Neurocomputing 464 (2021) 141–150

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Exploring user historical semantic and sentiment preference for microblog sentiment classification

Xiaofei Zhu [a,*], Jie Wu [a,b], Ling Zhu [a], Jiafeng Guo [c], Ran Yu [d], Katarina Boland [e,f], Stefan Dietze [e,f]

[a] College of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400054, China
[b] Baidu Inc., Beijing 100085, China
[c] Institute of Computing Technology, Chinese Academy of Science, Beijing 100190, China
[d] Institute of Computer Science III, University of Bonn, Bonn 53115, Germany
[e] Knowledge Technologies for the Social Sciences, Leibniz Institute for the Social Sciences, Cologne 50667, Germany
[f] Institute of Computer Science, Heinrich-Heine-University, Düsseldorf 40225, Germany

ARTICLE INFO

Article history:
Received 6 February 2021
Revised 2 May 2021
Accepted 20 August 2021
Available online 25 August 2021
Communicated by Zidong Wang

Keywords:
Microblog analysis
Sentiment classification
User historical preference

ABSTRACT

Microblog text is usually very short, thereby challenging existing sentiment classification methods by providing models with little context. Recently, historical user information has been widely used in many real-world applications, such as recommender systems. However, few research works consider user historical states in the loop of microblog sentiment analysis. In this work, we propose to involve historical user information for microblog sentiment analysis to alleviate the context sparsity problem. In particular, we propose a novel neural microblog sentiment classification method which learns informative representations of microblog posts by exploiting both a user's contextual information and his/her historical state information. The proposed method consists of four components, i.e., a micropost encoder, a user historical sentiment encoder, a User Historical Semantic Encoder, and a micropost sentiment classification component. Extensive experiments are conducted on real-world data collected from Weibo, and experimental results show that the proposed approach achieves superior performance as compared to state-of-the-art baselines.

1. Introduction

Sentiment analysis (also known as opinion mining) aims at classifying people's attitudes towards some topics or the overall polarity to a document. Due to the rise of social media such as blogs and social networks, significant research efforts have been devoted to sentiment analysis. With the proliferation of online reviews, ratings, and recommendations, mining user opinion has turned into a kind of virtual currency for many organisations looking to market their products, identifying new opportunities and managing their reputations. How to utilize machine learning technology to analyze the opinions of microblog posts has become one of the hotspots in the field of natural language research [1], and attracted considerable attention [2–5] from researchers in the past decade.

Traditional sentiment classification technology is primarily based on exploiting sentiment lexicons or leveraging feature extraction techniques. Methods based on sentiment lexicons [6,7] treat sentences as a combination of words, obtain statistical features based on expert generated opinions, emotional dictionaries and template rules, and then perform a multi-granular combination calculation of words in the text to conduct textual sentiment analysis. A shortcoming of these methods is that they heavily rely on user intervention, and the classification results obtained are sub-optimal [8]. Feature extraction based methods [9,10] usually construct feature vectors by extracting the feature information implied in the text, and then learn from the training set by using traditional methods, such as support vector machine [11], logistic regression [9], Naive Bayes [12] and etc. Although these methods tend to outperform lexicon-based approaches, the performance is still unsatisfactory.

Recently, deep learning-based methods have been studied extensively due to their excellent performance on many different tasks. It does not rely too much on feature extraction, and can exploit the feature information of the text through a deep neural network model. Most existing studies build sentiment classifiers by using deep neural networks, such as convolutional neural network (CNN) [13], recurrent neural network (RNN) [14], and attention-based methods [15–17]. Very recently, some research

* Corresponding author.
E-mail address: zxf@cqut.edu.cn (X. Zhu).

https://doi.org/10.1016/j.neucom.2021.08.089
0925-2312/

works also propose to exploit sentiment linguistic knowledge into deep neural networks via a multi-path attention mechanism [18]. Despite the promising progress achieved by previous works, how to effectively classify sentiment in micropost[1] text still remains an open research question. The main challenges of microblog sentiment classifications are that microposts are usually very short, and contain lots of noise, acronyms as well as informal words.

To address the aforementioned issues, in this paper, we leverage historical information about a user's prior posts as prior knowledge to enhance the representation learning of microposts, and propose a historical user states-enhanced microblog sentiment classification model, named as UHSE. In particular, we first design a micropost encoder, which is a stacked transformer capturing the information of the micropost content as well as word sentiment signals to learn a micropost representation. Then, we propose two encoders, i.e., historical user sentiment encoder and historical user semantic encoder. In the former, we attempt to capture user sentiment states by aggregating her historical records and sentiment information. Since user information preference and sentiment state would vary across different topics and evolve over time, we further exploit a time-aware attention mechanism to address the above issue. In the historical user semantic encoder, we aggregate historical semantic information of a user. Similar to the historical user sentiment encoder, the time dimension is also introduced into the encoder. After that we concatenate current text representation, its corresponding historical user semantic representation and historical user sentiment representation. At last, a softmax layer is employed for sentiment classification.

As there are few microblog sentiment datasets which have rich user contextual information to the scenario mentioned in this work, we create a dataset with user's contextual information by collecting data from Sina Weibo.[2] We conducted extensive experiments on this dataset, and results demonstrate that the proposed approach can effectively model a user's historical states across both semantic and sentiment preferences, and consistently outperforms the state-of-the-art baseline methods with a large margin. The main contributions of our work can be summarized as follows:

- We propose to involve both historical user semantic and sentiment preferences into the loop of microblog sentiment analysis to alleviate the sparsity issue.
- We propose a unified model which consists of a micropost encoder to learn representative micropost representations, a user sentiment state encoder to capture a user's historical emotion states and a user semantic state encoder to capture a user's historical semantic states.
- We create a new dataset with user's contextual information by collecting data from Sina Weibo, and make the dataset available athttps://github.com/zhucqut/UHSE to the research community.

2. Related work

Microblog sentiment classification has been a hot research topic in recent years. Previous work for microblog sentiment classification mainly focused on how to effectively model micropost text and external information. Some researchers propose to incorporate sentiment dictionaries to determine the sentiment of a text. As sentiment lexicons contain opinion words and sentiment phrases, it plays an important role in sentiment analysis tasks. For example, Kong et al. [19] create a microblog-specific sentiment lexicon from a massive microblog data, and build a neural architecture to train a sentiment-aware word embedding. Lei et al. [18] propose a multi-emotional resource enhancement attention network (MEAN), where three kinds of emotional language knowledges (i.e., emotional vocabulary, negative words, and word strength) are integrated into a deep neural network through the attention mechanism. Zhao et al. [20] propose a sentiment unit context propagation framework to extract task-specific explicit and implicit sentiment features. They mark a set of seed sentiment units with sentiment labels using general sentiment lexicons, and then conduct sentiment label propagation from seed sentiment units to unlabeled ones. Ito et al. [21] develop a sentiment interpretable neural network. They propose a novel learning strategy called lexical initialization learning and extract word-level contextual sentiment through extracting word-level original sentiment and its local and global word-level contexts.

