Final version of the accepted paper. Cite as: *"Muhammad Abulaish, Anuj Saraswat, and Mohd Fazil, A Multi-Task Learning Framework using Graph Attention Network for User Stance and Rumor Veracity Prediction, In Proceedings of the 2023 IEEE/ ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), Kusadasi, Turkey, 6-9 November 2023, pp. 1-5. DOI: https://doi.org/10.1145/3625007.3632289"*

A Multi-Task Learning Framework using Graph Attention Network for User Stance and Rumor Veracity Prediction

Muhammad Abulaish
Department of Computer Science, South Asian University
New Delhi, India
abulaish@sau.ac.in

Anuj Saraswat
Department of Computer Science, South Asian University
New Delhi, India
anujsaraswat741@gmail.com

Mohd Fazil
Department of Information Technology, Imam Muhammad Ibn Saud Islamic University
Riyadh, Saudi Arabia
mfazil@imamu.edu.sa

ABSTRACT

In this paper, we present a multi-task learning framework consisting of two interrelated components for the joint modeling of stance classification and rumor veracity prediction on Twitter. The proposed hierarchical framework models a conversation sequence using a graph attention network, leveraging a BERT-based word representation augmented with user credibility information. The lower component of the framework models the conversation sequence of a claim in addition to the BERT-based user content representation to predict the stance of the underlying tweets. It employs a modified graph attention network, which models a conversation thread by finding tweets *path-to-root* conversation sequences. Further, the learned stance representation is augmented with the users' credibility information and content representation to predict rumor veracity. The experimental evaluation results over two benchmark datasets show that the proposed approach outperforms the state-of-the-art methods.

KEYWORDS

Multitask Learning, Stance Detection, Rumor Veracity prediction, Joint Learning

ACM Reference Format:

Muhammad Abulaish, Anuj Saraswat, and Mohd Fazil. 2023. A Multi-Task Learning Framework using Graph Attention Network for User Stance and Rumor Veracity Prediction. In *International Conference on Advances in Social Networks Analysis and Mining (ASONAM'23), Nov 6-9, 2023, Kusadasi, Turkiye*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3625007.3632289

1 INTRODUCTION

The open nature and large-scale spreading potential of Online Social Networks (OSNs) have facilitated the anti-social elements in performing mischievous activities like rumor, fake news, hate speech, and spamming generally using fake profile and socialbots [8–10]. The anti-social elements use social engineering to conduct illicit activities [1]. The user-generated content on OSNs includes plenty of unverified information.

In this study, we define rumor as *a piece of information whose truth value is still undefined* [20]. A rumor can be *false rumor*, *true rumor*, or *unverified*. A rumor is labeled as fake news if it originates from a news publishing agency. There are growing cases of real-life incidents where rumor lead to lynchings and communal riots. Therefore, identification and segregation of rumors are vital for establishing and building the trust of general users on OSN platforms. There are well-known and established debunking services like `snopes.com` and `altnews.com`, which are tirelessly working to debunk the fake news and rumors. However, these services generally follow a manual procedure, which is resource-intensive and financially infeasible. Due to the abundance of rumors on OSNs, the development of automated methods for predicting the veracity of information is vital and essential. However, most of the early approaches generally use feature engineering-based classification models for rumor veracity prediction [2, 4]. Many studies also found that user attitude embedded within the replies of a claim contains useful features to label its veracity [25]. Our proposed approach presents a two-level graph attention network-based multi-task neural network model, wherein the lower part of the framework assigns the stance label to each tweet engaged in a rumor conversation thread. Further, stance representations are aggregated and concatenated with content representation and user credibility vector for veracity prediction of the conversation thread.

