different speakers with various accents, low resource languages, a cocktail party effect and more. In our paper we address the solution of a speech recognition task for Slovak language in which we identified only one publicly available speech corpus [12] and which is also more difficult for people to acquire. According to our knowledge, the solution using neural networks does not exist for Slovak language, even though a research is actively going on [6].

Log mel features and MFCCs belong to the most common representations of speech which are still used today. These features are hand-engineered from raw waveform. Solutions, which use this type of features, achieve a very high success rate. Although nowadays, the use of raw waveform and automated extraction of features is becoming more and more popular thanks to deep learning and *Convolutional neural networks* (CNN). This approach also achieves remarkable results. Advantages of classic hand-engineered features are that they will surely provide useful information. On the other hand, they can also throw away a lot of useful information from raw waveform. Raw waveform contains all the necessary information and therefore deep neural network can extract all features. The disadvantage can be longer training time, because a feature extraction can be tedious.

The use of *Long short-term memory recurrent neural networks* (LSTMs) is intuitive and it achieves excellent results. Drawback of LSTMs is their high computational complexity during training when enlarging the context and therefore impossibility of using raw waveform as an input. WaveNet [4] shows that CNNs are able to model sequences such as speech and what is more, they are able to extract features from raw waveform. Dilated convolutions proved that they can fully replace LSTMs and that they are even less computationally complex than LSTMs, which is a huge advantage that allows to use a larger context and raw waveform as an input.

As CNNs have proved to be successful in a speech recognition task [4, 9, 11, 13], we decided to use an architecture based on them. In our paper we modified a WaveNet [4] model from a generative model to a discriminative model. We also used *Connectionist temporal classification* (CTC) [1] loss function instead of two original loss terms. The experiments were performed with MFCCs and raw waveform as an inputs.

2 Related work

There are many solutions to the problem of the speech recognition in the English language. We focus on a

* Bachelor study programme in field: Informatics
Supervisor: Assoc. Professor Peter Lacko, Institute of Informatics, Information Systems and Software Engineering, Faculty of Informatics and Information Technologies STU in Bratislava

IIT.SRC 2019, Bratislava, April 17, 2019, pp. 25–28.

few, which affected our paper the most and we also analyze papers in the Slovak language.

Speech recognition using RNN

In 2013 deep *Bidirectional LSTM* (BLSTM) was used for the first time for speech recognition [2]. According to the authors, the addition of layers has caused a dramatic improvement over the single layer LSTM. Achieved *phoneme error rate* (PER) is 17,7% on *TIMIT* speech corpus, which were state-of-the-art results. Mel-spectrogram was used as an input and CTC [1] was used for alignment of acoustic frame labels and also as loss function.

Speech recognition using CNN

CNNs have several uses in a speech recognition task. CNNs can be used to reduce speech spectral variation [10] and to model correlation in a speech spectrum [9] and thus obtaining a better representation of speech. CNNs can also replace hand-engineered features created by filter banks [11] or even replace LSTM layers [4].

In 2017, the authors [13] reported CNN architecture (denoted as RCNN-CTC), which contains residual connections and uses CTC [1] loss function. The system also uses a language model and mel filter banks based features. The authors [13] emphasize that they used more than 40 hidden layers thanks to residual connections. The lowest achieved WER was 5.35% on *Wall Street Journal* speech corpus. On larger *Tencent Chat* speech corpus lowest WER was 14.26%.

WaveNet

WaveNet [4] is based on deep CNN and can generate sound waves. It achieves state-of-the-art results when it is used for text-to-speech (TTS) tasks. In addition to speech, WaveNet can also generate music and recognize speech. WaveNet demonstrates that dilated convolutions can expand receptive field more efficiently than LSTMs. WaveNet was tested for a speech recognition on *TIMIT* speech corpus and it achieved PER 18.8%. As we were inspired by this model, we described its architecture in Section 3.2.

Slovak language

In 2018, a deployment of the system in which a deep learning is expected has begun. The paper [6] deals with a speech recognition in Slovak, English and Mandarin languages, it is case of a multilingual speech recognition system. Unfortunately, at the time of writing this paper, the system has not yet been finished and it uses an open source system Kaldi which is based on GMM-HMM. To train a part of the Slovak language *TUKE-BNews-SK* [7] and *TEDxSK & JumpSK* [12] speech corpora were used. The best achieved WER was 16.89% using 3-gram language model. The authors plan to try an approach using DeepSpeech model.

3 Method of solution

Our method is based on CNNs and modified WaveNet model [4] to convert speech to text instead of generating sound. We used two forms of representation of an input sound signal, specifically MFCCs and raw waveform. We also utilized CTC [1] loss function to deal with the labelling of unsegmented sequential data.

3.1 Input representation

In our experiments we separately used MFCCs and raw waveform as inputs. Both ways of an input representation have pros and cons.

MFCCs are hand-engineered features from speech, which are often used and achieve very good results. It is obvious that these features hide relevant information. On the other hand, a lot of relevant information may be discarded during processing as well as several parameters must be suitably set to create MFCCs. To create MFCCs we used 25ms window and 10ms shift [5] (60% overlap), as a window function we used the Hamming window function and we utilized 26 coefficients.

The use of raw waveform has recently expanded because of rise of CNNs and deep learning. This is mainly because deep CNNs can extract very useful features from given input. Also, the use of raw waveform erases the setting of parameters that exists when creating MFCCs.

The way we used raw waveform in our paper differs from the original method used in WaveNet [4]. Original method uses raw waveform, which is not framed into smaller windows with some shifting. Unfortunately, we could not use this method because we were unable to obtain the Slovak speech corpus where labels have lower level than sentence level and the use of sound samples for whole sentences was impracticable due to GPU memory capacity (we used GTX 1080Ti with 11 GB of memory). For this reason, we framed raw waveform using 25ms window and 10ms shift, although this is not a common use.

3.2 Architecture

We were inspired by WaveNet [4] model, which is based on dilated causal convolutions, gated activation units, residual and skip connections. We modified this model due to different input representation (fig. 1). Thanks to this architecture we did not have to use LSTMs, which are computationally more complex.

Dilated convolutions (à trous) can more effectively use large context than ordinary convolutions. Receptive field increases linearly with each added layer in ordinary convolutions, if we assume that filter size does not change. By using dilated convolutions, dilation can be exponentially increased in each added layer, which results in exponentially increasing of the receptive field [4]. Since filter size is the same in each layer, the number of parameters increases only linearly [14].

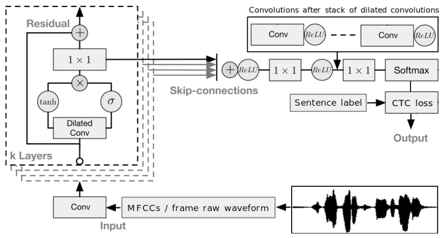

Figure 1. The architecture of model. We used MFCCs or framed raw waveform as separate inputs. The figure is taken from [4] and modified.

Other important elements in architecture are gated activation units, residual and skip connections. The authors [4] report, that non-linearity caused by gated activation units works significantly better than by ReLU and that residual and skip connections help to speed up the convergence and allow to train much deeper models. Because of the different input representation, we modified some parts of the model for our use. Firstly, as we used framed sound signal with 25ms window and 10ms shift, we could use smaller stack of dilated convolutions, because our one time step is 25ms instead of $(1/sample_rate * 1000)$ms. Secondly, we used only one CTC [1] loss function instead of two original loss terms. Thirdly, we did not use mean pooling layer after the stack of dilated convolutions again because of the framed input. Fourthly, we added several convolution layers after a stack of dilated convolutions.

3.3 Speech corpus

We identified several speech corpora [3, 7, 8, 12], but only [12] is publicly available. We used *TEDxSK & JumpSK* [12] speech corpus which contains about 46 hours of automatically and about 12 hours of manually transcribed talks and lectures. Automatically transcribed part is not flawless and WER of labels is about 14%. Created training set contains about 40 000 training examples. This speech corpus is also not properly balanced, although it contains 227 unique speakers so approximately only 30% are female speakers [12].

3.4 Speech augmentation

We tried adding two methods of speech augmentation because of the lack of data and for generalization improvement. The first method is called *Speed perturbation* and consists of a random change of speed of speech from 80% to 120%. The second method is a simple addition of white noise (to each training sample was added different white noise). Using augmentation, we have doubled the training set.

3.5 Experiments

Experiments were performed on *TEDxSK & JumpSK* [12] speech corpus using both MFCCs and raw waveform separately. We split a speech corpus into a training, validation and test set at ratio 90:5:5. This division is not ideal, but a speech corpus is quite small and even in such division we suffered from the lack of data.

The model has quite a lot of parameters which have to be set appropriately and optimized to achieve better results. Increasing the number of output channels improved results as well as the use of three stacks of dilated convolutions with dilations 1, 2, 4, 8, 16 (receptive field is ∼0.965s), which slightly improved results compared to two stacks. Another significant improvement was the addition of convolution layers after the stack of dilated convolutions (fig. 1). We set a kernel size of dilated convolutions to 2 as in original WaveNet [4] model.

The lack of data was manifested by overfitting and inability to improve on the validation set. We identified overfitting when a validation loss increased and training decreased. At that time, WER and CER stopped to decrease and sometimes also started to increase. At this point we stopped training. To prevent overfitting, we added a batch normalization and L2 regularization. We have experimented with optimization of L2 regularization, but the best results were achieved without it. We have also tried speech augmentation (see Section 3.4), that also improved results and decreased tendency to overfitting.

4 Evaluation

To evaluate our results, we used a WER and *character error rate* (CER). The results are shown in Table 1. As we can see, models with more channels were able to get better results, although they were more prone to overfitting. The experiments have shown that it was easier for the model to learn something from MFCCs than from raw waveform. Although the results of the smaller models that used MFCCs or raw waveform were comparable, the enlargement of the model did not improve results when raw waveform was used. We do not understand this phenomenon and have not had enough time to find out why it happened.

