Multiple-Language Approach To Fake News Classification

Weronika Borek[0000−0003−2426−9541] and Paweł Ksieniewicz[0000−0001−9578−8395]

Department of Systems and Computer Networks,
Faculty of Information and Communication Technology,
Wrocław University of Science and Technology,
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław, Poland
{weronika.borek,pawel.ksieniewicz}@pwr.edu.pl

Abstract. In the era of the open Internet, which opened the door to global, communication, the risk of encountering false information has increased significantly as they are unverifiable. Fake news often becomes practically indistinguishable from ordinary news and their multitude forces the use of automation. This paper addresses this need by presenting an approach based on the use of multiple languages to diversify classifiers ensemble.

Keywords: Fake News · Classifier Ensemble · Natural Language Processing

1 Introduction

New amounts of data and information are produced every day around the world. The Internet has made it possible for every person to be not only a recipient, but also a sender of communicates that reach a wide range of people. This allowed for a significant improvement in the functioning of many branches of society and industry, but it also entailed an escalation of negative phenomena, such as fake news. Their influence on recipients' decisions was proved, among others, during the US elections in 2020, when Twitter and other social media were flooded with a wave of unconfirmed information about the voting procedure[1]. The current events related to the war in Ukraine, which from the beginning is surrounded by fake news and propaganda, also allow taking closer look at the dangers of disinformation.

The scale of the phenomenon prioritize the task of verifying information. It becomes easier to observe that handling this problem without far-reaching automation is impossible [5]. This paper proposes an approach to recognize fake news, which increases the quality of classification not by using complex *Natural Language Processing* algorithms or *deep neural networks*, which are already widely discussed in the field [2], but by data analysis in many languages simultaneously. In this way, the classifier ensemble is provided with an appropriate

[1] https://blog.twitter.com/en_us /topics/company/2020/2020-election-changes.html

diversification. Thanks to it, method can achieve better quality, according to the theory that a suitably diverse pool of even weak experts is capable of performing better than a highly qualified single expert [3]. Additionally, the process of translating texts using automated tool reduces it to a more unified form, often eliminating the use of synonyms and generating repetitions instead. Something that is a disadvantage in the context of linguistic hygiene can become an important advantage in the context of automatic text analysis.

2 Experiments

The experimental evaluation was performed using the *scikit-learn* [6] library. *Gaussian Naive Bayes* was chosen as the base classifier due to the simplicity and speed of its operation. As a metric to evaluate the results, the *balanced accuracy score* was used, selected due to its immunity to potential data imbalance.

The analysis used a fragment of the NELA-GT-2020 [4] dataset, which contains two thousand objects marked with an appropriate label, consisting of the title and content of the article. It has been translated from English into 34 other languages using the Google API [1]. The translation process itself eliminated some texts that turned out to be too complex for the API to handle. Ultimately, the experiments were carried out on 1677 objects, 200 of which were extracted using *Term Frequency - Inverse Document Frequency* (TF-IDF).

To ensure the reliability of the results, the 5×2 CV protocol was used, and for each language the training and test sets were separated using the same masks. This approach provided the possibility of induction on different datasets while maintaining a common bias. In practice, this means that exactly the same objects are compared every time, and translation is

Table 1. Achieved *balanced accuracy score* for every language model and multilingual ensemble

CC	LANGUAGE	TITLE	TEXT	ENSEMBLE
de	German	0.822	0.820	0.871
da	Danish	0.832	0.832	0.874
no	Norwegian	0.780	0.841	0.859
sv	Swedish	0.821	0.827	0.869
en	English	0.840	0.864	0.897
is	Icelandic	0.823	0.821	0.879
yi	Yiddish	0.754	0.808	0.812
bg	Bulgarian	0.740	0.809	0.830
mk	Macedonian	0.824	0.851	0.892
bs	Bosnian	0.817	0.791	0.872
hr	Croatian	0.749	0.808	0.845
sl	Slovenian	0.786	0.752	0.827
sr	Serbian	0.808	0.795	0.869
pl	Polish	0.779	0.810	0.844
cs	Czech	0.715	0.810	0.790
sk	Slovakian	0.786	0.768	0.830
ru	Russian	0.786	0.794	0.842
uk	Ukrainian	0.798	0.820	0.872
be	Belorussian	0.832	0.817	0.872
he	Hebrew	0.811	0.832	0.875
ar	Arabic	0.816	0.864	0.897
et	Estonian	0.813	0.780	0.857
fi	Finnish	0.820	0.794	0.858
hu	Hungarian	0.776	0.791	0.851
es	Estonian	0.826	0.843	0.882
fr	French	0.825	0.841	0.868
pt	Portuguese	0.811	0.836	0.871
it	Italian	0.826	0.848	0.883
ro	Romanian	0.805	0.834	0.873
lv	Latvian	0.820	0.779	0.868
lt	Lithuanian	0.818	0.791	0.863
zu	Zulu	0.814	0.792	0.872
eo	Esperanto	0.813	0.804	0.861
sq	Albanian	0.816	0.799	0.848
tr	Turkish	0.802	0.835	0.876
ens	Ensemble	0.835	0.895	0.866

an additional preprocessing procedure
prior to vectorization of documents.

