

Microblog Sentiment Classification via Recurrent Random Walk Network Learning

Zhou Zhao¹, Hanqing Lu¹, Deng Cai², Xiaofei He², and Yueting Zhuang¹

¹College of Computer Science, Zhejiang University

²State Key Lab of CAD&CG, Zhejiang University

{zhaozhou, lhq110, yzhuang}@zju.edu.cn, {dengcai, xiaofeihe}@gmail.com

Abstract

Microblog Sentiment Classification (MSC) is a challenging task in microblog mining, arising in many applications such as stock price prediction and crisis management. Currently, most of the existing approaches learn the user sentiment model from their posted tweets in microblogs, which suffer from the insufficiency of discriminative tweet representation. In this paper, we consider the problem of microblog sentiment classification from the viewpoint of heterogeneous MSC network embedding. We propose a novel recurrent random walk network learning framework for the problem by exploiting both users' posted tweets and their social relations in microblogs. We then introduce the deep recurrent neural networks with random-walk layer for heterogeneous MSC network embedding, which can be trained end-to-end from the scratch. We employ the back-propagation method for training the proposed recurrent random walk network model. The extensive experiments on the large-scale public datasets from Twitter show that our method achieves better performance than other state-of-the-art solutions to the problem.

1 Introduction

Recently the microblogs have become a very popular Internet-based service that enables the users to share and exchange information and their opinions about all kinds of issues at real time. We have witnessed the popular microblogs such as Twitter and Sina Weibo. With the rapid growth of opinional microblog data, Microblog Sentiment Classification (MSC) becomes the essential problem in microblog mining [Liu, 2012], which arises in many real applications such as stock price prediction and crisis management.

The problem of microblog sentiment classification is to predict the sentiment class of users' posted tweets in microblogs, which has attracted considerable attention recently [Song *et al.*, 2015; Wu and Huang, 2016; Liu *et al.*, 2012]. The existing works mainly consider the problem as text classification task [Pak and Paroubek, 2010], which trains the user sentiment model from their past posted tweets, and

then predict the sentiment class of users' new tweets. Although the existing methods have achieved promising performance, most of them still suffer from the insufficiency of discriminative tweet representation in microblogs [Kim, 2014].

Currently, most of the existing microblog sentiment classification methods [Song *et al.*, 2015; Wu and Huang, 2016; Liu *et al.*, 2012] learn the semantic representation of users' posted tweets based on the hand-crafted feature (e.g., bag-of-words). However, the word sequence information is not fully utilized by the semantic representation of existing methods, which is critical for the understanding of users' posted tweets [Mao and Lebanon, 2006]. Since the users' posted tweets are always sequential data with variant length, it is natural to employ the deep recurrent neural networks to learn the semantic representation of the tweets [Hochreiter and Schmidhuber, 1997]. On the other hand, the existing deep neural networks based classification approaches [Glorot *et al.*, 2011; Kim, 2014] are mainly proposed for textual data, which may not be easily extended to heterogeneous networks for microblog sentiment classification.

In this paper, we consider the problem of microblog sentiment classification from the viewpoint of heterogeneous MSC network embedding. Specifically, we introduce the MSC network learning framework that exploits both the deep semantic representation of users' posted tweets and their social relations for the problem. We propose the deep recurrent neural networks with random-walk layer for learning the heterogeneous MSC network embedding, named as RRWNL, which can be trained end-to-end from the scratch. We then employ the back-propagation method for learning the proposed recurrent random walk network model. When a certain user's new tweet is posted, RRWNL can predict its sentiment class based on the trained recurrent random walk network embedding. The main contributions of this paper are as follows:

- Unlike the previous studies, we present the problem of microblog sentiment classification from the viewpoint of heterogeneous MSC network embedding. We learn the MSC network by exploiting both the deep semantic representation of users' posted tweets and their social relations.
- We propose the deep recurrent neural networks with random-walk layer for learning the proposed heterogeneous MSC network embedding. We then introduce

the back-propagation method for recurrent random walk network learning, which can be trained end-to-end from the scratch.

- We evaluate the performance of our method on the large-scale public datasets from Twitter. The extensive experiments show that our method can outperform several state-of-the-art solutions to the problem.

The rest of this paper is organized as follows. In Section 2, we introduce the problem of microblog sentiment classification from the viewpoint of the heterogeneous MSC network embedding, and propose the recurrent neural networks with random-walk layer. A variety of experimental results are presented in Section 3. We provide a brief review of the related work about sentiment classification and network embedding in Section 4. Finally, we provide some concluding remarks in Section 5.

2 Microblog Sentiment Classification via Recurrent Random Walk Network Learning

In this section, we first present the problem of microblog sentiment classification from the viewpoint of the heterogeneous MSC network embedding, and next introduce the recurrent random-walk network learning framework. We then propose the deep recurrent neural networks with random-walk layer for the heterogeneous MSC network learning, which can be trained end-to-end from the scratch.

