Modeling Conversation Structure and Temporal Dynamics for Jointly Predicting Rumor Stance and Veracity

Abstract

Automatically verifying rumorous information has become an important and challenging task in natural language processing and social media analytics. Previous studies reveal that people's stances towards rumorous messages can provide indicative clues for identifying the veracity of rumors, and thus determining the stances of public reactions is a crucial preceding step for rumor veracity prediction. In this paper, we propose a hierarchical multi-task learning framework for jointly predicting rumor stance and veracity on Twitter, which consists of two components. The bottom component of our framework classifies the stances of tweets in a conversation discussing a rumor via modeling the structural property based on a novel graph convolutional network. The top component predicts the rumor veracity by exploiting the temporal dynamics of stance evolution. Experimental results on two benchmark datasets show that our method outperforms previous methods in both rumor stance classification and veracity prediction.

Introduction

Social media websites have become the main platform for users to browse information and share opinions, facilitating news dissemination greatly. However, the characteristics of social media also accelerate the rapid spread and dissemination of unverified information, i.e., rumors BIBREF0. The definition of rumor is “items of information that are unverified at the time of posting” BIBREF1. Ubiquitous false rumors bring about harmful effects, which has seriously affected public and individual lives, and caused panic in society BIBREF2, BIBREF3. Because online content is massive and debunking rumors manually is time-consuming, there is a great need for automatic methods to identify false rumors BIBREF4.

Previous studies have observed that public stances towards rumorous messages are crucial signals to detect trending rumors BIBREF5, BIBREF6 and indicate the veracity of them BIBREF7, BIBREF8, BIBREF9, BIBREF10, BIBREF11. Therefore, stance classification towards rumors is viewed as an important preceding step of rumor veracity prediction, especially in the context of Twitter conversations BIBREF12.

The state-of-the-art methods for rumor stance classification are proposed to model the sequential property BIBREF13 or the temporal property BIBREF14 of a Twitter conversation thread. In this paper, we propose a new perspective based on structural property: learning tweet representations through aggregating information from their neighboring tweets. Intuitively, a tweet's nearer neighbors in its conversation thread are more informative than farther neighbors because the replying relationships of them are closer, and their stance expressions can help classify the stance of the center tweet (e.g., in Figure FIGREF1, tweets “1”, “4” and “5” are the one-hop neighbors of the tweet “2”, and their influences on predicting the stance of “2” are larger than that of the two-hop neighbor “3”). To achieve this, we represent both tweet contents and conversation structures into a latent space using a graph convolutional network (GCN) BIBREF15, aiming to learn stance feature for each tweet by aggregating its neighbors' features. Compared with the sequential and temporal based methods, our aggregation based method leverages the intrinsic structural property in conversations to learn tweet representations.

After determining the stances of people's reactions, another challenge is how we can utilize public stances to predict rumor veracity accurately. We observe that the temporal dynamics of public stances can indicate rumor veracity. Figure FIGREF2 illustrates the stance distributions of tweets discussing $true$ rumors, $false$ rumors, and $unverified$ rumors, respectively. As we can see, $supporting$ stance dominates the inception phase of spreading. However, as time goes by, the proportion of $denying$ tweets towards $false$ rumors increases quite significantly. Meanwhile, the proportion of $querying$ tweets towards $unverified$ rumors also shows an upward trend. Based on this observation,

we propose to model the temporal dynamics of stance evolution with a recurrent neural network (RNN), capturing the crucial signals containing in stance features for effective veracity prediction.

Further, most existing methods tackle stance classification and veracity prediction separately, which is suboptimal and limits the generalization of models. As shown previously, they are two closely related tasks in which stance classification can provide indicative clues to facilitate veracity prediction. Thus, these two tasks can be jointly learned to make better use of their interrelation.

Based on the above considerations, in this paper, we propose a hierarchical multi-task learning framework for jointly predicting rumor stance and veracity, which achieves deep integration between the preceding task (stance classification) and the subsequent task (veracity prediction). The bottom component of our framework classifies the stances of tweets in a conversation discussing a rumor via aggregation-based structure modeling, and we design a novel graph convolution operation customized for conversation structures. The top component predicts rumor veracity by exploiting the temporal dynamics of stance evolution, taking both content features and stance features learned by the bottom component into account. Two components are jointly trained to utilize the interrelation between the two tasks for learning more powerful feature representations.

The contributions of this work are as follows.

