Methods to Identify Fake News in Social Media Using Machine Learning*

Denis Zhuk, Arsenii Tretiakov, Andrey Gordeichuk
ITMO University
St. Petersburg, Russia
jukdenis@gmail.com, ars.tretyakov@gmail.com, a.gordeychuk@int-syst.com

Abstract—**Fake news (fake-news) existed long before the advent of the Internet and spread rather quickly by all possible means of communication being an effective tool for influencing public opinion. Currently, there are many definitions of fake news, but the professional community cannot fully agree on a single one, what creates a big problem for their detection. Many large IT companies, such as Google and Facebook, are developing their own algorithms to protect the public from informational falsification. At the same time, the lack of a common approach to understanding the essence of fake news makes the solution of this issue ideologically impossible. This problem requires to be seriously studied by specialized experts and scientists from different fields. This research analyzes the mechanisms of publication and distribution of fake-news, gives their classification, structure and algorithm of construction. The researchers decide on the methods of identifying this type of news in social media with the help of systems featuring the elements of artificial intelligence and machine learning.**

I. Introduction

In 2016, there was a great public response based on the assumption that fake-news strongly influenced the outcome of the presidential elections in the United States. Some sources provide information that the fake news about the US elections on Facebook was more popular among users than the articles of the largest traditional news sources. However, the scope of active use of fake news is not limited to politics. For example, the news story about Canadian, Japanese and Chinese scientists studying the effectiveness of the treatment of blood cancer using the root of an ordinary dandelion was transferred from user to user more than 1.4 million times in 2016.

False news is a concern, because they can affect minds of millions of people every day. Such coverage puts them in a single row with traditional methods of influence, such as advertising, and the latest ones – the search engine manipulation effect (Search Engine Manipulation Effect) and the effect of search options (Search Suggestion Effect). These have led to the emergence of the term *post-truth,* which in 2016 became the word of the year by the Oxford Dictionaries [10]. Thus, fake news is defined as a piece of news, which is stylistically written as real news but is completely or partially false [13].

Another problem that prevents users from getting the full news image of the day is so-called "informational separation" caused by filtration of information through news aggregators and social networks. The same thing happens when we use Facebook [5]. For example, if a user does not support Brexit, his news feed is likely to contain posts of those of his friends who have the same attitude towards Brexit. Thus, the user does not have any access to the opposite point of view, even if they try to find it on purpose. To avoid such a situation, Facebook began to mark each news story depending on whether the news is truthful or not. Facebook marks some posts as "disputed" and gives a list of websites that consider this information fake [5]. Mark Zuckerberg estimates the volume of such news at Facebook to be 1% [1]. In 2016 "Google News", a news aggregator, began to mark news about the USA and the United Kingdom. Then the company started checking news about Germany and France, and since February 2017 this feature has become available in Mexico, Brazil and Argentina [2]. Russian government also paid attention to this problem. In February 2017, Russian Ministry of Foreign Affairs started publishing examples of fake news by foreign mass-media companies [11]. Moreover, in August 2017 the President of USA Donald Trump offered his own decision about spreading fakes. He launched his own news program on his Facebook page "Real News" for posting only reliable facts there [6].

At first non-technic ways for analysis and detection fake news are worth mentioning. In 2017 the European Commission launched a public consultation on fake news and online disinformation and set up a High-Level Expert Group representing academics, online platforms, news media and civil society organisations [9]. The Expert Group includes citizens, social media platforms, news organisations, researchers and public authorities. Moreover, the International Federation of Library Associations and Institutions (IFLA) has published the instruction about fake news [4]. It contains eight rules helping to define what information is false. Among other things, authors recommend to pay attention to news headings, the place of their placement, date and formatting. Infographics with the tips can be downloaded in the PDF format in different languages. Besides, in 2017 a group of journalists in the Ukraine started "StopFake News" with the goal of debunking it. Started by professors and journalists from Kiev Mohyla University, "StopFake News" considers itself to be a media institution for providing public service journalism [7].

*The reported study was funded by RFBR according to the research project № 18-311-00125

II. Related work

Here we describe the general approaches for fake news detection of, classification, structure and construction algorithm.

