

MDFEND: Multi-domain Fake News Detection

Qiong Nan^{1,2,3}, Juan Cao^{1,2}, Yongchun Zhu^{1,2}, Yanyan Wang^{1,2}, Jintao Li¹

1. Key Lab of Intelligent Information Processing of Chinese Academy of Sciences
(CAS) , Institute of Computing Technology, CAS, Beijing, China

2. University of Chinese Academy of Sciences, Beijing, China

3. State Key Laboratory of Media Convergence Production Technology and Systems,
Beijing, China

Motivation



- Existing works almost focus on **single-domain fake news detection**
- Social media releases news from **many domains** every day.
- There exists serious **domain shift** between different domains (propagation pattern, topic distribution, et al.)

Multi-domain Fake News Detection



中国科学院计算技术研究所

INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES

Weibo21: A New Dataset

- Data statistics of nine domains

domain	Science	Military	Education	Disasters	Politics
real	143	121	243	185	306
fake	93	222	248	591	546
all	236	343	491	776	852

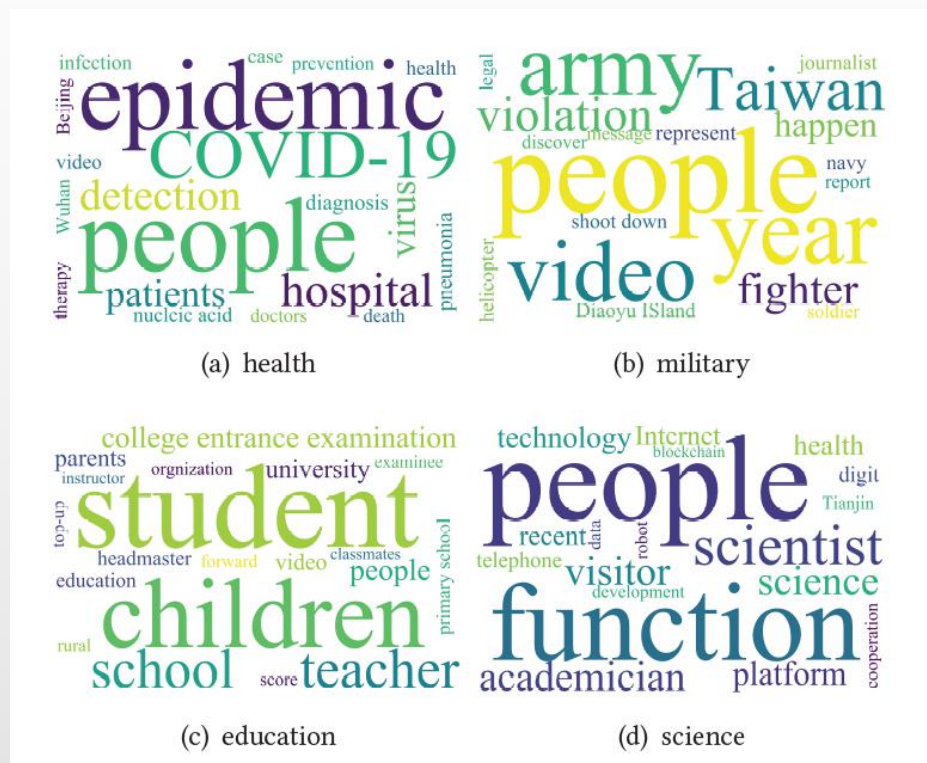
domain	Health	Finance	Entertainment	Society	All
real	485	959	1000	1198	4640
fake	515	362	440	1471	4488
all	1000	1321	1440	2669	9128

- **Data source:**
 - Sina Weibo
- **Time period:**
 - from December 2014 to March 2021
- **Data composition:**
 - News content
 - Corresponding pictures
 - Timestamp
 - Comments
 - Judgement information (for fake news)



