Advances in Intelligent Systems and Computing 1281

Ján Paralič
Peter Sinčák
Pitoyo Hartono
Vladimír Mařík *Editors*

Towards Digital Intelligence Society

A Knowledge-based Approach

Springer

Advances in Intelligent Systems and Computing

Volume 1281

The series "Advances in Intelligent Systems and Computing" contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing such as: computational intelligence, soft computing including neural networks, fuzzy systems, evolutionary computing and the fusion of these paradigms, social intelligence, ambient intelligence, computational neuroscience, artificial life, virtual worlds and society, cognitive science and systems, Perception and Vision, DNA and immune based systems, self-organizing and adaptive systems, e-Learning and teaching, human-centered and human-centric computing, recommender systems, intelligent control, robotics and mechatronics including human-machine teaming, knowledge-based paradigms, learning paradigms, machine ethics, intelligent data analysis, knowledge management, intelligent agents, intelligent decision making and support, intelligent network security, trust management, interactive entertainment, Web intelligence and multimedia.

The publications within "Advances in Intelligent Systems and Computing" are primarily proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Indexed by SCOPUS, DBLP, EI Compendex, INSPEC, WTI Frankfurt eG, zbMATH, Japanese Science and Technology Agency (JST), SCImago.

All books published in the series are submitted for consideration in Web of Science.

More information about this series at http://www.springer.com/series/11156

Ján Paralič · Peter Sinčák ·
Pitoyo Hartono · Vladimír Mařík
Editors

Towards Digital Intelligence Society

A Knowledge-based Approach

Springer

Editors
Ján Paralič
Department of Cybernetics and AI
Technical University of Košice
Košice, Slovakia

Peter Sinčák
Department of Cybernetics and AI
Technical University of Košice
Košice, Slovakia

Pitoyo Hartono
Department of Electrical
and Electronics Engineering
Chukyo University
Nagoya, Japan

Vladimír Mařík
Czech Institute of Informatics, Robotics
and Cybernetics
Czech Technical University in Prague
Prague 6, Czech Republic

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-3-030-63871-9 ISBN 978-3-030-63872-6 (eBook)
https://doi.org/10.1007/978-3-030-63872-6

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Humanity is now going through difficult times to fight the Covid-19 pandemic. Simultaneously, in these difficult times of physical separation, we can also realize how much digital society technology helps us cope with many difficulties that bring us this time. This book aims to provide readers with up-to-date knowledge on how to make these technologies smarter. We focus on selected research challenges for intelligent digital society and state-of-the-art methods how to face them.

The book's subtitle suggests that a core concept that the reader can study from various points of view in particular chapters is the knowledge. The knowledge that can help us intelligently face different digital society challenges (Part I of this book); the knowledge extracted from available big data employing intelligent analysis techniques (Part II). For efficient processing and analysis of data, there is a strong need for smart data and information modeling techniques (Part III).

This book was created within the context of the project entitled Knowledge-based Approaches for Intelligent Analysis of Big Data supported by the Slovak Research and Development Agency as project no. APVV-16-0213.

The process of preparation of this book had four phases. In the first one, potential authors could submit a one-page abstract of the proposed chapter in line with the main topics described in the call for chapters proposals. We received altogether 15 extended abstracts, all of them fitting pretty well into the book's theme. We afterward invited the authors to prepare their full chapter manuscripts, resulting in 14 chapter manuscripts submissions. In the next phase, two to three reviewers carefully reviewed each of the submitted chapter manuscripts. Based on the reviews, we accepted nine chapters. In the last phase, authors of accepted chapters prepared the final versions of their manuscripts, carefully incorporating the reviewers' suggestions. We are proud to present to you the result of this strict and demanding book preparation process.