2.1 The Problem

Before presenting the problem, we first introduce some basic notions and terminologies. Since the users’ posted tweets in microblogs are always the sequential data with variant length, we then encode the tweets into the fixed length feature vectors for deep semantic representation using recurrent neural networks [Le and Mikolov, 2014]. Given a set of users’ tweets $X = \{x_1, x_2, \dots, x_n\}$, we take the last hidden layer of neural networks as the semantic embedding of the tweets by $T = \{t_1, t_2, \dots, t_n\}$, where the vector $t_i \in R^d$ is the semantic representation of the i -th tweet x_i . We then denote the set of user sentiment model embeddings by $U = \{u_0, u_1, u_2, \dots, u_m\}$ where $u_0 \in R^{d \times c}$ is the global sentiment embedding matrix for all users, and $u_i \in R^{d \times c}$ is the user-specific embedding matrix for the i -th user. The value c is the number of sentiment classes in microblogs. Intuitively, the learning for global sentiment model embedding u_0 takes the advantage of the statistical power given by all users’ tweets, while the personalized embedding matrix u_i provides the flexibility to capture the user-specific sentiment preference. We then employ the popular softmax function $f_{u_0, u_i}(t_j)$ to convert the real-valued embedding of the j -th tweet posted by the i -th user to the conditional probabilities for sentiment classification. For notional convenience, we write the function $f_{u_0, u_i}(t_j)$ as $f_{u_i}(t_j)$ throughout the paper.

With the development of social science, it is well-established that the users in microblogs tend to catch the sentiment of their following users, which has been recognized

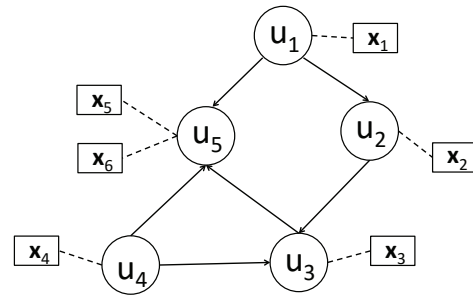


Figure 1: Heterogeneous MSC Network

as emotional contagion [Hatfield *et al.*, 1994; Zhao *et al.*, 2016a]. Thus, we propose the heterogeneous MSC network by exploiting both the users’ posted tweets and their social relations. We first denote the network connections between users in microblogs by matrix $S \in R^{m \times m}$. The entry $s_{ij} = 1$ if the i -th user is following the j -th user, otherwise, $s_{ij} = 0$. We then denote the relation between tweets and users by matrix $A \in R^{n \times m}$, where the entry $a_{ij} = 1$ if the i -th tweet is posted by the j -th user. Furthermore, we denote the set of tweets posted by the i -th user by set A_i . We now define the heterogeneous MSC network by $G = (V, E)$ where the set of nodes V is composed of tweets X and users U , and the set of edges consists of users’ network connections in S and tweet-user relations in A . We illustrate a simple example of heterogeneous MSC network modeling in Figure 1. We show the following relation between users u_1 and u_2 (i.e. $s_{12} = 1$), and tweet-user relation between tweet x_6 and user u_5 (i.e., $a_{65} = 1$) in Figure 1.

Using the notations above, we define the problem of microblog sentiment classification from the viewpoint of heterogeneous MSC network embedding as follows. Given the input users’ tweets X , and the MSC network G , our goal is to learn the personalized sentiment function $f_U(\cdot)$ for all users, and then predict the sentiment class of users’ new posted tweets.

2.2 Heterogeneous MSC Network Embedding with Recurrent Random Walk Network

In this section, we propose the recurrent random-walk network learning for the heterogeneous MSC network embedding, and then present the learning process in Figures 2(a), 2(b) and 2(c).

We first choose the proper semantic representation method for the users’ tweets in MSC networks. Given a sequence of words for the tweet x_i , we represent the t -th word by pre-training word embedding [Mikolov *et al.*, 2013] as x_{it} , and then use the sequence $(x_{i1}, x_{i2}, \dots, x_{ik})$ as the input of the corresponding recurrent neural network. However, simple recurrent neural network is difficult to train because of the resulting long term dependencies [Sutskever *et al.*, 2014]. Therefore, we choose the variant recurrent neural networks called long-short term memory (LSTM) [Hochreiter and Schmidhuber, 1997] to learn the tweet embedding by

