FEATURES TO DISTINGUISH BETWEEN TRUSTWORTHY AND FAKE NEWS

Daniel CHOVANEC, Ján PARALIČ
Department of Cybernetics and Artificial Intelligence, Faculty of Electrical Engineering and Informatics,
Technical University of Košice, Letná 9, 042 00 Košice, Slovak Republic,
Tel. +421 55 602 4128, E-mail: jan.paralic@tuke.sk

ABSTRACT

This paper discusses the necessity of fake news detection and how selected features can show differences between trustworthy and fake news. To demonstrate this concept, we first identified a set of features, that we believe can distinguish between fake and trustworthy news. We used these features to analyse two real datasets and evaluated our results in various ways. We first used visual analysis by means of boxplots and evidenced the significance of differences by means of the Wilcox singed-rank test. As next, we used three different classification algorithms to train models for distinguishing between trustworthy and fake news using all important features. Finally, we used Principal Component Analysis (PCA) to visualize relations between identified features.

Keywords: *analysis, fake news, fake news detection, text analysis*

1. INTRODUCTION

Information flow in the online world is persistently rising, providing vast amounts of new information every day [1]. For some news providers, this has become a business, in which factuality is overlooked in favour of profits. This has a negative effect on the credibility of some news, as the information can often be modified to such an extent, that it becomes fake and misleading. The need for fake news detection has been therefore rising, which has resulted in various projects [2], that had the detection of fake news as their main goal (e.g., Rebelion [3], FakeNewsTracker [4]).

We believe, that selecting features which could indicate the differences between fake and trustworthy news is essential in text analysis [3], as features that prove to be relevant can create a basis for fake news detection in the future. Therefore, in this paper we briefly outline our selected features for the experiment, the chosen dataset, evaluation of text analysis and the results of our evaluation.

2. SUBJECT

In this section we present identified features and text analysis on selected datasets.

2.1. Relevant features

The first and most important step of the text analysis was to gather features, that would be tested on a dataset. Collected features were split into three categories:

1. *Readability indexes* – Gunning Fog Index, The Flesch Readability Index, Automated Readability Index
2. *Linguistic features* – first-person pronouns, third-person pronouns, uppercase titles of articles, truth indicators, hard words
3. *Source* of the text

Readability indexes analyse each given text word by word, resulting in a number, which represents the level of reading difficulty of a given article. These indexes do not differentiate between specific words, which meant that there was a need for features, that would also analyse specific words (linguistic features).

Uppercase titles of articles are prone to indicate fake news articles, so it is relevant to analyse not just the whole article but also the title. All other linguistic features can also show us differences between trustworthy and fake news.

The source of articles can also be a suitable feature for fake news detection, as fake news sharing sites are more likely to be continuing this behaviour in the future [5]. It is in fact analogical situation as with sources of the toxic comments [6].

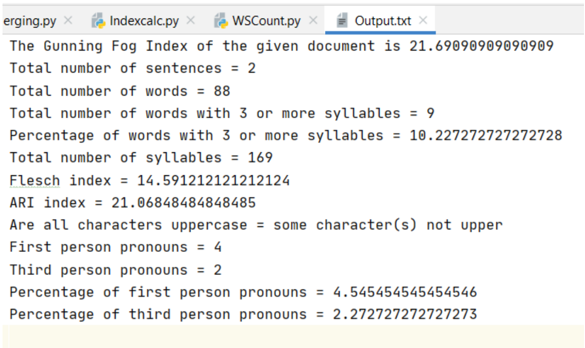
2.2. Datasets

For this experiment, we decided to use two similar datasets, which both are part of a bigger dataset called Benjamin D. Horne, found on Github [7]. The first dataset consists of 42 fake news articles and 39 real (trustworthy) news articles, along with their titles. The second dataset contains 61 fake news articles, 61 real news articles, but also has 69 articles, which are labelled as satire news. The satire news category allows to expand the experiment to also satire texts and compare them to real and fake news. The second dataset includes titles of all articles in each of the three categories as well. All articles and their titles were in .txt form file.

2.3. Text analysis

The implementation of all features and readability indexes was done in the *Python* programming language. Each article from each dataset was analysed separately, to give precise results that could be further analysed. The results of linguistic features analysis (except uppercase titles of articles) were presented as a percentage of these specific words in relation to the total number of words in an article.