Some recent researches are devoted to utilize emoji signals as weak sentiment labels to deal with the label scarcity issue. For example, Eisner et al. [22] estimate the representation of emojis by averaging the words from their description. The learnt embedding of emoji could be used in downstream tasks such as Twitter sentiment analysis. Chen et al. [23] propose to learn bi-sense emoji embeddings under positive and negative sentimental texts, and then train a sentiment classifier by attending an attention-based LSTM network on these bi-sense emoji embeddings. Chen et al. [24] learn representation for cross-lingual sentiment classification by employ emoji prediction as an instrument to learn respective representation for each language.

Very recently, several works propose to leverage context information for enhancing the representation of microblog. For example, Wu et al. [25] propose to boost microblog sentiment classification performance via combining social context information with textual content. Wang et al. [26] study the problem of sentiment spreading in social networks. Specifically, they explore the correlation between users' sentimental statuses and topic distributions embedded in the tweets, and then learn the influence strength between linked users. Zheng et al. [27] identify crucial contextual information with the help of syntactic structure and then perform a replicated random walk on a syntax graph to effectively focus on the informative contextual words. They exploit two kinds of social texts (i.e., social connections between micropost messages and social connections between users), and formulate the social context information as the graph structure over the sentiments of micropost messages. Feng et al. [28] regard microblogs as conversation streams with fragmented sentiment expression, and propose to integrate attention mechanism into hierarchical LSTM network models to classify context-aware sentiments in microblogs. They build a highrarchical LSTM network to generate tweet-level and conversation-level representation, and then incorporate an attention mechanism over conversation sequence. Pong et al. [29] propose a novel dual-view model for sentiment classification as well as summarization, and they introduce an inconsistency loss to jointly improve the performance of text summarization and sentiment classification by encouraging the sentiment information in the decoder states to be close to that in the text context representation.

Although great success has been achieved, existing methods ignore the user historical states (e.g., whether users are optimistic or pessimistic), which is essential when the current text is short and ambiguous. A significant difference between our approach and traditional methods is that we incorporate user historical emotion states into our model and propose a unified framework to simultaneously capture contextual information of text and user historical preference information.

[1] We refer to an individual microblog post of a user as a micropost.
[2] https://www.weibo.com

3. Our approach

In this section, we introduce the proposed User Historical States Enhanced Microblog Sentiment Classification Model (USHE), which consists of four components: a Microblog Encoder, a Historical User Sentiment Encoder, a Historical User Semantic Encoder, and a Microblog Sentiment Classification component. Next, we detail all components sequentially from bottom to top. Fig. 1 demonstrates the architecture of our proposed model.

3.1. Micropost encoder

The micropost encoder aims to learn the representation of a micropost from its textual content, which consists of five layers. Fig. 2 shows the framework of the micropost encoder. The first layer is a BERT embedding layer, which is used to obtain word embeddings from a pre-trained BERT model (BERT-Base) [30]. We denote the word sequence of a micropost text as $(w_1, w_2, \dots, w_L)$, where L is the length of the sequence, the BERT embedding layer is used to convert the text into a sequence of word embeddings $(e_1, e_2, \dots, e_L)$ via a BERT-based word embedding look-up table $W_e \in \mathbb{R}^{V \times d}$, where V and d represent vocabulary size and word embedding dimension, respectively.

The second layer is a concatenation layer which concatenates the word embedding and its corresponding sentiment score. This layer is used to enhance the word representations through combination with sentiment information and semantic information. Note that there are many ways to define the sentiment score. In this work, we compute the sentiment score sw_i of a word token w_i based on its occurrence in a microblog set with different sentiment polarity [31], which is formalized as follows:

$$Freq(w_i) = |\alpha \cdot pos(w_i) - \beta \cdot neg(w_i)| \tag{1}$$

$$Freq_{min} = \min_{w_j \in V} Freq(w_j) \tag{2}$$

$$Freq_{max} = \max_{w_j \in V} Freq(w_j) \tag{3}$$

$$sw_i = \lfloor \frac{Freq(w_i) - Freq_{min}}{Freq_{max} - Freq_{min}} \gamma \rfloor \tag{4}$$

where $pos(w_i)$ and $neg(w_i)$ represent the frequency of word w_i in positive documents and negative documents, respectively. $|\cdot|$ represents the absolute value and $\lfloor \cdot \rfloor$ represents the round down operation, α, β and γ are parameters. If w_i is not a sentiment word (not in the sentiment dictionary), then the sentiment score of the word is $Freq(w_i)$, otherwise the sentiment score of the word is sw_i. The output of the concatenation layer $\left(h_1^{cat}, h_2^{cat}, \dots, h_L^{cat}\right)$ with $h_i^{cat} = [e_i; sw_i]$, where $[\cdot;\cdot]$ denotes a concatenation operation.

The third layer is a BiLSTM layer which is used to incorporate the contextual information of the micropost text into the representation of each word. Specifically, we feed $\left(h_1^{cat}, h_2^{cat}, \dots, h_L^{cat}\right)$ to a BiLSTM to learn hidden representation $(h_1, h_2, \dots, h_L)$. A BiLSTM consists of a forward LSTM which reads from $\left(h_1^{cat}\text{ to } h_L^{cat}\right)$ and backward LSTM which reads from $\left(h_L^{cat}\text{ to } h_1^{cat}\right)$:

$$\overrightarrow{h}_i = LSTM\left(h_i^{cat}, \overrightarrow{h}_{i-1}\right) \tag{5}$$

$$\overleftarrow{h}_i = LSTM\left(h_i^{cat}, \overleftarrow{h}_{i-1}\right) \tag{6}$$

where $\overrightarrow{h}_i \in \mathbb{R}^{d/2}$ and $\overleftarrow{h}_i \in \mathbb{R}^{d/2}$ denote the hidden states of the forward LSTM and backward LSTM, respectively. Then we concatenate $\overrightarrow{h}_i$ and $\overleftarrow{h}_i$ to form the hidden representation h_i, i.e., $h_i = \left[\overrightarrow{h}_i; \overleftarrow{h}_i\right]$.

Next, the output $(h_1, h_2, \dots, h_L)$ will be fed into the fourth layer, a Multi-Head Self-Attention layer [17], which can attend to information from different representation subspaces. In particular, for the i_{th} attention head, we compute the representation vector a_j^i of the j_{th} word as follows:

$$\widehat{\alpha}_{j,k}^i = h_j^T W_i h_k, \tag{7}$$

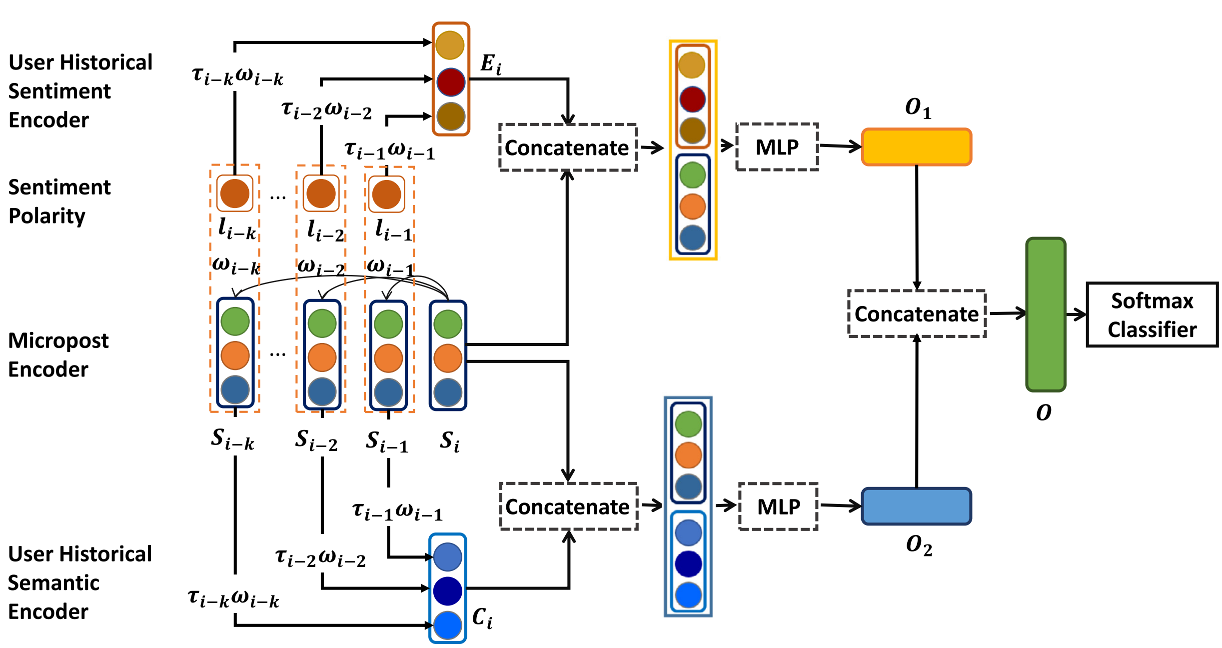


Fig. 1. The Architecture of User Historical States Enhanced Microblog Sentiment Classification (USHE) Model.

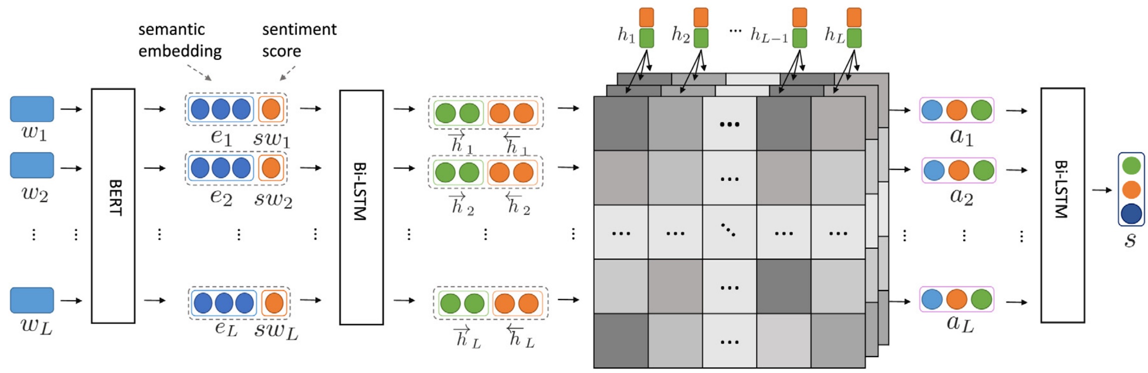


Fig. 2. Framework of the Micropost Encoder. The Micropost Encoder obtains an effective representation for a micropost based on modeling its textual content.

$$\alpha^i_{j,k} = \frac{exp\left(\widehat{\alpha}^i_{j,k}\right)}{\sum_{l=1}^{L} exp\left(\widehat{\alpha}^i_{j,l}\right)}, \tag{8}$$

$$a^i_j = V_i\left(\sum_{k=1}^{L}\alpha^i_{j,k}h_k\right), \tag{9}$$

where W_i and V_i are parameters corresponding to the i_{th} self-attention head. Then, the representation of word a_j is the concatenation of the representation vectors produced by all h self-attention heads, i.e., $a_j = \left[a^1_j, a^2_j, \ldots, a^h_j\right]$.

On top of the Multi-Head Self-Attention layer, we add another BiLSTM layer to model more complex word interactions within the micropost text. Finally, we use the hidden state of the last word token as the final representation S of the micropost.

3.2. User historical sentiment encoder

In the user historical sentiment encoder, we attempt to capture user sentiment states by aggregating her historical microposts. As user's information preference and sentiment state would vary across different topics and evolve over time, we propose a novel time-aware attention mechanism to address the above issue.

We represent user historical sentiment states by the sentiment labels of her preceding microposts. As it is impractical in real-world applications to generate sentiment labels, we propose to leverage weak labels as a surrogate. In this work, we utilize the prediction probability of a variant of Bert [30] model, named BERT-sentiment words (BASE) in this paper, as the weak label indicator. More details about this model is given in Section 4.6. Our preliminary experimental results show that utilizing weak label-based user historical sentiment states can achieve encouraging performance.

We denote the weak labels of a user historical sentiment states as $(l_{i-k}, \ldots, l_{i-1})$, and we define the user historical sentiment encoder as follows:

$$E_i = [\tau_{i-k}\omega_{i-k}l_{i-k}; \ldots; \tau_{i-1}\omega_{i-1}l_{i-1}] \tag{10}$$

where k represents the number of historical microposts, τ and ω represent the time weight and semantic weight, respectively. The time weights are used to model the influence of the time. Given a user, her micropost sequence is $(s_1, \ldots, s_i)$, where s_i denotes the current micropost, the time weights of the previous microposts of s_i are defined as follows:

$$\tau_p = \tau(s_p, s_i) = N_0 e^{-\mu(t+m)} \tag{11}$$

where $s_p$represents the p-th historical micropost, $i-k \leqslant p \leqslant i-1$. N_0 represents an initial value and t represents the time interval. m represents the leftward translation amount, which allows the value to not decay from N_0, but continues to decay from any position. Suppose that we need to decay from N_{init} to N_{finish} after z time units, then parameters μ and m can be estimated as follows:

$$\mu = \frac{1}{z}\ln\left(\frac{N_{init}}{N_{finish}}\right) \tag{12}$$

$$m = \frac{1}{\mu}\ln\left(\frac{N_0}{N_{init}}\right) \tag{13}$$

The semantic weight is used to model the semantic relationships between microposts. Specifically, the value of semantic weight ω_p of historical information is the normalized similarity between the historical micropost s_p and the current micropost s_i, which is defined as follows:

$$f(s_p, s_i) = \frac{S_p \cdot S_i}{||S_p|| \times ||S_i||} \tag{14}$$

$$\omega_p = \frac{exp(f(s_p, s_i))}{\sum_j exp(f(s_p, s_j))} \tag{15}$$

where S_i and S_p are the corresponding representations of s_i and s_p, respectively. These corresponding representations of microposts are obtained with the Microblog Encoder.

3.3. User Historical Semantic Encoder

In the User Historical Semantic Encoder, we aggregate semantic representations of a user's historical microposts. Similar to the user historical sentiment encoder, we introduce the time dimension into the encoder and define it as follows:

$$C_i = \sum_{p=i-k}^{i-1}(\tau_p\omega_p S_p) \tag{16}$$

where k represents the number of historical microposts, S_p represents the representation of the p-th historical micropost. τ_p and ω_p represent its corresponding time weight and semantic weight, respectively.