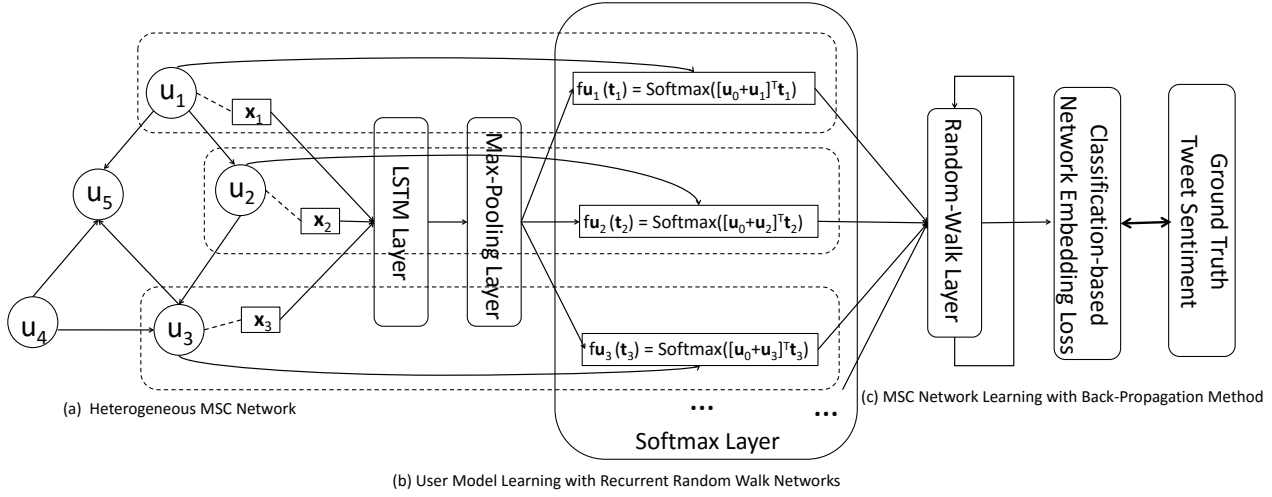


Figure 2: The framework of heterogeneous MSC network learning. (a) The heterogeneous MSC network is constructed by integrating both users' posted tweets and their network connections in microblogs. (b) The personalized user sentiment functions are trained by deep recurrent neural networks with random-walk layer. (c) The tweets and users in MSC networks are encoded into the fixed feature vectors based on the classification-based network embedding loss, which is optimized by the back-propagation method.

$$\begin{aligned}
 \mathbf{i}_t &= \delta(\mathbf{W}_i \mathbf{x}_t + \mathbf{G}_i \mathbf{h}_{t-1} + \mathbf{b}_i), \\
 \hat{\mathbf{C}}_t &= \tanh(\mathbf{W}_c \mathbf{x}_t + \mathbf{G}_f \mathbf{h}_{t-1} + \mathbf{b}_f), \\
 \mathbf{f}_t &= \delta(\mathbf{W}_f \mathbf{x}_t + \mathbf{G}_f \mathbf{h}_{t-1} + \mathbf{b}_f), \\
 \mathbf{C}_t &= \mathbf{i}_t \cdot \hat{\mathbf{C}}_t + \mathbf{f}_t \cdot \mathbf{C}_t, \\
 \mathbf{o}_t &= \delta(\mathbf{W}_o \mathbf{x}_t + \mathbf{G}_o \mathbf{h}_{t-1} + \mathbf{V}_o \mathbf{C}_t + \mathbf{b}_o) \\
 \mathbf{h}_t &= \mathbf{o}_t \cdot \tanh(\mathbf{C}_t).
 \end{aligned}$$

where δ represents the sigmoid activation function; \mathbf{W}_s , \mathbf{G}_s and \mathbf{V}_o are weight matrices, and \mathbf{b}_s are bias vectors. The architecture structure of LSTM can be found in [Hochreiter and Schmidhuber, 1997]. We train the LSTM networks for tweet embedding and then take the output of the last LSTM cell, \mathbf{h}_t , as the semantic embedding of tweets \mathbf{t} . Considering the fact that the tweets may be in the paragraph with several sentences in real microblogs, we split them into sentences for learning the semantic embedding by LSTM and then merge the embeddings by an additional max-pooling layer, illustrated in Figures 2(a) and 2(b).

We then devise the personalized user sentiment model based on softmax function. Given the semantic representation of the j -th tweet \mathbf{t}_j , the personalized sentiment function of the i -th user is given by

$$\hat{\mathbf{y}}_j = f_{\mathbf{u}_i}(\mathbf{t}_j) = \text{softmax}([\mathbf{u}_0 + \mathbf{u}_i]^T \mathbf{t}_j) = \begin{pmatrix} f_{\mathbf{u}_i,1}(\mathbf{t}_j) \\ f_{\mathbf{u}_i,2}(\mathbf{t}_j) \\ \vdots \\ f_{\mathbf{u}_i,c}(\mathbf{t}_j) \end{pmatrix},$$

where the vector $\hat{\mathbf{y}}_j \in R^c$ is the predicted sentiment class for the j -th tweet posted by the i -th user. Considering the sentiment function of the i -th user $f_{\mathbf{u}_i}(\mathbf{t}_j)$, the matrix \mathbf{u}_0 is the global sentiment embedding matrix for all users, the matrix \mathbf{u}_i is the user-specific predictive embedding matrix, and