Our best achieved results are 54.62% WER by using MFCCs on an augmented speech corpus. Achieved results are not very good. Unfortunately, to our knowledge, there is no solution which uses only

*Table 1. Results of experiments. # denoted number of channels (r - residual, s - skip, d - dilation). * denotes augmented dataset. CAD denotes convolutions after dilated convolutions (`number_of_layers x kernel_size`).*

input	dilations	#r	#s	#d	CAD	epochs	CER	WER
MFCCs	$(1, 2, 4, 8, 16) * 2$	26	26	26	-	300	36.22%	90.99%
MFCCs	$(1, 2, 4, 8, 16) * 2$	64	64	64	-	200	30.41%	85.03%
MFCCs	$(1, 2, 4, 8, 16) * 2$	128	128	128	-	60	27.43%	79.86%
MFCCs*	$(1, 2, 4, 8, 16) * 2$	128	128	128	-	38	25.81%	76.59%
MFCCs*	$(1, 2, 4, 8, 16) * 2$	256	256	256	5x7	23	20.79%	66.75%
MFCCs*	$(1, 2, 4, 8, 16) * 3$	256	256	256	5x7	25	20.16%	65.17%
MFCCs*	$(1, 2, 4, 8, 16) * 3$	256	256	256	10x7	53	22.16%	67.07%
MFCCs*	$\mathbf{(1, 2, 4, 8, 16) * 3}$	**256**	**256**	**256**	**5x21**	**13**	**17.33%**	**54.62%**
MFCCs	$(1, 2, 4, 8, 16) * 3$	256	256	256	5x21	7	17.58%	56.47%
raw	$(1, 2, 4, 8, 16) * 2$	128	128	128	-	85	33.74%	85.85%
raw	$(1, 2, 4, 8, 16) * 3$	256	256	256	5x21	10	37.02%	85.39%

TEDxSK & JumpSK [12] speech corpus. Although the authors [6] used *TEDxSK & JumpSK* [12] they also used *TUKE-BNews-SK* [7]. Their best results are 16.89% WER. However, it should be noted that *TUKE-BNews-SK* [7] speech corpus is significantly larger and authors also used a language model.

By gradual optimization of model parameters and by data augmentation we were able to get even better results. We assume that the biggest problem is the lack of data. Around 40 000 training examples are still not enough for the speech recognition task as other papers in the English language use about 5 times more training samples.

5 Conclusion

In this paper, we presented a speech recognition in the Slovak language using the modified WaveNet model based on dilated convolutions, gated activation units, residual and skip connections. As an input we separately used MFCCs and framed raw waveform. To provide alignment of acoustic frame labels and as loss function we used CTC. In order to achieve better results and for future work we would like to get another Slovak speech corpus and add other speech augmentation methods to prevent overfitting and improve generalization. We would also like to add a language model to avoid typos in transcripts and improve WER.

Acknowledgement. This work was supported by grant No. 002STU-2-1/2018 supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic.

References

[1] Graves, A., Fernández, S., Gomez, F., Schmidhuber, J.: Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks. In: *Proc. of the 23rd int. conf. on Machine learning*, 2006, pp. 369–376.

[2] Graves, A., Mohamed, A.r., Hinton, G.: Speech recognition with deep recurrent neural networks. In: *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, IEEE, 2013, pp. 6645–6649.

[3] Milan, R., Marian, T., Sakhia, D.: MobilDat-SK- a Mobile Telephone Extension to the SpeechDat-E SK Telephone Speech Database in Slovak. In: *SPECOM'2006*, 2006, pp. 449–454.

[4] Oord, A., et al.: WaveNet: A Generative Model for Raw Audio. *ICASSP, IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, 2016, pp. 2353–2356.

[5] Paliwal, K.K., Lyons, J.G., Wójcicki, K.K.: Preference for 20-40 ms window duration in speech analysis. In: *2010 4th Int. Conf. on Signal Processing and Communication Systems*, 2010, pp. 1–4.

[6] Pleva, M., et al.: Towards Slovak-English-Mandarin Speech Recognition Using Deep Learning. In: *2018 International Symposium ELMAR*, 2018, pp. 151–154.

[7] Pleva, M., Juhar, J.: TUKE-BNews-SK: Slovak Broadcast News Corpus Construction and Evaluation. In: *LREC*, 2014, pp. 1709–1713.

[8] Pollak, P., et al.: SpeechDat(E)-Eastern European Telephone Speech Databases, 2000.

[9] Sainath, T., Mohamed, A.R., Kingsbury, B., Ramabhadran, B.: Deep convolutional neural networks for LVCSR. In: *Acoustics, speech and signal processing (ICASSP), 2013 IEEE int. conf.*, 2013, pp. 8614–8618.

[10] Sainath, T., Vinyals, O., Senior, A., Sak, H.: Convolutional, Long Short-Term Memory, fully connected Deep Neural Networks. In: *2015 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 4580–4584.

[11] Sainath, T.N., Weiss, R.J., Senior, A., Wilson, K.W.: Learning the Speech Front-end With Raw Waveform CLDNNs. *In INTERSPEECH-2015*, 2015.

[12] Staš, J., Hládek, D., Viszlay, P., Koctúr, T.: TEDxSK and JumpSK: A New Slovak Speech Recognition Dedicated Corpus. *Journal of Linguistics/Jazykovedný casopis*, 2017, pp. 346–354.

[13] Wang, Y., Deng, X., Pu, S., Huang, Z.: Residual Convolutional CTC Networks for Automatic Speech Recognition. *arXiv preprint arXiv:1702.07793*, 2017.

[14] Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. *ICLR*, 2016.

Data Storytelling Using an In-Memory Columnar Database Accelerated by GPGPU

Andrej FÚSEK, Nikolas HAMRAN, Dávid KUBÍK, Martin PETRÁŠ
Alexandra SMOLOVÁ, Martin STAŇO, Veronika VČELKOVÁ, Jakub VESELÝ*

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
t01-tp2018@googlegroups.com

1 Introduction

Conventional relational databases store data in row based format. However, in data analytics, it is necessary to quickly retrieve individual columns. Columnar databases significantly reduce disk input and output requirements as well as the amount of data that needs to be loaded. Column-oriented databases have very good horizontal scaling, enabling use of low-cost hardware to increase throughput. For this reason, this type of database is ideal for our goal - big data processing.

Since 2007, when NVIDIA launched the CUDA Toolkit, using graphics cards for high performance and scientific computing has become increasingly popular. Graphics cards enable highly parallelised processing of data, which can be useful in many areas. One such area is graphics processing unit (hereafter GPU) accelerated databases. The first open source GPU database and SQL engine was launched by OmniSci in 2017. We are creating GPU-accelerated columnar database that will be suitable for large data analysis. As a reference, we chose OmniSci's popular columnar database - MapD. On large datasets, we want to achieve shorter query execution time than MapD. To see if our solution is also faster than traditional row-oriented databases, we compared our solution with Microsoft SQL (hereafter MSSQL) database with indexes. This project was developed in cooperation with the company Instarea s.r.o. [1].

2 Proposed solution

We have implemented an in-memory columnar database, accelerated using general-purpose computing on graphics processing units (hereafter GPGPU). The solution uses latest technology, which has made it possible to significantly speed up most of its components. The database is programmed in C++17, using the CUDA programming framework, version 10 for GPGPU acceleration. To accelerate transfers of data between system memory and GPU memory, data in system memory is compressed and only decompressed after it is transferred to the GPU. In order to minimize the number transfers of data between system memory and GPU memory, temporary results need for query evaluation are only stored on the GPU. Only input data and final query results are copied between system memory and GPU memory. Input data is cached on the GPU, using Least Recently Used caching algorithm, to further minimize number of transfers. CUDA primarily supports NVIDIA graphics cards, however wrappers are available for AMD GPUs. The solution supports both Windows and Unix-based operating systems. Web user interface built on the Angular web application framework is also provided. The user can visualise results of their queries using a variety of different graphs. We support most common SQL commands excluding "JOIN". In addition to common numeric and date time data types, we also support Polygon and Point types for geographic queries.

3 Architecture

The solution is divided into three main parts, each of them consisting of several components. "Database" part is show in figure 1. This part represents our columnar database. "Application server" serves as a backend for the web user interface. It also includes auxiliary SQL database for storing the configuration of the interface. The application server communi-

* Master study programme in field: Software Engineering, Artificial Intelligence
Supervisor: Marek Lóderer, Intelligent Information Systems, Faculty of Informatics and Information Technologies STU in Bratislava

[1] https://instarea.com/

cates wit the "Database" using custom protocol encapsulated in TCP. It also communicates with "User Interface" via REST calls.

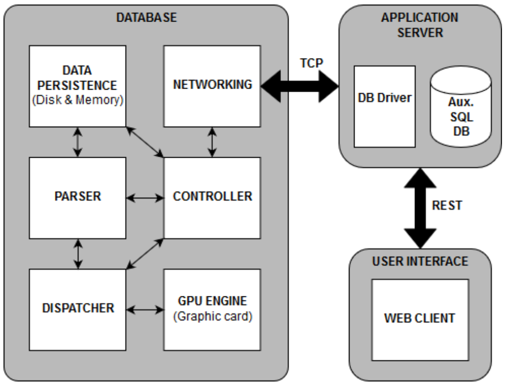

Figure 1. Architecture of the proposed solution.

The "Database" part consists of six components. "Data Persistence" is used to read and write data from/to disk.

The "Networking" part communicates with the application server. It uses TCP network protocol and Google Protocol Buffers for message serialization.

The "Parser" part processes queries into executable format using Antlr4 parser. The executable format has two parts: a list of function pointers and arguments, stored separately using custom memory stream.

After parsing is finished, "Dispatcher" executes functions from the function pointer list. These functions take care of the query execution steps, from loading, through filtering and arithmetic operations, to returning the final result. The currently executing function retrieves appropriate amount of arguments from the memory stream. Another responsibility of "Dispatcher" is to store query execution state which consist primary of currently allocated pointers to loaded columns of data or temporary results in GPU memory. For the sake of efficiency, we prevent re-evaluation of the same expression during query execution on a single block of data, by checking a list of known expression results.

The "Controller" part controls the individual instances of the "Dispatcher" component - each video card has its own instance.

The "GPU Engine" part consists of CUDA kernels and support functions called from the "Dispatcher". The tasks of "GPU Engine" is to filter by "WHERE" condition, calculate arithmetic operations with columns data and performs aggregation functions. An important role is also collection of keys and aggregate values of the "GROUP BY" function inspired by [2]. During this process, atomic functions are used to simplify these parallel computations. Since we support many different data types and operations, we use C++ templates and functors throughout the whole code of this part. Both templates and functors reduce the amount of code and therefore simplify code maintenance. We also support polygon union, intersect, subtraction implemented using [1].

The "Application Server" part creates REST endpoints to expose selected database operations (query execution and CSV file import) via HTTP requests. It has auxiliary SQL database for configuration storage.

The "User Interface" part is the web front-end with which the user interacts. It enables user to execute custom queries and visualise the results using various graphs.

4 Results

We compared our solution with the row-based MSSQL database and column-oriented GPU accelerated MapD. The hardware and software on which the two databases were tested are shown in the table 1.

Table 1. Hardware and software of the computer on which databases were tested.

OS	Windows 10 Pro x64
CPU	AMD Ryzen Threadripper 2950X 16-Core Processor 3.50 GHz
GPUs	2x NVIDIA GeForce GTX 1080 Ti
RAM	128 GB DDR4 3000 MHz

Table 2 shows the results of a comparison of MSSQL, MapD and our solution which is named DropDBase. The executed query had where and group by clauses and the dataset had 1 billion rows of data.

Table 2. First and second query execution times.

