Neurocomputing 383 (2020) 165–173

Contents lists available at ScienceDirect

Neurocomputing

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

BiLSTM with Multi-Polarity Orthogonal Attention for Implicit Sentiment Analysis

Jiyao Wei[a,1], Jian Liao[a,1], Zhenfei Yang[a], Suge Wang[a,b,*], Qiang Zhao[a]

[a] School of Computer & Information Technology, Shanxi University, China
[b] Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, China

ARTICLE INFO

Article history:
Received 15 August 2019
Revised 20 November 2019
Accepted 20 November 2019
Available online 5 December 2019

Communicated by Dr. Tie-Yan Liu

Keywords:
Implicit Sentiment Analysis
Multi-polarity Attention
Orthogonal Attention

ABSTRACT

Sentiment analysis has been a popular field in natural language processing. Sentiments can be expressed explicitly or implicitly. Most current studies on sentiment analysis focus on the identification of explicit sentiments. However, implicit sentiment analysis has become one of the most difficult tasks in sentiment analysis due to the absence of explicit sentiment words. In this article, we propose a BiLSTM model with multi-polarity orthogonal attention for implicit sentiment analysis. Compared to the traditional single attention model, the difference between the words and the sentiment orientation can be identified by using multi-polarity attention. This difference can be regarded as a significant feature for implicit sentiment analysis. Moreover, an orthogonal restriction mechanism is adopted to ensure that the discriminatory performance can be maintained during optimization. The experimental results on the SMP2019 implicit sentiment analysis dataset and two explicit sentiment analysis datasets demonstrate that our model more accurately captures the characteristic differences among sentiment polarities.

1. Introduction

With the rapid development of social media, sentiment analysis has become an attractive research topic due to its vast potential for practical applications such as review analysis, public opinion analysis, and content-based recommendation.

Sentiments are, by nature, subjective because they regard people's subjective views, appraisals, evaluations, and feelings [1]. Liu [2] first classified sentiments into explicit and implicit sentiments according to the subjectivity and the objectivity. Explicit sentiment analysis has received substantial attention from academia and industry and has yielded important results [1,3,4]. However, the analysis of implicit sentiment, which is defined as ”a language fragment (sentence, clause or phrase) [that] expresses subjective sentiment but contains no explicit sentiment word”, has attracted little attention [5]. Because there is no explicit sentiment word in an implicit sentiment expression, traditional sentiment dictionary-based methods [6] are not effective and the model cannot extract a definite sentiment clue from the words. Implicit analysis has become one of the core problems in natural language processing [5,7].

Consider the following examples of explicit and implicit sentiment sentences:

E1: 何止十年啊给你一辈子, 我们天天**开开心心**滴。 (For not just ten years but a lifetime, we live a **happy** life every day.)

E2: 内堂不能拍照, 里面有当年文成公主的陪嫁, 好像是一尊观音像, 历经岁月, 依然金光灿灿。 (Picture taking is prohibited in the inner hall. There is a statue of Avalokitesvara Bodhisattva inside, which is the dowry of Princess Wencheng. The statue is still shining through millennium vicissitudes.)

E3:8 月 7 日下午的奥运男子 110 米栏预赛中, 刘翔摔倒无缘半决赛。 (In the Olympic men's 110-meter hurdle preliminaries on the afternoon of August 7, Liu Xiang lost the semifinal due to falling.)

Sentence E1 is an explicit sentiment statement. The explicit word “开开心心 (happy)” can be a significant clue for identifying its sentiment polarity. Sentence E2 expresses the heartfelt admiration for a landscape by stating what the author sees. Sentence E3 contains an implicit derogatory sentiment. It expresses author's regret for Liu Xiang's defeat in the competition.

Because explicit sentiment sentences contain clear sentiment words that provide initial sentimental clues, it is relatively easy to distinguish their sentiment polarity via classical machine methods or popular deep learning methods. However, people tend to express their feelings in an implicit and concealed way. The

[*] Corresponding author at: School of Computer & Information Technology, Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, China.

E-mail addresses: wsg@sxu.edu.cn, 439967828@qq.com (S. Wang).

[1] Jiyao Wei and Jian Liao contribute equally to the article, and should be regarded as co-first authors of the article.

https://doi.org/10.1016/j.neucom.2019.11.054
0925-2312/© 2019 Elsevier B.V. All rights reserved.

proportions of implicit sentiment are approximately 15%-20% in subjective sentences among various areas [5,8].

BiLSTM(Bi-directional Long Short-Term Memory) with an attention mechanism has widely been proved to be an effective model for sentiment analysis. Traditional methods [9,10] typically adopt single attention for assigning the words' weights. However, since each implicit sentiment expression has no explicit sentiment words, it is highly difficult to capture the sentiment-related words using single attention. In contrast, we assume that the weights of neutral words in implicit sentiment expressions vary among sentiment polarity scenarios. For example, in sentence E3, "摔倒无缘半决赛 (lost the semifinal due to falling)" expresses a negative and regretful sentiment toward Liu Xiang without explicit sentiment words. However, the words "摔倒 (falling)" and "无缘 (miss)" should have much higher weights in negative sentiment scenarios than for other polarities. This will lead to differences among the attention weights that are associated with each sentiment polarity. This difference can be regarded as a significant feature for implicit sentiment identification and polarity classification.