A. A Subsection Sample

In the paper "Credibility Assessment of Textual Claims on the Web" [12] authors offered a general approach for credibility analysis of unstructured textual claims in an open domain setting. They used the language style and source reliability of articles reporting the claim to assess their credibility. There were experiments on analyzing the credibility of real-world claims. Authors in Fig.1 considered a set of textual claims C in the form of sentences or short paragraphs, and a set of web-sources WS containing articles A that report on the claims. If $a_{ij} \in A$ denotes an article of web-source $ws_j \in WS$ about claim ci ∈ C, each claim c_i is associated with a binary random variable y_i depicting its credibility label, where $y_i \in \{T,F\}$ (T stands for True, whereas F stands for Fake). Each article a_{ij} is associated with a random variable y_{ij} that depicts the credibility opinion (True or Fake) of the article a_{ij} (from ws_j) regarding c_i – when considering only this article. With the given labels of a subset of the claims (e.g., y_1 for c_1, and y_3 for c_3), the objective is to predict the credibility label of the remaining claims (e.g., y2 for c2). To learn the parameters in our credibility assessment model, we use Distant Supervision to attach observed true/fake labels of claims to corresponding reporting articles, and learn a Credibility Classifier. In this process, we need to (a) understand the language of the article, and (b) consider the reliability of the underlying web sources reporting the articles. Thereafter, we (c) compute the credibility opinion scores of individual articles, and finally, (d) aggregate these scores from all articles to obtain the overall credibility label of target claims.

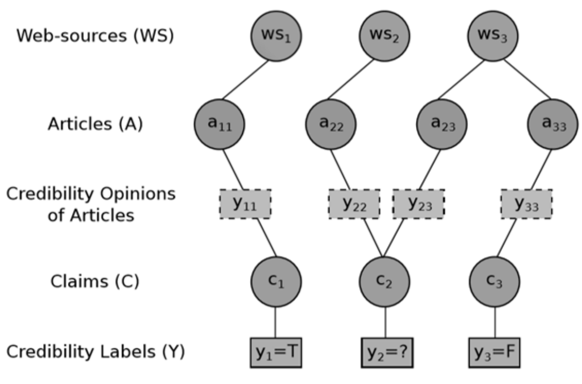

Fig. 1. The model of considering a set of textual claims

B. The reliability of web-sources

Then in Source Reliability, the web-source hosts, the article has also a significant impact on the credibility of the claim [12]. It means one should not believe a claim reported by an article from the one media source, as opposed to a claim on another website. To avoid modeling from sparse observations, authors combine all the web-sources having less than 10 articles in the dataset to a single web-source.

Moreover, in this approach for credibility aggregation from multiple sources Distant Supervision is used for training. Attaching the label y_i of each claim c_i to each article a_{ij} reporting the claim (i.e., setting labels $y_{ij} = y_i$) like in Figure 1 where $y_{11} = y_1 = T, y_{33} = y_3 = F$. Using these y_{ij} as the corresponding training labels for a_{ij}, with corresponding feature vectors $F^L(a_{ij}) \cup F^{SR}(a_{ij})$, we train L_1 – a regularized logistic regression model on the training data.

In addition, there is a misinformation detection model (MDM) that combines graph-based knowledge representation with algorithms for comparing text-derived graphs to each other, fuse documents to construct aggregated multi-source knowledge graph, detect conflicts between documents, and classify knowledge fragments as misinformation [8]. This model in Fig. 2 includes using probabilistic matching exploiting semantic and syntactic information contained in knowledge graphs, and inferring misinformation labels from reliability-credibility scores of corresponding documents and sources. Preliminary validation work shows the feasibility of the MDM in detecting conflicting and false storylines in text sources.

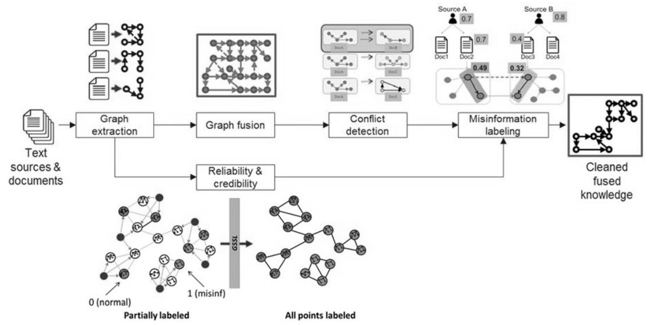

Fig. 2. Components of the Misinformation Detection Model

C. Language-independent approach

This approach, automatically distinguishing credible from fake news, is based on a rich feature set with using linguistic (*n*-gram), credibility-related (capitalization, punctuation, pronoun use, sentiment polarity), and semantic (embeddings and DBPedia data) features. The result was described in the research "In Search of Credible News" [3]. Authors experimented with the following linguistic features where n-grams: presence of individual uni-grams and bi-grams. The rationale is that some n-grams are more typical than credible vs. fake news, and vice versa. tf-idf is the same n-grams, but weighted using tf-idf. Vocabulary richness: the number of unique word types used in the article, possibly normalized by the number of word tokens [3].