Table 1 shows the *balanced accuracy score* results for all languages using each of the keys (TITLE and TEXT). Three types of classifier sets are also presented: (a) built for a single language, but consisting of all keys, (b) built for a single key, but using all languages, (c) aggregating all available sets and attributes.

In addition to analyzing the accuracy of the created models, it was decided to compare the predictions made by classifiers based on different languages. For this purpose, the *balanced accuracy score* metric was also used, in which, instead of indicating the correct labels, the predictions of the fitted models in different languages were cross-verified. The results for the key set are shown in Figure 1 in the form of a heatmap, where the lightest shade – white – means that the predictions were identical, and therefore the darker the color, the smaller the similarity. Languages within individual language groups were analyzed in detail, as it could be assumed that they would be the most similar to each other. According to this measure, the similarity does not fall below 80% for any language within the language group.

3 Conclusions And Future Works

The experiments presented in Section 2 showed that the same classification model used for many languages achieves results that differ in some cases by more than 10 percent. This proves the different level of complexity of individual translations, and thus, that some of the translations used in the experiment are characterized by a different degree of problem difficulty.

Thanks to this diversity, it is possible to build diversified classifier ensembles. Even if the pool composed of all tested sets does not achieve a significantly higher result than all its members, it can be concluded that it is possible to obtain higher accuracy. To do this, it would be worth carrying out an analysis based on the integration of classifiers by language groups, thanks to which they could reinforce each other more effectively. Alternatively, and also worth exploring, an integration would be possible using the *Mixture of Experts* approach.

It is also worth looking at the solutions using more complex classification algorithms than naive Bayes in subsequent works, which could result in obtaining higher-quality results and more extensive research.

Acknowledgement

This work was supported by the statutory funds of the Department of Systems and Computer Networks, Faculty of Electronics, Wroclaw University of Science and Technology.

References

1. Bisong, E.: An overview of google cloud platform services. Building Machine Learning and Deep Learning Models on Google Cloud Platform pp. 7–10 (2019)

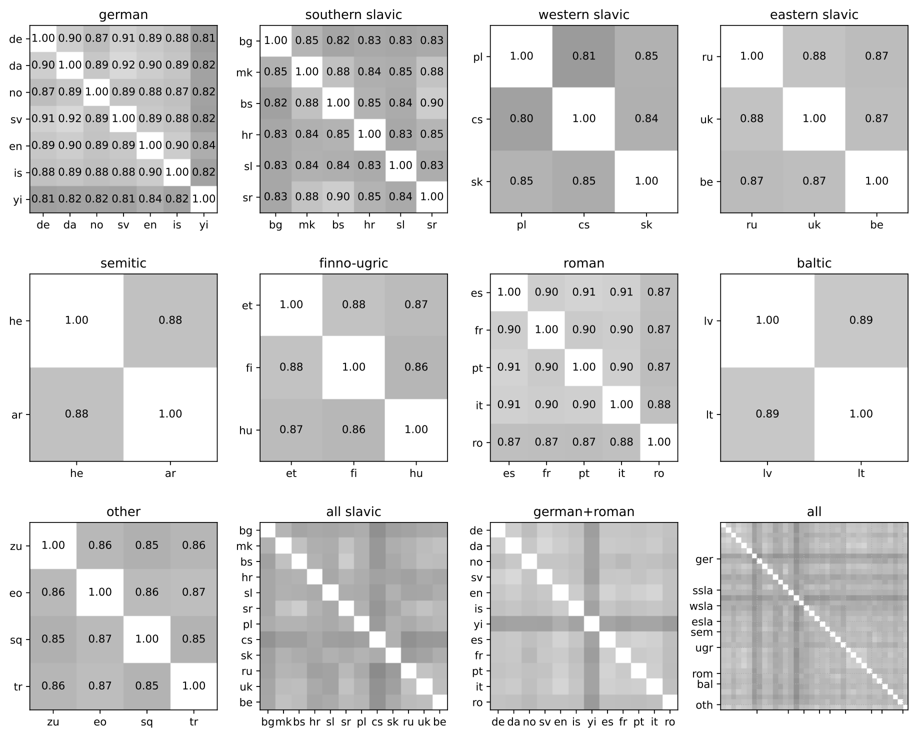

Fig. 1. Comparison of prediction cross compliance between languages from within different language groups according to balanced accuracy score

2. Choraś, M., Demestichas, K., Giełczyk, A., Herrero, Á., Ksieniewicz, P., Remoundou, K., Urda, D., Woźniak, M.: Advanced machine learning techniques for fake news (online disinformation) detection: A systematic mapping study. Applied Soft Computing **101**, 107050 (2021)
3. Galton, F.: Vox populi (the wisdom of crowds). Nature **75**(7), 450–451 (1907)
4. Gruppi, M., Horne, B.D., Adalı, S.: Nela-gt-2020: A large multi-labelled news dataset for the study of misinformation in news articles. arXiv preprint arXiv:2102.04567 (2021)
5. Oshikawa, R., Qian, J., Wang, W.Y.: A survey on natural language processing for fake news detection. arXiv preprint arXiv:1811.00770 (2018)
6. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E.: Scikit-learn: Machine learning in Python. Journal of Machine Learning Research **12**, 2825–2830 (2011)