An unreliable social media post ignites controversy among other users who react to it. The replies contain implicit and explicit clues, which can be leveraged to predict the claim's veracity. The stance embedded in user responses can be either *deny*, *support*, *query*, or *comment* with respect to the posted claim. The existing stance detection model use either the sequential [13, 26] or the temporal [22] organization of the conversation thread. In a multi-task neural network model, [23] exploited the conversation structure but aggregated the signals from all the neighbors of a *target tweet*. Unlike [23], our proposed approach employs the *path to root* context using the conversation sequence of a tweet rather than using all the neighboring tweets. Hence, our approach incorporates the context of a tweet using all the earlier tweets of its conversation sequence. All tweets of the sequence can not be equally

ASONAM'23, Nov 6-9, 2023, Kusadasi, Turkiye

ACM ISBN 979-8-4007-0409-3/23/11
https://doi.org/10.1145/3625007.3632289

important and should be assigned variable weights based on importance. Therefore, we model the textual content of each response in addition to their underlying conversation structure into a latent space using the graph attention network (GAT). To the best of our knowledge, the proposed approach is the first multi-task learning framework employing GAT for joint modeling of stance classification and rumor veracity prediction tasks. Unlike existing sequential- and temporal-based methods, our approach employs both content and context incorporated through conversation sequence for learning the representation of *target tweet*. Further, stance detection is vital because rumor classes have different stance distribution and evolution [17, 23].

1.1 Rumor Stance Detection and Veracity Prediction

Stance detection is the identification of a user's view on OSN posts related to a topic of interest. . The researchers also utilize the stance of user responses for veracity prediction of claims on OSNs. We group the existing stance detection approaches into two categories – *feature engineering*- and *deep learning*-based approaches. Mendoza et al. [18] analyzed the user stances when confronted with a rumor. Authors in [20] used content- and network-based features to classify users as a supporter or refuter of a claim. Likewise, [24] used linguistic features to train classifiers for stance prediction on Twitter. The authors classified a set of crisis-related tweets into *affirmative* and *denial* classes. The existing literature has several other feature engineering-based approaches using different set of features for rumor stance detection [3, 19]. Additionally, the existing literature has several approaches for stance detection using label propagation, and sequential classifiers [11, 26].

1.2 Joint Prediction of Stance and Veracity

The existing literature also has methods employing user stances as features in rumor veracity prediction [6, 16]. However, joint prediction of both stance and rumor veracity in a unified model is understudied. Ma et al. [17] presented a multi-task learning architecture to learn a unified feature representation employing task-invariant features. Additionally, the model also learns the task-specific feature representation. The model in [14] learns a shared representation of stance and rumor veracity features. It also includes task-specific layers to learn the task-specific representations. The existing literature has various multi-task learning schemes for joint modeling of stance and rumor veracity prediction tasks [12, 14, 15]. Wei et al. [23] jointly modeled the stance detection and rumor veracity prediction problem in a hierarchical fashion. Our proposed approach is also a hierarchical model, but it does not employ aggregation-based structure modeling. Rather, it uses conversation sequence-based structure modeling employing the attention mechanism to assign a variable weight to each of the contributing tweets of the sequence depending on their importance.

2 PROPOSED FRAMEWORK

Figure 1 illustrates the architecture of the proposed model, which consists of two components. The lower component learns the stance representation of each tweet of a conversation thread which is employed by the upper component for rumor veracity prediction through concatenating the user and content features. The bottom component considers structural information of the conversation thread and learns stance features using the GAT layer employing attention strategy. The output from the GAT layer is concatenated with additional content-based features including the *punctuation count*, *content similarity with the source*, and *parent tweet*, along with a set of *category-specific cue words*. The concatenated vector forms the stance feature. We use SBERT [21] to generate the representation of each tweet of the thread. Next, the encoded representation is concatenated with stance features and given to the top component, which along with the user credibility vector determines the veracity of the conversation thread.

2.1 Stance Detection

2.1.1 Conversational-GAT: Attention-aware Stance Prediction. The input to this bottom component is a conversation thread C in the form of nodes (tweets), i.e., C=$\{t_1, t_2, \ldots, t_{|N|}\}$. First, each tweet t_i: i∈$1, |C|$ of C is encoded using SBERT embedding to find its embedding representation denoted by $e_{t_i} \in \mathbb{R}^d$, where d represents the embedding dimension. Further, e_{t_i} of each tweet of C is given to the BiGRU layer to observe the encoded content-based feature representation $\mathbf{t}_i \in \mathbb{R}^{d'}$, where d' represents the dimension of BiGRU layer-based feature representation. Similarly, all the tweets of C are encoded $\mathcal{C}$=$\{\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_{|N|}\}$.