$\bullet $ We propose a hierarchical framework to tackle rumor stance classification and veracity prediction jointly, exploiting both structural characteristic and temporal dynamics in rumor spreading process.

$\bullet $ We design a novel graph convolution operation customized to encode conversation structures for learning stance features. To our knowledge, we are the first to employ graph convolution for modeling

the structural property of Twitter conversations.

$\bullet $ Experimental results on two benchmark datasets verify that our hierarchical framework performs better than existing methods in both rumor stance classification and veracity prediction.

Related Work

Rumor Stance Classification Stance analysis has been widely studied in online debate forums BIBREF17, BIBREF18, and recently has attracted increasing attention in different contexts BIBREF19, BIBREF20, BIBREF21, BIBREF22. After the pioneering studies on stance classification towards rumors in social media BIBREF7, BIBREF5, BIBREF8, linguistic feature BIBREF23, BIBREF24 and point process based methods BIBREF25, BIBREF26 have been developed.

Recent work has focused on Twitter conversations discussing rumors. BIBREF12 proposed to capture the sequential property of conversations with linear-chain CRF, and also used a tree-structured CRF to consider the conversation structure as a whole. BIBREF27 developed a novel feature set that scores the level of users' confidence. BIBREF28 designed affective and dialogue-act features to cover various facets of affect. BIBREF29 proposed a semi-supervised method that propagates the stance labels on similarity graph. Beyond feature-based methods, BIBREF13 utilized an LSTM to model the sequential branches in a conversation, and their system ranked the first in SemEval-2017 task 8. BIBREF14 adopted attention to model the temporal property of a conversation and achieved the state-of-the-art performance.

Rumor Veracity Prediction Previous studies have proposed methods based on various features such as linguistics, time series and propagation structures BIBREF30, BIBREF31, BIBREF32, BIBREF33. Neural networks show the effectiveness of modeling time series BIBREF34, BIBREF35 and propagation paths BIBREF36. BIBREF37's model adopted recursive neural networks to incorporate structure information

into tweet representations and outperformed previous methods.

Some studies utilized stance labels as the input feature of veracity classifiers to improve the performance BIBREF9, BIBREF38. BIBREF39 proposed to recognize the temporal patterns of true and false rumors' stances by two hidden Markov models (HMMs). Unlike their solution, our method learns discriminative features of stance evolution with an RNN. Moreover, our method jointly predicts stance and veracity by exploiting both structural and temporal characteristics, whereas HMMs need stance labels as the input sequence of observations.

Joint Predictions of Rumor Stance and Veracity Several work has addressed the problem of jointly predicting rumor stance and veracity. These studies adopted multi-task learning to jointly train two tasks BIBREF40, BIBREF41, BIBREF42 and learned shared representations with parameter-sharing. Compared with such solutions based on “parallel” architectures, our method is deployed in a hierarchical fashion that encodes conversation structures to learn more powerful stance features by the bottom component, and models stance evolution by the top component, achieving deep integration between the two tasks' feature learning.

Problem Definition

Consider a Twitter conversation thread $\mathcal {C}$ which consists of a source tweet t_1 (originating a rumor) and a number of reply tweets $\lbrace t_2,t_3,\ldots ,t_{|\mathcal {C}|}\rbrace $ that respond t_1 directly or indirectly, and each tweet t_i ($i\in [1, |\mathcal {C}|]$) expresses its stance towards the rumor. The thread $\mathcal {C}$ is a tree structure, in which the source tweet t_1 is the root node, and the replying relationships among tweets form the edges.

This paper focuses on two tasks. The first task is rumor stance classification, aiming to determine the

stance of each tweet in $\mathcal {C}$, which belongs to $\lbrace supporting,denying,querying,commenting\rbrace $. The second task is rumor veracity prediction, with the aim of identifying the veracity of the rumor, belonging to $\lbrace true,false,unverified\rbrace $.

Proposed Method

We propose a Hierarchical multi-task learning framework for jointly Predicting rumor Stance and Veracity (named Hierarchical-PSV). Figure FIGREF4 illustrates its overall architecture that is composed of two components. The bottom component is to classify the stances of tweets in a conversation thread, which learns stance features via encoding conversation structure using a customized graph convolutional network (named Conversational-GCN). The top component is to predict the rumor's veracity, which takes the learned features from the bottom component into account and models the temporal dynamics of stance evolution with a recurrent neural network (named Stance-Aware RNN).