Besides, this approach uses embedding vectors to model the semantics of the documents. The authors wanted to model implicitly some general world knowledge. For this purpose, they trained word2vec vectors on the text of the long abstracts and then built vectors for a document as an average of the word2vec vectors of the non-stop word tokens that composed them.

III. Using artificial intelligence technologies (machine learning) to identify fake news

In the framework of this study, we have solved the task of creating a model of a system capable to detect news content with inaccurate information (fake news) with high reliability (more than 90%) and distribute it in the appropriate categories. To deal with this problem, the module of analysis and preprocessing of facts Akil.io was used. In our case it is aimed at solving the tasks of automation of the execution of processes in software and technical complexes through recognition and analysis of tasks presented in the form of a system of facts in text format and their subsequent transformation into a ready solution according to the input data (Fig.3). The technical essence of the system is to automate the execution of processes in software and technical complexes, by recognizing and analyzing tasks presented as a system of facts in text format and their subsequent transformation into a ready solution according to the input data. This module provides the following functions:

- Input and recognition of the input system of facts;
- Analysis of data relationships in the graph;
- Identify the sufficiency / inadequacy of data;
- Formation of a request for additional data in case of their insufficiency;
- Formation of the algorithm for solving the problem;
- Formation of the execution plan for the solution;
- Ensure interactive execution of the plan;
- Representation of a ready solution in the form determined by the task manager.

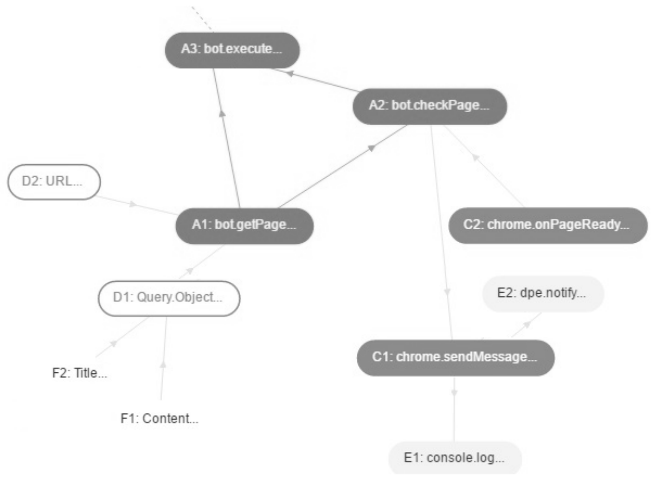

Fig. 3. The module of analysis and preprocessing of facts by Akil.io

A key feature is the consecutive system of the input allowing using unlimited quantity of conditions and rules for the description of the final system unlike classical neural networks. The second feature is that the neural network (a complex of neural networks) used in the decision structurally changes in the course of work and develops in the process of training of the system whereas all decisions existing now in the market are static in respect of their structure.

Since the training was done using the ready-made module for analyzing and preprocessing the facts of the system with the elements of artificial intelligence Akil.io, the most important stages were the collection of data for training and the subsequent verification of the reliability of learning outcomes.

To identify the categories of fake news, you need a large number of examples from different categories of texts that the model was able to recognize. As a result of the preliminary analysis, an average classification of fake news was compiled and used (misinterpretation of facts, pseudoscientific, author's opinions, humor and others). For the distribution of news by a category, two approaches were tested: automatic collection of data from a list of sources with a pre-determined category of all news on this source and manual collection and subsequent sorting by category.

To collect data from a list of sources with a predetermined category of all news on this source, a crawler was used, which allows information to be collected automatically. With the use of this tool, 35,000 articles were collected, which was sufficient for learning the model, but subsequent manual verification of the results of testing this method showed its unreliability (60% reliability). The reason for this is the heterogeneity of the data, combinations of fake and true news within a single resource and a short text length.

As its part the approbation included the manual collection approach and subsequent sorting by categories, a manual step-by-step review of each article, its category definition and subsequent entry into the database for analysis was performed. Based on the results of the training and the subsequent verification of the model, an accuracy of 70% was obtained.

Since the approaches distributing fake news in categories show low reliability, the approach with revealing non-fake news was tested, because there was much more information, generally accepted rules, classifications and other attributes for them. Reliable news appeared to be much easier to reduce to a single category. They tended to base on facts, set out briefly and clearly, and contain a minimum of subjective interpretation. Moreover, reliable resources, where you can publish news materials, they have enough.

