

INFORMACIJSKA DRUŽBA

Zbornik 26. mednarodne multikonference

Zvezek A

INFORMATION SOCIETY

Proceedings of the 26th International Multiconference

Volume A

**Slovenska konferenca o
umetni inteligenci**

**Slovenian Conference on
Artificial Intelligence**

Uredniki • Editors:

Mitja Luštrek, Matjaž Gams, Rok Piltaver

12. oktober 2023 | Ljubljana, Slovenija • 12 October 2023 | Ljubljana, Slovenia

IS2023

Zbornik 26. mednarodne multikonference
INFORMACIJSKA DRUŽBA – IS 2023
Zvezek A

Proceedings of the 26th International Multiconference
INFORMATION SOCIETY – IS 2023
Volume A

Slovenska konferenca o umetni inteligenci
Slovenian Conference on Artificial Intelligence

Uredniki / Editors

Mitja Luštrek, Matjaž Gams, Rok Piltaver

<http://is.ijs.si>

12. oktober 2023 / 12 October 2023
Ljubljana, Slovenia

Uredniki:

Mitja Luštrek
Odsek za inteligentne sisteme
Institut »Jožef Stefan«, Ljubljana

Matjaž Gams
Odsek za inteligentne sisteme
Institut »Jožef Stefan«, Ljubljana

Rok Piltaver
Outfit7, Ljubljana

Založnik: Institut »Jožef Stefan«, Ljubljana
Priprava zbornika: Mitja Lasič, Vesna Lasič, Mateja Mavrič
Oblikovanje naslovnice: Vesna Lasič

Dostop do e-publikacije:
<http://library.ijs.si/Stacks/Proceedings/InformationSociety>

Ljubljana, oktober 2023

Informacijska družba
ISSN 2630-371X

Kataložni zapis o publikaciji (CIP) pripravili v Narodni in univerzitetni knjižnici v Ljubljani
[COBISS.SI-ID 170717955](#)
ISBN 978-961-264-275-4 (PDF)

PREDGOVOR MULTIKONFERENCI INFORMACIJSKA DRUŽBA 2023

Šestindvajseta multikonferenca Informacijska družba se odvija v obdobju izjemnega razvoja za umetno inteligenco, računalništvo in informatiko, za celotno informacijsko družbo. Generativna umetna inteligenca je s programi kot ChatGPT dosegla izjemen napredek na poti k superinteligenci, k singularnosti in razcvetu človeške civilizacije. Uresničujejo se napovedi strokovnjakov, da bodo omenjena področja ključna za obstoj in razvoj človeštva, zato moramo pozornost usmeriti na njih, jih hitro uvesti v osnovno in srednje šolstvo in vsakdan posameznika in skupnosti.

Po drugi strani se poleg lažnih novic pojavljajo tudi lažne enciklopedije, lažne znanosti ter »ploščate Zemlje«, nadaljuje se zapostavljanje znanstvenih spoznanj, metod, zmanjševanje človekovih pravic in družbenih vrednot. Na vseh nas je, da izzive današnjice primerno obravnavamo, predvsem pa pomagamo pri uvajanju znanstvenih spoznanj in razčiščevanju zmot. Ena pogosto omenjanih v zadnjem letu je eksistencialna nevarnost umetne inteligence, ki naj bi ogrožala človeštvo tako kot jedrske vojne. Hkrati pa nihče ne poda vsaj za silo smiselnega scenarija, kako naj bi se to zgodilo – recimo, kako naj bi 100x pametnejši GPT ogrozil ljudi.

Letošnja konferenca poleg čisto tehnoloških izpostavlja pomembne integralne teme, kot so okolje, zdravstvo, politika depopulacije, ter rešitve, ki jih za skoraj vse probleme prinaša umetna inteligenca. V takšnem okolju je ključnega pomena poglobljena analiza in diskurz, ki lahko oblikujeta najboljše pristope k upravljanju in izkoriščanju tehnologij. Imamo veliko srečo, da gostimo vrsto izjernih mislecev, znanstvenikov in strokovnjakov, ki skupaj v delovnem in akademsko odprtem okolju prinašajo bogastvo znanja in dialoga. Verjamemo, da je njihova prisotnost in udeležba ključna za oblikovanje bolj inkluzivne, varne in trajnostne informacijske družbe. Za razcvet.

Letos smo v multikonferenco povezali deset odličnih neodvisnih konferenc, med njimi »Legende računalništva«, s katero postavljamo nov mehanizem promocije informacijske družbe. IS 2023 zajema okoli 160 predstavitev, povzetkov in referatov v okviru samostojnih konferenc in delavnic, skupaj pa se je konference udeležilo okrog 500 udeležencev. Prireditve so spremljale okrogle mize in razprave ter posebni dogodki, kot je svečana podelitev nagrad. Izbrani prispevki bodo izšli tudi v posebni številki revije Informatica (<http://www.informatica.si/>), ki se ponaša s 46-letno tradicijo odlične znanstvene revije. Multikonferenco Informacijska družba 2023 sestavljajo naslednje samostojne konference:

- Odkrivanje znanja in podatkovna središča
- Demografske in družinske analize
- Legende računalništva in informatike
- Konferenca o zdravi dolgoživosti
- Miti in resnice o varovanju okolja
- Mednarodna konferenca o prenosu tehnologij
- Digitalna vključenost v informacijski družbi – DIGIN 2023
- Slovenska konferenca o umetni inteligenci + DATASCIENCE
- Kognitivna znanost
- Vzgoja in izobraževanje v informacijski družbi
- Zaključna svečana prireditve konference

Soorganizatorji in podporniki konference so različne raziskovalne institucije in združenja, med njimi ACM Slovenija, SLAIS za umetno inteligenco, DKZ za kognitivno znanost in Inženirska akademija Slovenije (IAS). V imenu organizatorjev konference se zahvaljujemo združenjem in institucijam, še posebej pa udeležencem za njihove dragocene prispevke in priložnost, da z nami delijo svoje izkušnje o informacijski družbi. Zahvaljujemo se tudi recenzentom za njihovo pomoč pri recenziranju.

S podelitvijo nagrad, še posebej z nagrado Michie-Turing, se avtonomna stroka s področja opredeli do najbolj izstopajočih dosežkov. Nagrado Michie-Turing za izjemen življenjski prispevek k razvoju in promociji informacijske družbe je prejel prof. dr. Andrej Brodnik. Priznanje za dosežek leta pripada Benjaminu Bajdu za zlato medaljo na računalniški olimpijadi. »Informacijsko limono« za najmanj primerno informacijsko tematiko je prejela nekompatibilnost zdravstvenih sistemov v Sloveniji, »informacijsko jagodo« kot najboljšo potezo pa dobi ekipa RTV za portal dostopno.si. Čestitke nagrajencem!

Mojca Ciglarič, predsednica programskega odbora
Matjaž Gams, predsednik organizacijskega odbora

FOREWORD - INFORMATION SOCIETY 2023

The twenty-sixth Information Society multi-conference is taking place during a period of exceptional development for artificial intelligence, computing, and informatics, encompassing the entire information society. Generative artificial intelligence has made significant progress towards superintelligence, towards singularity, and the flourishing of human civilization with programs like ChatGPT. Experts' predictions are coming true, asserting that the mentioned fields are crucial for humanity's existence and development. Hence, we must direct our attention to them, swiftly integrating them into primary, secondary education, and the daily lives of individuals and communities.

On the other hand, alongside fake news, we witness the emergence of false encyclopaedias, pseudo-sciences, and flat Earth theories, along with the continuing neglect of scientific insights and methods, the diminishing of human rights, and societal values. It is upon all of us to appropriately address today's challenges, mainly assisting in the introduction of scientific knowledge and clearing up misconceptions. A frequently mentioned concern over the past year is the existential threat posed by artificial intelligence, supposedly endangering humanity as nuclear wars do. Yet, nobody provides a reasonably coherent scenario of how this might happen, say, how a 100x smarter GPT could endanger people.

This year's conference, besides purely technological aspects, highlights important integral themes like the environment, healthcare, depopulation policies, and solutions brought by artificial intelligence to almost all problems. In such an environment, in-depth analysis and discourse are crucial, shaping the best approaches to managing and exploiting technologies. We are fortunate to host a series of exceptional thinkers, scientists, and experts who bring a wealth of knowledge and dialogue in a collaborative and academically open environment. We believe their presence and participation are key to shaping a more inclusive, safe, and sustainable information society. For flourishing.

This year, we connected ten excellent independent conferences into the multi-conference, including "Legends of Computing", which introduces a new mechanism for promoting the information society. IS 2023 encompasses around 160 presentations, abstracts, and papers within standalone conferences and workshops. In total about 500 participants attended the conference. The event was accompanied by panel discussions, debates, and special events like the award ceremony. Selected contributions will also be published in a special issue of the journal *Informatica* (<http://www.informatica.si/>), boasting a 46-year tradition of being an excellent scientific journal. The Information Society 2023 multi-conference consists of the following independent conferences:

- Data Mining and Data Warehouse - SIKDD
- Demographic and Family Analysis
- Legends of Computing and Informatics
- Healthy Longevity Conference
- Myths and Truths about Environmental Protection
- International Conference on Technology Transfer
- Digital Inclusion in the Information Society - DIGIN 2023
- Slovenian Conference on Artificial Intelligence + DATASCIENCE
- Cognitive Science
- Education and Training in the Information Society
- Closing Conference Ceremony

Co-organizers and supporters of the conference include various research institutions and associations, among them ACM Slovenia, SLAIS for Artificial Intelligence, DKZ for Cognitive Science, and the Engineering Academy of Slovenia (IAS). On behalf of the conference organizers, we thank the associations and institutions, and especially the participants for their valuable contributions and the opportunity to share their experiences about the information society with us. We also thank the reviewers for their assistance in reviewing.

With the awarding of prizes, especially the Michie-Turing Award, the autonomous profession from the field identifies the most outstanding achievements. Prof. Dr. Andrej Brodnik received the Michie-Turing Award for his exceptional lifetime contribution to the development and promotion of the information society. The Achievement of the Year award goes to Benjamin Bajd, gold medal winner at the Computer Olympiad. The "Information Lemon" for the least appropriate information move was awarded to the incompatibility of information systems in the Slovenian healthcare, while the "Information Strawberry" for the best move goes to the RTV SLO team for portal dostopno.si. Congratulations to the winners!

Mojca Ciglarič, Chair of the Program Committee
Matjaž Gams, Chair of the Organizing Committee

KONFERENČNI ODBORI

CONFERENCE COMMITTEES

International Programme Committee

Vladimir Bajic, South Africa
Heiner Benking, Germany
Se Woo Cheon, South Korea
Howie Firth, UK
Olga Fomichova, Russia
Vladimir Fomichov, Russia
Vesna Hljuz Dobric, Croatia
Alfred Inselberg, Israel
Jay Liebowitz, USA
Huan Liu, Singapore
Henz Martin, Germany
Marcin Paprzycki, USA
Claude Sammut, Australia
Jiri Wiedermann, Czech Republic
Xindong Wu, USA
Yiming Ye, USA
Ning Zhong, USA
Wray Buntine, Australia
Bezalel Gavish, USA
Gal A. Kaminka, Israel
Mike Bain, Australia
Michela Milano, Italy
Derong Liu, Chicago, USA
Toby Walsh, Australia
Sergio Campos-Cordobes, Spain
Shabnam Farahmand, Finland
Sergio Crovella, Italy

Organizing Committee

Matjaž Gams, chair
Mitja Luštrek
Lana Zemljak
Vesna Koricki
Mitja Lasič
Blaž Mahnič
Mateja Mavrič

Programme Committee

Mojca Ciglarič, chair
Bojan Orel
Franc Solina
Viljan Mahnič
Cene Bavec
Tomaž Kalin
Jozsef Györköös
Tadej Bajd
Jaroslav Berce
Mojca Bernik
Marko Bohanec
Ivan Bratko
Andrej Brodnik
Dušan Caf
Saša Divjak
Tomaž Erjavec
Bogdan Filipič
Andrej Gams
Matjaž Gams
Mitja Luštrek
Marko Grobelnik
Nikola Guid

Marjan Heričko
Borka Jerman Blažič Džonova
Gorazd Kandus
Urban Kordeš
Marjan Krisper
Andrej Kuščer
Jadran Lenarčič
Borut Likar
Janez Malačič
Olga Markič
Dunja Mladenič
Franc Novak
Vladislav Rajkovič
Grega Repovš
Ivan Rozman
Niko Schlamberger
Stanko Strmčnik
Jurij Šilc
Jurij Tasič
Denis Trček
Andrej Ule
Boštjan Vilfan

Baldomir Zajc
Blaž Zupan
Boris Žemva
Leon Žlajpah
Niko Zimic
Rok Piltaver
Toma Strle
Tine Kolenik
Franci Pivec
Uroš Rajkovič
Borut Batagelj
Tomaž Ogrin
Aleš Ude
Bojan Blažica
Matjaž Kljun
Robert Blatnik
Erik Dovgan
Špela Stres
Anton Gradišek

KAZALO / TABLE OF CONTENTS

<i>Slovenska konferenca o umetni inteligenci / Slovenian Conference on Artificial Intelligence</i>	1
PREDGOVOR / FOREWORD	3
PROGRAMSKI ODBORI / PROGRAMME COMMITTEES	5
Time-Series Cutmix Data Augmentation for Heart Sound Classification Using Neural Networks / Susič David, Gams Matjaž	7
Social Interaction Prediction from Smart-Phone Sensor Data / Martinšek Marcel Franse, Lukan Junoš, Bolliger Larissa, Clays Els, Šiško Primož, Luštrek Mitja	11
Comparison of Advanced Processing Methods for PPG Denoising using a Novel Signal Quality Metric / Kawai Kitoshi, Slapničar Gašper, Hirose Akira, Luštrek Mitja	15
Machine-learning methods for analysis of gene expression data / Jordan Marko, Valič Jakob, Luštrek Mitja ...	19
Identifying Bumblebee Buzzes Using Neural Networks / Šket Tilen, Susič David, Galen Candace, Schul Johannes, Arbetman Marina, Campopiano Robinson Victoria, Villagra Gil Cristian Alfonso, Herrera Valentina, Gradišek Anton	20
Recognition of Bee Activity in the Hive Entrance Using Machine Vision and Other Methods / Rotar Oskar, Žnidaršič Maks, Vesel Tian, Silan Darja, Božič Janko	24
Implementation of a Virtual Assistant ChatGPT into the Medical Platform / Zadobovšek Matic, Kocuvan Primož, Gams Matjaž.....	28
Types of Democracy Defined: Keyword Extraction from Eleven Different Text Descriptions / Kolar Žiga, Lukšič Andrej A., Vozlič Andrej A., Gams Matjaž.....	32
Evaluation of the Effects of On-demand Dynamic Transportation of Employees to Their Workplaces in Ljubljana / Bohanec Marko, Guček Marko, Kontić Davor, Sirk Karina, Ženko Bernard, Žnidaršič Martin..	36
Usability Test of a Modified WHCA* Algorithm for Multi-Agent Pathfinding in a Real-time Strategy Game / Antešić Ivan, Sadikov Aleksandar	40
An Attempt at Predicting Algorithm Performance on Constrained Multiobjective Optimization Problems / Andova Andrejaana, Vodopija Aljoša, Cork Jordan, Tušar Tea, Filipič Bogdan	44
A Multi-Step Evaluation Process in Electric Motor Design / Tušar Tea, Korošec Peter, Filipič Bogdan	48
<i>Indeks avtorjev / Author index</i>	53

Zbornik 26. mednarodne multikonference
INFORMACIJSKA DRUŽBA – IS 2023
Zvezek A

Proceedings of the 26th International Multiconference
INFORMATION SOCIETY – IS 2023
Volume A

Slovenska konferenca o umetni inteligenci
Slovenian Conference on Artificial Intelligence

Uredniki / Editors

Mitja Luštrek, Matjaž Gams, Rok Piltaver

<http://is.ijs.si>

12. oktober 2023 / 12 October 2023
Ljubljana, Slovenia

PREDGOVOR

Kar se tiče umetne inteligence, živimo v vznemirljivih časih. V času Informacijske družbe 2022 ChatGPT še ni bil na voljo, ko se je pojavil, pa je bilo vse drugače – umetna inteligenca se je iz nečesa, o čemer so govorili predvsem strokovnjaki, prelevila v nekaj, o čemer govorijo vsi. Spremembe v dojemanju (generativne) umetne inteligence so še nekoliko večje kot napredek njenih dejanskih zmožnosti, a tudi slednji je velik: nesporno je postala zelo uporabno orodje za pisanje besedil in programske kode ter generiranje slik, uveljavlja se kot osebni pomočnik in še več. Če bo razvoj umetne inteligence še naprej tako hiter, kot je trenutno, si v prihodnosti zlahka predstavljamo dramatičen napredek v produktivnosti gospodarstva, raziskavah in vsakdanjem življenju. A marsikdo umetno inteligenco vidi tudi kot grožnjo: od tega, da bo generirala velike količine posamezniku prilagojenih škodljivih vsebin in ljudem prevzela delovna mesta, do tega, da nas bo vse pobila. Tehtanje, kaj od tega se res utegne zgoditi, je malo prehud zalogaj za tale predgovor, tako da si raje pogledjmo, kaj tak razvoj dogodkov pomeni za naše delo.

Dostopnost velikih jezikovnih modelov pomeni, da so naloge, ki zahtevajo razumevanje ali generiranje naravnega jezika, lažje izvedljive kot kdajkoli prej. To je odlično, nerodno pa je, da je gradnja tovrstnih modelov zaradi velike računske zahtevnosti izven dosega velike večine organizacij – tako raziskovalnih kot podjetij. A še vedno se lahko ukvarjamo z njihovim prilagajanjem, vključno z uporabo spodbujevanega učenja s človekovim povratnim odzivom (angl. reinforcement learning with human feedback), in tehnikami za tvorbo pozivov (angl. prompt engineering), ki dajo zelena izhodna besedila. Veliko raziskovalnih izzivov nudijo metode za zagotavljanje zaupanja v umetno inteligenco in njeno uravnavanje s cilji snovalcev (angl. alignment), za kar v splošnem še ni dobrih rešitev. Lahko pa se ukvarjamo tudi z vprašanji, kako umetno inteligenco regulirati, kako bo preobrazila (informacijsko) družbo in kaj storiti, da jo bo na bolje, kaj inteligenca sploh je in kdaj bo njena umetna različica dobila zavest.

Če se ozremo k Slovenski konferenci o umetni inteligenci, opazimo, da se velikih jezikovnih modelov dotika le en prispevek od 12, družbenih vidikov umetne inteligence pa nobeden. Da bi bolje naslovili vroče tematike zadnjega časa, smo povabili Mykolo Pechenizkiyja s Tehnične univerze v Eindhovnu, da nam predava o praktičnih problemih in metodah zaupanja vredne umetne inteligence. Po lanski uspešni izvedbi Data Science Meetupa v sklopu konference – dogodka, kjer imajo strokovnjaki iz industrije kratke predstavitve svojega dela – ga na podoben način organiziramo tudi letos. Povezovanje z industrijo je dandanes namreč bolj pomembno kot kdajkoli prej, saj se tam dogaja vedno več raziskav in razvoja umetne inteligence.

Mitja Luštrek
Matjaž Gams
Rok Piltaver
predsedniki Slovenske konference o umetni inteligenci

FOREWORD

With regards to artificial intelligence, we live in exciting times. During Information Society 2022, ChatGPT was not yet available, but when it appeared, everything changed – artificial intelligence turned from a subject mainly discussed by experts to something everyone is talking about. The changes in the perception of (generative) artificial intelligence are somewhat greater than the changes in its actual capabilities, but the latter are also substantial: it has definitely become a highly useful tool for writing text and software code as well as for generating images, it is increasingly capable as a personal assistant, and more. If it continues to advance at the current pace, we can easily imagine the future to bring dramatic improvements in economic productivity, research and everyday life. But many also see it as a threat: it may generate personalised harmful content at scale and take away jobs from people, and some even believe it may kill us all. Considering which of these things are actually likely to happen is beyond the scope of this preface, so let us look instead at what these developments mean for our work.

The accessibility of large language models means that tasks requiring understanding or generating natural language are easier than ever. This is excellent, but on the other hand building such models is beyond the reach of the vast majority of organisations – both research and business – due to its huge computational cost. However, we can still engage in fine-tuning of such models, including with reinforcement learning with human feedback, and develop techniques for engineering prompts that yield desired outputs. A lot of research challenges can be found in developing methods to ensure trust in artificial intelligence and align it with the goals of the creators, for which there are no generally good solutions yet. We can also tackle questions of how to regulate artificial intelligence, how it will transform (information) society and how to ensure it will transform it for the better, what intelligence is, and when its artificial version will gain consciousness.

Looking at the Slovenian Conference on Artificial Intelligence, we notice that only one out of the 12 papers touches upon large language models, and none addresses the social aspects of artificial intelligence. To better address the hot topics of recent times, we invited Mykola Pechenizkiy from Eindhoven University of Technology to give a keynote on practical problems and methods of trustworthy artificial intelligence. Last year we successfully combined the conference with Data Science Meetup – an event where industry experts give short presentations of their work – so we are going to organise it in a similar manner this year. Connecting with the industry is more important than ever, as more and more research and development of artificial intelligence is happening there.

Mitja Luštrek
Matjaž Gams
Rok Piltaver
Slovenian Conference on Artificial Intelligence chairs

PROGRAMSKI ODBOR / PROGRAMME COMMITTEE

Mitja Luštrek

Matjaž Gams

Rok Piltaver

Cene Bavec

Marko Bohanec

Marko Bonač

Ivan Bratko

Bojan Cestnik

Aleš Dobnikar

Erik Dovgan

Bogdan Filipič

Borka Jerman Blažič

Marjan Krisper

Marjan Mernik

Biljana Mileva Boshkoska

Vladislav Rajkovič

Niko Schlamberger

Tomaž Seljak

Peter Stanovnik

Damjan Strnad

Miha Štajdohar

Vasja Vehovar

Martin Žnidaršič

Time-Series Cutmix Data Augmentation for Heart Sound Classification Using Neural Networks

David Susič
Jožef Stefan Institute
Ljubljana, Slovenia
david.susic@ijs.si

Matjaž Gams
Jožef Stefan Institute
Ljubljana, Slovenia
matjaz.gams@ijs.si

ABSTRACT

In cutmix augmentation, a synthetic data instance is created by mixing parts of a pair of original instances. In this paper, a domain-specific time-series cutmix approach was employed for a heart sound classification task. The approach utilized neural networks and was tested on a publicly available heart sound dataset. To assess the efficacy of generating realistic instances, we implemented three distinct constraints for pairing requirements. Our main focus of interest was evaluation of performance of the approaches on datasets of varying sizes, thus we performed the experiments using different fractions of the train set. Cutmix showed promising results as best improvements in accuracy over the no augmentation baseline reached 5.61% and 1.46% when 10% and 100% of the training data were used, respectively.

KEYWORDS

cutmix, synthetic data generation, time-series augmentation, phonocardiogram, heart sound classification, neural network

1 INTRODUCTION

The accurate classification of heart sounds plays a crucial role when it comes to early detection of cardiovascular diseases (CVDs). Although heart auscultation is a cost-effective and efficient diagnostic method, the accurate identification of heart abnormalities requires proficient auscultation skills. Auscultation, typically performed with a stethoscope, involves listening to the various sounds produced by the heart, such as heartbeats and murmurs, to detect any irregularities or issues. To reduce the burden on healthcare staff and to mitigate the medical costs of delayed CVDs detection, there is a growing demand for automated approaches for identification of heart sound abnormalities. The integration of cutting-edge machine learning (ML) and signal processing technologies has led to the widespread adoption of automated computer methods for tackling this challenge.

When it comes to heart sound classification from phonocardiograms (PCGs) using ML technologies, a variety of approaches have been tried and established, most of which draw inspiration from the broader domains of general audio and image data analysis. These encompass strategies such as utilizing features derived from time-series signals as inputs for classical ML models or neural networks (NNs). Additionally, techniques involving spectrograms as inputs for NNs, employing time-series directly as NN inputs, extracting deep features from NNs for use in ML models, and even applying wavelet analysis have all been experimented with.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Information Society 2023, 9–13 October 2022, Ljubljana, Slovenia
© 2023 Copyright held by the owner/author(s).

In alignment with this research landscape, our study revolves around utilizing time-series PCGs as inputs for a NN. This particular approach has been previously explored and reported in [11, 3, 13, 16, 5, 6, 17, 10, 1, 18, 12].

When it comes to time-series augmentation, the basic approaches include augmentations in time, frequency and time-frequency domains such as jittering, flipping, scaling, warping, cutout, cutmix and others [15, 7, 19]. To the best of our knowledge, the only works that specifically analyse augmentation methods for heart sound classification deal with spectral imaged based approaches [8, 2, 21]. In this paper, we present a cutmix approach that has been specifically adapted to time-series PCG domain. In a cutmix approach, a synthetic instance is created from two original PCGs. To assess the efficacy of generating realistic instances, we implemented three distinct constraints for pairing requirements. These constraints help limit the various ways in which the original PCGs can be distributed in pairs for the creation of the synthetic PCGs.

2 DATASET

In this study, we used heart sounds from the PhysioNet/CinC Challenge 2016 dataset [9, 4], which consists of six (a-f) distinct PCG databases. Recordings were labeled as either normal or abnormal heart sounds and were split into train and validation set by the competition organizers. For our experiments, we utilized the validation set as the test set, while adjusting the original train set to achieve class balance. To ensure fair evaluation, both the train and the test subsets were balanced with respect to the two classes. In total, we used 277 normal and abnormal train recordings and 145 normal and 151 abnormal test recordings as depicted in Figure 1. The total lengths of train and test recordings are 3h 18min and 1h 37min, respectively.

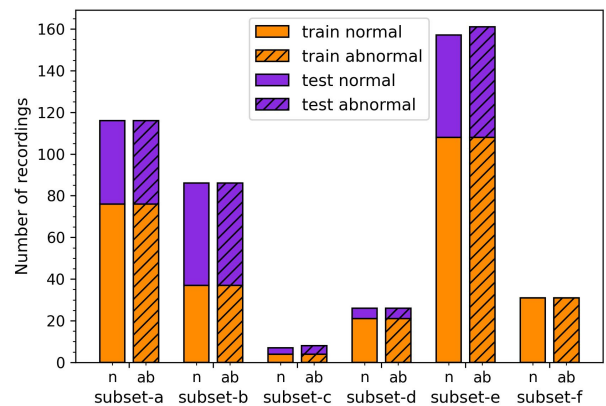


Figure 1: Data distribution used in our experiments.

The dataset also includes segmentation annotations denoting the locations of the fundamental heart states (S1, systole,

S2, and diastole) for each of the PCGs. They were computed using Springer’s segmentation algorithm [14] and additionally manually checked and corrected by the experts.

3 METHODOLOGY

The outcome of interest used to evaluate the cutmix approaches was a binary variable indicating whether a PCG represents a normal or abnormal heart. The methodology steps included pre-processing of the PCGs, selection of NN architecture and determination of cutmix augmentations techniques.

3.1 Preprocessing

The recordings were first down-sampled to 1000 Hz, filtered using Butterworth filter of order four to four frequency bands, 25-45Hz, 45Hz-80Hz, 80-200Hz, 200-400Hz, and finally normalized using root mean square normalization with the target amplitude of -20 dBFS. The recordings were then split into separate heart cycles (segments) according to segmentation annotations and zero-padded to 2.5s. Each instance was thus a time-series 2-D array of shape (4, 2500), with the first dimension denoting the number of frequency band channels and the second dimension denoting the segment length.

The goal of our study was to also find out how well the cutmix augmentation approaches perform on datasets of smaller sizes, therefore we split the training set into a variety of differently-sized smaller sets, specifically, 10%, 20%, 40%, 60%, and 80%. The selected smaller training sets were randomly selected and were stratified with respect to the data subsets and class labels within each subset. First, the smallest (10%) subset was created. Then, new subjects were added without reselection of the smaller set, meaning that the first next larger set (20%) included all of the subjects from the smaller set. This process was repeated until all of the subsets were created.

3.2 Convolutional Neural Network

The neural network structure used in our experiments was inspired by the one proposed in [11]. It consists of four convolutional neural networks (CNN) where each takes a different frequency band as the input. Besides the input layer, the CNNs consists of two additional layers, each including a convolution, ReLU activation and max-pooling. The last CNN layer also includes 25% dropout. The output of the four CNNs are flattened and input to a dense network with 20 hidden neurons with dropout of 50% and two output neurons.

3.3 PCG Cutmix Augmentation

We tested different cutmix approaches of data augmentation for classification of the PCGs. A cutmix data augmentation technique is originally an image data augmentation strategy that replaces a patch of pixels from one image with a patch from another image [20]. In our case, a section of one PCG was replaced by a section of another PCG. A synthetic instance is generated from a pair of original instances, input 1 and input 2.

In the cutmix technique presented in this study, a synthetic instance was generated using the following steps. Firstly, a pair of input instances with the same class label was selected. Secondly, both input instances were cut at the same point within the four heart beat stages (S1, systole, S2, diastole). Subsequently, the first part of input 1 was mixed with the second part of input 2. This process resulted in a new synthetic instance encompassing all four heart beat stages that had the same class label as the inputs.

Note that only one synthetic instance was generated per input instance. Cutmix scheme is shown in Figure 2.

Two cutting variations were considered, deterministic and random. In a deterministic variation, the cutting point was always right after systole, as depicted with red dashed line in Figure 2, while in random variation, a cutting point was randomly chosen to be either right after S1, systole, or S2 as depicted with dashed black lines.

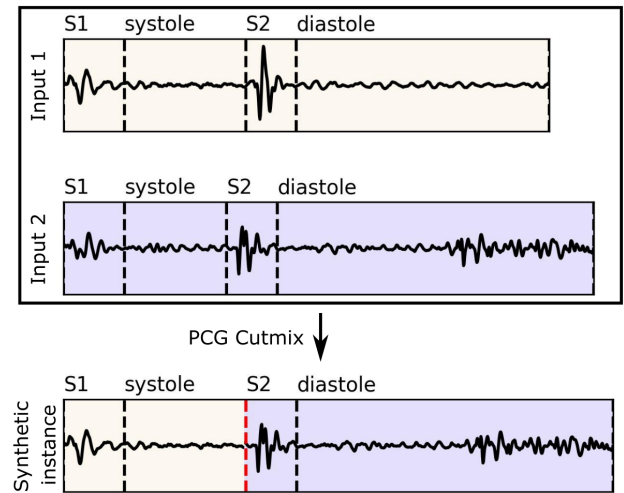


Figure 2: Visualization of PCG cutmix. Inputs 1 and 2 have the same class label. The black dashed lines denote possible cutting points, while the red dashed line denotes the selected cutting point. Note that only one synthetic instance is generated per input pair.

Data augmentation was performed during the model training. If a batch underwent augmentation (this was determined by the cutmix probability parameter for each batch separately at random), then the whole batch was replaced with the synthetic batch. The synthetic batch was created in such a way, that each instance of the original batch played a role of input 1 and input 2 exactly once, which results in the synthetic batch having the same size as the original and no original instance being over- or under-represented. In all of the cases, the synthetic instances were generated using PCGs of the same class label, thus eliminating any uncertainties regarding the labeling of the generated instance. Furthermore, we implemented some additional pairing requirements. In the Subset cutmix, the input pairs were constrained to be drawn from the same data subset, while in the Subject cutmix, the input pairs had to belong to the same subject (same PCG). In the Length cutmix, the instances in the batch were sorted into 10 bins based on their heart beat length, and each pair was drawn from the same bin. The unconstrained cutmix methodology is referred to as Basic cutmix. The idea behind the limitations was to check if creating more realistic synthetic PCGs by forcing the input PCGs of each synthetic instance to have the same recording settings, come from the same subject and/or be of similar length improves the model performance.

3.4 Hyperparameter Finetuning

In our experiments, there were four parameters to be tuned. Three hyperparameters, e.g., batch size, epoch number and learning rate, as well as the augmentation probability parameter. The latter denotes the probability of applying augmentation to a batch during

training. To optimize these parameters, we employed a grid-search methodology, which involved evaluating performance on the training set through three-fold cross-validation. To reduce search time, batch size was fixed to 512, which demonstrated effectiveness across all scenarios. Meanwhile, we explored epoch numbers within the range of 50 to 100, learning rates ranging from 10^{-4} and 10^{-1} , and cutmix probabilities spanning 0.1 to 1.0.

In order to minimize the number of experiments required for identifying the optimal configurations, we divided the fine-tuning process into stages. Initially, we determined the best combination of epoch number and learning rate for the baseline scenario without augmentation. Subsequently, utilizing these identified settings, we fine-tuned the augmentation probability for each specific augmentation technique. Ultimately, the chosen parameters were utilized to train the model using the complete training set, with final evaluation carried out on the test set.

It is important to note that additional improvement of both baseline and the methods' accuracies could be achieved by expanding the hyperparameter search grid. Although we fixed the batch sized at 512 due to computational constraints, our experimentation revealed that smaller batch sizes tend to yield slightly improved outcomes.

4 RESULTS

To explore the impact of generating synthetic data on datasets of varying sizes, we assessed the effectiveness of cutmix augmentation technique across various proportions of the training data: 10% (56 PCGs), 20% (112 PCGs), 40% (222 PCGs), 60% (334 PCGs), 80% (444 PCGs), and 100% (554 PCGs). Alongside the cutmix approaches, we present results for the no augmentation baseline scenario as well. Accuracy served as the key performance evaluation metric, with findings based on the mean and standard deviation (SD) derived from three runs. The results are shown in Figure 3 and are given in Table 1.

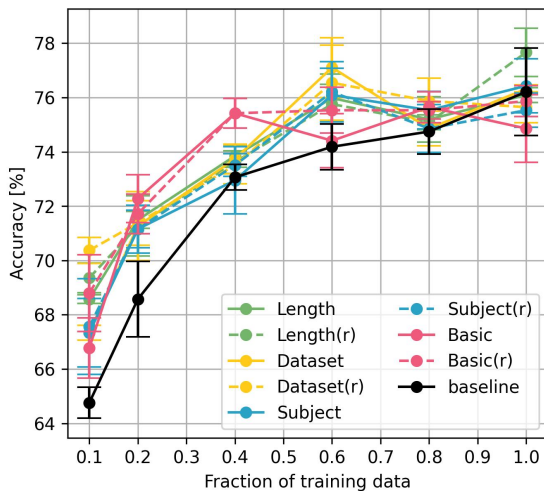


Figure 3: Accuracy vs. training data fraction of the cutmix approaches and the baseline.

We see that all of the methods work decently with some of the methods providing big improvements in accuracy over the baseline. At 10% and 100% of the train data, accuracy improvements over the baseline are 5.61% and 1.46%, respectively. To

check the statistical significance of the results, we included accuracy SD derived from a set of three experiment runs. In scenarios involving 10% and 20% of the training data, all cutmix approaches demonstrated statistically significant improvements compared to the baseline approach. However, for higher percentages of training data, not all of the approaches showed statistically significant improvements. The augmentation techniques seem to have insignificant effect when the whole train data is used.

When comparing deterministic and random cutting variations, the outcomes show that in the majority of the cases there are no statistically significant differences between the two.

In terms of the pairing requirements constraints, the most constrained method, Subject, performs the worst. Conversely, the least constrained method, Subject and unconstrained Basic approach, perform the best, whereas the moderately constrained Length method shows intermediate performances. Consequently, we deduce that the pairing diversity holds substantial significance, whereas the strategy of how the pair instances are mixed together holds lesser importance. Comparing the Dataset and Basic approaches does not give a straightforward answer on which one works better.

Interesting view of performance of the synthetic data generation is also measuring how much the dataset size appears to expand due to synthetic data generation technique relative to the initial dataset size. For this purpose, we first linearly extrapolated the baseline method above the 100% of the training data as well as interpolated the values below 100%. After that, the accuracy of each method at given fraction of training data was compared to the baseline to see at which data fraction the (inter/extrapolated) baseline achieves the same performance. The expansion percentages were then calculated by dividing the two data fractions. For example, the Dataset approach at 10% of training data achieves the same accuracy as the (interpolated) baseline at 28% of the training data, thus the apparent data size expansion due to synthetic data generation equals to 280%. Complete results are given in Figure 4.