We divided the content of the book into three parts, each one consisting of three chapters:

I. Digital Intelligence: Selected Research Challenges
II. Intelligent Analysis of Big Data
III. Data and Information Modelling

Part I presents selected challenges toward digital intelligence society. As the first challenge, the authors K. Machova et al. describe various forms of antisocial behavior in the online environment (mostly on social media), especially the creation and spreading of false information and using abusive language. Moreover, they also present intelligent approaches on how to cope with this challenge. The second chapter in this part of the book, written by I. Cik et al., focuses on five research challenges in advanced computer vision, where convolutional neural networks can help cope with these challenges intelligently. The last challenge presented by P. Blažek et al. is the combination of human intelligence with the power of machines to facilitate workloads or enable the mobility of their users.

Part II focuses on intelligent data analysis, as one of the pillars for the digital intelligence society. P. Bednar et al. present in their chapter two main types of knowledge-based approaches to data analysis. One of them is semantic modeling of data analytics processes, which can efficiently cover the explicit form of background knowledge. The second one is typical for medical applications, where a principal amount of background knowledge tends to stay tacit. In such a situation, human-in-the-loop approach is a way how to perform data analysis intelligently. V. Rozinajova et al. focus on intelligent analysis of data streams, particularly on two frequent data mining tasks, prediction and optimization. They provide an excellent summarization of their research work's selected outcomes, verified in the power engineering area in solving tasks of smart grid optimization. D. Andrešič et al. dive deeper into astronomical time series data analysis. They present an original approach that includes artificial neural networks enhanced by an evolutionary algorithm to find and learn the best performing classification model for pre-processed light curves.

Part III picks up some important data and information modeling challenges and solutions. V. Jirkovsky et al. present one of the key concepts of Industry 4.0, the seamless integration of various information resources providing a ubiquitous knowledge management system. L. Antoni et al. describe the computational methods for the generation of attribute implications in formal concept analysis and illustrate the process of attribute exploration for special types of object-attributes tables. Moreover, they proposed an own test for measuring pupils' computational thinking in several consecutive stages and collected data from pupils' national testing in Slovakia. M. Kvet and K. Matiasko cope in their chapter with one particular problem of database optimization. They propose an extension of the join operation method if the foreign key index is not present.

We want to thank all the authors for their valuable contributions to this book and all the reviewers for careful assessment and detailed reviews provided to the authors. All of them helped make this book an exciting work about selected topics on knowledge-based approaches suitable for solving some of the challenges of digital intelligence society.

Ján Paralič
Peter Sinčák
Pitoyo Hartono
Vladimír Mařík

Organization

Program Chairs

Ján Paralič	Technical University of Koáice, Slovakia
Vladimír Mařík	Czech Technical University in Prague, Czech Republic
Pitoyo Hartono	Chukyo University, Japan
Peter Sinčák	Technical University of Košice, Slovakia

Program Committee

Gabriela Andrejková	Pavol Jozef Šafárik University in Košice, Slovakia
Ľubomír Antoni	Pavol Jozef Šafárik University in Košice, Slovakia
František Babič	Technical University of Košice, Slovakia
Peter Bednár	Technical University of Košice, Slovakia
Ivana Budinská	Slovak Academy of Sciences, Slovakia
Marek Bundzel	Technical University of Košice, Slovakia
Radek Burget	Brno University of Technology, Czech Republic
Ľuboš Buzna	University of Žilina, Slovakia
Dalibor Fiala	University of West Bohemia, Czech Republic
Pitoyo Hartono	Chukyo University, Japan
Pavol Jancura	Eindhoven University of Technology, Netherlands
Aleš Janota	University of Žilina, Slovakia
Petr Kadera	Czech Technical University in Prague, Czech Republic
Stanislav Krajč	Pavol Jozef Šafárik University in Košice, Slovakia

