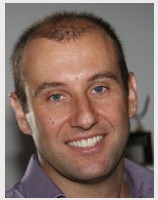
CITY
UNIVERSITY OF LONDON
EST 1894

16th International AAAI Conference on Web and Social Media

Evaluation of Fake News Detection with Knowledge-Enhanced Language Models



Chenxi Whitehouse, Tillman Weyde, Pranava Madhyastha and Nikos Komninos

City, University of London

June 2022

Overview

- Background and Motivation
- Knowledge-Enhanced Models
- Datasets
- Experimental Results
- Discussion and Conclusion

Background and Motivation

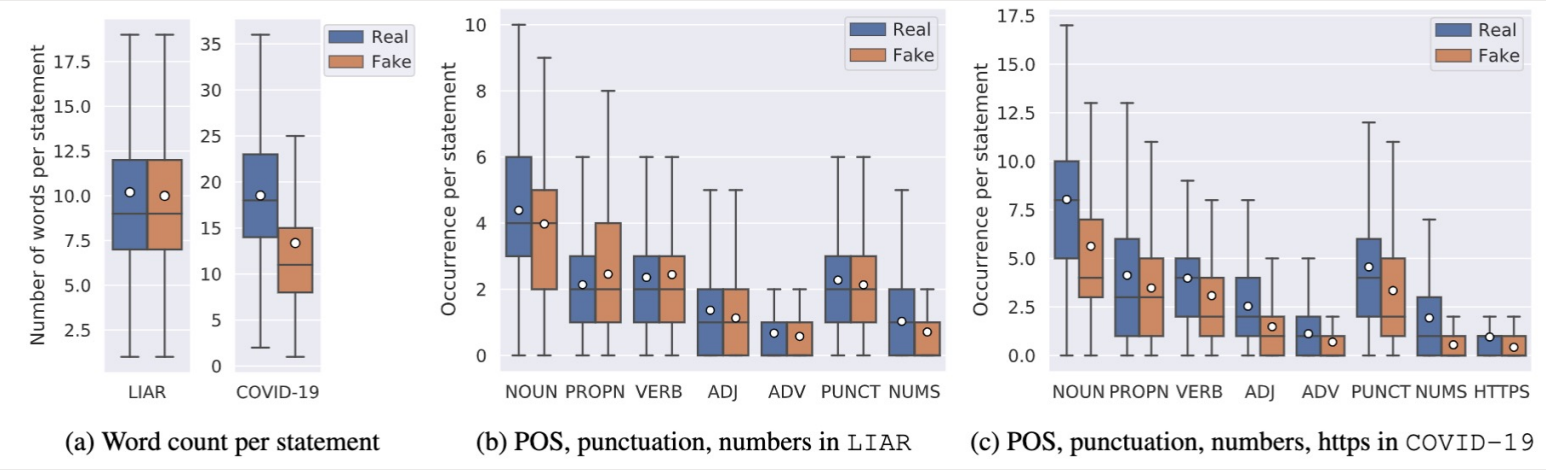
- Large Pre-trained Language Models (PLMs) achieve SOTA performance on NLP tasks and show potential for fake news detection
- PLMs usually lack explicit grounding to factual knowledge i.e., entities and relations in knowledge bases (KBs)
- Integrating knowledge into PLMs has been studied and shows promising results on entity- and knowledge-centric tasks
- This paper:
  - Expand the use of knowledge-enhanced PLMs on fake news detection
  - Empirically study and evaluate the effectiveness of various knowledge-enhanced PLMs on distinct fake news dataset

Knowledge-Enhanced Language Models

- KnowBert: BERT + WikiData and WordNet
- ERNIE: BERT + WikiData
- KEPLER: Roberta + WikiData
- K-ADAPTER: Roberta + WikiData and Linguistic features

Fake News Datasets

- LIAR (2017) and COVID-19 (2020):
  - COVID-19 has more distinct linguistic and stylistic features between two classes



