



Introduction

Task: Fake news detection

Motivation:

1. Large Pre-trained Language Models (PLMs) achieve SOTA performance on NLP tasks and show potential for fake news detection.
2. PLMs usually lack explicit grounding to factual knowledge i.e., real world entities and relations in knowledge bases (KBs), which is also important for fake news detection.
3. Integrating knowledge into PLMs has been studied and shows promising results on entity- and knowledge-centric tasks.

Approach: Expand the use of knowledge-enhanced PLMs to fake news detection, empirically study and evaluate the effectiveness of various knowledge-enhanced PLMs on distinct fake news dataset.

Knowledge-Enhanced PLMs

4 recent Knowledge-Enhanced PLMs:

- ERNIE: BERT + WikiData
- KnowBert: BERT + WikiData and WordNet
- KEPLER: RoBERTa + WikiData5M
- K-ADAPTER: RoBERTa + WikiData and linguistic knowledge

Fake News Dataset

2 different publicly available fake news datasets:

LIAR (2017) and COVID-19 (2020)

Linguistic Features: COVID-19 shows differences in stylistic features between classes, especially word count and https links.

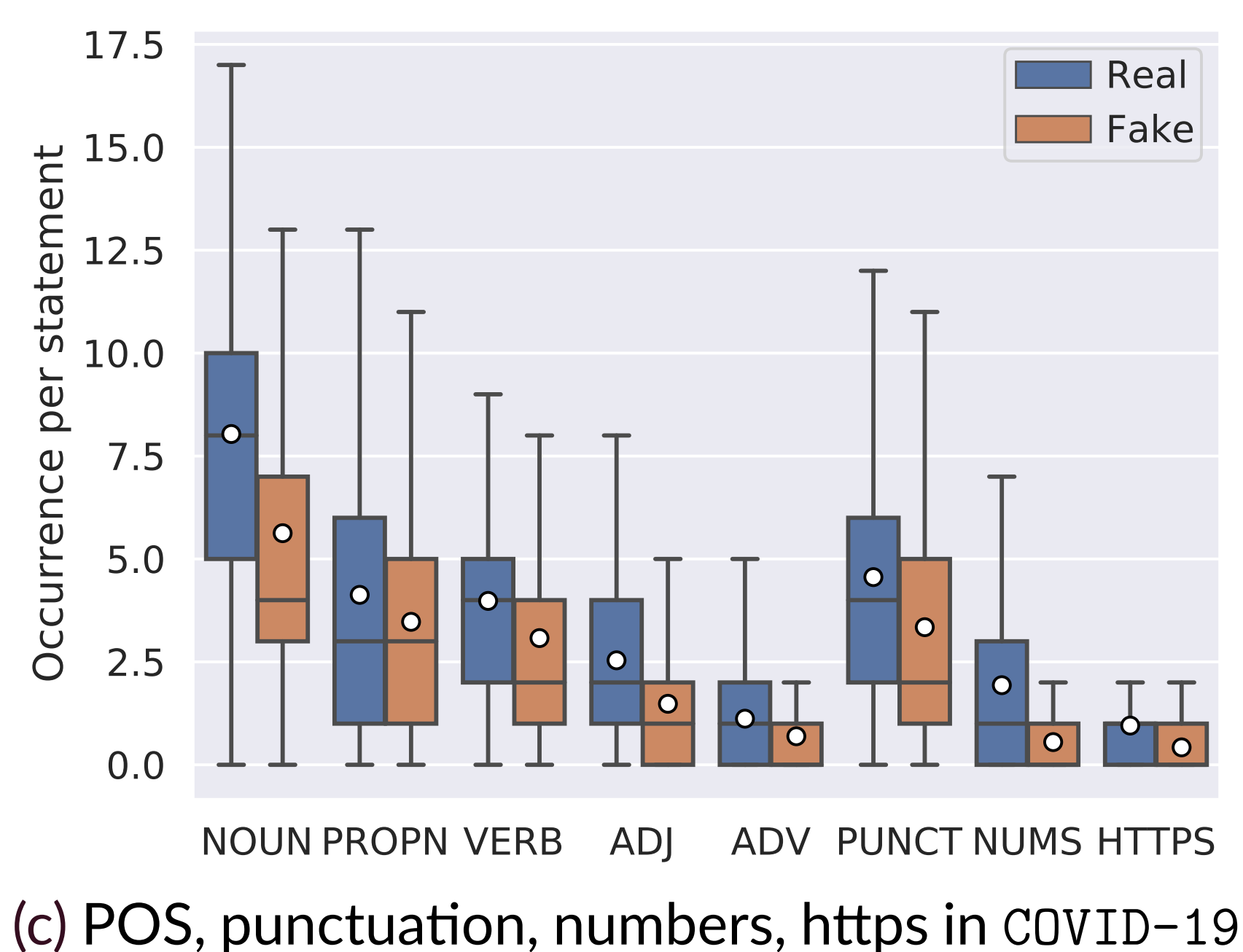
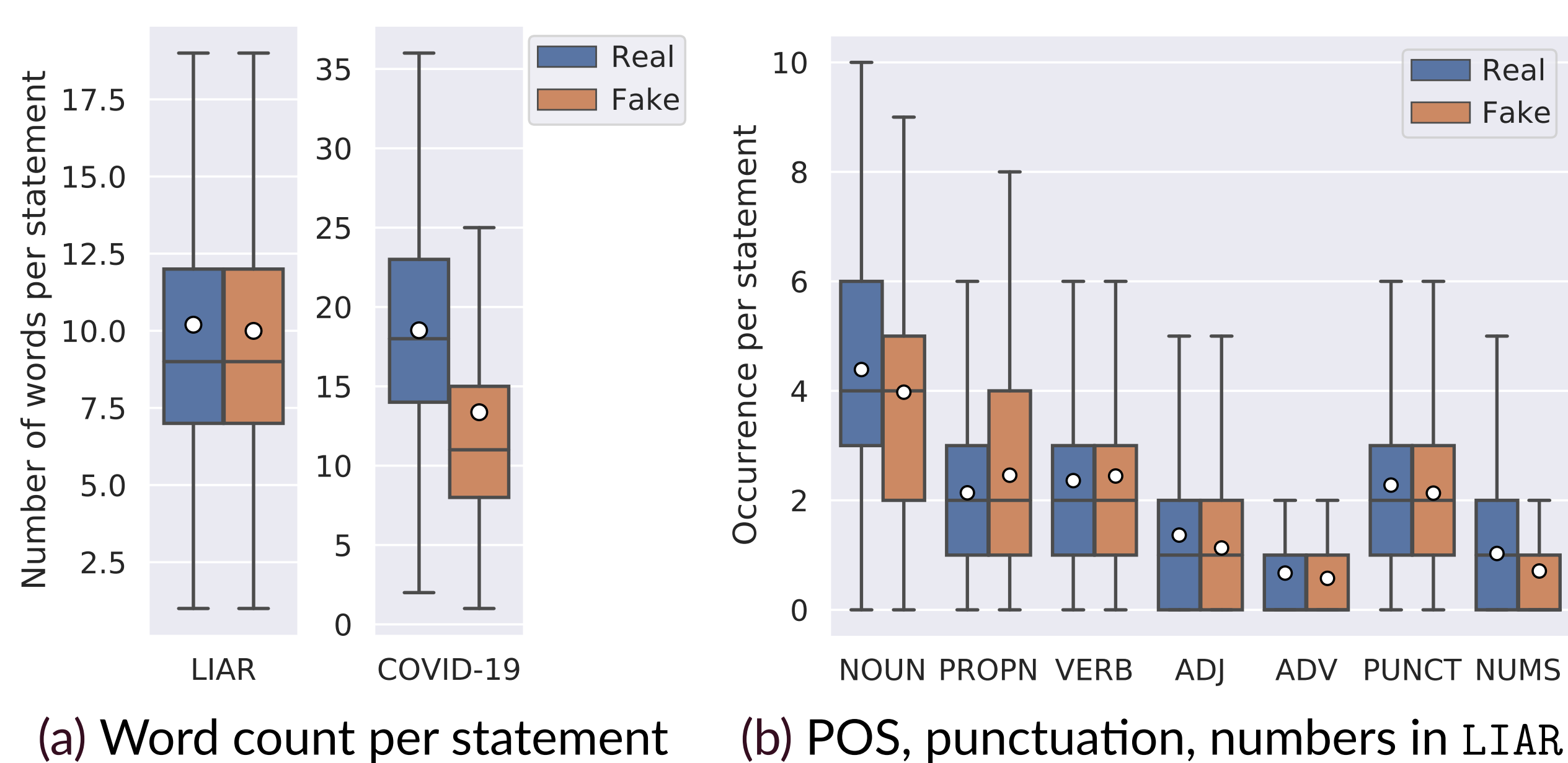


Figure 1. Word count, POS tags, punctuation and numbers per statement in real and fake news in LIAR and COVID-19, and number of https-links per statement in COVID-19. The mean values are shown as white filled circles.

Linked Knowledge Base Entities

Linked KB entities are essential for knowledge integration.