Lenka Lhotská	Czech Technical University in Prague, Czech Republic
Vladimír Mařík	Czech Technical University in Prague, Czech Republic
Karel Mls	University of Hradec Králové, Czech Republic
Petr Novák	Czech Technical University in Prague, Czech Republic
Marek Obitko	Rockwell Automation, Czech Republic
Ján Paralič	Technical University of Košice, Slovakia
Libor Přeučil	Czech Technical University in Prague, Czech Republic
Jan Rauch	University of Economics, Prague, Czech Republic
Viera Rozinajová	Slovak University of Technology in Bratislava, Slovakia
Martin Sarnovský	Technical University of Košice, Slovakia
Peter Sinčák	Technical University of Košice, Slovakia
Pavel Tichý	Rockwell Automation, s.r.o., Czech Republic
Ján Vaščák	Technical University of Košice, Slovakia
Hiroaki Wagatsuma	Kyushu Institute of Technology, Japan
Petr Šaloun	Palacky University Olomouc, Czech Republic
Jan Šedivý	Czech Technical University in Prague, Czech Republic

Contents

Data and Information Modelling

Digital Intelligence: Selected Research Challenges

Addressing False Information and Abusive Language in Digital Space Using Intelligent Approaches

Kristina Machova[1](✉), Ivan Srba[2], Martin Sarnovský[1], Ján Paralič[1], Viera Maslej Kresnakova[1], Andrea Hrckova[3], Michal Kompan[2], Marian Simko[2], Radoslav Blaho[4], Daniela Chuda[2], Maria Bielikova[2], and Pavol Navrat[2]

[1] Faculty of Electrical Engineering and Informatics, Technical University of Košice, Letná 9, 042 00 Košice, Slovakia
{kristina.machova,martin.sarnovsky,jan.paralic, viera.maslej.kresnakova}@tuke.sk

[2] Kempelen Institute of Intelligent Technologies, Mlynské Nivy II.18890/5, 811 09 Bratislava, Slovakia
{ivan.srba,michal.kompan,marian.simko,daniela.chuda, maria.bielikova,pavol.navrat}@kinit.sk

[3] Faculty of Informatics and Information Technologies, Slovak University of Technology, Ilkovičova 2, 842 16 Bratislava, Slovakia
andrea.hrckova@stuba.sk

[4] Faculty of Arts, Department of Psychology, Comenius University, Gondova Ulica 2, 811 02 Bratislava, Slovakia
radoslav.blaho@uniba.sk

Abstract. While digital space is a place where users communicate increasingly, the recent threat of COVID-19 infection even more emphasised the necessity of effective and well-organised online environment. Therefore, it is nowadays, more whenever in the past, important to deal with various unhealthy phenomena, that prohibit effective communication and knowledge sharing in the digital space. Undesired user behaviour and user-generated content in the online environment (mostly on social media) can have various forms, probably, the most harmful is the creation and spreading of *false information* (e.g., fake news) and using *abusive language* (e.g. hate speech). While a notable amount of research effort has been already dedicated to reducing and mitigating the negative consequences of such phenomena, a number of additional challenges and open problems remain unsolved. In this book chapter, at first, we provide a summary of existing research works, challenges and open problems. Consequently, we introduce our research results addressing false information and abusive language. Our approaches are based on intelligent and knowledge-based methods, mainly machine learning, natural language processing, and semi-automatic approaches. Besides the detection of undesired *content*, we present also less studied approaches to identify *users*, who contribute or spread such content. Our approaches are also complemented by the methods that are specific for computational social science and humanities as hyperlink network analysis, conceptual analysis, and interviews. Finally, we

J. Paralič et al. (Eds.): DISA 2020, AISC 1281, pp. 3–32, 2021.
https://doi.org/10.1007/978-3-030-63872-6_1

present some of our applications for monitoring and mitigating such undesired content and behaviour.

Keywords: False information · Abusive language · Digital space · Text processing · Machine learning

1 Introduction

Internet is a place where users communicate and express their thoughts. While most users tend to accept social norms in digital communication, others engage in undesired behaviour (misbehaviour or even anti-social behaviour), which negatively affects the rest of the community and its goals. Such malign behaviour in the online environment (mostly on social media) can have various forms, probably, the most harmful is the creation and spreading of *false information* and using *abusive language*. False information, which can be either spread accidentally (i.e., *misinformation*) or deliberately to deceive (i.e., *disinformation*), can be further categorised into more specific types, such as *fake news*, *fake reviews*, *rumours*, or *hoaxes*. Similarly, we distinguish different types of abusive language, such as *hate speech*, *trolling*, *cyberbullying*, *profanity* and in some cases, also *sexting*.