In this paper, we propose a novel multi-polarity orthogonal attention mechanism for modeling these differences and improve the BiLSTM model for implicit sentiment analysis. By using multi-polarity attention, we adopt the embeddings of words that differ in terms of sentiment polarity as the initialization of attentions; hence, our model can better capture the characteristics of each sentiment polarity. An orthogonality restriction is imposed during model optimization to maintain the performance in distinguishing among attentions.

The main contributions of this paper are summarized as follows:

1. We proposed the multi-polarity attention mechanism, which enables an attention-based model to more accurately capture differences in sentiment polarity and use it as a significant feature for classification, especially for those implicit sentiment expressions without sentiment words as clues.
2. We imposed the orthogonal restriction on the multi-polarity attention layer, which can effectively preserve the differences among the attentions' representations and maintain the performance in distinguishing among the attentions during the optimization process.
3. We proposed a new identification and classification model for a novel Chinese implicit sentiment analysis task. The results demonstrate our model realizes satisfactory performance in implicit sentiment analysis and can serve as a baseline for further studies. The code can be available at https://github.com/SXU-CICI/bilstm_mpoa.

The remainder of this paper is organized as follows: Section 2 introduces the related work. The proposed multi-polarity orthogonal attention methods are detailed in Section 3. The experiment and analysis are discussed in Section 4. Finally, the conclusions of the paper are presented in Section 5.

2. Related work

Much of the existing research has focused only on explicit sentiments and only limited work has been conducted on implicit sentiments. Studies of implicit sentiment analysis are still in their initial stages.

The related work mainly consists of studies on implicit sentiment analysis and representation learning models.

2.1. Implicit sentiment analysis

Most corpus resources are constructed for explicit sentiment analysis tasks [8,11–14]. Chen and Chen [7] constructed a double-implicit corpus. They aimed at identifying aspects and polarities of opinion statements that do not contain sentiment words or aspect terms and they observed that an implicit sentiment and its neighboring explicit sentiment tend to have the same aspect and polarity. Deng and Wiebe [15] analyzed implicit sentiments via the detection of explicit sentiment expression clues and the inference of events. Shutova et al. [16] designed a pattern-based recognition algorithm for identifying metaphors and experimented on datasets in English, Spanish and Russian. Zhang et al. [17] built a Chinese metaphor corpus that contains annotations of linguistic metaphors, emotional categories (joy, anger, sadness, fear, love, disgust, and surprise), and intensity. Liao et al. [5] constructed a small fact-implied implicit sentiment corpus and proposed a multilevel semantic fused model; however, the model is limited by the syntax structure. Van de Kauter et al. [18] proposed a new fine-grained approach to mining explicit and implicit emotions in the financial news. Shutova et al. [16] focused on the implicit semantics of metaphors; they regarded a metaphor as a mapping projection from the source domain to the target domain and by exploiting statistical techniques, they started from a small seed set of metaphorical expressions, learned the analogies that were involved in their production and extended the set of analogies via verb and noun clustering. Their study focused more on implicit semantics than on sentiments. Based on this study, they later used weak supervised and unsupervised methods to learn the distribution of metaphor conception, explored clustering methods at various levels and under various limit conditions, and designed a pattern-based identification algorithm for recognizing metaphor sentences in English, Spanish and Russian [19].

2.2. Representation learning

Recently, representation learning has attracted a large amount of attention in natural language processing(NLP) research. Many neural-based representation learning methods have been proposed for encoding the semantics of words, sentences, documents or relations into distributed embeddings, and have exhibited satisfactory performances. Among them, CNN(Convolutional Neural Networks) [20,21] and RNN(Recurrent Neural Networks) [22,23] are widely used in many sentence embedding learning tasks. The CNN model uses a fixed-width window feature detector to slide over the sentence to extract the "local" semantic features effectively. RNN are built upon text sequences and better model the "global" semantic information. To overcome gradient disappearance or explosion in the RNN model, gated mechanisms, such as LSTM(Long Short-Term Memory) [24] and GRU(Gated Recurrent Unit) [25], are introduced into the original RNN model. Several variants have been proposed that are based on the LSTM model, such as the BiLSTM model [26] and the LSTM-CNN model [27]. The BiLSTM model uses reverse coding features, which overcomes the inability to encode information from back to front. By combining the regional CNN and LSTM, LSTM-CNN can consider both local (regional) information within sentences and long-distance dependencies across sentences in the prediction process.

An attention mechanism simulates the human brain when dealing with information overload. It filters out irrelevant information via feature selection and focuses more attention on key information. Adding an attention mechanism to the network layer will effectively improve the performance in extracting sentiment features to yield more accurate representations [28,29]. The commonly used attention mechanisms can be divided into global and local mechanisms according to the alignment position of attention [30]. They can also be divided into source target attention, self-attention and multihead attention [31,32]. Peng et al. [33] proposed a multilayer attention mechanism that is based on BILSTM, which can capture long-distance information