	MSSQL	MapD	DropDBase
1st [ms]	12885	50128	545
2nd [ms]	300	47	59

The first query execution time was much longer because of cold cache. The second query used the cached data from the first query to speed up execution.

Acknowledgement. This work was supported by grant No. 002STU-2-1/2018 supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic.

References

[1] Greiner, G., Hormann, K.: Efficient clipping of arbitrary polygons. *ACM Transactions on Graphics (TOG)*, 1998, vol. 17, no. 2, pp. 71–83.

[2] Karnagel, T., Müller, R., Lohman, G.M.: Optimizing GPU-accelerated Group-By and Aggregation. *ADMS@ VLDB*, 2015, vol. 8, p. 20.

Breyslet - IoT Healthcare Monitoring System

Radovan ANDRÁŠ, Michaela CHLEPKOVÁ, Daniel GÁBRIŠ, Viktor LANČARIČ, Juraj LAUKO, Ján PETRÁŠ, Veronika UHNÁKOVÁ, Alexander VALACH*

Internet Technologies
Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
timproj3@gmail.com

1 Introduction

Nowadays, several wearable devices are used to measure and record data such as heart rate, step count or burned calories. Most of these devices are utilized by active people. They can track the progress of their training or the results over a certain period.

At the doctor we could see several devices which are used to determine whether we are healthy or there are some problems. Over time, these devices have shrunk to such an extent that today they can be turned into watch size.

People whose health is often aggravated by illness, injury or other influence should have access to information about their health as quickly as possible. If there is a serious situation, human need to know how to quickly get help.

For example, there are many situations when older people stay at home because their relatives have to go to work. At that moment, the older people have to take care of themselves. If something happens to them, they must know how to get help [2].

As care for our relatives is highly important, we have focused on a development of a wearable device that is able to monitor basic physiological parameters of a person on regular basis. The device is intended mainly for older people and people with health problems. As these people need to keep in touch with their relatives, the device is equipped with SOS button which could be pressed in case of emergency. The device notifies relatives. This will simplify the call for help.

The purpose of the device is to measure certain indicators and send their values to the health center periodically. These values are constantly being evaluated to avoid protentional threats or they can be used in case of emergency to call for help.

2 Our work

In this paper we propose a watch-size device - Breyslet. It contains sensors for pulse, heart-rate, blood oxygen saturation, temperature measurement. There are also a gyroscope and an accelerometer that can be used for fall detection. The Breyslet is lightweight battery-constrained device, so we have put GSM and GPS to separate device called the Breysbox. To keep communication between Breysbox and Breyslet we use a Bluetooth module. Sigfox is used as primary data communication channel for Breyslet [1]. The role of Breysbox is to process data from Breyslet and send them via GSM network when there is no Sigfox network available.

Sigfox is low power wide area network created for IoT devices. It operates at low frequencies (868MHz in Europe) making it easy to communicate over a long distance. The advantage of this technology is the low power consumption and a large coverage. It also has large community and lot of useful examples for developers. It is commonly used when large number of devices would like to send small data over long distances. These advantages make it appropriate for our scenario.

The measured values are sent through the Sigfox network. The Breyslet also features the SOS button. It is used in critical situations when the person needs help. When button is pressed, the data are measured from the sensors and sent via the Sigfox network or via the GSM network, depending on availability. The relatives are then notified, and help can be called too.

2.1 Functionality

The system consists of two devices. One device is attached to the user's wrist – Breyslet (Figure 1). The second device is a box – Breysbox.

The Breyslet is easy to use. Its user has to do no additional configuration. Configuration is done only

* Master's degree study programme in field: Information Technology
Supervisor: Vladimír Kunštár, Institute of Informatics, Institute of Computer Engineering and Applied Informatics, Faculty of Informatics and Information Technologies STU in Bratislava

IIT.SRC 2019, Bratislava, April 17, 2019, pp. 31–32.

by relatives who registers Breyslet with corresponding ID. Since older people do not use smartphones, the device does not need to be paired with one.

The Breyslet mobile application is primarily intended for close relatives who want to be informed about the condition of the person. They can view the measured user data. After SOS button has been pressed, the application automatically informs you that there may be emergency.

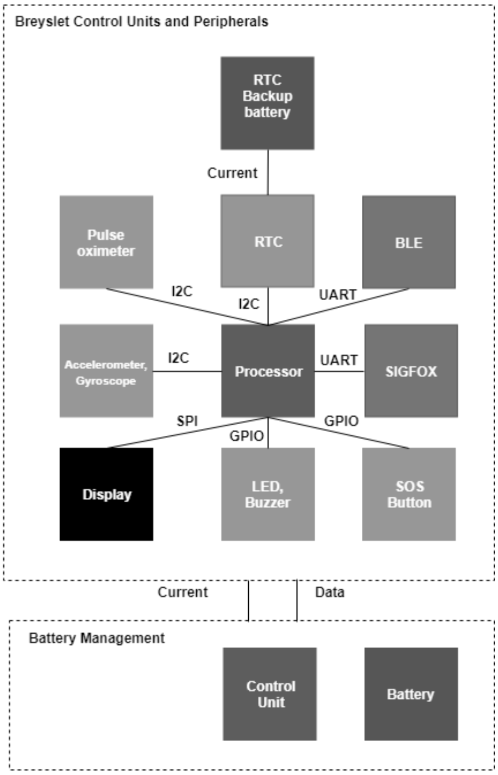

Figure 1. Breyslet.

2.2 Usage

Breyslet is intended for older people and those with health problems. They should be supervised if something happens to them. This is often a challenging task. For such cases, Breyslet has great potential. It provides information of the Breyslet's user to the relatives, so they can act according to it. The older people tend to be more confident when moving or working at home because they are supervised.

2.3 Goal

Our goal is to create a standalone device that monitors health and wellbeing. The Breyslet is able to send measured data through Sigfox network. This fact eliminates a need to be paired with a smartphone. Comparing to Apple Watch [3], Breyslet consumes less energy, so it can last longer.

Another advantage is that Apple Watch has to be paired with iPhone using the latest iOS [3] which makes it also vendor-locked. There is also a cellular version of Apple Watch which allows user to call for help immediately using dedicated SIM card, but it is not available in Slovakia.

Another concept focuses on improving proper breathing techniques during panic situation [4]. It does not try to solve emergency but provide a guide for user when stress situation occurs, and holder needs to calm down.

3 Future work

Next steps are to improve the device hardware and add additional sensors. Since development is still ahead, individual components of the device could be time-shifted, which would result in even smaller size of the device. The functionality will be extended, if there are new sensors like smoke detector, EKG, blood pressure sensor etc., or there are smaller or more accurate ones available on the market.

4 Conclusion

Health care has many options to integrate into the everyday routine of people, especially in the form of wearable devices. Breyslet is one of those options. It provides regular information about the health of the user, alerts the close ones of the fall and allows you to call assistance via the SOS button. As its development is only at the beginning, it has the potential for improvement in the future.

References

[1] Sigfox - The Global Communications Service Provider for the Internet of Things (IoT). [Online; accessed February 23, 2019]. Available at: https://www.sigfox.com/en

[2] BAKER, Stephanie B.; XIANG, Wei; ATKINSON, Ian. Internet of things for smart healthcare: Technologies, challenges, and opportunities. IEEE Access, 2017, 5: 26521-26544.

[3] https://support.apple.com/kb/SP778?viewlocale=en_US&locale=sk_SK

[4] https://maas.museum/app/uploads/2018/02/Austin-Marandos-Miner-smart-watch-Project-Development-Realisation_A4-LANDSCAPE.pdf

GridBox: A Cloud-based Microgrid Modeling Solution

Martin Činčurak, František Ďurana, Richard Mocák, Michal Ostrodický
Dávid Pavelka, Peter Pavlík, Matej Procházka*

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
`tp_06_2018@googlegroups.com`

In the recent years, usage of renewable energy sources has substantially increased. Many households have installed solar panels or other alternatives to supplement the power received from energy providers. However, wider adoption of renewable sources is still hampered by the variability of the energy output of these sources based on the time of day or weather. Currently, excess energy produced by the consumer is often even penalized by the energy provider, which further discourages the renewable energy source adoption.

The move towards self-sustaining microgrids with energy trading can solve these problems by directing the excess energy within communities and greatly reducing the total power consumed from the energy providers. Many different energy producers can operate within the microgrid system distributing the energy to the endpoints. However, to design an effective distributed power system, a software to model, simulate and monitor the microgrid is required.

1 GridBox

Our software, GridBox, developed in cooperation with Sféra a. s., the industry leader in power grid information systems, offers a user-friendly toolbox for creation and modification of graph models representing microgrids and their integration with the simulation and monitoring software. These functions allow the users to easily modify the microgrid and test these changes in simulated environments.

The main tenets of GridBox are modularity and expandability. While the core of the project is the intuitive drag-and-drop user interface for microgrid modelling through web browser, it is not by far its only functionality.

2 Model

Our browser-based application allows the user to easily create a microgrid by inserting microgrid elements such as batteries, solar panels or households into topological (Figure 1) or topographical view (Figure 2) and connecting them into a network. All created networks are saved on the remote repository and allow simple collaboration between colleagues.

The available microgrid element and connection types are also stored in a remote repository and custom elements with specific behavior and graphics can be added by users. These elements can each have a defined set of attributes which can be modified from within the GridBox modeling interface.

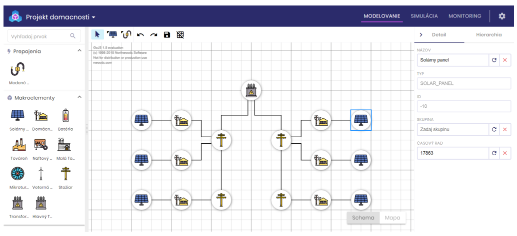

Figure 1. Schema modeling view.

Figure 2. Map modeling view.

* Master study programme in field: Software Engineering, Artificial Intelligence
Supervisor: Marek Lóderer, Institute of Informatics, Information Systems and Software Engineering, Faculty of Informatics and Information Technologies STU in Bratislava

IIT.SRC 2019, Bratislava, April 17, 2019, pp. 33–34.

3 Simulate

The networks saved in the cloud can also be used for running various simulations over the network structure. The system has a universal interface for requesting element networks along with other parameters by simulation engines, which can execute in the cloud and send the results back to the GridBox user interface for visualization (see Figures 3 and 4). The simulation results are sent back as events which can be aggregated into charts for the whole simulation interval or observed chronologically directly within the network structure in the map or the schema view.

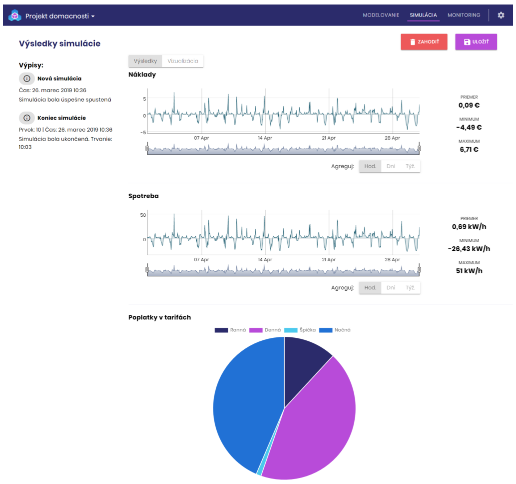

Figure 3. Simulation results view.