The negative consequences of such undesired online content on society are not negligible, particularly when a new coronavirus causing COVID-19 is threatening humans that shift their communication to digital space. Since some forms of regulation are in demand, computer-science research in connection with information science and psychology focuses on the development of *characterisation*, *detection* and *mitigation* methods and solutions. Among these, the detection task is being studied the most.

In this book chapter, we will at first define the terms and categorisation of concepts related to false information and abusive language. We will summarise solutions, challenges, and open problems, that characterise the state-of-the-art research on the characterisation, detection, and mitigation tasks. Consequently, we will present our contribution to this area by describing various approaches addressing these challenges and open problems.

Our main contributions are as follows:

- We provide an overview of previous research efforts and identify challenges and open problems that point out to future directions in the research focused on characterisation, detection and mitigation of false information and abusive language.
- We describe our research results while addressing such challenges and open problems.
- Finally, we provide a multidisciplinary perspective – while taking computer science as the primary point of view, we supplement it with our results from information science and psychology.

2 False Information

2.1 Definitions and Categorisation

In this chapter, we are mostly dealing with two of the most widespread forms of false information: fake news and fake reviews.

The concept of *fake news* is a neologism, often used to refer to a fictive message or disinformation [27] that are intentionally and veritably false and could mislead readers [3]. In practice, fake news is, however, a biased and widely-used term, often describing any false stories spreading on social media. Many authors therefore limit the definition of fake news to false news articles in the media, e.g. [27, 66]. In our conceptual analysis [39], we tended to adopt the latter definition, as computer science research often uses another term for false or mixed information on social media – rumours. Zubiaga [90] utilise the term rumours to denote items of information that are unverified at the time of posting. This definition is too general and does not help to determine, which posts on social media can be considered fake news and which rumours. Traditionally, rumours are characterised by the propagation of unauthorised messages that are of universal interest and are disseminated diffusely in social networks [9].

Due to the ambiguous utilisation of the term fake news and their overlap with some definitions of rumours, we differentiated the terms according to the production and distribution [39]. Fake news was distinguished as disinformation that was deliberately produced on the media by its author and rumours as misinformation that were distributed by users via social media. This is very important to discern as producers and distributors of false information have different intents. Producers create fake news either for profit, power or irresponsible reporting. Distributors (sometimes called also "useful idiots" [75]) are usually unaware of sharing false information. Their intents vary from the lack of knowledge or ignorance to the quest to socialise.

The fake news can be created and distributed on traditional media, partisan media or social networking sites. According to Fletcher [24], partisan media have a significantly lower number of readers and time spent on the web than traditional media. The impact of fake news, however, increased significantly with social networking sites. Any social network account can create professionally looking posts, which are spread quickly and for free [80]. Fake news distributed by Facebook have five times higher interaction than news from traditional media [50]. Spreading is affected by influencers much more than the quality or truth of the information.

Many users are ravished by emotions when sharing, as fake news often involve sensational headlines touching the basic emotions like fear, anger or surprise. If a misleading information comes from multiple sources in a similar period, it is not difficult to believe that it is serious information. The lack of information literacy also contributes to the quick spread of false information. Most social networks users do not dedicate enough time and space to confront the source of information and give them only a quick look. Another reason why users do not verify the truth of information is that the information is already enjoyed by thousands of users, and their sharing is perceived as some kind of recommendation or filter.

The producers of *fake reviews* have similar goals and methods as the producers of fake news, although the definitions and scope of news and reviews are different. Fake reviews can be defined as opinions on products or services that have been made to look valuable or genuine, often to deceive or mislead users. Lappas [48] perceives writing fake reviews as a form of attack, performed to purposefully harm or boost the reputation of an item (company, product, or person). The intents of fake news and reviews producers are similar – to make a profit, whereas in the case of fake news the intent is mixed with a need for power and/or ignorance [39]. The goals of both are to influence the opinions of recipients, and this could be a reason why these unethical activities are also called *opinion spamming*. Nevertheless, the term opinion spamming is mostly used for fake reviewers.