$f_{\mathbf{u}_i,k}(\mathbf{t}_j) = \frac{\exp([\mathbf{u}_{0,k} + \mathbf{u}_{i,k}]^T \mathbf{t}_j)}{\sum_{p=1}^c \exp([\mathbf{u}_{0,p} + \mathbf{u}_{i,p}]^T \mathbf{t}_j)}$ predicts the conditional probability of tweet \mathbf{t}_j on the k -th sentiment class.

We next conduct the supervised learning for training the personalized user sentiment function, where the sentiment class of each tweet in the training data is accompanied with its ground truth. We employ the well-known cross-entropy loss term \mathcal{L}_c between the predicted sentiment probability distribution and the ground truth to learn the user sentiment function for classification by

$$\begin{aligned}
 \mathcal{L}_c &= - \sum_{i=1}^m \sum_{j \in A_i} \mathbf{y}_j^T \log \hat{\mathbf{y}}_j \\
 \text{s.t. } \hat{\mathbf{y}}_j &= f_{\mathbf{u}_i}(\mathbf{t}_j),
 \end{aligned} \tag{1}$$

where the set A_i consists of the posted tweets by the i -th user, and the vector $\mathbf{y}_j \in R^c$ is the sentiment class ground truth of the j -th tweet. Note that only the dimension of the vector \mathbf{y}_j corresponding to the ground truth is 1, with all others being 0.

We now introduce the heterogeneous MSC network learning based on both users' posted tweets and their network connections in microblogs. We first denote the relation between tweets by matrix $\mathbf{B} \in R^{n \times n}$, which can be represented by the users' network connection matrix \mathbf{S} and the tweet-user relation matrix \mathbf{A} . We then model the first-order tweet relation matrix by $\mathbf{B} = \mathbf{A} \mathbf{S} \mathbf{A}^T$. The entry $b_{ij} = 1$ if both the i -th and the j -th tweets are posted by the same user or by the two users with one-hop network connection in microblogs. We next introduce a transition matrix $\mathbf{W} \in R^{n \times n}$ between tweets. We note that the entry $b_{ij} = 1$ in tweet relation matrix \mathbf{B} , then the transition probability between them is given by $w_{ij} = \frac{1}{|A_i|}$, where $|A_i|$ is the number of tweets associated with the i -th tweet in \mathbf{B} . If there is no network connection between two tweets, then their transition

probability is set to 0. We then denote the diagonal matrix by $\mathbf{D} = \text{diag}(|A_1|, |A_2|, \dots, |A_n|)$, and thus the first-order tweet transition matrix is given by $\mathbf{W} = \mathbf{D}^{-1}\mathbf{B}$. Following the emotional contagion theory [Hatfield *et al.*, 1994], the sentiment consistency preserving of the j -th tweet can be mathematically formulated by $\hat{\mathbf{y}}_i \approx \sum_{w_{ij} > 0, w_{ij} \in \mathbf{W}} w_{ij} \hat{\mathbf{y}}_j$, where $\hat{\mathbf{y}}_i$ is the predicted sentiment of the i -th tweet, and $w_{ij} > 0$ means the i -th and the j -th tweets are connected in the tweet relation matrix \mathbf{B} . By minimizing the reconstruction error $\|\hat{\mathbf{y}}_i - \sum_{w_{ij} > 0, w_{ij} \in \mathbf{W}} w_{ij} \hat{\mathbf{y}}_j\|_F^2$, we can preserve the first-order sentiment consistency in microblogs.

Inspired by the random walk process, we propose to preserve the high-order sentiment consistency in microblogs to further improve the performance of microblog sentiment classification. Let the vector $\hat{\mathbf{y}}_j^{(k)}$ denote the predicted sentiment of the j -th tweet at time k , then its sentiment contagion at time $k+1$ can be modeled by performing one step random walk process by $\hat{\mathbf{y}}_j^{(k+1)} = \mathbf{W} \hat{\mathbf{y}}_j^{(k)} = \mathbf{W}^{(k)} \hat{\mathbf{y}}_j$, where the vector $\hat{\mathbf{y}}_j$ is the predicted sentiment of the j -th tweet at time 0. Thus, the k -th order sentiment consistency of the j -th tweet is preserved by minimizing the error $\|\hat{\mathbf{y}}_i - \sum_{w_{ij}^{(k)} > 0, w_{ij}^{(k)} \in \mathbf{W}^{(k)}} w_{ij}^{(k)} \hat{\mathbf{y}}_j\|_F^2$, where the weight $w_{ij}^{(k)}$ is from the k -th order transition matrix $\mathbf{W}^{(k)}$, and $\|\cdot\|_F^2$ is the Frobenius norm. Finally, the problem of recurrent random-walk network learning for heterogeneous MSC network embedding can be mathematically formulated by