Proposed Method ::: Conversational-GCN: Aggregation-based Structure Modeling for Stance Prediction

Now we detail Conversational-GCN, the bottom component of our framework. We first adopt a bidirectional GRU (BGRU) BIBREF43 layer to learn the content feature for each tweet in the thread $\mathcal {C}$. For a tweet t_i ($i\in [1,|\mathcal {C}|]$), we run the BGRU over its word embedding sequence, and use the final step's hidden vector to represent the tweet. The content feature representation of t_i is denoted as $\mathbf {c}_i\in \mathbb {R}^{d}$, where d is the output size of the BGRU.

As we mentioned in Section SECREF1, the stance expressions of a tweet t_i's nearer neighbors can provide more informative signals than farther neighbors for learning t_i's stance feature. Based on the above intuition, we model the structural property of the conversation thread $\mathcal {C}$ to learn

stance feature representation for each tweet in $\mathcal {C}$. To this end, we encode structural contexts to improve tweet representations by aggregating information from neighboring tweets with a graph convolutional network (GCN) BIBREF15.

Formally, the conversation $\mathcal {C}$'s structure can be represented by a graph $\mathcal {C}_{G}=\langle \mathcal {T}, \mathcal {E} \rangle $, where $\mathcal {T}=\lbrace t_i\rbrace _{i=1}^{|\mathcal {C}|}$ denotes the node set (i.e., tweets in the conversation), and $\mathcal {E}$ denotes the edge set composed of all replying relationships among the tweets. We transform the edge set $\mathcal {E}$ to an adjacency matrix $\mathbf {A}\in \mathbb {R}^{|\mathcal {C}|\times |\mathcal {C}|}$, where $\mathbf {A}_{ij}=\mathbf {A}_{ji}=1$ if the tweet t_i directly replies the tweet t_j or $i=j$. In one GCN layer, the graph convolution operation for one tweet t_i on $\mathcal {C}_G$ is defined as:

where $\mathbf {h}_i^{\text{in}}\in \mathbb {R}^{d_{\text{in}}}$ and $\mathbf {h}_i^{\text{out}}\in \mathbb {R}^{d_{\text{out}}}$ denote the input and output feature representations of the tweet t_i respectively. The convolution filter $\mathbf {W}\in \mathbb {R}^{d_{\text{in}}\times d_{\text{out}}}$ and the bias $\mathbf {b}\in \mathbb {R}^{d_{\text{out}}}$ are shared over all tweets in a conversation. We apply symmetric normalized transformation $\hat{\mathbf {A}}={\mathbf {D}}^{-\frac{1}{2}}\mathbf {A}{\mathbf {D}}^{-\frac{1}{2}}$ to avoid the scale changing of feature representations, where ${\mathbf {D}}$ is the degree matrix of $\mathbf {A}$, and $\lbrace j\mid \hat{\mathbf {A}}_{ij}\ne 0\rbrace $ contains t_i's one-hop neighbors and t_i itself.

In this original graph convolution operation, given a tweet t_i, the receptive field for t_i contains its one-hop neighbors and t_i itself, and the aggregation level of two tweets t_i and t_j is dependent on $\hat{\mathbf {A}}_{ij}$. In the context of encoding conversation structures, we observe that such operation can be further improved for two issues. First, a tree-structured conversation may be very deep,

which means that the receptive field of a GCN layer is restricted in our case. Although we can stack multiple GCN layers to expand the receptive field, it is still difficult to handle conversations with deep structures and increases the number of parameters. Second, the normalized matrix $\hat{\mathbf {A}}$ partly weakens the importance of the tweet t_i itself. To address these issues, we design a novel graph convolution operation which is customized to encode conversation structures. Formally, it is implemented by modifying the matrix $\hat{\mathbf {A}}$ in Eq. (DISPLAY_FORM6):

where the multiplication operation expands the receptive field of a GCN layer, and adding an identity matrix elevates the importance of t_i itself.

After defining the above graph convolution operation, we adopt an L-layer GCN to model conversation structures. The l^{th} GCN layer ($l\in [1, L]$) computed over the entire conversation structure can be written as an efficient matrix operation:

where $\mathbf {H}^{(l-1)}\in \mathbb {R}^{|\mathcal {C}|\times d_{l-1}}$ and $\mathbf {H}^{(l)}\in \mathbb {R}^{|\mathcal {C}|\times d_l}$ denote the input and output features of all tweets in the conversation $\mathcal {C}$ respectively.