The stance feature representation of each tweet t_i is computed based on the sequence of conversation up to the tweet using a graph attention layer employing the multi-head attention strategy. For a *target tweet* t_i, we use the sequence of conversation representing the path up to the root because these tweets are informative, unlike original GAT, which learns the hidden representation based on all the neighboring tweets. Therefore, the stance feature representation of t_i is learned based on the sequential structure of its conversation sequence up to the source node.

The conversation thread C is modeled using a graph G_C = V, E, where V represents the tweets (nodes) of C and E is the edge set between tweets based on *reply* relationship. The edge set is converted into an adjacency matrix $\mathbf{A} \in \mathbb{R}^{|N| \times |N|}$ where $\mathbf{A}_{ij}$=1 Next, the adjacency matrix is converted into a *path-to-root* matrix $\mathbf{A}_p$, as given in equation 1, where PR is a function to convert an adjacency matrix into a *path-to-root* matrix $\mathbf{A}_p$. In $\mathbf{A}_p$, diagonal elements are zero, which will void the effect of the *target tweet* t_i. Therefore, to elevate the importance of t_i itself, an identity matrix is added to $\mathbf{A}_p$, as given in equation 2.

$$\mathbf{A}_p = PRA \tag{1}$$

$$\mathbf{A}_p = \mathbf{A}_p \;\; I \tag{2}$$

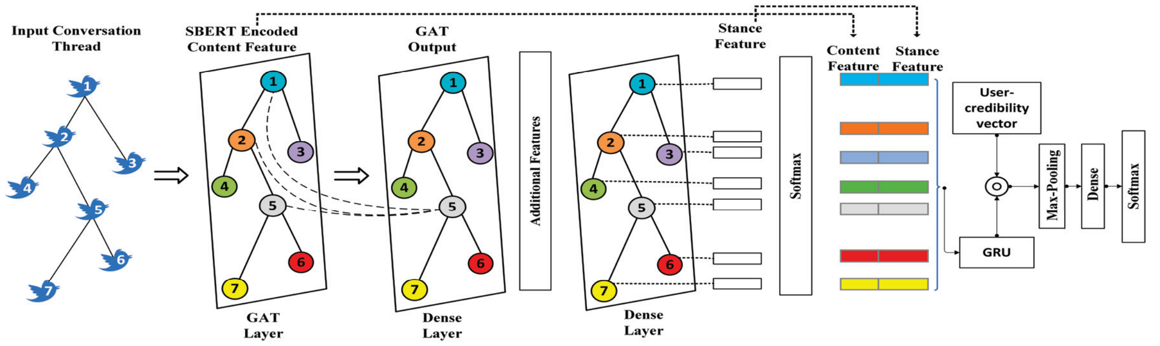

Figure 1: Architecture of the proposed framework

Finally, the BiGRU-based encoded set $\mathcal{C}=\{\mathbf{t}_1,\mathbf{t}_2,\ldots,\mathbf{t}_{|N|}\}$: $\mathbf{t}_i\in\mathbb{R}^{d'}$ of N tweets of C is given to a graph attention layer, which produces a new set of node features $\mathcal{C}' = \{\mathbf{t}'_1,\mathbf{t}'_2,\ldots,\mathbf{t}'_N\}$: $\mathbf{t}'_i\in\mathbb{R}^d$, where d represents the dimension of the newly encoded representation which is same as the initial dimension. Further, we apply a shared linear transformation using weight matrix $\mathbf{W}\in\mathbb{R}^{d\times d'}$ to all the tweets of C. Next, self-attention, a, is employed to calculate the attention coefficients, e_{ij}, using equation 3, where e_{ij} represents the importance of tweet t_j to tweet t_i after applying the a. The attention mechanism $a:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}$ comprises of a single layer feed-forward neural network employing a *Leaky Relu* activation function.

$$e_{ij} = a\{\mathbf{W}\mathbf{t}_i,\mathbf{W}\mathbf{t}_j\} \tag{3}$$

The graph attention layer uses the graph structure into the mechanism by using masked attention to compute the attention coefficients (e_{ij}) for the tweets (nodes) that are in the neighborhood of t_i in the graph. However, we exploit tweets which are on the conversation sequence *path-to-root*, as opposed to using all neighboring nodes of t_i. We calculate e_{ij} for nodes $j\in\mathcal{N}_i$, where $\mathcal{N}_i$ is the set of nodes in the path from t_i to the root node. Attention coefficients are normalized using the softmax function defined in equation 4, where $\cdot^T$ represents matrix transposition and $||$ represents concatenation operation.