$$\mathcal{L} = \mathcal{L}_c + \alpha \|\hat{\mathbf{y}}_i - \sum_{w_{ij}^{(k)} > 0, w_{ij}^{(k)} \in \mathbf{W}^{(k)}} w_{ij}^{(k)} \hat{\mathbf{y}}_j\|_F^2, \quad (2)$$

where the term \mathcal{L}_c is the sentiment classification loss, parameter k is the number of random-walk layer, and coefficient α is the trade-off parameter.

For training the proposed recurrent random walk network, we denote all the model coefficients including neural network parameters and the result embeddings by Θ . Therefore, the objective function in our learning process is given by

$$\min_{\Theta} \mathcal{L}(\Theta) = \mathcal{L} + \lambda \|\Theta\|^2, \quad (3)$$

where λ is the trade-off parameter between the training loss and regularization. To optimize the objective, we employ the stochastic gradient descent (SGD) with diagonal variant of AdaGrad in [Duchi *et al.*, 2011].

3 Experiments

In this section, we conduct several experiments on the Twitter datasets, and show the effectiveness of our approach RRWNL for microblog sentiment classification.

3.1 Data Preparation

We evaluate the performance of our method using two public Twitter datasets for microblog sentiment classification: Stanford Twitter Sentiment (STS) [Go *et al.*, 2009] and Obama-McCain Debate (OMD) [Shamma *et al.*, 2009], which are composed of raw tweets with their ground truth sentiment

Table 1: Experimental results on microblog sentiment classification using STS dataset with different proportions of training data.

Methods	Accuracy			
	10%	25%	50%	100%
SVMSC	0.6814	0.7203	0.7336	0.7425
CNNSC	0.7816	0.7945	0.8046	0.8105
PMSC	0.7556	0.7755	0.7849	0.7936
UPNN	0.7841	0.7970	0.8053	0.8093
SANT	0.7643	0.7782	0.7933	0.7963
LFM	0.6998	0.7345	0.7531	0.7585
ISCNN	0.7799	0.7926	0.7980	0.8096
RRWNL	0.8009	0.8122	0.8193	0.8264

Table 2: Experimental results on microblog sentiment classification using OMD dataset with different proportions of training data.

Methods	Accuracy			
	10%	25%	50%	100%
SVMSC	0.6583	0.6623	0.6965	0.7158
CNNSC	0.6985	0.7195	0.7549	0.7704
PMSC	0.6849	0.7136	0.7464	0.7551
UPNN	0.6950	0.7228	0.7553	0.7754
SANT	0.6853	0.7173	0.7483	0.7633
LFM	0.6662	0.6819	0.7064	0.7189
ISCNN	0.6908	0.7222	0.7513	0.7800
RRWNL	0.7152	0.7460	0.7798	0.7954

classes. The STS dataset contains 22,262 labeled tweets, where 11,959 tweets are labeled positive and 10,303 tweets are labeled negative. The average number of labeled tweets per user in STS is 2.63. The OMD dataset is composed of 1,827 tweets, in which 747 have positive sentiment and 1,080 have negative sentiment. The average number of labeled tweets per user in OMD is 2.49. Following the experimental setting in [Shamma *et al.*, 2009; Hu *et al.*, 2013], we take 80% of the dataset as training data, other 10% of the data for validation and the remaining 10% ones for testing. So the training and testing data do not have overlap.

3.2 Performance Comparisons

We compare our propose method with other seven state-of-the-art methods for the problem of microblog sentiment classification as follows:

- SVMSC [Pang *et al.*, 2002] is the support vector machine based model with bag-of-word representation for sentiment classification.
- CNNSC [Kim, 2014] is the convolutional neural network based model for sentiment classification.
- PMSC [Wu and Huang, 2016] is the social regularized multi-task model for sentiment classification in microblogs.
- UPNN [Tang *et al.*, 2015a] is the neural network models that incorporates both the user and product information for sentiment analysis.
- SANT [Hu *et al.*, 2013] is the supervised method with