3.4. Microblog sentiment classification

The microblog sentiment classification component is used to classify the sentiment polarity of a micropost. It first applies MLP to transform the hidden representation as follows:

$$O^1_i = MLP([C_i; S_i]) \tag{17}$$

$$O^2_i = MLP([E_i; S_i]) \tag{18}$$

Then, the representation of a micropost is represented as the concatenation of O_1 and O_2, formulated as:

$$O_i = \left[O_i^1; O_i^2\right] \tag{19}$$

Finally, we utilize a softmax layer to predict the sentiment label distribution of the micropost.

$$\hat{y}_i = Softmax\left(W_o^T O_i + b_o\right) \tag{20}$$

where $\hat{y}_i$ is the output vector of the model, M represents the number of labels, W_o and b_o are trainable parameters.

In the training of the model, our goal is to minimize the cross entropy between the true label of microposts and their predicted results. The loss function of the model is defined as follows:

$$L(y,\hat{y}) = -\sum_{i=1}^{N}\sum_{c=1}^{M} y_i^c \log(\hat{y}_i^c) + \lambda\left(\sum \theta \in \Theta\theta^2\right) \tag{21}$$

where y_i represents the true label of the i-th micropost, which is a one-hot vector, i.e., $y_i^c = 1$ if the c-th item of y_i is the target class. θ represents a regularization of parameters, Θ denotes the parameter set, and λ is a regularization coefficient of L2. The detailed learning procedure of the proposed UHSE is shown in Algorithm 1.

Algorithm 1: User Historical States Enhanced Microblog Sentiment Classification Model.

Input A user's current micropost s_i; Her corresponding historical microposts $(s_{i-k}, \ldots, s_{i-1})$; Weak sentiment labels of these historical microposts $(l_{i-k}, \ldots, l_{i-1})$; Time intervals of these historical microposts to the current micropost $(t_{i-k}, \ldots, t_{i-1})$.

Output The predicted sentiment label $\hat{y}_i$ of the current micropost.

1: Learn the representation S_i of the current micropost s_i and the corresponding representations $(S_{i-k}, \ldots, S_{i-1})$ of the historical microposts $(s_{i-k}, \ldots, s_{i-1})$ with the Micropost Encoder;

2: Compute the time weights $(\tau_{i-k}, \ldots, \tau_{i-1})$ of these historical microposts based on the time intervals $(t_{i-k}, \ldots, t_{i-1})$;

3: Compute the semantic weights $(\omega_{i-k}, \ldots, \omega_{i-1})$ of these historical microposts based on the normalized cosine similarity between each historical micropost and the current micropost;

4: Obtain the corresponding historical semantic representation C_i for the current micropost s_i with the User Historical Semantic Encoder;

5: Obtain the corresponding historical sentiment representation E_i for the current micropost s_i with the User Historical Sentiment Encoder;

6: Both C_i and E_i are merged with S_i using concatenation and MLP, respectively. The outputs O_i^1 and O_i^2 are further concatenated and fed into a softmax layer to predict sentiment label $\hat{y}_i$ of the current micropost;

7: **return** $\hat{y}_i$

4. Experiments

4.1. Experimental setting

In our experiments, we use a 768-dim pre-trained word embedding [30], and the Bi-LSTM network has 2×100 units. We use 200 heads in the multi-head self-attention network, and set the dropout rate and epoch number to 0.5 and 50, respectively. The initial attenuation value N_{init} and the completion attenuation value N_{finish} are empirically set to 0.2 and 0.8, respectively. We set the parameter k which is the number of historical microposts to 5, and set the parameter z which is the time decay length for controlling time weights to 30. More discussion about the influence of the parameters k and z will be given in Section 4.8.

4.2. Dataset

The sentiment dictionary we used in this work consists of two parts. The first part includes both Hownet[3] and NTUSD's Chinese sentiment dictionary,[4] and the second part contains a manually added set of common emotional terminology in microposts. At last, we obtain a sentiment dictionary with 9231 positive emotional words and 14022 negative emotional words.

As far as we know, there are few microblog sentiment datasets, which have rich user context information to the scenario we study in this work, such as user's historical posted contents as well as time-stamps for each of her micropost. To this end, we create a dataset with user's context information by collecting data from Sina Weibo.[5] Specifically, we select a set of candidate users with the number of followers between 100 and 20,000, which are considered as general users. For each user, we crawl her user id, microposts and time-stamp of each micropost from January 1, 2016 to October 31, 2018, which results in 638,859 microposts. Then, we process the data by removing microposts which are considered as retweets or the main content consists of an image or advertisements.

After that we randomly selected a set of users who have more than 100 processed microposts, and manually label all of their microposts as positive, negative and neutral. In particular, we annotate the label of each micropost by two annotators from computer science background. A label is assigned to a micropost when both of them give the same label to this micropost. Otherwise, another annotator will be involved to annotate the micropost. In this work, we focus on binary sentiment classification, thus we only utilize microposts with sentiment polarity and discard neutral ones. At last, we gather 10,089 microposts, and the average number of microposts of each user is 121.6. We released our dataset athttps://github.com/zhucqut/UHSE to facilitate reproducing of our the experimental results. The detailed statistics of the dataset are illustrated in Table 1.

4.3. Baselines

In order to fully evaluate the performance of the proposed model, we compare it with ten competitive baselines:

SO-CAL: Taboada et al. [6] propose a lexicon-based method by extracting sentiment from text. This method considers the semantic orientation (polarity and strength) of words, and also takes into account intensification and negation.

C-WL: Dredze et al. [9] use a traditional regression analysis method for the emotion classification task. They propose a confidence-weighted linear classifier by introducing parameter confidence information to the linear classifier, which maintains a Gaussian distribution over parameter vectors and update the mean and covariance of the distribution with each instance.

AENB: Narayanan et al. [12] present a sentiment classification method based on Naive Bayes. This method aims to choose the right type of features and removing noise by appropriate feature selection and results in a significant performance improvement.

CNN: Kim et al. [13] employ convolutional neural networks (CNN) trained on top of pre-trained word vectors for the task of

[3] http://www.keenage.com/
[4] https://rdrr.io/rforge/tmcn/man/NTUSD.html
[5] https://www.weibo.com

Table 1
Statistics of the datasets.

Num. of microposts	Total 10089	Positive 4111	Negative 5918
Num. of each user's historical microposts	Average 120.6	Min 23	Max 274
Length of microposts	Average 29.5	Min 3	Max 1256
Time interval of each user's microposts (days)	Average 5.9	Min 1.2	Max 28.5

sentiment classification. This method leverages both pre-trained and task-specific vectors by having multiple channels.

Bi-LSTM: Xu et al. [32] take the Bidirectional Long Short-Term Memory (Bi-LSTM) to capture the context information effectively for learning better representation. It obtains the sentiment tendency of the text through the feedforward neural network and softmax mapping.

Self-attention: Lin et al. [17] propose a self-attention mechanism to learn sentence embedding by extracting different aspects of the sentence into multiple vector representations. It performs on top of an LSTM embedding model and relieves some long-term memorization burden from LSTM.