The materials were distributed only in two groups: true and false. To the untrue belonged all possible categories of fake news and everything else that did not contain strictly factual information and did not follow the standards of journalistic ethics developed back in the last century with the direct participation of UNESCO. The final sample was 14,300 fake articles and another 25,000 reliable ones. As a result of manual verification of this approach, the accuracy in 92% is fixed. The high accuracy of the approach is due to the ability to provide a large array of reliable information for analysis, which is represented in the stylistics and language typical for a reliable news article.

V. CONCLUSION

Ultimately, the model has learned to analyze how the text is written, and to determine whether it has evaluative vocabulary, an author's judgments, and words with emotional coloring or obscene expressions. If it gives a very low score, it means that the text is not a fact-based news item in its classic form: it can be misinformation, satire, subjective opinion of the author or something else. This method has proven to be quite effective.

Naturally, this method does not solve the problem of fake news itself, but it helps with high confidence to determine non-news by the style of writing that in combination with other available methods such as crowdsourcing, classification of sources and authors, fact checking and numerical analysis. In addition, the method is likely to increase its accuracy close to 100% if to be developed. Further we are planning working on our method and using real dataset by companies for checking our hypothesis and system.

REFERENCES

[1] Fiveash K.: Zuckerberg claims just 1% of Facebook posts carry fake news. Arstechnika (2016). https://arstechnica.com/information-technology/2016/11/zuckerberg-claims-1-percent-facebook-posts-fake-news-trump-election/, last accessed 2018/01/25.

[2] Gingras R.: Richard: Expanding Fact Checking at Google. VP NEWS GOOGLE (2017). https://blog.google/topics/journalism-news/expanding-fact-checking-google/, last accessed 2018/01/25.

[3] Hardalov M., Koychev I., Nakov P.: In Search of Credible News. In: Dichev C., Agre G. (eds) Artificial Intelligence: Methodology, Systems, and Applications. AIMSA 2016. Lecture Notes in Computer Science, vol 9883, pp.172-180, Springer, Cham (2016).

[4] How To Spot Fake News, IFLA (2018), https://www.ifla.org/publications/node/11174, last accessed 2018/02/20

[5] Kafka P.: Facebook has started to flag fake news stories. ReCode (2017), http://www.recode.net/2017/3/4/14816254/facebook-fake-news-disputed-trump-snopespolitifact-seattle-tribune, last accessed 2018/01/26.

[6] Koerner C.: Trump Has Launched A “Real News” Program On His Facebook, Hosted By His Daughter-In-Law. BuzzFeed News (2017), https://www.buzzfeed.com/claudiakoerner/trumps-daughter-in-law-hosting-real-news-videos-for-the?utm_term=.slXDwM080a#.dlEOLWxqx9, last accessed 2018/02/15.

[7] Kramer A.E.: To Battle Fake News, Ukrainian Show Features Nothing but Lies. New York Times (2017), nyti.ms/2mvR8m9, last accessed 2018/02/19

[8] Levchuk G., Shabarekh C.: Using soft-hard fusion for misinformation detection and pattern of life analysis in OSINT. Proc. SPIE 10207, Next-Generation Analyst V (2016), https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10207/1/Using-soft-hard-fusion-for-misinformation-detection-and-pattern-of/10.1117/12.2263546.full?SSO=1#ArticleLink, last accessed 2018/02/20.

[9] Next steps against fake news: Commission sets up High-Level Expert Group and launches public consultation, European Commission (2017), http://europa.eu/rapid/press-release_IP-17-4481_en.htm, last accessed 2018/02/19

[10] Norman M.: Whoever wins the US presidential election, we've entered a post-truth world – there's no going back now. Independent (2016) http://www.independent.co.uk/voices/us-election-2016-donald-trump-hillary-clinton-who-wins-post-truth-world-no-going-back-a7404826.html, last accessed 2018/02/18.

[11] Ministry of Foreign Affairs will publish fake news and their disclosures, RIA Novosti (2016). https://ria.ru/mediawars/20170215/1488012741.html, last accessed 2018/01/25.

[12] Popat K., Mukherjee S., Strötgen J., Weikum G.: Credibility Assessment of Textual Claims on the Web. Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, pp. 2173-2178, Indianapolis, Indiana, USA (2016).

[13] Sukhodolov A.P.: THE PHENOMENON OF “FAKE NEWS” IN THE MODERN MEDIA SPACE, pp. 87-106, Evroaziatskoe sotrudnichestvo: gumanitarnye aspekty (2017).