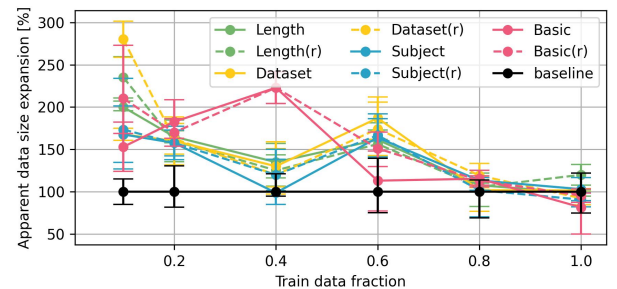


Figure 4: Apparent data size expansion vs. training data fraction.

5 DISCUSSION AND CONCLUSIONS

Our analysis indicates that the proposed cutmix augmentations substantially improve the no augmentation baseline when it comes to heart sound classification tasks using variously-sized training datasets. The results also show that there is no statistically significant difference between the random and deterministic cutting variations of the approaches. In terms of the pairing requirements constraints, the methods with no or very little constraints, Basic and Dataset, show superior performance.

Table 1: Methods' accuracies for different percentages of the training set. The numbers are given as mean (SD) derived from three runs. The best results for each considered percentage of training data are written in bold.

Method	Percentage of training data					
	10% (N=56)	20% (N=112)	40% (N=222)	60% (N=334)	80% (N=444)	100% (N=554)
Basic	66.78±1.11	72.28±0.88	75.42±0.55	74.41±0.99	75.65±0.57	74.86±1.24
Basic(r)	68.8±1.41	71.72±0.73	75.42±0.55	75.53±0.84	75.53±0.32	75.87±0.57
Length	68.57±0.16	71.49±0.32	73.85±0.42	75.98±0.88	75.2±0.84	76.09±0.27
Length(r)	69.36±0.55	71.27±1.11	73.63±0.42	75.76±0.73	75.08±0.27	77.67±0.88
Dataset	67.34±0.27	71.27±1.27	73.74±0.55	77.1±1.1	74.86±0.63	76.32±0.16
Dataset(r)	70.37±0.48	71.38±0.82	73.51±0.69	76.54±1.38	75.87±0.84	75.65±0.57
Subject	67.34±1.26	71.16±0.88	72.95±1.24	76.09±0.99	75.53±0.69	76.43±0.99
Subject(r)	67.56±1.77	71.16±0.69	73.51±0.42	76.21±1.11	74.86±0.88	75.53±0.63
baseline	64.76±0.57	68.57±1.38	73.06±0.48	74.19±0.84	74.75±0.82	76.21±1.61

Conversely, the most constrained method, Subject, performs the worst, whereas the moderately constrained Length method shows intermediate performances. As a result, we deduce that the variability in pairings carries considerable importance, whereas the cutting approach, whether deterministic or random, holds comparatively less importance.

Although the results of the best augmentation methods look promising, further experimentation is essential. These experiments should include various selections of fractions of the training data. This is important due to intrinsically stochastic nature of selecting a fraction of the training dataset. This influence of randomness becomes stronger as the fraction size decreases. For instance, if we repeatedly choose 10% of the training set at random, the resulting subsets will exhibit significantly less overlap compared to the scenario where we repeatedly select 80% of the training set randomly.

The cutmix augmentation methods have demonstrated their ability to yield superior results compared to the baseline. Nevertheless, further experiments are needed to compare these methods with standard time-series augmentations, including techniques such as cutout, manifold mixup, noise induction, polarity flip, gain adjustment, and more.

ACKNOWLEDGMENTS

The authors acknowledge the funding from the Slovenian Research and Innovation Agency, Grant (PR-10495) (DS) and Basic core funding P2-0209 (MG).

REFERENCES

- [1] Neeraj Baghel, Malay Kishore Dutta, and Radim Burget. 2020. Automatic diagnosis of multiple cardiac diseases from pcg signals using convolutional neural network. *Computer Methods and Programs in Biomedicine*, 197, 105750. doi: <https://doi.org/10.1016/j.cmpb.2020.105750>.
- [2] Neeraj Baghel, Malay Kishore Dutta, and Radim Burget. 2020. Automatic diagnosis of multiple cardiac diseases from pcg signals using convolutional neural network. *Computer Methods and Programs in Biomedicine*, 197, 105750. doi: <https://doi.org/10.1016/j.cmpb.2020.105750>.
- [3] Martin Gjoreski, Anton Gradišek, Borut Budna, Matjaž Gams, and Gregor Poglajen. 2020. Machine learning and end-to-end deep learning for the detection of chronic heart failure from heart sounds. *IEEE Access*, 8, 20313–20324. doi: [10.1109/ACCESS.2020.2968900](https://doi.org/10.1109/ACCESS.2020.2968900).
- [4] Ary L. Goldberger et al. 2000. Physiobank, physiotoolkit, and physionet. *Circulation*, 101, 23, e215–e220. eprint: <https://www.ahajournals.org/doi/pdf/10.1161/01.CIR.101.23.e215>. doi: [10.1161/01.CIR.101.23.e215](https://doi.org/10.1161/01.CIR.101.23.e215).
- [5] Ahmed Imtiaz Humayun, Shabnam Ghaffarzadegan, Md. Istiaq Ansari, Zhe Feng, and Taufiq Hasan. 2020. Towards domain invariant heart sound abnormality detection using learnable filterbanks. *IEEE Journal of Biomedical and Health Informatics*, 24, 8, 2189–2198. doi: [10.1109/JBHI.2020.2970252](https://doi.org/10.1109/JBHI.2020.2970252).
- [6] Ahmed Imtiaz Humayun, Shabnam Ghaffarzadegan, Zhe Feng, and Taufiq Hasan. 2018. Learning front-end filter-bank parameters using convolutional neural networks for abnormal heart sound detection. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 1408–1411. doi: [10.1109/EMBC.2018.8512578](https://doi.org/10.1109/EMBC.2018.8512578).
- [7] Brian Kenji Iwana and Seiichi Uchida. 2021. An empirical survey of data augmentation for time series classification with neural networks. *PLOS ONE*, 16, 7, (July 2021), 1–32. doi: [10.1371/journal.pone.0254841](https://doi.org/10.1371/journal.pone.0254841).
- [8] Tomoya Koike, Kun Qian, Björn W. Schuller, and Yoshiharu Yamamoto. 2021. Transferring cross-corpus knowledge: an investigation on data augmentation for heart sound classification. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 1976–1979. doi: [10.1109/EMBC46164.2021.9629714](https://doi.org/10.1109/EMBC46164.2021.9629714).
- [9] Chengyu Liu et al. 2016. An open access database for the evaluation of heart sound algorithms. *Physiological Measurement*, 37, 12, (Nov. 2016), 2181. doi: [10.1088/0967-3334/37/12/2181](https://doi.org/10.1088/0967-3334/37/12/2181).
- [10] Shu Lih Oh, V. Jahmunah, Chui Ping Ooi, Ru-San Tan, Edward J Ciaccio, Toshitaka Yamakawa, Masayuki Tanabe, Makiko Kobayashi, and U Rajendra Acharya. 2020. Classification of heart sound signals using a novel deep wavenet model. *Computer Methods and Programs in Biomedicine*, 196, 105604. doi: <https://doi.org/10.1016/j.cmpb.2020.105604>.
- [11] Cristhian Potes, Saman Parvaneh, Asif Rahman, and Bryan Conroy. 2016. Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds. In (Dec. 2016). doi: [10.22489/CinC.2016.182-399](https://doi.org/10.22489/CinC.2016.182-399).
- [12] Ali Raza, Arif Mehmood, Saleem Ullah, Maqsood Ahmad, Gyu Sang Choi, and Byung-Won On. 2019. Heartbeat sound signal classification using deep learning. *Sensors*, 19, 21. doi: [10.3390/s19214819](https://doi.org/10.3390/s19214819).
- [13] Heechang Ryu, Jinkyoo Park, and Hayong Shin. 2016. Classification of heart sound recordings using convolution neural network. In *2016 Computing in Cardiology Conference (CinC)*, 1153–1156.
- [14] David B. Springer, Lionel Tarassenko, and Gari D. Clifford. 2016. Logistic regression-hsmm-based heart sound segmentation. *IEEE Transactions on Biomedical Engineering*, 63, 4, 822–832. doi: [10.1109/TBME.2015.2475278](https://doi.org/10.1109/TBME.2015.2475278).
- [15] Qingsong Wen, Liang Sun, Fan Yang, Xiaomin Song, Jingkun Gao, Xue Wang, and Huan Xu. 2021. Time series data augmentation for deep learning: a survey. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*. Zhi-Hua Zhou, editor. Survey Track. International Joint Conferences on Artificial Intelligence Organization, (Aug. 2021), 4653–4660. doi: [10.24963/ijcai.2021/631](https://doi.org/10.24963/ijcai.2021/631).
- [16] Bin Xiao, Yunqiu Xu, Xiuli Bi, Weisheng Li, Zhuo Ma, Junhui Zhang, and Xu Ma. 2020. Follow the sound of children's heart: a deep-learning-based computer-aided pediatric chds diagnosis system. *IEEE Internet of Things Journal*, 7, 3, 1994–2004. doi: [10.1109/JIOT.2019.2961132](https://doi.org/10.1109/JIOT.2019.2961132).
- [17] Bin Xiao, Yunqiu Xu, Xiuli Bi, Junhui Zhang, and Xu Ma. 2020. Heart sounds classification using a novel 1-d convolutional neural network with extremely low parameter consumption. *Neurocomputing*, 392, 153–159. doi: <https://doi.org/10.1016/j.neucom.2018.09.101>.
- [18] Te-chung Issac Yang and Haowei Hsieh. 2016. Classification of acoustic physiological signals based on deep learning neural networks with augmented features. In *2016 Computing in Cardiology Conference (CinC)*, 569–572.
- [19] Hong Yang and Travis Desell. 2022. Robust augmentation for multivariate time series classification. (2022). arXiv: 2201.11739 [cs.LG].
- [20] Sangdoon Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chum, Junsuk Choe, and Youngjoon Yoo. 2019. Cutmix: regularization strategy to train strong classifiers with localizable features. (2019). arXiv: 1905.04899 [cs.CV].
- [21] George Zhou, Yunchan Chen, and Candace Chien. 2022. On the analysis of data augmentation methods for spectral imaged based heart sound classification using convolutional neural networks. *BMC Medical Informatics and Decision Making*, 22, (Aug. 2022). doi: [10.1186/s12911-022-01942-2](https://doi.org/10.1186/s12911-022-01942-2).

Social Interaction Prediction from Smart-Phone Sensor Data

Marcel Franse Martinšek
Jožef Stefan Institute
Department of Intelligent Systems
Ljubljana, Slovenia
marcel.franse.martinsek@ijs.si

Junoš Lukan*
Jožef Stefan Institute
Department of Intelligent Systems
Ljubljana, Slovenia
junos.lukan@ijs.si

Larissa Bolliger
Department of Public Health and
Primary Care
Ghent University
Ghent, Belgium
larissa.bolliger@ugent.be

Els Clays
Department of Public Health and
Primary Care
Ghent University
Ghent, Belgium
els.clays@ugent.be

Primož Šiško
Jožef Stefan Institute
Department of Intelligent Systems
Ljubljana, Slovenia
sisko.primoz@gmail.com

Mitja Luštrek*
Jožef Stefan Institute
Department of Intelligent Systems
Ljubljana, Slovenia
mitja.lustrek@ijs.si

ABSTRACT

Occupational stress is often associated with social interactions. We used data collected as a part of the larger study to predict whether a person is interacting with another while at work and at home. The dataset consisted of three weeks of data of 55 participants and included information on application and screen usage, calls, location, and Bluetooth and Wi-Fi data. We exploited a question about work activities to obtain approximate labels of social interactions. Additionally, we derived a feature indicating indoor location, which did not turn out to be useful in our case. In a binary classification problem tackled with a random forest model, we achieved an F_1 score of about 0.57.

KEYWORDS

social interactions, mobile sensing, indoor localization, stress detection

1 INTRODUCTION

The relationship between stress and social interactions is a complex one. Social interactions can both offer relief from psychological distress and be influenced by stressors [18]. For instance, distress can trigger support from empathetic individuals, but persistent distress can erode such support over time. Interpersonal encounters, especially instrumental support (e.g., help with tasks), can act as a protective factor against occupational stress [16] and burnout [4], but they can also be sources of conflict.

In the field of artificial intelligence, researchers have recognized the significance of social interactions, often using smartphones for monitoring due to their widespread use. Various smartphone sensors, including GPS, Bluetooth, Wi-Fi, accelerometer, microphone, and on-device analytics like call and message monitoring, application usage, screen activity, and battery status, have been employed for this purpose [20].

This paper outlines a method for detecting a person's involvement in a social interaction. We utilize a dataset of smartphone

sensor data labelled with the number of people a person interacted with in the last ten minutes. Importantly, unlike many studies that focus on remote social interactions (e.g., [2]), controlled settings (e.g., [13, 8]), or use dedicated hardware (e.g., [7, 14]), we predict everyday face-to-face interactions using personal smartphone data. Our focus is solely on predicting the number of people present during interactions. We intend to use this as a feature to predict social support quality and aspects of occupational stress.

2 DATA COLLECTION

We obtained this dataset as a part of a larger study called *Stress at work* (STRAW; [1]), where we developed an Android application [12] based on the AWARE framework [6]. This app collected participants' self-reports and smartphone data, including screen and application use, GPS location, calls, and Bluetooth and Wi-Fi access points. Participants completed short questionnaires approximately every 90 minutes, which included questions from psychometrically validated scales. They also contained questions about work activities from the previous 10 minutes.

Participants reported diverse work activities, including breaks, transit, individual work, and working with others. In the case of working with others, they specified in-person, telecommunication-based, lecture, or other interactions, also indicating the number of people involved as one, two, or more than two.

In the primary study, we collected data from 55 participants employed at research organizations in Belgium and Slovenia, with diverse genders (26 women) and ages (mean age 34.9 years, ranging from 24 to 63 years). Each participant provided data for 15 working days, responding to the questionnaires roughly every 90 min as designed [11].

3 TARGET AND FEATURE EXTRACTION

To create a predictive model for social interactions, we first defined the target variable by processing answers to the question work activity questions and then selected the most informative features. With these labelled data and the chosen features at hand, we trained a supervised random forest model, as elaborated in Section 5.

3.1 Label Extraction

We processed answers to the question about work activity in the last ten minutes and extracted the following attributes for these 10-minute segments of collected data:

*Also with Jožef Stefan International Postgraduate School.

- `n_others`: Number of interactions with others in last 10-minutes (-1: exact number unknown, 0: alone, 1: one additional person, 2: two additional people, 3: more than two additional people).
- `inperson`: Interaction in person (True/False).
- `formal`: Formality of interaction (True/False)

Due to question interdependence some attribute values were impossible to determine, resulting in missing values. For instance, participants mentioned activities like “coffee, lunch or toilet break” for which we didn’t collect the number of people present.

These attributes served two purposes: data filtering (`inperson` and `formal`) and segment labelling (`n_others`), resulting in a labelled dataset of 3371 10-minute segments with features listed in the next section. The target variable distribution was highly imbalanced (see Table 1).

Label	Examples Count	Binary Merged Count
3	479	1157
2	179	
1	390	
-1	109	
0	2214	2214

Table 1: Class distributions of extracted and merged labels.

To simplify the problem, we merged labels into a binary representation, indicating whether the number of people involved was greater than 0 (True) or not (False).

3.2 Features

3.2.1 Initial Feature Set. Following the literature referenced in Section 1, we selected sensors from the collected data that could be informative in predicting interactions. We computed first and second-order features from these sensors during the 10-minute segments. Below are the sensors we utilized, along with brief descriptions and feature counts in parentheses:

- `Activity_recognition` (1): Physical activity (walking, running, cycling).
- `Applications_foreground` (99): Use of various categories of phone applications.
- `Bluetooth` (30): Count and variance of visible Bluetooth devices (own and foreign).
- `Calls` (29): Call duration, quantity, and entropy of duration.
- `Locations` (21): GPS location features like variance and average speed.
- `Messages` (10): Number of sent/received messages.
- `Screen` (7): Screen details, such as unlock duration.
- `Speech` (5): Detection of human speech via microphone input.
- `Wi-Fi` (8): Visible Wi-Fi access point count, Wi-Fi localization (see Section 4).

Here, the speech sensor was a custom-implemented method that was running on the device and classified audio data online (see our previous work, [9], for more details)

3.2.2 Feature selection. In previous work [12, 11], we implemented numerous features from the sensors mentioned earlier. To address high feature correlation, we reduced them to a smaller subset before classification. We applied a threshold of $r = 0.8$ and retained only one feature from each highly correlated group.

In cases where features were highly similar and had correlations close to 1, the selection was arbitrary. We chose the feature based on its simplicity and ease of interpretation. Figures 1 and 2 depict correlation matrices before and after this feature selection step.

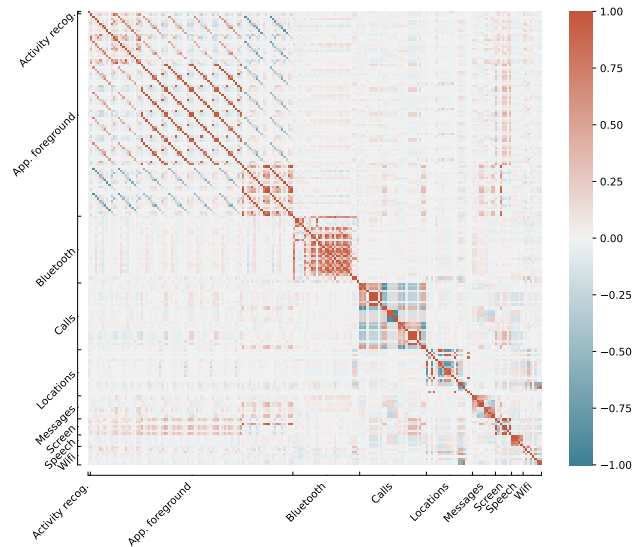


Figure 1: Correlation matrix of all features grouped by sensor, excluding categorical features

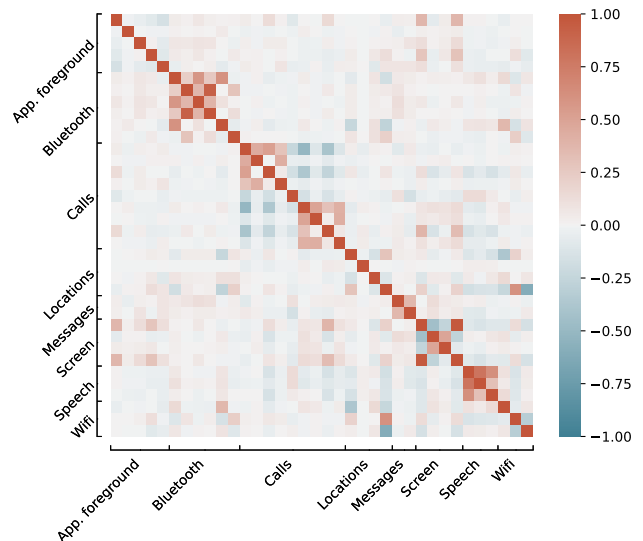


Figure 2: Correlation matrix after manual elimination of highly correlated features.

We utilized sklearn’s [15] random forest (RF) implementation to reduce the dimensionality of feature space, employing the Gini impurity metric. We explored various RF hyperparameters (`max_depth`: 10, 20, 30 and 40; `n_estimators`: 100, 500 and 1000) to observe feature selection and their impact on Gini impurity. Some features were infrequently chosen, and some of the selected ones had minimal impact (less than 0.02 reduction) on Gini impurity. We performed 10-fold cross-validation to assess the selected feature set’s predictive quality using the F_1 metric. By evaluating how features influenced the F_1 score and retaining those significantly impacting it while removing others with minor effects on Gini impurity, we derived the final feature set:

- `phone_screen`: Screen unlocks and total screen unlocked time.
- `phone_speech`: Mean and standard deviation of human voice proportion in audio.
- `phone_applications_foreground`: Duration of communication and tools app usage.
- `phone_locations`: Average speed, time in significant locations, location variance logarithm, and time at the most visited location.
- `phone_bluetooth`: Number of unique devices, total scans of others’ devices, and their mean.
- `phone_wifi_visible`: Number of unique sensed Wi-Fi access points.
- `phone_calls`: Mean call duration (outgoing and incoming).

Most of the features used were included in RAPIDS [19] with implementation of Doryab’s [5, 3] Bluetooth and location features.

4 WI-FI LOCALIZATION

In addition to the simple features described in the previous section, we aimed to incorporate indoor location data into our models, focusing on a subject’s home and work settings based on GPS location. We identified these settings by distinguishing between the two most frequented locations within each setting.

4.1 Rough GPS location

For rough GPS location determination, we utilized GPS cluster labels provided by Doryab [5]. We employed a simple heuristic, considering the most common cluster label between 00:00 and 6:00 as home and between 6:00 and 20:00 as work. Subsequently, we segregated each subject’s Wi-Fi scan data into home and work categories, excluding data recorded outside of these time segments. This allowed us to analyze each setting independently.

4.2 Indoor Wi-Fi location

We adapted Wi-Fi localization from a supervised to an unsupervised learning approach due to a lack of calibration steps during data collection. This adjustment was made individually for each subject’s data, given the differences in their environments.

The Wi-Fi sensor data included three key variables: timestamp, detected device’s media access control (MAC) address, and received signal strength indicator (RSSI). We filtered out entries with uncommon MAC addresses (appearing fewer than 100 times) and grouped the remaining entries into 1-minute segments to create Wi-Fi fingerprints. Each fingerprint, specific to a subject and setting, comprised a combination of MAC addresses and corresponding RSSI values recorded during that minute. This often resulted in 70 or more unique values, presenting a high-dimensional clustering challenge.

4.3 Clustering

We applied k -means clustering with the silhouette measure to determine the optimal value of k . Given the uncertainty about the number of significant locations per subject, we started with $k = 12$ and reduced it, ultimately stopping at a minimum of $k = 4$, as we expected at least four significant locations (e.g., desk, lunch, meeting room, other).

For each clustering result, we computed the silhouette measure [17] and selected the k value yielding the highest silhouette score. Additionally, we manually assessed the clustering results

by visualizing the relationship between different combinations of principal components (see Fig. 3).

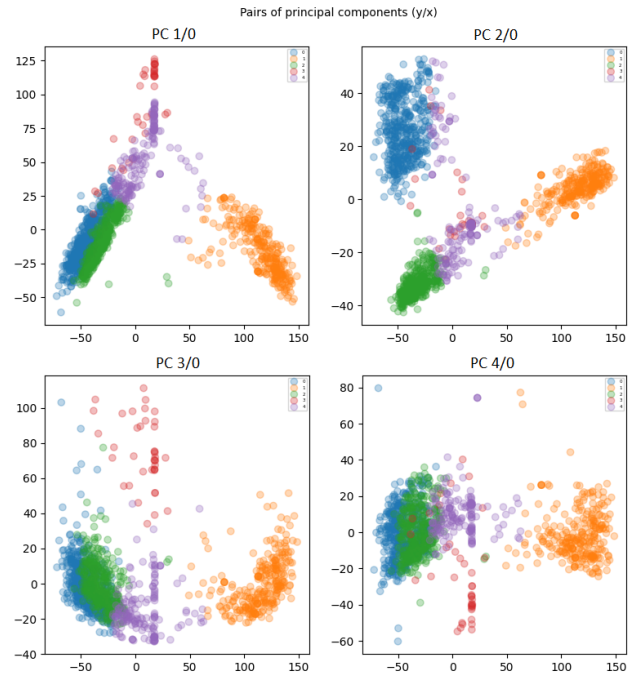


Figure 3: Visualization of a subject’s Wi-Fi fingerprint clustering results using principal component pairs as returned by principle component analysis with colours representing different clusters.

After clustering individual-specific data, we assigned an indicator label unique to each person, which couldn’t be generalized. For each location, we determined whether it represented the most or second-most frequent location in a given setting. We split the data into home and work subsets, performing separate Wi-Fi data clustering for each, resulting in four localization features.

- **Scope:** General location classification as most common (1), second most common (2), or neither (0), alongside personal cluster labels.
- **Setting:** Home or work (determined by rough GPS location). If outside these settings, corresponding features were set to -1.

Ultimately, these features proved uninformative for our use case and were removed during feature selection in Section 3.2.2. This may be attributed to limitations in our modified clustering approach without initial calibration steps during data collection.

5 INTERACTION CLASSIFICATION

We selected the random forest model for its versatility in handling both categorical and numerical data, capturing complex, non-linear feature-target relationships through ensemble techniques. To tackle dataset imbalance, we initially used imblearn’s [10] random oversampling and undersampling, with undersampling proving superior in our F_1 score evaluation. Additionally, we imputed missing values with zeros when it logically represented the intended feature content, e.g., missing duration values indicating non-usage of specific applications.

Table 2 presents LOSO and 10-fold cross-validation results for models, comparing undersampling, oversampling, and the

original class distribution. We explored various random forest configurations by adjusting maximum tree depth (Max_d) and the number of trees (N_estim).

Dataset imbalance posed challenges; oversampling and the base approach led to numerous false classifications in the majority class, while undersampling misclassified more majority class examples as the minority. A more balanced class distribution yielded a less biased model and a slight increase in the average F_1 score with undersampling. LOSO was notably affected by sampling methods, with undersampling performing better due to the problem's subject-dependent nature and overlap in examples between train and test splits in 10-fold CV.

Hyperparameters		LOSO			10-Fold		
Max_d	N_estim	Under	Over	Base	Under	Over	Base
40	100	0.573	0.504	0.471	0.553	0.530	0.471
40	500	0.562	0.508	0.482	0.549	0.551	0.476
40	1000	0.557	0.512	0.481	0.556	0.556	0.479
30	1000	0.555	0.513	0.483	0.542	0.551	0.480
20	1000	0.561	0.519	0.477	0.547	0.556	0.482
10	1000	0.558	0.524	0.462	0.559	0.552	0.447

Table 2: Table displaying F_1 scores for the model with various hyperparameter combinations. It presents results for both CV and LOSO approaches using undersampling (Under), oversampling (Over), and no class imbalance mitigation (Base).

6 CONCLUSIONS

In this paper, we developed an RF model to detect personal interactions using various phone sensors, achieving an F_1 score of 0.57 after feature filtering. The most informative features included the speech sensor and the subject's application usage statistics. Surprisingly, the indoor location did not improve results, possibly due to the COVID-19 pandemic's remote work context. Despite social interactions not being the focus of the primary study, we achieved a moderate F_1 value. We anticipate that with a more focused data collection approach and a balanced dataset, further improvements can be made, potentially enabling accurate estimation of the number of people in social interactions.

ACKNOWLEDGEMENTS

The stress at work (STRAW) study was supported by the Research Foundation – Flanders, Belgium (FWO) under Grant (project no. G.0318.18N); and the Slovenian Research and Innovation Agency (ARIS) under Grant (project ref. N2-0081).

REFERENCES

- [1] Larissa Bolliger, Junoš Lukan, Mitja Luštrek, Dirk De Bacquer, and Els Clays. 2020. Protocol of the STRESS at work (STRAW) project: how to disentangle day-to-day occupational stress among academics based on EMA, physiological data, and smartphone sensor and usage data. *International Journal of Environmental Research and Public Health*, 17, 23, (Nov. 2020), 8835. doi: 10.3390/ijerph17238835.
- [2] Mehdi Boukhechba, Alexander R. Daros, Karl Fua, Philip I. Chow, Bethany A. Teachman, and Laura E. Barnes. 2018. DemonicSalmon: monitoring mental health and social interactions of college students using smartphones. *Smart Health*, 9-10, (Dec. 2018), 192–203. doi: 10.1016/j.smhl.2018.07.005.
- [3] Luca Canzian and Mirco Musolesi. 2015. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, Osaka, Japan, (Sept. 2015), 1293–1304. ISBN: 9781450335744. doi: 10.1145/2750858.2805845.
- [4] J. D. DeFreese and Jason P. Mihalik. 2016. Work-based social interactions, perceived stress, and workload incongruence as antecedents of athletic trainer burnout. *Journal of Athletic Training*, 51, 1, (Jan. 2016), 28–34. doi: 10.4085/1062-6050-51.2.05.
- [5] Afsaneh Doryab, Prerna Chikarsel, Xinwen Liu, and Anind K. Dey. 2018. Extraction of behavioral features from smartphone and wearable data. abs/1812.10394, (Dec. 18, 2018). arXiv: 1812.10394 [cs.CY]. doi: 10.48550/ARXIV.1812.10394.
- [6] Denzil Ferreira, Vassilis Kostakos, and Anind K. Dey. 2015. AWARE: Mobile context instrumentation framework. *Frontiers in ICT*, 2, 6, 1–9. doi: 10.3389/fict.2015.00006.
- [7] Michele Girolami, Fabio Mavilia, and Franca Delmastro. 2020. Sensing social interactions through BLE beacons and commercial mobile devices. *Pervasive and Mobile Computing*, 67, (Sept. 2020), 101198. doi: 10.1016/j.pmcj.2020.101198.
- [8] Kleomenis Katevas, Katrin Hänsel, Richard Clegg, Ilias Leontiadis, Hamed Haddadi, and Laurissa Tokarchuk. 2019. Finding dory in the crowd. In *Proceedings of the 1st Workshop on Machine Learning on Edge in Sensor Systems*. ACM, (Nov. 2019). doi: 10.1145/3362743.3362959.
- [9] Marko Katrašnik, Junoš Lukan, Mitja Luštrek, and Vitomir Štruc. 2019. Razvoj postopka diarizacije govorcev z algoritmi strojnega učenja. In *Proceedings of the 22nd International Multiconference INFORMATION SOCIETY – IS 2019*. Slovenian Conference on Artificial Intelligence (Ljubljana, Slovenia, Oct. 7–11, 2019). Mitja Luštrek, Rok Piltaver, and Matjaž Gams, editors. Vol. A, 57–60. http://library.ijs.si/Stacks/Proceedings/InformationSociety/2019/IS2019_Volume_A.pdf.
- [10] Guillaume Lemaitre, Fernando Nogueira, and Christos K. Aridas. 2017. Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, 18, 17, 1–5. <http://jmlr.org/papers/v18/16-365.html>.
- [11] Junoš Lukan, Larissa Bolliger, Els Clays, Oscar Mayora, Venet Osmani, and Mitja Luštrek. 2021. Participants' experience and adherence in repeated measurement studies among office-based workers. In *Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers*. Afsaneh Doryab, Qin Lv, and Michael Beigl, editors. ACM, (Sept. 2021), 528–531. doi: 10.1145/3460418.3479367.
- [12] Junoš Lukan, Marko Katrašnik, Larissa Bolliger, Els Clays, and Mitja Luštrek. 2020. Straw application for collecting context data and ecological momentary assessment. In *Proceedings of the 23rd International Multiconference INFORMATION SOCIETY – IS 2020*. Slovenian Conference on Artificial Intelligence (Ljubljana, Slovenia). Mitja Luštrek, Rok Piltaver, and Matjaž Gams, editors. Vol. A, (Oct. 2020), 63–67.
- [13] Aleksandar Matic, Venet Osmani, and Oscar Mayora-Ibarra. 2012. Analysis of social interactions through mobile phones. *Mobile Networks and Applications*, 17, 6, (Aug. 2012), 808–819. doi: 10.1007/s11036-012-0400-4.
- [14] Niklas Palaghias, Seyed Amir Hoseinitabatabaei, Michele Nati, Alexander Gluhak, and Klaus Moessner. 2015. Accurate detection of real-world social interactions with smartphones. In *2015 IEEE International Conference on Communications (ICC)*. IEEE, (June 2015). doi: 10.1109/icc.2015.7248384.
- [15] F. Pedregosa et al. 2011. Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- [16] Maria C. W. Peeters, Bram P. Buunk, and Wilmar B. Schaufeli. 1995. Social interactions, stressful events and negative affect at work: a micro-analytic approach. *European Journal of Social Psychology*, 25, 4, (July 1995), 391–401. doi: 10.1002/ejsp.2420250404.
- [17] Peter J. Rousseeuw. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, (Nov. 1987), 53–65. doi: 10.1016/0377-0427(87)90125-7.
- [18] Marybeth Shinn, Stanley Lehmann, and Nora W. Wong. 1984. Social interaction and social support. *Journal of Social Issues*, 40, 4, (Jan. 1984), 55–76. doi: 10.1111/j.1540-4560.1984.tb01107.x.
- [19] Julio Vega, Meng Li, Kwesi Aguilera, Nikunj Goel, Echhit Joshi, Kirtiraj Khandekar, Krina C. Durica, Abhineeth R. Kunta, and Carissa A. Low. 2021. Reproducible analysis pipeline for data streams. Open-source software to process data collected with mobile devices. *Frontiers in Digital Health*, 3, doi: 10.3389/fdgh.2021.769823.
- [20] Heng Zhang, Ahmed Ibrahim, Bijan Parsia, Ellen Poliakoff, and Simon Harper. 2022. Passive social sensing with smartphones: a systematic review. *Computing*, 105, 1, (Aug. 2022), 29–51. doi: 10.1007/s00607-022-01112-2.

Comparison of Advanced Processing Methods for PPG Denoising using a Novel Signal Quality Metric

Kitoshi Kawai*
kawai-kitoshi38@eis.t.u-tokyo.ac.jp
The University of Tokyo
Tokyo 113-8656, Japan

Akira Hirose
ahirose@ee.t.u-tokyo.ac.jp
The University of Tokyo
Tokyo 113-8656, Japan

Gašper Slapničar
gasper.slapnicar@ijs.si
Jožef Stefan Institute
Jožef Stefan International Postgraduate School
Jamova cesta 39
Ljubljana, Slovenia

Mitja Luštrek
mitja.lustrek@ijs.si
Jožef Stefan Institute
Jožef Stefan International Postgraduate School
Jamova cesta 39
Ljubljana, Slovenia

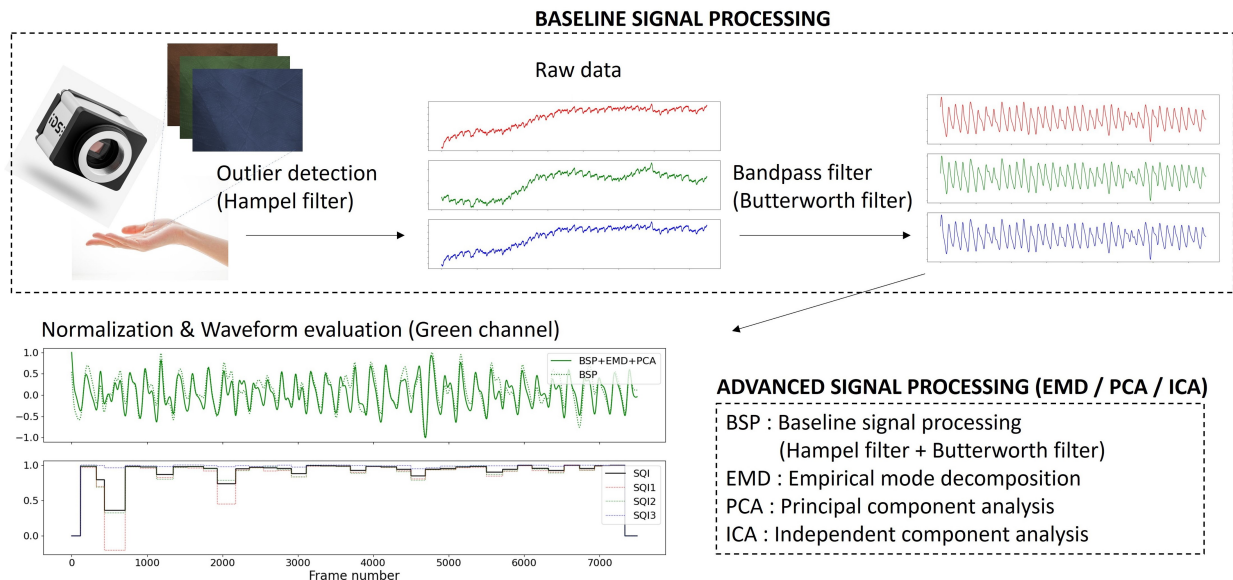


Figure 1: The top dashed box is baseline signal processing, the bottom right is advanced signal processing. The Upper left subplot is the raw recording of RGB channels. Both baseline and advanced signal processing were applied, and the lower subplot shows the denoised signal and its waveform evaluation.

