Predicting Rumor Veracity on Social Media with Graph Structured Multi-task Learning

Yudong Liu[1,2], Xiaoyu Yang[1], Xi Zhang[1(✉)], Zhihao Tang[1], Zongyi Chen[1], and Zheng Liwen[1]

[1] Key Laboratory of Trustworthy Distributed Computing and Service (MoE), Beijing University of Posts and Telecommunications, Beijing, China
{yudong.liu,littlehaes,zhangx,innerone,zongyi_chen,zhenglw}@bupt.edu.cn
[2] Beijing Electronic and Science Technology Institute, Beijing, China

Abstract. Previous studies have shown that the multi-task learning paradigm with the stance classification could facilitate the successful detection of rumours, but the shared layers in multi-task learning tend to yield a compromise between the general and the task-specific representation of structural information. To address this issue, we propose a novel **M**ulti-**T**ask **L**earning framework with **S**hared **M**ulti-channel **I**nteractions (MTL-SMI), which is composed of two shared channels and two task-specific graph channels. The shared channels extract task-invariant text features and structural features, and the task-specific graph channels, by interacting with the shared channels, extract the task-enhanced structural features. Experiments on two realworld datasets show the superiority of MTL-SMI against strong baselines.

Keywords: Rumor veracity · Stance classification · Multi-task learning · Graph neural network

1 Introduction

A rumor is an item of circulating information whose veracity status has not been verified at the time of posting. Recently, related research has proved the high pertinence between the rumor verification and stance classification tasks. Most of these studies aim to use various deep neural networks, such as LSTM [2], GRU [8] and Transformer [18], to extract textual features. However, they failed to make full use of the network structure existed in social media, which has been proven effective for rumor detection tasks [5–7].

As shown in Fig. 1 (a), a conversation thread consists of a source post with a veracity label, comments with stance labels, and the users. Figure 1 (b) is the constructed conversation network. Comment 4 and 2 come from the same user and have similar stance toward the source post, hence having less impact on

Y. Liu and X. Yang—Equal contribution.

A. Bhattacharya et al. (Eds.): DASFAA 2022, LNCS 13247, pp. 230–237, 2022.
https://doi.org/10.1007/978-3-031-00129-1_16

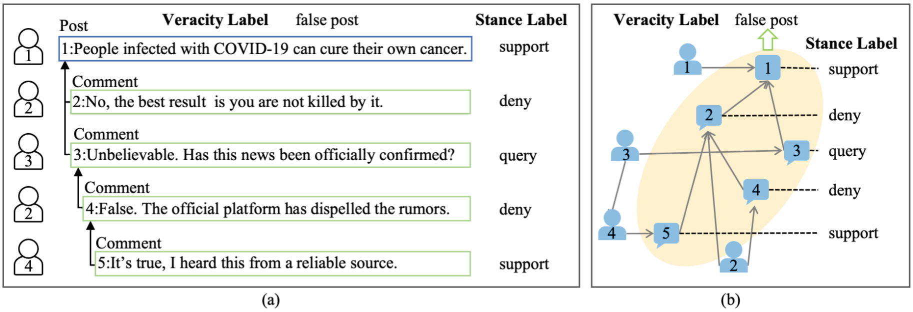

Fig. 1. (a) A conversation thread with four different users. The leading numbers are the timestamp order. (b) The conversation network based on the left thread.

rumor verification task than two independent comments, which motivates us to employ extra task-specific channels in MTL-SMI. Besides, we use combination of different neural networks in the shared channels and interactions between channels to enhance the ability of representation learning.

Our contributions can be summarized as follows:

– To improve the accuracy of rumor verification, we propose a multi-task learning framework MTL-SMI, which utilizes a shared text channel combining different neural networks, three graph channels, and multi-channel interactions to enhance the ability of task-specific representation learning.
– We conduct extensive experiments on two benchmark datasets and the results demonstrate the effectiveness of our key components.

2 Related Work

Rumor verification task focuses on the truthfulness of a rumor post. Previous studies [12–14] commonly rely on traditional hand-crafted features. Nowadays, deep learning methods are more popular. Ma et al. [15] propose two recursive neural models respectively for rumor representation learning and classification. Dungs et al. [16] model the temporal changes in stance information with multi-spaced Hidden Markov Model. Li et al. [17] generates the representation of posts through an attention-based LSTM network and use an ensemble of the traditional classification algorithms and neural network models for rumor verification.