The repetition of lies or one-sided truths is called *propaganda* in politics, and the same applies for propagation in marketing. Both fake news and fake reviews are far beyond the ethical border. In both fields, tools like forged documentaries, planted evidence, staged media spectacles, and PR articles/content farms can be spotted [38]. Besides that, Sorgatz [75] designates other methods for describing discussion manipulation in the communities used in unethical marketing as well as for shifting public opinion in politics. Flooding the discussions with numerous comments of fake accounts to spread mass confusion and create the illusion of widespread support of a product or political party is called *astroturfing* and can be spotted in some campaigns. These fake accounts - *sock puppets* promote their ideology, create fake reviews and involve in the discussions. We classify both astroturfing and sockpuppeting in the anti-social behaviour, explained further in the Sect. 3.1.

2.2 State-of-the-Art Solutions, Challenges, and Open Problems

Fake News. The necessity for detection of false information in digital space significantly increased in the last years. We can witness growth in many research papers focused on (semi-)automatic detection of fake news, that employ a wide scale of different approaches – mostly based on machine learning. At the highest level, these approaches can be categorised according to the input data, which are considered during the extraction.

Probably the largest body of research derives for the purpose of the detection a number of *content-based features*, mostly describing text, while multimedia is used in the lesser extent (such as a number of words, a number of references, sentiment of the text, or various text representations). These approaches rely on an assumption, that style of fake news is different from true news. However, this kind of approaches has some drawbacks – features of style do not reflect the real veracity of the presented content, the predictions are difficult to explain, and finally, producers of fake news can quite easily mimic characteristics of true news and thus deceive trained detection models. At the same time, this group of approaches are widely applicable on different types of fake news (even coming from various sources, that were not covered by the training data) and do not require any domain-specific knowledgebases.

Another group of approaches analyses *contextual features* (such as information about the article's source, author or attached discussion). In comparison with content-based features, such contextual data cannot be influenced by the false information producer so easily. Thus they may provide useful supplementary information (contextual features are usually used in combination with content-based ones).

Some approaches focus specifically on *spreading of fake news* (such as a speed of spreading on social networks, spreading paths). Such spreading information has been confirmed to be a valuable feature that distinguishes between true and fake news. At the same time, it is usually difficult to obtain such data as they are available mostly only on social networking sites and not publicly available.

Finally, some authors aim to detect fake news more precisely (not only according to style or spreading characteristics), and thus they apply some kind of *existing knowledgebases* and perform some sort of fact-checking of articles' content against prior verified knowledge. Such approaches may provide a detection with a higher precision and also a better explainability, however, they are fundamentally dependent on the extent of available knowledgebases and, moreover, mapping articles' content to the existing knowledge is a challenging task.

From the second perspective, existing approaches can be distinguished by the type of machine learning algorithms employed for the purpose of detection. At first, there are some approaches based on *traditional machine learning models*. For example, the work [88] is devoted to research the realistic mechanism of the fake news propagation. They tracked large databases of fake news and truth news from two different cultures – Japan and China. Spreading of fake news was monitored at early stages of propagation, e.g. five hours after the first posting. They declared that the information propagation at early stages offers novel features for the early detection of fake news. The novel features were set on the base of identifying topological properties of social networks. Another interesting result is that there are real differences in the propagation of fake news and true news. Another traditional approach to fake news detection based on machine learning methods is a work presented in [29]. It uses a vector representation of text data, TF-IDF weighting scheme and PCFG (Probabilistic Context-Free Grammars). The following machine learning methods were tested for use in solving the problem of recognition of fake news: Support Vector Machines, Stochastic Gradient, Random Forests and Naive Bayes. The best model was learned by Stochastic Gradient (the highest Accuracy = 0.687), while the baseline was a Naive Bayes model (Accuracy = 0.679).