The results of analysis of an article were being exported to an Output.txt file, which was supplemented with new information each time a new article was analysed. An example of the set of result metrics saved in the Output.txt file can be seen in Fig. 1.


erging.py Indexcalc.py WSCount.py Output.txt

```
The Gunning Fog Index of the given document is 21.69090909090909
Total number of sentences = 2
Total number of words = 88
Total number of words with 3 or more syllables = 9
Percentage of words with 3 or more syllables = 10.227272727272728
Total number of syllables = 169
Flesch index = 14.591212121212124
ARI index = 21.06848484848485
Are all characters uppercase = some character(s) not upper
First person pronouns = 4
Third person pronouns = 2
Percentage of first person pronouns = 4.545454545454546
Percentage of third person pronouns = 2.272727272727273
```

Fig. 1 Output.txt file with metrics from text analysis of an article

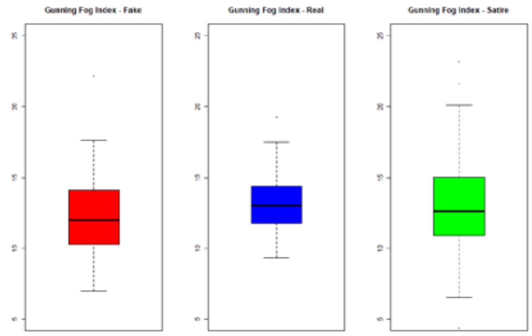


Fig. 2 Boxplot comparison for the Gunning Fog Index feature, second dataset

3. EXPERIMENTS AND RESULTS

Comparison of results, statistical tests and classification tasks were all done in the *R* programming language (except decision tree classification done in *Python*), for easy graphical representation and for the availability of libraries, which can easily be used in *Rstudio*. For all these tasks, tests, and comparisons we chose six measured features.

- Gunning Fog Index,
- Automated Readability Index,
- The Flesch Readability Index,
- Percentage of first-person pronouns,
- Percentage of third-person pronouns,
- Percentage of hard words

First three are various readability indexes (i.e. from the first category defined in previous section 2.1) and the other three are linguistic features (second category defined above). We did not have information about the sources of the articles (i.e. no feature from the third category has been included).

3.1. Feature differences between target classes

Comparing the results of fake, real and satire news (in case of the second dataset) analysis was firstly done using boxplot graphs, which in most cases showed us visible differences between analyzed features. Fig. 2 shows as an example the Gunning Fog Index feature values comparison for three target categories in the second dataset.

To prove whether the visual differences in particular feature values are also statistically significant between target classes, we used the Wilcox singed-rank test to test the difference in feature medians between target classes. The p-value was used to determine, whether there is evidence to reject the H_0 hypothesis, which states that both measures are similar and that they are not statistically significant [8]. The hypothesis has been rejected if $p < 0.05$. We tested the Gunning Fog Index and the Automated Readability Index results from the first dataset, and the Flesch Readability Index, as these indexes were visually the most different amongst other features. The Wilcox signed-rank test showed us, that in all three cases the measurements and their medians were different, therefore these differences are statistically significant.

3.2. Classification models

We selected three different types of classification models to see, how good are we able to predict the target class by means of identified features. We included k-NN, naïve Bayes and decision tree classification. We experimented with different subsets of features as input for the training phase to achieve the highest possible accuracy in the resulted model. The first dataset proved to show the best results with only the readability indexes as input, occasionally along with the percentage of hard words. The average achieved accuracies were around 50%, however climbing up to 67.6% for k-NN classification.

The second dataset achieved the best results in classification tasks when all six features were used, due to satire news showing differences in first and third-person pronouns in comparison with fake and real news. The achieved accuracies were lower, ranging from 43% (naïve Bayes) to 57% (k-NN).

The number of uppercase titles of articles proved to be higher in fake news articles compared to real news articles (6:1 in the first dataset, 7:0 in the second dataset). Satire news however, had 29 out of 69 article titles written all in uppercase letters, surpassing fake news.

3.3. PCA analysis

The principal component analysis (PCA) method helped us understand our data better and showed the differences between our features and our datasets. Fig. 3 shows the PCA for the second dataset.

4. USER-INTERFACE FOR PRESENTATION OF RESULTS

Even though classification tasks, statistical tests, PCA and comparisons were existent and had a graphical output, they were all results of runnable chunks of code. To present these graphs and confusion matrixes (k-NN, naïve Bayes) in a more user-friendly way, a user-interface was created using *RShiny*. The application has a simple interface, which allows the user to choose which data from which dataset s/he wants to display. This application can be run from *Rstudio*.