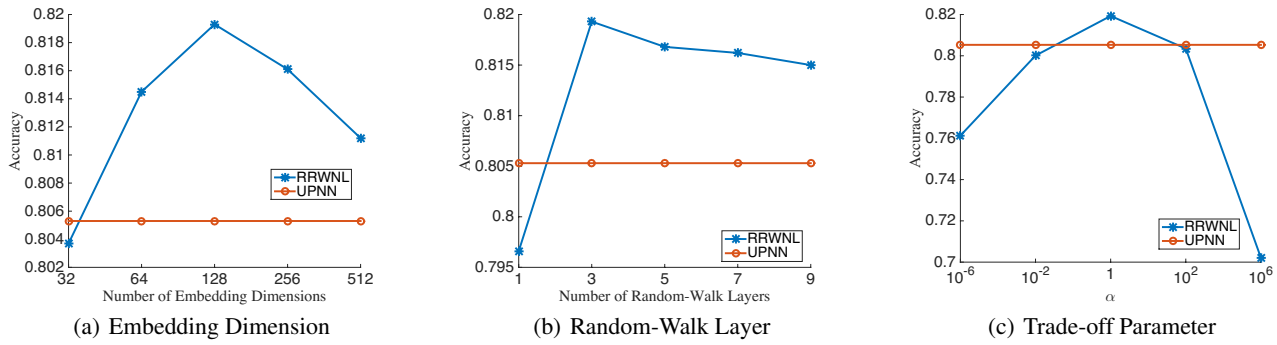


Figure 3: The effect of embedding dimensions, random-walk layers and trade-off parameter to microblog sentiment classification accuracy using STS dataset.

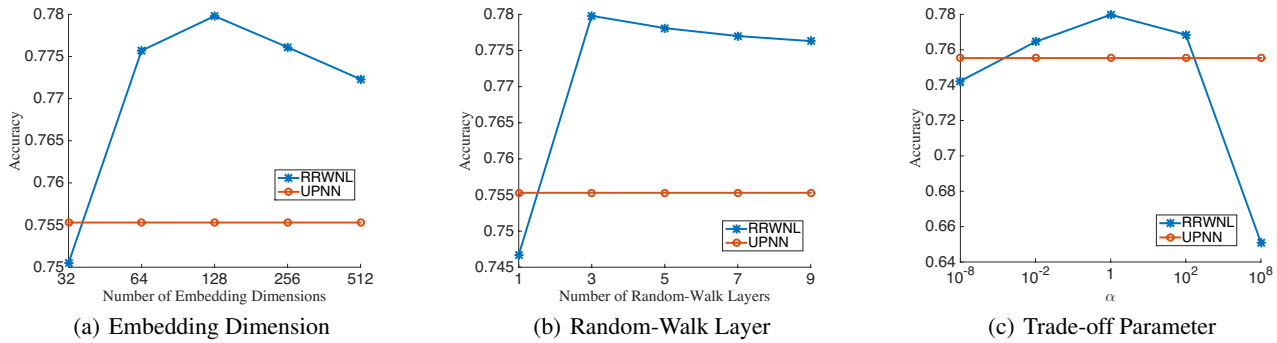


Figure 4: The effect of embedding dimensions, random-walk layers and trade-off parameter to microblog sentiment classification accuracy using OMD dataset.

users’ social relations for sentiment analysis in microblogs.

- LFM [Song *et al.*, 2015] is the latent factor model that maps users and posts into a low-dimensional factor space for sentiment analysis.
- ISCNN [Gui *et al.*, 2016] is the Intersubjectivity convolutional neural network that leverages the pre-trained writer embeddings into a convolutional neural network for sentiment analysis.

Among them, the SVMSC, PMSC, PMSC and LFM methods learn the sentiment classification model based on hand-crafted feature while the CNNSC, UPNN and ISCNN methods are the sentiment classification with neural networks. Unlike the previous studies, our method RRWNL learns the user model from the proposed MSC network, which learns the MSC network embedding for microblog sentiment classification. The input words of our methods are initialized by pre-trained word embeddings [Mikolov *et al.*, 2013] and the weights of LSTMs are randomly by a Gaussian distribution with zero mean. We utilize the recurrent random-walk network learning with LSTM models for the problem of microblog sentiment classification.

Tables 1 and 2 show the evaluation results on microblog sentiment classification accuracy using STS dataset and OMD dataset, respectively. The evaluation were conducted

with different ratio of the training data from 10%, 25%, 50% to 100%. For example, using 25% of the training data means using 16% of the whole dataset as the training data for training the model. The hyperparameters and parameters which achieve the best performance on the validation set are chosen to conduct the testing evaluation. We report the average value of all the methods on both datasets. The experiments reveal a number of interesting points:

- The deep neural networks based text classification models CNNSC, UPNN and ISCNN outperform the methods SVMSC, PMSC, SANT and LFM with hand-crafted features on classification accuracy, which suggests that the semantic representation for tweets with deep neural networks is critical for the problem of microblog sentiment classification.
- The methods PMSC and SANT achieve better performance than the SVMSC method. This suggests that leveraging the users’ network connections can also improve the performance of microblog sentiment classification.
- In all the cases, our RRWNL method achieves the best performance. This fact shows that the microblog sentiment classification network embedding framework that exploits both deep representation of users’ posted tweets on microblogs and their network connections can further

improve the performance of microblog sentiment classification.