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 linked entities are irrelevant (e.g. *HTTPS*, *Twitter*)

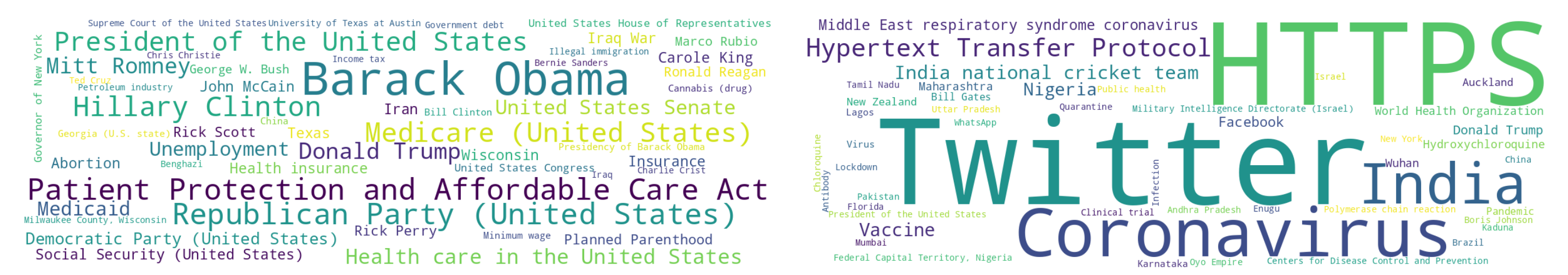


Figure 2. Word clouds for the 50 most frequent KB entities linked by ERNIE in LIAR (left) and COVID-19 (right).

Experimental Results

Knowledge-enhanced models compared to baseline:

significant improvement on LIAR, mixed results on COVID-19.

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

Table 1. Detection accuracy results (average of five runs). Results with * indicate statistically significant improvements over the baseline, both for mean (t-test, one-sided, $p < .05$) and median (Wilcoxon signed rank test, one-sided, $p < .05$)

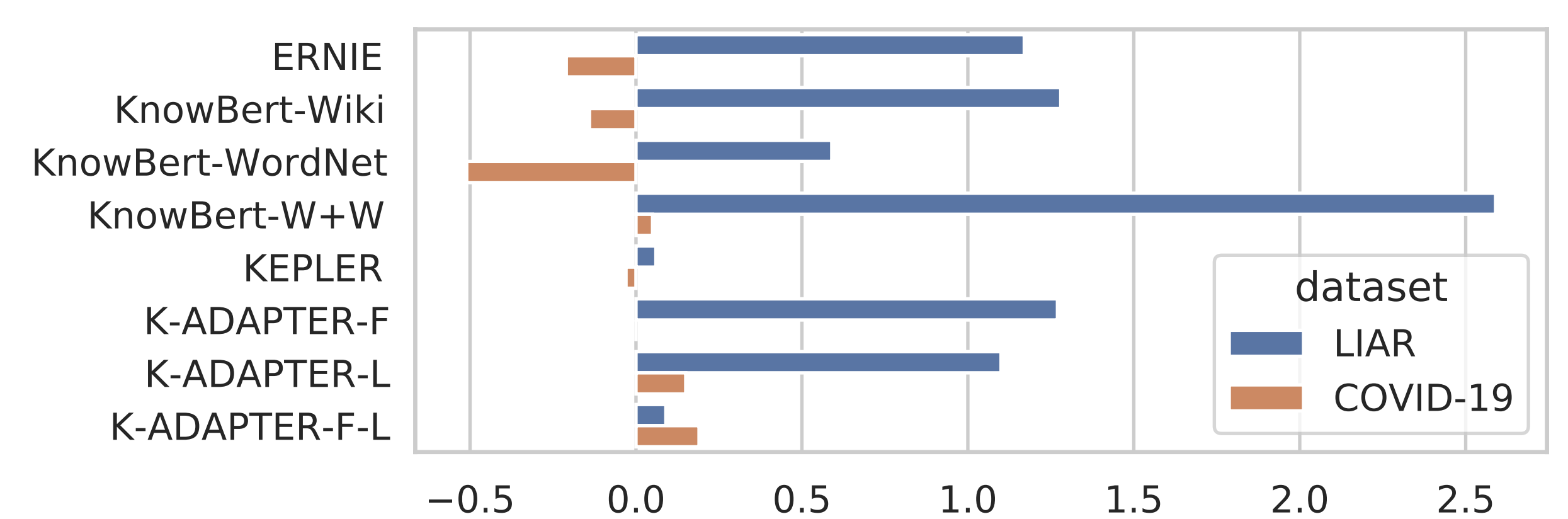


Figure 3. Detection Accuracy Delta vs. corresponding baselines.

Discussion and Conclusions

Integrating knowledge into PLMs can improve the fake news detection accuracy on static datasets, but it depends on the data quality, and 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