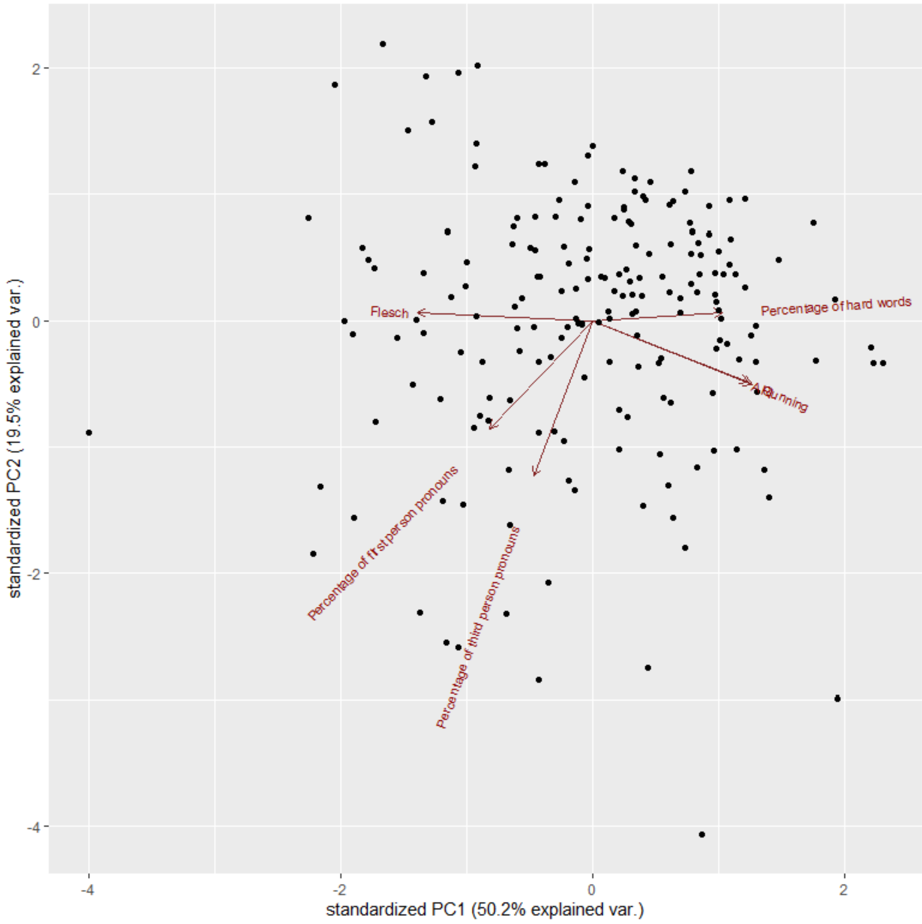


Fig. 3 Principal component analysis for the second dataset

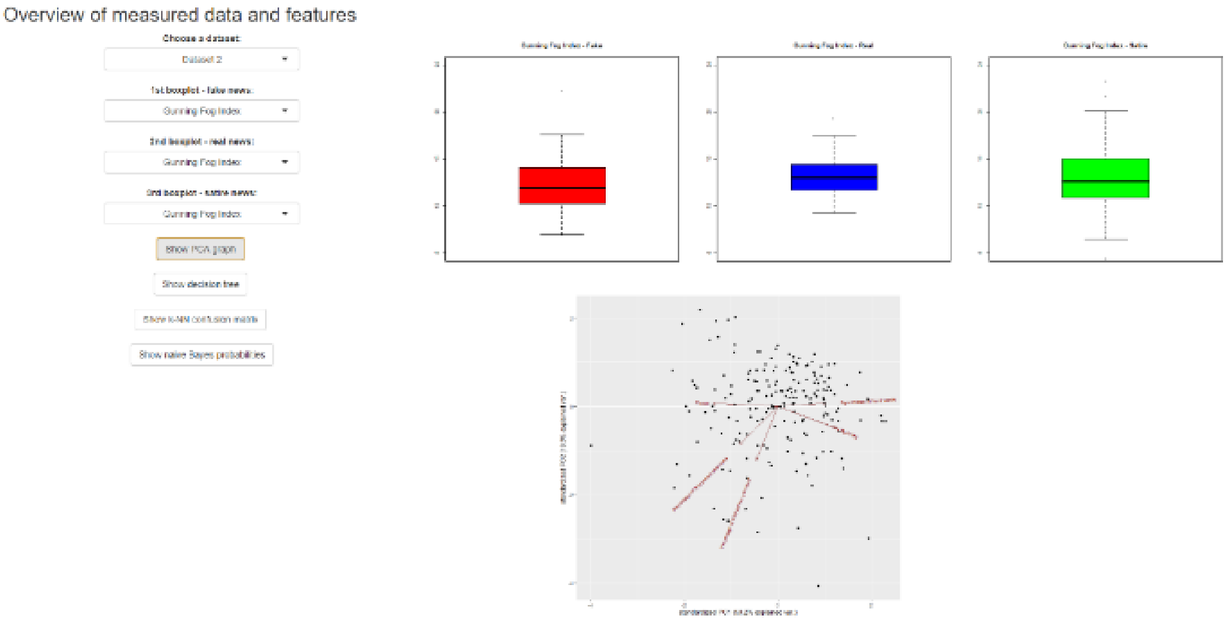


Fig. 4 User interface of our *RShiny* application

Fig. 4 shows the user interface of our application with displayed PCA and boxplot graphs for the second dataset.

5. CONCLUSIONS

Fake news detection is a necessary process, which can protect users from hoaxes and fake information on the internet. We showed that identified features proved to be relevant in distinguishing between fake and real news, as they showed numerical, statistical and graphical differences, which in the end, were presented to the used in an interactive *RShiny* application.

ACKNOWLEDGMENTS

This work was supported by the Slovak Grant Agency of the Ministry of Education and Academy of Science of the Slovak Republic under grant no. 1/0685/21.

REFERENCES

[1] ARIELY, D.: Controlling the Information Flow: Effects on Consumers' Decision Making and Preferences. *Journal of Consumer Research*, Vol. 27, No. 2, 2000.

[2] SIMKO, J. – RACSKO, P. – TOMLEIN, M. – HANAKOVA, M. – MORO, R. – BIELIKOVA, M.: A study of fake news reading and annotating in social media context. *New Review of Hypermedia and Multimedia*, 2021.

[3] MACHOVA K. et al.: Addressing False Information and Abusive Language in Digital Space Using Intelligent Approaches. In: Paralič J., Sinčák P., Hartono P., Mařík V. (eds) *Towards Digital Intelligence Society*. DISA 2020. Advances in Intelligent Systems and Computing, Vol. 1281. Springer, Cham, 2021.

[4] SHU, K. – MAHUDESWARAN, D. – LIU, H.: FakeNewsTracker: a tool for fake news collection, detection and visualization. Data. *Computational and Mathematical Organization Theory* 25 (11), 2019.