$$a_{ij} = \frac{\exp\Big(LeakyReLU\big(a^T\mathbf{W}\mathbf{t}_i||\mathbf{W}\mathbf{t}_j\big)\Big)}{_{k\in\mathcal{N}_i}\exp\Big(LeakyReLU\big(a^T\mathbf{W}\mathbf{t}_i||\mathbf{W}\mathbf{t}_k\big)\Big)} \tag{4}$$

$$\mathbf{t}'_i = \sigma\Big(_{j\in\mathcal{N}_i} a_{ij}\mathbf{W}\mathbf{t}_j\Big) \tag{5}$$

The final output features are a linear combination of the attention coefficients and respective features. We have used a multi-headed attention to stabilize the process of self-attention and to perform the linear transformation described above using K independent attention mechanisms. Thereafter, features are concatenated to get the resulting output feature, as given in equation 6. Finally, at GAT layer, the output of the multi-headed attention is averaged to get the final representation of the tweets of C.

$$\mathbf{t}'_i = ||_{k=1}^{K}\sigma\Big(_{j\in\mathcal{N}_i} a_{ij}^k\mathbf{W}^k\mathbf{t}_j\Big) \tag{6}$$

$$\mathbf{t}'_i = \sigma\Big(\frac{1}{K}\,_{k=1}^{K}\,_{j\in\mathcal{N}_i} a_{ij}^k\mathbf{W}^k\mathbf{t}_j\Big) \tag{7}$$

2.1.2 Additional Features. To enrich the representation of the output feature from GAT, we concatenate some informative additional features to it. First, we calculate each tweet's *Cosine* similarity with source and previous (parent) tweets. We also concatenate the *punctuation count*, *hashtag count*, *is tweet a source tweet?* and *is tweet a reply tweet?*. We also use various rumor indicating words and count their occurrence belonging to *support*, *deny*, *query*, and *comment* categories of stances, creating four sets of signal words having 43, 52, 55, and 13 words. Next, for each tweet t_i of C, we apply softmax to the stance features to get the stance probabilities and their respective labels.

$$s_i = softmax t'_i\epsilon\ \mathbb{R}^4, i\epsilon 1, N \tag{8}$$

2.2 Rumor Veracity Prediction

2.2.1 Stance-aware GRU. This layer couples the e_t from SBERT with the stance feature learned from the bottom component, to capture the temporal evolution of stance in the conversation thread. Given a conversation thread $C=\{t_1,t_2,....,t_N\}$, we concatenate the encoded tweet feature and stance feature for each tweet and pass it as an input to a GRU layer to learn the hidden representation, as defined in equation 9.

$$h_i = GRU\big(e_{t_i}\ ||\ s_i, e_{t_{i-1}}\big), i\in 1, N, \tag{9}$$

2.2.2 User Credibility Vector. Authors in [4, 15] highlighted that credibility of users posting the content is crucial in determining the veracity of the content. Therefore, we use a set of 7 users' profile and activity-based features like *is verified* and *has a profile description*. The value of these features is normalized using Z-score normalization.

2.2.3 Rumor Verification. This integrates the user-credibility vector $\mathbf{u}=\{u_1,u_2,\ldots,u_N\}, i\in 1, N$ with the output of the

Table 1: A brief statistics of datasets

Dataset	#CT	#tweets	Rumor label statistics			Stance label statistics			
			True	False	Unverified	S	D	Q	C
RumorEval	325	5568	145	74	106	1004	415	464	3685
PHEME	1972	31430	1008	393	571	-	-	-	-

stance-aware GRU to create a rumor feature vector for veracity prediction task. Equation 10 defines it for an example tweet t_i.

$$v_i = \{h_i \mid\mid u_i\}, i \in 1, N \tag{10}$$

Further, we perform a max-pooling operation over the rumor feature vectors of C to compute a final rumor vector $v = maxpooling v_1, v_2, ..., v_N$. The final feature vector embeds the global information of stance evolution, user-credibility, and content features. Finally, v is passed to a FCC layer, followed by a softmax function for prediction.