ABSTRACT

Photoplethysmography (PPG) is a non-invasive method measuring blood volume changes using light. By illuminating tissue and observing light variations, PPG captures blood flow fluctuations. Vital physiological parameters can be obtained by analyzing the PPG signals, such as heart rate and oxygen saturation. In this study, we acquired PPG signals from the palm of a hand using camera-based remote sensing. However, this approach is especially sensitive to noise due to contact-free nature. We propose novel metrics for waveform evaluation of PPG signals and a unique pipeline combining several advanced methods for PPG

denoising. Applying our state-of-the-art pipeline to our dataset demonstrated a 5-12% improvement in the resting state and a 27-28% improvement in the active state in terms of our proposed Signal Quality Index metric. This was compared to the baseline denoising which employed only outlier removal and bandpass filtering.

KEYWORDS

photoplethysmography, noise removal, empirical mode decomposition, principal component analysis, independent component analysis, signal quality assessment

1 INTRODUCTION

Non-invasive physiological monitoring has grown essential in healthcare. Photoplethysmography (PPG) uses light to track blood volume changes. PPG often employs contact sensors, capturing light variations through the skin, which unveil blood flow characteristics and allow the extraction of metrics like heart rate (HR)

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Information Society 2023, 9–13 October 2023, Ljubljana, Slovenia

© 2023 Copyright held by the owner/author(s).

and oxygen saturation. However, obtaining such parameters is not straightforward due to PPG being susceptible to noise such as motion artifacts.

Remote PPG (rPPG) acquires PPG signals without direct contact, avoiding the discomfort of attached devices. However, rPPG signals are more vulnerable to noise, especially from unintended movements. This leads to degradation of the PPG signal. When estimating blood pressure using PPG, accurate detection of systolic peaks is crucial [12]. Motion artifacts, however, can induce problematic peak shifts in the PPG signal.

The aim of this study was to refine the rPPG signal and assess its waveform with our proposed Signal Quality Index (SQI) detailed in Section 3.3. By examining different processing techniques for noisy rPPG signals, we wanted to enhance PPG technology and improve physiological monitoring reliability.

2 RELATED WORK

Several methods have been proposed to remove motion artifacts from PPG signals, with statistical approaches like Principal component analysis (PCA), Independent component analysis (ICA), and Empirical mode decomposition (EMD) [3, 9, 6]. In estimating the PPG-derived respiratory rate, separate utilization of PCA and EMD has demonstrated errors of 1.48 and 0.07 breaths/min, respectively, suggesting their efficacy in noise removal. However, the performance without any filter remained unassessed, so it is difficult to evaluate its performance. In another study, Motin et al. integrated EMD and PCA for PPG-based breath rate estimation. They reported absolute errors of 0.9 breaths and 9.9 breaths in 5-minute recordings, depending on the dataset [6]. In contrast, the exclusive use of a bandpass filter resulted in errors of 5.4 and 10.5 breaths in 5-minute respectively [8], indicating that combining different processing methods might be effective in improving robustness. In these studies, the PPG signal was not evaluated directly but rather through the quality of variables extracted from it. In contrast, Slapničar et al. focused on the waveform of the PPG signal itself and evaluate the signal in a data-driven manner [11]. In their study, a metric called Signal Quality Indices (SQIs) was defined and a threshold was set to extract only good waveforms.

3 METHODOLOGY

Our methodology for processing PPG data was divided into two stages: baseline signal processing (BSP) and advanced signal processing. In the BSP stage, we first applied the Hampel filter followed by the Butterworth filter to detect outliers and extract specific frequency components. In the advanced signal processing stage, additional noise removal was accomplished using either EMD, PCA, ICA, or a combination of these methods.

We then assessed the performance of previously described PPG processing methods using custom SQIs, which we describe in detail in Section 3.3. In this study, it was assumed that the HR was between 50 and 140 beats per minute ($50 \leq \text{HR} \leq 140$) as the HR of the subjects in our experiment ranged from about 60 to 130.

3.1 Baseline Signal Processing

We first used the Hampel filter to detect outliers. Afterwards, since the PPG data was expected to have some frequency range, a Butterworth filter was employed to extract specific frequency components. Based on the data obtained from these processes, we detected peaks using peak detection algorithm detailed by

Lazaro et al. [4] and if the number of peaks did not correspond to the HR specified previously, the data was discarded as too noisy during recording. For peak detection, we set a window (the size of which was based on HR), identified the highest gradient within, and marked the subsequent local maxima as the systolic peak. To isolate a PPG signal cycle, we also discerned the valley points, taken as the minimum between two systolic peaks [11].

3.1.1 Hampel Filter. The Hampel filter identifies outliers in time-series data through the following procedures.

- Take three points before and after the target data point and calculate the median within this window.
- Compute the median absolute deviation (MAD) by determining the median of the absolute differences between each data point in the window and the window's median.
- Multiply the MAD by a constant to estimate the standard deviation under the assumption that the data follows a normal distribution.
- If the absolute difference between a data point and the median of its window exceeds three times the MAD, the data point is deemed an outlier and is replaced with the window's median.

3.1.2 Butterworth Filter. After processing the data using the Hampel filter, we used the Butterworth filter to extract components from 0.5Hz to 3.0Hz. By using this filter, we eliminated both low-frequency and high-frequency components.

3.2 Advanced Signal Processing

Within the Butterworth filter's cutoff frequency range, we further denoised using either EMD, ICA, PCA, or a combination of these methods. In this study, we considered the following combinations:

- | | |
|-----------------|----------------------|
| (i) BSP | (iv) BSP + EMD |
| (ii) BSP + PCA | (v) BSP + EMD + PCA |
| (iii) BSP + ICA | (vi) BSP + EMD + ICA |

3.2.1 Empirical Mode Decomposition. EMD is a method of decomposing a signal into physically meaningful components and is used to analyze nonlinear or non-stationary signals. Using EMD, signals can be decomposed into "intrinsic mode functions" (IMFs), which correspond to components of different frequency. We referred to the detailed algorithm in this paper [1]. For each component of the IMFs, only the first intrinsic mode function (IMF) consistently had peaks within the predefined HR range. Thus, we selected this first IMF as the PPG signal.

3.2.2 Principal Component Analysis. PCA compresses multidimensional data and extracts essential features [10]. In this study, each frame captured by the camera was divided into a 3x3 grid, giving 9 regions. Each region's average pixel value formed a nine-dimensional input. This was reduced to two dimensions, representing the PPG signal and noise. By comparing the number of peaks in the post-PCA data with the expected HR, the first component of the PCA was determined to be the PPG data. However, it should be noted that PCA can occasionally produce inverted outputs because the sign of eigenvectors, which determine the direction of principal components, is arbitrary and can be positive or negative. To account for this, we checked the correlation coefficient between the PCA output and the PPG signal which was processed with BSP after averaging the entire frame.

If the correlation coefficient was negative, indicating inversion, the output was then multiplied by -1.

3.2.3 Independent Component Analysis. ICA is a technique used to decompose multivariate signals into statistically independent components [2]. Like PCA, each frame was divided into nine regions for input, and the output consisted of two dimensions: the PPG signal and noise. However, since ICA does not define the order or sign of the separated sources, it is ambiguous to identify which outputs are related to PPG or noise, and also the components can be inverted. Therefore, considering both the output data sets and their inverses (making a total of four potential sets), the data with the highest correlation coefficient to the PPG signal which was processed with BSP after averaging the entire frame was selected as the PPG data.

3.3 PPG Waveform Quality Assessment

For PPG waveform evaluation, a template was first created. This template was then used to calculate the SQI by comparing the denoised PPG data with the template.

3.3.1 Template Wave Formation. The length of one cycle template waveform was computed using autocorrelation analysis. Given the previously defined HR range, the potential cycle length range was denoted by Eq. (1)

$$\frac{\text{fps} \times 60}{140} \leq L \leq \frac{\text{fps} \times 60}{50} \quad (1)$$

where fps is the sampling frequency in Hertz (fps = 250 Hz) and L is the template length in samples.

We then shifted the signal by all the lengths within that range, the correlation coefficients of the original and shifted signals were compared and the shift length with the highest correlation coefficient was selected as the template width. With the length L of the template determined, we extracted segments of width L from each valley point of the PPG signal. The template waveform was then created by averaging these segments.

3.3.2 Signal Quality Indices. We defined SQIs based on method proposed by Slapničar et al. [11]. However, in our study, we made some modifications to SQI3 to ensure normalization between -1 and 1. The template, created as detailed above, was compared and evaluated against each cycle of the PPG signal using the three following SQIs.

- SQI1: Data of length L starting from each valley point is directly compared with the template to calculate the Pearson's correlation coefficient.
- SQI2: Data between two adjacent valleys is considered as one cycle. If the waveform length of one cycle is different from the template, the data is compared with the template by resampling to determine Pearson's correlation coefficient.
- SQI3: Data between two adjacent valleys is considered as one cycle, dynamic time warping (DTW) is employed to find similar points with the template. DTW is a method used to align two sequences by warping their time axis to best match each sequence to the other [7]. The corresponding points are then used to calculate the Pearson's correlation coefficient.

Finally, the final SQI was computed by taking the average of SQI1, SQI2, and SQI3, as expressed in Eq. (2).

$$SQI = \frac{1}{3}(SQI1 + SQI2 + SQI3) \quad (2)$$

4 EXPERIMENTS

4.1 Recording Setup

In this study, we collected rPPG data from 11 subjects, both male and female, aged 22 to 45 years old. The iDS 3040SE-Q RGB camera set at 250 fps, equipped with the Sony IMX273 1/3" CMOS image sensor and the iDS-5M23-C1618 16 mm lens, was used for the recordings. Each subject underwent four 30-second recordings. For each recording, rPPG was obtained from red (R), green (G), and blue (B) channels. Of the four sessions, the initial two recordings were conducted in a "rest state" while the subsequent two were in an "active state". The rest state entailed subjects being in a relaxed condition for the recording, achieved by prompting them to engage in meditation or deep breathing prior to the session. Conversely, the active state referred to recordings taken after subjects performed physical activities like jumping or squats to elevate their HR. Consequently, it was expected that the PPG signal would be more stable in the rest state and the noise level would be higher in the active state.

4.2 Evaluation Pipeline

Using the PPG data obtained by the signal processing described in Section 3, the SQI of each channel (R, G, and B) was computed for each recording. The mean and standard deviation of the SQI were computed across both rest and active states. Our analysis entailed comparing the BSP with advanced methods to identify the most effective processing technique in terms of SQI. Additionally, we evaluated the performance differences between the rest (stable PPG) and active (potentially noisier PPG) states.

To compare channels, we also computed the mean and standard deviation of SQI using the PPG data obtained only by the best-performing processing method for both states. R, G, and B channels were then evaluated based on these SQI values as given in Table 1 (Best method).

5 RESULTS

The primary objective of this study was to investigate which combination of methods was most effective in noise removal. We computed the SQI for each RGB channel in both rest and active states, as detailed in Table 1.

In the rest state with minimal motion artifacts, the highest SQI was achieved using the BSP, EMD, and PCA filters. Conversely, post-motion data, which had pronounced motion artifacts, exhibited maximal SQI when BSP, EMD, and ICA filters were applied. Relative to the baseline SQI obtained by the BSP, enhancements were: 12% (R), 5% (G), and 10% (B) in the rest state, and 27% (R and G) and 28% (B) in the active state. In the rest state with minimal original noise, filtering slightly enhanced the PPG signal. In contrast, for active states with pronounced noise, the improvement was substantial. The SQI disparity between states was minimized with EMD, highlighting its efficiency in motion artifact removal. Besides, combining EMD with ICA and PCA further improved robustness. Moreover, as highlighted by Jihyoung et al.[5], the green channel consistently provided superior PPG signals compared to the other channels.

A specific view of the evaluation of the denoised PPG signal and waveform is shown in Fig. 2. The figure depicts the denoised PPG signal using the optimal pipeline (BSP + EMD + PCA) and the corresponding SQI assessment for the green channel during the rest state. For instance, during the collapse of the PPG waveform (approximately at frame number 700), the SQI manifests a diminished value.

	R_SQI		G_SQI		B_SQI	
	rest	active	rest	active	rest	active
BSP	0.73 ± 0.14	0.65 ± 0.13	0.80 ± 0.13	0.66 ± 0.12	0.74 ± 0.15	0.65 ± 0.13
BSP+PCA	0.74 ± 0.16	0.61 ± 0.12	0.80 ± 0.13	0.63 ± 0.11	0.75 ± 0.15	0.61 ± 0.12
BSP+ICA	0.71 ± 0.17	0.64 ± 0.13	0.77 ± 0.16	0.70 ± 0.14	0.72 ± 0.17	0.65 ± 0.13
BSP+EMD	0.79 ± 0.14	0.81 ± 0.14	0.83 ± 0.13	0.82 ± 0.14	0.80 ± 0.13	0.82 ± 0.13
BSP+EMD+PCA	0.82 ± 0.14	0.79 ± 0.14	0.84 ± 0.13	0.81 ± 0.15	0.82 ± 0.15	0.79 ± 0.14
BSP+EMD+ICA	0.79 ± 0.15	0.82 ± 0.14	0.83 ± 0.15	0.84 ± 0.13	0.80 ± 0.15	0.83 ± 0.14
Best method	0.82 ± 0.14		0.84 ± 0.13		0.83 ± 0.15	

Table 1: The mean and standard deviation of SQI across different processing methods, scenarios, and color channels. The final "Best method" is computed only from best-performing methods (BSP+EMD+PCA and BSP+EMD+ICA).

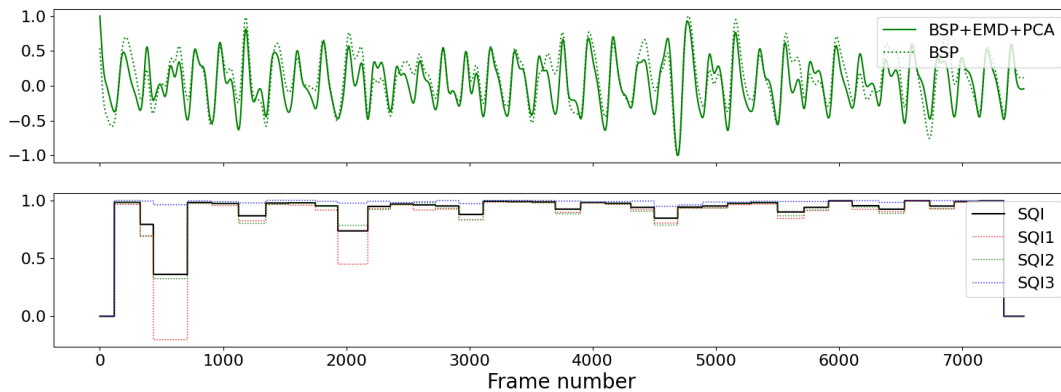


Figure 2: The top subplot shows the PPG signal (green channel) using basic signal processing and the best combination of advanced signal processing. The bottom subplot shows the corresponding waveform evaluation.

SQI values, presented in Eq. (2) as the mean of SQI1, SQI2, and SQI3, should be adjusted based on the application. For applications where precise waveform morphology are vital [12], SQI2 might be more crucial. Conversely, when precise morphology is not as important, but only the dominant peak matters, SQI1 (with its minimal computational demand) or SQI3 might be enough. Moreover, the template for SQI was created using the PPG signal obtained from each processing method. This can affect the waveform, as some systolic peaks get shifted by signal processing, as observed in Fig. 2. This is an important limitation, that requires further future investigation on how noise removal affects the details of the PPG waveform.

6 CONCLUSION

In this study, we assessed PPG signal processing techniques on in-house rPPG dataset in terms of SQI. The BSP+EMD+PCA combination outperformed the BSP by 5-12% in the rest state, while the BSP+EMD+ICA combination improved by 27-28% in the active state. The superior quality and stability of the green channel in PPG reaffirms the findings of previous studies [5].

The results highlight the potential for more advanced noise removal by integrating different signal processing, which is crucial for estimating HR and other physiological information from PPG signal. Future work will explore the impact of proposed noise removal techniques on precise morphology and subtle details of the PPG waveform.

ACKNOWLEDGEMENTS

This work was supported by the Jožef Stefan Institute and by The University of Tokyo.

REFERENCES

- [1] Norden E Huang, Zheng Shen, Steven R Long, Manli C Wu, Hsing H Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, and Henry H Liu. 1998. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society of London. Series A: mathematical, physical and engineering sciences*, 454, 1971.
- [2] Aapo Hyvärinen and Erkki Oja. 2000. Independent component analysis: algorithms and applications. *Neural networks*, 13, 4-5.
- [3] Byung S Kim and Sun K Yoo. 2006. Motion artifact reduction in photoplethysmography using independent component analysis. *IEEE transactions on biomedical engineering*, 53, 3.
- [4] Jesús Lázaro, Eduardo Gil, José María Vergara, and Pablo Laguna. 2013. Pulse rate variability analysis for discrimination of sleep-apnea-related decreases in the amplitude fluctuations of pulse photoplethysmographic signal in children. *IEEE journal of biomedical and health informatics*, 18, 1.
- [5] Jihyoung Lee, Kenta Matsumura, Ken-ichi Yamakoshi, Peter Rolfe, Shinobu Tanaka, and Takehiro Yamakoshi. 2013. Comparison between red, green and blue light reflection photoplethysmography for heart rate monitoring during motion. In *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE.
- [6] Mohammad Abdul Motin, Chandan Kumar Karmakar, and Marimuthu Palaniswami. 2019. Selection of empirical mode decomposition techniques for extracting breathing rate from ppg. *IEEE Signal Processing Letters*, 26, 4.
- [7] Meinard Müller. 2007. Dynamic time warping. *Information retrieval for music and motion*.
- [8] Lena Nilsson, Anders Johansson, and Sigga Kalman. 2000. Monitoring of respiratory rate in postoperative care using a new photoplethysmographic technique. *Journal of clinical monitoring and computing*, 16.
- [9] B Prathyusha, T Sreekanth Rao, and D Asha. 2012. Extraction of respiratory rate from ppg signals using pca and emd. *International Journal of Research in Engineering and Technology*, 1, 2.
- [10] Jonathon Shlens. 2014. A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*.
- [11] Gašper Slapničar, Mitja Luštrek, and Matej Marinko. 2018. Continuous blood pressure estimation from ppg signal. *Informatica*, 42, 1.
- [12] Gašper Slapničar, Wenjin Wang, and Mitja Luštrek. 2022. Feasibility of remote pulse transit time estimation using narrow-band multi-wavelength camera photoplethysmography. In *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers*.

Machine-learning Methods for Analysis of Gene Expression Data

Marko Jordan

Department of Intelligent Systems,
Jožef Stefan Institute
Ljubljana, Slovenia
marko.jordan@ijs.si

Jakob Valič

Department of Intelligent Systems,
Jožef Stefan Institute
Ljubljana, Slovenia

Mitja Luštrek

Department of Intelligent Systems,
Jožef Stefan Institute
Jožef Stefan International
Postgraduate School
Ljubljana, Slovenia
mitja.lustrek@ijs.si

ABSTRACT

Gene expression and similar types of biological data are often studied because they provide rich information about the state of an organism, and machine-learning models can be built to predict the organism's state from such data. A common challenge is that the number of genes, which correspond to features for machine learning, is typically large compared to the number of samples. This is tackled by feature-selection and dimensionality-reduction methods. The former have the advantage of providing information on important features, which may allow reducing the number of features that have to be collected prospectively. We present two feature-selection methods: an ensemble of established filter methods, and a custom bi-directional wrapper designed specifically for problems where the number of features is large compared to the number of instances and there may be interactions between the features. We compare the methods on a dataset consisting of multiple cohorts, by training models on some cohorts and testing on others, which best approximates real-life use. We find that some informative features can be identified, and while the wrapper does not outperform the filter ensemble, it does work better on the more challenging cases.

KEYWORDS

Gene expression, machine learning, feature selection

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Information Society 2023, 9–13 October 2023, Ljubljana, Slovenia

© 2023 Copyright held by the owner/author(s).

Identifying Bumblebee Buzzes Using Neural Networks

Tilen Šket, David Susič
Department of Intelligent Systems,
Jožef Stefan Institute
Ljubljana, Slovenia

Candace Galen, Johannes
Schul
Division of Biological Sciences,
University of Missouri
Columbia, Missouri, United States

Marina Arbetman, Victoria
Campopiano Robinson
Universisad Nacional del Comahue
Bariloche, Argentina

Cristian Alfonso Villagra Gil,
Valentina Herrera
Universidad Metropolitana de
Ciencias de la Educación
Santiago, Chile

Anton Gradišek
anton.gradisek@ijs.si
Department of Intelligent Systems,
Jožef Stefan Institute
Ljubljana, Slovenia

ABSTRACT

Bumblebees as important pollinators are keystone species and as such are crucial for functioning of the ecosystem. In Patagonia (in Argentina and Chile), the native species *Bombus dahlbomii* is under threat by the spread of invasive European species that were introduced for agricultural purposes. An important aspect of conservation efforts is monitoring of the presence of native and invasive species. Here we report on the analysis of sound recordings using neural networks, with the aim of detecting the presence of bumblebee buzzes in the recordings.

KEYWORDS

bumblebees, neural networks, buzz detection, spectrograms

1 INTRODUCTION

Bumblebees (genus *Bombus*) are a group of social insects from the bee family Apidae. They are important pollinators, often more efficient than honeybees. This comes in part due to their different morphology and lifestyle. They can forage in cold and rainy weather when honeybees will not even exit the hive, and due to a special technique called "buzz pollination" they can pollinate flowers where the pollen needs to be extracted - this is for example relevant for tomatoes, where bumblebee pollination has an extremely important commercial role as well. In ecosystems where honeybees are absent, such as in the mountains, plants rely on bumblebees and other wild pollinators.

The largest bumblebee species, *Bombus dahlbomii*, lives in temperate forests of South America, in southern Argentina and Chile [1]. It is an important pollinator of local plant species, however, it is being threatened by the introduced species (*B. terrestris* and *B. ruderatus*) that have been brought from Europe for agricultural purposes in the past decades [12, 10, 2]. As these species are expanding in range and increasing in numbers, the population of *B. dahlbomii* is diminishing and is faced with possible extinction. In order to boost the conservation efforts, careful monitoring of populations of local and introduced bumblebee species is needed as a starting point for policy makers and conservationists to plan their actions.

Bumblebee monitoring can be carried out either manually, with people in the field observing target plants and writing down notes, or with the use of technology. Clearly, even if manual monitoring has the advantage of an expert observer being able to produce high-quality records, this approach is time and resource consuming. Therefore, we explore the possibilities that smart sensors can offer, in particular sound recordings with microphones coupled with signal analysis with AI algorithms.

When it comes to utilizing machine learning (ML) for insect detection and/or classification from sound recordings, a variety of approaches have been established, many of which are influenced by the broader fields of general audio and image data analysis. These include strategies such as using features derived from time-series signals as inputs for classical ML models [11, 7], using spectrograms as inputs for neural networks (NNs) [14, 9, 8] and employing time-series directly as NN inputs [13].

In the previous studies of bumblebee sounds, some of the co-authors of this paper have investigated whether it is possible to distinguish bumblebee species and type based on flight buzzing sound [4], where, in brief, a larger body size of a bumblebee will likely result in a lower buzzing frequency. In another study, we monitored "bumblebee traffic" using microphones next to the nest-box entrance, where we used sound analysis to count the workers flying in and out [5, 3]. In those studies, the approach with microphones turned out to be highly efficient in detecting buzzes, as well as distinguishing between the arrivals and departures. In the present paper, we go one step further. We wanted to develop an algorithm that would identify bumblebee buzzes from a long recording at several chosen plants during a field study in South America. The task is more complex than the one with bumblebee traffic as the bees do not necessarily pass close by the microphone, and the buzzes sounds are furthermore masked by various ambient noises. Here, we present the initial results of an approach using convolutional neural networks for buzz detection. We discuss the accuracy of the algorithm and outline the future steps.

2 DATASET

Recordings of bumble bees were obtained in November, 2022 at forest understory and rural field sites in Argentina and Chile. Audiomoth (Open Acoustic Devices) and DB-9 USB recording devices with a detection range of 1-2 m were used to collect single channel audio recordings at 16 kHz. In this study, we used roughly 750 minutes of such recordings. This is a labeled dataset, however, the size of all the recordings obtained during the field study is substantially larger. It should be stressed that the labels

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Information Society 2023, 9–13 October 2023, Ljubljana, Slovenia

© 2023 Copyright held by the owner/author(s).

for buzzes were all approximate, as the expert noted when they saw the bumblebee, not necessarily when the bumblebee sound was picked by the microphone.

As the recordings included various ambient sounds such as lawnmowers and passing cars (which produce somewhat similar spectrograms to those of bumblebees, but more on this later), we underwent a process of isolating highly noisy segments. These segments were then combined to create a supplementary dataset totaling 2 minutes in duration, referred to as the machine sounds dataset.

3 METHODOLOGY

The objective of our study was to identify timestamps within the recordings that denote the occurrence of bumblebee buzzes. The methodology included initial recording preprocessing and data partitioning, selection of neural network architecture, and determination of the model training settings.

3.1 Data Preprocessing

Bumblebees exhibit a natural flying frequency in the range of 200 Hz, alongside higher harmonics that can extend up to 1500 Hz (at least those we can detect). Similarly, when the bumblebee engages in sonication on a flower, the emitted sounds resonate at a natural frequency of about 300 Hz, with corresponding higher harmonics [4]. It is important to stress that bumblebees are much larger and heavier than most of the other pollinating insects present in the area (such as honeybees or solitary bees), so their buzzing frequencies will be lower than those of other pollinators.

For the initial preprocessing step, recordings were subjected to a frequency filtering process, limiting frequencies up to 1500 Hz. Next, the recordings were segmented into 4-second intervals with a 50% overlap. These discrete intervals were then transformed into Mel spectrograms using the fast Fourier transform. Consequently, each instance was represented as a 2-dimensional array of dimensions 128x128. Unlike conventional spectrograms, Mel spectrograms utilize the Mel frequency scale to mimic the human ear's perception of distinct frequencies (as an ear has a logarithmic response, not a linear one). The examples of preprocessed spectrograms to be used in our model are given in Figure 1. Note that the added machine sounds were preprocessed equally.

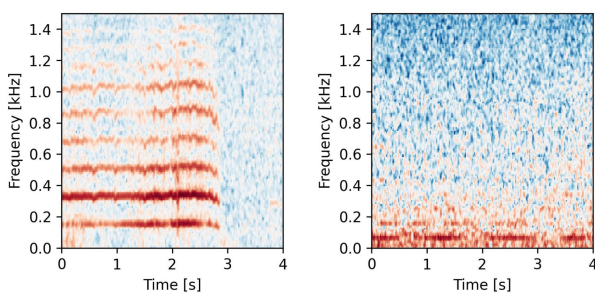


Figure 1: Spectrograms of a bumblebee buzz (left) and no buzz (right)

3.2 Data Partitioning

Our dataset comprises recordings of varying lengths: some are short (3-10 seconds), while others are longer (around 10 minutes). The training dataset was composed of the spectrograms extracted

Table 1: Test, validation, and test data set distributions.

Split	Sound	#Spectrograms
Train	buzz	96
	no buzz	440
	machine	24
Validation	buzz	24
	no buzz	110
	machine	6
Test	buzz	1984
	no buzz	20564
	machine	0

from the short recordings which either included buzz throughout the whole length or did not include a buzz at all. Conversely, the test dataset was composed of spectrograms extracted from the longer recordings. Spectrograms containing the bumblebee buzz sound were assigned with a label of 1, whereas the ones without buzzes were assigned a label of 0. The training data recordings were all manually checked and labeled by the experts. In contrast, the labeling process for the testing dataset involved experts identifying the presence of buzzing in the long recordings by providing a single time-stamp per bumblebee buzz. This time-stamp indicated the approximate detection time, meaning some spectrograms might have been mislabeled. In addition, the cases where the buzzing sound persisted for longer than 4 seconds, at least one spectrogram included a (part of a) buzz but was labeled with 0.

The training data was split into train and validation sets with the ratio of 4:1. The folds were stratified with respect to the dataset (bumblebee, machines) and with respect to the class label (buzz or no buzz). Our train, validation, and test set distributions are shown in Table 1.

As expected, we see that the test set is greatly unbalanced (most of the time, there are no bumblebees). In the approximately 710 minutes of the test recordings, the experts detected and marked down 992 time-stamps of bumblebee buzzes, resulting in 1984 of the spectrograms being labeled with a buzz due to the 50% temporal overlap of the spectrograms (each time-stamp was covered by two spectrograms).

3.3 Neural Network Architecture

From the machine learning perspective, our study was an image classification task, thus we implemented a convolutional neural network (CCN) as they are regarded as the state-of-the-art in the image classification domain [6].

Our model architecture comprises three convolutional layers, followed by a fully connected network with two dense layers and a single output neuron. The convolutional layers included, starting from the initial layer, 128, 256, and 512 filters all size at 3x3. Each convolutional layer was accompanied by a 2x2 max pooling layer. The output of the last convolutional layer was flattened and input to a dense layer consisting of 8192 neurons, followed by a batch normalization, and a dense layer consisting of 512 neurons. Ultimately, the final output was consolidated into a single neuron, generating a numeric value between 0 and 1 that signified the degree of confidence in predicting a buzz. The activation function of all convolutional layers and the first dense layer was ReLu, while the last dense layer used a sigmoid

activation function. The schematic of the the CNN used in our analysis is given in Figure 2.

As the train and validation data volume is fairly small, we tried implementing dropout regularization to the CNN layers. Surprisingly, our findings revealed that dropout did not yield any performance improvement. Consequently, we decided not to incorporate any regularization into the final model architecture.

3.4 Model Training Settings

The selected loss function to be optimized during the neural network training was the binary cross-entropy. The validation data was used to control and adjust the settings during the training. For the optimization algorithm, we employed the Adam optimizer along with a learning rate scheduler. The initial learning rate was set to 10^{-4} , which was dynamically adjusted as training progressed. Specifically, the learning rate was reduced by a factor of 10 after each epoch where there was no observed improvement in the validation loss. This adaptive approach helped the model navigate towards convergence. The batch size was 32 and the model trained for 50 epochs.

4 RESULTS

Considering the significant imbalance inherent in our dataset, we employed a heuristic approach to evaluate our model’s performance effectively. Because the test set spectrograms were extracted from 10 minute continuous intervals, there was a significant amount of temporal correlation. The buzz sounds sometimes lasted continuously for a few minutes, however during that period, only two spectrograms were labeled with a buzz, as only one time-stamp was assigned for that bumblebee and there is 50% overlap between the spectrograms.

Hence, our devised methodology operated as follows: when our model detected a buzz across a series of consecutive spectrograms and at least one of these spectrograms has been expert-annotated as a buzz, all the spectrograms within this contiguous sequence were counted as true positives.

We assess the effectiveness of the buzz detection model using a set of four key metrics: F1-score, precision, recall, and accuracy. How the metrics are calculated from true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values is given in Eqs. (1)–(4). We compare the results against a baseline model that predicted randomly assuming the class probabilities of the training data. The models’ performance metrics are compared to the baseline in Table 2. We see that our model demonstrates significant improvements across all metrics, with the exception of accuracy, where it shows a modest advantage over the baseline.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1\text{-score} = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (4)$$

The confusion matrix depicting the model’s predictions is presented in Figure 3. It’s important to note that the false negative rate of the predictions might be underestimated. This rate signifies the instances where the model failed to identify buzzes based on time-stamps, rather than indicating instances where the

Table 2: CNN model and the baseline performance results. Baseline predicted randomly assuming train data class probabilities.

Model	Metric	Result
CNN	Accuracy	76%
	Precision	60%
	Recall	96%
	F1-score	74%
Baseline	Accuracy	74%
	Precision	9%
	Recall	21%
	F1-score	12%

model missed buzz-containing spectrograms entirely. This discrepancy arises, same as before, due to the characteristic nature of buzzes, which often span across multiple spectrogram durations. Conversely, instances where the model correctly predicted a buzz, but with a slight delay compared to the expert-annotated time-stamp, were marked as incorrect predictions due to minor label misalignment.

While our model missed 298 (30%) buzz events, it often identified events as buzz despite the label indicating otherwise. This discrepancy was attributed to label misalignment, absence of buzz labels alongside bumblebee presence, and primarily the model’s unfamiliarity with noise types not encountered in training (e.g., cars, lawn mowers, distant music). Figure 4 displays examples of a false negative (left) and a false positive (right) predictions.

In cases like this, where the event one is trying to detect is rare, it is crucial to catch as many occurrences as possible. Missing out on detections (false negatives) is more problematic than having a some incorrect ones (false positives). Mistaken predictions can be manually reviewed by humans, which is quicker than re-screening the entire recordings. Therefore, although improving event detection rate and minimizing the false positive rate are crucial, the relatively high occurrence of false positive outcomes is of lesser concern.

5 DISCUSSION AND CONCLUSIONS

As demonstrated in the analysis of the algorithm performance, the neural network does detect buzzes. Just looking at the comparison with the validation set where (nominally) true positive events are labeled, the performance is still not close to the values that would be of strong use for actual field monitoring applications.

Nevertheless, we should stress that there are several improvements possible. As outlined above, we are (i) dealing with a highly unbalanced dataset, which is (ii) labeled following a protocol used in biological field studies where the event labels only approximately corresponding to the buzzes picked up by microphones (or several segments contain buzzes and there is only one label). After manually inspecting several miss-classified segments, the next step is to improve the training process of the CNN with larger training sets, which will allow us to more accurately analyze the extensive dataset recorded during the field study. Ultimately, this approach will allow us to identify bumblebee buzzes in the area and, following an approach similar to the one in Gradišek et al. [4], distinguishing between the native and introduced species.

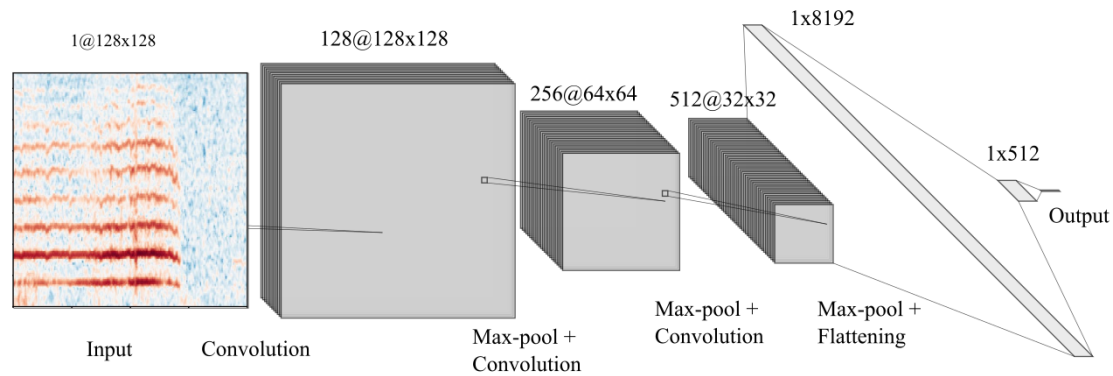


Figure 2: Convolutional neural network architecture used in our study.

Actual	no buzz	9720	5020
	buzz	298	7564
		no buzz	buzz
		Prediction	

Figure 3: Confusion matrix of the model's predictions.