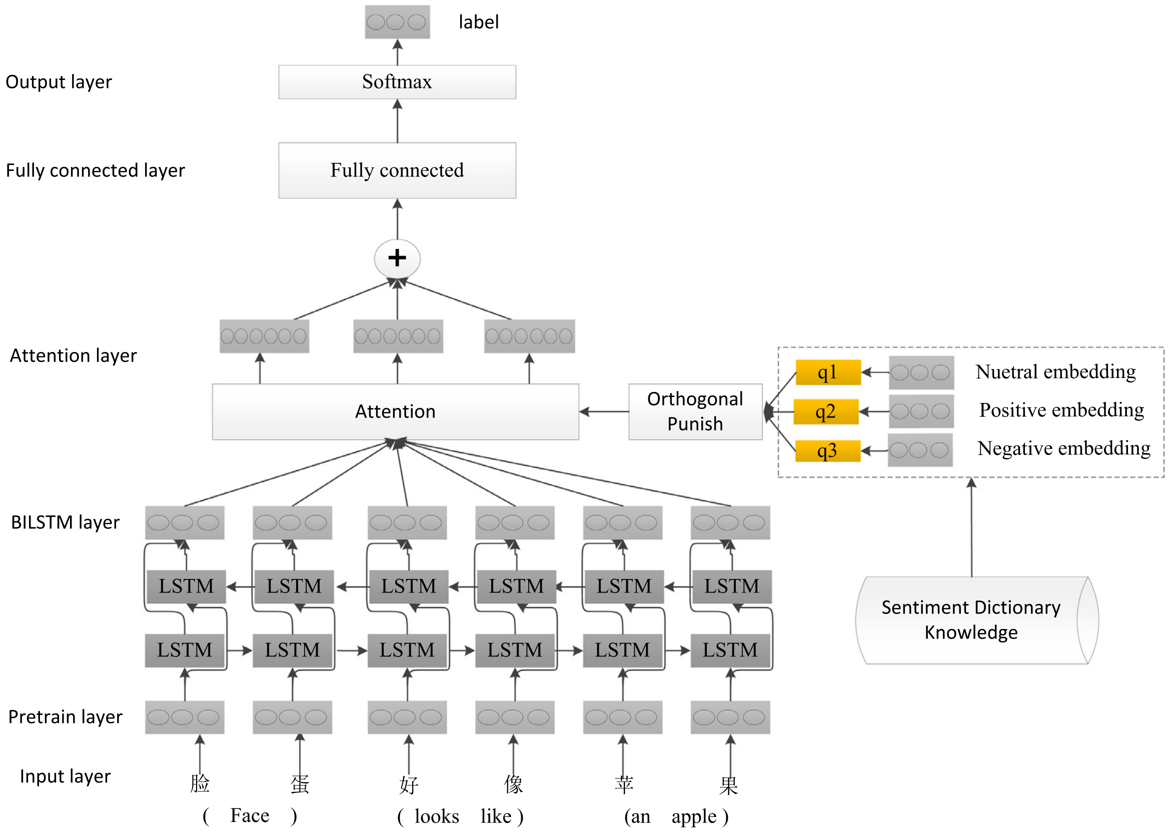


Fig. 1. The framework of BiLSTM with multi-polarity orthogonal attention model.

and has stronger robustness to irrelevant information. Qian et al. [34] attempted to improve the accuracy of sentiment analysis by using negative words and intensity words in the LSTM model via mandatory localization based on language. However, traditional attention-based methods have difficulty capturing the sentiment-related words by using single attention in implicit sentiment analysis. Except for NLP, multi-attention mechanism is also widely used in other domains like computer vision. Sun et al. [35] proposed an attention-based convolutional neural network which regulates multiple object parts among different images. Song et al. [36] proposed an end-to-end spatial and temporal attention model for human action recognition from skeleton data. Liu et al. [37] proposed a global context-aware attention LSTM network which has strong attention capability for skeleton based action recognition. These research inspired us to introduce sentiment-specific attention which embedded extra information for capturing the characteristic differences among sentiment polarities.

3. BiLSTM with multi-polarity orthogonal attention model

Because there is no explicit sentiment word to serve as a sentiment clue for identification, traditional dictionary-based methods are not effective. We must identify other features and methods for implicit sentiment analysis. In this paper, we propose a novel multi-polarity orthogonal attention mechanism for modeling the differences and improve the BiLSTM model for feature representation and implicit sentiment analysis.

3.1. Framework of the proposed model

This section introduces the proposed framework of the BiLSTM with multi-polarity orthogonal attention model, which is illustrated in Fig. 1.

According to Fig. 1, the proposed model contains 6 layers: the input layer, the pretraining layer, the BiLSTM layer, the multi-polarity attention layer, the fully connected layer, and the output layer. We describe each layer in detail as follows:

(1) Input layer. In this layer, each word in the sentence is represented as an index embedding via one-hot representation, which is denoted as $\boldsymbol{x_t} \in R^{|V|}$ (the character in bold represents the embedding or matrix of the corresponding variable, similarly hereinafter), where $|V|$ is the size of the vocabulary set in the experimental corpus. Then, the index embeddings are fed to the pretraining layer.

(2) Pretraining layer. The processing in the pretraining layer can be divided into two parts: the static part and the dynamic part, respectively.

In the static part, the model generates a shared lookup table $\boldsymbol{L} \in R^{d_e \times |V|}$, which adopts a pretrained word embedding resource as initialization, to encode more initial semantic information, where d_e is the dimension of the word embedding. The words are represented with continued embeddings via formula (1).

$$\boldsymbol{w_t} = \boldsymbol{x_t}^T \boldsymbol{L} \tag{1}$$

In the dynamic part, we use Elmo [38] to dynamically encode the context information of each word, which can improve the precision of the word representation for a task. The Elmo layer represents the words via formula (2).

$$\boldsymbol{w_t} = [\overrightarrow{\boldsymbol{h_t}^{LM}} : \overleftarrow{\boldsymbol{h_t}^{LM}}] \tag{2}$$

We use the same word embedding $\boldsymbol{L}$ in the static part to initialize it. $\overrightarrow{\boldsymbol{h_t}^{LM}}$ and $\overleftarrow{\boldsymbol{h_t}^{LM}}$ represent the hidden-layer outputs of forward and backward BiLSTM encoders, respectively, that have been optimized via a language model LM.