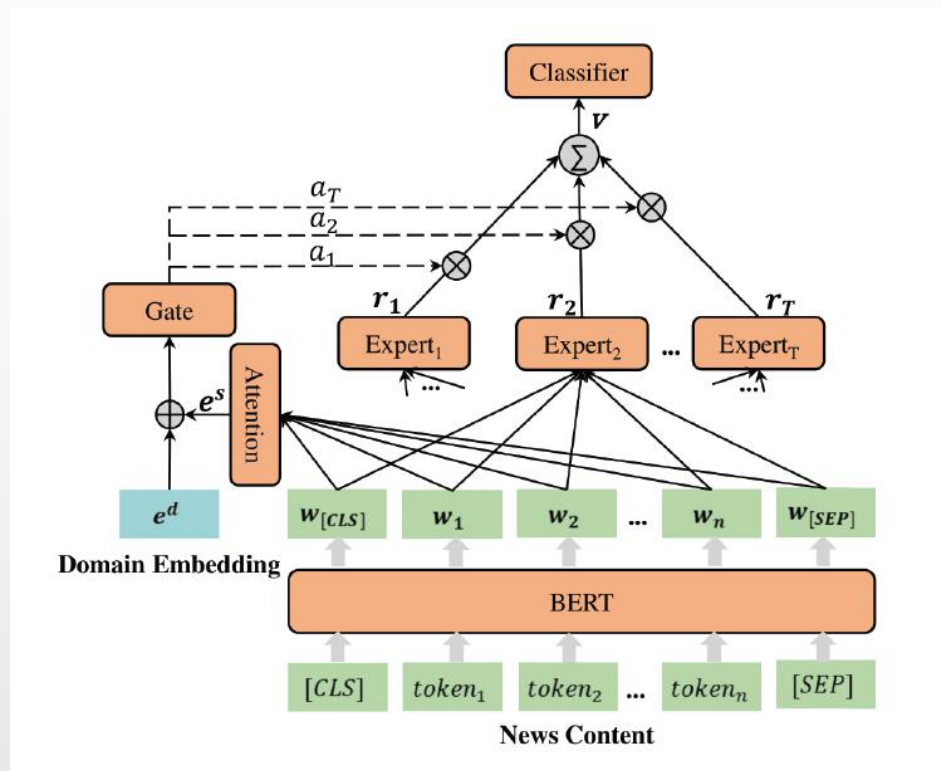
Weibo21: A New Dataset

- Preliminary analysis



- As the most straightforward clue, we analyze the topic distribution of news among different domains. We choose four domains to show off, and it can be observed that different domains have different frequently used words, which illustrates the phenomenon of domain shift.

MDFEND framework



- Expert module: $r_i = \Psi_i(W; \theta_i)$
- Domain gate: $\mathbf{a} = \text{softmax}(G(e^d \oplus e^s; \phi))$
- Aggregation: $v = \sum_{i=1}^T a_i r_i$
- Prediction: $\hat{y} = \text{softmax}(MLP(v))$

Experiments

single-domain

model	Science	Military	Education	Accidents	Politics	Health	Finance	Entertainment	Society	All
TextCNN_single	0.7470	0.778	0.8882	0.8310	0.8694	0.9053	0.7909	0.8591	0.8727	0.8380
BiGRU_single	0.4876	0.7169	0.7067	0.7625	0.8477	0.8378	0.8109	0.8308	0.6067	0.7342
BERT_single	0.8192	0.7795	0.8136	0.7885	0.8188	0.8909	0.8464	0.8638	0.8242	0.8272
TextCNN_all	0.7254	0.8839	0.8362	0.8222	0.8561	0.8768	0.8638	0.8456	0.8540	0.8686
BiGRU_all	0.7269	0.8724	0.8138	0.7935	0.8356	0.8868	0.8291	0.8629	0.8485	0.8595
BERT_all	0.7777	0.9072	0.8331	0.8512	0.8366	0.9090	0.8735	0.8769	0.8577	0.8795
EANN	0.8225	0.9274	0.8624	0.8666	0.8705	0.9150	0.8710	0.8957	0.8877	0.8975
MMOE	0.8755	0.9112	0.8706	0.8770	0.8620	0.9364	0.8567	0.8886	0.8750	0.8947
MOSE	0.8502	0.8858	0.8815	0.8672	0.8808	0.9179	0.8672	0.8913	0.8729	0.8939
EDDFN	0.8186	0.9137	0.8676	0.8786	0.8478	0.9379	0.8636	0.8832	0.8689	0.8919
MDFEND	0.8301	0.9389	0.8917	0.9003	0.8865	0.9400	0.8951	0.9066	0.8980	0.9137

mixed-domain

multi-domain

- Metric: f1-score
- Baselines: single-domain baselines; mixed-domain baselines; multi-domain baselines



中国科学院计算技术研究所



INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES

Conclusion

- We constructed Weibo21, the first multi-domain fake news detection (MFND) dataset collected from one platform with richest domains;
- We proposed a simple but effective method named MDFEND for MFND, which utilizes domain gate to aggregate multiple representations extracted by mixture-of-experts;
- We evaluated MFND performance with different methods on our proposed Weibo21 dataset, and experiments show the effectiveness of our MDFEND model.



Thanks Q&A

 Qiong Nan |  nanqiong19z@ict.ac.cn



中国科学院计算技术研究所
INSTITUTE OF COMPUTING TECHNOLOGY, CHINESE ACADEMY OF SCIENCES