Stance classification has gradually matured in recent years. Kochkina et al. [1] build a multi-task learning framework with hard parameter sharing. Ma et al. [3] propose a model with both shared layers and task-specific layers to enhance each task. Li et al. [4] take the user credibility into consideration. Wei et al. [7] consider both the text feature, and structural features of the conversation network. Yu et al. [10] propose a Coupled Transformer Module to capture the interactions among tasks and promote the accuracy of rumor verification by using the predicted stance labels. Table 1 shows the differences among these multi-task learning frameworks.

Table 1. Multi-task learning frameworks for rumor verification task. "N-Interaction" refers to the number of interaction channels to extract representative features.

MTL approch	Post	Comment	User	Network	#N-Interaction
Kochkina [1]	✓	✓	×	×	1
Yu [10]	✓	✓	×	×	1
Ma [3]	✓	✓	×	×	1
Li [4]	✓	✓	✓	×	1
Wei [7]	✓	✓	×	✓	0
MTL-SMI (Ours)	✓	✓	✓	✓	2

3 Problem Statement

We follow the problem setting as previous studies [7,10] and denote the dataset as a set of conversations $\mathbb{D} = \left\{C_1, C_2, \ldots, C_{|\mathbb{D}|}\right\}$, where C_i is composed of a source post and corresponding comments. For conversation $C_i = \left\{S_0^i, R_1^i, R_2^i, \ldots, R_n^i\right\}$, the source post is denoted as S_0^i, and the attached n comments are denoted as $R_1^i, R_2^i, \ldots, R_n^i$. The goal of stance classification is to learn a classifier $g : (S_0^i, R_j^i) \rightarrow s_i$, and s_i takes one of the four possible stance labels: support, deny, query and comment. The goal of rumor verification is to learn a classifier $f : S_0^i \rightarrow y_i$, where y_i is of three possible labels: true, false and unverified.

Given a conversation network $G = (V, E, A)$, V is the node set including user, post, and comment nodes, E is the edge set and $\mathbf{A} \in \left\{0, 1\right\}^{|V| \times |V|}$ is the adjacency matrix. And an edge is established between the following node pair: (1) a user and the comment or post that he or she published; (2) two users according to the following or followed relationship; (3) two comments if one comment is commented by the other. Our conversation network G is an undigraph.

4 Methodology

In this section, we propose the MTL-SMI framework as Fig. 2, which consists of four channels: RV Graph Channel, Shared Graph Channel, Shared Text Channel, SC Graph Channel; and two MLP classfiers: RV and SC.

SGC (Shared Graph Channel) is a two-layer GCN [9]. Initial nodes, H^0, are represented by average of word vectors of text content or user profile.

STC (Shared Text Channel). Each text sequence is truncated or filled to fixed-length L, to which three steps are applied: (1) the pre-trained BERT [11] to get the initial word vectors; (2) SIGNED-TRANS, which is a signed co-attention layer [19] with the "-softmax" channel to capture both the positive and negative correlation among words; (3) text-cnn to aggregate the words vector into the final feature vector.

RVGC (RV Graph Channel). The difference with the SGC is in the input of the 2^{nd} layer of the GCN. To interact with two shared channels, three vectors

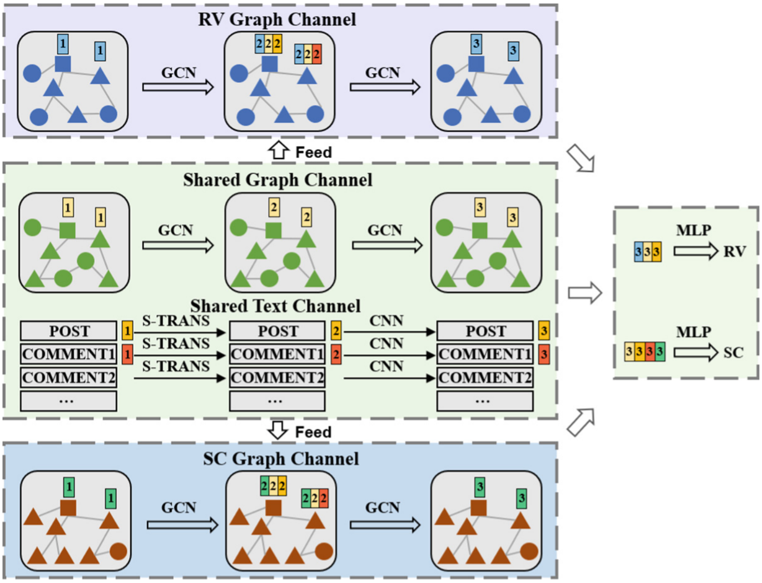

Fig. 2. Multi-task learning framework MTL-SMI. The circle, square, and triangle stand for user node, source post node, and comment node respectively (Best viewed in color).