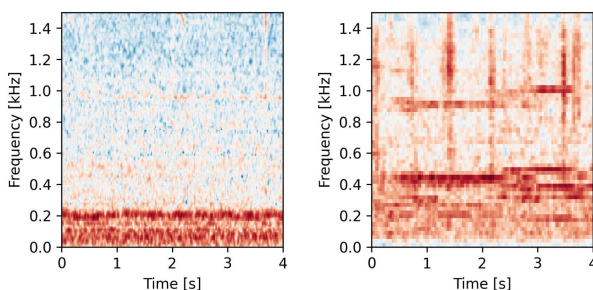


Figure 4: False negative (left) and false positive (right) spectrograms. The false negative spectrogram includes an audible buzz, however, the signal strength in the buzz frequency range is overpowered by the prevailing background noise in the lower parts of the spectrogram. The false positive spectrogram, on the other hand, exhibits distinctive horizontal lines that are similar to the patterns generated by the bumblebees, but were generated from unrelated sources that misled the model.

ACKNOWLEDGEMENTS

The authors acknowledge the funding from the Slovenian Research and Innovation Agency, Grant (PR-10495) and Basic core funding P2-0209. The research was partially funded by the National Geographic Society. We also acknowledge the students who assisted with field data collection.

DATA AND CODE AVAILABILITY STATEMENT

The data and the code that support the findings of this study are available from the authors upon reasonable request.

REFERENCES

- [1] A.H. Abrahamovich, Maria C. Telleria, and N.B. Diaz. 2001. Bombus species and their associated flora in argentina. *Bee World*, 82, 2, 76–87.
- [2] Benoit Geslin and Carolina L. Morales. 2015. New records reveal rapid geographic expansion of *bombus terrestris* linnaeus, 1758 (hymenoptera: apidae), an invasive species in argentina. *Check List*, 11, 1620–1620.
- [3] Anton Gradišek, Nicolas Cheron, David Heise, Candace Galen, and Janez Grad. 2018. Monitoring bumblebee daily activities using microphones. In *Proceedings of the 21st Annual International Multiconference Information Society–IS*, 5–8.
- [4] Anton Gradišek, Gašper Slapničar, Jure Šorn, Mitja Luštrek, Matjaž Gams, and Janez Grad. 2017. Predicting species identity of bumblebees through analysis of flight buzzing sounds. *Bioacoustics*, 26, 1, 63–76.
- [5] David Heise, Zachary Miller, Ellie Harrison, Anton Gradišek, Janez Grad, and Candace Galen. 2019. Acoustically tracking the comings and goings of bumblebees. In *2019 IEEE Sensors Applications Symposium (SAS)*. IEEE, 1–6.
- [6] Kavi B. Obaid, Subhi R. M. Zeebaree, and Omar M. Ahmed. 2020. Deep learning models based on image classification: a review. doi: 10.5281/ZENODO.4108433.
- [7] Satoshi Kawakita and Kotaro Ichikawa. 2019. Automated classification of bees and hornet using acoustic analysis of their flight sounds. *Apidologie*, 50, 1. doi: 10.1007/s13592-018-0619-6.
- [8] Ali Khalighifar, Daniel Jiménez-García, Lindsay P Campbell, Koffi Mensah Ahadj-Dabla, Fred Aboagye-Antwi, Luis Arturo Ibarra-Juárez, and A Townsend Peterson. 2021. Application of Deep Learning to Community-Science-Based Mosquito Monitoring and Detection of Novel Species. *Journal of Medical Entomology*, 59, 1, (Sept. 2021), 355–362. eprint: <https://academic.oup.com/jme/article-pdf/59/1/355/41963852/tjab161.pdf>. doi: 10.1093/jme/tjab161.
- [9] Ivan Kiskin, Bernardo Pérez Orozco, Theo Windebank, Davide Zilli, Marianne Sinka, Kathy Willis, and Stephen Roberts. 2017. Mosquito detection with neural networks: the buzz of deep learning. (2017). arXiv: 1705.05180 [stat. ML].
- [10] Josefin A. Madjidian, Carolina L. Morales, and Henrik G. Smith. 2008. Displacement of a native by an alien bumblebee: lower pollinator efficiency overcome by overwhelmingly higher visitation frequency. *Oecologia*, 156, 835–845.
- [11] Quoc Viet Phung, Iftekhar Ahmad, Daryoush Habibi, and Steven Hinckley. 2017. Automated insect detection using acoustic features based on sound generated from insect activities. *Acoustics Australia*, 45, 2. doi: 10.1007/s40857-017-0095-6.
- [12] Juliana Ordoñez Rego, Clemens Schlindwein, Ruben Garrido, and Victor H Monzón. 2021. Low fruit set in an endangered tree: pollination by exotic bumblebees and pollen resource for relictual native bees. *Arthropod-Plant Interactions*, 15, 491–501.
- [13] Myat Su Yin, Peter Haddawy, Borvorntat Nirandmongkol, Tup Kongthaworn, Chanaporn Chaisumritchoke, Akara Supratak, Chaitawat Sa-ngamuang, and Patchara Sriwichai. 2021. A lightweight deep learning approach to mosquito classification from wingbeat sounds. In *Proceedings of the Conference on Information Technology for Social Good (GoodIT '21)*. Association for Computing Machinery, Roma, Italy, 37–42. ISBN: 9781450384780. doi: 10.1145/3462203.3475908.
- [14] Myat Su Yin et al. 2023. A deep learning-based pipeline for mosquito detection and classification from wingbeat sounds. *Multimedia Tools and Applications*, 82, 4. doi: 10.1007/s11042-022-13367-0.

Prepoznavanje aktivnosti čebel na panjskem žrelu s pomočjo strojnega vida in drugih metod

Recognition of Bee Activity in the Hive Entrance Using Machine Vision and Other Methods

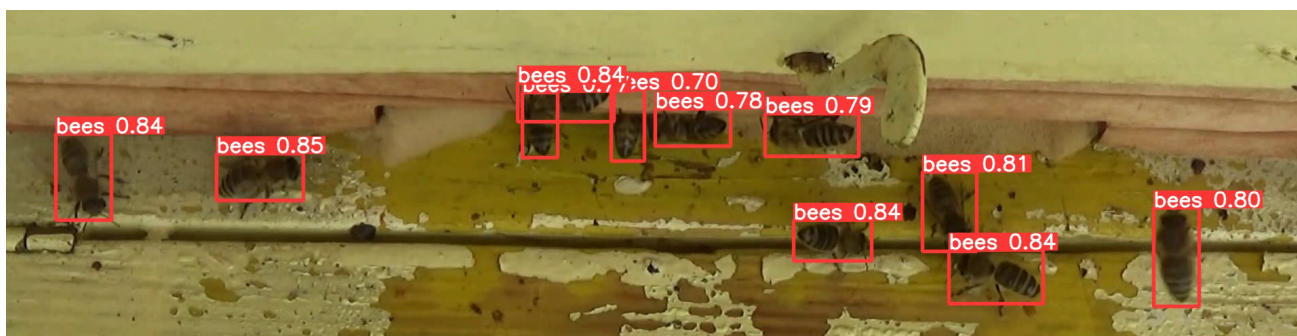
Oskar Rotar
oskar.rotar@dijak.gjp.si
Gimnazija Jožeta Plečnika, Ljubljana
Šubičeva ulica 1
Ljubljana, Slovenija

Maks Žnidaršič
maks.znidarsic@dijak.gjp.si
Gimnazija Jožeta Plečnika, Ljubljana
Šubičeva ulica 1
Ljubljana, Slovenija

Tian Vesel
tian.vesel@dijak.gjp.si
Gimnazija Jožeta Plečnika, Ljubljana
Šubičeva ulica 1
Ljubljana, Slovenija

Mag. Darja Silan
darja.silan@gjp.si
Gimnazija Jožeta Plečnika, Ljubljana
Šubičeva ulica 1
Ljubljana, Slovenija

Dr. Janko Božič
janko.bozic@bf.uni-lj.si
Biotehniška fakulteta - Univerza v
Ljubljani
Jamnikarjeva ulica 101
Ljubljana, Slovenija



Slika 1: Zaznave čebel narejene z uporabo YOLOv5 modela.

POVZETEK

V tem delu smo raziskovali možnost uporabe najnaprednejših tehnologij pri analizi gibanja čebel med vhom v panj. Čebele imajo ključno vlogo pri opravljanju in so bistvene za ohranjanje uravnoteženega ekosistema, vendar je njihovo proučevanje zaradi majhne velikosti in hitrega gibanja zahtevno. Z uporabo strojnega vida, tehnologije, ki omogoča računalniško interpretacijo in razumevanje vizualnih podatkov ter metod sledenja večim objektom (MOT) nam je uspelo razviti moderno rešitev za njihovo proučevanje. Raziskava je pokazala, da uporaba teh tehnologij ponuja nov, obetaven pristop k zbiranju velikega števila natančnih podatkov o gibanju čebel.

ABSTRACT

In this article we researched the possible use of advanced technologies for the analysis of bee movement in front of a beehive. Bees play a key role in the pollination of plants, they have an immense part in keeping our ecosystem balanced. However, due

to their small size and fast movements, it is challenging for researchers to study them. With the use of computer vision, a technology that interprets and understands visual information, as well as Multi Object Tracking (MOT), we were able to develop a modern solution to the otherwise difficult task of bee tracking. With this, the study of bees becomes less constrained and far more effective. The research shows that the use of numerous modern technologies offer a promising new approach to the collection of a large number of accurate data on bee movement.

KLJUČNE BESEDE

zaznava čebel, računalniški vid, MOT, YOLO, *Apis mellifera carnica*

KEYWORDS

bee recognition, computer vision, MOT, YOLO, *Apis mellifera carnica*

1 UVOD

V nedavni zgodovini smo bili priča velikim tehnološkim dosežkom in inovacijam. Svet hitro napreduje, vendar nekatera področja še zmeraj ostajajo nespremenjena. V Sloveniji je čebelarstvo velik del naše bogate zgodovine in kulture. Zaradi podnebni sprememb, porasta parazitov in boleznih čebel ni bilo čebelarstvo

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
Information Society 2023, 9–13 October 2023, Ljubljana, Slovenia
© 2023 Copyright held by the owner/author(s).

nikoli tako zahtevno, zato želimo z uporabo sodobnih metod in moderne tehnologije prispevati k uspešnejšemu uvajanju informacijske tehnologije za upravljanje čebeljih družin in uspešnejše delo čebelarjev in raziskovalcev. Želeli smo razviti učinkovit, stabilen in splošen model strojnega učenja za zaznavo čebel, ki vzdrži visoko uspešnost tudi v primerih, ko so algoritmične metode neuspešne. Model te podatke nato poda algoritmu za sledenje večim objektom, ki natančno sledi gibanju čebel. Ves program omogoča samodejno zbiranje velikih količin podatkov o gibanju na panjskem žrelu, kar bi lahko pomagalo pri preučevanju njihovega obnašanja in komuniciranja.

2 METODOLOGIJA

2.1 Izbira modela zaznavanja objektov

V nalogi smo za zaznavo čebel uporabili konvolucijske nevronske mreže. Pri učenju modela zaznave čebel smo uporabili modele dveh družin prednaučenih nevronskih mrež, namenjenih zaznavi objektov. To sta YOLOv5[1], ki je eden najbolj razširjenih in najboljših modelov za zaznavanje objektov in YOLOv7[2], ki je novejša arhitektura. YOLOv7 domnevno opisuje boljše rezultate, kot YOLOv5. Modele teh dveh družin modelov smo dodatno učili na lastnih podatkih. Učili smo prednaučene modele YOLOv5s, YOLOv5m, YOLOv5x in modela YOLOv7 ter YOLOv7x.

2.1.1 Poimenovanje modelov. Imena vseh modelov smo začeli z akronimom BDM (model zaznave čebel - *angl. Bee Detection Model*). Temu sledita identifikacijska številka modela. Za tem je oznaka za vrsto modela. Modeli YOLOv5s imajo oznako s, YOLOv5m oznako m, YOLOv5x oznako x, YOLOv7 oznako 7 in YOLOv7x oznako 7x. Primer poimenovanja modela je **BDM1-s**. To je model z identifikacijsko številko 1, ki je naučen na prednaučenem modelu YOLOv5s (Tabela 1, stolpec za ime).

2.2 Obdelava video posnetkov

Uporabljali smo 10 različnih video posnetkov panjskega žrela. Izbrali smo zahtevne in raznolike posnetke (različna povečava, osvetljenost, barva panja, gostota čebel), da bi dosegli čim bolj splošene modele. Posnetki so bili posneti na Urbanem učnem čebelnjaku botaničnega vrta v Ljubljani[3]. Za učenje smo iz video posnetkov na enakomernih časovnih intervalih vzeli 2100 slik. Te slike smo označili v orodju Roboflow[4]. Pri učenju se je ločljivost slik samodejno zmanjšala na ločljivost modela. 70 % slik smo postavili v učno množico, 20 % v validacijsko in 10 % v testno množico.

2.3 Učenje modelov

Vse modele smo lahko učili na istih podatkih, saj YOLOv5 in YOLOv7 uporabljata isti format za označevanje slik. Da bi prišli do čim boljšega končnega modela, smo naučili več modelov. Pri učenju modelov smo spreminjali velikost paketa, število epoh, izbiro prednaučenega modela in število slik. Ostale hiperparametre smo ohranili na standardni nastavitvi. Pri optimizaciji modela je uporabljen stohastični gradientni spust. Isti algoritem je bil uporabljen pri učenju prednaučenih modelov [1].

Velikost paketa, pri kateri smo lahko učili, je neposredno omejena z velikostjo grafičnega pomnilnika na grafični kartici (8 GB). Pri učenju manjših modelov (posebej YOLOv5s) se proces samodejno predčasno konča, ko se mAP@[0.5:0.95] modela ne spreminja, kar omeji prekomerno prilagajanje modela. To temelji na validacijski množici.

Tabela 1: Specifikacije naučenih modelov. Vse ločljivosti so kvadratne. Negativna velikost paketa označuje samodejno izbiro

ime	epohe	velikost paketa	model	ločljivost
BDM1-s ¹	100	16	v5s	640
BDM2-s	100	16	v5s	416
BDM3-s	100	16	v5s	640
BDM4-s	174	-1	v5s	640
BDM5-m	268	-1	v5m	640
BDM6-s ¹	174	-1	v5s	640
BDM7-x	27	-1	v5x	640
BDM8-s	20	32	v5s	640
BDM9-s	184	42	v5s	640
BDM10-7x	87	8	v7x	640
BDM11-7	166	10	v7	640
BDM12-7	150	11	v7	640
BDM13-7x	150	11	v7x	640
BDM14-m	135	16	v5m	640
BDM15-m	163	-1	v5m	640

¹Ta modela sta bila naučena na procesorju in ne grafični kartici.

Skupaj smo učili 15 modelov, štiri iz družine YOLOv7 in 11 iz družine YOLOv5 (Tabela 1). Pri izdelavi modelov smo imeli dva glavna cilja - hitrost izvajanja in uspešnost zaznavanj. Večinoma smo učili manjše, YOLOv5s modele. Za to smo imeli dva razloga: hitrejšo učenje teh modelov in majhno število učnih podatkov. Po prvih testnih učenjih smo ugotovili, da so modeli že zelo hitro začeli zaznavati velik delež čebel. Kljub temu smo učili tudi nekaj v5m modelov in en v5x model. Učili smo tudi YOLOv7 modele, saj smo upali, da bodo nudili višjo hitrost in kvaliteto, kot napisano v članku[2].

Pri učenju model računa svojo uspešnost s funkcijo izgube. Pri YOLOv5 in YOLOv7 modelih je ta funkcija seštevek (pri YOLOv7 obtežen seštevek) funkcije izgube objektivnosti, regresijske funkcije izgube omejevalnega okvira in funkcije izgube klasifikacije (1). Slednji je v našem primeru 0, ker uporabljamo le en razred objektov.

$$loss = l_{obj} + l_{box} + l_{cls} \quad (1)$$

2.4 Sledenje gibanju čebel

Za sledenje smo uporabili knjižnici Norfair[5] ter ByteTrack[6]. Sledenje gibanja čebel smo izvedli s pomočjo podatkov, ki nam jih je vrnil model. Pridobljene vrednosti iz modela in sledilca smo primerjali ter vsakemu objektu določili identifikacijsko številko. Knjižnica Norfair[5] je to opravila sama, pri knjižnici ByteTrack[6] pa smo to dosegli z zunanjo knjižnico Onemetric[7].

2.5 Analiza modelov

Hitrost delovanja modelov smo testirali na dveh posnetkih. Izbrali smo posnetek, ki ima majhno število čebel in počasno gibanje ter posnetek, ki ima več čebel ter bolj kaotično gibanje. Po učenju nam je program vrnil dve verziji modela - tisto, ki je bila shranjena po zadnji epohi in tisto, ki je dosegla najvišjo metriko mAP@[0.5:0.95] na validacijski množici. Vedno smo uporabili model z najvišjo mAP@[0.5:0.95], saj smo se s tem izognili prekomernemu prilagajanju modela.

2.5.1 *Uspešnost zaznavanja modelov.* Pri izračunu uspešnosti modela se uporablja več metod in funkcij[8]. Modele smo primerjali po uspešnosti, priklicu, $mAP@0.5$ in $mAP@[0.5:0.95]$. Merila smo izračunali na testni množici, torej slikah na katerih modeli niso bili naučeni.

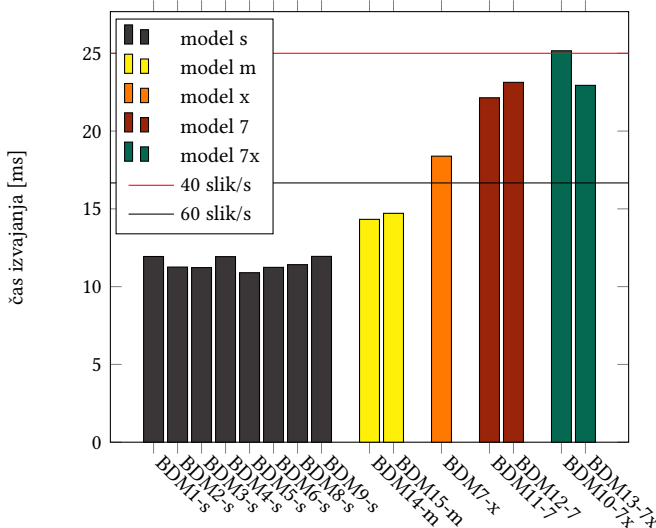
2.6 Analiza metod sledenja premikanja čebel

Pri analizi premikanja čebel smo, kot pri analizi modelov, uporabili dva posnetka. Za zaznavo čebel smo uporabljali model BDM15-m, saj se ta model izvaja hitro z dobro uspešnostjo. Za ocenjevanje natančnosti zaznave smo z obema knjižnicama prešteli čebele, ki so se približale vhodu oziroma oddaljile od vhoda v panj. Te podatke smo nato primerjali z ročno preštetimi količinami.

3 REZULTATI

3.1 Hitrost izvajanja modelov

Pri primerjavi modelov lahko vidimo, da gostota čebel na posnetku ne vpliva na hitrost zaznav. Na hitrost je imela bistven vpliv le velikost modela. Najhitrejši so bili vsi modeli prednaučenega modela YOLOv5s, ki so tudi najmanjši, najpočasnejši pa modeli YOLOv7x (Slika 2). Skoraj vsi modeli so dovolj hitri, da z zmogljivo grafično kartico posnetke obdelajo sproti pri hitrosti vsaj 40 slik na sekundo. Testiranje je potekalo na grafični kartici RTX 3070 Ti.

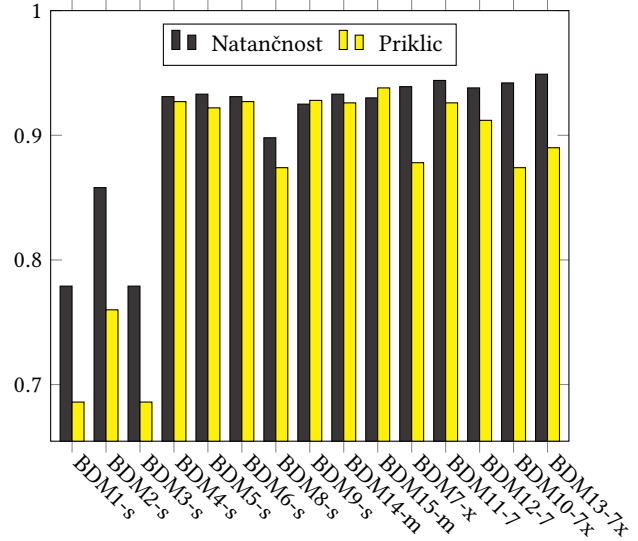


Slika 2: Povprečne hitrosti modelov.

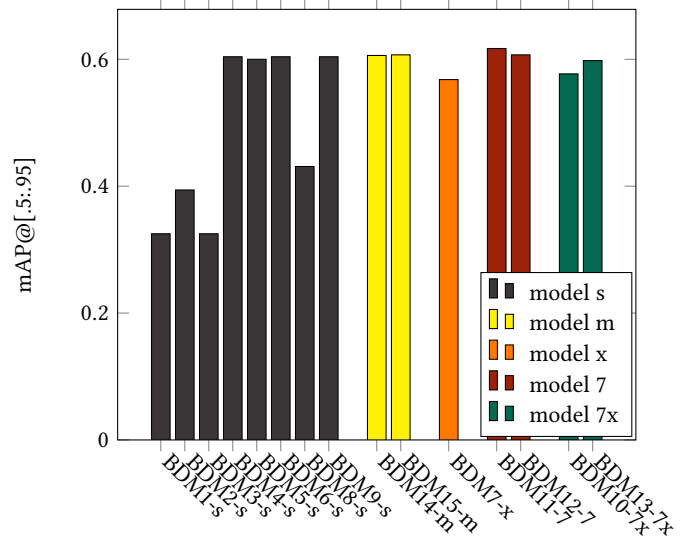
3.2 Uspešnost detekcije modelov

Pri primerjavi modelov po priklicu in natančnosti (Slika 3) vidimo, da bistveno izstopata modela BDM1-s in BDM3-s, saj več kot tretjine čebel ne zaznata, petina zaznav je pa napačnih. Slabi rezultati so očitni tudi po merilu mAP . Merilo $mAP@[.5:.95]$ najbolje prikaže dejansko razliko med modeli (Slika 4).

Vsi slabši modeli imajo arhitekturo YOLOv5s, vendar imajo drugi YOLOv5s modeli (BDM4-s, BDM5-s in BDM6-s) dobro uspešnost. Ni neposredne povezave med arhitekturo in kvaliteto modela. Najbolje se je izkazal model BDM11-7. 8 modelov ima podoben $mAP@[.5:.95]$, malo nad 0.6. Kljub močno počasnejšemu



Slika 3: Natančnost in priklic modelov strojnega učenja



Slika 4: Kvaliteta zaznavanja na testni množici po metriki $mAP@[.5:.95]$

času izvajanja (Slika 2), je kvaliteta zaznav YOLOv7 primerljiva z modeli YOLOv5.

Pri slabših modelih je najpogostejša napaka združevanje več čebel v eno in zaznavanje grč. To lahko najlažje vidimo na primerjavi med najboljšim in najslabšim modelom (Slika 5).

3.3 Rezultati sledenja

Pri primerjavi natančnosti štetja vidimo, da je Norfair bolj natančen od ByteTracka (Slika 6). Norfair[5] se je še posebej izkazal pri štetju čebel, ki vstopajo v panj, saj na posnetku z več čebelami od realnih rezultatov odstopa le za 2,5 %. Obe knjižnici zaznamuje slabše štetje čebel, ko se oddaljujejo od panja. Sledenje je bilo v vseh primerih hitrejše od 50 slik/sekundo (20 ms na sliko). Norfair je rahlo počasnejši na posnetku z več čebelami in rahlo hitrejši na posnetku z manj čebelami, vendar je ta razlika majhna, okoli 0,8 milisekunde.

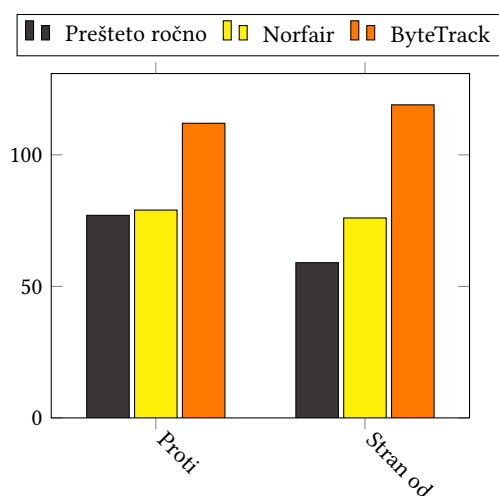


(a) Zaznave najboljšega modela BDM11-7



(b) Zaznava najslabšega modela BDM1-s

Slika 5: Primerjava boljšega in slabšega modela



Slika 6: Število letov proti in stran od vhoda v panj, prešteti na posnetku z več čebelami.

4 DISKUSIJA

Obstajajo različni načini opazovanja gibanja, od preprostega sistematičnega opazovanja do uporabe modelov strojnega vida. Kljub sorazmerno visoki natančnosti modelov, ki smo jih naučili menimo, da bi se dalo naučiti še bolj natančne modele. Večje število podatkov in boljše eksperimentalno okolje (grafična kartica z več grafičnega pomnilnika), bi omogočila še boljše rezultate. Eno od možnosti predstavlja uporaba novejših arhitektur YOLOv8[9], ki je bila objavljena med izdelovanjem naloge in je zato še nismo uporabljali. Obstaja tudi možnost izvajanja programa na

preprostejši, specializirani strojni opremi, kar bi omogočilo zaznavo na kraju, kjer so bili posnetki pridobljeni[10]. Uspešnost zaznave in sledenja bi lahko izboljšali tudi s sestavo poligona, ki omeji hitrost gibanja čebel na posnetem območju. Z vključitvijo termalne kamere v zaznavo, bi lahko model še izboljšali. Ne nazadnje, bi lahko natančnost algoritmičnega sledenja gibanja izboljšali z uporabo umetne inteligence, z učenjem nevronske mreže, ki bi zagotavljala boljše časovno doslednost sledenja.

5 ZAKLJUČEK

Štetje čebel, ki priletijo iz oziroma v panj lahko čebelarjem nudi vpogled v stanje panja. Avtomatizirano zbiranje podatkov o čebelah ni uporabno le za čebelarje, ampak lahko tudi podpira mnogo drugih raziskav o vedenju in delovanju teh čudovitih živali. Z uporabo računalniškega vida in metodami sledenja večim objektom (MOT) nam je uspelo dokazati, da je taka avtomatizacija mogoča, natančna in uporabna.

ZAHVALA

Zahvalili bi se radi naši mentorici Mag. Darji Silan, ki nam je nudila podporo in pomagala organizirati celotno izdelavo naloge. Radi bi se tudi zahvalili somentorju Dr. Janku Božiču, ki nam je pomagal s strokovno literaturo in splošnim znanjem na tem področju ter za posnetke čebel.

LITERATURA

- [1] Rač. prog. Glenn Jocher in sod., ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation nov. 2022. URL: <https://github.com/ultralytics/yolov5> Retrieved 26. feb. 2023 from.
- [2] Chien-Yao Wang, Alexey Bochkovskiy in Hong-Yuan Mark Liao. 2022. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. (jul. 2022). <https://github.com/WongKinYiu/yolov7>.
- [3] 2023. Urbani učni čebelnjak, botanični vrt ljubljana. (2023). Retrieved 28. feb. 2023 from <http://www.botanicni-vrt.si/urbani-ucni-cebelnjak>.
- [4] Rač. prog. Brad Dwyer, Joseph Nelson (2022), Jacob Solawetz in sod., ver. 1.0. URL: <https://roboflow.com> Retrieved 26. feb. 2023 from.
- [5] Rač. prog. Tryolabs, Norfair. URL: <https://github.com/tryolabs/norfair> Retrieved 26. feb. 2023 from.
- [6] Yifu Zhang, Peize Sun, Yi Jiang, Dongdong Yu, Fucheng Weng, Zehuan Yuan, Ping Luo, Wenyu Liu in Xinggang Wang. 2022. Bytetrack: multi-object tracking by associating every detection box.
- [7] Rač. prog. Piotr Skalski, onemetric 2021. URL: <https://github.com/SkalskiP/onemetric> Retrieved 26. feb. 2023 from.
- [8] Mark Everingham in John Winn. 2011. The pascal visual object classes challenge 2011 (voc2011) development kit. *Pattern Analysis, Statistical Modelling and Computational Learning, Tech. Rep.*, 8.
- [9] Rač. prog. Glenn Jocher, Ayush Chaurasia in Jing Qiu, YOLO by Ultralytics ver. 8.0.0, jan. 2023. URL: <https://github.com/ultralytics/ultralytics> Retrieved 26. feb. 2023 from.
- [10] Ratko Pilipović, Vladimir Risojević, Janko Božič, Patricio Bulić in Uroš Lotrič. 2021. An approximate gemm unit for energy-efficient object detection. *Sensors*, 21, 12. <https://www.mdpi.com/1424-8220/21/12/4195>.

Vpeljava virtualnega asistenta ChatGPT v medicinsko platformo

Implementation of a Virtual Assistant ChatGPT into the Medical Platform

Matic Zadobovšek
matic.zadobovsek@gmail.com
Univerza v Ljubljani
Fakulteta za računalništvo in
informatiko
Večna pot 113
Ljubljana, Slovenija

Primož Kocuvan
primoz.kocuvan@ijs.si
Institut "Jožef Stefan"
Jamova cesta 39
Ljubljana, Slovenija

Matjaž Gams
matjaz.gams@ijs.si
Institut "Jožef Stefan"
Jamova cesta 39
Ljubljana, Slovenija

POVZETEK

V prispevku predstavimo vpeljavo ChatGPT v platformo za elektronsko in mobilno zdravje Insieme, ki uporabnikom omogoča učinkovito pridobivanje informacij s področja medicine, spletno človeško pomoč s strani zdravstvenih izvedencev in uporabo virtualnega asistenta, ki je nastal z integracijo najnovejših tehnologij na področju obdelave naravnega jezika. Opišemo delovanje platforme Insieme in podamo razlago ter opis implementacije virtualnega asistenta. Prototipna vpeljava ChatGPT služi testiranju zmogljivosti z namenom revolucije slovenskega zdravstva.

ABSTRACT

In this paper, we present the introduction of ChatGPT into the Insieme platform for electronic and mobile health, which enables users to efficiently acquire information in the field of medicine, receive online human assistance from healthcare professionals, and utilize a virtual assistant created using the latest natural language processing technologies. We report on the functionalities of the Insieme platform and provide an explanation and description of the implementation of the virtual assistant. The prototype implementation of ChatGPT into a medical platform serves as a test for potential advancement of the Slovenian healthcare system.

KLJUČNE BESEDE

virtualni asistenti, vektorske podatkovne baze, besedne vložitve, GPT-4, obdelava naravnega jezika

KEYWORDS

virtual assistants, vector databases, word embeddings, GPT-4, natural language processing

1 UVOD

V današnjem svetu, kjer se količina podatkov in informacij nenehno povečuje, je dostop do zanesljivih virov informacij in strokovnih nasvetov postal ključnega pomena. Še posebej na področju medicine, ki je eno izmed temeljnih področij družbe, je pomembno, da uporabnikom zagotovimo enostaven in učinkovit

način za pridobivanje zdravstvenih informacij. Na spletu in mobilnih telefonih je vrsta medicinskih aplikacij, ki nudijo informacije in nasvete. Termin 'Dr. Google' opisuje iskanje zdravniških informacij in diagnosticiranja s pomočjo iskalnika Google.

Na drugi strani pa ChatGPT predstavlja zelo obetavno orodje, ki je neprestano dostopno vsakomur, kar omogoča uporabnikom, da lahko dobijo zanesljive odgovore na svoja vprašanja v realnem času. S hitrim napredkom tehnologije in nenehnim izboljševanjem ChatGPT se zdi, da bo uporaba le še rasla.

V okviru projekta Interreg je bila razvita platforma za elektronsko in mobilno zdravje Insieme, ki jo je razvilo nekaj partnerjev, ključni del pa je prispeval Odsek za inteligentne sisteme na Institutu 'Jožef Stefan'. Platforma ne le olajšuje iskanja zdravstvenih storitev, ampak hkrati ponuja spletno človeško pomoč s strani zdravstvenih izvedencev, pridobitev dodatnih koristnih informacij in ogled video vsebin s področja medicine. Ključno vlogo na tej platformi pa ima virtualni asistent, ki temelji na najnovejših dosežkih na področju obdelave naravnega jezika. V platformi so virtualni asistenti starejše generacije, pred nastankom ChatGPT, tj. botov z generativno splošno inteligenco.

V nadaljevanju je opisano delovanje platforme Insieme, vloga virtualnega asistenta ChatGPT za medicino ter razlaga, kako smo združili tehnološko znanje in domensko strokovnost, da bi zagotovili natančne in uporabniku prijazne odgovore na vprašanja s področja zdravja. Predstavljene so besedne vložitve in razlogi za uporabo vektorskih podatkovnih baz, vse skupaj pa je povezano z orodjem LangChain in velikim jezikovnim modelom GPT-4. Vse omenjene enote omogočajo, da razvijamo zdravstvene aplikacije, ki prej niso obstajale in močno presegajo npr. dr. Googla. Ta študija je služila tudi kot prvi preizkus nove tehnologije ChatGPT za medicinske namene.

2 CHATGPT

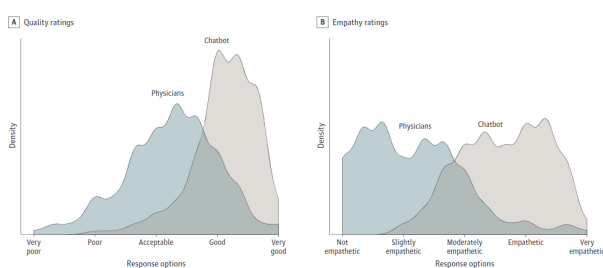
Po nekaj testih s konkurenčnimi produkti smo se odločili za uporabo velikega jezikovnega modela GPT-4, ki smo ga obogatili z znanjem naše platforme Insieme. GPT-4 je bil razvit marca 2023 in predstavlja velik napredek na področju obdelave naravnega jezika, predvsem pa je koristen za odgovarjanje na vprašanja, generiranje besedil in prevajanje v druge jezike. V primerjavi z modeli prejšnjih generacij se je povečala zanesljivost in pravilnost odgovorov, hkrati pa je izboljšano upravljanje glede na uporabnikove ukaze (naprimer, da povemo, v kakšnem slogu naj bo generirani odgovor). Številna testiranja [6] (različni izpiti in preizkusi znanja z različnih področij) so pokazala, da GPT-4 dosega rezultate, ki so povsem primerljivi s tistimi, ki jih dosegamo ljudje. V primerjavi z zdravniki je general daljše odgovore, ki

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Information Society 2023, 9–13 October 2023, Ljubljana, Slovenia

© 2023 Copyright held by the owner/author(s).

so hkrati bili s strani ostalih zdravstvenih izvedencev označeni kot boljši tako po kvaliteti odgovorov kot tudi empatiji [2]. Posledično smo se odločili, da bo naš virtualni asistent temeljil na modelu GPT-4. Namesto spreminjanja obstoječega modela smo omogočili, da lahko uporabnik povprašuje po podatkih, ki so pridobljeni iz naše platforme in ostalih dokumentov, ki jih lahko poljubno mnogo podamo. Na ta način ločimo jezikovni model in našo bazo znanja, omogočimo uporabniku, da komunicira s podanimi dokumenti, pri generiranju odgovora pa so uporabljene zgolj informacije, ki se nahajajo znotraj naših podanih dokumentov, kar omogoča, da se uporabniku zagotovi najbolj relevanten odgovor. Z omenjenim pristopom lahko enostavno dodajamo nove vire informacij in prilagodimo model za specifične naloge; brez treniranja obstoječega modela, kar bi sicer seveda bilo časovno in računsko zahtevno. Postopek je podrobneje opisan v nadaljevanju referata.



Slika 1: V več testih je ChatGPT presegel človeške zdravniške. Vir: John Ayers in sod. 2023. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA internal medicine, 183, (apr. 2023).