3 EXPERIMENTAL SETUP AND RESULTS

3.1 Dataset and Evaluation Metrics

We evaluate the proposed approach over two benchmark datasets – RumorEval [5] and PHEME [27]. We use the RumorEval dataset for the evaluation of both the stance and rumor veracity prediction tasks. The second dataset PHEME has 1972 rumorous conversation threads, which are labeled as either true, false, or unverified denoting their veracity. We use only the English version of the PHEME dataset and perform leave-one-out cross validation. Table 1 presents a brief statistics of both the datasets, where S, D, Q, and C represent the *Support*, *Deny*, *Query*, and *Comment* stance categories, respectively and CT stands for conversation threads. We evaluate the proposed approach using the macro-averaged F1 score and accuracy metrics due to class imbalanced-dataset

3.2 Stance Evaluation Results

BranchLSTM:. It is an LSTM-based neural network architecture to process the sequential branches of conversation threads. The input to the LSTM layer also includes handcrafted features. The LSTM layer is followed by several dense layers and a softmax layer for output prediction.

Affective + SVM:. It extracts four categories of features including *affective*- and *dialog-act*-based features and trained SVM as a mutli-class classifier.

Temporal Attention: It is an attention aware CNN-LSTM model, which incorporates the context of a tweet using its neighboring tweets.

Conversational GCN:. In this, the stance prediction component employs a novel graph convolution operation to model the structural property.

Table 2 presents the comparision results for stance detection. It shows that our approach performs best in terms of macro-averaged F1. All the existing methods either fail or show very poor performance in predicting *deny* stance, whereas our approach shows the best performance. However,

Table 2: Performance evaluation results for stance detection over RumorEval dataset, where F_s, F_d, F_q, and F_c denote the F1 score of *support*, *deny*, *query*, and *comment* stance categories, respectively. A dash represents that the value of the respective evaluation metric is zero.

Approach	Macro-F1	F_s	F_d	F_q	F_c	Accuracy
BranchLSTM [13]	0.434	0.403	0.000	0.462	0.873	0.784
Affective+SVM [19]	0.470	0.410	0.000	0.580	**0.880**	0.795
Temporal Attention [22]	0.482	-	-	-	-	**0.820**
Conversational GCN [23]	0.499	0.311	0.194	**0.646**	0.847	0.751
Proposed approach	**0.508**	**0.428**	**0.260**	0.494	0.849	0.750

Table 3: Comparative performance evaluation results for rumor veracity prediction

Approach	RumorEval		PHEME	
	Macro-F1	Accuracy	Macro-F1	Accuracy
BranchLSTM [13]	0.491	0.500	0.336	0.454
NileTMRG [7]	0.539	0.570	0.339	0.438
MTL2 [14]	0.558	0.571	0.376	0.441
MTL3 [14]	-	-	0.396	0.492
Hierarchical-PSV [23]	0.588	0.643	0.361	0.433
Shared LSTM + Attention [15]	0.638	0.606	0.418	0.483
Proposed Approach	**0.825**	**0.821**	**0.514**	**0.676**

in terms of F_q, F_c, and accuracy, comparison approaches perform better. The better performance in predicting *support* and *denying* stances is important because they are the good indicators of rumor veracity [23]. The *Conversational-GAT* component effectively encodes each tweet based on its conversational sequence along with the attention strategy. The attention aware aggregation of information from the sequence enables the learning of powerful and discriminative stance features.

3.3 Rumor Veracity Evaluation Results

NileTMRG:. It first learns a bag-of-word representation and further concatenate other features. Next, a linear SVM classifier is trained for prediction.

MTL2 (Stance + Veracity): It has a shared LSTM layer to learn a unifying representation employing the common features of all the tasks. It also has task-specific layers to learn the discriminatory feature representation.

MTL3 (Detection + Stance + Veracity): It is like MLT2, but includes one more task of rumor detection.

Hierarchical-PSV:. It uses graph convolution operation-based stance features, stance evolution, along with content representation for rumor veracity prediction.

Shared LSTM + Attention: It jointly predict stance and rumor veracity. However, it does not present results for stance detection. It uses a shared LSTM layer for unifying representation and two task-specific layers.