We also use Bert [39] to dynamically encode the context information of each word, which can improve the precision of the word representation for most tasks. The Bert layer represents the words via formula (3)-(5).

$$\boldsymbol{w_t} = \text{Concat}(\boldsymbol{head}_1, \ldots, \boldsymbol{head}_\text{h})\boldsymbol{W}^O \tag{3}$$

$$\boldsymbol{head}_\text{i} = \text{Attention}(\boldsymbol{QW}_i^Q, \boldsymbol{KW}_i^K, \boldsymbol{VW}_i^V) \tag{4}$$

$$\text{Attention } (\boldsymbol{q}, \boldsymbol{k}, \boldsymbol{v}) = \text{softmax}\left(\frac{\boldsymbol{qk}^T}{\sqrt{\boldsymbol{d_k}}}\right)\boldsymbol{v} \tag{5}$$

$\boldsymbol{W_i}^Q \in R^{d_{model} \times d_k}$, $\boldsymbol{W_i}^K \in R^{d_{model} \times d_k}$, $\boldsymbol{W_i}^V \in R^{d_{model} \times d_v}$, d_k is the dimension of attention, d_{model} is the dimension of embedding, $\boldsymbol{K}$, $\boldsymbol{Q}$, $\boldsymbol{V}$ are hidden states.

(3) BiLSTM layer. The BiLSTM model has been widely used and has realized magnificent performance in many natural language processing tasks. It represents a substantial improvement of LSTM that effectively solves the problem of gradient disappearance or explosion in simple RNN. The BiLSTM layer consists of two LSTM layers with opposite directions; hence, the model can capture the context information of a sequence more precisely. The basic LSTM unit consists of three gates and a conveyor belt that preserves the state of each neuron. It controls the path of information transfer via the gating mechanism. The state calculation formulas for each LSTM unit are presented as formulas (6)–(11).

$$\boldsymbol{f_t} = \sigma(\boldsymbol{W_f} \cdot [\boldsymbol{h_{t-1}}, \boldsymbol{w_t}] + \boldsymbol{b_f}) \tag{6}$$

$$\boldsymbol{i_t} = \sigma(\boldsymbol{W_i} \cdot [\boldsymbol{h_{t-1}}, \boldsymbol{w_t}] + \boldsymbol{b_i}) \tag{7}$$

$$\boldsymbol{o_t} = \sigma(\boldsymbol{W_o} \cdot [\boldsymbol{h_{t-1}}, \boldsymbol{w_t}] + \boldsymbol{b_o}) \tag{8}$$

$$\tilde{\boldsymbol{c}}_{\boldsymbol{t}} = \tanh(\boldsymbol{W_c} \cdot [\boldsymbol{h_{t-1}}, \boldsymbol{w_t}] + \boldsymbol{b_c}) \tag{9}$$

$$\boldsymbol{c_t} = \boldsymbol{f_t} \odot \boldsymbol{c_{t-1}} + \boldsymbol{i_t} \odot \tilde{\boldsymbol{c}}_{\boldsymbol{t}} \tag{10}$$

$$\boldsymbol{h_t} = \boldsymbol{o_t} \odot \tanh(\boldsymbol{c_t}) \tag{11}$$

where $\boldsymbol{f_t}$, $\boldsymbol{i_t}$, and $\boldsymbol{o_t}$ represent the forget gate, the input gate, and the output gate, respectively; $\boldsymbol{c_t}$ is internal information status; $\boldsymbol{h_t}$ is the output at time t; $\boldsymbol{W_f}$, $\boldsymbol{W_i}$, and $\boldsymbol{W_o} \in R^{d_h \times d_k}$ are the weight matrices; $\boldsymbol{b_f}$, $\boldsymbol{b_i}$, $\boldsymbol{b_o}$, and $\boldsymbol{b_c} \in R^{d_h}$ are the biases; d_h is the number of units in the LSTM hidden layer; $\boldsymbol{w_t} \in R^{d_e}$ denotes the word embedding from the pretraining layer; $d_k = d_e + d_h$; σ is the sigmoid function; and $\cdot$ and $\odot$ denote matrix multiplication and element multiplication, respectively.

In the LSTM model, the hidden-layer state $\boldsymbol{h_t}$ of each location encodes only the context information in the forward direction and ignores the backward context information. The BiLSTM model utilizes both the positive and reverse orders of the sequence information by concatenating the hidden-layer outputs of each model, as expressed in formulas (12) and (13).

$$\boldsymbol{m_t} = [\overrightarrow{\boldsymbol{h_t}} : \overleftarrow{\boldsymbol{h_t}}] \tag{12}$$

$$\boldsymbol{M} = \{\boldsymbol{m_1}, \ldots, \boldsymbol{m_t}, \ldots, \boldsymbol{m_T}\} \tag{13}$$

$\boldsymbol{m_t}$ is the concatenation of the forward and backward hidden layer output of the tth word in a sentence and $\boldsymbol{M}$ is the representation of a sentence with T words.