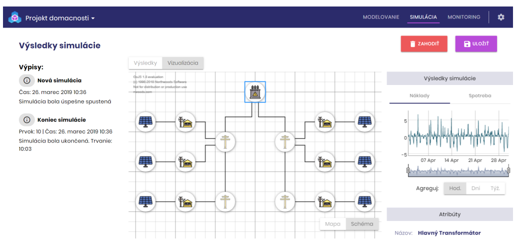

Figure 4. Simulation replay view.

4 Monitor

The other possible use of the GridBox software is the real-time monitoring of current microgrid state. The elements of the network model can be directly linked to real-life power grid elements and show the current state of the microgrid in the visual language matching the simulation view.

5 Architecture

The modular architecture of GridBox is centered around the browser-based user interface built in Angular with GoJS graph library for network modeling and visualization.

The user interface is connected to the database server for storing projects and microgrid element data such as consumption time series and icons. Everything is stored as JSON objects within a NoSQL MongoDB database.

The simulation engines access the database server to retrieve the data about network structure and element types and then save the simulation results to the server. From here, the user interface retrieves them and visualizes the results.

6 Usage outside power industry

Even though GridBox was developed with the needs of microgrid modeling in mind, its usage is not necessarily limited to this area. Since the software is modular, the microgrid specific elements and simulation engines can easily be swapped for different endpoints adhering to the prescribed interfaces. This means that GridBox could also be useful for users in need of software for modeling and simulation of other kinds of graph structures, e.g. telecommunication and computer networks or logic and electric circuits.

Acknowledgement. This work was supported by grant No. 002STU-2-1/2018 supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic.

Holographic Eyes

Zsuzsanna BERNÁTH*, Jakub DOMIAN*, Andrej HUCKO*, Dušan JANEČEK*,
Ján KARAFFA*, Ľudovít POPELKA*, Ľubomíra TRNAVSKÁ*

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
tp8fiit@gmail.com

According to the World Health Organization (WHO), there are about 36 million blind people among 1.3 billion people having some form of visual impairment ranging from mild to severe [1]. This group of people, which forms a significant fraction of the world population, is often limited to visiting only the places they are well familiar with, for instance, their own home, workplace and other places they frequently go to. A personal aide is needed when going to a less familiar place, including many public buildings such as a library, a bureau, a hospital, a courthouse, and many others. Another challenge is the fact that to find one´s way around at these public places can be a difficult task even for sighted people and more for those visually impaired. This paper describes a product that aims to bring a solution to navigating in buildings for the disadvantaged.

The presented product, Caneless, is an application for holographic glasses, more precisely, Microsoft HoloLens. Its core feature, navigation, can be easily used. A user puts glasses on and starts the application. After the start, a synthetic voice welcomes the user and gives them a choice of navigation. The user can browse through the 3D map of the building by having available all the names of rooms or sections they can visit in the building. When they choose their destination, the way is calculated and there are sound points placed along the path (*see Figure 1*). The user can just simply follow the sound and arrive at the given point at the building. The sound was chosen based on other published researches that explored navigation possibilities [2][3][4], taking into consideration some feedback from a blind 12-year-old potential user that participated as an early tester of this product. Another important information gained from the tester was the fact that while walking, there might be some unexpected or temporary obstacles in the way; if users were warned about these they could safely follow the path and feel more confident.

All Caneless' communication with the user is generated by a synthetic voice, which introduces user options and confirms the made choices. Users control the application by voice commands, which they are presented with when given the options, e.g., the synthetic voice announces the option of starting navigation and then presents the command by saying: "To start navigation, say 'navigate'!"

In order to be able to use the navigation, a 3D map has to be created first. The public space of a building is mapped within the technician mode which can be entered right at the beginning after the application launches. The technician, a sighted person responsible for creating the map, needs to walk through all the parts of the building that are to be on the map and scan the space. They can mark different places that will be listed as possible destinations for navigation. These marks, referred to as landmarks, are visualized after their creation as objects in front of the operator. The technician can use voice commands or hand gestures to add or name points and control the application. Hand gestures can be used in the user mode as well.

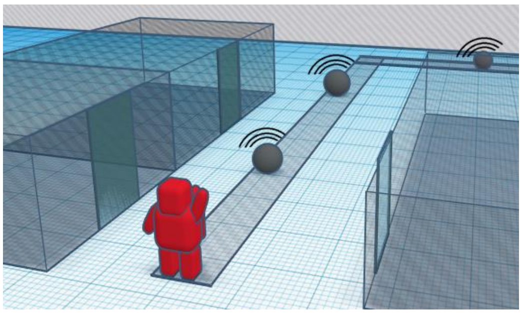

Figure 1 A model of sound navigation: the red figure is a user wearing HoloLens standing in a hallway and having a path generated in front of them. The black spheres represent sound points which user follows to the desired destination.

* Master degree study programme in field: Software Engineering and Artificial Intelligence
Supervisor: Martin Tamajka, Institute of Computer Engineering and Applied Informatics, Faculty of Informatics and Information Technologies STU in Bratislava

IIT.SRC 2019, Bratislava, April 17, 2019, pp. 35–36.

As aforementioned, there are two user roles: a regular user - supposedly a visually impaired person who needs navigation and a technician who administers 3D maps.
Caneless helps the regular user to:

- Define their destination using voice command and/or choosing from the offered list
- Start sound navigation

A technician can:

- Scan space and create a new 3D map
- Add landmarks with names – to denote possible destinations, e.g., rooms, caffé points, toilets, etc.
- Store scanned maps
- Add, remove, edit 3D maps on the web portal

The product itself consists of two main parts: a web portal, where users can access their 3D maps and edit them, and an application for holographic glasses, Microsoft HoloLens. These holographic glasses were used because of their suitable features, such as spatial sound, spatial mapping, and head-mounted cameras. The application for HoloLens devices is built in Unity with HoloLens toolkit using C# scripts, and the web portal is written in Angular using WebGL which encapsulates a unity application for work with 3D maps. The web application is also wrapped in a docker container. In order to provide map editing online, maps are stored at Azure blob storage.

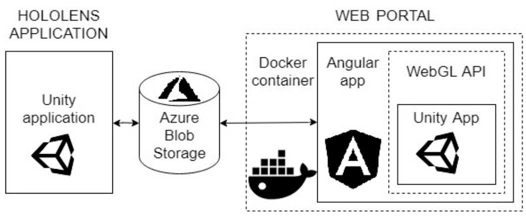

Figure 2 A top view of Caneless' architecture and used technologies. The HoloLens application is deployed on a device and directly communicates with Azure blob storage. The web application, communicating with Azure, wrapped in a docker container, uses Angular framework and includes a unity application that allows working with maps. All communication with the storage is via RESTful APIs.

Microsoft HoloLens brought plenty of usable features for this product as well as some challenges. One of the greatest challenges was that the head-mounted computer has very limited operating memory, and the navigation requires some calculations which are quite memory intensive, therefore the application had to be optimized to run on HoloLens. Another challenge was creating a 3D map in such a way that it can be stored and edited outside the device preserving the coordinates and all necessary metadata so that it corresponds with the reality.

There is a big potential for further development and application improvement, for instance, features as object recognition or text reading could be a great asset.

Caneless literally opens doors for visually impaired people to new, unexplored places. It offers confidence and independence when visiting such places, and thus mildens limitations they have to deal with every day. The main aim of this product is to create new opportunities by comfortable navigation for the blind and those with impaired vision.

Acknowledgement. This work was partially supported by the company Siemens Healthineers.

References

[1] Vision impairment and blindness, 2018. *Who.int* [online], [Accessed: 23- Feb- 2019].

[2] ZIJLSTRA, ARJEN TEIJO, 2017, Using the HoloLens' Spatial Sound System to aid the Visually Impaired when Navigating Indoors. Master's thesis. University of Groningen

[3] LIU, YANG, STILES, NOELLE R. B. and MEISTER, MARKUS, 2018, Augmented Reality Powers a Cognitive Prosthesis for the Blind. . 2018. DOI 10.1101/321265. Cold Spring Harbor Laboratory.

[4] BÅGLING, MAXIMILIAN, 2017, Navigating to real life objects in indoor environments using an Augmented Reality headset. Master's thesis. Umeå University.

ImageSearch: Deep Learning based Visual Search for E-commerce

Dominika DOLHÁ, Peter MALÍK, Tomáš POHL
Jozef SCHNEIDER, Andrej ŠČASNÝ, Dominik ŠTEVLÍK*

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
`2018tim2@gmail.com`

1 Introduction

On-line shopping has become more popular than ever. More and more people purchase items through e-commerce sites, and the e-shops expand the variety of products to attract more people. With the increasing amount of products, it has become challenging to retrieve only the relevant products for the customer.

Traditional text-based search engines might not be sufficient enough to retrieve the most relevant products and therefore, the e-shops might lose a potential customer. This is especially notable in the fashion domain, where it might be tough to search for a specific product with textual input, especially since a lot of fashion products lack a qualitative and informative description. Not only the products lack their description, but also the customers do not always have to know how to create a good textual query. This can be caused by customer's lack of information about the desired item, or simply his inability to phrase well a text input for the search engine.

One way to solve this problem is to use a content-based image retrieval, where a search engine analyzes the content of the images. For example, in the fashion domain the content might refer to color, shape, or texture of the clothing.

The main goal of our work is to make the searching process easier by enabling users to search for products by uploading an image of an item they are interested in. This image could be a celebrity wearing a certain dress, or simply just a photo of our well-clothed friend.

It is also worth to mention the age of social networks we live in, which results in an increasing number of so called influencers. These influencers often invoke in people a need to buy the same products as those seen in the photos on social networks.

2 Related work

In this section we examine existing approaches of image auto-cropping and visual search.

The approach for image cropping in Stentiford's work [2] is based on finding a window with the highest average pixel attention score in a saliency map of the desired image and still maintaining the aspect ratio.

Authors of Deep Learning based Large Scale Visual Recommendation and Search for E-Commerce [1] used a different approach unlike us and designed a large scale visual search by training a convolutional neural network. Their training data elements were represented by two different types of triplets of images. These triplets consisted of a query, positive and a negative image. Finally, they evaluated their network by humans rating the results of search queries on the Exact Street2Shop dataset.