Specifically, the first GCN layer takes the content features of all tweets as input, i.e., $\mathbf {H}^{(0)}=(\mathbf {c}_1,\mathbf {c}_2,\ldots ,\mathbf {c}_{|\mathcal {C}|})^{\top }\in \mathbb {R}^{|\mathcal {C}|\times d}$. The output of the last GCN layer represents the stance features of all tweets in the conversation, i.e., $\mathbf {H}^{(L)}=(\mathbf {s}_1,\mathbf {s}_2,\ldots ,\mathbf {s}_{|\mathcal {C}|})^{\top }\in \mathbb {R}^{|\mathcal {C}|\times 4}$, where $\mathbf {s}_i$ is the unnormalized stance distribution of the tweet t_i.

For each tweet t_i in the conversation $\mathcal {C}$, we apply softmax to obtain its predicted stance

distribution:

The ground-truth labels of stance classification supervise the learning process of Conversational-GCN. The loss function of $\mathcal {C}$ for stance classification is computed by cross-entropy criterion:

where s_i is a one-hot vector that denotes the stance label of the tweet t_i. For batch-wise training, the objective function for a batch is the averaged cross-entropy loss of all tweets in these conversations.

In previous studies, GCNs are used to encode dependency trees BIBREF44, BIBREF45 and cross-document relations BIBREF46, BIBREF47 for downstream tasks. Our work is the first to leverage GCNs for encoding conversation structures.

Proposed Method ::: Stance-Aware RNN: Temporal Dynamics Modeling for Veracity Prediction

The top component, Stance-Aware RNN, aims to capture the temporal dynamics of stance evolution in a conversation discussing a rumor. It integrates both content features and stance features learned from the bottom Conversational-GCN to facilitate the veracity prediction of the rumor.

Specifically, given a conversation thread $\mathcal {C}=\lbrace t_1,t_2,\ldots ,t_{|\mathcal {C}|}\rbrace $ (where the tweets t_* are ordered chronologically), we combine the content feature and the stance feature for each tweet, and adopt a GRU layer to model the temporal evolution:

where $[\cdot ;\cdot]$ denotes vector concatenation, and $(\mathbf {v}_1,\mathbf {v}_2,\ldots ,\mathbf {v}_{|\mathcal {C}|})$ is the output sequence that represents the temporal feature. We then transform the sequence to a vector $\mathbf {v}$ by a max-pooling function that captures the global information of stance evolution, and feed it into a one-layer feed-forward neural network (FNN) with softmax

normalization to produce the predicted veracity distribution $\hat{\mathbf {v}}$:

The loss function of $\mathcal {C}$ for veracity prediction is also computed by cross-entropy criterion:

where v denotes the veracity label of $\mathcal {C}$.

Proposed Method ::: Jointly Learning Two Tasks

To leverage the interrelation between the preceding task (stance classification) and the subsequent task (veracity prediction), we jointly train two components in our framework. Specifically, we add two tasks' loss functions to obtain a joint loss function $\mathcal {L}$ (with a trade-off parameter $\lambda $), and optimize $\mathcal {L}$ to train our framework:

In our Hierarchical-PSV, the bottom component Conversational-GCN learns content and stance features, and the top component Stance-Aware RNN takes the learned features as input to further exploit temporal evolution for predicting rumor veracity. Our multi-task framework achieves deep integration of the feature representation learning process for the two closely related tasks.

Experiments

In this section, we first evaluate the performance of Conversational-GCN on rumor stance classification and evaluate Hierarchical-PSV on veracity prediction (Section SECREF21). We then give a detailed analysis of our proposed method (Section SECREF26).

Experiments ::: Data & Evaluation Metric

To evaluate our proposed method, we conduct experiments on two benchmark datasets.

The first is SemEval-2017 task 8 BIBREF16 dataset. It includes 325 rumorous conversation threads, and has been split into training, development and test sets. These threads cover ten events, and two events of that only appear in the test set. This dataset is used to evaluate both stance classification and veracity prediction tasks.

The second is PHEME dataset BIBREF48. It provides 2,402 conversations covering nine events. Following previous work, we conduct leave-one-event-out cross-validation: in each fold, one event's conversations are used for testing, and all the rest events are used for training. The evaluation metric on this dataset is computed after integrating the outputs of all nine folds. Note that only a subset of this dataset has stance labels, and all conversations in this subset are already contained in SemEval-2017 task 8 dataset. Thus, PHEME dataset is used to evaluate veracity prediction task.