3 INSIEME

Za testno medicinsko platformo smo izbrali Insieme, ki je bila nedavno razvita v sodelovanju s slovenskimi in italijanskimi partnerji v okviru čezmejnega projekta ISE-EMH [7]. Opremljena je s prijaznim uporabniškim vmesnikom, ki omogoča, da lahko uporabnik na enem spletnem mestu na enostaven in eleganten način pridobi koristne informacije s področja zdravstva.

Glavne funkcionalnosti so možnost iskanja storitev z uporabo stranske menijske vrstice oziroma z uporabo iskalne funkcije, spletna človeška pomoč (klepet v živo s klicnim centrom ali zdravstvenim izvedencem), ogled video vsebin s področja zdravja ter uporaba virtualnega pomočnika, ki je v nadaljevanju članka predstavljena kot osrednja tematika. Vsa omenjena vsebina je uporabniku na voljo v treh oz. štirih jezikih.

Na levi strani slike 2 se nahaja seznam storitev, ki jih ponuja platforma Insieme. Ta menijska vrstica omogoča izbiranje med različnimi vejami medicine, zatem pa se nadaljnji izbor razširi na bolezni in bolezenska stanja, ki pripadajo izbrani veji medicine, poleg tega pa so prikazane še informacijske storitve, ki pripadajo izbrani specializaciji medicine. S klikom na eno izmed bolezenskih stanj je uporabnik preusmerjen na ustrezno podstran. Tu so mu na voljo bistveni podatki o poteku bolezni, simptomih, morebitni preventivi in nadaljnjemu ukrepanju. Spodaj se nahaja še več preusmeritev na zunanja spletna mesta, ki uporabniku omogočijo, da pridobi ustrezno znanje o izbrani bolezni.

Poleg ročnega prehajanja med podstranmi platforme je uporabniku na voljo še funkcionalnost iskanja, ki uporabniku prikaže vse storitve na platformi, ki so ujemajoče glede na niz iskanja.



Slika 2: Prikaz glavne strani ob obisku platforme Insieme.

Uporabnikom je na voljo tudi spletna človeška pomoč. Na vstopni strani so nanizani klicni centri in aktivni uporabniki, ki jih lahko kontaktirate preko spletnega klepeta v živo, ki je vgrajen v platformo.

Insieme platforma nudi več vgrajenih asistentov: čakalne vrste, IJS asistenta, iskanje po storitvah, virtualnega asistenta za medicino, hkrati pa so priložene še povezave do ostalih ne-vgrajenih asistentov. Npr. asistent za čakalne vrste omogoča, da vnesemo ime posega oziroma storitve, podamo približno nujnost, kdaj izbrani poseg potrebujemo ter zeleno regijo v Sloveniji za opravljanje posega. Asistent za medicino odgovarja na poljubna vprašanja uporabnika s področja zdravstva, kot odgovor pa mu poda ustrezne napotke in nasvete.

4 VIRTUALNI ASISTENT ZA MEDICINO

4.1 Ozadje

Virtualni asistent v platformi Insieme je namenjen odgovarjanju na vprašanja o zdravju. Obstoječe asistente smo dogradili z asistentom tipa ChatGPT. Tak asistent ima ogromno svojega splošnega znanja s spleta, na voljo pa ima tudi dodatne lokalne informacije, povezane s projektom Insieme. Prvi problem pri uporabi je, če bi želeli, da bi podali neko večjo količino besedila (morda kar celo knjigo), ali pa imamo več dokumentov, ki bi jih radi uporabili. Veliki jezikovni modeli imajo običajno omejitve, koliko besedila lahko sprejmejo [4]. Zato je pomembno, da velikemu jezikovnemu modelu podamo le informacije, ki so bistvene. Pri tem so ključne besedne vložitve in vektorske podatkovne baze.

4.2 Besedne vložitve in vektorske podatkovne baze

Vložitve (angl. embeddings) so način, kako lahko predstavimo besede, povedi ali pa kar celotne dokumente. Za njihov izračun potrebujemo ustrezne modele, ki so bili trenirani na ogromni količini podatkov in znajo poiskati razmerja med besedami s pomočjo analiziranja vzorcev, ki se pojavljajo v podatkih [3]. V našem primeru smo uporabili model, ki ga ponuja OpenAI – text-embedding-ada-002. Na ta način, da pridobimo vektor za vsako izmed besed, lahko predstavimo pomen besedila. Besedne vložitve lahko predstavimo v večdimenzionalnih prostorih, kjer so si besede oziroma povedi s podobnim pomenom blizu – med

vektorji lahko izračunamo razdalje in tako poiščemo pomensko sorodne besede.

Vektorske podatkovne baze shranjujejo informacije v obliki vektorjev, kar pogosto imenujemo kar besedne (vektorske) vložitve. To omogoča, da lahko indeksiramo in preiščemo ogromno količino nestrukturiranih podatkov, kot so slike, surovo besedilo ali pa senzorski podatki. Vektorska baza organizira podatke z uporabo visoko-dimenzionalnih vektorjev, ki vsebujejo nešteto dimenzij, vsaka dimenzija pa opisuje točno določeno lastnost podatkovnega objekta, ki ga predstavlja. Vektorske baze se torej od tradicionalnih baz, ki shranjujejo podatke v tabelarični obliki, razlikujejo v tem, da vrnejo rezultate na podlagi podobnosti (tradicionalne baze vrnejo popolnoma ujemajoče se objekte) [9]. Za merjenje podobnosti med vektorji v vektorskem prostoru se lahko uporablja različne mere — pogosto uporabljamo kosinusno podobnost. Te mere uporabljamo, da lahko primerjamo vektorje, ki so shranjeni v naši vektorski bazi, in poiščemo tiste, ki so najbolj podobni vektorju, ki ustreza vnosu uporabnika. Omogočajo torej delo s kompleksnimi podatki in hitro iskanje, kar bi sicer tradicionalnim bazam povzročalo težave. Recimo, da imamo dokument, ki bi ga radi indeksirali. Uporabili bomo model, ki omogoča ustvarjanje besednih vložitev (zgoraj smo omenili `text-embedding-ada-002`) [5]. Shranili jih bomo v izbrano vektorsko bazo, pri tem pa se shrani referenca na dokument, iz katerega je bila vložitev ustvarjena. Kadarkoli bo naš uporabnik poslal poizvedbo, bo uporabljen enak model za ustvarjanje vložitev — uporabili jih bomo, da v vektorski bazi poiščemo najbolj podobne besedne vložitve, ki so zaradi omenjene reference povezane z originalnim dokumentom, kjer so bile ustvarjene. Pridobljene dokumente lahko zatem podamo velikemu jezikovnemu modelu — dokumenti bodo uporabljeni kot kontekst za generiranje odgovora. Ravno zaradi vseh omenjenih lastnosti so vektorske podatkovne baze odlična izbira, da obogatimo naše generativne modele.

4.3 Implementacija

Eno izmed pomembnejših orodij za implementacijo ChatGPT v Insieme je knjižnica LangChain, ki omogoča delo z velikimi jezikovni modeli (LLM). LLM lahko učinkovito opravljajo ogromno število različnih opravil, vendar obstaja verjetnost, da ne bodo zmožni pravilno odgovarjati na vprašanja s specializiranih področij, kot je na primer medicina. LangChain pomaga, da lahko naše modele nadgradimo z znanjem specifičnih področij in jim omogočimo zavedanje o podatkih ter kontekstu pogovora. LangChain predstavlja zmogljivo orodje, ki zapolnjuje praznino v povezavi med jezikovnimi modeli in domenskim znanjem, kar je tudi razlog, zakaj se LangChain vse pogosteje uporablja v aplikacijah, ki opravljajo naloge, povezane z obdelavo naravnega jezika. LangChain vsebuje številne module, ki pomagajo pri razvoju [1]:

- LLM: omogoča uporabo zmožnosti specifičnega velikega jezikovnega modela.
- Verige (angl. chains): glavna enota, kar je razvidno že iz imena LangChain, ki združuje več LLM klicev. Primer tega bi bil, da najprej preberemo uporabnikov vnos, ta vnos pa uporabimo, da sestavimo nov vnos (angl. prompt), ki se poda velikemu jezikovnemu modelu, ki zatem generira odgovor.
- Vnosi, pozivi (angl. prompts): LangChain omogoča veliko različnih načinov, kako lahko spreminjamo vnos, ki je posredovan jezikovnemu modelu. Lahko uporabljamo

predloge (angl. prompt templates), kjer točno definiramo obliko vnosa.

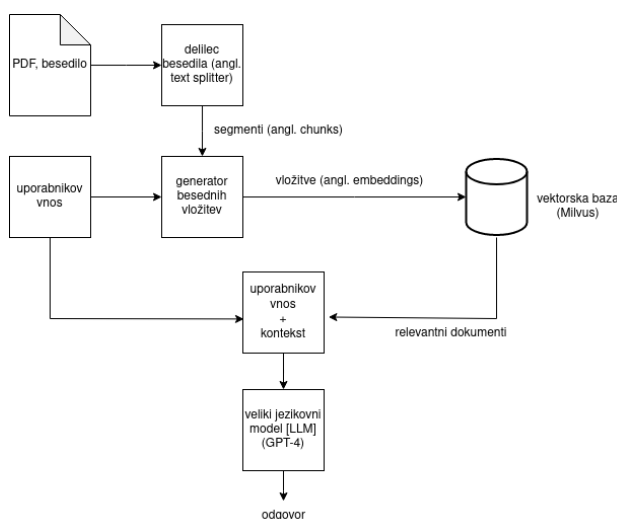
- Moduli za nalaganje dokumentov: omogočajo pretvorbo različnih vrst podatkov (PDF dokumenti, HTML spletne strani, slikovno gradivo) v besedilo, ki ga je mogoče procesirati.
- Agenti: za aplikacije, kjer zaporedje klicev ni določeno vnaprej, LangChain zagotavlja agente, ki lahko ukrepajo na podlagi vhodov, namesto da bi imeli vnaprej določeno zaporedje.

Naslednja bistvena komponenta sistema so vektorske podatkovne baze, ki so predstavljene v razdelku 4.2. Odločili smo se za uporabo odprtokodne vektorske baze Milvus, ki omogoča učinkovito shranjevanje vektorjev, njihovo indeksiranje, hkrati pa ponuja API (angl. application programming interface), ki omogoča enostavno integracijo z različnimi programskimi jeziki. Tudi sama povezava med vektorsko bazo Milvus in jezikovnim modelom GPT-4 je preprosta, saj ni potrebe, da bi podatke posebej označevali ali pa ponovno trenirali model. Podatke je potrebno pretvoriti v vektorsko obliko in jih shraniti v Milvus. Končni odgovor, ki ga model generira, je na takšen način ustvarjen z referenciranjem vsebine v naši zbirki dokumentov, kar zagotavlja, da virtualni asistent pridobi prave podatke in posledično zmanjša verjetnost napak.

Slika 3 grafično nakazuje postopek. Prvi korak v tem razvoju je nalaganje podatkov v 'Dokumente', ki so pravzaprav besedilni kosi. Modul za nalaganje dokumentov v orodju LangChain poenostavi to nalogo in omogoča enostavno nalaganje in predobdelavo naših podatkov — uporabimo lahko `DirectoryLoader`, ki omogoča, da shranimo vse uporabljene dokumente v skupen direktorij. Sledi razdeljevanje dokumentov na manjše kose — delilec besedila (angl. text splitter) omogoča razbijanje dolgih besedilnih delov na manjše, pomensko smiselne koščke [8]. Ta naloga se morda zdi preprosta, vendar vključuje nekaj kompleksnosti. Cilj je besedilo razdeliti na način, ki ohranja pomensko povezane dele skupaj, pri čemer je 'pomenska povezanost' odvisna od vrste besedila, ki se obdeluje. Delilci besedila razdelijo besedilo na majhne kose, pogosto na podlagi mej med stavki. Te majhne kose združijo v večje kose, dokler ne dosežejo določene velikosti, ki jo določi predhodno določena funkcija za merjenje velikosti kosa — ko doseže kos želeno velikost, postane samostojen kos besedila. Zatem se ustvari nov kos z nekaj prekrivanja (angl. chunk overlap), da se ohrani kontekst med posameznimi kosi.

Sledi generiranje besednih vložitev, ki igrajo ključno vlogo pri predstavitvi besedilnih informacij. Razred `Embedding` v orodju LangChain služi kot standardiziran vmesnik za različne ponudnike vložitev, vključno z OpenAI. S pomočjo generiranja besednih vložitev se besedilo pretvori v vektorsko predstavitev, ki omogoča semantično analizo in opravljanje nalog, kot je semantično iskanje. Vse to se s pomočjo vgrajenih metod shrani v našo vektorsko podatkovno bazo kot nov indeks — nad tem objektom lahko opravljamo semantično iskanje in pridobimo dokumente, ki so relevantni glede na uporabnikov vnos. Dobljeni dokumenti se zatem posredujejo jezikovnemu modelu, ki dokumente obravnava kot kontekst za generiranje odgovora. Na sliki 4 lahko vidimo prikaz odgovarjanja na uporabnikova vprašanja. ChatGPT tekoče odgovarja na zdravstvena vprašanja z upoštevanjem splošnega znanja, znanja medicine s celotnega spleta in posebnih znanj s platforme Insieme.

Preverjanje in zagotavljanje informativnih ter ustreznih odgovorov je potekalo s strani avtorjev prispevka. ChatGPT je bil testiran v vprašanji, ki so v celoti pokrili znanje, ki ga vsebuje platforma Insieme. Odgovore, ki so bili generirani, smo kritično ovrednotili in uvedli popravke ob morebitnemu odstopanju od pričakovanj. Točnost pridobljenih odgovorov je za uporabnika ključnega pomena, saj lahko zavajanje in netočni odgovori v primeru, da jih uporabnik upošteva, v hujših situacijah privedejo celo do poslabšanja zdravstvenega stanja. Vsebina platforme Insieme, ki virtualnemu asistentu predstavlja kontekst za generiranje odgovorov, je bila predhodno pripravljena s strani zdravstvenih izvedencev, kar uporabniku omogoča pridobitev preverjenih in koristnih informacij.



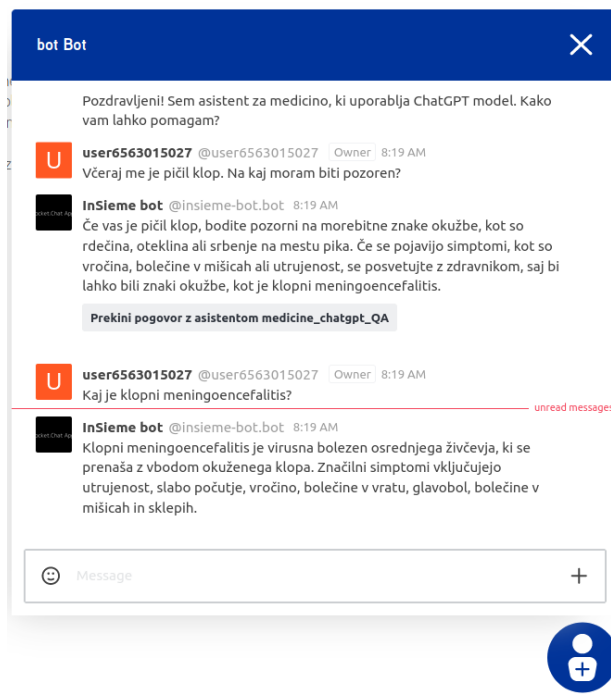
Slika 3: Prikaz integracije vseh komponent sistema.

5 ZAKLJUČEK

Predstavljena je implementacija pogovornega virtualnega asistenta ChatGPT v platformo za elektronsko in mobilno zdravje, imenovano Insieme. Eksperimenti kažejo, da je sistem sposoben nuditi tako običajne odgovore ChatGPT kot dodatne na osnovi informacij, dosegljivih s platforme. Vpeljava je bila uspešna.

Platforma ima še veliko možnosti za nadaljnje izboljšave, predvsem bi bila koristna integracija z drugimi zdravstvenimi informacijskimi sistemi in viri v Sloveniji, kar bi omogočilo, da bi platforma Insieme postala še bolj celovit vir informacij za uporabnike – pri tem se bomo povezali z zdravstvenimi institucijami po Sloveniji. V okviru vpeljave virtualnega asistenta je bil preizkušen tudi najšibkejši jezikovni model iz družine Llama 2, ki je bil lokalno nameščen. Pojavila se je težava, saj je večina učnih podatkov, ki so bili uporabljeni, bila v angleščini, zato je model kot tak neustrezen za uporabo v slovenskem jeziku. Tu se nam odpira možnost, da v prihodnje preizkusimo zmogljivejši model iz družine Llama 2, ki bo prav tako lokalno nameščen. Pri tem se pojavlja nov vidik uporabe, saj bi na ta način vsi podatki bili dostopni lokalno in bi izpodrinili potrebo po zunanemu dostopanju do podatkov, kot je to sedaj nujno z uporabo ChatGPT.

Naš glavni namen je, da bi platforma predstavila temeljne ideje, kako bi se lahko moderniziral zdravstveni sistem, pri tem pa stremimo k razbremenitvi ljudi, ki delajo v tej stroki, hkrati pa želimo vsem uporabnikom omogočiti dostop do učinkovitega in



Slika 4: Prikaz delovanja virtualnega asistenta ChatGPT nad zdravstveno platformo Insieme.

nenehno dostopnega vira informacij, ki temelji na najsodobnejših pridobitvah raziskav.

Eksperimenti v tej študiji kažejo, da je generativna umetna inteligenca dejansko uporabna in obeta radikalne izboljšave, če jo bomo uspeli vpeljati v slovensko zdravstvo.

ZAHVALA

Platforma Insieme je nastala v okviru čezmejnega projekta Interreg ISE-EMH, ki ga financira Program sodelovanja Italija-Slovenija iz Evropskega sklada za regionalni razvoj.

LITERATURA

- [1] Amogh Agastya. 2023. Harnessing retrieval augmented generation with langchain. (Sep. 2023). Retrieved September 5, 2023 from <https://betterprogramming.pub/harnessing-retrieval-augmented-generation-with-langchain-2eae65926e82>.
- [2] John Ayers in sod. 2023. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA internal medicine*, 183, (apr. 2023). DOI: 10.1001/jamainternmed.2023.1838.
- [3] Qilu Jiao in Shun Yao Zhang. 2021. A brief survey of word embedding and its recent development. V *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*. Zv. 5, 1697–1701. DOI: 10.1109/IAEAC50856.2021.9390956.
- [4] Humza Naveed, Asad Ullah Khan, Shi Qiu, Muhammad Saqib, Saeed Anwar, Muhammad Usman, Naveed Akhtar, Nick Barnes in Ajmal Mian. 2023. A comprehensive overview of large language models. (2023). arXiv: 2307.06435 [cs. CL].
- [5] Arvind Neelakantan in sod. 2022. Text and code embeddings by contrastive pre-training. (2022). arXiv: 2201.10005 [cs. CL].
- [6] OpenAI. 2023. Gpt-4 technical report. (2023). arXiv: 2303.08774 [cs. CL].
- [7] Insieme Platform. 2023. Insieme platform. (Sep. 2023). Retrieved September 3, 2023 from <https://ise-emh.eu>.
- [8] Adith Sreeram A S in Pappuri Jithendra Sai. 2023. An effective query system using llms and langchain. *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT)*, 12, 06, (jun. 2023).
- [9] Jianguo Wang in sod. 2021. Milvus: a purpose-built vector data management system. V *Proceedings of the 2021 International Conference on Management of Data (SIGMOD '21)*. Association for Computing Machinery, Virtual Event, China, 2614–2627. ISBN: 9781450383431. DOI: 10.1145/3448016.3457550.

Types of Democracy Defined: Keyword Extraction from Eleven Different Text Descriptions

Žiga Kolar
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenia
ziga.kolar@ijs.si

Andrej Vozlič
Independent Researcher
Kersnikova ulica 3
Ljubljana, Slovenia
hommage.artists@gmail.com

Andrej A. Lukšič
Faculty of Social Sciences
Kardeljeva ploščad 5
Ljubljana, Slovenia
andrej.luksic@fdv.uni-lj.si

Matjaž Gams
Jožef Stefan Institute
Jamova cesta 39
Ljubljana, Slovenia
matjaz.gams@ijs.si

ABSTRACT

The various forms of democracy worldwide display a range of distinct traits, influenced by regional, historical, and cultural backgrounds. This research undertook the challenge of extracting keywords from textual descriptions of eleven distinct democracy types to identify and categorize their core principles and nuances. By employing Natural Language Processing (NLP) techniques, specifically, the Text Rank, RAKE and TF-IDF (Term Frequency-Inverse Document Frequency) methodologies, the study aimed to extract meaningful keywords that encapsulate the foundational principles governing each democracy type. The results provide a deeper understanding of how various democratic structures operate and differentiate from one another. Such insights are crucial for researchers, policymakers, and political enthusiasts to navigate the intricate landscape of global democratic models.

KEYWORDS

keywords, data mining, natural language processing, democracy

1 INTRODUCTION

Democracy, a term rooted in the Greek words "dēmos" (people) and "kratos" (power), has been a guiding principle of governance for centuries [18]. However, the interpretation and application of democracy have evolved and diversified across regions, taking shapes influenced by historical experiences, cultural nuances, socio-political challenges, and economic contexts. Today, the world does not see a monolithic form of democracy but rather an array of forms, each with its unique characteristics and mechanisms [1].

Understanding these distinct democratic forms is crucial for several reasons. Firstly, it offers a glimpse into the diverse ways in which societies prioritize and ensure the participation of their citizenry in governance. Secondly, it helps in identifying the checks and balances incorporated in each system to protect against potential excesses and abuses of power. Lastly, it provides a roadmap for nations looking to either adopt a democratic model or refine

their existing systems in response to changing circumstances and needs [6].

However, given the vast amount of textual data describing these democratic forms, there arises a need for a systematic approach to identify their defining features. Keyword extraction, an established technique in the realm of Natural Language Processing (NLP) [4], offers a solution. By extracting keywords from descriptions of different democratic models, we aim to distill their essence, thereby providing a concise yet comprehensive overview of each type's foundational principles.

Keyword extraction serves as a fundamental tool for information retrieval. Researchers, academics, policymakers, and practitioners often rely on keywords to quickly identify relevant documents amidst an overwhelming volume of textual data. By extracting keywords from eleven different text descriptions of democracy, this research facilitates more effective and precise information retrieval in the realm of democratic studies. Un addition, by extracting keywords, this paper helps enhance the accessibility and comprehension of your research findings. It enables readers to gain a rapid overview of the various democracy types under study, making it easier for them to delve deeper into specific aspects of interest.

In this paper, we delve deep into the intricacies of eleven specific democratic models, unraveling their characteristics through the lens of extracted keywords. The structure of this paper unfolds in this manner: Section 2 delves into concise explanations of various democratic types. In Section 3, we describe and apply machine learning algorithms for keyword extraction to the realm of text descriptions associated with types of democracy. In Section 4, we present the results and provide commentary on them. We wrap up the paper with a conclusion and suggestions for subsequent research.

2 DEMOCRACY TYPES

In this section, we explore various forms of democracy, highlighting their unique characteristics. Democracy, which promotes individual freedoms and collective decisions, differs based on context. These descriptions, initially from a PhD study [19], are condensed here for clarity. The dataset is available at [10]. Each text description corresponds to one democratic type and is characterized by an English text description containing a maximum of 10,000 words. It's worth noting that while some democratic terms discussed here are widely accepted, others still lack a uniform definition.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
Information Society 2023, 9–13 October 2023, Ljubljana, Slovenia
© 2023 Copyright held by the owner/author(s).

2.1 Participative Democracy

Participative democracy involves direct citizen participation in decision-making, using methods like town halls and online platforms. While fostering inclusivity and informed decisions, it might be resource-intensive and not always ensure equal representation [17].

2.2 Deliberative Democracy

Deliberative democracy emphasizes informed discussion among citizens to make collective decisions. While it can promote nuanced decision-making and civic engagement, ensuring equal voice, especially for marginalized groups, remains a challenge [7].

2.3 Transdemocracy

Transdemocracy proposes expanding democratic principles to global governance, recognizing the need for collective action on global challenges. It seeks more direct citizen involvement in global decisions but faces challenges like reconciling diverse values and ensuring equal global representation [12].

2.4 Guided Democracy

Guided democracy, also referred to as authoritarian democracy, centralizes power with restricted participation from citizens. Though it can provide stability, it might compromise on authentic democratic liberties. There's also an inherent risk of corruption and potential power misuse [5].

2.5 Modern Direct Democracy

Modern direct democracy empowers citizens with immediate control over political decisions, using platforms like electronic voting. It encourages civic engagement and responsive governance but may struggle with equal access and ensuring well-informed choices [16].

2.6 E-democracy

E-democracy uses digital tools to enhance citizen participation in politics. This includes online voting, social media, digital petitions, and e-forums. It encourages political involvement, especially among the youth, and offers transparency. However, concerns arise regarding online security, digital literacy, and accessibility [9].

2.7 Representative democracy

In representative democracy, citizens elect officials to act on their behalf, as direct involvement in every decision isn't feasible. This system promotes informed decisions and offers stability. Yet, true representation and influence of special interests are concerns [3].

2.8 Liquid democracy

Liquid democracy merges direct and representative democracy. Citizens can vote directly or delegate their vote. This system promotes active engagement but faces challenges in equal representation and decision quality [8].

2.9 Blockchain democracy

Blockchain democracy employs blockchain technology for a secure, transparent voting process, allowing remote voting. Despite its security, potential system vulnerabilities and accessibility issues persist [13].

2.10 Source democracy

This democracy focuses on the foundational decision-making structures. It advocates transparency and collective input. However, inclusive participation and decision quality remain challenges [15].

2.11 Ideal typical democracy

This normative concept outlines the optimal features of a democracy, including universal suffrage, free elections, representation, civil liberties, rule of law, power separation, and government transparency.

3 METHODOLOGY

Our exploration delved deep into three contemporary keyword extraction methodologies, namely TextRank, Rake, and TF-IDF, each offering its unique algorithmic underpinning for processing textual data.

The TextRank methodology stands out for its holistic approach to text analysis, where the textual content undergoes a series of rigorous preprocessing steps to distill its core meaning. These steps include tokenization, which breaks the text into individual words or phrases, and the removal of irrelevant or stop words that carry little semantic value. Once the text is prepared, TextRank constructs a graph based on term co-occurrence. This graph elucidates the intricate relationships between words, uncovering the underlying semantic structure of the text. In this graph, each term becomes a node, and the strength and number of their connections determine their significance. The top-scoring terms, which act as representative keywords, find applications in various domains. These applications extend far beyond traditional keyword extraction and encompass areas such as text classification, where keywords play a pivotal role in categorizing documents, sentiment analysis, which relies on keywords to gauge the emotional tone of text, and document clustering, where keywords aid in grouping similar documents together [11].

Transitioning to RAKE (Rapid Automatic Keyword Extraction), this method lives up to its name by offering an expedient approach to keyword extraction. Unlike the graph-based structure of TextRank, RAKE's algorithmic focus lies in dissecting textual information into individual components, often referred to as phrases or keyphrases. Each of these components is then scored based on its occurrence and relation to other words within the text. This nuanced approach to keyword extraction results in the isolation of top-scoring keyphrases, providing a richer and more contextually meaningful representation of the underlying text's content. The efficiency of RAKE makes it particularly useful in scenarios where quick and accurate keyword extraction is essential for tasks such as summarization, information retrieval, and content indexing [14].

Lastly, the TF-IDF (Term Frequency-Inverse Document Frequency) methodology delves into the granular essence of a term's importance within a document and across a broader corpus. This method combines two essential components: "TF" (Term Frequency) and "IDF" (Inverse Document Frequency). The "TF" metric captures the frequency of a term within a particular document, highlighting its significance in that context. However, TF alone may emphasize commonly occurring words that appear significant in isolation but lack uniqueness.

The "IDF" component evaluates the term's rarity or uniqueness across a broader corpus of documents. This measure ensures that commonly occurring terms, which might be abundant within a

single document, are appropriately contextualized against their prevalence across multiple documents. The multiplicative result of TF and IDF offers a composite score that indicates the weight or significance of the term. Terms with notable TF-IDF scores emerge as the textual frontrunners, illuminating the primary themes and subjects of the source material. TF-IDF is widely used in information retrieval, document ranking, and text mining tasks, making it a foundational method in natural language processing and information science [2].

4 RESULTS

The essential keywords for the Text Rank, Rake, and TF-IDF algorithms have been laid out in Tables 1, 2, and 3. A striking observation is the omnipresence of the keyword "Democracy" across all types of democracies and within all three algorithms. Given its recurrent use throughout the texts, the prominence of this keyword was anticipated.

Furthermore, a pattern emerges wherein many keywords bear close resemblance or are derivatives of the nomenclature of their respective democracy types. For instance, for Deliberative democracy, we see "Deliberation" and "Deliberative" making appearances. Similarly, "Participation" and "Participatory" are highlighted for Participative democracy, "Direct" and "Modern" for Modern direct democracy, "Representative" for Representative democracy, and "Liquid" for Liquid democracy. The frequency of these terms in the texts underscores their significance, thereby warranting their selection as keywords.

The keyword "Participation" is shared among Participative democracy, Deliberative democracy, and E-democracy in all three keyword extraction algorithms. This common keyword suggests alignment in themes among these democracy forms, potentially reflecting shared core principles. In contrast, other democracy types exhibit more distinct keywords. Additionally, the keyword "Citizens" is present in Source democracy, Participative democracy, and Blockchain democracy using the Text Rank algorithm. It appears only in Representative democracy with RAKE and in both Blockchain democracy and Representative democracy with TF-IDF. These variations in results across Text Rank, RAKE, and TF-IDF stem from their distinct methodologies. Text Rank considers relational context between terms, extracting keywords with conceptual relationships. RAKE focuses on local frequency and co-occurrence, sensitive to document structure. Finally, TF-IDF emphasizes term uniqueness across a document set. These differences mean that each algorithm yields a somewhat different set of keywords, even when analyzing the same content.

In all three algorithms, a set of 12 keywords serves as unique identifiers for distinguishing each type of democracy. These keywords are "Ancient", "Rule", "Laws", "Deliberation", "Deliberative", "Modern", "Liquid", "Vote", "Information", "Blockchain", "New", and "Government". They are depicted by red color in Tables 1, 2, 3.

Further delving into the specifics, the Text Rank algorithm possesses its own exclusive set of four keywords that differentiates it from the others. These are "Theory", "Communication", "Systems", and "Power". They are represented by blue color in Table 1. Conversely, the TF-IDF algorithm also has its unique identifiers with two sole keywords: "Axis" and "Counter democracy". In Table 3, they are represented using the color green. These individual sets of keywords not only underscore the differences in each algorithm's framework but also hint at their distinctive analytical approaches.

In summary, the Text Rank algorithm ranks highest with four unique keywords. Following closely is the TF-IDF algorithm with two distinct keywords. Lastly, the Rake algorithm doesn't have any unique keywords, placing it at the bottom of this comparison. Text Rank works best because dataset contains a lot of nuanced contextual information and that's why TextRank might be better suited to extract these nuances due to its graph-based approach.

5 CONCLUSION

In this research, we delved into the intricate fabric of descriptions for eleven distinctive types of democracies. The aim was to elucidate and capture the most salient keywords that epitomize the essence of each democracy type. By employing three different keyword extraction algorithms — Text Rank, RAKE, and TF-IDF — our study has not only shed light on the specific linguistic constructs and terminologies inherent to each democratic type but also the differential efficacy of the algorithms in context.

Our findings reveal that the keyword "Democracy" consistently appears in all types of democracies across various keyword extraction methods. This is expected since each text description provides definitions for its respective democracy type, leading to the frequent use of the term "Democracy." In the analysis of democracy types, keywords often closely align with or derive from the corresponding democracy nomenclatures. Notably, the term "Participation" consistently emerges across Participative democracy, Deliberative democracy, and E-democracy in all algorithms, indicating shared principles among these democracies. However, the keyword "Citizens" varies in its occurrence across different algorithms and democracy types. These discrepancies in keyword results stem from the distinct methodologies of the Text Rank, RAKE, and TF-IDF algorithms. Each approach leads to unique sets of keywords, even when applied to the same content.

This research underscores the importance of algorithm selection tailored to the specific requirements of the textual dataset in hand. Moreover, the keywords extracted present an invaluable repository for political scientists, historians, and policymakers in understanding the foundational pillars of varied democracy forms. Future work might consider combining every three keyword extraction algorithms into one single table, removing duplicates and extracting only unique keywords that belong to the corresponding democracy type. Additionally, by converting keywords into numerical values, clustering algorithms could be applied. This would enable the identification of democracy types that fall within the same group or cluster. Furthermore, in future endeavors, experts possessing a background in political education will assess the accuracy and validity of the extracted keywords.

ACKNOWLEDGEMENTS

The authors acknowledge the funding from the Slovenian Research and Innovation Agency (ARRS), Grant (PR-10495) and Basic core funding P2-0209.

REFERENCES

- [1] Christopher H Achen and Larry M Bartels. 2017. *Democracy for realists*. In *Democracy for Realists*. Princeton University Press.
- [2] Akiko Aizawa. 2003. An information-theoretic perspective of tf-idf measures. *Information Processing & Management*, 39, 1, 45–65.
- [3] Sonia Alonso, John Keane, and Wolfgang Merkel. 2011. *The future of representative democracy*. Cambridge University Press.
- [4] KR1442 Chowdhary and KR Chowdhary. 2020. Natural language processing. *Fundamentals of artificial intelligence*, 603–649.
- [5] Peter Christoff. 2003. Ecological citizens and ecologically guided democracy. In *Democracy and green political thought*. Routledge, 159–176.

Table 1: 5 keywords for each democracy type extracted with Text Rank algorithm.

Type of Democracy	Keyword 1	Keyword 2	Keyword 3	Keyword 4	Keyword 5
Source democracy	Democracy	Society	Direct	Ancient	Citizens
Ideal-typical dem.	Democracy	Society	Rule	Executive	Laws
Participative dem.	Democracy	Political	Society	Participation	Citizens
Deliberative dem.	Democracy	Deliberation	Deliberative	Participation	Process
Modern direct dem.	Democracy	Direct	Society	Theory	Modern
Liquid democracy	Democracy	Liquid	Vote	Representative	Representatives
E-democracy	Democracy	Information	Participation	Public	Communication
Blockchain dem.	Democracy	Blockchain	New	Systems	Citizens
Representative dem.	Democracy	Representative	Government	Political	Power
Control democracy	Democracy	Civil	Democratic	Society	Monitoring
Transdemocracy	Democracy	Political	System	Concept	Form

Table 2: 5 keywords for each democracy type extracted with Rake algorithm.

Type of Democracy	Keyword 1	Keyword 2	Keyword 3	Keyword 4	Keyword 5
Source democracy	Democracy	Society	Executive	Ancient	Direct
Ideal-typical dem.	Democracy	Society	Rule	Executive	Laws
Participative dem.	Democracy	Participation	Political	Participatory	Society
Deliberative dem.	Democracy	Deliberation	Deliberative	Participation	Participatory
Modern direct dem.	Democracy	Direct	Modern	Referendum	Instrumental
Liquid democracy	Democracy	Liquid	Vote	Control	Representative
E-democracy	Democracy	Information	Participation	Public	Decision-making
Blockchain dem.	Democracy	Blockchain	New	Decentralized	Political
Representative dem.	Democracy	Representative	Government	Political	Citizens
Control democracy	Democracy	Civil	Society	Monitoring	Representative
Transdemocracy	Democracy	System	Political	Form	Representative

Table 3: 5 keywords for each democracy type extracted with TF-IDF algorithm.