3 Design and technologies

In our work, we propose a search engine API for managing and searching in a catalog of products. Our solution offers many types of actions over the products, like adding, modifying and removing it by our customer. Customers can also send us data about their users and view statistics about queries done in their e-shop. The data consist of statistics from their users about types of queries which have led to acquiring products. For example, if a user of e-shop uses a specific query and uses its result to add specific product to cart (and furthermore buys it), our customer should send us this information. We use this data for data mining to find patterns in user behaviour, and there-

* Master study programme in field: Software Engineering, Artificial Intelligence
Supervisor: Peter Gašpar, Institute of Informatics, Information Systems and Software Engineering, Faculty of Informatics and Information Technologies STU in Bratislava

IIT.SRC 2019, Bratislava, April 17, 2019, pp. 37–38.

fore improve and personalize the results for queries.

Users can use either traditional text input or our innovative image input. Moreover, our approach is also capable of combining both the inputs. If an image is used as an input from a user query, we allow users to manually crop it and therefore select the desired part of the image to be used in the search process.

The feature extraction process is handled by the ResNet50 model from the Tensorflow library, since it gave the best results on our clothing dataset. The extracted feature vector from the query image is used for finding the products, whose images resemble the uploaded image the most. The similarity is determined using the cosine similarity of these vectors. Finally, the search results can be adjusted by other filters and sorted by multiple sort options.

For feature extraction of products is used the following approach. First, each image of a given product is auto-cropped, since there is not always just the clothing in the picture. We utilize the Smartcrop library to find and rate the possible crops. In addition, we adjust the weights such that the crops focus primarily on clothing. Then, the feature vectors are extracted from each of the cropped images. Lastly, the final weighted feature vector is calculated for the product, giving different importance to each image based on their order, since usually the first-few images are the most representative ones. For this purpose, a logarithmic scale was used. The whole process of searching the top N most similar products is illustrated in Figure 1.

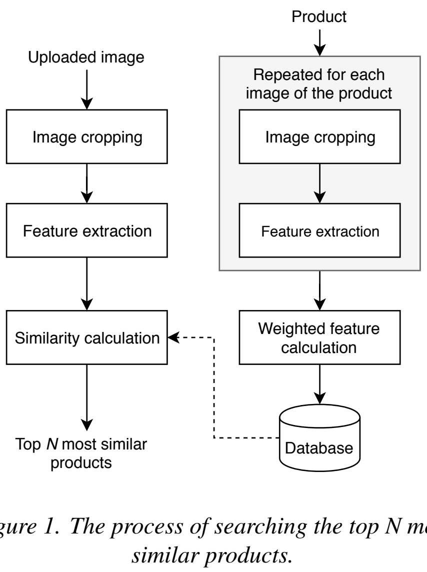

Figure 1. The process of searching the top N most similar products.

The application is built on top of the Django framework. For the text-based search Elasticsearch is used for its powerful full-text search functionality. The visual search primarily utilizes MongoDB.

Our unique value proposition is that we provide both image and text based search engine through an API, which is an innovative approach since e-shops still use only traditional text based search engines. We also plan to provide personalized searching, which should result in better outcomes.

Our plan is to provide our search engine in a range of prices with multiple monthly bundles varying in a number of queries available for use. After exceeding the limit of the bundle, additional queries may be charged separately. We also plan to offer our potential customers a one week free trial, so they would be able to try out our services. We expect our service to perform comparably fast as traditional search engines.

4 Conclusion

Our innovative solution provides a way to substitute traditional text-based search engines with content-based image retrieval. We retrieve feature vectors of each product image which are automatically cropped and then combine them to produce one feature vector for each product. We let users use both text and image based query. If they choose the image-based search, they can crop their uploaded images and then we extract a feature vector, which is then used to find the most similar products. Approach like ours has not been done before and we aspire to make shopping and searching for products easier and more pleasant for everyone.

Acknowledgement. This work was supported by grant No. 002STU-2-1/2018 supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic.

References

[1] Shankar, D., Narumanchi, S., Ananya, H.A., Kompalli, P., Chaudhury, K.: Deep Learning based Large Scale Visual Recommendation and Search for E-Commerce. *CoRR*, 2017, vol. abs/1703.02344.

[2] Stentiford, F.: Attention based auto image cropping. In: *Workshop on Computational Attention and Applications, ICVS*. Volume 1., Citeseer, 2007, pp. 253–261.

Webable: Web Browser for Visually Impaired People

Michal FABIŠ, Katarína RAFČÍKOVÁ, Daniela SITÁROVÁ, Andrej SLANINKA, Maroš VAŠŠ, Andrej ZAŤKO, Martin ŽÁK*

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
tp@webable.solutions

Introduction

Nowadays, it is almost impossible to imagine living without the internet. Mainly because we are using it to accomplish a lot of our daily tasks, ranging from reading the news and shopping to communicating with others.

Navigation throughout the information is much simpler for users without any impairment compared to the ones who are visually impaired. These people told us about their problems with web browsing. One of the main problems is that the website structure differs from site to site. Also, they are reliant on all senses but sight. They are using screen readers at most, although every website has its own unique structure and therefore it can be hard for a blind user to navigate within the website. Although, standards for writing a website accessible also for disabled people exist, some authors do not comply with them at all or respect very narrow subset.

As a result of findings mentioned above, we decided to create a web browser, which enables blind people to browse web more easily, even the websites that are not accessible out of the box. Our goal is to create a web browser that provides all of the functionality which is provided by all commonly used browsers like Chrome, Firefox, Microsoft Edge or Safari. The meaning of the word functionality in this context is that a user can use keyboard shortcuts, open and work with more tabs, show history of browsing or customize the browser. Our solution is enriched by special features, which make web content more accessible.

Solution description

Our application is designed for users of each common operating system - Windows, MacOS or Linux. The core of the browser is based on the Electron, running on Node.js, framework for building cross-platform desktop applications. Browser itself is a web application which is rendered using Chromium[1], open-source browse engine, which makes up a core of the Electron. Browser itself is written in framework Angular. Angular is effective when it comes to working with abstraction of the DOM.

As in Figure 1, visiting web pages seems similar to any browser. But there is a very important intermediate step, thanks to which our browser is unique and displayed websites are more accessible. It is automatic correction of inaccessible elements in HTML document, which is described in more detail below.

Validation and correction of inaccessible elements in HTML

The general idea of making web pages more accessible is to find elements not following accessibility guidelines and correct them. It is possible to find many tools on the web which are automatically analyzing and evaluating web content considering WCAG standards[2]. In our web browser we used analyzer named Pa11y[3], which is compatible with used technologies and fulfills requirements.

First step in our solution is validation of visited web page by this tool. It is configured to find only some types of accessibility mistakes, for example form fields without label or images with no alternative text. The result of validation is a list of issues, which need to be corrected. Every issue consists of code of mistake, some description and the most important is selector of inaccessible element. This list is used as input data into our correction method. This method is working actually on incorrect form fields, but our goal is to extend it on more types of mistakes and make this algorithm more general.

In the next step we iterate through list of issues which are corrected by algorithm. Each element is

* Master degree study programme in field: Intelligent software systems
Supervisor: Dr. Jakub Šimko, Institute of Informatics, Information Systems and Software Engineering, Faculty of Informatics and Information Technologies STU in Bratislava

[1] https://www.chromium.org/Home

[2] https://www.w3.org/TR/WCAG21/

[3] https://github.com/pa11y/pa11y

IIT.SRC 2019, Bratislava, April 17, 2019, pp. 39–40.

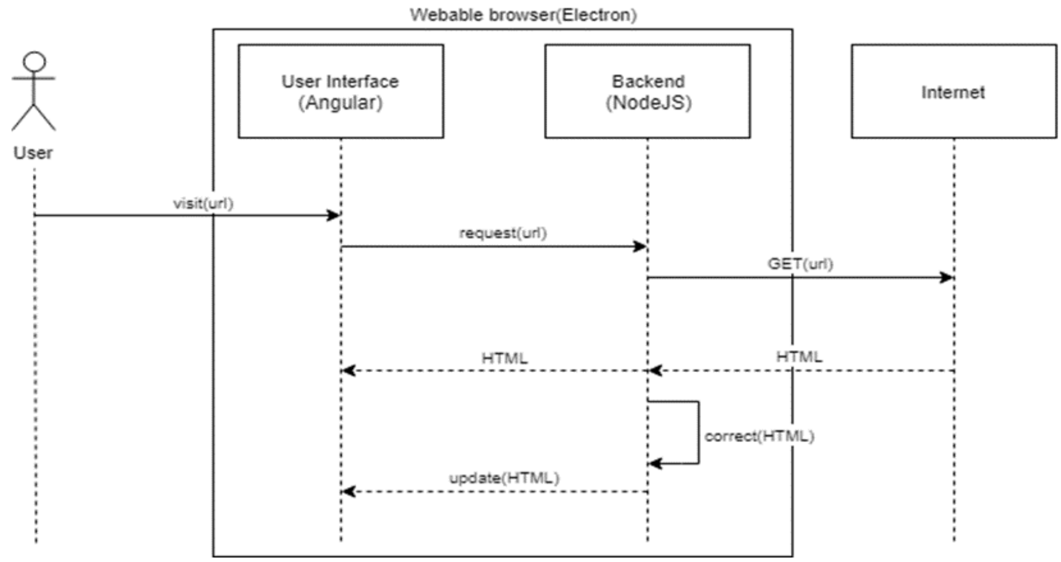

Figure 1. Flow of visiting webpage.

found in DOM structure using selector from issue. After then is looked for the nearest text to this element in DOM structure. This text is then assigned as label to concrete element.

Evaluation

Our idea and later prototype validation with visually impaired people were necessary from the very beginning. At first, we got in touch with several blind people, who helped to develop our idea and gave us first feedback as well. Later, we found organization associating all visually impaired people in Slovakia thanks to which we could test our prototype with many users.

Our browser has been tested by 5 visually impaired people since we released the prototype. The test scenario consists of 2 parts and each has 5 tasks. First part is about to get familiar with the browser environment and test accessibility of the browser. Second part consists of several more complex tasks, which helped us to understand better how visually impaired people interact with a web browser. This part of testing is designed also to imitate everyday usage of web browser, so the tasks are use cases inspired by scenarios described by visually impaired people.

The result of every testing session has been useful for next development of our browser and we really appreciate feedback from visually impaired.

Future work

In near future, we will focus on improving of code correction algorithm as we see it is really helpful. Later we would like to make the navigation of the user faster by the sitemap. Not the regular sitemap but intended for user to decrease number of steps needed to find an information. In general, this also increases speed of browsing throughout the website. Next feature will be intent. It is something the user is looking for or wants to do. In this manner the solution will offer a search within a site that looks for a typed in intent. If there will be multiple actions and information fulfilling the criteria, they will be listed, and user will be able to go through them one by one. If just one fulfills them, the browser opens subpage that corresponds with this search.