Table TABREF19 shows the statistics of two datasets. Because of the class-imbalanced problem, we use macro-averaged F_1 as the evaluation metric for two tasks. We also report accuracy for reference.

Experiments ::: Implementation Details

In all experiments, the number of GCN layers is set to $L=2$. We list the implementation details in Appendix A.

Experiments ::: Experimental Results ::: Results: Rumor Stance Classification

Baselines We compare our Conversational-GCN with the following methods in the literature:

$\bullet $ Affective Feature + SVM BIBREF28 extracts affective and dialogue-act features for individual tweets, and then trains an SVM for classifying stances.

$\bullet $ BranchLSTM BIBREF13 is the winner of SemEval-2017 shared task 8 subtask A. It adopts an LSTM to model the sequential branches in a conversation thread. Before feeding branches into the LSTM, some additional hand-crafted features are used to enrich the tweet representations.

$\bullet $ TemporalAttention BIBREF14 is the state-of-the-art method. It uses a tweet's “neighbors in the conversation timeline” as the context, and utilizes attention to model such temporal sequence for learning the weight of each neighbor. Extra hand-crafted features are also used.

Performance Comparison Table TABREF20 shows the results of different methods for rumor stance classification. Clearly, the macro-averaged F_1 of Conversational-GCN is better than all baselines.

Especially, our method shows the effectiveness of determining $denying$ stance, while other methods can not give any correct prediction for $denying$ class (the F_{D} scores of them are equal to zero). Further, Conversational-GCN also achieves higher F_1 score for $querying$ stance (F_{Q}). Identifying $denying$ and $querying$ stances correctly is crucial for veracity prediction because they play the role of indicators for $false$ and $unverified$ rumors respectively (see Figure FIGREF2). Meanwhile, the class-imbalanced problem of data makes this a challenge. Conversational-GCN effectively encodes structural context for each tweet via aggregating information from its neighbors, learning powerful stance features without feature engineering. It is also more computationally efficient than sequential and temporal based methods. The information aggregations for all tweets in a conversation are worked in parallel and thus the running time is not sensitive to conversation's depth.

Experiments ::: Experimental Results ::: Results: Rumor Veracity Prediction

To evaluate our framework Hierarchical-PSV, we consider two groups of baselines: single-task and multi-task baselines.

Single-task Baselines In single-task setting, stance labels are not available. Only veracity labels can be used to supervise the training process.

$\bullet $ TD-RvNN BIBREF37 models the top-down tree structure using a recursive neural network for veracity classification.

$\bullet $ Hierarchical GCN-RNN is the single-task variant of our framework: we optimize $\mathcal {L}_{\rm {veracity}}$ (i.e., $\lambda =0$ in Eq. (DISPLAY_FORM16)) during training. Thus, the bottom Conversational-GCN only has indirect supervision (veracity labels) to learn stance features.

Multi-task Baselines In multi-task setting, both stance labels and veracity labels are available for training.

$\bullet $ BranchLSTM+NileTMRG BIBREF41 is a pipeline method, combining the winner systems of two subtasks in SemEval-2017 shared task 8. It first trains a BranchLSTM for stance classification, and then uses the predicted stance labels as extra features to train an SVM for veracity prediction BIBREF38.

$\bullet $ MTL2 (Veracity+Stance) BIBREF41 is a multi-task learning method that adopts BranchLSTM as the shared block across tasks. Then, each task has a task-specific output layer, and two tasks are jointly learned.

Performance Comparison Table TABREF23 shows the comparisons of different methods. By comparing

single-task methods, Hierarchical GCN-RNN performs better than TD-RvNN, which indicates that our hierarchical framework can effectively model conversation structures to learn high-quality tweet representations. The recursive operation in TD-RvNN is performed in a fixed direction and runs over all tweets, thus may not obtain enough useful information. Moreover, the training speed of Hierarchical GCN-RNN is significantly faster than TD-RvNN: in the condition of batch-wise optimization for training one step over a batch containing 32 conversations, our method takes only 0.18 seconds, while TD-RvNN takes 5.02 seconds.

Comparisons among multi-task methods show that two joint methods outperform the pipeline method (BranchLSTM+NileTMRG), indicating that jointly learning two tasks can improve the generalization through leveraging the interrelation between them. Further, compared with MTL2 which uses a “parallel” architecture to make predictions for two tasks, our Hierarchical-PSV performs better than MTL2. The hierarchical architecture is more effective to tackle the joint predictions of rumor stance and veracity, because it not only possesses the advantage of parameter-sharing but also offers deep integration of the feature representation learning process for the two tasks. Compared with Hierarchical GCN-RNN that does not use the supervision from stance classification task, Hierarchical-PSV provides a performance boost, which demonstrates that our framework benefits from the joint learning scheme.