Type of Democracy	Keyword 1	Keyword 2	Keyword 3	Keyword 4	Keyword 5
Source democracy	Democracy	Executive	Society	Ancient	Direct
Ideal-typical dem.	Democracy	Rule	Society	Executive	Laws
Participative dem.	Democracy	Participation	Political	Participatory	Society
Deliberative dem.	Democracy	Deliberation	Deliberative	Process	Participation
Modern direct dem.	Democracy	Direct	Modern	Instrumental	Referendum
Liquid democracy	Democracy	Liquid	Vote	Representative	Control
E-democracy	Democracy	Information	Decision-making	Participation	Axis
Blockchain dem.	Democracy	Blockchain	Citizens	New	Decentralized
Representative dem.	Democracy	Representative	Government	Political	Citizens
Control democracy	Democracy	Civil	Counter democracy	Democractic	Society
Transdemocracy	Democracy	System	Political	Concept	Form

- [6] Klaus Gründler and Tommy Krieger. 2016. Democracy and growth: evidence from a machine learning indicator. *European Journal of Political Economy*, 45, 85–107. On Institutions and Well Being. DOI: <https://doi.org/10.1016/j.ejpolco.2016.05.005>.
- [7] Amy Gutmann and Dennis F Thompson. 2004. *Why deliberative democracy?* Princeton University Press.
- [8] Anson Kahng, Simon Mackenzie, and Ariel Procaccia. 2021. Liquid democracy: an algorithmic perspective. *Journal of Artificial Intelligence Research*, 70, 1223–1252.
- [9] Marianne Kneuer. 2016. E-democracy: a new challenge for measuring democracy. *International Political Science Review*, 37, 5, 666–678.
- [10] Žiga Kolar and Matjaž Gams. 2023. Democracy dataset. Available at: https://drive.google.com/drive/folders/1W_5qa9scDXvQwR7TjBzbMan91o2-gAfe?usp=sharing. 11.9.2023. (2023).
- [11] Rada Mihalcea and Paul Tarau. 2004. TextRank: bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing*, 404–411.
- [12] Marcelo Neves. 2017. From transconstitucionalism to transdemocracy. *European Law Journal*, 23, 5, 380–394.
- [13] Peter Racsko. 2019. Blockchain and democracy. *Society and Economy*, 41, 3, 353–369.
- [14] Stuart Rose, Dave Engel, Nick Cramer, and Wendy Cowley. 2010. Automatic keyword extraction from individual documents. *Text mining: applications and theory*, 1–20.
- [15] Douglas Rushkoff. 2003. *Open source democracy: How online communication is changing offline politics*. Vol. 10753. Demos.
- [16] Theo Schiller. 2011. *Local direct democracy in Europe*. Springer.
- [17] Marius H Smit and Izak J Oosthuizen. 2011. Improving school governance through participative democracy and the law. *South African Journal of Education*, 31, 1, 55–73.
- [18] Charles Tilly. 2007. *Democracy*. Cambridge University Press.
- [19] Andrej Vozlič. 2020. *Transdemokracija kot rekonceptualizacija demokracije v okolju digitalne države: doktorska disertacija*. PhD thesis. Univerza v Ljubljani, Fakulteta za družbene vede.

Evaluation of the Effects of On-Demand Dynamic Transportation of Employees to Their Workplaces in Ljubljana

Marko Bohanec
marko.bohanec@ijs.si
Jožef Stefan Institute, Department of
Knowledge Technologies
Jamova cesta 39
Ljubljana, Slovenia

Marko Guček
marko@goopti.com
GoOpti d.o.o.
Ljubljana, Slovenia

Davor Kontić
davor.kontic@ijs.si
Jožef Stefan Institute, Department of
Environmental Sciences and Centre
for Participatory Research

Karina Sirk
karina.sirk@ipop.si
Institute for Spatial Policies
Ljubljana, Slovenia

Bernard Ženko
bernard.zenko@ijs.si
Jožef Stefan Institute, Department of
Knowledge Technologies

Martin Žnidaršič
martin.znidarsic@ijs.si
Jožef Stefan Institute, Department of
Knowledge Technologies

ABSTRACT

On-demand dynamic transportation is an innovative information-technology supported service that enables passengers to book and configure their rides. It is foreseen as a promising service to improve the sustainable mobility of citizens and alleviate traffic problems. In this paper, we present the results of a three-month pilot implementation of dynamic transportation of employees from their homes to their workplaces. The transports were managed by the company GoOpti, d.o.o., they were free and took place using vans on the routes connecting the cities Kamnik and Kranj with two areas (BTC and UKC) in Ljubljana. The project was very successful in terms of sustainable mobility: it attracted users that normally drive passenger cars, travel times were comparable to conventional modes of transportation and users were very satisfied with the service, while substantially reducing (from 30% to 70%) the harmful emissions of CO₂, NO_x and solid particles. However, two challenges for the future still remain: improving the occupancy rate of vehicles and bridging the gap between the economic price and users' willingness to pay for the service.

KEYWORDS

sustainable mobility, on-demand dynamic transport, employee transportation, Ljubljana Urban Region

1 INTRODUCTION

Within the last thirty years information technologies (IT) profoundly changed many aspects of our life and the society. IT provide us with tools that enable creation of new services, which have the potential to significantly improve our every day lives. One of such services is on-demand dynamic passenger transport [11], which typically makes use of vans or mini busses that operate without fixed itineraries or fixed stops, enabling passengers to book their ride and select their pick up and drop off locations. On-demand dynamic transport bridges the gap between the conventional public transport and private car transport, and promises to reduce green house gas and other pollutant

emissions, improve modal split, reduce traffic congestions and alleviate related problems.

The focus of this paper are not the IT needed to implement an on-demand dynamic transport (e.g., mobile apps, databases, scheduling and optimization systems), but rather we present the results of a real life evaluation of such a system that has been carried out in Ljubljana Urban Region, Slovenia, within the SmartMOVE project [10]. SmartMOVE addresses the challenges of sustainable mobility in the Ljubljana Urban Region with the capital city of Ljubljana, which is the primary destination of daily migration flows in Slovenia.

The aim of this study is to answer two questions that are crucial if such a service is to be implemented in practice: (1) Can on-demand dynamic transport really help in reducing harmful emissions and traffic congestion? and (2) Under what conditions can such a service sustainably operate in a given economic environment? It is important to note that before this study no evaluation of on-demand dynamic passenger transport in comparable environments has been done, and its benefits and limitations used to be assessed only theoretically through simulations [1].

2 RELATED WORK

Ljubljana is home to over 220,000 jobs, which accounts for over 25% of all jobs in Slovenia. As a result, over 120,000 people flock to Ljubljana daily from elsewhere. This means approximately 100,000 vehicles entering and exiting Ljubljana on a daily basis (with an average occupancy of vehicles at 1.2 pers/car). Since the majority of this is associated to the personal car transport, the main goal is to transfer the car drivers/passengers to public transport. However, the road public transport is faced with the same problems as car transport – congestions due to absence of dedicated lanes, low travel speeds, poor occupation of vehicles.

The problem of public transport in Ljubljana was tackled by studies that focused on various aspects. Some of them outline the overall context of the problem and the related phenomena [4]. The current public transport organization in Ljubljana is reflected well in a paper that describes the public transport plans from a decade ago [6]. There are studies about the state of the public transport in Ljubljana, such as accessibility [12], speed [7], studies about the effects of public transport management measures [13] and about its potential future developments.

On-demand dynamic transportation, which is in the focus of this paper, shares several characteristics with carpooling [14]. However, we could find no previous studies about on-demand

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Information Society 2023, 9–13 October 2023, Ljubljana, Slovenia

© 2023 Copyright held by the owner/author(s).

dynamic transportation in Ljubljana and even in general, the studies of on-demand dynamic transportation are to a large extent dedicated to theoretical models [8, 5] and simulations [2, 1]. A good showcase of dynamic transport practices are the on-demand airport shuttles, which have indicated a potential solution to the downsides of regular public lines; they support the users' needs and commodity as one of the most important factors for choosing the preferred means of transport.

3 STUDY DESIGN

The main objective of this pilot study was to test using on-demand dynamic transportation of employees as a sustainable alternative to existing modes of transport, especially in comparison with using passenger cars. The study took place in the trial period from February to April 2023, when selected passengers were transported free of charge by the company GoOpti, d.o.o., employing vans that can carry up to eight passengers. Transportation costs were covered from the SmartMOVE project. Two pilot routes were established that connected two nearby cities Kamnik (14,000 inhabitants, about 23 km from Ljubljana) and Kranj (38,000, 28 km) with the areas of two large employer organisations located in Ljubljana: UKC and BTC. UKC, the University Medical Centre of Ljubljana, is with 8,000 employees the largest employer in Slovenia; daily, it is visited by additional 20,000 people. BTC, the Business Trade Centre, is the largest shopping area in Slovenia. It does not have many direct public transportation links, but is located close to a highway, inducing high volumes of car traffic.

Before the study, the opportunity to join the experiment was advertised using different channels, particularly in the UKC and BTC areas. Among more than 500 interested individuals, 131 were eventually selected and invited to participate. All the operation, including the IT solution, customer management and logistics, was carried out by the project partner GoOpti.

Two data sources were collected during the study:

- *Traffic data*: Collected by GoOpti while providing the service. This included detailed data about the travelled routes (distances, times and GPS locations) and provided services (anonymized individual user's rides).
- *User survey*: Collected using a survey questionnaire once per each individual user at the end of the study period. The questions mainly addressed users' current mobility habits (with more detailed questions for users using cars) and their experience with the service. A full version of the questionnaire (in Slovene language) is available in [3].

Using this data, we carried out the following analyses: basic traffic and demographical statistics, average occupancy of vans and individual users' rides, users' current mobility habits, and user satisfaction with and willingness to pay for the service. By combining the data sources, we estimated the differences between the current and dynamic means of transportation in terms of travelling time and contribution to lower emissions of CO₂, NO_x (nitrogen oxides) and PM10 (particles with a diameter of 10 µm or less).

In order to assess environmental burdens, we analyzed the difference in air emissions between the "BaU - Business-as-Usual" scenario (i.e., trips without the introduction of dynamic transport) and the GoOpti service. We took into account the distances traveled by the types of transport that users used before the introduction of GoOpti: mainly driving passenger cars of different types (gasoline, diesel), taking into account the age of the vehicles and corresponding average emissions. Data sources included

Table 1: Basic ride statistics.

Route	Passengers [total]	Distance [km]	Time [min]	Speed [km/h]	Occupancy [%]
Kamnik	5440	21.20	42.81	29.72	38.0
Kranj	1742	33.66	45.62	44.27	25.9
Total	7182	25.18	43.70	34.57	34.1

the Statistical Yearbook of the Republic of Slovenia [9] and EU emission standards (EURO standards¹). We considered the users' distances from home to workplace and the number of journeys they would have made if dynamic transport had not been employed. As a weighting factor for the calculation, we considered whether the users mainly drive with personal vehicles (several times a week) or perhaps combine the drives with other modes of transportation to work (public passenger transport, bicycle, etc.). The emission factors for GoOpti vehicles were obtained from the manufacturer's specifications.

4 RESULTS

4.1 Traffic Data Analysis

Traffic statistics. In three months of the pilot study, there were 2,629 rides of the total distance of 66,199 km and time of 1,915 hours (almost 32 days). Here, each "ride" means picking up the passengers at one or more origin locations and dropping them down at one or more destination locations. Table 1 shows, grouped by the routes and in total, the total number of passengers and average ride distances, times, speed and vehicle occupations. The term "passenger" refers to one ride of a single person.

User statistics. The service was used by 131 individual users, who used the service daily or less. The most active user used the service 121 times, and the average was 54.88 times per user. Female users prevailed over males (73% vs. 27%). The prevailing age groups were 31–42 (43%) and 45–64 (46%), while the distribution of users' education levels was close to uniform.

Vehicle occupancy. As large as possible vehicle occupancy is essential for the effectiveness of dynamic transportation. Figure 1 displays the occupancy achieved in the study.

4.2 User Survey Analysis

The survey was completed by users at the end of the trial period, so that each user completed the survey at most once. Out of the 131 users, the completed survey was submitted by 88: 30 travelling from/to Kranj and 58 in the Kamnik direction.

Current mobility habits. Figure 2 shows relative proportions of transportation modes used by the survey respondents. The prevailing mode is using personal cars: 70% as sole drivers and 20% as fellow passengers. The train and city bus come next at approximately 30% each, the bicycle at 10%, while the remaining modes are barely indicated. Both routes exhibit similar usage patterns.

The respondents that use cars estimated their average occupancy at 1.37 per ride. Most of the cars have diesel or gasoline engines (about 45% each). Hybrid and electric cars account only for about 2%. The average age of cars is 9.15 years (7.28 on route Kranj and 10.04 on route Kamnik).

User satisfaction. The users were generally very satisfied with the service. They most appreciated the easiness and comfort of

¹ Euro 1 (1992): 91/441/EEC, 93/59/EEC; Euro 2 (1996): 94/12/EC, 96/69/EC; Euro 3 (2000): 98/69/EC; Euro 4 (2005): 98/69/EC (& 2002/80/EC); Euro 5 (2009): 715/2007/EC; Euro 6 (2014)

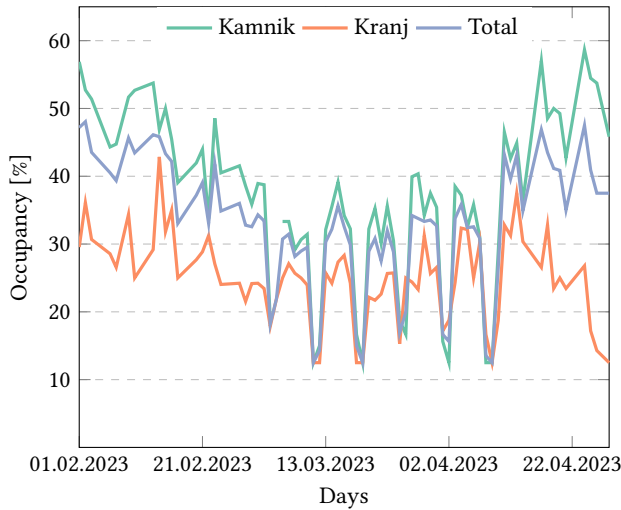


Figure 1: Average vehicle occupancy, per days and routes.

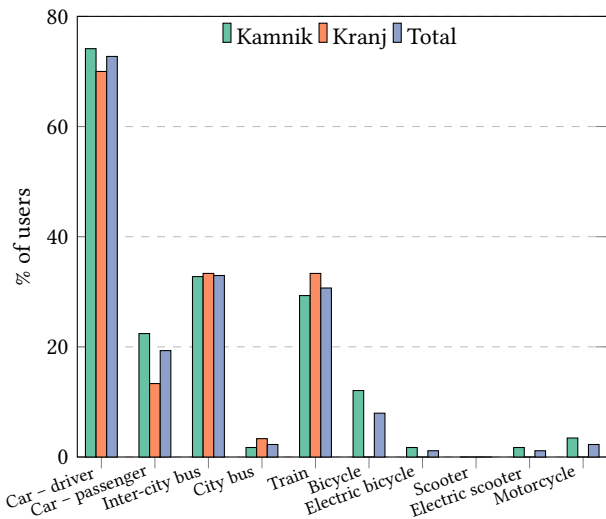


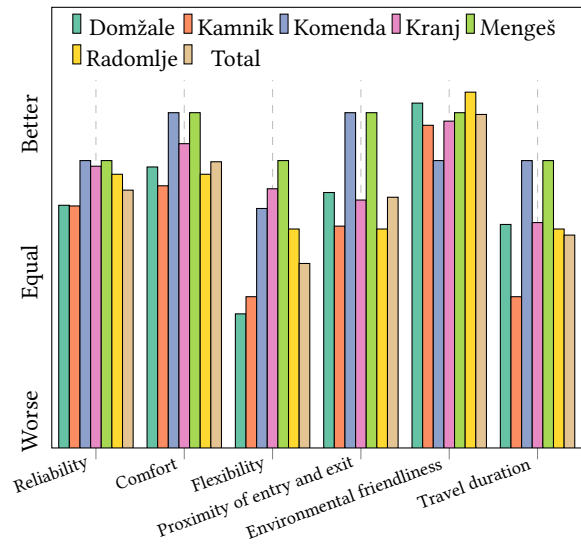
Figure 2: Current modes of transportation, normalized by the number of respondents. Since multiple answers were possible, the values shown are not true relative values and do not add up to 100%.

this mode of transport, its reliability, the proximity of stations, the possibility to effectively use the time for themselves and the fact that they did not need to search for the parking place. Complaints, on the other hand, were scarce and mainly addressed too long collection/drop-off times, variable collection times and the need to check the collection information every day. When asked to compare their usual means of transportation with the dynamic one (Figure 3), they evaluated the latter better in all points except the flexibility. It is very likely (on average 9 on the 0–10 scale) that the respondents will recommend the service to their friends and relatives.

Willingness to pay. When asked about their willingness to pay for the service, the average response was about 70€ per month (64€ for Kamnik and 86€ for Kranj).

4.3 Comparison of Traveling Times

We compared the time of traveling between the dynamic transportation and the usual modes of respondents’ transportation.



Transportation quality aspect

Figure 3: Comparison of the respondents’ usual transportation mode with the dynamic one.

Table 2: Comparison of average traveling times, in minutes.

Entry location	Usual transportation	Dynamic transportation	Difference
Domžale	30.97	26.67	-4.30
Kamnik	42.00	44.16	2.16
Komenda	59.00	36.91	-22.09
Kranj	54.72	38.97	-15.75
Mengeš	36.00	27.80	-8.20
Radomlje	40.71	40.09	-0.63
Average	41.22	35.07	-6.15

We used both data sources. It should be noted that the comparison includes only passengers who have completed the survey (otherwise we have no information about their normal time traveling time). Additionally, we had to exclude two respondents because of a mismatch with traffic data. Traffic data do not include any time of walking and waiting for a van. The survey data was collected only once for each passenger ($n = 86$), while there was substantially more traffic data (5,476 passenger rides). Therefore, there is a substantial difference in the amount and quality of the data, and a cautious interpretation is advised.

Results are shown in Table 2. Considering all entry locations, except Kamnik, the dynamic rides are faster (in average by 6.15 minutes). This comparison is not entirely fair because the dynamic transportation times do not include possible waiting and walking times, but we can safely assume that the times are at least comparable.

4.4 Analysis of Environmental Burdens

The analysis of the environmental burdens focused on CO₂ and NO_x emissions and particulate matter (PM) for human health reasons. The results in Table 3 show a significant reduction in CO₂ emissions after the introduction of the dynamic transportation service: reduction of emissions by approximately 27%, or over 50% in case of higher occupancy/exclusive use of the dynamic service. Similar reductions are also expected with NO_x (26% to 68%) and PM (27% to 79%). It is important to emphasize that

Table 3: Comparison of CO₂, NO_x and PM emissions.

	BaU	Dynamic (weighted)	Dynamic (total)
CO ₂ [t]	23.90	17.30 (-27%)	10.90 (-54%)
NO _x [kg]	23.21	17.20 (-26%)	5.30 (-69%)
PM [kg]	1.44	1.05 (-27%)	0.30 (-79%)

the absolute values of PM particle emissions are low, since the average vehicle age is about 9 years, when the minimum emission standards for solid particles have already been introduced.

Generally, the results show a significant reduction in emissions when implementing the dynamic transport compared to the BaU scenario. However, it should be noted that the data refer to the average number of users and that these may vary depending on vehicle occupancy, driving dynamics (driving speed), road conditions (e.g., duration of traffic jams), etc. These are preliminary estimates that do not yet take into account possible errors in data collection and interpretation of uncertainty. The numbers/results correspond to the implementation of the pilot project only – the potential up-scaling analysis would show the actual potential of such measures for solving the urban mobility issues.

5 CONCLUSIONS

The results of this pilot study confirm that on-demand dynamic transportation of employees to their jobs has a great potential for improving the sustainable mobility in the Ljubljana Urban Region. As part of the pilot project, which was taking place from the 1st February 2023 to 28th April 2023, 2,629 van rides were carried out with a total length of 66,199 km and a total duration of almost 32 days. The average ride was about 25 km long and lasted 44 minutes. 131 individual users were involved, who made a total of 7,182 individual commutes. Some used the service in both directions on a daily basis, others less often.

Most users of the service (about 70%) usually use a passenger car for their transport to work. The transition to dynamic transportation puts all these cars "off the road", while only moderately competes with other forms of public transportation. This also causes a significant reduction in harmful emissions of CO₂, NO_x and solid particles into the environment: from 30% to 70%, depending on the substance and transportation. Also, the dynamic transportation is comparable with other modes of transportation considering the speed and time spent. The satisfaction of users was also highly positive, as they liked that somebody else took care of their everyday commute to work, which was done without stress, accurately and reliably. They especially appreciated the "door-to-door" aspect and avoiding the search for a parking lot. However, from the data collected we cannot actually determine to which extent was the users' enthusiasm caused by the fact that the transport was free of charge during the pilot study.

Two problems were identified that might jeopardize the permanent operation of the dynamic transportation. The average occupancy of the vehicles was only around 2.73 passengers per ride (34% of 8 seats). Of the two routes, Kamnik and Kranj, the former were better occupied – by almost one passenger in average per ride. According to GoOpti [3], a higher occupancy (75%) could be achieved during the traffic peak times around the city of Domžale, which lies halfway to Kamnik.

Another problem is the price that the users are willing to pay for the service, which amounts to around 70€ per month. This is significantly less than the economic price estimated at 250€ at 50% occupancy and 160€ at 75% [3]. It seems that car users are not really aware of the actual costs and underestimate

their expenses for commuting to work. Better awareness of users would be needed on the costs and environmental impacts of different transportation modes.

In summary, dynamic transportation seems a feasible and effective alternative to using cars for commuting to work. However, in Slovenia, it requires a careful consideration, at the levels of communities, cities, regions and the whole country, of how to attract the users and support the transition to this and other sustainable transportation means. In perspective, it will also require intelligent software for supporting the service. The tasks that are particularly interesting for applying artificial intelligence methods, include the prediction of customers' requests, optimization of costs and dynamic planning of routes.

ACKNOWLEDGMENTS

This work has been funded by the SmartMOVE project, which is co-financed by Iceland, Lichtenstein and Norway funds from the EEA Financial Mechanism and corresponding Slovenian participation within the Climate Change Mitigation and Adaptation programme in the amount of 1 609 167€. The contents of this document are the sole responsibility of the authors and can in no way be taken to reflect the views of the Climate Change Mitigation and Adaptation Program Authority.

The authors also acknowledge the financial support from the Slovenian Research Agency for research core funding for the programme Knowledge Technologies (No. P2-0103).

REFERENCES

- [1] Claudia Archetti, M. Grazia Speranza, and Dennis Weyland. 2017. A simulation study of an on-demand transportation system. *International transactions in operational research*, 25, 4, 1137–1161. doi: 10.1111/itor.12476.
- [2] Carlos Lima Azevedo et al. 2016. Microsimulation of demand and supply of autonomous mobility on demand. *Transportation Research Record*, 2564, 1, 21–30. doi: 10.3141/2564-03.
- [3] Marko Bohanec, Marko Guček, Davor Kontić, Karina Sirk, Bernard Ženko, and Martin Žnidaršič. 2023. Vrednotenje učinkov dinamičnih skupinskih prevozov zaposlenih na delovna mesta. Tech. rep. Institut Jožef Stefan, Delovno poročilo DP-14427 (in Slovene), Ljubljana, Slovenija.
- [4] David Bole, Matej Gabrovec, Janez Nared, and Nika Razpotnik Viskovič. 2012. Integrated planning of public passenger transport between the city and the region: the case of Ljubljana. *Acta geographica Slovenica*, 52, 1, 141–163.
- [5] Malcolm Egan and Michal Jakob. 2016. Market mechanism design for profitable on-demand transport services. *Transportation Research Part B: Methodological*, 89, 178–195. doi: https://doi.org/10.1016/j.trb.2016.04.020.
- [6] Cveto Gregorc and David Krivec. 2012. Networking of public passenger transport modes, a step towards sustainable mobility in ljubljana urban region. *Procedia-Social and Behavioral Sciences*, 48, 3009–3017.
- [7] Simon Koblar and Luka Mladenovič. 2020. Calculating the speed of city bus trips: the case of Ljubljana, slovenia. *Urbani izziv*, 31, 1, 112–122.
- [8] Wengen Li, Jiannong Cao, Jihong Guan, Shuigeng Zhou, Guangqing Liang, Winnie K. Y. So, and Michal Szczecinski. 2019. A general framework for unmet demand prediction in on-demand transport services. *IEEE Transactions on Intelligent Transportation Systems*, 20, 8, 2820–2830.
- [9] Statistical Office RS. 2023. Statistical yearbook of RS. accessed on 23.8.2023. (2023). <https://www.stat.si/Pages/en/goals/goal-12.-ensure-sustainable-consumption-and-production-patterns/12.5-average-co2-emissions-per-km-from-new-passenger-cars>.
- [10] SmartMOVE. 2022–24. Smart mobility measures for sustainable mobility in slovenia. project financed by iceland, liechtenstein and norway funds from the eea financial mechanism and the norwegian financial mechanism. (2022–24). <https://kt.ijs.si/project/smartmove/>.
- [11] Amirmahdi Tafreshian, Neda Masoud, and Yafeng Yin. 2020. Frontiers in service science: ride matching for peer-to-peer ride sharing: a review and future directions. *Service Science*, 12, 2-3, 44–60. doi: 10.1287/serv.2020.0258.
- [12] Jernej Tiran, Luka Mladenovič, and Simon Koblar. 2015. Accessibility to public transport using the ptal method: the case of ljubljana. *Geodetski Vestnik*, 59, 4, 723–735.
- [13] G. Titos, H. Lyamani, L. Drinovec, F.J. Olmo, G. Močnik, and L. Alados-Arboledas. 2015. Evaluation of the impact of transportation changes on air quality. *Atmospheric Environment*, 114, 19–31.
- [14] Steve Wright, John D. Nelson, and Caitlin D Cottrill. 2020. Maas for the suburban market: incorporating carpooling in the mix. *Transportation Research Part A: Policy and Practice*, 131, 206–218.

Test uporabnosti prilagojenega WHCA* algoritma za iskanje poti za več agentov v strateški igri v realnem času

Usability Test of a Modified WHCA* Algorithm for Multi-Agent Pathfinding in a Real-time Strategy Game

Ivan Antešić

ivan.v.antesic@gmail.com

Laboratorij za umetno inteligenco
Fakulteta za računalništvo in informatiko
Univerza v Ljubljani, Slovenia

Aleksander Sadikov

aleksander.sadikov@fri-uni-lj.si

Laboratorij za umetno inteligenco
Fakulteta za računalništvo in informatiko
Univerza v Ljubljani, Slovenia

POVZETEK

V zadnjih letih je bilo predlaganih več algoritmov za iskanje poti za več agentov (*angl. kratica MAPF*), ki naj bi bili primerni za vodenje enot v strateških igrah v realnem času. Toda algoritmi so predstavljeni in testirani brez upoštevanja ključnih lastnosti kompleksnega okolja iger, kot so dinamični zemljevidi, različne lastnosti in hitrosti agentov, prisotnost sovražnikov. Da bi ugotovili, ali je MAPF pristop res primeren za uporabo v igrah, smo v obstoječem igralnem pogonu za strateške igre v realnem času implementirali in prilagodili WHCA* algoritem ter ga primerjali s standardnim LRA* pristopom iskanja poti za posameznega agenta. Eksperimentalni rezultati kažejo, da naša implementacija WHCA* algoritma znatno izboljša kakovost poti ter lahko reši težke primere, ki jih LRA* ne more. Čeprav je čas iskanja poti z WHCA* veliko daljši, menimo, da MAPF ima potencial v razvoju iger.

ABSTRACT

Over the years, several multi-agent pathfinding (MAPF) algorithms have been proposed as suitable solutions for guiding units in real-time strategy (RTS) games. However, algorithms are tested without considering the crucial properties of a complex game environment, such as dynamic maps, different unit properties and agent speeds, the presence of enemies. To determine whether MAPF approach really is suitable for use in games, we implemented and modified the seminal WHCA* algorithm in an existing RTS game engine and compared it to the common LRA* single-agent pathfinding approach. Our experimental results show that our WHCA* implementation greatly improves the path quality and can solve difficult scenarios that the single-agent approach cannot. WHCA*'s search times are much longer, but we still think MAPF has potential in game development.

KLJUČNE BESEDE

iskanje poti za več agentov, strateške igre v realnem času, heuristično preiskovanje, WHCA*

KEYWORDS

multi-agent pathfinding, real-time strategy games, heuristic search, WHCA*

1 UVOD IN MOTIVACIJA

NP-težek problem iskanja poti za več agentov (*angl. multi-agent pathfinding, MAPF*) je definiran z grafom $G = (V, E)$ ter množico n sodelovalnih agentov a_1, \dots, a_n . Veljavna rešitev problema je množica n poti, ki pripelje vsakega agenta a_i od njegovega začetnega vozlišča $s_i \in V$ do ciljnega vozlišča $g_i \in V$ po povezavah $e_i \in E$, ne da bi dve poti prišli v konflikt. Konflikt oz. trk nastane, ko se agenta a_i in a_j oba nahajata v vozlišču $v \in V$ v istem časovnem koraku t_i . Optimalna rešitev je po navadi tista, ki minimizira skupno vsoto cen vseh poti [2].

Več znanstvenih del predstavlja MAPF algoritme, za katere trdijo, da so primerni za uporabo v igrah, še posebno za strateške igre v realnem času (*angl. real-time strategy, RTS*). Vendar predstavljeni algoritmi pri razvoju in testiranju ne upoštevajo ključne lastnosti kompleksnega okolja RTS iger. RTS igre vsebujejo agente različnih hitrosti in značilnosti. Med igro se obstoječim agentom kadarkoli lahko dodeli nov ukaz, poleg tega pa se neprenehoma ustvarjajo novi agenti. Značilni so tudi dinamični zemljevidi, ki jih igralci lahko tekom igre spreminjajo z gradnjo objektov, ter sovražni agenti, s katerimi ne sodelujemo, ampak jih moramo še vedno upoštevati pri iskanju poti, kot neprehodne, premikajoče ovire.

Izkaže se, da v praksi večina iger išče pot za vsakega agenta posebej (*angl. single-agent pathfinding, SAPF*), po navadi z A* algoritmom [3], brez upoštevanja načrtov ostalih agentov. Konflikte, ki nastanejo kasneje med premikanjem, poskusijo igre lokalno razrešiti z ad-hoc preverbami in pravili. Ta pristop je znan kot Lokalno popravljanje A* (*angl. Local-Repair A*, LRA**) [10]. Ko zahtevnost iger narašča lahko preprosti LRA* algoritem postane nezadosten. Zmogljivost je mogoče izboljšati na več načinov, na primer s hierarhično abstrakcijo [9] ali z upoštevanjem simetrije preiskovalnega prostora [5]. Marsikateri SAPF algoritmi se lahko danes uporabijo za vodenje velikega števila agentov v RTS igri. Ker pa pri iskanju ne upoštevajo načrtov drugih agentov, so opazni pogosti zastoji ter nasploh okorno in nenaravno premikanje, še posebej v ozkih hodnikih.

V želji, da bi premostili razkorak med znanstvenim raziskovanjem in praktičnim razvojem iger, smo implementirali enega ključnih MAPF algoritmov, Okvirjeni hierarhični kooperativni A* (*ang. Windowed Hierarchical Cooperative A*, WHCA**) v OpenRA igralnem pogonu za RTS igre in ocenili, ali je MAPF pristop res uporaben za RTS igre. WHCA* namreč velja za enega temeljnih sub-optimalnih MAPF algoritmov, ki služi kot izhodišče mnogim novejšim algoritmom in nadgradnjam [8, 1, 6].

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Information Society 2023, 9–13 October 2023, Ljubljana, Slovenia

© 2023 Copyright held by the owner/author(s).

2 METODE

2.1 OpenRA

OpenRA [4] je odprtokodni pogon za strateške igre v realnem času, ki med drugimi poganja na novo ustvarjeno igro Dune 2000. Le ta predstavlja dokaj tipičen primer RTS igre. Igralci iz različnih perspektiv opazujejo mrežni, dvodimenzionalni zemljevid in nadzirajo svoje enote. Cilj igre je nabirati surovine, razširiti bazni tabor z novimi zgradbami in proizvesti enote, s katerimi poskušamo napasti in uničiti nasprotnikov tabor. Enote imajo različne hitrosti, so zmožne več akcij in se gladko premikajo iz ene celice v drugo. To predstavlja kompleksno okolje za navigacijo enot, še posebej, ker se vse odvija v realnem času.

2.2 LRA*

OpenRA za iskanje poti uporablja različico standardnega LRA* algoritma. Ker se agenti lahko premikajo med celicami diagonalno, LRA* za heuristično funkcijo uporablja oktilno razdaljo, definirano kot $D_{oct} = (|x_2 - x_1| + |y_2 - y_1|) + (\sqrt{2} - 2) \cdot \min(|x_2 - x_1|, |y_2 - y_1|)$, med dvema celicama (x_1, y_1) in (x_2, y_2) . Oktilna razdalja je monotona in dopustna heuristika.

Pred premikom agenta v novo celico algoritem požene zaporedje preverb, s katerimi poskuša ujeti in preprečiti vse možne trke s čakanjem ali iskanjem nove poti. Ta sistem je zapleten, ker med igranjem lahko pride do množice različnih robnih pogojev. To je ena glavnih slabosti LRA* pristopov.

2.3 Osnovni WHCA*

Osnovni WHCA* algoritem je bil v izvirnem članku [7] označen za nadvse primerne za RTS igre. Z A* preiskuje 3D prostor (2D mrežni zemljevid in diskretna časovna dimenzija) za vsakega agenta posebej ter rezervira najdene poti v skupno časovno-prostorsko rezervacijsko tabelo. Vnos vozlišča (x, y, t_i) v tabeli oznanja, da je celica (x, y) že zasedena v časovnem koraku t_i in zato ni na voljo ostalim agentom. Med razvijanjem vozlišča se preveri, katere sosednje celice so proste ovir in rezervacij, kar implicitno omogoči izogibanje trkom in kooperativno premikanje. WHCA* za heuristično oceno razdalj uporablja Obratno nadaljevalno A* (*angl. Reverse Resumable A*, RRA**). Pred začetkom iskanja kooperativnih poti se najprej za vsakega agenta požene RRA* iskanje, ki z A* poišče najkrajšo pot od agentovega cilja do začetka. Zaradi monotonosti oktilne razdalje imamo sedaj na voljo točno ceno do agentovega cilja za vsako razvito vozlišče. Če med kooperativnim iskanjem agent zaide z najkrajše poti in potrebuje heuristično oceno za novo vozlišče, nadaljujemo RRA* iskanje, dokler le-ta ne razvije potrebnega vozlišča in pridobi točno razdaljo. WHCA* išče poti le za naslednjih w korakov, kar pripomore k zmogljivosti. Ko agent izvede $w/2$ premikov, še enkrat poženemo iskanje za naslednjih w korakov. To imenujemo iskanje z okvirjem.

2.4 Prilagojeni WHCA*

Zaradi kompleksnosti in narave OpenRA okolja je bilo potrebno prilagoditi osnovni WHCA* algoritem. Igralec lahko kadarkoli ustvari nove agente in ukazuje nove premike obstoječim agentom. Zaradi tega se kooperativno iskanje poti ne sme nikoli končati. Posledično mora biti rezervacijska tabela krožna. Agenti morajo po premiku v naslednjo celico izbrisati rezervacijo prejšnje celice, da ne bi vplivala na iskanje poti v prihodnosti.

Vsi agenti ne dobijo ukaza v istem trenutku zato je potrebno sinhronizirati individualna iskanja in za vse agente izračunati

novo delno pot vsakih $R = w/2$ taktov (*angl. ticks*, osnovna časovna enota pogona in časovni korak našega algoritma). Ena sekunda ponavadi vsebuje okoli 25 taktov. Zaradi večje zveznosti premikov en časovni korak ne ustreza več enemu premiku celice – agenti, glede na njihovo hitrost, zasedejo celico med premikom za različno število taktov. To lahko povzroči nemirno obnašanje agentov, ki želijo zapolniti iskalno okno z nepotrebnimi premiki, kar odpravimo z znižanjem cene čakanja na mestu.