of the same node from the 1^{st} layer of SGC, STC, and RVGC, are concatenated as input of the 2^{nd} layer of RVGC. In this way, we extend the input information for the second layer and thus extract efficient features with GCN.

SCGC (SC Graph Channel). The SCGC has the same structure as the RVGC, but dedicated to the SC task.

RV (Rumor Verification). We concatenate post node embedding from the RVGC, SGC, and STC to predicate the veracity category with a fully connected layer.

SC (Stance Classification). We concatenate comment features from the SGC, STC, SCGC, and the post feature from the STC, to predicate the stance category with a fully connected layer.

Overall Loss. A cross-entropy based loss function is defined as the overall loss of MTL-SMI:

$$\mathcal{L}(\theta_{rv}, \theta_{sc}) = \sum_{a=1}^{N_{rv}} \sum_{c_{rv} \in \{0,1,2\}} -\mathbf{y}_a^{rv} \log \hat{\mathbf{y}}_a^{rv} + \lambda \sum_{b=1}^{N_{sc}} \sum_{c_{sc} \in \{0,1,2,3\}} -\mathbf{y}_b^{sc} \log \hat{\mathbf{y}}_b^{sc} \tag{1}$$

where N_{rv} and N_{sc} represent the data volume. c_{rv} and c_{sc} are the categories. $\mathbf{y}_a^{rv}$ and $\mathbf{y}_b^{sc}$ represent the one-hot labels. λ is used to control the proportion of the stance classification task loss in the overall loss.

Table 2. Statistics of the datasets.

Dataset	Threads	Tweets	Stance labels				Rumor veracity labels		
			#Support	*#Deny*	*#Query*	*#Comment*	*#True*	*#False*	*#Unverified*
SemEval	325	5,568	1,004	415	464	3,685	145	74	106
PHEME	2,402	105,534	–	–	–	–	1,067	638	697

STL-GT (Single Task Learning Setting on the Shared Graph and Text Channel). After removing the two task-specific graph channels in Fig. 2, the left shared channels can be trained in a single task setting, which is used to illustrate the efficiency of the combined neural networks and serve as one of the base models in the ablation test.

5 Experiments

5.1 Datasets and Experiment Setup

According to previous studies [1,7,10], we experiment on two datasets: (1) The SemEval [20], whose training and validation sets are related to eight events and the test set covers another two events, with stance and veracity labels. (2) The PHEME [1] dataset contains nine events with only verification labels. We perform leave-one-event-out cross-validation on this dataset for rumor verification task. The statistical information of the two datasets is shown in Table 2.

We use Macro-F_1 as the main evaluation metric and Accuracy as the secondary evaluation metric, aiming to improve the performance of rumor verification without considering the metrics of stance classification. GCN in different channels all have two layers. λ in formula 1 is set to 1. We use the Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and learning rate is 0.00005.

5.2 Results of Rumor Verification with STL-GT and MTL-SMI

The following baselines are choosen to compair with STL-GT and MTL-SMI respectively.

- **Single Task Baselines**
  - **BranchLSTM** [?] decomposes the tree conversation structure into linear structure, and uses a sequence model based on LSTM to incorporate structural information for classification.
  - **TD-RvNN** [15] models the top-down conversation tree structure to capture complex propagation patterns and classifies rumors with RNN.
  - **Hierarchical GCN-RNN** [7] is a variant of Conversational-GCN for single tasks, which does not use the stance labels in the training process.
- **Multi-Task Baselines**
  - **BranchLSTM + NileTMRG** [21] extracts the textual features with BranchLSTM, and uses SVM for classification.

Table 3. Rumor verification results with STL-GT and MTL-SMI.