Besides traditional machine learning models, *deep learning methods* have also been used to detect fake news. Several studies are focused on training a fake news classifier which would be able to detect the fake news solely based on the content of the article using NLP [6, 30]. Textual features represent the information extracted from the actual content of the article, which mostly includes the actual text of the article and its title.

While some of the deep learning models rely solely on the text when classifying the fake news, several approaches combine the content with the other contextual features. Linguistic features can be enhanced with multimedia content contained in the article [82]. Zhang in [87] presents the deep diffusive network based on Recurrent Neural Networks (RNN) and Gated recurrent unit (GRU) complemented by regularisation techniques. Authors use information related to the subject and also the author of a news article. More papers suggest that incorporation of integrated linguistic, syntactic, and semantic information about the articles could be beneficial, e.g., in [68] authors also used the data about the source of the article and response the article gets. More recently, a combination of *textual entailment* was used to improve the deep learning models. Authors in [70] presented such a system based on a combination of statistical machine learning and deep learning, which gained superior performance when compared to original approaches. An interesting approach to tackle the problem of obtaining a correctly annotated data was presented in [63]. Authors used crowdsourcing to collect the human-labelled data from seven different domains, and then compared the performance of various deep learning models when using original and human-labelled data. When using the social networks to spread the fake news, very important is also to study, how they are being propagated through the network. Long-Short Term Memory (LSTM) model was presented in [85] to represent the path, how a message spread.

Fake Reviews. The second problem, fake reviews identification is usually based on the machine learning methods used for training of prediction models. For example, work [47] presents algorithm REV2 based on Bayes classifier. The approach uses the following independent metrics: the credibility of an author, reliability of evaluation and quality of a product. The best testing result was promising (Precision = 0.846). The REV2 was employed in the detection system Flipkarte [47].

Another interesting question connected with the fake reviews' recognition is following. Is it necessary to process whole text of the reviews, or is it sufficient to process just the headlines of these reviews? The results obtained in [54] confirm that the models learned from the whole body of texts of reviews are more precise than those learned only from headlines. Another result was that Random Forest models were the best when learned from bodies of reviews (Accuracy = 0.983) and Naive Bayes models were best on the headlines data (Accuracy = 0.812).

The work [35] is also focused on the detection of fake reviews using a text analysis approach based on n-gram models and machine learning techniques. It compares six different classification techniques, namely, K-Nearest Neighbours (KNN), Logistic Regression, linear Support Vector Machine (SVM), Decision Trees and Stochastic Gradient Descent. To reduce the size of the lexical profile of texts, two methods were used: TF and TF-IDF weightings. The highest accuracy was achieved using the SVM algorithm and the lowest using the KNN algorithm. The paper [19] solved the problem of fake reviews uncovering using Naive Bayes classifier and Random Forest. The experiments showed that the Random Forest model achieved better results than the Naive Bayes classifier.

Authors of the work [16] studied an effectivity of a performance of the classification methods for fake reviews filtering when they are used in real-world scenarios that require online learning. Their datasets were from various domains as a trip advising, a recommendation of hotels, or an evaluation of restaurants. They used Bernoulli and Multinomial Naïve Bayes, KNN, Decision Trees, Random Forests and SVM. The best results were achieved by the SVM model (F1 measure = 0.899).

Challenges and Open Problems. The detection of fake news is accompanied with many challenges and open issues. At first, while some studies already analysed the characteristics of fake news (and how they differ from true news), there is a plenty room for additional especially large-scale studies that will focus on describing and understanding their characteristics and thus providing valuable input for feature engineering process performed as a part of detection methods. Despite the first approaches based on deep learning, a huge potential of such approaches (already demonstrated on many different NLP tasks) is still not completely revealed. Also, most of the existing approaches focus on the detection of individual pieces of content, while the research on detection of users who are possible producers or susceptible to share fake news is lacking. Last but not least, the advance in fake news detection is hampered by the absence of large well-annotated benchmark datasets – the existing datasets are labelled manually by human experts and very small (and thus unsuitable for learning some advanced models, such as deep learning ones) or labelled by some kind of simplified heuristic (e.g., the veracity of articles is determined by the reliability of source), what provides us with a large dataset but the labels may contain a lot of noise. This challenge could be addressed by appropriate interdisciplinary cooperation with the professionals from social sciences and humanities. For example, we identified professionals from the library and information science as suitable experts for labelling fake news articles within the preparation of our medical misinformation dataset.