TextGCN: Yao et al. [33] propose to use graph convolutional networks for text classification. It constructs a single graph with global relations between documents and words, then jointly learns the embeddings for both words and documents.

LR-Bi-LSTM: Qian et al. [34] introduce linguistically regularized LSTMs for sentence-level sentiment classification. This method aims to enhance the performance of sentiment classification by imposing linguistic roles of sentiment lexicons, negation words, and intensity words into neural networks.

TextING: Zhang et al. [35] extend TextGCN by building an individual graph for each document and then apply a GNN to capture the fine-grained word representations based on the local structure.

MEAN: Lei et al. [18] propose a coupled word embedding obtained from character-level embedding and word-level embedding to capture both the character-level morphological information and word-level semantics. They further propose a multi-sentiment-resource attention module to learn better representation from modeling three kinds of sentiment resources including sentiment lexicon, intensity words, negation words.

4.4. Evaluation metrics

For evaluation, we use the Macro-averaged Precision (P), Recall (R), and F1-score (F1) [36] which evaluates averaged P, R and F1 of all different class-labels, respectively. It gives equal weight to each label. Formally, Macro-averaged P, R, and F1 are defined as:

$$P = \frac{1}{|C|}\sum_{t \in C} P_t \tag{22}$$

$$R = \frac{1}{|C|}\sum_{t \in C} R_t \tag{23}$$

$$F1 = \frac{1}{|C|}\sum_{t \in C} \frac{2P_t R_t}{P_t + R_t} \tag{24}$$

where $P_t = \frac{TP_t}{TP_t + FP_t}, R_t = \frac{TP_t}{TP_t + FN_t}$, and TP_t, FP_t, FN_t denote the true-positives, false-positives, and false-negatives for the t-th label in a label set C, respectively.

4.5. Results

Table 2 shows the performance of our model and baseline models. It can be observed that our proposed model achieves the best performance among all methods. It is obvious that the SO-CAL method gets the worst performance, because lexicon-based approaches cannot effectively capture the semantic information. C-WL is slightly better than SO-CAL as it further leverages the semantic signals of the text. The performance of AENB is better than C-WL, which indicates that the Bayesian method can better fit nonlinear data.

Compared with traditional methods, deep neural network-based methods demonstrate superior performance. For example, the F1-score of Bi-LSTM and CNN are considerable higher than AENB. In addition, the Self-attention model outperforms both Bi-LSTM and CNN, and achieves a F1-score of 0.8257, as it can effectively capture the complex connection information among micropost words. The performance of TextGCN is comparable to that of Self-attention. It models the global structure information by conducting a GCN over a textgraph. LR-Bi-LSTM outperforms aforementioned methods as it integrates linguistic roles of sentiment, negation and intensity words into neural networks via the linguistic regularization. TextING achieves a better performance than LR-Bi-LSTM, mainly since TextING captures a fine-grained word representation by applying GCN on the individual graph for each document. The best performing baseline method is MEAN, which achieves the highest F1-score of 0.8789. The reason is that it can effectively learn representations from multiple kinds of sentiment resources. Our method outperforms all baselines. Compared with the best performing baseline MEAN, the absolute F1-score improvement is 0.031. The results verify the effectiveness of our proposed model, which leverages user historical states in microblog sentiment classification.

4.6. Ablation study

In order to analyse the effectiveness of each component of our model, we also conduct an ablation study and the result is shown in Table 3.

W2V is the model which leverages word2vec to obtain word embeddings, and then employs a BiLSTM to learn the contextual information of words. The hidden vector of the last word will be used for prediction. W2V-sentiment words expands W2V by introducing sentiment words. BERT-sentiment words is the model which replaces word2vec with BERT. BASE-historical sentiment and BASE-historical semantic are the models which utilize user historical sentiment information and user historical semantic information, respectively. UHSE is our proposed model. UHSE w/o time weight and UHSE w/o semantic weight are the models

Table 2
Performance of different models. Best performance is indicated in bold.

Methods	Precision	Recall	F1
SO-CAL [6]	0.7114	0.6580	0.6569
C-WL [9]	0.6756	0.6695	0.6714
AENB [12]	0.7979	0.7988	0.7967
CNN [13]	0.8029	0.8043	0.8018
Bi-LSTM [14]	0.8208	0.8181	0.8189
Self-attention [17]	0.8274	0.8280	0.8257
TextGCN [33]	0.8305	0.8290	0.8294
LR-Bi-LSTM [34]	0.8348	0.8350	0.8349
TextING [35]	0.8576	0.8574	0.8570
MEAN [18]	0.8794	0.8796	0.8789
UHSE	**0.9099**	**0.9098**	**0.9099**

Table 3
Ablation study results in terms of Precision, Recall and F1-score.

Methods	Precision	Recall	F1
W2V	0.8111	0.8117	0.8113
W2V-sentiment words	0.8698	0.8603	0.8595
BERT-sentiment words (BASE)	0.8794	0.8795	0.8788
BASE-historical sentiment	0.8973	0.8974	0.8970
BASE-historical semantic	0.8970	0.8964	0.8958
UHSE w/o time weight	0.8963	0.8949	0.8953
UHSE w/o semantic weight	0.8980	0.8964	0.8968
UHSE	**0.9099**	**0.9098**	**0.9099**

which remove time weight and semantic weight from UHSE, respectively.

In Table 3, we observe that the performance of W2V-sentiment words is better than W2V. It verifies that adding sentiment words is beneficial for improving the classification performance. BERT-sentiment words performs better than W2V-sentiment since BERT can better learn word vector expressions through the transformer mechanism. BASE-historical sentiment and BASE-historical semantic further boost the performance as compared to BERT-sentiment words. This shows that involving user historical sentiment information and user historical semantic information is very important for sentiment classification. In addition, we also observe that both time weights and semantic weights play a critical role in our model, and removing either of them will lead to a significant performance drop. Compared with all variants, our proposed model UHSE achieves the best performance. This is because that UHSE can effectively capture both user historical semantic and sentiment preference information in an end-to-end unified framework. In particular, UHSE models user historical semantic preference by utilizing the User Historical Semantic Encoder, and captures user historical sentiment preference by employing the User Historical Sentiment Encoder. The results are also consistent with that in Table 2 where UHSE outperforms all other existing methods which do not take user historical information into consideration.

4.7. Case study

In this section, we conduct a case study to analyze the influence of historical emotional information on our approach. Table 4 shows a list of user historical posted texts as well as her current text. For the current text "8 个月， 你开心就 好 (Eight months, as long as you're happy) ", we cannot get the real emotion of the user. However, when we go through her previous microposts, we can observe that she is experiencing a process of losing a loved one. Therefore, it is clear that "Eight months, as long as you're happy" are just words of comfort and her real emotion is negative as she is broken heart at this moment.

From Table 4, we observe that the best performing baseline MEAN incorrectly classified it as positive while our proposed approaches correctly assigned a negative label to the current micropost. The is mainly because that MEAN focuses on modeling the information within the current micropost while ignoring the rich information from the user's historical microposts. When the positive words "happy" appears, MEAN would be misleaded and is prone to classify the micropost as positive. As compared to MEAN, our proposed method UHSE involves user's historical emotion states into the loop of sentiment classification. When the sentiment polarity of the current micropost is ambiguous, such as the one mentioned above, UHSE can correctly identify the real emotion of the user by exploring her historical emotion states.