(4) multi-polarity attention layer. Attention is a mechanism for capturing significant features in a sequence by learning their weights automatically. It can effectively fuse the context-sensitive information of the keywords during sequence encoding. The attention mechanism can be formally expressed as follows:

$$\boldsymbol{v} = \sum_{i=1}^{T} \alpha_i \boldsymbol{h_i} \tag{14}$$

$$\alpha_i = \frac{\exp(e_i)}{\sum_{k=1}^{T} \exp(e_k)} \tag{15}$$

$$e_i = \boldsymbol{qWh_i} \tag{16}$$

where $\boldsymbol{v}$ is the weighted context embedding, α_i is the weight that corresponds to the hidden-layer output $\boldsymbol{h_i}$ at index i, e_i is the attention score at i, and W is a weight matrix that connects attention queries $\boldsymbol{q}$ and $\boldsymbol{h_i}$ via a bilinear function.

As discussed in Section 1, for the task of implicit sentiment analysis, since implicit sentiment expressions do not contain explicit sentimental words, there may be no significant differences among the weights of words in the same sentiment polarity scenario. However, considering various sentiment polarities will lead to differences among the attention weights that are related to each orientation. This difference can be regarded as a significant feature for implicit sentiment identification and polarity classification. Inspired by this observation, we propose a multi-polarity attention mechanism that can learn and embed the difference into the representation.

The proposed multi-polarity attention mechanism is inspired by the multihead attention mechanism. For each sentiment polarity, we introduce a separate query $\boldsymbol{q_i}$ for capturing the words' features under a specified sentiment orientation scenario. Multi-polarity attention makes each query vector pay more attention to the characteristics of various sentiment polarities via formulas (17)–(20).

$$\boldsymbol{v_j} = \sum_{i=1}^{T} \alpha_{ij} \boldsymbol{h_i} \tag{17}$$

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^{T} \exp(e_{kj})} \tag{18}$$

$$e_{ij} = \boldsymbol{q_j W h_i} \tag{19}$$

$$\boldsymbol{v} = [\boldsymbol{v_1}, \boldsymbol{v_2}, \ldots, \boldsymbol{v_n}] \tag{20}$$

where $\boldsymbol{v_j}$ is the weighted context embedding under sentiment polarity j's scenario. Finally, we concatenate the weighted embeddings of the sentiment polarities as the final representation of the implicit sentiment sentence. n is the number of attention queries that correspond to sentiment polarity classes.

An automatically initialized attention query $\boldsymbol{q_j}$ will learn and embed the sentiment-specific information during model optimization. Moreover, we posit that the pretrained sentiment words' embeddings already imply the subjective information, which will render the attention learning more precise.

To obtain the original subjective information, we introduce a sentiment dictionary [40] and initialize the attention query vectors with the fusion representation of words' embeddings that share the same sentiment polarity. The initialization is expressed via formula (21).

$$\boldsymbol{q_j} = \frac{1}{N_j} \sum_{i=1}^{N_j} \boldsymbol{w_{ij}} \tag{21}$$

where $\boldsymbol{w_{ij}}$ is the word embedding of word i with sentiment polarity class j. We regard $\boldsymbol{q_j}$ as the initialization of the ith attention query vector and N_j is the number of sentiment polarity classes.

(5) Fully connected layer and output layer. In the fully connected layer, the model maps the polarity-related fused distributed feature representation to the instance label space. It acts as a classifier throughout the neural network model. In the output layer, a softmax function is adopted for normalization and transforms the output of the fully connected layer into an approximate probability value $\boldsymbol{y}$ for each sentiment polarity category. The calculation is expressed in formula (22).

$$y = softmax(\boldsymbol{A}\boldsymbol{v}^{\mathrm{T}} + \boldsymbol{b}) \tag{22}$$

where $\boldsymbol{A}$ is the parameter matrix in the fully connected layer, $\boldsymbol{v}$ is the polarity-related fused distributed feature representation that was obtained via formula (20), $\boldsymbol{b}$ is the bias, and $softmax(\cdot)$ is the normalization function, which is calculated as $softmax(x_i) = e^{x_i} / \sum_{i\in n} e^{x_i}$.

3.2. Orthogonal attention mechanism

Sentiment-polarity-specific attention query is beneficial for learning the differences among sentiment polarities. However, during network optimization, we found that the basic model has difficulty learning the true differences among sentiment polarities. This is likely to lead to a scenario in which the distributed representations of the attention queries of different sentiment polarities are learned as almost the same embedding. To solve this problem, we proposed the orthogonal attention mechanism and incorporated it into the model optimization.

The orthogonal attention mechanism is inspired by the constraint on the similarity among latent topics in [41]'s study. For example, in a practical application, the expected set of latent topics are "military", "economy", and "health", among others; however, without using a similarity constraint, the generated set of topics could be "national defense", "weapons", "economy", and "health". Thus, it is important and necessary to introduce the similarity constraint strategy into the latent topic detection layer.

The orthogonal attention mechanism that is proposed in this paper is designed based on the assumption that a corpus is generated on a hyperspace in which each sentence is characterized by its semantic and sentiment polarity information and the sentiment polarity of each sentence is represented by a hypersurface in the space. In this paper, the sentiment polarities are represented in attention queries and can be treated as the basis of the space. As a result, the comprehensive sentiment polarity information of a sentence can be mathematically calculated via the concatenation of all queries' embeddings. According to the assumption, it may be necessary to learn distinctive and diverse attention queries.