Table 3 presents the comparative results for rumor veracity prediction. In this table, MTL3 results over RumorEval dataset is not presented because it contains only rumorous threads (no *potential rumors*). The table shows that our framework performs significantly better in comparison to state-of-the-art models. Hierarchical-PSV [23] learns the tweet representation aggregating information from all the neighbors assigning equal weight to all. On investigating the evaluation results between multi-task and single-task learning approaches, we found that jointly learning improves the performance due to the feedback from one task (independent) to the second task (dependent). The best performance by MTL3 establishes that as the number of tasks is increased, the performance improves due to the inter-relation between the tasks. Further, the only multi-task approach [15] which uses the attention mechanism shows the best performance among the comparison approaches. It establishes the efficacy of graph attention for better context representation.

4 CONCLUSION AND FUTURE WORK

In this paper, we have presented a multi-task learning framework to jointly model the prediction of user stance and rumor veracity in a conversation thread using structural, content, and temporal dynamics information. The proposed framework employs the novel GAT architecture with *path-to-root* conversation sequence of a *target tweet* to learn stance feature through a multi-head attention strategy, and uses SBERT for better context incorporation. The generated tweet representation along with the stance dynamics is augmented with the users' credibility information and content-based features for rumor veracity prediction. The experimental evaluation results over two benchmark datasets show that the proposed framework outperforms the existing state-of-the-art approaches for stance and rumor veracity prediction. Enhancement of the proposed framework with structural features and its evaluation over larger datasets seems one of the future directions of research.

REFERENCES

[1] Muhammad Abulaish and Mohd Fazil. 2020. Socialbots: Impacts, threat-dimensions, and defense challenges. *IEEE Technology and Society Magazine* 39, 3 (2020), 52–61.

[2] Muhammad Abulaish, Nikita Kumari, Mohd Fazil, and Basanta Singh. 2019. A Graph-Theoretic Embedding-Based Approach for Rumor Detection in Twitter. In *Proceedings of the 18th Int'l Conf. on Web Intelligence*. Thessaloniki, Greece, 466–470.

[3] Ahmet Aker, Leon Derczynski, and Kalina Bontcheva. 2017. Simple Open Stance Classification for Rumour Analysis. In *Proceedings of the Int'l Conf. on Recent Advances in Natural Language Processing*. Varna, Bulgaria, 31–39.

[4] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on Twitter. In *Proceedings of the Int'l Conf. on WWW*. Hyderabad India, 675–684.

[5] Leon Derczynski, Kalina Bontcheva, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Arkaitz Zubiaga. 2017. Semeval-2017 task 8: Rumoureval: Determining Rumour Veracity and Support for Rumours. In *Proceedings of the 11th Int'l Workshop on Semantic Evaluation*. Vancouver, Canada, 69–76.

[6] Sebastian Dungs, Ahmet Aker, Norbert Fuhr, and Kalina Bontcheva. 2018. Can Rumour Stance Alone Predict Veracity?. In *Proceedings of the Int'l Conf. on Computational Linguistics*. New-Mexico, USA, 3360–3370.

[7] Omar Enayet and Samhaa R El-Beltagy. 2017. NileTMRG at SemEval-2017 Task 8: Determining Rumour and Veracity Support for Rumours on Twitter. In *Proceedings of the 11th Int'l Workshop on Semantic Evaluations*. Vancouver, Canada, 470–474.

[8] Mohd Fazil and Muhammad Abulaish. 2017. Why a Socialbot is Effective in Twitter? A Statistical Insight. In *Proceedings of the 9th COMSNETS Social Networking Workshop*. Bengaluru, India, 562–567.

[9] Mohd Fazil, Amit Kumar Sah, and Muhammad Abulaish. 2021. Deepsbd: a deep neural network model with attention mechanism for socialbot detection. *IEEE Transactions on Information Forensics and Security* 16, 8 (2021), 4211–4223.

[10] Michael Fire, Dima Kagan, Aviad Elyashar, and Yuval Elovici. 2014. Friend or Foe? Fake Profile Identification in Online Social Networks. *SNAM* 194, 4 (2014), 1–23.

[11] Georgios Giasemidis, Nikolaos Kaplis, Ioannis Agrafiotis, and Jason R Nurse. 2020. A Semi-Supervised Approach to Message Stance Classification. *IEEE TKDE* 32, 1 (2020), 1–11.