Conclusion

The main aim of this project is to create a complex solution which allows blind people to browse websites, which are not accessible when rendered in conventional web browsers. Webable focuses mainly on elimination of defects caused by weak using of WCAG standards. Our next step is to test our solution a lot to make our browser user friendly for visually impaired people and to improve algorithm of code correction. Secondly, we will focus on developing algorithm for creation of sitemap for each domain, which enhance speed during navigation on website.

Acknowledgement. This work was supported by grant No. 002STU-2-1/2018 supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic.

Intelligent Importer of Public Data Sets

Filip VARGA[†], Milan VAŠKO[†], Jana VRABĽOVÁ[*]
Ladislav BARI[†], Adam TALIAN[†], Adam ŠEVČÍK[†]

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
tp.tim16@gmail.com

Public data sets generated on a daily basis by government offices and institutions contain data of varying importance and quality in different formats. This data may hold useful information, collectible from it using methods of machine learning and data mining algorithms. It can also manifest various relationships between its elements or entities. So far this data has mostly been stored in formats used or processed by manual editors, thus making it not machine readable.

The role and purpose of the Intelligent Importer of Public Data Sets is to provide a platform of services, among which the most important is the ability to upload and transform data. The uploaded data is mainly stored in formats such as CSV files or Microsoft Excel workbooks. The desired format recommended by the government is RDF (Resource Description Framework), which is machine readable and exposes the contents of the original files for public use.

1 Use case

The Importer is a web application that can be opened in a browser window, in which users can access its services. As an example, a typical workflow of the application's usage is described below (and can be seen on Figure 1 on the next page):

The user uploads data on a designated web page of the application in his browser using an upload button. The system stores the data on the server temporarily for editing and displays the preview of the data in the browser. The application prompts the user to select column names of the data based on entities located in the Central model (a standardized model approved by the government), thus confirming the credibility of the uploaded data.

Once the user is done editing the data, he confirms his changes. The system then inserts the edited data into a database, where it will be stored, as well as any changes done to it in the future. Afterwards the user can access the data from the application by simply clicking a download button on the corresponding link.

In that case, the data will be pulled from the database and based on the desired format of download, it can be downloaded either as the original file that will recreated from the retrieved data, or the recommended machine readable RDF format, in which the used data entities are exposed and linked to the Central model through references.

2 Used technologies

In order for the application to execute the abovementioned use case and provide other services for the data processing, it uses various technologies and services specifically suited for this role.

The Central Model is represented by ontologies in OWL (Web Ontology Language) and uses XML presentation syntax to define it. It is provided on government servers so that public data sets an reference entities and attributes contained within the ontologies. The application uses an external service - Elasticsearch - to retrieve data from the Central Model.

The application's notable feature is the ability to generate RDF files containing input data from provided source files, which come in different formats. Thus, through uniting the data format, it makes the data very well machine readable and interchangeable on the web. RDF (Resource Description Framework) is a model designed specifically for data interchange on web and is used to reference the Central Model through contained URIs and literals. Through this, it can display relationships, describe data types and

[*] Master study programme in field: Computer Engineering
[†] Master study programme in field: Software Engineering, Artificial Intelligence
Supervisor: Assoc. prof. Jakub Simko, Institute of Informatics, Information Systems and Software Engineering, Faculty of Informatics and Information Technologies STU in Bratislava

IIT.SRC 2019, Bratislava, April 17, 2019, pp. 41–42.

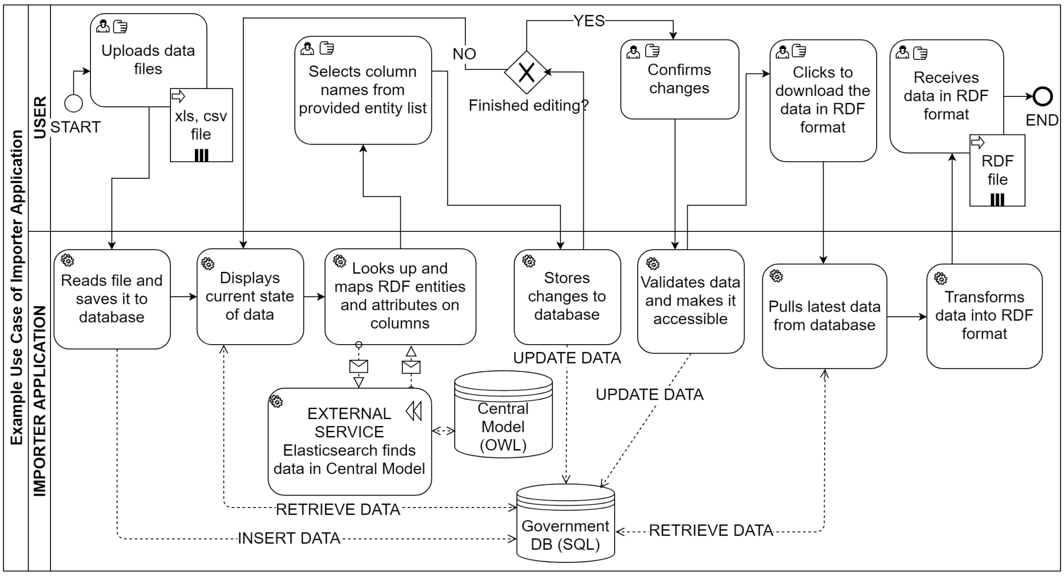

Figure 1. Example Use Case of the Importer application, which involves uploading a file, processing its data, storing it in a database and transforming it into an RDF file.

identify the entities and allows checking the credibility of the contained data against the Central Model.

All inserted data is, in the meantime, stored in an SQL database, along with the information about all changes done to the data. This means that the user can access multiple versions of the same file and see modifications done to the file. He is provided with an identifying token which enables him to access his files in the application. The user can share the token (by sending it embedded in an URL) with other users so they can easily access and modify the data themselves (a support for multi-user work use case).

Turning the attention to the design of the application, it was designed according to the Unified Design Manual for Government Web Services [1] and so can be deployed as a government web service once undergone thorough acceptance testing.

As stated in the project's title ”intelligent”, the importer uses methods of artificial intelligence to make data processing more automated. For example, the application uses knowledge discovery to provide data type detection. We can see in the Use Case on Figure 1, that the user has to select which entities from the Central Model are in his uploaded file. The application helps him complete this task by using a trained model, which can determine the entities based on the data and its features. It also provides the suggestions listed from more likely to less likely.

The application also provides an auto-completion service using Elasticsearch, when the user has to manually input the data entity, which provides entity suggestions based on typed letters or keywords. There is potential for more future work in this area though. We can improve the trained model's accuracy by collecting feedback on its suggestions by comparing them to the actually selected values.

3 Conclusion

The Intelligent Importer of Public Data Sets is a web application which can greatly improve the accessibility of public data, unify its format and make the process of uploading public data sets simple and able to detect potential data corruptions and faults in the uploaded files. The application's current production instance at the time is already deployed online and can be accessed on address 35.237.0.28. Should it be approved by the government for usage, it can contribute to the process of transforming publicly accessible data sets into machine readable open data, and increase data transparency of state offices.

Acknowledgement. This work was supported by grant No. 002STU-2-1/2018 supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic. We thank the Office of the Deputy Minister of the Slovak Republic for Investments and Informatization for supporting this project.

References

[1] Unified Design Manual for Government Web Services. `https://idsk-elements.herokuapp.com`.

MonAnt - Monitoring Online Antisocial Behaviour

Milan CÁK, David TRAN DUC, Dung LAM TUAN, Matúš BARABÁS
Peter LIPTÁK, Veronika ŽATKOVÁ, Patrik ŽIDULIAK*

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
fiit.tp14@gmail.com

Millions of people all over the world using the Internet everyday are, whether being aware of it or not, surrounded by antisocial behaviour. One of the most serious types of such behaviour is especially misinformation and false information spreading [1].

The huge number of online social networks and news portals has drastically changed the way of communications and information sharing. Being a part of this cyber space and knowing how to consume the right information is the most difficult part of all. The advent of social networks makes every user a source of information without checking for any facts. Because of that, the misinformation spreads very fast and everyone can access this information without any difficulty. As the usage of internet increases, the chance to spread misinformation and disinformation also increases many fold. The spread of misinformation has been revealed in many topics such as health, politics, finance and technology. Online environment provides a very good opportunity to misuse it.

Antisocial behaviour in online environment is one of the most recent and serious problems. This topic is researched by a permanently increasing body of research. In many publications, the three main crucial steps of regulation and elimination of antisocial behaviour are identified [2]:

- Characterization: in this step, we try to characterize antisocial behaviour by analyzing its characteristics.
- Detection: the goal of this step is to automatically or semi-automatically detect a antisocial behaviour.
- Elimination: we try to reduce the antisocial behaviour in internet. Many approaches are used, e.g. banning and filtering antisocial content.

As mentioned above about crucial step of regulation, we identified the important step that must be done before all is, how we can collect data for characterization, detection and elimination of antisocial behavior. Our project focuses on collecting antisocial behaviour data with purpose for further research and studies in the monitoring, detection and mitigation of antisocial behaviour area. The main goals of the project are:

- Having as much as possible data in one place, where we can easily apply text analysis and machine learning algorithms to identify the correctness and trustworthiness of data.
- Creating an application used to effectively manage sources of this data. Our goal is to collect as much data as possible, but we do not try to collect all data that is shared in Internet. Having a management application, we can focus in specific fields and have data well organized.
- Managing storage, which is used to store all collected data. We expect external sources, which can push their collected data to our storage. Therefore we develop API, which will be primary interface to the storage. Users, applications with correct authentication and authorization rights can make requests to the storage.

To achieve our goals, we propose a platform named MonAnt. Architecture of our project is shown in figure 1. While designing our architecture, we focus on availability, robustness and scalability of our system. Our system consists of three main modules:

- Web management: Standalone web application for web monitoring management. Users can predefine rules by setting web pages, which information will be retrieved from. The application is being developed in Python3 with usage of Django framework. The main reason, we chose Django is its portability and simplicity of Python as pro-

* Master study programme in field: Information Systems
Supervisor: Dr. Ivan Srba, Institute of Informatics, Information Systems and Software Engineering, Faculty of Informatics and Information Technologies STU in Bratislava

gramming language.

- Web monitoring: Implementations of data providers for specific topics. We use Python3 as programming language for our implementation. To simplify implementations we use Scrapy library for web crawling, BeautifulSoup for scraping HTML files and feedparser for parsing RSS feeds.
- Central storage: API layer for storage of collected data. This plays crucial role in our system. It also mediates communication and data transfer between other platform modules. As database, we choose to use Postgres, one of the most used relational databases currently.

