Supplementary Material for Unsupervised Cross-Domain Rumor Detection with Contrastive Learning and Cross-Attention

Hongyan Ran[1] **, Caiyan Jia**[1*]
[1] School of Computer and Information Technology & Beijing Key Lab of Traffic Data Analysis and Mining
Beijing Jiaotong University, Beijing 100044, China
{hongyran,cyjia}@bjtu.edu.cn

Abstract

This is the supplementary material for the paper titled "Unsupervised Cross-Domain Rumor Detection with Contrastive Learning and Cross-Attention".

PHEME Dataset and Relative Experiments

This section presents the reasons that the PHEME dataset (Zubiaga et al. 2015) is divided into two domain data, and we use the original PHEME dataset to construct the relative experiment to further verify our proposed method.

PHEME Dataset

The original PHEME dataset (Zubiaga et al. 2015) contains nine breaking events are listed in Table 1. Firstly, we briefly describe each event to show its properties.

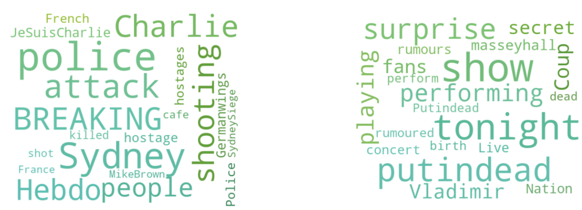
Table 1: Statistics of PHEME Dataset.

Events	Tweet	Rumors	Non-rumors
Charlie Hebdo (C)	2079	458	1621
Sydney Siege (S)	1221	522	699
Ferguson Unrest (F)	1143	284	859
Ottawa shooting (O)	890	470	420
Germanwings Crash (G)	469	238	231
The Missing of Putin (M)	238	126	112
Toronto Prince Show (T)	233	229	4
Gurlitt Trove (R)	138	61	77
Ebola Essien (E)	14	14	0
Total	6425	2402	4023

- Charlie Hebdo (C): A terrorist attack in which two brothers forced their way into the offices of the French satirical weekly newspaper Charlie Hebdo in Paris, killing 11 people and wounding 11 more, on January 7, 2015.
- Sydney Siege (S): A hostage situation in which a gunman held hostage ten customers and eight employees of a Lindt chocolate caf é located at Martin Place in Sydney, Australia, on December 15, 2014.

[*]Corresponding authors

- Ferguson Unrest (F): Citizens of Ferguson in Michigan, USA, protested after the fatal shooting of an 18-year-old African American, Michael Brown, by a white police officer on August 9, 2014.



(a) Terrorist Domain. (b) Gossip Domain.

Figure 1: Word clouds of the terrorist domain and Gossip domain is used to visualize the topic information.

- Ottawa Shooting (O): A serial shooting occurred on Ottawa's Parliament Hill in Canada, resulting in the death of a Canadian soldier on October 22, 2014.
- Germanwings Crash (G): A plane crash event that a passenger plane from Barcelona to Düsseldorf crashed in the French Alps on March 24, 2015, killing all passengers and crew.
- The Missing of Putin (M). An online farce that numerous rumors emerged in March 2015 when the Russian president Vladimir Putin did not appear in public for 10 days.
- Toronto Prince Show (T). A celebrity gossip says that a rumor started circulating on November 3, 2014, that the singer Prince would play a secret show in Toronto that night.
- Gurlitt Trove (R). A politically sensitive event that a rumor in November 2014 that the Bern Museum of Fine Arts was going to accept a collection of modernist masterpieces kept by the son of a Nazi-era art dealer.
- Ebola Essien (E). A celebrity gossip that a post by a Twitter user on October 12, 2014, stated that the AC Milan footballer Michael Essien had contracted Ebola.

From the above description, we find these nine events mainly focus on two types of topics including the 'Terrorist' topic and the ‘Gossip’ topic, where the C, S, F, O, G, and R events talk about the ‘Terrorist’ topic, and other events discuss the ‘Gossip’ topic. We construct the WordCloud to visualize the top 30 topic words for the two types of topics as shown in Fig 1. From Fig 1, we can observe that the topic words of the ‘Terrorist’ topic are ‘breaking’, ‘shooting’, ‘attack’, ‘killed’, and ‘police’, etc, the topic words of the ‘Gossip’ topic are ‘surprise’, ‘secret’, ‘performing’, ‘playing’, and ‘fans’, etc. This indicates that the two types of topics talk about different content and have a large discrepancy in rumor content. Therefore, we divide the PHEME dataset as cross-domain rumor data, the ‘Terrorist’ topic as the source domain and ‘Gossip’ topic as the target domain to verify the performance of our UCD-RD model.

Table 2: Rumor detection results (%) on cross-events in the PHEME dataset, where C→S represents the source events to the target events, and so do the others.

Method	Rumor-GAN	PPA-WAE	PLAN	UCD-RD
C→S	44.01	60.31	**76.01**	74.48
C→F	55.89	74.33	64.56	**77.88**
C→O	32.67	75.59	77.87	**80.74**
C→G	30.02	55.56	**67.59**	53.74
C→M	51.24	48.74	52.10	**56.10**
C→T	48.62	1.72	3.43	**76.59**
S→C	42.68	82.31	83.07	**83.39**
S→F	47.23	**75.24**	75.24	74.64
S→O	71.25	58.41	66.85	**74.20**
S→G	66.01	53.73	59.06	**67.53**
S→M	33.23	49.16	47.06	**64.63**
S→T	41.25	7.30	1.72	**73.95**
F→C	65.22	77.25	80.95	**82.62**
F→S	50.12	57.85	58.80	**68.90**
F→O	42.75	50.48	58.99	**63.83**
F→G	32.33	50.53	53.09	**72.66**
F→M	**58.72**	47.90	47.06	57.26
F→T	42.58	2.15	3.00	**67.78**
O→C	36.25	81.21	**85.23**	83.34
O→S	59.42	66.78	66.01	**75.15**
O→F	43.25	46.23	75.68	**76.25**
O→G	**74.31**	57.14	64.18	72.13
O→M	43.32	54.20	48.32	**67.07**
O→T	46.94	15.45	2.15	**87.53**
Average	46.75	52.88	54.91	**72.18**

The Experimental Results on Cross-Events

In order to further verify the performance of the UCD-RD model, we use the nine events of the original PHEME dataset as the cross-event data to verify our model’s performance. We divide the nine events into three parts, (1) The four largest events (C, S, F, O) which have enough data as source events; (2) the small events (G, M, T) which have relatively fewer data as target events; (3) The last two events R and E are abandoned since the data is too small. We construct 24 tasks for the cross-event experiment, the results are shown in Table 2.

From Table 2, we can observe that the UCD-RD model has achieved the best performance in 18 cases of 24 tasks compared with these baseline methods. Moreover, the average accuracy of the UCD-RD model on 24 tasks is far beyond the baseline models and achieves 31.45% improvement, which demonstrates that the cross-domain contrastive learning and cross-attention mechanism can learn the event-invariant features to improve the detection performance in a cross-event scenario. Therefore, our proposed UCD-RD method not only learns the event-invariant features but also acquires the domain-invariant features to alleviate the problem of domain shift.

References

Zubiaga, A.; Liakata, M.; Procter, R.; Bontcheva, K.; and Tolmie, P. 2015. Crowdsourcing the annotation of rumourous conversations in social media. In *Proceedings of the 24th international conference on World Wide Web*, 347–353.