Zaradi kompleksnega premikanja agentov in arhitekture pogona je preverjanje rezervacij oteženo. Potrebno je izračunati čas prihoda in odhoda agenta za sosednjo celico. Čas odhoda je težko predvideti, saj ne vemo z gotovostjo v katero od naslednjih sosednjih celic se bo napotil, od tega pa je odvisno koliko časa bo preživel v celici. Netočni čas prihoda in odhoda lahko povzroči sprejem konfliktnih poti ali zavračanje veljavnih rešitev.

RRA* heuristika je bila prvotno namenjena za uporabo na mreži, kjer se agenti premikajo v štirih glavnih smereh. V našem primeru se agenti lahko premikajo še diagonalno, kar povzroči potrebo po dodatnem nadaljevanju iskanja RRA* z oktilno razdaljo, tudi ko agent sledi prvotni poti. Da bi to optimizirali, v glavnem kooperativnem iskanju filtriramo sosedne in obdržimo le tiste, ki so že bili razviti z RRA*. To vpelje dodatne zaplete in povzroči, da agent včasih ne uspe najti poti okoli dinamičnih ovir.

WHCA* ima problem z vrstnim redom obdelave agentov, na kar sta opozorila tudi Sturtevant [8] in Bnaya [1]. Lahko se zgodi, da agent z višjo prioriteto prvi rezervira pot in zasede vse celice okoli naslednjega agenta z nižjo prioriteto, ki nato ne more najti poti. V takem primeru je konflikt neizogiben. Problem postane še bolj pogost zaradi heterogenih agentov v OpenRA, kjer lahko hitrejši agent rezervira več celic okoli počasnejšega agenta. Težavo smo poskusili rešiti s prepisovanjem rezervacij, kar pa lahko privede do kaskadne izgube sodelovanja med agenti.

Obdržali smo nekatere principe lokalnih preverb za izogibanje trkom, ker so prisotne tudi enote s katerimi ne sodelujemo, na primer zgradbe in sovražniki. Dodali smo tudi heuristično posodabljanje cilja. Agent zamenja za cilj prvo prosto celico v radiju 8 celic okoli prvotnega cilja, če je le ta med iskanjem poti že zaseden ali pa za cilj vzame svojo trenutni položaj, če je že v bližini cilja (8 celic) in med premikanjem naleti na oviro.

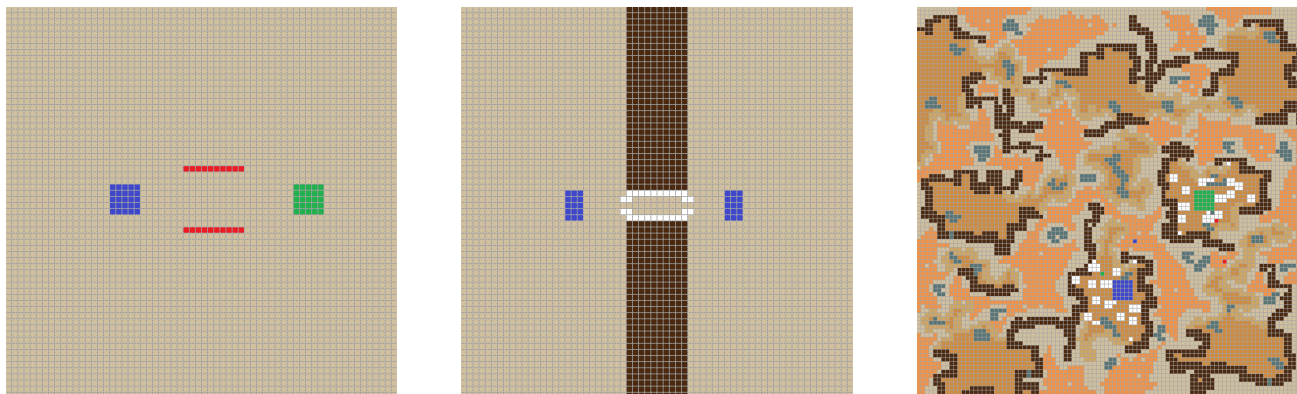
3 EKSPERIMENTI

Časovno zmogljivost in kakovost poti prilagojenega WHCA*(R) algoritma smo testirali v igri OpenRA pogona za različne dolžine iskalnega okvirja $R = w/2$ in primerjali rezultate s prvotnim algoritmom pogona LRA*.

Posamezni eksperiment sestoji iz 50 iteracij. Ena iteracija je premik skupine 25 agentov iz začetnega območja na levi strani zemljevida do ciljnega območja na desni. Agentom se pripadajoča začetna in ciljna celica znotraj območij določi naključno. Iteracija se konča, ko vsi agenti prispejo na cilj ali pa ko se izteče vnaprej določena časovna limita testa. Agenti so naključno izbrani iz nabora treh enot: hitro vozilo s kratkim časom obračanja, srednje hitrim tankom z dolgim časom obračanja in počasni pešak s takojšnjim obratom.

Iskanje poti smo testirali na treh različnih zemljevidih vidnih na Sliki 1. Vsak od njih predstavlja določen scenarij, ki lahko nastane tekom igre:

- SOVRAŽNIKI: prečkanje skupine premikajočih sovražnih agentov, časovna limita 1,150 taktov, velikost 64x64 celic



Slika 1: Sheme zemljevidov eksperimentalnih scenarijev: levo SOVRAŽNIKI, v sredini GRLO, desno IGRA. Modra barva označuje začetne celice, zelena pa ciljne. Skupini agentov v scenariju GRLO, si izmenjata začetni poziciji. Bele, temno sive in temno rjave celice predstavljajo neprehodne zgradbe ter naravne ovire. Ostale barve onačujejo prosto prehodna območja. Scenarij SOVRAŽNIKI vsebuje sovražnike, ki patrolirajo med zgornjimi in spodnjimi rdečimi celicami. IGRA vsebuje še v igrah pogosto, nabiralniško enoto – prijateljska patroljira ob vhodu spodnje baze, sovražna pa ob vhodu zgornje.

- GRLO: premik dveh manjših skupin iz nasprotnih smeri skozi ozko grlo, časovna limita 1,150 taktov, velikost 64x64 celic
- IGRA: simulacija povprečnega premika v nasprotnikovo bazo med igranjem igre, časovna limita 2,350 taktov, velikost 69x69 celic

Število agentov in dimenzije zemljevida smo izbrali skladno s povprečnim primerom igre.

Rezultati eksperimentov so vidni na Slikah 2 in 3. Trajanje iskanja poti smo merili s številom razvitih vozlišč (zgornja vrstica Slike 3) in z milisekundami (spodnja vrstica Slike 3). Da lažje analiziramo vpliv RRA* heuristike, smo, kot v ostalih člankih, ločili časovne meritve na prvi inicializacijski takt in na vse naslednje takte. Mera prvega takta je povprečje vseh prvih taktov iteracije. Za mero naslednjih taktov smo vzeli povprečje maksimalnih vrednosti iteracije po prvem taktu. Mera nam tako predstavlja kako dolgo, v povprečju, algoritem išče pot v najtežjih situacijah.

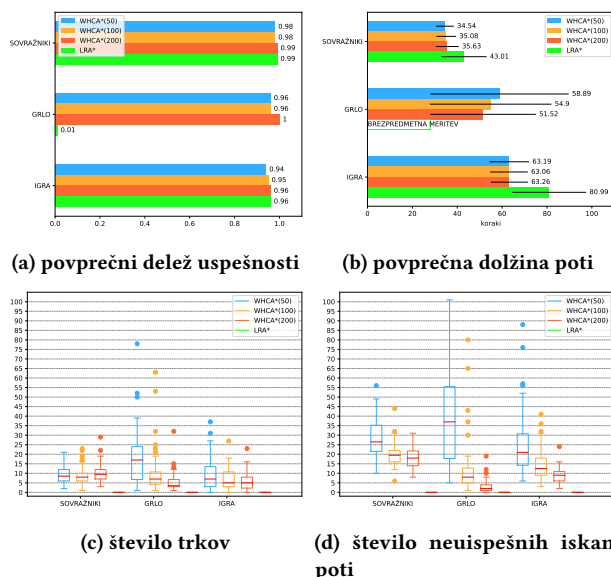
Dolžino poti smo merili s povprečnim številom korakov, ki jih potrebuje agent, da prvič prispe do cilja (Slika 2b).

Izmerili smo tudi delež agentov, ki uspešno prispejo do cilja (Slika 2a), kolikokrat algoritem ne uspe najti veljavne kooperativne poti (Slika 2d) ter število trkov (Slika 2c). Agentov po trku ne odstranimo iz igre.

4 DISKUSIJA

Na Sliki 2b lahko hitro odčitamo, da ima WHCA* v primerjavi z LRA* manjšo povprečno dolžino poti. Boljša kakovost rešitve je očitna tudi, ko s prostim očesom opazujemo premikanje agentov, tudi za scenarij IGRA, ki predstavlja dober vpogled v delovanje algoritma v OpenRA okolju. WHCA* gladko vodi agente skozi kompleksno okolje v bolj strnjjenih skupinah. Agenti so zmožni sodelovanja, se umikajo, da razrešijo konflikte in lahko obvozijo počasnejše agente. Agenti v LRA* eksperimentih se premikajo okorno, razpršeno in imajo navado slediti eden drugemu v dolgih, ozkih kolonah do cilja.

Zaradi sodelovalnega iskanja poti je WHCA* zmožen rešiti tudi težek scenarij z ozkim grlom. Večji kot je iskalni okvir, bolje bo WHCA* vodil agente skozi majhno odprtino. Pri LRA* agenti vedno obstanejo v gneči sredi prehoda in ne dosežejo svojih ciljev

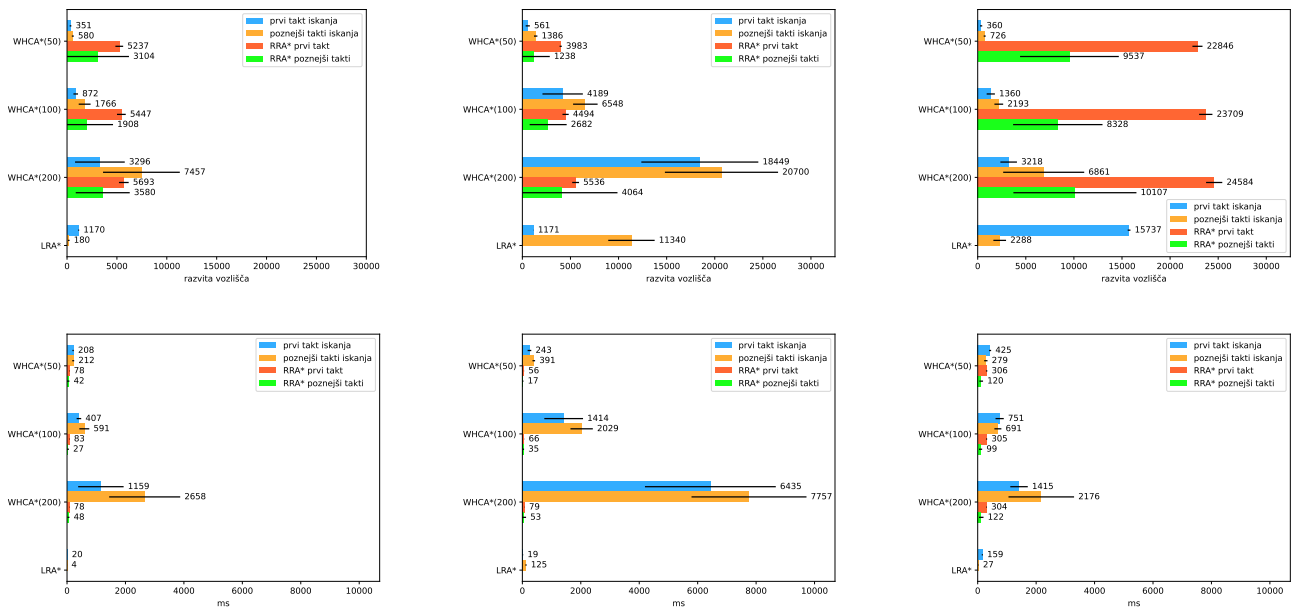


Slika 2: Rezultati eksperimentov za različne scenarije.

(zaradi izjemno nizke uspešnosti LRA* za scenarij GRLO, dolžine poti ni smotno primerjati z meritvami za WHCA*). V preostalih primerih oba algoritma do cilja privedeta skoraj vse agente.

Na žalost WHCA* potrebuje veliko več časa za pridobitev kooperativne poti za vse agente. Čas je odvisen od širine iskalnega okvirja, vendar je tudi najmanjši okvir počasnejši kot LRA* v vseh primerih. V povprečni OpenRA igri WHCA* s kompromisnim okvirjem $R = 100$ lahko porabi 700ms za iskanje poti, kar je v RTS igri moteče in opazno igralcem, tudi če ne prekine igro za toliko časa ob čisto vsakem ukazu.

Če na Sliki 3 primerjamo rezultate razvitih vozlišč s časom v milisekundah opazimo, da RRA* predstavlja le majhen delež časa iskanja poti z WHCA*, tudi v primerih, ko je število razvitih vozlišč z RRA* bistveno večje kot število vozlišč razvitih z WHCA*. Računanje dodatnih korakov potrebnih za iskanje poti v kompleksnem OpenRA okolju, povzroči, da razvijanje vozlišč traja dlje



Slika 3: Rezultati dolžine iskanja poti v razvitih vozliščih (zgornje slike) in v milisekundah (spodnje slike) za različne scenarije: levi sliki SOVRAŽNIKI, sredinski GRLO, desni IGRA.

za glavno WHCA* iskanje kot pa za bolj preprosto RRA* hevrstiko. Izkazuje se, da WHCA* razvije največ vozlišč, ko računa drugi okvir poti, saj se takrat kompleksnost primera poveča, medtem ko RRA* razvije večino vozlišč že v prvem taktu. Posledično naša implementacija potrebuje manj časa v prvem inicializacijskem taktu kot pa v naslednjih taktih.

Dinamični scenarij SOVRAŽNIKI povzroči težave WHCA* algoritmu, saj pokvari točnost RRA* hevrstike ter oteži kooperacijo s prisotnostjo nepričakovanih sovražnih agentov, s katerimi ne sodelujemo. Rezultati na sliki 2c in 2d prikazujejo številne trke in neuspela iskanja poti WHCA* algoritma, ki jih v nasprotju z drugima scenarijema, znatno ne zniža niti največja širina iskalnega okvirja. V ostalih primerih večji iskalni okvirji preprečijo dovolj neuspešnih iskanj poti in trkov, da le-ti niso opazni med igro. Zanimivo je, da se med opazovanjem preprostejši LRA* v primeru sovražnih agentov izkaže za bolj uspešnega, saj le počaka nekaj trenutkov, da se agent umakne, preden nadaljuje prvotno pot. Nasprotno, WHCA* nemudoma poišče novo pot okoli ovire. Kljub temu, dinamične ovire ne vplivajo občutno na časovno zmogljivost WHCA* algoritma.

5 POVZETEK

Implementirali smo prilagojeni WHCA* algoritem v obstoječem RTS igralnem pogonu OpenRA, da bi dognali, ali je MAPF pristop res primeren za uporabo v RTS igrah.

Naš prilagojeni WHCA* smo testirali in primerjali s standardnim SAPF algoritmom LRA*, ki je že bil implementiran v pogonu. Izkazuje se, da WHCA* najde veliko bolj kakovostne poti kot LRA* in je zmožen gladkega, kooperativnega premikanja agentov v zahtevnih scenarijih, kjer LRA* agenti običijo. Toda WHCA* je dosti počasnejši od preprostega LRA* in v realnih primerih potrebuje nekaj sto milisekund za pridobitev poti, tako v prvem kot tudi v naslednjih okvirjenih iskanjih. Zato menimo, da trenutna WHCA* implementacija ni primerna za premikanje agentov v

večini RTS iger. Lahko bi jo uporabili za igre v realnem času, ki vsebujejo manj agentov in preprosto okolje, ali za potezne igre, kjer je na voljo več časa za pridobitev kakovostne rešitve.

Kljub temu mislimo, da to ni dokončna zavrnitev MAPF pristopa za RTS igre. Z nadaljnjim praktičnim delom bi lahko postal bolj primeren. WHCA* bi lahko nadgradili s hierarhično abstrakcijo [8] ali pa z dinamično postavitvijo okvirja okoli možnih konfliktov [1]. Pomagalo bi tudi, če bi RTS igro razvili od začetka z WHCA* v mislih, namesto da prilagajamo algoritem obstoječi SAPF arhitekturi.

Kolikor nam je znano, je to edina delujoča implementacija MAPF algoritma v okolju realnočasovne igre. Izvorna koda je prosto dostopna na <https://github.com/ia6382/OpenRA>.

LITERATURA

- [1] Zagy Bnaya in Ariel Felner. 2014. Conflict-oriented windowed hierarchical cooperative A*. V *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 3743–3748.
- [2] Ariel Felner, Roni Stern, Solomon Eyal Shimony, Eli Boyarski, Meir Goldenberg, Guni Sharon, Nathan Sturtevant, Glenn Wagner in Pavel Surynek. 2017. Search-based optimal solvers for the multi-agent pathfinding problem: summary and challenges. V *Tenth Annual Symposium on Combinatorial Search*.
- [3] Peter E Hart, Nils J Nilsson in Bertram Raphael. 1968. A formal basis for the heuristic determination of minimum cost paths. *IEEE transactions on Systems Science and Cybernetics*, 4, 2, 100–107.
- [4] [n. d.] OpenRA. V <https://www.openra.net/>.
- [5] Steve Rabin in Nathan Sturtevant. 2016. Combining bounding boxes and jps to prune grid pathfinding. V *Proceedings of the AAAI Conference on Artificial Intelligence* številka 1. Zv. 30.
- [6] Devon Sigurdson, Vadim Bulitko, Sven Koenig, Carlos Hernandez in William Yeoh. 2019. Automatic algorithm selection in multi-agent pathfinding. *arXiv preprint arXiv:1906.03992*.
- [7] David Silver. 2005. Cooperative pathfinding. *AIIDE*, 1, 117–122.
- [8] Nathan R Sturtevant in Michael Buro. 2006. Improving collaborative pathfinding using map abstraction. V *AIIDE*. Marina del Rey, 80–85.
- [9] Nathan R Sturtevant, Devon Sigurdson, Bjorn Taylor in Tim Gibson. 2019. Pathfinding and abstraction with dynamic terrain costs. V *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment* številka 1. Zv. 15, 80–86.
- [10] A. Zelinsky. 1992. A mobile robot exploration algorithm. *IEEE Transactions on Robotics and Automation*, 8, 6, 707–717. DOI: 10.1109/70.182671.

An Attempt at Predicting Algorithm Performance on Constrained Multiobjective Optimization Problems

Andrejaana Andova

Aljoša Vodopija

Jordan Cork

Tea Tušar

Bogdan Filipič

andrejaana.andova@ijs.si

Jožef Stefan Institute and Jožef Stefan International Postgraduate School
Ljubljana, Slovenia

ABSTRACT

When solving new optimization problems, it is crucial that algorithms are selected capable of both finding the best solutions and computing them in reasonable amounts of time. However, testing multiple algorithms is time-consuming and impractical. A solution to this would be to build a model that automatically selects the algorithm that performs best on a new problem. In this work, we build machine learning models to automatically predict algorithm performance on constrained multiobjective optimization problems (CMOPs) using exploratory landscape analysis (ELA) features. The results showed a high mean absolute error, which indicates that, with the currently available benchmarks and ELA features, automatically predicting algorithm performance on CMOPs is a very hard task.

KEYWORDS

constrained multiobjective optimization, evolutionary algorithms, exploratory landscape analysis, machine learning, algorithm performance prediction

1 INTRODUCTION

The common way of solving black-box constrained multiobjective optimization problems (CMOPs) is to use multiobjective optimization algorithms with constraint-handling techniques (CHTs). However, deciding which specific algorithm to use, which CHT to include, and which setting of the algorithm parameters to apply is not trivial.

In the last few years, several authors have tried to find ways of automatically selecting evolutionary algorithms for solving single-objective optimization problems [10, 13, 7]. The core concept behind their work is to extract features of benchmark single-objective optimization problems and construct a model for predicting which algorithm performs best for each individual problem. When dealing with a new problem then, the model is able to automatically decide which algorithm to use for solving the problem.

Extracting optimization problem features can be done using exploratory landscape analysis (ELA). This is a technique that takes a sample of solutions and their fitness values as input and, based on this, extracts statistical features about the problem.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Information Society 2023, 9–13 October 2023, Ljubljana, Slovenia

© 2023 Copyright held by the owner/author(s).

These features ideally characterize the problems so that similar problems have similar feature values.

In this work, we propose a first step towards automatic algorithm selection for CMOPs. This task is much harder for constrained multiobjective optimization, because, in this area, there are fewer benchmark problems available, and the ELA methods are not as well developed as in single-objective optimization.

Although the ultimate goal of our work is automatic algorithm selection, we here focus on predicting the algorithm performance of three widely used algorithms. By proposing a method for predicting algorithm performance for a few well-known algorithms, researchers can easily extend the set of algorithms, in the future.

The paper is further organized as follows. In Section 2, we introduce the theoretical background of constrained multi-objective optimization. In Section 3, we briefly describe the ELA features used in this study. In Section 4, we describe the algorithm performance measure used as the prediction target value. In Section 5, we present the experimental setup and, in Section 6, the obtained results. Finally, in Section 7, we summarize the findings and outline ideas for future work.

2 THEORETICAL BACKGROUND

A CMOP can be formulated as:

$$\begin{aligned} & \text{minimize} && f_m(\mathbf{x}), \quad m = 1, \dots, M \\ & \text{subject to} && g_k(\mathbf{x}) \leq 0, \quad k = 1, \dots, K, \end{aligned} \quad (1)$$

where $\mathbf{x} = (x_1, \dots, x_D)$ is a *solution vector* of dimension D , $f_m(\mathbf{x})$ are *objective functions*, $g_k(\mathbf{x})$ are *constraint functions*, and M and K are the numbers of objectives and constraints, respectively.

In multiobjective optimization, we use the term search space S , representing a D dimensional space where all possible solution vectors \mathbf{x} are located. Additionally, we can define the M dimensional objective space $P = \{f(\mathbf{x}) \mid \mathbf{x} \in S\}$ which represents the space consisting of objective values for solutions.

A solution \mathbf{x} is *feasible*, if it satisfies all constraints, $g_k(\mathbf{x}) \leq 0$, for $k = 1, \dots, K$. A feasible solution \mathbf{x} is said to *dominate* another feasible solution \mathbf{y} if $f_m(\mathbf{x}) \leq f_m(\mathbf{y})$ for all $1 \leq m \leq M$, and $f_m(\mathbf{x}) < f_m(\mathbf{y})$ for at least one $1 \leq m \leq M$. A feasible solution \mathbf{x}^* is a *Pareto-optimal solution* if there exists no feasible solution $\mathbf{x} \in S$ that dominates \mathbf{x}^* . All feasible solutions constitute the *feasible region* F . All nondominated feasible solutions form the *Pareto set* S_0 , and the image of the Pareto set in the objective space is the *Pareto front*, $P_0 = \{f(\mathbf{x}) \mid \mathbf{x} \in S_0\}$.

Nondomination ranking is a concept in multiobjective optimization that helps sort the solutions in a population into fronts, based on their dominance. Thus, all nondominated solutions get a nondomination rank of 1, solutions that are dominated only by

the nondominated solutions get a nondomination rank of 2, and so on.

The point in the objective space with the best objective values is the *ideal point* $z_I = (\min_{\mathbf{x} \in S_0} f_1(\mathbf{x}), \dots, \min_{\mathbf{x} \in S_0} f_M(\mathbf{x}))$.

The *nadir point* represents the point in the objective space with the worst fitness values across all solutions in the Pareto front $z_N = (\max_{\mathbf{x} \in S_0} f_1(\mathbf{x}), \dots, \max_{\mathbf{x} \in S_0} f_M(\mathbf{x}))$.

The most widely used quality indicator in multiobjective optimization is the *hypervolume indicator* [17]. It maps the set of solutions found by an algorithm to a measure of the region dominated by that set and bounded by a given reference point.

3 ELA FEATURES FOR CONSTRAINED MULTI-OBJECTIVE OPTIMIZATION PROBLEMS

ELA is a methodology that extracts the features of an optimization problem from a sample of its solutions. These features are usually statistical relations between the solutions and are designed by experts. Many ELA feature sets were designed for single-objective optimization problems. However, only a few feature sets exist for CMOPs.

State-of-the-art features for CMOPs were collected by Alsouly et al. [1], who adopted all of the fast-computing features for CMOPs from the related work, and also proposed some additional features. The set of all features can be divided into three groups that describe: the multiobjective landscape, the violation landscape, and the combination of the two landscapes – the multiobjective violation landscape.

All three groups of features consist of global and random walk features. The global features were calculated on a sample of size $1000 \cdot D$. The random walk features are computed during a random walk, where statistics are derived from neighboring solutions that form a sequence within the random walk. The random walk neighborhood is of size $N = 2 \cdot D + 1$, the length of the random walk is equal to $(D/N) \cdot 10^3$, and the step size is 2% of the range of the search space.

In the *multiobjective landscape group*, the features are designed to describe the objectives and the relations between them. Thus, the global features in this group include the proportion of unconstrained Pareto optimal solutions, the hypervolume of the unconstrained Pareto front, the correlation between the objective values, statistics on the unconstrained ranks, etc. The random walk features in this group include statistics on the distance between random walk neighbors in the objective space.

In the *violation landscape group*, the features are designed to describe the constraints of the problem. Thus, the global features in this group include statistics of the constraint violations, while the random walk features include statistics of the constraint violations between random walk neighbors.

In the *multiobjective violation landscape group*, the features are designed to describe the relations between the objectives and the constraints. Thus, the global features in this group include the proportion of feasible solutions, the proportion of Pareto optimal solutions, the hypervolume, statistics on the correlations between objectives and constraints, statistics on the distance between solutions in the Pareto front, etc. The random walk features in this group include statistics on the dominance relations between random walk neighbors.

Another state-of-the-art feature set for CMOPs is the one proposed by Vodopija et al. [14]. This feature set includes important information about CMOPs, including their multimodality and

other landscape characteristics. However, to calculate these features one needs a larger sample size (a sample size of 250,000), which makes these features computationally very demanding.

In our study, we used both the features by Alsouly et al. and Vodopija et al.

4 EMPIRICAL CUMULATIVE DISTRIBUTION FUNCTIONS

One drawback of using hypervolume as the quality indicator in constrained multiobjective optimization is that it does not take into consideration infeasible solutions. For this reason, Vodopija et al. [15] proposed a new quality indicator designed specifically for constrained multiobjective optimization that generalizes the hypervolume-based quality indicator I_{HV+} from [5] as follows:

- (1) When there are no feasible solutions in the set, the quality indicator takes on the value of the smallest constraint violation of all solutions in the set plus a threshold τ^* .
- (2) When the set contains at least one feasible solution, the quality indicator equals the value of I_{HV+} bounded above by the threshold τ^* , i.e., it equals $\min\{I_{HV+}, \tau^*\}$.

The threshold value τ^* ensures that any infeasible solution will be deemed worse than any feasible one.

To measure algorithm performance during the algorithm run, we keep track of how many function evaluations, called *run-times*, are needed to reach a particular quality indicator value, called *target*. We do so for a number of targets and visualize these runtimes using the Empirical Cumulative Distribution Function (ECDF) [5]. The ECDF measures the proportion of achieved targets at a given runtime by the given algorithm. Whenever an algorithm achieves a target, the value of the measure rises. Thus, the maximum value that can be achieved by an algorithm is equal to 1, meaning that the algorithm achieved all the targets.

In our work, we want to express algorithm performance in a single value which will serve as the target of our machine learning (ML) problem. However, the ECDF is given for any number of function evaluations (up to a maximum value). To end up with a single value, we use the area under the curve (AUC) of the ECDF, in short AUC-ECDF. This way, the ML method needs to predict a single target variable, which also includes information about the convergence of the algorithm over time. To normalize the AUC-ECDF value, we divide it by the maximum number of function evaluations.

5 EXPERIMENTAL EVALUATION

We focus on constrained bi-objective optimization problems with 2D, 3D, and 5D search spaces and, thus, use three widely used benchmarks for constrained multiobjective optimization – MW [9], CF [16], and C-DTLZ [6]. Because some benchmark problems are only defined for more than 3D or more than three objectives, the total number of problems per dimension differs. Specifically, for 2D, we have 8 out of 14 MW problems, 0 out of 10 CF problems, and 5 out of 6 C-DTLZ problems. For 3D, we have 14 out of 14 MW problems, 5 out of 10 CF problems, and 6 out of 6 C-DTLZ problems. For 5D, we have 14 out of 14 MW problems, 7 out of 10 CF problems, and 6 out of 6 C-DTLZ problems.

The focus of this work is on predicting the algorithm performance of three multiobjective optimization algorithms – NSGA-III [6], MOEA/D-Epsilon [4], and C-TAEA [8]. Each algorithm is equipped with a different constraint-handling technique. Due to the stochastic nature of the algorithms, we conduct 31 individual runs of each algorithm on every given problem. This approach

allows us to obtain more precise values for algorithm performance. The target of the ML task for each problem is the mean AUC-ECDF value over all 31 runs of the algorithm. To facilitate the comparison of results, we use the same parameter settings for all algorithms – the population size $100 \cdot M$, and the number of generations $60 \cdot D$.

The ELA features are calculated stochastically; each time the feature calculation is started, a different sample of solutions is selected. To handle this, we created 30 samples using the Latin hypercube sampling method, resulting in 30 sets of features (learning instances) for each problem.

Predicting algorithm performance is a regression task and, therefore, we use regression ML methods – Linear Regression, Random Forest Regression (RF Regression) [2], and Epsilon-Support Vector Regression (SVR) [3]. We also included a dummy model in the comparison, which predicts the mean value of the target variable in the training data. We utilized the scikit-learn implementations [11] of these methods with default parameter settings. We tested algorithm parameter tuning as well, but there was no significant improvement of the results.

To evaluate the performance of the ML models, we use two evaluation methodologies – leave-one-sample-out and leave-one-problem-out. In the leave-one-sample-out evaluation, we use one instance as test data and the rest of the instances (including instances from other problems) as training data. We repeat this process for each instance in our dataset and take the average mean absolute error as an evaluation metric. Since all remaining instances of the problem are used during the training of the model, we expected the results from this evaluation methodology to be overly optimistic.

The leave-one-problem-out evaluation methodology is more fairly designed. In the real world, we have no information about the target problem available in the training data. Thus, in the leave-one-problem-out evaluation, we use all instances of a problem as test data and the instances from the rest of the problems as training data. This process is repeated for each problem in the dataset and the average mean absolute error is used as an evaluation metric.

6 RESULTS

The results showed a mean absolute error in the leave-one-sample-out evaluation lower than 0.01. This result is overly optimistic and shows that same-problem instances are similar to each other.

The results obtained in the leave-one-problem-out evaluation are presented in Table 1. These results show that none of the ML models performs significantly better than the dummy model. Moreover, because the target variable was normalized to $[0,1]$, a mean absolute error between 0.09 and 0.22 is large. This indicates that the tested models trained on the current benchmarks with the current ELA features cannot be used to predict algorithm performance accurately. Also, we note that for each problem dimensionality, there is a different ML method that performs best. For 2D problems this is Linear Regression, for 3D problems RF Regression, and for 5D problems SVR.

A significantly worse performance is achieved by Linear Regression on 5D problems. When attempting to understand the cause of this, we noticed that Linear Regression achieves similar results to the other models for all problems except for one, for which it performs very poorly. A possible explanation for this could be that Linear Regression is a simple and unbounded regression method, and when a problem is different from the

Table 1: Mean absolute error of the predicted AUC-ECDF values with respect to the true values for 2D, 3D, and 5D problems in leave-one-problem-out evaluation.

Dim	ML method	NSGA-III	MOEA/D	C-TAEA
2D	Dummy	0.18	0.17	0.18
	Linear Regression	0.16	0.14	0.18
	RF Regression	0.19	0.18	0.18
	SVR	0.20	0.21	0.19
3D	Dummy	0.14	0.12	0.13
	Linear Regression	0.22	0.12	0.15
	RF Regression	0.12	0.09	0.11
	SVR	0.14	0.12	0.13
5D	Dummy	0.14	0.10	0.12
	Linear Regression	0.70	0.42	0.75
	RF Regression	0.13	0.09	0.12
	SVR	0.10	0.09	0.10

rest in the training set it predicts very high target values, which increase the mean absolute error.

To better understand why the ML models performed poorly under the leave-one-problem-out evaluation, we used t-SNE [12] to reduce the dimensionality of the ELA features to 2D and visualized the results, as shown in Figure 1. Here we notice that samples from the same problem form clusters. This explains why the results of the leave-one-sample-out evaluation are significantly better than the leave-one-problem-out evaluation – in the former case, the ML task transforms into predicting the specific problem to which a sample belongs.

Analyzing the colors indicating the AUC-ECDF values of the three algorithms in Figure 1, we notice all algorithms perform similarly on almost all problems. This raises the question of whether a different algorithm parameter setting should be considered, emphasizing the differences in the performance of the algorithms. For example, we could check the algorithm performance on a smaller number of generations.

When analyzing the colors showing the AUC-ECDF values of an algorithm in a single dimension, we notice there is no visible pattern. This holds for each problem dimension-algorithm combination. Notably, we often find high and low AUC-ECDF values appearing close to each other in the plot.

The results show that, with the current benchmarks and ELA features, predicting algorithm performance is very difficult.

7 CONCLUSION

In this work, we attempted to predict the algorithm performance on CMOPs, using three well-known multiobjective optimization algorithms. For this purpose, we used ELA features specially designed for CMOPs as inputs to a ML model. To calculate the ELA features, we used 30 samples for each problem, resulting in 30 learning instances per problem. The target of prediction was the algorithm’s AUC-ECDF value, computed using the quality indicator designed explicitly for constrained multiobjective optimization [15].

We tested three ML regression methods – Linear Regression, RF Regression, and SVR. To compare the results from these methods, we also used a dummy model, which always predicts the mean value of the target variable in the training data. To evaluate the results, we used two evaluation methodologies – leave-one-sample-out and leave-one-problem-out.

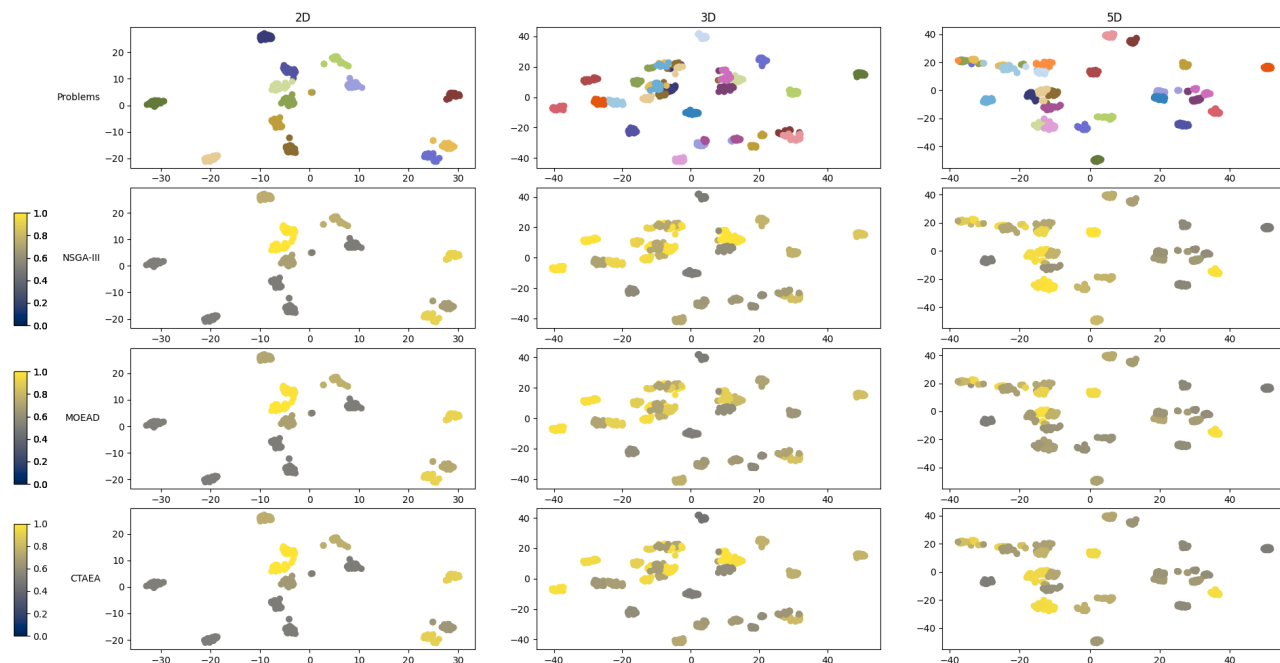


Figure 1: t-SNE visualizations of 2D, 3D and 5D problems. The colors in the first row of the plots represent the problems included in the benchmark. In the remaining rows, the colors represent the algorithm performance measured by AUC-ECDF for each algorithm considered.