In our approach, there are three essential parameters, which are the number of network dimensions, the number of random-walk layers and the trade-off parameter α . We vary the number of embedding dimensions from 32 to 521, the number of random-walk layers from 1 to 9, and the trade-off parameter α from 10^{-6} to 10^6 and 10^{-8} to 10^8 , respectively. We use 50% of the training data to illustrate the performance of our method using both datasets. We notice that the UPNN method consistently outperforms other methods in most of the evaluation criteria. Thus, we mainly compare our methods with UPNN by varying the number of embedding dimensions, the number of random-walk layers and the trade-off parameter α . Figures 3(a), 3(b) and 3(c) illustrate the performance of our method by varying these three parameters using STS dataset. Figures 4(a), 4(b) and 4(c) illustrate the performance of our method by varying these three parameters using OMD dataset. We observe that our method achieves the best performance when the number of embedding dimensions is set to 128, the number of random-walk layers is set to 3 and the trade-off parameter is set to 1 on both datasets. This suggests that by leveraging both the deep semantic representation of users' tweets by LSTM model, and the users' networking connections by random-walk layer, we can further improve the performance of microblog sentiment classification.

4 Related Work

In this section, we briefly review some related work on microblog sentiment classification and network learning.

The existing work for the problem of microblog sentiment classification can be categorized as graph-based approaches and neural network-based approaches. The graph-based approaches mainly leverage the users' network connections for training the sentiment classification model. Hu et al. [Hu et al., 2013] present a supervised method with the social relations for sentiment analysis. Wu et al. [Wu and Huang, 2016] propose a personalized approach for microblog sentiment classification, which capture the individual personality and the common sentiment knowledge shared by all users. Song et al. [Song et al., 2015] propose the personalized sentiment classification method based on latent factor model that captures personal sentiment variations by mapping users and posts into a low-dimensional factor space. Tang et al. [Tang et al.,] investigate the propagation-based sentiment analysis for microblogging data by iteratively alternating a propagation process and a fitting process. The neural network-based approaches mainly consider the problem as text classification and learn the unified classification model for predicting the sentiment class of all users' tweets. Gui et al. [Gui et al., 2016] propose the intersubjectivity network model that incorporates the pre-trained writer embeddings into a convolutional neural network for sentiment analysis. Tang et al. [Tang et al., 2015a] propose the sentiment analysis model that incorporates user- and product- level information into a neural network for document level sentiment classification. Unlike the previous studies, we formulate the problem from the viewpoint of microblog sentiment classification network em-

bedding, which exploiting both deep semantic representation and users' network connections to further improve the performance.

The network learning methods exploit the network structure for embedding. Perozzi et al. [Perozzi et al., 2014; Tang et al., 2015b] propose the network structure embedding method. Chang et al. [Chang et al., 2015] propose the embedding method for heterogeneous networks. Yang et al. [Yang et al.,] develop the learning method for attributed networks. Zhao et al. [Zhao et al., 2016b] propose the ranking metric embedding method for expert finding [Zhao et al., 2015]. Zhao et al. [Zhao et al., 2017] develop the multi-faceted ranking metric embedding method for community-based question answering. However, the objective of recurrent random walk network learning in our problem is different from these deep learning methods. Thus, these methods may not be suitable for our problem.

5 Conclusion

In this paper, we formulated the problem of microblog sentiment classification from the viewpoint of learning microblog sentiment classification network embedding. We propose the heterogeneous MSC network for the problem. We then introduce a novel recurrent random walk network learning framework for microblog sentiment classification by exploiting both users' posted tweets and their network connections in microblogs. We next develop the deep recurrent neural networks with random-walk layer for heterogeneous MSC network embedding, which can be trained end-to-end from the scratch. We employ the back-propagation method for training the proposed recurrent random-walk network model. We evaluate the performance our method using the well-known public datasets from Twitter. The extensive experiments demonstrate that our method can achieve better performance than several state-of-the-art solutions to the problem.

Acknowledgements

This work was supported by National Basic Research Program of China (973 Program) under Grant 2013CB336500, and National Natural Science Foundation of China under Grant 61602405, Fundamental Research Funds for the Central Universities 2016QNA5015 and the China Knowledge Centre for Engineering Sciences and Technology. The Project is also Supported by the Key Laboratory of Advanced Information Science and Network Technology of Beijing (XDXX1603).

References

- [Chang et al., 2015] Shiyu Chang, Wei Han, Jiliang Tang, Guo-Jun Qi, Charu C Aggarwal, and Thomas S Huang. Heterogeneous network embedding via deep architectures. In *SIGKDD*, pages 119–128. ACM, 2015.
- [Duchi et al., 2011] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *JMLR*, 12(Jul):2121–2159, 2011.