Setting	Method	SemEval		PHEME	
		$\#Macro\text{-}F_1$	$\#Acc$	$\#Macro\text{-}F_1$	$\#Acc$
Single-Task	BranchLSTM	0.491	0.500	0.259	0.314
	TD-RvNN	0.509	0.536	0.264	0.341
	Hierarchical GCN-RNN	0.540	0.536	0.317	0.356
	STL-GT	**0.618**	**0.607**	**0.359**	**0.430**
Multi-Task	BranchLSTM+NileTMRG	0.539	0.570	0.297	0.360
	MTL2(Veracity+Stance)	0.558	0.571	0.318	0.357
	Hierarchical PSV	0.588	0.643	0.333	0.361
	MTL-SMI	**0.685**	**0.679**	**0.409**	**0.468**

- **MTL2 (Veracity+Stance)** [1] adopts a multi-task learning framework with hard parameter sharing at the bottom layers, and use the output of task-specific layer for classification.
- **Hierarchical PSV** [7] uses Conversational-GCN to obtain features of the source post and comments. Then it extracts the temporal features of the conversation with GRU for rumor verification.

Performance Analysis. Table 3 shows the results of STL-GT and MTL-SMI in rumor verification task.

STL-GT achieves the best results on both datasets in the single-task setting. (1) Sequence models like BranchLSTM and TD-RvNN, perform worse than Hierarchical GCN-RNN and STL-GT, which demonstrates the importance of the structural features of conversation graph. (2) STL-GT, using BERT, SIGNED-TRANS in text processing, perform better than Hierarchical GCN-RNN with the traditional text processing technique.

MTL-SMI achieves the best results on both datasets in multi-task settings. (1) BranchLSTM+NileTMRG performs the worst in that it is not trained in an end-to-end manner. (2) Hierarchical PSV and MTL-SMI consider the structural information of conversation networks, hence performing better. (3) MTL-SMI with shared multi-channel interactions could effectively capture the relationship between two tasks and generates more powerful features for rumor verification.

5.3 Ablation Tests

In order to test the effectiveness of different components in MTL-SMI, we set up the following ablation experiments. (1) **M/INTER** is MTL-SMI without considering the interactions between task-specific channels and shared channels. (2) **M/GRAPH** only considers the interaction between Shared Text Channel. (3) **M/TEXT** only considers the interaction between Shared Graph Channel. (4) **M/SIGNED** is MTL-SMI without adding the "-softmax" channel in the multi-head self-attention mechanism. (5) **STL-GT** is the single task model.

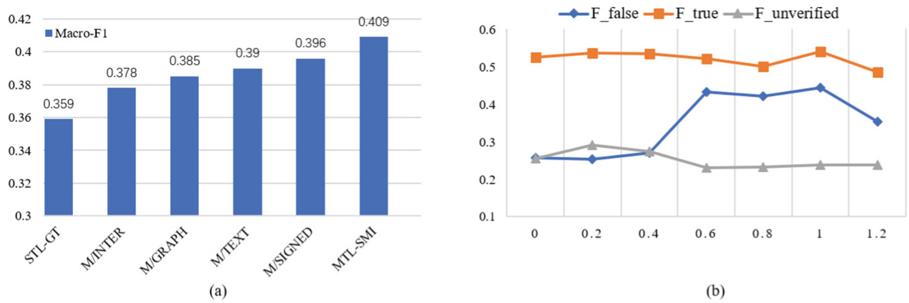

Fig. 3. (a) Results of ablation tests on SemEval dataset. (b) F_1 scores of each category under different λ values on SemEval dataset.

Performance Analysis. It can be seen from Fig. 3 (a) that: (1) Compared with the STL-GT, the MTL-SMI performs better, proving the effectiveness of multi-task learning. (2) M/GRAPH and M/TEXT, considering only one kind of interaction, still perform better than M/INTER, showing that the multi-channel interaction could benefit the performance. (3) M/TEXT and M/GRAPH perform similarly, showing that textual and network structural information are of equal importance. (4) MTL-SMI perform better than M/SIGNED, indicating that negative correlation may extract complementary semantics to help the task.