In the leave-one-sample-out evaluation, very optimistic results were found, with a mean absolute error lower than 0.01. However, the results from the leave-one-problem-out evaluation were poor; none of the ML models significantly outperforms the dummy model. To explain why this occurs, we used the t-SNE method to reduce the dimensionality of the ELA features and plotted them in a color scheme indicating the performance of the algorithms. These visualizations show no visible patterns in the algorithm performance figures. Thus, we conclude that, with the currently available ELA features and benchmark problems, predicting algorithm performance is a hard task.

In future work, we aim to address two distinct aspects of the problem. The first is to improve the ELA features via automatic construction using an end-to-end deep neural network. The second is to reduce the complexity of the ML task by simplifying the target. This could be achieved by changing the task to a classification task or by changing the target to the number of generations required for an algorithm to reach a feasible solution.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support from the Slovenian Research and Innovation Agency (young researcher program, research core funding No. P2-0209, and project No. N2-0254 “Constrained Multiobjective Optimization Based on Problem Landscape Analysis”).

REFERENCES

- [1] Hanan Alsouly, Michael Kirley, and Mario Andrés Muñoz. 2022. An instance space analysis of constrained multi-objective optimization problems. *IEEE Transactions on Evolutionary Computation*. Early Access. <https://ieeexplore.ieee.org/abstract/document/9899751>.
- [2] Leo Breiman. 2001. Random forests. *Machine Learning*, 45, 5–32.
- [3] Harris Drucker, Christopher J Burges, Linda Kaufman, Alex Smola, and Vladimir Vapnik. 1996. Support vector regression machines. *Advances in Neural Information Processing Systems*, 9.
- [4] Zhun Fan, Wenji Li, Xinye Cai, Han Huang, Yi Fang, Yugen You, Jiajie Mo, Caimin Wei, and Erik Goodman. 2019. An improved epsilon constraint-handling method in MOEA/D for CMOPs with large infeasible regions. *Soft Computing*, 23, 12491–12510.
- [5] Nikolaus Hansen, Anne Auger, Dimo Brockhoff, and Tea Tušar. 2022. Any-time performance assessment in blackbox optimization benchmarking. *IEEE Transactions on Evolutionary Computation*, 26, 6, 1293–1305.
- [6] Himanshu Jain and Kalyanmoy Deb. 2013. An evolutionary many-objective optimization algorithm using reference-point based nondominated sorting approach, part II: Handling constraints and extending to an adaptive approach. *IEEE Transactions on Evolutionary Computation*, 18, 4, 602–622.
- [7] Anja Jankovic, Tome Eftimov, and Carola Doerr. 2021. Towards feature-based performance regression using trajectory data. In *Applications of Evolutionary Computation: 24th International Conference*. Springer, 601–617.
- [8] Ke Li, Renzhi Chen, Guangtao Fu, and Xin Yao. 2018. Two-archive evolutionary algorithm for constrained multiobjective optimization. *IEEE Transactions on Evolutionary Computation*, 23, 2, 303–315.
- [9] Zhongwei Ma and Yong Wang. 2019. Evolutionary constrained multiobjective optimization: Test suite construction and performance comparisons. *IEEE Transactions on Evolutionary Computation*, 23, 6, 972–986.
- [10] Ana Nikolikj, Carola Doerr, and Tome Eftimov. 2023. Rf+clust for leave-one-problem-out performance prediction. In *Applications of Evolutionary Computation: 26th International Conference*. Springer, 285–301.
- [11] F. Pedregosa et al. 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- [12] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9, 11.
- [13] Diederick Vermetten, Hao Wang, Thomas Bäck, and Carola Doerr. 2020. Towards dynamic algorithm selection for numerical black-box optimization: Investigating bbob as a use case. In *Proceedings of the 2020 Genetic and Evolutionary Computation Conference*, 654–662.
- [14] Aljoša Vodopija, Tea Tušar, and Bogdan Filipič. 2022. Characterization of constrained continuous multiobjective optimization problems: A feature space perspective. *Information Sciences*, 607, 244–262.
- [15] Aljoša Vodopija, Tea Tušar, and Bogdan Filipič. 2023. Characterization of constrained continuous multiobjective optimization problems: A performance space perspective. *arXiv preprint arXiv:2302.02170*.
- [16] Qingfu Zhang, Aimin Zhou, Shizheng Zhao, Ponnuthurai Nagaratnam Suganthan, Wudong Liu, Santoshi Tiwari, et al. 2008. Multiobjective optimization test instances for the CEC 2009 special session and competition, 1–30.
- [17] Eckart Zitzler, Lothar Thiele, Marco Laumanns, Carlos M Fonseca, and Viviane Grunert Da Fonseca. 2003. Performance assessment of multiobjective optimizers: An analysis and review. *IEEE Transactions on Evolutionary Computation*, 7, 2, 117–132.

Večstopenjski postopek vrednotenja rešitev pri načrtovanju elektromotorja

A Multi-Step Evaluation Process in Electric Motor Design

Tea Tušar
tea.tusar@ijs.si
Institut "Jožef Stefan" in
Mednarodna podiplomska šola
Jožefa Stefana
Ljubljana, Slovenija

Peter Korošec
peter.korosec@ijs.si
Institut "Jožef Stefan"
Ljubljana, Slovenija

Bogdan Filipič
bogdan.filipic@ijs.si
Institut "Jožef Stefan" in
Mednarodna podiplomska šola
Jožefa Stefana
Ljubljana, Slovenija

POVZETEK

Pri načrtovanju elektromotorja je potrebno poiskati vrednosti optimizacijskih spremenljivk tako, da izdelek izpolnjuje tehnične zahteve in je njegova cena minimalna. V ta namen uporabljamo optimizacijske postopke z iterativnim vrednotenjem rešitev na osnovi numerične simulacije. Ti so računsko zahtevni, zato je glavni izziv načrtovanja, kako najti kakovostne rešitve v sprejemljivem času. V prispevku predstavljamo računalniško podprto načrtovanje sinhronskega elektromotorja za servovolanske sisteme, s poudarkom na prijemih za pohitritev optimizacijskega postopka. Med njimi je tudi posebej za ta problem razvit večstopenjski postopek vrednotenja rešitev, ki omogoča učinkovitost optimizacije in robustnost rešitev. S tem postopkom razviti elektromotor je boljši od prvotnega prototipa, dobljenega z enostavnejšim optimizacijskim postopkom, tako po tehničnih lastnostih kot stroškovno.

ABSTRACT

In the design of an electric motor, one has to find the values of the optimization variables such that the product satisfies the technical requirements and its price is minimal. For this purpose, we deploy optimization procedures with iterative evaluation of solutions based on numerical simulation. These are time-consuming, hence the key challenge of the design is how to find high-quality solutions in an acceptable time. In this paper, we present the computer-aided design of a synchronous electric motor for power steering systems, with an emphasis on measures for speeding up the optimization process. Among them, a multi-step solution evaluation procedure has been developed particularly for this problem. It enables the efficiency of optimization and the robustness of solutions. The resulting electric motor outperforms the original prototype obtained by a simpler optimization procedure both in technical characteristics and cost efficiency.

KLJUČNE BESEDE

načrtovanje, elektromotor, numerična simulacija, optimizacija, evolucijski algoritem, robustnost

KEYWORDS

design, electric motor, numerical simulation, optimization, evolutionary algorithm, robustness

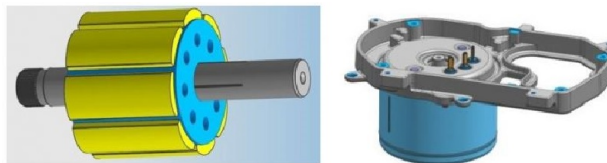
Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Information Society 2023, 9–13 October 2023, Ljubljana, Slovenia

© 2023 Copyright held by the owner/author(s).

1 UVOD

Podjetje MAHLE Electric Drives Slovenija d.o.o. proizvaja zagajalnike in alternatorje za motorje z notranjim izgorevanjem, avtonomno napajane enosmerne električne pogonske sisteme in druge zahtevnejše komponente za avtomobilsko industrijo. Eden izmed pomembnejših izdelkov podjetja je sinhronski elektromotor s površinsko nameščenimi magneti, ki poganja avtomobilski servovolanski sistem (slika 1).



Slika 1: Elektromotor za servovolanske sisteme (Vir: arhiv Mahle Electric Drives Slovenija d.o.o.).

Razvoj takšnega elektromotorja zahteva določitev geometrije in materialnih lastnosti njegovih komponent tako, da bo izpolnjeval vse tehnične zahteve in bo njegova cena minimalna. Ker se pri vrednotenju načrtov elektromotorja uporablja numerični simulator po metodi končnih elementov, v katerega nimamo vpogleda (gre za t.i. problem črne škatle, ang. black box-problem), optimizacijski postopek terja iterativno vrednotenje številnih načrtov. Numerične simulacije so dolgotrajne, zato je glavni izziv razvoja elektromotorja optimizacijski postopek zastaviti tako, da bo lahko našel dobre rešitve v doglednem času.

V nadaljevanju prispevka opisujemo, kako smo se načrtovanja elektromotorja (2. razdelek) lotili na Inštitutu "Jožef Stefan" v sodelovanju s podjetjem MAHLE. Optimizacijski postopek smo pohitrili s tremi prijemi:

- uporabili smo optimizacijski algoritem s hitro izračunljivimi nadomestnimi modeli (3. razdelek);
- vrednotenje rešitev smo razdelili na več korakov in tako omogočili izločanje nedopustnih rešitev pred dolgotrajnimi simulacijami (4. razdelek);
- računsko najzahtevnejši korak vrednotenja rešitev smo poenostavili in paralelizirali (4. razdelek).

Metodologijo smo preizkusili na konkretnem tipu elektromotorja (5. razdelek). Tako optimiran načrt elektromotorja je dosegel 10 % nižjo ceno komponent v primerjavi z različico, ki jo je podjetje predhodno razvilo s preprostejšim optimizacijskim postopkom. V prihodnje nameravamo računalniško implementacijo postopka razširiti tako, da bo omogočala prosto izbiro optimizacijskega kriterija in omejitev ter bo uporabna za raznovrstne elektromotorje (6. razdelek).

2 NAČRTOVANJE ELEKTROMOTORJA

Pri načrtovanju elektromotorja moramo nastaviti vrednosti številnih parametrov, ki določajo njegovo geometrijo in materialne lastnosti njegovih komponent. To so na primer dimenzije zoba na statorju, število navojev tuljave, dimenzije magnetov ipd. Vsak parameter ima podano spodnjo in zgornjo mejo ter najmanjši smiselni korak znotraj teh meja. Cilj optimizacije načrtovanja je najti načrt elektromotorja, ki zadošča vsem tehničnim zahtevam in je ob enem najcenejši.

Načrt elektromotorja lahko ocenimo na več načinov, z različno stopnjo zaupanja. Njegove lastnosti lahko najzanesljiveje preverimo, če na podlagi načrta izdelamo prototip elektromotorja in ga preizkusimo v praksi. Vendar to zahteva veliko dela, materiala in časa ter s tem povezane visoke stroške, ki ne dovoljujejo, da bi podjetje v fazi razvoja izdelalo večje število prototipov. Zato si pri reševanju tega problema pomagamo z računalniško podprtimi numeričnimi simulacijami, na osnovi katerih lahko izpeljemo ključne lastnosti elektromotorja. Računalniški programi, kot je Ansys Maxwell [1], omogočajo simulacijo elektromagnetnega polja elektromotorja z uporabo metode končnih elementov [2]. Ta deluje na podlagi mreže objekta; gostejša mreža omogoča večjo točnost simulacije, a je ta dolgotrajnejša. Zanesljivost numeričnih simulacij je tako deloma nastavljiva – odvisna je od računalniških zmogljivosti in časa, ki jih imamo na voljo.

Vendar zanesljivost simulacij znižujejo praktični vidiki izdelave elektromotorja, saj lahko ujemanje izdelanega elektromotorja z načrtom zagotovimo samo v okviru določenih toleranc. Na primer, če za velikost odprtine reže nastavimo vrednost 2 mm, lahko v proizvodnji zagotovimo le, da bo ta na intervalu [1.95 mm, 2.05 mm]. To pomeni, da je za načrt elektromotorja zelo pomembno, da je *robusten*, to je, da ob majhnih spremembah vrednosti parametrov znotraj toleranc lastnosti elektromotorjev ne odstopajo bistveno. Robustnost načrta je najlažje preveriti s simuliranjem številnih načrtov, ki se malo razlikujejo od izhodiščnega. Vendar to zahteva še več računsko zahtevnih simulacij in podaljšuje trajanje optimizacijskega postopka.

3 OPTIMIZACIJSKI POSTOPEK

V optimizacijskem postopku skušamo čim učinkoviteje rešiti dani optimizacijski problem. Formalno (in brez škode za splošnost) lahko optimizacijski problem načrtovanja elektromotorja zapišemo v obliki

$$\begin{aligned} &\text{minimiziraj } f(x), \\ &\text{ob pogojih } g_i(x) \leq 0, \quad i = 1, \dots, k, \end{aligned} \quad (1)$$

kjer je $x = (x_1, \dots, x_n) \in X \subseteq \mathbb{R}^n$ rešitev iz n -dimenzionalnega prostora rešitev X , $f : X \rightarrow \mathbb{R}$ kriterijska funkcija, funkcije $g_i : X \rightarrow \mathbb{R}$, $i = 1, \dots, k$, pa so *stroge omejitve*. Rešitev problema (načrt elektromotorja) je *dopustna*, če zadošča vsem strogim omejitvam. Sicer pravimo, da je *nedopustna*.

Pri nedopustnih rešitvah ne moremo vedno izračunati vrednosti kriterija. Zato takrat kriterij f nadomestimo s funkcijo

$$f_g(x) = p + \sum_{i=1}^k \max\{g_i(x), 0\}, \quad (2)$$

kjer je p konstanta in $\max\{g_i(x), 0\}$ nenegativna kazen za kršitev stroge omejitve g_i (ko omejitev, ni kršena, kazen znaša 0). Konstanta p mora biti dovolj velika, da je vrednost kriterija f_g za katerokoli nedopustno rešitev vedno višja (slabša) od vrednosti kriterija f za katerokoli dopustno rešitev.

Poleg strogih omejitev imajo optimizacijski problemi v praksi pogosto tudi *šibke omejitve*. To so funkcije $h_i : X \rightarrow \mathbb{R}$, $i = 1, \dots, l$, za katere želimo, da velja $h_i(x) \leq 0$, ni pa to pogoj, da je rešitev dopustna. V optimizacijskem problemu jih upoštevamo tako, da jih vgradimo v kriterijsko funkcijo na naslednji način

$$f_h(x) = f(x) + \sum_{i=1}^l \max\{h_i(x), 0\}, \quad (3)$$

kjer je f prvotna kriterijska funkcija, $\max\{h_i(x), 0\}$ pa nenegativna kazen za kršitev šibke omejitve h_i . Ker seštevamo prvotno kriterijsko funkcijo in kazni za kršitev šibkih omejitev, moramo zagotoviti, da so njihove vrednosti primerljive. V ta namen jih je potrebno normalizirati oz. primerno utežiti.

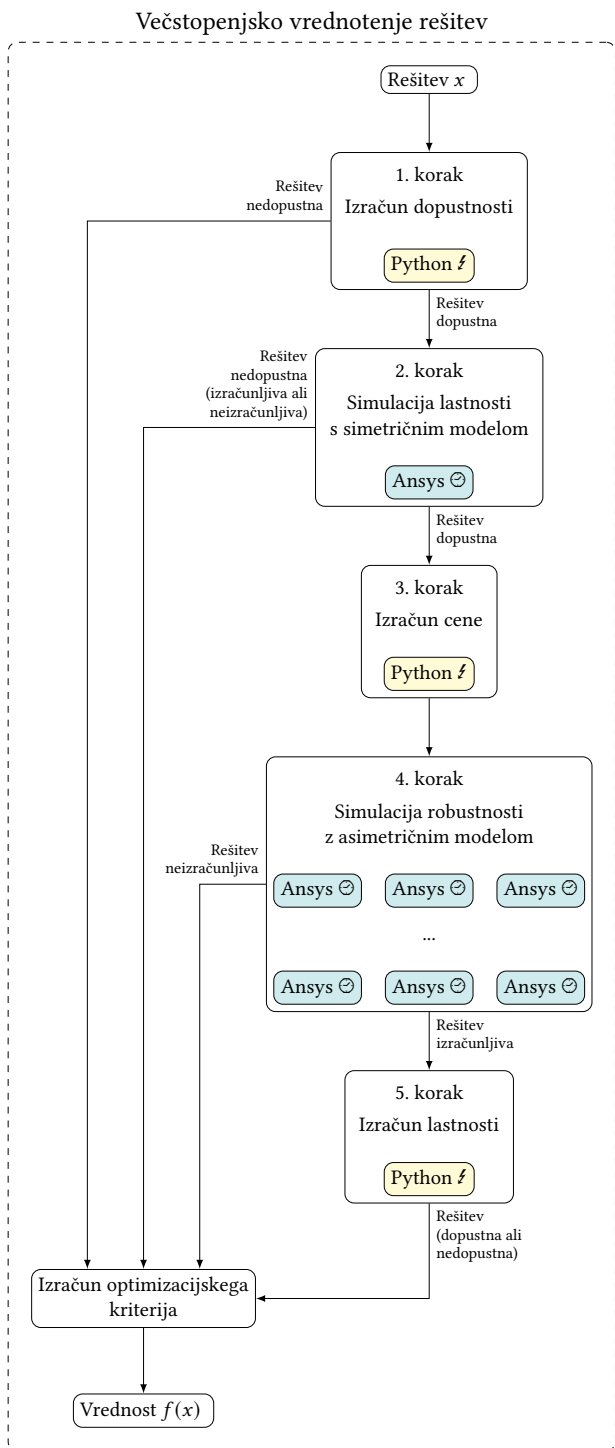
Pri problemu načrtovanja elektromotorja ne poznamo analitične oblike kriterijske funkcije in omejitev, zato za njegovo reševanje uporabimo algoritem, ki se dobro obnese na problemih črne škatle. To je evolucijska strategija s prilagajanjem kovarične matrike, oz. CMA-ES (ang. Covariance Matrix Adaptation Evolution Strategy) [5]. Natančneje, poslužujemo se različice algoritma CMA-ES imenovane Iq-CMA-ES [3], ki vrednotenja z dolgotrajnimi simulacijami delno zamenja s hitro izračunljivimi linearno-kvadratičnimi nadomestnimi modeli. Uporaba nadomestnih modelov je pogosto uporabljen pristop pri reševanju problemov z računsko zahtevnim vrednotenjem rešitev.

Optimizacijski algoritem z nadomestnimi modeli Iq-CMA-ES deluje v dveh fazah. Najprej na podlagi začetnih rešitev, ovrednotenih z numerično simulacijo, zgradi nadomestni model. Nato iterativno predlaga nove rešitve in jih, dokler so te dovolj podobne obstoječim, vrednoti z nadomestnim modelom ter tako prihrani na času. Ko pride do rešitev, ki jih nadomestni model ne opisuje več dovolj dobro, pa jih ovrednoti z numerično simulacijo in z njihovim rezultatom posodobi nadomestni model. To ponavlja, dokler ne izpolni zaustavitvenega pogoja.

4 VREDNOTENJE POSAMEZNEGA NAČRTA ELEKTROMOTORJA

Uporaba nadomestnih modelov zmanjša število izvedenih vrednotenj s simulacijami, a so te pri načrtovanju elektromotorja vseeno potrebne. Da bi vrednotenje posameznega načrta pohitrili, smo ga razdelili na pet korakov. Če se rešitev slabo izkaže po katerem od korakov, jo takoj zavržemo in s tem prihranimo čas, ki bi ga sicer porabili za izvedbo preostalih korakov. Vrednotenje je orisano na sliki 2 in opisano v nadaljevanju.

- (1) V prvem koraku s hitro izračunljivo skripto, ki so jo pripravili domenski eksperti, preverimo zadoščanje nekaterim strogim omejitvam. To nam pomaga izločiti precejšnje število nedopustnih rešitev, med njimi tudi take, ki bi lahko zaradi slabo zasnovane geometrije povzročale težave pri izvedbi simulacije. Samo dopustne rešitve gredo v naslednji korak vrednotenja.
- (2) V drugem koraku izvedemo simulacijo, s katero pridobimo informacije o določenih pomembnih lastnostih elektromotorja. Ta uporablja poenostavljeno, simetrično formulacijo geometrije elektromotorja, zato je relativno hitra (za izbrani problem traja približno 1 minuto). Trajanje simulacije je sicer odvisno od posameznih rešitev. Pri načrtih, za katere je geometrija zaradi neposrečene kombinacije vrednosti parametrov nemogoča, simulator lahko po dolgem času ne vrne nobenega rezultata ali pa se (redko) celo zruši. Če rešitev ni izvedljiva ali ne zadošča strogim omejitvam, jo zavržemo. Sicer nadaljujemo z naslednjim korakom.



Slika 2: Vrednotenje rešitev v petih korakih.

- (3) V tretjem koraku iz vseh dobljenih podatkov izračunamo ceno elektromotorja.
- (4) V četrtem koraku preverimo robustnost rešitve. Pri tem se simulacije izvajajo na celotni, asimetrični geometriji elektromotorja, zato so časovno bolj potratne. Pohitrimo jih z uporabo manj natančne mreže (predhodni poskusi so pokazali, da lahko nekaj zanesljivosti žrtvujemo za precejšnji prihranek časa; tako za izbrani problem ena simulacija traja 7 namesto 15 minut). Ker moramo za preverjanje

robustnosti ene rešitve izvesti več simulacij, ki so med seboj neodvisne, to izkoristimo za njihovo paralelizacijo. To izvedemo tako, da vse simulacije z asimetričnim modelom elektromotorja poženemo hkrati na računalniku z večjedrnim procesorjem in tako prihranimo veliko časa. Če preverjanje robustnosti mine brez težav, nadaljujemo z zadnjim korakom postopka.

- (5) V zadnjem, petem koraku izračunamo še dodatne lastnosti elektromotorja in preverimo dopustnost preostalih strogih omejitev.

Za vsako rešitev x po izvedbi opisanega postopka izračunamo vrednost optimizacijskega kriterija. Ta je odvisna od koraka, do katerega se je izvedlo vrednotenje rešitve, in njene kakovosti:

$$f(x) = \begin{cases} 300 + \sum_{i=1}^{k_1} \max\{g_i^1(x), 0\} & x \text{ nedopustna po 1. kor.} \\ 250 & x \text{ neizračunljiva v 2. kor.} \\ 200 + \sum_{i=1}^{k_2} \max\{g_i^2(x), 0\} & x \text{ nedopustna po 2. kor.} \\ 150 & x \text{ neizračunljiva v 4. kor.} \\ 100 + \sum_{i=1}^{k_5} \max\{g_i^5(x), 0\} & x \text{ nedopustna po 5. kor.} \\ c(x) + \sum_{i=1}^l \max\{h_i(x), 0\} & x \text{ dopustna} \end{cases} \quad (4)$$

Pri tem g^1, g^2 in g^5 po vrsti predstavljajo stroge omejitve v 1., 2. in 5. koraku, c cena rešitve in h_i njene šibke omejitve. Kazni za kršitev obeh vrst omejitev g_i in h_i so normalizirane tako, da njihova vsota nikoli ne preseže vrednosti 50. Na ta način poskrbimo, da so dopustne rešitve vedno ocenjene boljše od nedopustnih in je rešitev, ki je šla čez več korakov postopka vrednotenja, ocenjena boljše od tiste, ki se je slabo izkazala v katerem od prejšnjih korakov.

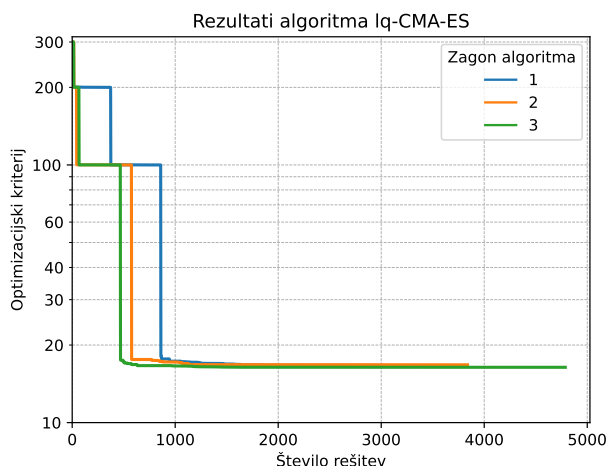
Opisani večstopenjski postopek vrednotenja rešitev s pohitritvami je glavna novost, ki smo jo v okviru sodelovanja med inštitutom in podjetjem uvedli v optimizacijo načrtovanja elektromotorja. Razvijalci v podjetju so že prej uporabljali optimizacijo z nadomestnimi modeli v okviru programskega orodja Ansys [1], vendar pa so v njej lahko upoštevali samo 2. in 3. korak vrednotenja. Začetnega preverjanja dopustnosti (1. koraka) ni bilo, zadnjih dveh korakov pa se ni dalo enostavno vključiti v optimizacijski postopek, zato sta se izvedla šele po zaključku optimizacije na izbranih (najboljših) rešitvah optimizacijskega problema.

5 NUMERIČNI POSKUSI IN REZULTATI

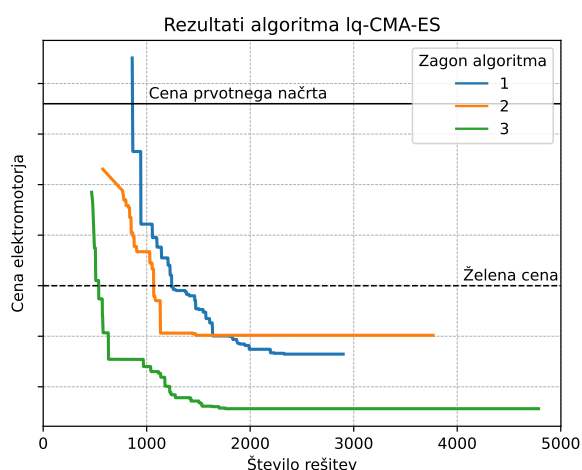
Optimizacijski postopek smo preizkusili na konkretnem primeru elektromotorja za servovolanske sisteme. Ta ima 13 optimizacijskih spremenljivk, od katerih je 12 'zveznih' in ena celoštevilska¹. Optimizacijski kriterij (za dopustne rešitve) sestavljata cena in vsota kršitev dveh šibkih omejitev, medtem ko je vseh strogih omejitev 10. Pri simulaciji robustnosti z asimetričnim modelom (4. korak) smo hkrati poganjali 15 simulacij.

V optimizacijskih izračunih smo uporabili knjižnico pycma, ki nudi implementacijo algoritma lq-CMA-ES v programskem jeziku Python [4]. Algoritmu smo posredovali informacijo o tem, da je ena od spremenljivk celoštevilska, ostale je obravnaval kot zvezne. Velikost začetnega koraka σ_0 smo nastavili na 0,2, dovoljeno število avtomatskih ponovnih zagonov algoritma pa na 5. Vrednosti preostalih parametrov algoritma so bile enake privzetim.

¹Tudi 'zvezne' spremenljivke so zaradi najmanjšega smiselnega koraka v resnici diskretne, a jih obravnavamo kot zvezne s stališča algoritma (preden rešitev ovrednotimo, jo zaokrožimo na najbližjo diskretno vrednost).



Slika 3: Najnižja vrednost optimizacijskega kriterija tekom treh zagonov algoritma lq-CMA-ES.



Slika 4: Primerjava najnižje cene elektromotorja tekom treh zagonov algoritma lq-CMA-ES s ceno prvotnega optimiranega načrta in zeleno ceno. (Dejanske cene niso navedene zaradi varovanja poslovne skrivnosti.)

Zaradi časovnih omejitev smo uspeli izvesti samo tri zagone algoritma lq-CMA-ES, ki se razlikujejo v semenu generatorja ključnih števil. Njihovi rezultati so prikazani na slikah 3 in 4. Obe kažeta, kako se opazovana veličina zmanjšuje s časom (številom pregledanih rešitev – štejemo vsako vrednotenje rešitev, tudi če se je končalo že s prvim korakom). Na sliki 3 vidimo zmanjševanje optimizacijskega kriterija (vrednosti pod 100 pomenijo, da je algoritem našel dopustne rešitve), na sliki 4 pa pripadajoče (do takrat najboljše) cene elektromotorjev.

Vidimo, da so rezultati treh zagonov algoritma precej raznoliki. Pri tako zahtevnem problemu in majhnem številu vrednotenj je bilo to pričakovano. Vseeno pa za vse tri zagone velja, da potrebujejo manj kot 1000 pregledanih rešitev, da najdejo dopustne rešitve in hkrati izboljšajo najboljši načrt, najden s prvotnim optimizacijskim postopkom v podjetju. V vseh treh primerih je potrebnih manj kot 1500 vrednotenj, da algoritem najde rešitev s ceno, ki je boljša od zelene. Po približno 2500 vrednotenjih se cena elektromotorja neha bistveno izboljševati.

Kvantitativnega ovrednotenja doprinosa posameznih pohitritev nismo izvedli, vemo pa, da se je vrednotenje rešitev končalo po 1. koraku v 17,5 % primerov (8 min prihranka na rešitev) in po 2. koraku v 29,4 % primerov (7 min prihranka na rešitev).

Ker smo analizo robustnosti, ki se izvede v 4. koraku, precej poenostavili, da smo jo pohitрили in vključili v optimizacijo, smo preverili, če so rezultati 4. in 5. koraka skladni s tistimi, ki bi jih dobili z prvotno analizo. Zato smo nekaj izbranih (najboljših) rešitev podrobneje analizirali in njihove rezultate primerjali s tistimi, ki veljajo za prvotni optimirani načrt. Ugotovili smo, da dobimo kakovostne rešitve, ki niso le dopustne, ampak so za 10 % cenejše od prvotnega najboljšega načrta.

6 ZAKLJUČKI

Problem načrtovanja elektromotorja je zahteven praktičen optimizacijski problem, saj je mnogo načrtov nedopustnih, njihovo vrednotenje pa temelji na numeričnih simulacijah in je zato dolgotrajno. Da bi ga lahko uspešno reševali z uporabo optimizacijskih algoritmov, smo vrednotenje rešitev razdelili na pet korakov, s katerimi želimo čim prej izločiti nedopustne rešitve, da na njih ne tratimo časa. Poleg tega smo optimizacijski postopek pohitрили z uporabo algoritma s hitro izračunljivimi nadomestnimi modeli ter paralelizacijo in poenostavitvijo računsko najzahtevnejšega koraka vrednotenja rešitev.

Predlagani postopek smo preizkusili na konkretnem primeru elektromotorja za servovolanske sisteme, za katerega smo želeli minimizirati ceno in obenem zagotoviti, da bo zadoščal vsem tehničnim zahtevam. Primerjava tako dobljenih načrtov z najboljšim, ki so ga v podjetju prvotno našli s pomočjo enostavnejšega optimizacijskega postopka, je pokazala, da dosežemo dopustne rešitve, ki so cenovno za 10 % ugodnejše od obstoječih. Ob dejstvu, da se v celotnem obdobju proizvodnje takšnega izdelka proizvede več milijonov kosov, to za podjetje predstavlja bistven prihranek in močno izboljšuje njegovo konkurenčnost na trgu.

V prihodnje želimo implementacijo postopka razširiti tako, da bo omogočala enostavno izbiro optimizacijskega kriterija in omejitev ter dodajanje skript za preverjanje dopustnosti rešitev. Cilj je izdelati računalniško orodje, ki ga bodo inženirji lahko samostojno uporabljali pri optimizaciji raznovrstnih elektromotorjev brez poseganja v sam postopek ali optimizacijski algoritem.

ZAHVALA

Zahvaljujemo se podjetju MAHLE Electric Drives Slovenija d.o.o. za financiranje projekta razvoja elektromotorja in sodelovanje pri njegovi izvedbi. Naše temeljne raziskave evoliucijskega računanja in večkriterijske optimizacije sofinancira Javna agencija za znanstvenoraziskovalno in inovacijsko dejavnost Republike Slovenije z raziskovalnima programoma št. P2-0098 in P2-0209 ter projekti št. J2-4460, N2-0239 in N2-0254.

LITERATURA

- [1] Ansys, Inc. 2023. Ansys Maxwell. <https://www.ansys.com/products/electronics/ansys-maxwell>. Dostopano 15. 8. 2023. (2023).
- [2] Klaus-Jürgen Bathe. 2007. Finite element method. *Wiley encyclopedia of computer science and engineering*, 1–12.
- [3] Nikolaus Hansen. 2019. A global surrogate assisted CMA-ES. *V Proceedings of the 2019 Genetic and Evolutionary Computation Conference*. ACM, New York, NY, USA, 664–672. doi: 10.1145/3321707.3321842.
- [4] Nikolaus Hansen, Youhei Akimoto in Petr Baudis. 2019. CMA-ES/pycma on Github. (Feb. 2019). doi: 10.5281/zenodo.2559634.
- [5] Nikolaus Hansen in Andreas Ostermeier. 1996. Adapting arbitrary normal mutation distributions in evolution strategies: the covariance matrix adaptation. *V Proceedings of 1996 IEEE International Conference on Evolutionary Computation*. IEEE, 312–317. doi: 10.1109/ICEC.1996.542381.

Indeks avtorjev / Author index

Andova Andrejaana.....	44
Antešić Ivan	40
Arbetman Marina	20
Bohanec Marko	36
Bolliger Larissa	11
Božič Janko	24
Campopiano Robinson Victoria	20
Clays Els.....	11
Cork Jordan	44
Filipič Bogdan.....	44, 48
Galen Candace	20
Gams Matjaž	7, 28, 32
Gradišek Anton	20
Guček Marko.....	36
Herrera Valentina	20
Hirose Akira.....	15
Jordan Marko	19
Kawai Kitoshi.....	15
Kocuvan Primož.....	28
Kolar Žiga	32
Kontić Davor.....	36
Korošec Peter	48
Lukan Junoš	11
Lukšič Andrej A.	32
Luštrek Mitja.....	11, 15, 19
Martinšek Marcel Franse.....	11
Rotar Oskar	24
Sadikov Aleksandar.....	40
Schul Johannes	20
Silan Darja.....	24
Sirk Karina	36
Šiško Primož	11
Šket Tilen	20
Slapničar Gašper	15
Susič David	7, 20
Tušar Tea.....	44, 48
Valič Jakob.....	19
Vesel Tian	24
Villagra Gil Cristian Alfonso	20
Vodopija Aljoša	44
Vozlič Andrej A.	32
Zadobovšek Matic	28
Ženko Bernard.....	36
Žnidaršič Maks.....	24
Žnidaršič Martin.....	36



**Slovenska konferenca o
umetni inteligenci**

**Slovenian Conference on
Artificial Intelligence**

Uredniki • Editors:

Mitja Luštrek, Matjaž Gams, Rok Piltaver