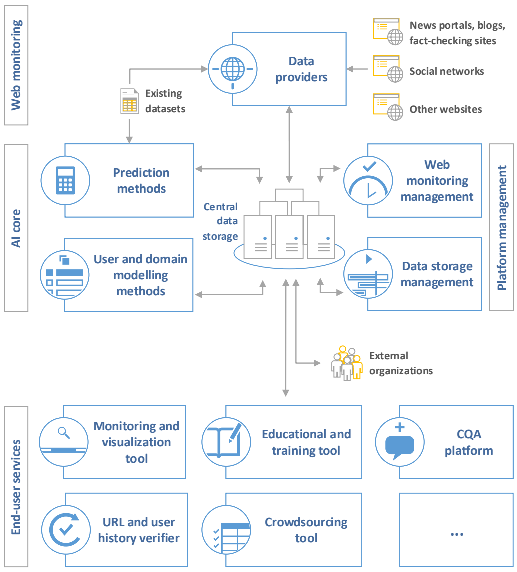

Figure 1. Project architecture [3].

By integration of all three modules, we have scalable system, where we can define rules for monitoring of information spread. Collected datasets are then well organized, easily to use and available for all potential researchers.

The trigger of system is a user, who wants to monitor some sources of information, e.i news portals. The user defines the list of sources, which he wants to monitor the spread of information on. This action activates data providers, which start crawling data from predefined web sources. The collected data are then sent by REST API to storage and be stored by database schemes. With stored data, we provide them to people, who want to analyze them to identify the trustworthiness.

Our platform provides a possibility to be extended by AI core modules. AI core modules are components that ensure handling, scheduling etc. on methods over data from central storage that are collected by our system. It will support whole machine learning flow, search and natural language processing. They can be added in anytime, so many AI algorithms can use our datasets at one moment.

In the final version of system, we want to integrate whole system with Docker[1]. By doing so, from the technological point of view, we can painless manage resources, performance, some aspects of security of the whole system.

This platform is proposed and will be implemented in projects Rebelion and Misdeed. Rebelion is research project that focuses on Automatic Recognition of Antisocial Behaviour in Online Communities. Misdeed is research project that focuses on Misinformation Detection in Healthcare Domain.

Conclusion

Antisocial behaviour in online world is one of the most recent and serious problems. We propose the platform MonAnt for monitoring, detection and mitigation of antisocial behaviour.

Acknowledgement. This work was supported by grant No. 002STU-2-1/2018 supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic.

References

[1] Fortuna, P., Nunes, S.: A Survey on Automatic Detection of Hate Speech in Text. *J. ACM*, 2018.

[2] Kumar, S., Shah, N.: False Information on Web and Social Media:A Survey, CRC press., 2018.

[3] Srba, I., et al.: Monant: Universal and Extensible Platform for Monitoring, Detection and Mitigation of Antisocial Behavior, 2019.

[1] provides a uniform interface for application isolation to containers

TrafficWatch: Road Traffic Monitoring Solution for Smart Cities

Matej GROMA*, Matej HORVÁTH*, Peter JURKÁČEK†, Jozef KAMENSKÝ†, Adam KŇAZE†, Kristína MACKOVÁ†, Lenka PEJCHALOVÁ†, Jakub SEDLÁŘ†

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
tim21.2018.fiit@gmail.com

For road traffic management and planning, a detailed information about what is happening on the road network is needed. It is crucial to know how many vehicles are using it, at what time of the day, and at least a rough estimate of where from and where to are they heading. Here are few examples of urban development decisions and needed information:

– Heavily used street in the city centre is going to be closed due to construction works. An alternate route has to be found to prevent a traffic collapse. Prerequisities are analyses of how many vehicles actually need to be diverted, what are the capacities of possible detours, and what are their current traffic levels.
– One of the intersections in the city is congested. It may be rebuilt and expanded, first however, the cause of the congestion needs to be identified. Therefore, it is important to know what are currently the most heavily used areas of the intersection and where do cars spend the most time waiting.
– Routes and timetables for public transportation are created in order to reduce car usage. The starting point is knowledge of how many cars are transiting to particular borough at what time of the day.

All of the strategic decisions described above require data collection and analyses. Our aim is to create a system able to provide knowledge for a variety of road-traffic management&planning use-cases like these. Furthermore, our system aims to be easily deployable, maintainable, scalable and centered on insightful, user-friendly outputs.

Our solution consists of two main parts - a network of smart cameras observing several points of the road network, and a server gathering recorded data and serving it in a meaningful form to the end-user. There are several types of information which can be analyzed in our system.

First one are *transits*. Users are able to define several zones in the camera view. When a vehicle passes the view, we track its movement and log all the zones it passed and times of passage. For example on an intersection, users may mark all points of entry and exit. Our system then provides them not only with information about how many vehicles at what time entered the intersection from what direction, but also where they were heading. This gives them a general sense of traffic flow and its trends (e.g. at 15:00 there is a big spike in traffic coming from industrial zone going to nearby village).

We are working on further extending transit information in several useful ways. For example, for each vehicle a *type classification* will be performed (to the predefined vehicle types, e.g. car, van, bus, truck), so we can provide more detailed transit analyses per individual vehicles types. This will open up a whole new class of knowledge to be utilized.

We record also an average time of passage between zones. This is useful, because pure traffic volume itself does not help in identifying the traffic congestion. If traffic is congested number of passing vehicles is low (they move slowly), which could be falsely interpreted just as a low amount of vehicles with no congestion.

---

* Master study programme in field: Computer Engineering
† Master study programme in field: Artificial Intelligence, Software Engineering
Supervisor: Dr. Ivan Srba, Institute of Informatics, Information Systems and Software Engineering, Faculty of Informatics and Information Technologies STU in Bratislava

IIT.SRC 2019, Bratislava, April 17, 2019, pp. 45–46.

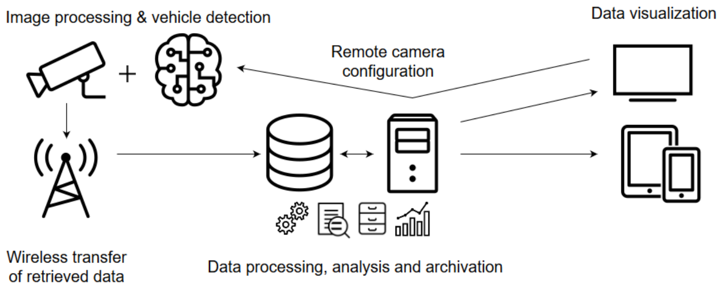

Figure 1. An overall architecture of the proposed system.

Transit events mentioned until now ignore *precise trajectories* of vehicles while moving between zones. However, we do record and evaluate these as well opening yet another range of possibilities. It gives us more information about the quality of design of observed road. We record and display:

– heatmap showing volume of passing vehicles
– precise trajectories of passing vehicles
– heatmap showing stationary traffic (vehicles stopped/waiting)
– heatmap showing relative speeds of vehicles

This kind of information shows general driving behavior of passing vehicles and may possibly identify problematic or dangerous areas.

Cherry on the top of this all is the fact, that all of the above analytics are constantly updated and may be used for *real-time monitoring*.

Easy maintainability and deployability of our system mentioned before is achieved by remote camera configuration via simple admin interface. It allows both analytics setup (e.g. transit zones definition) and detection algorithm tuning with live feedback (parameter tweaking with real-time stream from camera to observe effects of changes).

All the knowledge our system delivers is obtained by the methods of a computer vision. Basis of our detection algorithm is the background substraction. However, to improve robustness and accuracy we extended the detection process by various means, including keypoint tracking, optical flow calculation and movement prediction.

There are multiple alternatives in the domain of traffic analytics. To begin with, there is a plethora of vehicle counting products, ranging from simple human-based counting with pencil and paper (which is still the most common solution in our region), through mechanical devices physically installed on the road, to computer vision based systems like ours. Other group of products do trajectory centered analysis, or speed detection (usually using expensive equipment with radar). Our solution has the advantages of providing full package plus some more (our approach to transits is much more insightful than pure vehicle counting), as well as a refined computer vision algorithm.

First part of our solution (see overview in the Figure 1). is the network of smart cameras. As does word “smart” indicate, these cameras perform much more than just record a video. All the computer vision, vehicle detection and event generation happens here. This arrangement is called edge computing and it is opposite to traditional way of streaming video feed to server that does all the heavy-lifting. It greatly reduces amount of data sent to server plus it scales really well.

Main programming language of the camera module is currently Python (with plans to move to Cython or C++ in the future). We use the OpenCV library. All this is running on Nvidia’s Jetson TX2, one of the few edge devices allowing us to leverage graphics card acceleration.

The second part of our architecture is the server which uses Java (Spring) for backend, JavaScript (React) for frontend and Postgres (TimescaleDB) as database. Server communicates with cameras via the mqtt, handled by pahoMQTT library. Also, to enable real-time communication (e.g. augmented video live-stream) we use the WebRTC technology.

In the future we plan to take it one step further and perform vehicle re-identification. It means multiple cameras on different locations with ability to identify that a particular car had previously appeared on another camera. We plan to do this without reading vehicle registration plates because it requires expensive or specially positioned cameras. Our solution would then be the first and only one with this functionality.

Acknowledgement. This work is enrolled in the project DA-SPACE, a part of Danube Transnational Programme. It was conducted in collaboration with partners Unicorn Systems Slovakia and Orange Slovakia. The work was also supported by grant No. 002STU-2-1/2018 supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic.

TextMania: Intelligent Text Analytics

Dávid CSOMOR*, Adam ĎURIŠ*, Alan KOVÁČ*, Daniel KOVÁČ*, Peter KRIŽAN*, Patrik MELICHERÍK*, Krištof ORLOVSKÝ*

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
team5fiit@gmail.com

Nowadays, there is a huge amount of digitized written information, which surrounds us and is growing every minute. Our inability and lack of time to carefully read and organize every new article and categorize it, requires a tool for automated text analysis and tag recommendation. This tool should be able to import new texts, categorize them based on provided tags train the machine learning model and allow future texts to be categorized based on that pre-trained model. There are already several tools like this, but none of them is useful for Slovak language.

With this purpose in mind, we are developing a web-based environment with Node.js backend for analysis of text documents in Slovak language. We use technologies and advancements from the field of machine learning and automated text processing to achieve automated tagging and categorization.

In this paper, we focus on importing, analysing and automated categorization of new articles based on their content. Using our solution, we can easily import and categorize new articles based on our pretrained machine learning model.

Our proposed text analysis environment is designed with intent to be user friendly above all, so it can be used not only by data scientists, but also by regular users (without programming or scientific background). This should lead to improvements of our machine learning model, because we should have more input data. It is very important for us to make the process as easy and accessible as possible, so users can help us grow our datasets and improve our algorithms.

Our environment is focused on working with large collections of texts in Slovak language (e.g. various news articles, encyclopedia articles or any other text). For initial tracking, we firstly obtained articles from Slovak instance of Wikipedia[1] and Slovak popular news site Webnoviny[2]. Then we tagged them with keywords from the source page. Texts of these articles are stored in MongoDB, which is a flexible and scalable NoSQL database useful for textual data.