Similarly, the detection of fake reviews poses several challenges and open problems. In general, fake reviews detection is dominantly solved by traditional machine learning approaches, e.g. SVM. There is a lack of annotated datasets with reliable identification of fake reviews. Similarly, as in the case of fake news, this is partially addressed by incorporating heuristic information, e.g. considering marginal opinions to be fake. However, even such datasets are often insufficient in terms of size. Deep learning approaches necessitate much large text corpora to fully leverage the potential of neural network. As a result of lacking large benchmark datasets, only a few approaches to fake news detection employ deep learning. The potential of state-of-the-art techniques (e.g. contextualised language models, transfer learning) in similar tasks are yet to be explored.

In the following subsections, we will present how we addressed some of these challenges and open problems as a part of our previous or ongoing work.

2.3 Characterisation of Fake News by Their Hyperlink Network

As mentioned above, there is a considerable room for research aimed at the characterisation of fake news. Specifically, we recognized an underestimation of the role of hyperlinks in fake news detection. Paradoxically, the quality of the links (resources) traditionally serves: 1) as indicators of the reliability of the research and also 2) as the input to website spam detection employed by some search engines. One of the mechanisms of the spam detection - Trust Rank is based on the premise that good websites seldom link to bad websites, and bad websites often link to good websites in an effort to increase the score in the hub. Trust Rank operates on the principle of selection of a small set of seed pages to be evaluated by an expert. When the reliable seed pages are manually identified, a crawl extending outward from the seed set seeks out similarly reliable and trustworthy pages [34]. The logic works oppositely as well, the closer a site is to spam resources, the more likely it is to be spam. Krishnan and Rashmi [46] call this opposite indicator Anti-Trust Rank.

The aim of our research [39] was to identify whether this linking principle is valid also for the fake news. A method of hyperlink network analysis was utilised to analyse the inbound and outbound hyperlinks of partisan news media (also labelled as fake news media by our local database). The dataset involved the hyperlinks of eighteen most popular mainstream news media and fifteen most popular partisan news media in Slovakia and Czech Republic. More than 171 million of domains of inbound and outbound hyperlinks of selected online news media were collected with Ahrefs tool[1], analysed and visualised with Gephi[2]. Ahrefs is a crawler, the original purpose of which is to

[1] https://ahrefs.com/.

[2] https://gephi.org/.

serve search marketing professionals. It collects, processes and stores data, consisting of hyperlinks, keywords and user behaviour. AhrefsBot is also mentioned as the second most active crawler on the Internet after Googlebot [86]. Gephi is an open-source software for visualisation and exploration of graphs and networks, applicable also for link analysis. Thanks to Gephi visualizations, the networks of directed hyperlinks from and to the online news media and the importance of some nodes, based on the automatic calculation of the sum of hyperlinks were obtained.

We concluded that there are some differences between the linking patterns of both types of media. The most significant one is that mainstream media rarely link to partisan media with dofollow attribute. There is also a little number of hyperlinks between mainstream media of different ownership, partisan media link to established mainstream media more often. We also identified some similarities between the links of mainstream and partisan news media. The vast majority of hyperlinks in both types of media is formed by directories or recommended articles from the same domain. Nevertheless, we confirmed the principle of Anti-Trust Rank and the role of hyperlinks for fake news detection. One of the possibilities for fake news detection could be manually and professionally selected blacklists of fake news media that would serve as seeds for other fake news media identification.

