A proposal for the Machine Learning Application
Research Paper

April 9, 2020

Optimising online documents for fact-checking

Group:

Alpha

Members:

Alexander Peikert	alexpeikert@uni-koblenz.de	218200812
Clemens Steinmann	csteinmann@uni-koblenz.de	218200209
Erwin Letkemann	erwinletkemann@uni-koblenz.de	218200352
Julian Dillmann	juliandillmann@uni-koblenz.de	218100919

WeST
Universität Koblenz-Landau
Germany

1 Introduction

1.1 Problem

With media consumption rapidly growing on the web, it is fairly difficult to distinguish between news sources that are genuine and ones that are not, due to the nature of the internet. Easy access and possibilities to publish more news articles quickly overshadow efforts of proper fact checked journalism and in the worst case scenario even distribute miss information. Even though growing efforts to counteract the spread of fake news in the past years and measures to fact check in order to detect miss information are overwhelmed by the consistent stream of data.
A fully automated system to detect and fact check claims on the web sounds like a good idea, but in practice the task of fact checking is to complex for such a system. The amounts of relevant contexts with the needed sensitivity to judge claims are out of reach, for non human instances, to detect at this moment. Therefore the focus right now should be on improving the manual progress of fact checking by developing and providing tools. Such as checking claims against an authoritative source or checking against previous fact checked claims[8]. These automated progresses decrease the time needed for manual fact checking to verify claims.

1.2 Our solution

In the course of the project we are going to develop a web-based application to help the optimisation of future fact checking by developing an automated tool to match already verified claims, by trusted fact checking Organisations such as PolitiFact or Snopes, to claims, from to be examined articles. Furthermore we are evaluating a check worthiness score for the English news articles. The score will be computed given the accordance of the already verified claims and the help of "Information nutritional labels" [1] extracted from the article content. At the same time the web application can be used to suggest similar non verified news articles, on the same topic ranked, corresponding to score.
The service is aimed at two different user groups with different benefits:

1. Mainly, our web service can be used, by fact checkers, to quickly determine if a claim was already verified before. This saves time in the progress of verifying among individuals or between entities. The ranked verified claims provide help to quickly identify if a claim was checked before. The check worthiness score also gives guidance on which articles the fact checker should focus on.
   A smaller subcategory includes researchers that want to further research the possibilities of nutrition labels in article/claim verification. They could use part of the application or the database to create new collections for research.

2. On the other side, ordinary web users can use the labels and matched verified claims as guidance to self judge a news article before reading. The suggested articles of unverified news articles gives opportunity to discover different articles on the same topic.

2 Related work

On the subject of "claim retrieval" is currently worked on for example in CLEF CheckThat 2020! [9]. Most work in the area of automatic fact checking is done on the topic of identification of "check worthy" statements or articles. ClaimPortal [6] is such an instance where tweets from several US politicians are analysed and assigned with a check worthiness score, in the regard if the factual claim in the post, if present, is of importance for the public.
Research on determining the check worthiness of political claims was also done with the help of nutritional labels and word embeddings [5]. The authors used models in natural language processing to determine the nutritional labels and also used Word2Vec model to determine key words in sentences. They hypothesized a correlation between the different implications of the labels and the veracity of claims, for example high emotions in political statements indicate an attempt to deceive the audience. Hence the statement should be check worthy. All three of these were used in different combinations in several different ML algorithms to evaluate the performance on the Checkthat! 2018 [7] dataset. The authors concluded that the nutritional labels with word-embedding with stochastic gradient descend logless without oversampling archived the best MAP of all the combinations of possible inputs with different algorithms. At the same time their solution outperformed or matched the best participants' methods on MAP, MRR, MR-P and MP@5.
These results showcase that nutritional labelling could be of use in further research in computational detection of misinformation in a bigger range of topics or automated help tools for manual fact checking. Therefore we develop a solution to assist verification of claims by providing a tool to retrieving already verified claims.