The proposed orthogonal attention mechanism imposes orthogonal constraints on each query's embedding during the training process. It can effectively preserve the differences among attentions by minimizing the sum of the pairwise cosine similarities of all the attention queries, as expressed in formula (23).

$$\mathcal{L}_1 = \sum \left| \frac{\boldsymbol{q_i} \cdot \boldsymbol{q_j}}{|\boldsymbol{q_i}| \cdot |\boldsymbol{q_j}|} \right| \qquad (i, j \in [1, 2, ..., n], i \neq j) \tag{23}$$

where n is the number of attentions and $\boldsymbol{q_i}$ is the query embedding of attention i.

3.3. Model optimization

Since we design the orthogonal attention mechanism as a part of the objective function, we use two Losses in the proposed model to learn the representations of implicit sentiment sentences. $\mathcal{L}_1$ uses the cosine similarity to ensure the orthogonality of attention query embedding q to preserve the differences among attentions, as expressed in formula (23). The other is the classic cross-entropy loss function of the output layer, which is expressed in formula (24).

Table 1
The proportion of the implicit experimental dataset.

Subset	Neutral	Positive	Negative
training	6695	3946	3977
development	1654	990	1000
testing	1663	989	996

$$\mathcal{L}_2 = \sum_{(x,y)\in D} \sum_{c\in C} y^c log f^c(x;\theta) \tag{24}$$

where y is the golden label of an instance, x is the input sentence, f represents the model, its output is the corresponding probability of each category, c is the category, and θ is the set of model parameters.

During the training process, for each training batch, we jointly minimize $\mathcal{L}_1$ and $\mathcal{L}_2$ with a weight parameter γ, which is expressed in formula (25).

$$\mathcal{L} = \gamma\mathcal{L}_1 + (1-\gamma)\mathcal{L}_2 \tag{25}$$

4. Experiment and analysis

4.1. Datasets and evaluation index

(1) Datasets. Our experiments are conducted on the dataset of the evaluation of the Chinese implicit sentiment analysis task in SMP2019 (one of the top academic conferences on social media processing in China).[2]. The data in the dataset are crawled from two main sources: The first source is the largest and most popular social media platform, namely, Weibo.[3]. These parts of sentences focus on event implicit sentiments and are domain-independent with wide coverage; examples are "AlphaGo," "Letv bankruptcy," "Olympic game," "haze pollution," "national examination for admissions to the civil service," and "Spring Festival Gala". The other source is famous forums in China, namely, Mafengwo(tourism)[4], Ctrip(tourism)[5], and Autohome(automobile)[6]. The forum dataset mainly focuses on products and services. It is much longer than the Weibo dataset and can provide more context information.

Since there are some same sentences with different labels for their different contexts in the dataset, we remove the duplicate sentences from the dataset. The detailed statistics of the dataset are presented in Table 1.

In order to show the effectiveness and generalization of our proposed method on different datasets, we test our method on COAE(2015) and SemEval(2013-2017) datasets. The dataset of COAE(2015) task 2 social media sentiment analysis[7] contains the reviews of cars, digital, insurance, etc. The task of sentiment analysis in twitter in SemEval[8] is a famous and ongoing series of evaluations since 2013. Because of the large number of duplicate data in the SemEval dataset, we consolidate the dataset from 2013 to 2017 and divide all the data into training/development/testing subsets on the proportion of 8:1:1. The detailed statistics of the datasets are presented in Table 2.

(2) Evaluation index. We use the F1 scores of every sentiment polarity and the macro F1 to evaluate the performance of our

[2] https://biendata.com/competition/smpecisa2019/.
[3] https://weibo.com.
[4] https://www.mafengwo.cn/.
[5] https://www.ctrip.com/.
[6] https://www.autohome.com.cn/.
[7] http://tsaop.com:8066/web/resource.html.
[8] http://alt.qcri.org/semeval2017/task4/.

Table 2
The proportion of the explicit experimental datasets.

subset	COAE			SemEval		
	Neutral	positive	negative	Neutral	positive	negative
training	1233	9161	4752	23732	18289	9952
development	153	1133	593	2931	2359	1211
testing	156	1130	587	3026	2245	1213

proposed model. The calculation is expressed in formulas (26) and (27).

$$F-macro = \frac{1}{N}\sum_{i=1}^{N} F1_i \tag{26}$$

$$F1_i = \frac{2*P_i*R_i}{P_i+R_i} \tag{27}$$

where i is the index of the sentiment polarity and P_i and R_i are the precision and recall, respectively, of instances with sentiment polarity i.

4.2. Implementation details

(1) Input. In the experiment of SMP and COAE datasets, the initialization of the network input word embedding adopts the TENCE-AI-LAB word embedding[9]. The original sentiment words are collected from the sentiment dictionary of Dalian University of Technology[10] and use the average of words' embeddings, which are grouped by their sentiment polarity, to initialize the attention queries. The dimension of each word embedding is 200.

In the experiment of SemEval dataset, the initialization of the network input word embedding adopts the GLOVE word embedding[11]. The original sentiment words are collected from the sentiment dictionary of CMU of Twitter[12] and use the average of words' embeddings, which are grouped by their sentiment polarity, to initialize the attention queries. The dimension of each word embedding is 300.

