Exposé for a Bachelor's thesis

Elitsa Pankovska

16.11.2020

Working title:

Fact Checking using trusted Knowledge Bases

1. Introduction

The spread of misinformation, especially online, is easier than ever since online content has become the main source for obtaining up-to-date information. Particularly during a time of a global pandemic, it is crucial to find an efficient way of determining the credibility of information found online. From fake news to conspiracy theories, it is hard to fight the "infodemic" which eventually makes the pandemic even more dangerous. Currently, fact-checking websites such as Snopes use manual fact-checking to validate claims from articles, speeches, or even social media posts from known figures, etc. Their work could be made more efficient by automatic claim-extraction from the information sources that need to be fact-checked.

A lot of research has been invested in both claim verification [2] and fact-check-worthiness [10], but there is no work done yet on extraction of dubious claims, combined with their fact checking using external information sources, such as knowledge graphs and knowledge bases, especially on the COVID-19 domain. The goal of this thesis is to develop a high-performance component for fact checking of small- to medium-sized documents in English and/or German language on the topic of COVID-19. In the scope of the thesis, fact checking refers to checking the veracity of claims, statements or sentences in a text (document).

2. Methodology

2.1. Sentence classification

Given a text document as input, the system should classify each sentence into one of two possible categories: suspicious sentences with potentially dubious claims and sentences that do not seem irregular. Alternatively, the sentences might also be classified into three groups: regular, suspicious and relevant, suspicious but irrelevant, the latter probably containing claims that have little to do with COVID-19. This step would increase the performance and the speed of the component since the majority of the sentences in a text are either not necessarily relevant or do not contain (potentially dubious) claims and therefore do not need to be further evaluated in the next step. This is the main machine learning part in the thesis, which is supposed to make use of current neural approaches. The initial idea is to use pre-trained

language model and using a training dataset, fine-tune it to suite the task challenges better. Since the work is to be done on the COVID-19 domain, a suitable model might be SciBERT (a BERT model trained on scientific text) [1], but it might be useful to compare the performance of alternative models such as the classical BERT or RoBERTa. The following broader research question: can – on a higher level – specific patterns, words, terms, constructions be identified in dubious sentences, can be explored in the thesis using the results from this step.

2.2. Claim extraction

The next step in the pipeline would be to extract the claims in the sentences, which were identified as suspicious. For this step, typical NLP techniques such as tokenization, lemmatizing, named entity recognition, etc. are to be used. This can be done with either Stanza [6] or spaCy [5], which provide the most common NLP tasks.

2.3. Claim verification

Once the (dubious) claims in the document have been identified, they must be verified with external information sources, such as knowledge graphs or knowledge bases, with which we can determine their veracity (generally, in the form of a numerical value). Fact checking tools will include the Google Fact Check Tools [7] and possibly Politifact [3], both of which have granted access to their knowledge bases via REST APIs. One relevant technical component for the fact checking is ClaimReview markup [4], which allows the annotation of manually checked claims in a standardized way. ClaimReview markup is used by many big publishers and news agencies, including by Google Fact Check.

Once the three main steps are implemented, the best suited models will be determined, and the system should be deployed as a whole. Optionally, a simple Streamlit app [10] can be built, which will cover the whole process pipeline and can also be dockerized.

3. The Data

The work will be done on the wider COVID-19 domain, for which many different data sets and also publications exist. The main purpose of the data will be to fine-tune the pre-trained language models for the sentence classification. For this reason, it makes sense to use a dataset of sentences, which are labeled either as suspicious or regular/not relevant. A dataset of this exact sort does not exist yet, so a combination of different ones will be used to achieve the desired dataset structure. The initial idea is to get the suspicious sentences from the archive of fact checking websites, such as Snopes [9] and Politifact and the regular/irrelevant ones from scholarly articles on the topic of COVID-19 [8] and claims on other topics.

4. Timeline

Duration	Implementation	Thesis
4 weeks	Related work research and Dataset preparation	Literature overview
4 weeks	Fine-tuning of the models, model evaluation	Writing the thesis
2 weeks	Claim extraction	
2 weeks	Veracity checking using external sources	
2 weeks	Deployment of the final system	
2 weeks	Optional: visualization	
3 weeks	-	
1 week	-	Proofreading

5. References

[1] https://github.com/allenai/scibert

[2] https://github.com/allenai/scifact

[3] https://politifact-py.readthedocs.io/en/latest/

[4] https://schema.org/ClaimReview

[5] https://spacy.io/

[6] https://stanfordnlp.github.io/stanza/

[7] https://toolbox.google.com/factcheck/explorer

[8] https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge

[9] https://www.snopes.com/

[10] https://www.streamlit.io/

[11] Shaar, S., Nikolov, A., Babulkov, N., Alam, F., Barr´on-Cede, A., Elsayed, T., Hasanain, M., Suwaileh, R., Haouari, F., Da San Martino, G., Nakov, P.: Overview of the CLEF-2020 CheckThat! Automatic identification and verification of claims English: Automatic Identification and Verification of Claims in Social Media