3 Our Application

3.1 Needed Components:

- Databases
- Content extractor
- Content natural language processing algorithms
- Web-application

3.2 General procedure

Given an input URL to an English news article, the applications' database will check whether the url has already been checked by our system before.

1. If the news article is present in our secondary database we will display the articles that help to verify the articles content and if requested find the similar articles.

2. Otherwise first the news article content extractor will crawl the html to extract the plain text of the news article. Afterwards NLP algorithms assign the labels, extract the claims, check for the similarity to the verified claims, assign the check worthiness score. In the end these information will be stored in the secondary database.

3.3 Qualities of the components

- The main database will contain claims that have been verified before, such as from ClaimsKG a collection of claims from different fact check organisations. The additional database is used to store information about the already checked articles such as url, metadata, date of the check, topic, labels etc. The databases should take advantage of the use of full text search, such as Elasticsearch.
- The extracted main claims of the article and the verified claims are compared by using word embeddings of keywords, and ranked. The method will be trained with the CLEF CheckThat 2020 Task [9] 2 dataset (and if available compared to the results)[1].
- Similarity for the suggested articles should be based on topic. Articles should be ranked with respect to the labels, the contained claims the check worthiness score.
- We hypothesise certain combinations of label scores indicate check worthiness of articles or provide data for research on specific topic, for example high technicality and bad readability, emotional language and controversial topic.

3.4 Basic ideas & implementation ideas for labels.

The following labels should be considered for the estimation of the check worthiness and as guidance for web users:

- Factuality/ Opinion: The ratio over the article text therefore more factual statements should increase the check worthiness; [2] NLP to calculate the ratio of factual to opinion sentences over the article content.
- Emotion: high usage emotional language can be used to deceive consumers therefore increasing check worthiness; [3], [4] NLP of sentences and sum over the text will tell the ratio of positive and negative emotions.
- Controversy: Hence the nature of contradicting viewpoints, the interest to check for factual content increases; comparing topics per paragraph in the article to known controversial topics to determine the number of controversial topics over the text.
- Credibility of the source: As we are likely dealing with text from a big spectrum of topics, the origin of the article has to be considered too; checking the metadata with available resources like "Wikipedia:Reliable sources/Perennial sources" to find out if we are have a reliable sources.
- Virality: viral topics should be considered as these topics circulate quicker, which leads bigger spread of miss information if not verified.

4 References

[1] Fuhr, Norbert et al. "An Information Nutritional Label for Online Documents." SIGIR Forum 51 (2018): 46-66.

[2] Sahu, Ishan & Majumdar, Debapriyo. (2017).Detecting Factual and Non-Factual Content in News Articles. 1-12. 10.1145/3041823.3041837.

[3] AFFIN Database Informatics and Mathematical Modelling, Technical University of Denmark (2011)

[4] Cambria, Erik et al. "The Hourglass of Emotions." COST 2102 Training School (2011).

[5] Cédric Lespagnol, Josiane Mothe, and Md Zia Ullah. 2019. Information Nutritional Label and Word Embedding to Estimate Information Check-Worthiness. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'19). Association for Computing Machinery, New York, NY, USA, 941–944. DOI:https://doi.org/10.1145/3331184.3331298

[6] Majithia, Sarthak & Arslan, Fatma & Lubal, Sumeet & Jimenez, Damian & Arora, Priyank & Caraballo, Josue & Li, Chengkai. (2019). ClaimPortal: Integrated Monitoring, Searching, Checking, and Analytics of Factual Claims on Twitter. 153-158. 10.18653/v1/P19-3026.

[7] Agez, Romain et al. "IRIT at CheckThat! 2018." CLEF (2018).

[8] "Understanding the Promise and Limits of Automated Fact-Checking." (2018).

[9] Barron-Cedeno, Alberto & Elsayed, Tamer & Nakov, Preslav & Martino, Giovanni & Hasanain, Maram & Suwaileh, Reem & Haouari, Fatima. (2020). CheckThat! at CLEF 2020: Enabling the Automatic Identification and Verification of Claims in Social Media.

[1] the conference will take place on the 22th-25th of September; not sure if the results will be published before.