2.4 Detection of Fake News with Deep Learning

Deep learning has become one of the most popular approaches in machine learning in many fields during recent years. Especially in natural language processing (NLP) or computer vision, deep learning models have shown excellent results and proved to be very useful to tackle the problems from those domains.

In the past, many of the NLP problems were solved using traditional machine learning models. Many of these approaches faced issues related to very high-dimensionality of linguistic information. More recently, with the utilisation of more advanced representations such as word embeddings, neural networks gained popularity in NLP tasks thanks to the excellent results when compared to traditional machine learning models. Deep learning models have brought significant improvements over the n-grams model language. Such models can extract features on their hidden layers and learn the representations with high accuracy. These models can capture both semantic and syntactic patterns.

Many researchers have been using the popular deep learning methods such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) or Long-Short Term Memory (LSTM). Especially RNN and LSTM models are popular when applying on the sequential data, as the model supports to memorise the results of previous computations and use that information in the current computation. This makes recurrent networks very well equipped to capture the context dependencies and makes them particularly useful in natural language processing because it helps to gain a deeper understanding of a language. Deep neural networks are also known for a fact that the larger the dataset is available for training, the better is the quality of extracted features. On the other hand, one of the significant drawbacks in certain types of such deep learning models could be the cost of computational resources needed to train these models. Training time increases dramatically, mostly due to the size of the datasets used to train such models.

Used Methods. *Convolutional neural networks* (CNN) are one of the deep neural networks which gained popularity in 2012, thanks to ImageNet [20]. However, they originated long before, in 1989, when they were first introduced to the world by computer scientist Yann LeCun in his work *Generalisation and network design strategies* [49]. CNNs can recognise and analyse not only images or video but have achieved success in NLP tasks or time sequence processing. During training, the individual layers are created as different degrees of abstraction of input data. Filter weights are initially initialised randomly. Gradually, the lower layers reveal low-level features, and the higher ones create more complex concepts. CNN topology is defined by three types of layers: a *convolutional layer*, a *pooling layer*, and a *fully connected layer*. A set of filters (kernel) represents the convolutional layer. The neuron's activation on the convolutional layer is calculated by moving the filter sequentially in the horizontal and vertical directions. The mathematical formula of the function:

$$S(i,j) = (K * I)(i,j) = \sum_m \sum_n I(m,n)K(i-m, j-n), \tag{1}$$

where K is the kernel, i.e., filter, and I is a digital image [31]. In the NLP tasks, usually one-dimensional convolutional neural networks (1D CNN) are used. An example of 1D CNN is depicted in Fig. 1.

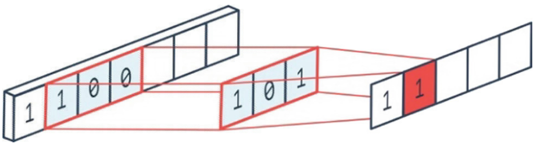

Fig. 1. One-dimensional convolutional process. Input data on the left, the filter in the middle, and the convolution output on the right.

Other types of deep neural networks are *recurrent neural networks* (RNN). RNNs are an excellent tool for NLP. Unlike standard neural networks, RNNs uses the word as an input, not the whole sample. RNN provides flexibility when working with the text of different lengths. RNN treats each word of the sentence as a separate input at time t, in addition to which it takes into account the activation value at time $t - 1$. A special type of RNN is *Long Short-Term Memory* (LSTM). LSTM has a complex structure of individual neurons and represents a way to deal with the vanishing gradient problem (see Fig. 2. The first figure shows the standard recurrent neural network and the vanishing gradient problem, which results in the loss of context. In contrast, the second figure shows the preservation of information and context in LSTM [32]). Using LSTM in any sequential task will ensure that long-term information and context is maintained.

The LSTM [37] network consists of memory blocks that are connected to layers instead of neurons. Each block contains gateways that manage the state and output of the block to regulate the information flow. Gateways can learn which data in a sequence is relevant and needs to be preserved. Thus, only relevant information will be included in the predictions.