(2) Network configuration. The number of units in each hidden layer is set as 200. When do experment in Chinese dataset, the word embedding dimension as 200 (initialized by the TENCE-AI-LAB word embedding). When do experment in English dataset, the word embedding dimension as 300 (initialized by the GLOVE840B300D word embedding). The middle layer has two layers and the size of the LSTM cell state is 200. The dropout from the hidden layer to the output layer is 0.3. Among the optimizers that are discussed in Section 3.3, we adopt the stochastic gradient descent optimizer, where the learning rate is 0.1, the momentum is 0.9, the batch size is 64, the weight parameter γ is 0.1, and the maximum number of training iterations is 100. The dimensions of the fully connected layers is 1200. In the Bert model, we use the Bert-adam optimizer and set the dropout as 0.5, the number of network neurons as 768, the maximum number of training iterations as 10, which is the official configuration of Bert.

4.3. Baselines

Implicit sentiment analysis in this paper is essentially a type of text classification problem. We use popular models that have exhibited excellent performances in explicit sentiment analysis and text classification as the baselines, which are described as follows:

Table 3
The performance of implicit sentiment analysis.

Method	F-neutral	F-positive	F-negative	F-macro
GRU	0.763	0.595	0.715	0.691
LSTM	0.783	0.605	0.733	0.707
BiLSTM+att	0.810	0.634	0.747	0.730
BiLSTM+multi-att	0.800	0.630	0.749	0.726
proposed model(dynamic-elmo)	0.808	0.684	0.669	0.663
proposed model(static)	0.804	0.644	0.752	0.733
proposed model(dynamic-bert)	**0.834**	**0.690**	**0.802**	**0.775**

LSTM\GRU [13] LSTM and GRU are widely adopted models for many NLP tasks. In this paper, the batch size is set as 128, the number of units in the hidden layer of the neural network is 128, and the middle layer has two layers. The size of the LSTM cell state is 200. The dimension of the fully connected layers is 200. We add a dropout of 0.8 to each RNN cell. The model uses the Adam optimizer with a learning rate of 0.001.

BiLSTM+attention [42][14] BiLSTM with an attention mechanism has been the most popular and effective model in recent years. We set the batch size of the model as 64, the number of middle layer as 2 and the size of the LSTM cell state as 200. Bilinear attention is adopted and the dropout is 0.3. The dimension of the fully connected layers is 400. The model uses a stochastic gradient descent optimizer and the learning rate and the momentum are 0.01 and 0.9, respectively.

BiLSTM+multi-attention BiLSTM with a multi-attention mechanism model can learn the differences among objects. We set the batch size, the number of middle layer, the size of the LSTM cell state, dropout, optimizer parameters of this model are the same as BILSTM+attention. We randomly initialize the multi-attention queries as a baseline model for evaluating the performance and improvement of our proposed model. The dimensions of the fully connected layers is 1200. The configuration of this model is the same as that of our proposed model, as specified in Section 4.2.

4.4. Results and analysis

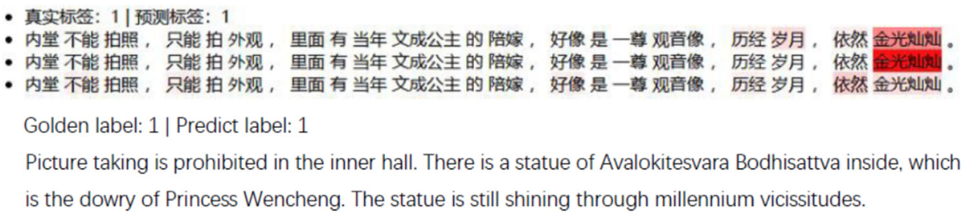
(1) Implicit sentiment analysis task

The implicit sentiment analysis task requires the model to identify the sentences that contain implicit sentiments and to classify their sentiment polarities. The performances of the experimental models are listed in Table 3. The results in the table are the averages of 5 repeated experiments. F-neutral, F-positive, and F-negative are the F1 scores of sentences with neutral, positive, and negative sentiment polarities, respectively as calculated via formula (27).

In Table 3, att denotes attention, static or dynamic-elmo/bert refers to the input being initialized by a static pretrained word embedding or being repretrained by Elmo[15]/ Bert[16] dynamically in the pretraining layer, respectively.

The results demonstrate that our proposed model substantially outperformed the baseline models in this task. Compared to the basic GRU, LSTM, BiLSTM with attention, BiLSTM with multi-attention, our proposed models can more accurately capture the differences among polarities in the sentence. Our proposed models realize improvements of 4.5% and 4.9% over the BiLSTM+att model and the BiLSTM+multi-att model, respectively, which demonstrates

[9] https://ai.tencent.com/ailab/nlp/embedding.html.
[10] http://ir.dlut.edu.cn/EmotionOntologyDownload.
[11] https://nlp.stanford.edu/projects/glove/.
[12] http://www.saifmohammad.com/WebPages/SCL.html.
[13] https://github.com/gaussic/text-classification-cnn-rnn.
[14] https://github.com/gao-g/metaphor-in-context.
[15] https://github.com/HIT-SCIR/ELMoForManyLangs.
[16] https://github.com/dbiir/UER-py.