[12] Mohammad Raihanul Islam, Sathappan Muthiah, and Naren Ramakrishnan. 2019. RumorSleuth: Joint Detection of Rumor Veracity and User Stance. In *Proceedings of the Int'l Conf. on ASONAM*. Vancouver, Canada, 131–136.

[13] Elena Kochkina, Maria Liakata, and Isabelle Augenstein. 2017. Turing at SemEval-2017 Task 8: Sequential Approach to Rumour Stance Classification with Branch-LSTM. In *Proceedings of the Int'l Workshop on Semantic Evaluations*. Vancouver, 475–480.

[14] Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. 2018. All-in-one: Multi-task Learning for Rumour Verification. In *Proceedings of the Int'l Conf. on Computational Linguistics*. New-Mexico, USA, 3402–3413.

[15] Quanzhi Li, Qiong Zhang, and Luo Si. 2019. Rumor Detection By Exploiting User Credibility Information, Attention and Multi-task Learning. In *Proceedings of the Annual Meeting of the ACL*. Florence, Italy, 1173–1179.

[16] Xiaomo Liu, Armineh Nourbakhsh, Quanzhi Li, Rui Fang, and Sameena Shah. 2015. Real-time Rumor Debunking on Twitter. In *Proceedings of the Int'l Conf. on Knowledge Management*. Melbourne, Australia, 1867–1870.

[17] Jing Ma, Wei Gao, and Kam-Fai Wong. 2018. Detect Rumor and Stance Jointly by Neural Multi-task Learning. In *Proceedings of the Int'l Conf. on WWW*. Lyon, France, 585–583.

[18] Marcelo Mendoza, Marcelo Mendoza, and Carlos Castillo. 2010. Twitter Under Crisis: Can we trust what we RT?. In *Proceedings of the 1st Workshop on Social Media Analytics*. ACM, Washington, DC, USA, 71–79.

[19] Endang Wahyu Pamungkas, Valerio Basile, and Viviana Patti. 2018. Stance Classification for Rumour Analysis in Twitter: Exploiting Affective Information and Conversation Structure. In *Proceedings of the Int'l Workshop on Rumours and Deception in Social Media*. 1–7.

[20] Vahed Qazvinian, Emily Rosengren, Dragomir R. Radev, and Qiaozhu Mei. 2011. Rumor has it: Identifying Misinformation in Microblogs. In *Proceedings of the Int'l Conf. on EMNLP*. Edinburgh, Scotland, 1589–1599.

[21] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the EMNLP*. Austin, USA, 3982–3992.

[22] Amir Pouran Ben Veyseh, Javid Ebrahimi, Dejing Dou, and Daniel Lowd. 2017. A Temporal Attentional Model for Rumor Stance Classification. In *Proceedings of the Int'l Conf. on Information and Knowledge Management*. Singapore, 2335–2338.

[23] Penghui Wei, Nan Xu, and Wenji Mao. 2019. Modeling Conversation Structure and Temporal Dynamics for Jointly Predicting Rumor Stance and Veracity. In *Proceedings of the Int'l Conf. on EMNLP-IJCNLP*. Hong Kong, China, 4787–4798.

[24] Li Zeng, Kate Starbird, and Emma S. Spiro. 2016. #Unconfirmed: Classifying Rumor Stance in Crisis-Related Social Media Messages. In *Proceedings of the Int'l Conf. on Web and Social Media*. Cologne, Germany, 747–750.

[25] Zhe Zhao, Paul Resnick, and Qiaozhu Mei. 2015. Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts. In *Proceedings of WWW*. Florence, Italy, 1395–1405.

[26] Arkaitz Zubiaga, Elena Kochkina, Maria Liakata, Rob Procter, Michal Lukasik, Kalina Bontcheva, Trevor Cohn, and Isabelle Augenstein. 2018. Discourse-aware Rumour Stance Classification in Social Media using Sequential Classifiers. *Information Processing & Management* 54, 11 (2018), 273–290.

[27] Arkaitz Zubiaga, Maria Liakata, and Rob Procter. 2017. Exploiting Context for Rumour Detection in Social Media. In *Proceedings of the Int'l Conf. on Social Informatics*. Oxford, 109–123.