(a) Word count per statement

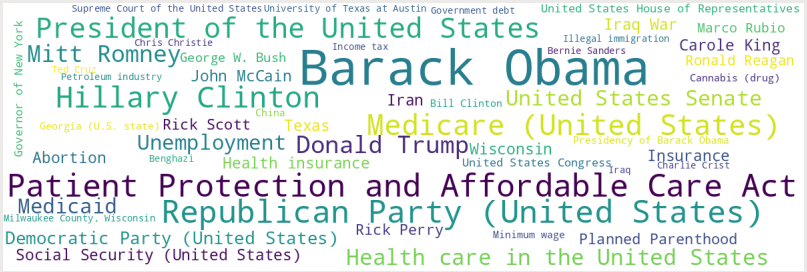
(b) POS, punctuation, numbers in LIAR

(c) POS, punctuation, numbers, https in COVID-19

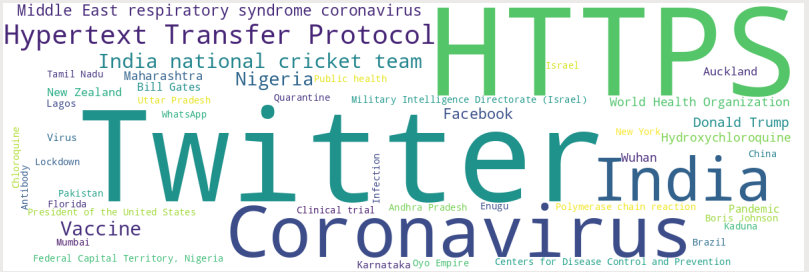
Linked Knowledge Base Entities

- Most frequent linked entities by ERNIE

  `COVID-19` show worse quality:
  - The most frequent term in the dataset, *COVID-19*, is not linked (KB used for ERNIE is not up-to-date for `COVID-19`)
  - Most frequently linked entities are irrelevant (e. g. *HTTPS*, *Twitter*)



LIAR



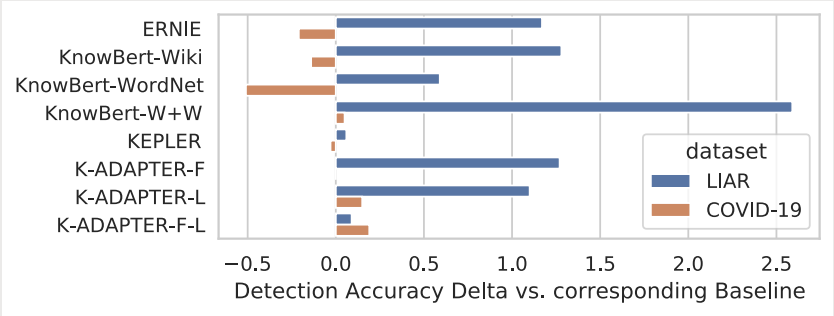
COVID-19

Experimental Results

- Knowledge-Enhanced models can improve detection accuracy on `LIAR`, where knowledge bases are current and relevant, and the dataset is not heavily skewed by stylistic features

MODEL	BASE	LIAR	COVID-19
BERT-Base (BB)	-	26.36 ±0.58	97.51 ±0.19
RoBERTa-Base (RB)	-	26.71 ±0.93	97.61 ±0.26
RoBERTa-Large (RL)	-	**27.36** ±0.79	**97.92** ±0.17
ERNIE	BB	27.53 ±0.13	97.30 ±0.18
KnowBert-Wiki	BB	27.64 ±0.09	97.37 ±0.09
KEPLER	RB	26.77 ±1.15	97.58 ±0.15
K-ADAPTER-F	RL	**28.63** ±0.90*	**97.92** ±0.10
KnowBert-WordNet	BB	26.95 ±0.45	97.00 ±0.06
KnowBert-W+W	BB	**28.95** ±0.64*	97.56 ±0.15
K-ADAPTER-L	RL	28.46 ±0.87*	98.07 ±0.09
K-ADAPTER-F-L	RL	27.45 ±0.78	**98.11** ±0.14

Detection Accuracy over five runs



Discussion and Conclusion

- Knowledge-enhanced PLMs can improve fake news detection on static datasets, but it depends on
  - Data quality
  - Suitable and current KBs
- Real-World Application Aspects
  - Dynamic adaptation
  - Adversarial robustness
  - More explainability and interpretability
- Challenges and future work:
  - Need for relevant and up-to-date knowledge bases
  - Realistic test scenarios with dynamic knowledge and adversarial and automatic fake news generators


Thank you!