4.8. Parameter sensitivity

In this section, we study the sensitivity of our proposed method USHE to the two parameters k and z, and explore how the different values for parameters would affect the model performance.

We first analyze the parameter k, which is the number of historical microposts. The left column of Fig. 3 shows the performance of USHE by varying k from 1 to 10 with a step size 1. From the figure, we can see that the value of k significantly affects the performance of USHE in all evaluation metrics. Specifically, with the increase of the k value, the performance first increases until it reaches the highest performance at k = 5, and then starts to decline. The reason behind this trend is that for smaller k, less user's historical microposts are leveraged and can not capture a user's historical sentiment preference effectively. When k becomes too large, more irrelevant and outdated information will be incorporated into the model and result in sub-optimal performance.

Then, we study the influence of the parameter z which is the time decay length for controlling time weights. The right column of Fig. 3 demonstrates the performance of USHE by varying z from 10 to 50 with a step size 10. We can observe that the model performance increases as we employ a larger value of z, and reach a peak when z is 30. When we further increase z, the performance will drop gradually. This is mainly because that if we set a small value to z, the time weights will decay quickly and several recent relevant microblogs will be underestimated. However, when the value of z is too large, all historical texts will have similar time weights

Table 4
Case Study. Results of the best performing baseline MEAN and our proposed method UHSE on testing examples.

Micropost Position	Micropost Content	MEAN	UHSE	Label
Historical Micropost 1	时光的绝情之处是，它让你熬到真相，却不给你任何补偿。 The heartless part of time is that it allows you to stay up till the truth, but it does not give you any compensation.	Neg	Neg	Neg
Historical Micropost 2	这些年我交了很多不再联系的朋友，而我没有想过你也是这里面的。 These years, I have made many friends who I don't contact anymore, while I never imagine that you are among them.	Neg	Neg	Neg
Historical Micropost 3	你给我一个绚烂的秘密，沉睡在仲夏夜里的美丽梦境中。 You give me a gorgeous secret, sleeping in the midsummer night with a beautiful dream.	**Pos**	Neg	Neg
Historical Micropost 4	我做过最坚强的事，就是听你说你爱她，笑容不减，眼泪不掉。 The strongest thing I have ever done is to hear you say you love her, my smile didn't diminish and my tears didn't fall.	Neg	Neg	Neg
Current Micropost	8个月，你开心就好 Eight months, as long as you're happy.	**Pos**	Neg	Neg

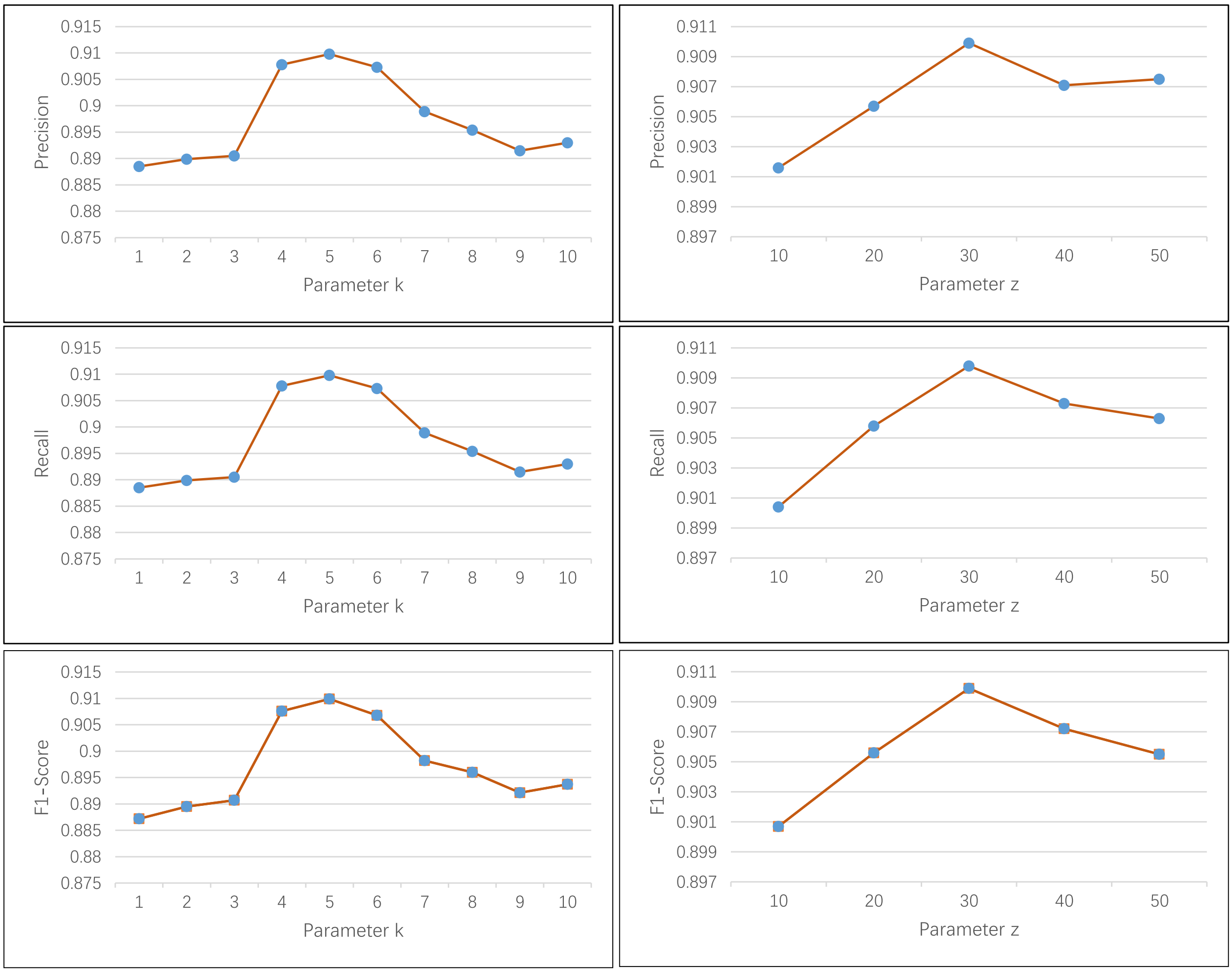


Fig. 3. The sensitivity of USHE to parameter k (the number of historical microposts) and parameter z (the time decay length for controlling time weights).

and the effectiveness from time dimension will vanish, which is consistent with the results in the ablation study.

5. Conclusion

In this paper, we attempt to involve user's historical semantic and sentiment preferences into the loop of microblog sentiment analysis. To this end, we propose a unified model which consists of a micropost encoder to learn representative micropost representation, a user sentiment state encoder to capture user-specific historical emotion states, as well as a user semantic state encoder to capture user-specific historical semantic states. Since there are few microblog sentiment datasets, which have rich user historical information to the scenario we study in this work, we create a new dataset with user historical information by collecting data from Weibo. We release the dataset for the research community and enable reproducibility of research. Extensive experiments are conducted, and results show that the proposed approach achieves superior performance compared with state-of-the-art baseline methods.