Relationship of Two Tasks. We adjust the proportion of stance classification loss by changing λ values in formula 1 to investigate the influence of SC on RV task. From Fig. 3(b), we can see that: (1) the SC task helps the RV task when $\lambda < 1$ since the F_1 scores are increasing with the increase of λ; (2) the F_1 scores show a downward trend when $\lambda > 1$. The possible reason is that the model pay more attention to the stance classification. Therefore, adjusting $\lambda = 1$ is crucial for the multi-task learning.

6 Conclusion

In this paper, we construct the conversation networks with three types of nodes and employ three graph channels to best utilize the structure information. MTL-SMI is designed to improve the accuracy of rumor verification. It has two shared channels, two task-specific graph channels and use feature interaction among channels to improve the representation of features. Results of the experiments on two datasets demonstrate the effectiveness of the MTL-SMI. Further investigation analyse the potency of key components and correlation of the two tasks.

Acknowledgements. This work was supported by the Natural Science Foundation of China (No. 61976026, No. 61902394) and 111 Project (B18008).

References

1. Kochkina, E., Liakata, M., Zubiaga, A.: All-in-one: multi-task learning for rumour verification. In: COLING, pp. 3402–3413 (2018)
2. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
3. Ma, J., Gao, W., Wong, K.: Detect rumor and stance jointly by neural multi-task learning. In: Companion of the the Web Conference 2018, pp. 585–593 (2018)
4. Li, Q., Zhang, Q., Si, L.: Rumor detection by exploiting user credibility information, attention and multi-task learning. In: ACL, pp. 1173–1179 (2019)
5. Yang, X., et al.: Rumor detection on social media with graph structured adversarial learning. In: IJCAI, pp. 1417–1423 (2020)
6. Bian, T., Xiao, X., Xu, T.: Rumor detection on social media with bi-directional graph convolutional networks. In: AAAI, pp. 549–556 (2020)
7. Wei, P., Xu, N., Mao, W.: Modeling conversation structure and temporal dynamics for jointly predicting rumor stance and veracity. In: EMNLP, pp. 4786–4797 (2019)
8. Cho, K., van Merrienboer, B.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: EMNLP, pp. 1724–1734 (2014)
9. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: ICLR, Conference Track Proceedings, OpenReview.net (2017)
10. Yu, J., et al.: Coupled hierarchical transformer for stance-aware rumor verification in social media conversations. In: EMNLP (2020)
11. Devlin, J., Chang, M., Lee, K., Toutanova, K.: BERT: pre-training of deep bidirectional transformers for language understanding. In: NAACL-HLT, pp. 4171–4186 (2019)
12. Castillo, C., Mendoza, M.: Information credibility on Twitter. In: Proceedings of the 20th International Conference on World Wide Web, pp. 675–684 (2011)
13. Ma, J., Gao, W., Wei, Z.: Detect rumors using time series of social context information on microblogging websites. In: CIKM, pp. 1751–1754 (2015)
14. Wu, K., Yang, S.: False rumors detection on Sina Weibo by propagation structures. In: 31st IEEE International Conference on Data Engineering, pp. 651–662 (2015)
15. Ma, J., Gao, W., Wong, K.: Rumor detection on Twitter with tree-structured recursive neural networks. In: ACL, Volume 1: Long Papers, pp. 1980–1989 (2018)
16. Dungs, S., Aker, A., Fuhr, N., Bontcheva, K.: Can rumour stance alone predict veracity? In: COLING, pp. 3360–3370 (2018)
17. Li, Q., Zhang, Q., Si, L.: eventAI at SemEval-2019 task 7: rumor detection on social media by exploiting content, user credibility and propagation information. In: Proceedings of the 13th International Workshop on Semantic Evaluation, pp. 855–859 (2019)
18. Vaswani, A., Shazeer, N., Parmar, N.: Attention is all you need. In: Annual Conference on Neural Information Processing Systems 2017, pp. 5998–6008 (2017)
19. Tian, T., et al.: QSAN: a quantum-probability based signed attention network for explainable false information detection. In: CIKM, pp. 1445–1454 (2020)
20. Derczynski, L., Bontcheva, K., Liakata, M.: SemEval-2017 task 8: rumoureval: determining rumour veracity and support for rumours. In: Proceedings of the 11th International Workshop on Semantic Evaluation, pp. 69–76 (2017)
21. Liu, X., Nourbakhsh, A., Li, Q.: Real-time rumor debunking on Twitter. In: CIKM, pp. 1867–1870 (2015)