Automatic Fact Checking based on External Evidence

Student: Erxue Min Supervisor: Sophia Ananiadou

Research Group: NaCTeM

The recently increased focus on misinformation has stimulated research in fact checking, the task of assessing the truthfulness of a claim. Research in automating this task has been conducted in a variety of disciplines including natural language processing, machine learning, knowledge representation, databases, and journalism. While there has been substantial progress, existing methods mostly make use of the features of claims or news themselves, i.e., surface-level linguistic features, without considering valuable evidences on the web or more reliable sources like wikipeida. In our work, we explore how to retrieve related evidences for claims or news and then to check their veracity. We mainly focus on text evidence and knowledge graphs, and we are interested in applying state-of-the-art graph neural networks and pretraining models like BERT, to automatically learn features from textual inputs and make the decision as to whether a claim is true or false.

Introduction

Misinformation such as fake news is one of the big challenges of our society. Research on automated fact-checking has proposed methods based on supervised learning, but these approaches mostly rely on the information of claims or news themselves, or related replies on social media. We define these information as inner evidences. They can be categorized as follows:

- Surface-level linguistic features. The linguistic cues and features in written text are useful in identifying deceptive language.
- Additional metadata in fact checking such as the originator of the claim, speaker profile and the media source in which the claim is presented.
- The distribution of posts on social networks and users' stance towards the posts also provide useful insights.

In our work, we explore to use external knowledge to verify information. We mainly focus textual evidences and knowledge graph. Text evidences include encyclopedia articles, verified news and scientific journals contain information that can be used to check claims. Knowledge graphs in general domain include Wordnet, ConceptNet, or DBpedia, providing valuable commonsense knowledge, while in specific domain like biomedical domain, biomedical knowledge bases like UMLS can be reliable evidence source.

Datasets and Existing Results

FEVER is a benchmark dataset for fact extraction and verification. Each instance in it consists of a claim, groups of ground-truth evidence from Wikipedia and a label indicating its Veracity. Furthermore, it is attached with a dump of Wikipedia, which contains 5,416,537 preprocessed documents.

Split	SUPPORTED	REFUTED	NEI
Training	80,035	29,775	35,659
Dev	6,666	6,666	6,666
Test	6,666	6,666	6,666

Method	Label Acc. (%)	FEVER Score (%)
Athene	65,46	61.58
UCL Machine Reading Group	67.62	62.52
UNC-NLP	68.21	64.21
GEAR-single	71.60	67.10
DREAM	76.85	70.60

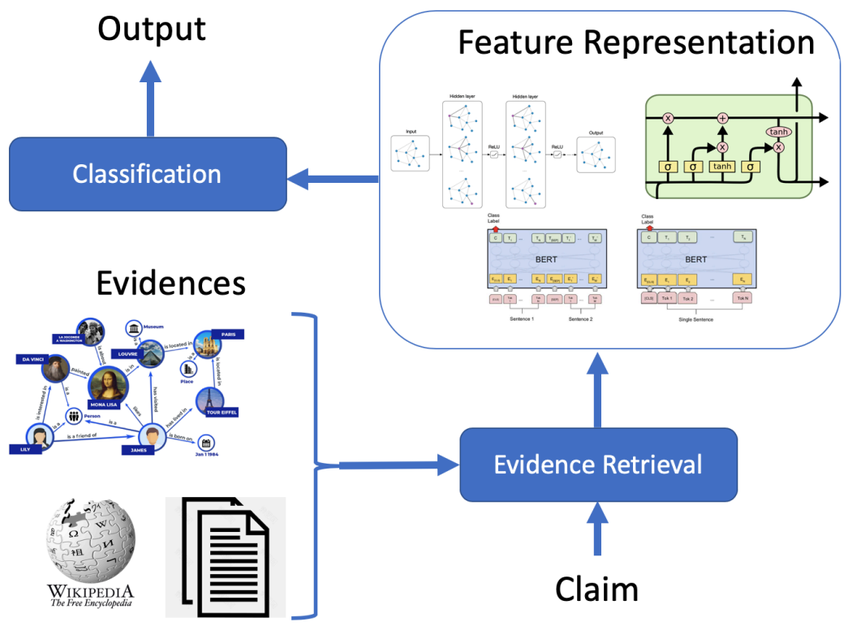
Conclusion

Fact checking is important due to the recently increased misinformation on social media. Inner information like surface-level linguistic features of claims or news is not enough for checking their veracity, thus in our work, we focus on checking information with the help of external documents and knowledge bases, which contains valuable evidences.

Methodology

The procedure of automatic fact checking system can be illustrated as follows:

- Evidence Retrieval:Given a claim as input, the evidence retrieval model retrieve top *k* related evidences from external knowledge sources. For textual sources like wikipedia, the retrieval procedure can includes document retrieval and sentence retrieval. Metrics like TF-IDF vector similarity can be applied for selection. For knowledge base, entity linking models can be used to retrieve the related concepts.
- Feature Representation: After we collected related evidences from corpus base and knowledge base, we require models to represent these evidences as well as claim. For textual input, we can use off-the-shelf NLP models such as LSTM, CNN, BERT to get context-aware representation, while for graph-structure evidences, we can use graph neural networks to keep their graph relations.
- Claim Classification When we get the feature representation of claim and evidences, we can use a final classifier to get the veracity class, which can be True, False or Not Enough Information (NEI).



References

Thorne J, Vlachos A.
Automated fact checking: Task formulations, methods and future directions.
arXiv preprint arXiv:1806.07687, 2018.

Thorne J, Vlachos A, Christodoulopoulos C, et al.
FEVER: a large-scale dataset for fact extraction and verification.
arXiv preprint arXiv:1803.05355 (2018).

Lv S, Guo D, Xu J, et al.
Graph-Based Reasoning over Heterogeneous External Knowledge for Commonsense Question Answering.
arXiv preprint arXiv:1909.05311 (2019).

Zhong W, Xu J, Tang D, et al.
Reasoning Over Semantic-Level Graph for Fact Checking.
arXiv preprint arXiv:1909.03745, 2019.

Devlin J, Chang M W, Lee K, et al.
Bert: Pre-training of deep bidirectional transformers for language understanding.
arXiv preprint arXiv:1810.04805, 2018.