[5] SITAULA, N. (et al.): Credibility Signals in Coauthorships. In: K. Shu et al. (eds), Disinformation, Misinformation, and Fake News in Social Media, *Lecture Notes in Social Networks*, Springer 2019.

[6] MASLEJ-KRESNAKOVA, V. – SARNOVSKY, M. – BUTKA, P. – MACHOVA, K.: Comparison of Deep Learning Models and Various Text Pre-Processing Techniques for the Toxic Comments Classification. *Applied Sciences*, Vol. 10, No. 23, Dec. 2020.

[7] BENJAMIN D. HORNE: fakenewsdata1. Available at: https://github.com/BenjaminDHorne/fakenewsdata1

[8] GRINČOVÁ, A. – PETRILLOVÁ, J.: *Applied statistics* (in Slovak). Technical University of Košice, 2019.

Received July 21, 2021, accepted September 17, 2021

BIOGRAPHIES

Daniel Chovanec graduated in 2021 (MSc) at the Department of Cybernetics and Artificial Intelligence of the Faculty of Electrical Engineering and Informatics at Technical University in Košice. This article summarizes the main results of his master thesis.

Ján Paralič graduated in 1992 (MSc) at the Department of Cybernetics and Artificial Intelligence of the Faculty of Electrical Engineering and Informatics at Technical University in Košice. He defended his PhD in the field of Artificial Intelligence in 1998. In 2004 he was promoted to Associate Professor in Artificial Intelligence from the Technical University in Košice, and in 2011 he was appointed Full Professor in Business Information Systems. He serves as Deputy's head at the Dept. of Cybernetics and Artificial Intelligence, TUKE and leads the Data Science research team of researchers and doctoral students in the field of (big) data analytics, including text data analysis using knowledge-based approaches. He was master thesis advisor of Daniel Chovanec.