The first text processing task in our pipeline is word-by-word analysis. For this we use a third-party web-application NLP4SK[3], which is a set of tools for natural language processing. NLP4SK provides tools for text tokenization, sentence identification, lemmatization, part-of-speech (POS) tagging, named entity recognition (NER) and some other features.

After processing, all data are available for browsing and visualization in our web-application. Our main use case of text analysis is:

1. Import article(s) from several sources (*.txt or URL).
2. Store article(s) in MongoDB database.
3. Obtain basic information about imported text using the NLP4SK.
4. Insert analysis results into database.
5. Create and update inverted index.
6. Retrain machine learning model used for automatic categorization.
7. All articles and its attributes are available for browsing in our web-application.

Articles added during the development were manually tagged by our team, which provided a starting point for our machine learning process. We analyzed the inserted text and stored the retrieved data such as grammatical categories (*POS* – part-of-speech), identified entities (*NER* – named-entity-recognition) such as location and person, lemmas and N-Grams.

Our solution is based on two main parts. Figure 1 shows an overview of the system architecture. Client side of the application is built on

* Master degree study programme in field: Software Engineering
Supervisor: Dr. Miroslav Blšták, Institute of Informatics, Information Systems and Software Engineering, Faculty of Informatics and Information Technologies STU in Bratislava

[1] https://sk.wikipedia.org

[2] https://www.webnoviny.sk

[3] Link to NLP4SK: http://arl6.library.sk/nlp4sk/

IIT.SRC 2019, Bratislava, April 17, 2019, pp. 47–48.

Angular[4] framework, which provides flexibility for the frontend. User interface currently allows user to:

- Add new articles from supported sources.
- Show visualized statistical data based on stored articles.
- Search and browse articles based on specified criteria.

Inverted index section, shown on Figure 2, allows user to search our database based on:

- Category name
- Selected words
- Article names
- Part of speech specification
- Named entity specification
- Word's lemma

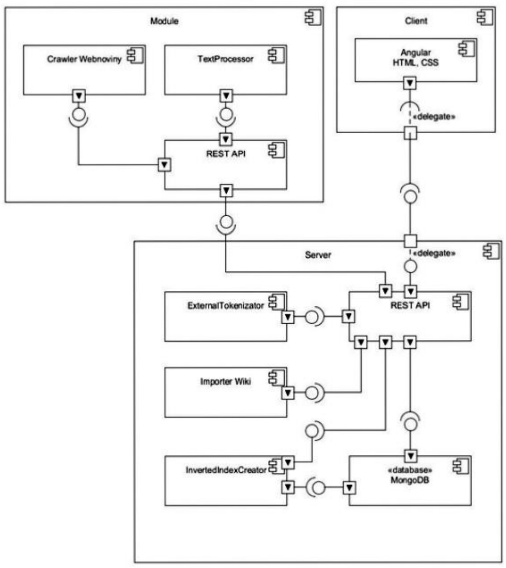

Figure 1. System architecture.

Server side of the application is built on Express[5] framework and running in Node.js environment. This component provides database access and routing. In this part are implemented "Wikipedia importing", "inverted index creation" and all other features mandatory for the client side.

There is also component powered by Django[6] framework which is the backbone of our machine learning processes. In the Python module, there is also implemented "articles importing from "Webnoviny".

All these components are communicating with each other through the HTTP and are independent. This allows you to deploy the Python component to another server and thus facilitate processing on the JavaScript component (the client-side searching through articles will be still smooth).

Final product will allow users to add new articles to

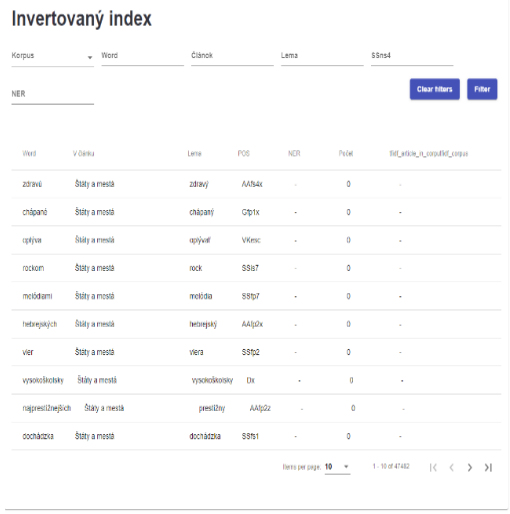

Figure 2. Application preview – inverted index view.

our solution through our web page interface, where it will be processed, tagged and stored for later analysis by the user. Regular users could use it for categorization and researchers can use it for data analysis purposes, since they will have detailed information about this process and so they should be able improve their algorithms.

Acknowledgement. This work was supported by grant No. 002STU-2-1/2018 supported by the Ministry of Education, Science, Research and Sport of the Slovak Republic.

[4] Link to angular framework: https://angular.io

[5] Link to express framework: https://expressjs.com

[6] Link to Django: https://www.djangoproject.com

WiFi Funtoro

František BEKEŠ*, Jakub HUBERT*, Marek OTRUBA*, Samuel GEDERA*,
Nikolas VIROSTEK*, Radovan KOHÚT*, Patrik BRANDÝS*

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
`Funtoro2018@fiit.stuba.sk`

Over the past few years, travelling has become very popular and electronic devices became thinner, lighter and with bigger screens. These changes have a big impact on bus transportation, trains and also airlines. Almost every traveller enjoys watching a movie, listening to music or playing a simple game on the screen located in the seat. But, on the other hand, every full-size bus has at least 60 tablets which need to be maintained on a daily basis, and this can be very expensive for every transportation company. Not to mention, technology is improving so fast, that upgrading tablets in every bus every 5 to 7 years is quite a big investment.

Since now we have big screen devices such as smartphones, tablets and also laptops, companies can benefit from using WiFi technology to distribute digital content to all these devices for a fraction of the previous costs. This change can dramatically lift the potential of the entertainment system in every transportation vehicle.

Our potential WiFi Media on Demand (*MOD*) benefits are (innovations):

- Removed seat tablets – no maintenance
- Only vehicle server maintenance
- Updates / changes on server side only
- More complex MOD – adaptive to new technology and trends
- Scalable – easy to add new features

Known risks:

- Digital Rights Management – media leaks
- Responsive design implementation – needs to support all devices (*screen size, browser*), including different operating systems (*Android, iOS, Windows, OS X)*
- Vehicle WiFi network stability

Media on Demand is a new generation of Video on Demand. We can watch or listen video and audio content like TV shows or movies and we can also see real time information about the current route along with advertising system. This technology was used for the first time by FUNTORO [1], using monitors with MOD system embedded in seatback or armrest. This is a good opportunity for improvements. In the past, there was more projects which were dealing with the MOD system, but with other features. In the Europe, FUNTORO is represented by company Molpir s.r.o. which is cooperating with our team.

In our MOD system we are offering many functions, such as:

- Watching movies
- Listening to music
- Streaming TV
- News
- Reading E-Books
- Sightseeing
- Actual location of the bus
- Live bus cameras

Sightseeing is a specific functionality of the system which stands out from the other functions, because it offers many useful information about the route. The main objective is to inform the passenger about Point of interests (*POI*) along the route, such as castles, historical points and many others. The system notifies the passenger about upcoming POI and it will be upon the passenger to show or ignore the POI information window. This feature will be especially interesting for companies which offer sightseeing tours, but also for regular travel companies, to entertain passengers during journey and to improve customer satisfaction. Sightseeing system works not only inside the bus on local bus network, but also outside the bus, where passengers can connect to the server using mobile internet in their devices and continue using sightseeing feature while walking around the city and seeing individual points of interest.

* Master degree study programme in field: Software Engineering & Applied Informatics
Supervisor: Dr. Peter Pištek, Institute of Computer Engineering and Applied Informatics, Faculty of Informatics and Information Technologies STU in Bratislava

Because companies which provide movies and other multimedia content are sensitive to misuse by third parties, it's important to think about Digital rights management, especially in this type of system, where media are served directly to the customer's device. Nowadays, there are several approaches to protect multimedia content from abuse (Visible watermark, Forensic watermark, Digital fingerprint). Using a specific method depends on many aspects [2]. Some of them can not be applied to all multimedia contents, others are too sophisticated for bus conditions. With our implementation of DRM, we will prevent the unwanted leakage of medial content. It is obvious, that it is not possible to completely prevent leakage, but we also provide mechanism to identify source of leakage, in case it will happen.

Bus travellers have to be able to use our MOD system with their own devices that are connected to bus local network through WiFi. To cover most of the devices and fight against platform specific issues, we have decided to implement WiFi Funtoro system as a web application. The application uses client-server architecture. Client part is responsible for displaying media content and other features mentioned above. We have also decided to implement it as a rich client web application and it is based on TypeScript framework in Angular. There is very high possibility that the application will be extended in the future with more features. Angular is very easily extendable by clear component-oriented structure.

Frontend side is based on Angular 6. We use bootstrap 4.1.3., which contains HTML and CSS-based design templates to style web application. We also use Font Awesome toolkit, which is based on CSS, to make fonts and icons more beautiful. Application is fully responsive for all devices.
Server side of the application is based on Node.js library Express. Its main responsibilities are:

- persisting MOD data
- webserver - serving files built by Angular
- transcoding
- REST API
- logging

We have decided to go use Node.js because it is easily scalable and has good support for web sockets. Client components communicate with server part of application through REST API. In our application we use sequelize.js ORM, because it provides easy access to our PostgreSQL database by mapping database entries to objects and vice versa. It also helps us to create database structure by specifying the model structure. It significantly saves our time and effort while creating database structure because we do not have to write SQL queries. Other advantages are also better readability, consistency, automatization or ability to connect with different databases, thus it provides easy switch from one database to another. This could be useful for future development of our application. One of the disadvantages of using sequelize.js is that it can be slower when creating complex data models. However, we do not assume such very complex models for our application.

During winter semester, we have implemented data model for movies and routes for POIs. We had to consider Molpir's structure of storage files in their file system. Structure was specified by XML file provided by Molpir. We had to parse the XML file to get file paths leading to files in storage. To accomplish this, we have used DOMParser.js which provides the ability to parse XML or HTML source code from a string into a DOM document.

Overall, the application architecture and used technologies are chosen wisely and we have spent many hours to figure out the best technology for such project and the vehicle environment, which is very uncomfortable for any computer based devices. Dust, rapid change of temperatures, vibrations are definitely the worst conditions for computers. It was important to properly choose technologies which do not require very powerful servers and a lot of computation memory to server all travellers in the vehicle. Despite all these negative reasons, WiFi funtoro has a big advantage for all travel companies and it could reduce the costs and offer more entertainment, which can be extended very easily.

References

[1] Yang, C. M. (2014). U.S. Patent No. 8,726,155. Washington, DC: U.S. Patent and Trademark Office.

[2] Ma, Zhaofeng. *"Digital rights management: model, technology and application."* China Communications 14.6 (2017): 156-167.