Experiments ::: Further Analysis and Discussions

We conduct additional experiments to further demonstrate the effectiveness of our model.

Experiments ::: Further Analysis and Discussions ::: Effect of Customized Graph Convolution

To show the effect of our customized graph convolution operation (Eq. (DISPLAY_FORM7)) for modeling conversation structures, we further compare it with the original graph convolution (Eq.

(DISPLAY_FORM6), named Original-GCN) on stance classification task.

Specifically, we cluster tweets in the test set according to their depths in the conversation threads (e.g., the cluster “depth = 0” consists of all source tweets in the test set). For BranchLSTM, Original-GCN and Conversational-GCN, we report their macro-averaged F_1 on each cluster in Figure FIGREF28.

We observe that our Conversational-GCN outperforms Original-GCN and BranchLSTM significantly in most levels of depth. BranchLSTM may prefer to “shallow” tweets in a conversation because they often occur in multiple branches (e.g., in Figure FIGREF1, the tweet “2” occurs in two branches and thus it will be modeled twice). The results indicate that Conversational-GCN has advantage to identify stances of “deep” tweets in conversations.

Experiments ::: Further Analysis and Discussions ::: Ablation Tests

Effect of Stance Features To understand the importance of stance features for veracity prediction, we conduct an ablation study: we only input the content features of all tweets in a conversation to the top component RNN. It means that the RNN only models the temporal variation of tweet contents during spreading, but does not consider their stances and is not “stance-aware”. Table TABREF30 shows that “– stance features” performs poorly, and thus the temporal modeling process benefits from the indicative signals provided by stance features. Hence, combining the low-level content features and the high-level stance features is crucial to improve rumor veracity prediction.

Effect of Temporal Evolution Modeling We modify the Stance-Aware RNN by two ways: (i) we replace the GRU layer by a CNN that only captures local temporal information; (ii) we remove the GRU layer. Results in Table TABREF30 verify that replacing or removing the GRU block hurts the performance, and thus modeling the stance evolution of public reactions towards a rumorous message is indeed necessary for

effective veracity prediction.

Experiments ::: Further Analysis and Discussions ::: Interrelation of Stance and Veracity

We vary the value of $\lambda $ in the joint loss $\mathcal {L}$ and train models with various $\lambda $ to show the interrelation between stance and veracity in Figure FIGREF31. As $\lambda $ increases from 0.0 to 1.0, the performance of identifying $false$ and $unverified$ rumors generally gains. Therefore, when the supervision signal of stance classification becomes strong, the learned stance features can produce more accurate clues for predicting rumor veracity.

Experiments ::: Case Study

Figure FIGREF33 illustrates a $false$ rumor identified by our model. We can observe that the stances of reply tweets present a typical temporal pattern “$supporting\rightarrow querying\rightarrow denying$”. Our model captures such stance evolution with RNN and predicts its veracity correctly. Further, the visualization of tweets shows that the max-pooling operation catches informative tweets in the conversation. Hence, our framework can notice salience indicators of rumor veracity in the spreading process and combine them to give correct prediction.

Conclusion

We propose a hierarchical multi-task learning framework for jointly predicting rumor stance and veracity on Twitter. We design a new graph convolution operation, Conversational-GCN, to encode conversation structures for classifying stance, and then the top Stance-Aware RNN combines the learned features to model the temporal dynamics of stance evolution for veracity prediction. Experimental results verify that Conversational-GCN can handle deep conversation structures effectively, and our hierarchical framework

performs much better than existing methods. In future work, we shall explore to incorporate external context BIBREF16, BIBREF50, and extend our model to multi-lingual scenarios BIBREF51. Moreover, we shall investigate the diffusion process of rumors from social science perspective BIBREF52, draw deeper insights from there and try to incorporate them into the model design.

Acknowledgments

This work was supported in part by the National Key R&D Program of China under Grant #2016QY02D0305, NSFC Grants #71621002, #71472175, #71974187 and #71602184, and Ministry of Health of China under Grant #2017ZX10303401-002. We thank all the anonymous reviewers for their valuable comments. We also thank Qianqian Dong for her kind assistance.