As the sentiment would vary considerably with respect to different time intervals. In this paper, we simplify this issue and leverage users' k most recent microposts to capture their sentiment states. In the future work, we will model the time interval information in a fine-grained manner. In addition, we will also investigate the influence of user historical state information under a different scenario, such as Twitter, where users have different background as compared to users in Weibo.

CRediT authorship contribution statement

Xiaofei Zhu: Conceptualization, Methodology, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition. **Jie Wu:** Software, Investigation, Writing - original draft. **Ling Zhu:** Investigation, Data curation. **Jiafeng Guo:** Supervision, Project administration. **Ran Yu:** Visualization. **Katarina Boland:** Investigation. **Stefan Dietze:** Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (No. 61722211); the Federal Ministry of Education and Research (No. 01LE1806A); the Beijing Academy of Artificial Intelligence (No. BAAI2019ZD0306); the Technology Innovation and Application Development of Chongqing (No. cstc2020jscx-dxwtBX0014).

Appendix A. Influence of time

In this work, we leverage the Newton's Law of Cooling to infer the time weights of the previous microposts of s_i as mentioned in Eq. (11). Here we discuss how to estimate the two parameters μ and m.

Suppose that we need to decay from N_{init} to N_{finish} after z time units, as follows:

$$N_{init} = N_0 e^{-\mu m}$$

$$N_{finish} = N_0 e^{-\mu(z+m)}$$

then we have,

$$\mu = \frac{1}{z} \ln\left(\frac{N_{init}}{N_{finish}}\right)$$

$$m = \frac{1}{\mu} \ln\left(\frac{N_0}{N_{init}}\right)$$

Appendix B. Key factors of data

There are two key factors of data which would affect the performance of the proposed method:

(1) The number of available user historical microposts. A large number of historical microposts can provide rich information about user historical sentiment as well as semantic preference, which will be used to enhance the representation of user current micropost. It is important especially when the current micropost is sparse and ambiguous;
(2) The quality of labels for all historical microposts. As the proposed method relies on user historical sentiments, the correctness of the labels will affect the results of identifying user sentiment preference. As manually annotating the labels for all historical microposts is impractical in real application, we leverage a base sentiment classifier for providing surrogate labels, and experimental results based on the surrogate labels verify the effectiveness of the strategy.

References

[1] B. Pang, L. Lee, et al., Opinion mining and sentiment analysis, Foundations and Trends in Information Retrieval 2 (1–2) (2008) 1–135..
[2] H. Li, Y. Chen, H. Ji, S. Muresan, D. Zheng, Combining social cognitive theories with linguistic features for multi-genre sentiment analysis, in: Proceedings of the 26th Pacific Asia Conference on Language, Information, and Computation, 2012, pp. 127–136..
[3] L. Gui, Y. Zhou, R. Xu, Y. He, Q. Lu, Learning representations from heterogeneous network for sentiment classification of product reviews, Knowledge-Based Systems 124 (2017) 34–45.
[4] X. Sun, C. Li, F. Ren, Sentiment analysis for chinese microblog based on deep neural networks with convolutional extension features, Neurocomputing 210 (2016) 227–236.
[5] D. Lin, L. Li, D. Cao, Y. Lv, X. Ke, Multi-modality weakly labeled sentiment learning based on explicit emotion signal for chinese microblog, Neurocomputing 272 (2018) 258–269.
[6] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, M. Stede, Lexicon-based methods for sentiment analysis, Computational linguistics 37 (2) (2011) 267–307.
[7] J. Wiebe, T. Wilson, C. Cardie, Annotating expressions of opinions and emotions in language, Language Resources and Evaluation 39 (2–3) (2005) 165–210.
[8] A. Joshi, A. Balamurali, P. Bhattacharyya, R. Mohanty, C-feel-it: a sentiment analyzer for micro-blogs, in: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Systems Demonstrations, Association for Computational Linguistics, 2011, pp. 127–132..
[9] M. Dredze, K. Crammer, F. Pereira, Confidence-weighted linear classification, in: Proceedings of the 25th International Conference on Machine learning, ACM, 2008, pp. 264–271.
[10] A.G. Shirbhate, S.N. Deshmukh, Feature extraction for sentiment classification on twitter data, International Journal of Science and Research 5 (2) (2016) 2183–2189.
[11] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up?: sentiment classification using machine learning techniques, in: Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing-Volume 10, Association for Computational Linguistics, 2002, pp. 79–86..
[12] V. Narayanan, I. Arora, A. Bhatia, Fast and accurate sentiment classification using an enhanced naive bayes model, in: International Conference on Intelligent Data Engineering and Automated Learning, Springer, 2013, pp. 194–201.
[13] Y. Kim, Convolutional neural networks for sentence classification, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 2014, pp. 1746–1751.
[14] K. Cho, B. van Merrienboer, Ç. Gülçehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning phrase representations using RNN encoder-decoder for statistical machine translation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 2014, pp. 1724–1734.
[15] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, E. Hovy, Hierarchical attention networks for document classification, in: Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies, 2016, pp. 1480–1489.
[16] X. Zhou, X. Wan, J. Xiao, Attention-based lstm network for cross-lingual sentiment classification, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016, pp. 247–256.
[17] Z. Lin, M. Feng, C.N. dos Santos, M. Yu, B. Xiang, B. Zhou, Y. Bengio, A structured self-attentive sentence embedding, in: Proceedings of the 5th International Conference on Learning Representations, 2017.
[18] Z. Lei, Y. Yang, M. Yang, Y. Liu, A multi-sentiment-resource enhanced attention network for sentiment classification, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018, pp. 758–763.
[19] L. Kong, C. Li, J. Ge, Y. Yang, F. Zhang, B. Luo, Construction of microblog-specific chinese sentiment lexicon based on representation learning, in: Proceedings of the 15th Pacific Rim International Conference on Artificial Intelligence, 2018, pp. 204–216.
[20] C. Zhao, S. Wang, D. Li, Exploiting social and local contexts propagation for inducing chinese microblog-specific sentiment lexicons, Computer Speech & Language 55 (2019) 57–81.
[21] T. Ito, K. Tsubouchi, H. Sakaji, T. Yamashita, K. Izumi, Word-level contextual sentiment analysis with interpretability, in: Proceedings of the 34th AAAI Conference on Artificial Intelligence, 2020, pp. 4231–4238.
[22] B. Eisner, T. Rocktäschel, I. Augenstein, M. Bosnjak, S. Riedel, emoji2vec: Learning emoji representations from their description, in: Proceedings of The Fourth International Workshop on Natural Language Processing for Social Media, 2016, pp. 48–54.
[23] Y. Chen, J. Yuan, Q. You, J. Luo, Twitter sentiment analysis via bi-sense emoji embedding and attention-based LSTM, ACM Multimedia Conference on Multimedia Conference 2018 (2018) 117–125.
[24] Z. Chen, S. Shen, Z. Hu, X. Lu, Q. Mei, X. Liu, Emoji-powered representation learning for cross-lingual sentiment classification, in: The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13–17, 2019, 2019, pp. 251–262..
[25] F. Wu, Y. Huang, Y. Song, Structured microblog sentiment classification via social context regularization, Neurocomputing 175 (2016) 599–609.
[26] X. Wang, D. Jin, K. Musial, J. Dang, Topic enhanced sentiment spreading model in social networks considering user interest, in: Proceedings of the 34th AAAI Conference on Artificial Intelligence, 2020, pp. 989–996.
[27] Y. Zheng, R. Zhang, S. Mensah, Y. Mao, Replicate, walk, and stop on syntax: An effective neural network model for aspect-level sentiment classification, in: Proceedings of the 34th AAAI Conference on Artificial Intelligence, 2020, pp. 9685–9692..
[28] S. Feng, Y. Wang, L. Liu, D. Wang, G. Yu, Attention based hierarchical lstm network for context-aware microblog sentiment classification, World Wide Web 22 (2019) 59–81.
[29] H.P. Chan, W. Chen, I. King, A unified dual-view model for review summarization and sentiment classification with inconsistency loss, in: Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020, 2020, pp. 1191–1200..
[30] J. Devlin, M. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, in: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2019, pp. 4171–4186..
[31] L. Cai, C. Peng, S. Chen, L. Guo, Sentiment analysis based on multiple features convolutional neural networks, Computer Engineering 45 (4) (2019) 169–174.
[32] G. Xu, Y. Meng, X. Qiu, Z. Yu, X. Wu, Sentiment analysis of comment texts based on bilstm, IEEE Access 7 (2019) 51522–51532.
[33] L. Yao, C. Mao, Y. Luo, Graph convolutional networks for text classification, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, 2019, pp. 7370–7377..
[34] Q. Qian, M. Huang, J. Lei, X. Zhu, Linguistically regularized lstms for sentiment classification, in: arXiv preprint arXiv:1611.03949, 2016..
[35] Y. Zhang, X. Yu, Z. Cui, S. Wu, Z. Wen, L. Wang, Every document owns its structure: Inductive text classification via graph neural networks, in:

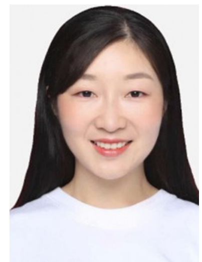
Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2018, pp. 334–339.
[36] H. Peng, J. Li, Y. He, Y. Liu, M. Bao, L. Wang, Y. Song, Q. Yang, Large-scale hierarchical text classification with recursively regularized deep graph-cnn, in: Proceedings of the 27th International Conference on World Wide Web, 2018, pp. 1063–1072.

Prof. Dr. Xiaofei Zhu is currently a full professor at Chongqing University of Technology. He received his Ph. D. degree at the Institute of Computing Technology, Chinese Academy of Science (ICT-CAS), Beijing, China, in 2012. Then he spent four years as a Postdoctoral Research Fellow at the L3S Research Center, Leibniz University Hannover, Hanover, Germany. His research interests include web search, data mining and machine learning, and he has published more than 30 papers in international conferences and journals, including the top conferences like SIGIR, WWW, CIKM, TKDE, etc. He has won the First Place Award in CIKM AnalytiCup (2020) and the Best Paper Awards of CIKM (2011). He has served as the Area Chair of top conferences such as CIKM, as well as the PC member and editorial board member of numerous international conferences and journals, including SIGKDD, SIGIR, AAAI, CIKM, WSDM, COLING, EMNLP, ECIR, etc.

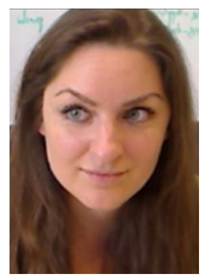
Jie Wu is currently a researcher in Baidu Inc. She received her B.E. degree in Software Engineering from Luliang University, Shanxi, China, in 2017 and received her M.E. degree in Computer Science from Chongqing University of Technology, Chongqing, China, in 2020. Her research interests include natural language processing, user modeling, sentiment analysis and machine learning.

Ling Zhu is currently a master student at the College of Computer Science and Engineering from Chongqing University of Technology, Chongqing, China. She received her B.E. degree in Computer Science from Neijiang Normal University, Sichuan, China, in 2014. Her research interests include web mining, machine learning and natural language processing.

Prof. Dr. Jiafeng Guo is currently a Professor in Institute of Computing Technology, Chinese Academy of Sciences, and University of Chinese Academy of Sciences. He received his Ph.D. in Computer Software and Theory from the University of Chinese Academy of Sciences, Beijing, China, in 2009. He has worked on a number of topics related to web search and data mining, including query representation and understanding, learning to rank, and text modeling. His current research focuses on representation learning and neural models for information retrieval and filtering. He has published more than 80 papers in several top conferences/journals such as SIGIR, WWW, CIKM, IJCAI, and TKDE. His work on information retrieval has received the Best Paper Award in ACM CIKM (2011), Best Student Paper Award in ACM SIGIR (2012) and Best Full Paper Runner-up Award in ACM CIKM (2017). Moreover, he has served as the PC member for the prestigious conferences including SIGIR, WWW, KDD, WSDM, and ACL, and the associate editor of TOIS.



Dr. Ran Yu is currently a senior research fellow in the Data Science & Intelligent Systems (DSIS) research group at the University of Bonn. Prior to that, she worked as a postdoctoral researcher at GESIS - Leibniz Institute for the Social Sciences in Cologne, Germany. She received her B.E. degree in Computer Science and Technology from Fuzhou University, Fujian, China, in 2011, and received her M.E. degree in Computer Science from Institute of Computing Technology(ICT), Chinese Academy of Sciences, Beijing, China, 2014, and received her PhD in Computer Science from GESIS – Leibniz Institute for the Social Sciences & Heinrich Heine University Düsseldorf. Her research interests include user modeling, information retrieval and semantic Web. She has published more than 20 papers in international conferences and journals, including SIGIR, CIKM, ISWC, WWW, ICDE, etc.



Katarina Boland is currently a PhD candidate at GESIS - Leibniz Institute for the Social Sciences in Köln, Germany. She received her M.A. degree in Computational Linguistics, Computer Science and Psychology from Heidelberg University, Heidelberg, Germany, in 2011. Her research interests include natural language processing, text mining, with a particular emphasis in the analysis of public discourse in social media and the extraction of information from scientific publications.

Prof. Dr. Stefan Dietze is a professor at the University of Düsseldorf and the scientific director of Knowledge Technologies for the Social Sciences (WTS), GESIS . He received his Ph.D. degree from the Institute for Computer Science of the University of Potsdam (Ph.D. / Dr. rer. nat. in Applied Computer Science), Germany, in 2004. His professional career led him to the Fraunhofer Institute for Software and Systems Engineering (ISST) in Berlin. Then, he spent five years as a Postdoctoral Research Fellow at the Knowledge Media Institute (KMi) of The Open University in Milton Keynes, UK. Before joining GESIS, he led a research group at the L3S research center at Leibniz University Hannover. His research interests include semantic web technologies, web intelligence, semantic search, and information retrieval. He has published numerous papers in prestigious journals and international conferences, including SIGIR, WWW, CIKM, ESWC, CHIIR and so on. He has been general / poster / track chair, program committee and editorial board member of numerous international conferences and journals.