- [Glorot *et al.*, 2011] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Domain adaptation for large-scale sentiment classification: A deep learning approach. In *ICML*, pages 513–520, 2011.
- [Go *et al.*, 2009] Alec Go, Richa Bhayani, and Lei Huang. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1:12, 2009.
- [Gui *et al.*, 2016] Lin Gui, Ruifeng Xu, Yulan He, Qin Lu, and Zhongyu Wei. Intersubjectivity and sentiment: From language to knowledge. 2016.
- [Hatfield *et al.*, 1994] Elaine Hatfield, John T Cacioppo, and Richard L Rapson. *Emotional contagion*. Cambridge university press, 1994.
- [Hochreiter and Schmidhuber, 1997] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [Hu *et al.*, 2013] Xia Hu, Lei Tang, Jiliang Tang, and Huan Liu. Exploiting social relations for sentiment analysis in microblogging. In *WSDM*, pages 537–546. ACM, 2013.
- [Kim, 2014] Yoon Kim. Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*, 2014.
- [Le and Mikolov, 2014] Quoc V Le and Tomas Mikolov. Distributed representations of sentences and documents. In *ICML*, volume 14, pages 1188–1196, 2014.
- [Liu *et al.*, 2012] Kun-Lin Liu, Wu-Jun Li, and Minyi Guo. Emoticon smoothed language models for twitter sentiment analysis. In *AAAI*, 2012.
- [Liu, 2012] Bing Liu. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1):1–167, 2012.
- [Mao and Lebanon, 2006] Yi Mao and Guy Lebanon. Sequential models for sentiment prediction. In *ICML Workshop on Learning in Structured Output Spaces*, 2006.
- [Mikolov *et al.*, 2013] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- [Pak and Paroubek, 2010] Alexander Pak and Patrick Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. In *LREc*, volume 10, pages 1320–1326, 2010.
- [Pang *et al.*, 2002] Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. Thumbs up?: sentiment classification using machine learning techniques. In *EMNLP*, pages 79–86, 2002.
- [Perozzi *et al.*, 2014] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. Deepwalk: Online learning of social representations. In *SIGKDD*, pages 701–710. ACM, 2014.
- [Shamma *et al.*, 2009] David A Shamma, Lyndon Kennedy, and Elizabeth F Churchill. Tweet the debates: understanding community annotation of uncollected sources. In *SIGMM workshop on Social media*, pages 3–10. ACM, 2009.
- [Song *et al.*, 2015] Kaisong Song, Shi Feng, Wei Gao, Dal-ing Wang, Ge Yu, and Kam-Fai Wong. Personalized sentiment classification based on latent individuality of microblog users. In *IJCAI*, 2015.
- [Sutskever *et al.*, 2014] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *NIPS*, pages 3104–3112, 2014.
- [Tang *et al.*,] Jiliang Tang, Chikashi Nobata, Anlei Dong, Yi Chang, and Huan Liu. Propagation-based sentiment analysis for microblogging data.
- [Tang *et al.*, 2015a] Duyu Tang, Bing Qin, and Ting Liu. Learning semantic representations of users and products for document level sentiment classification. In *ACL*, 2015.
- [Tang *et al.*, 2015b] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. Line: Large-scale information network embedding. In *WWW*, pages 1067–1077. WWW, 2015.
- [Wu and Huang, 2016] Fangzhao Wu and Yongfeng Huang. Personalized microblog sentiment classification via multi-task learning. In *AAAI*, 2016.
- [Yang *et al.*,] Cheng Yang, Zhiyuan Liu, Deli Zhao, Maosong Sun, and Edward Y Chang. Network representation learning with rich text information.
- [Zhao *et al.*, 2015] Zhou Zhao, Lijun Zhang, Xiaofei He, and Wilfred Ng. Expert finding for question answering via graph regularized matrix completion. *TKDE*, 27(4):993–1004, 2015.
- [Zhao *et al.*, 2016a] Zhou Zhao, Hanqing Lu, Deng Cai, Xiaofei He, and Yueting Zhuang. User preference learning for online social recommendation. *TKDE*, 28(9):2522–2534, 2016.
- [Zhao *et al.*, 2016b] Zhou Zhao, Qifan Yang, Deng Cai, Xiaofei He, and Yueting Zhuang. Expert finding for community-based question answering via ranking metric network learning. In *IJCAI*, pages 3000–3006. AAAI Press, 2016.
- [Zhao *et al.*, 2017] Zhou Zhao, Hanqing Lu, Vincent W Zheng, Deng Cai, Xiaofei He, and Yueting Zhuang. Community-based question answering via asymmetric multi-faceted ranking network learning. In *AAAI*. AAAI Press, 2017.