- 真实标签：1 | 预测标签：1
- 内堂 不能 拍照， 只能 拍 外观， 里面 有 当年 文成公主 的 陪嫁， 好像 是 一尊 观音像， 历经 岁月， 依然 金光灿灿 。
- 内堂 不能 拍照， 只能 拍 外观， 里面 有 当年 文成公主 的 陪嫁， 好像 是 一尊 观音像， 历经 岁月， 依然 金光灿灿 。
- 内堂 不能 拍照， 只能 拍 外观， 里面 有 当年 文成公主 的 陪嫁， 好像 是 一尊 观音像， 历经 岁月， 依然 金光灿灿 。

Golden label: 1 | Predict label: 1

Picture taking is prohibited in the inner hall. There is a statue of Avalokitesvara Bodhisattva inside, which is the dowry of Princess Wencheng. The statue is still shining through millennium vicissitudes.

Fig. 2. Attention Visualization Example1.

Table 4
The performance of explicit sentiment analysis.

Dataset	COAE		SemEval	
Method	ACC	F-macro	ACC	F-macro
GRU	0.813	0.702	0.658	0.647
LSTM	0.789	0.667	0.661	0.646
BiLSTM+att	0.846	0.705	0.692	0.688
BiLSTM+multi-att	0.845	0.711	0.695	0.690
proposed model(dynamic-elmo)	0.808	0.684	0.669	0.663
proposed model(static)	0.843	0.720	0.704	0.698
proposed model(dynamic-bert)	**0.887**	**0.778**	**0.711**	**0.701**

Table 5
The performances of different pretraining layers.

Method	F-neutral	F-positive	F-negative	F-macro
BILSTM+att(random)	0.771	0.562	0.683	0.671
BiLSTM+att(dynamic-elmo)	0.774	0.566	0.694	0.664
BiLSTM+att(static)	0.810	0.634	0.747	0.730
BiLSTM+att(dynamic-bert)	0.827	0.686	0.793	0.769
proposed model(random)	0.764	0.569	0.691	0.674
proposed model(dynamic-elmo)	0.804	0.523	0.712	0.679
proposed model(static)	0.804	0.644	0.752	0.733
proposed model(dynamic-bert)	**0.834**	**0.690**	**0.802**	**0.775**

Table 6
The effectiveness of multi-polarity and orthogonal attention mechanism

Method	F-neutral	F-positive	F-negative	F-macro
BMA	0.800	0.630	0.749	0.726
BMA+pre_ort	0.803	**0.646**	0.747	0.732
BMA+pre_ort+pun	0.805	0.630	0.749	0.728
BMA+pol-untrain	0.805	0.616	0.749	0.723
BMA+pol-train	**0.807**	0.629	0.749	0.727
proposed model	0.804	0.644	**0.752**	**0.733**

the effectiveness of multi-polarity and the orthogonal attention mechanism.

(2) Explicit sentiment analysis task

To validate the effectiveness and generalization of our model, we conducted further experiments on two datasets of explicit sentiment analysis, COAE and SemEval. The performances of the experimental models are listed in Table 4.

Compared to the basic baselines, the multi-polarity and orthogonal attention mechanism also benefit the explicit sentiment analysis task significantly. Our proposed models realize improvements of 6.7% and 1.1% over the BiLSTM+multi-att model in F-macro on the two datasets, respectively.

(3) Comparison among pretraining layers

Since the pretraining layer shows great improvement and may cover up the effectiveness of our proposed multi-polarity and orthogonal attention mechanism. We conducted a further comparison between our proposed methods and BiLSTM+attention models with the same pretraining layer. Considering the dynamic and static pretraining layers, we present the results for comparison in Table 5.

Random refers to random initialization of the word embedding for input. According to Table 5, using the dynamic Elmo pretraining layer, the proposed model realizes an improvement of 0.4% (line 6-line 2) over the baseline model BiLSTM+att. A much larger improvement (line 8-line 4) is realized if we adopt the dynamic Bert pretraining layer. Using a static pretrained word embedding that is well-trained on a huge corpus with wide coverage will yield more accurate results than dynamic repretraining. Moreover, the dynamic Elmo realizes higher performance on the basic and simple BiLSTM+att model. This may be due to the sizes of the parameters and the corpus. The proposed model with multi-polarity attention, which contains many more parameters than the baseline model, the experimental dataset may not be sufficient for Elmo to obtain well-trained representations.

(4) The effects of multi-polarity and the orthogonal attention mechanism

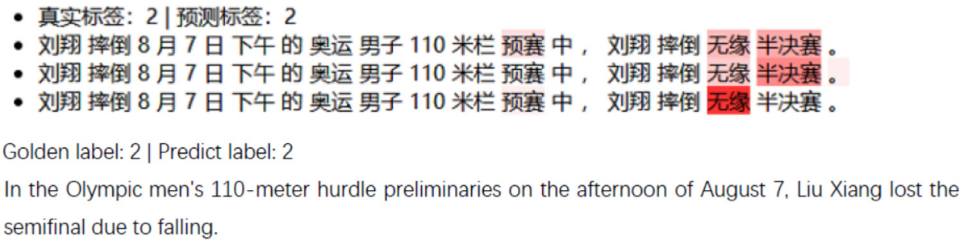
The proposed model uses the sentiment words with various orientations to initialize sentiment-polarity-specific attentions. The orthogonal attention mechanism can preserve the differences among sentiment polarities during model optimization. Table 6 presents the multi-polarity and orthogonal attention mechanism performance results. The results in the table are the averages of 5 repeated experiments.

In Table 6, BiLSTM with the multi-attention model is denoted as BMA. pre_ort denotes that we randomly and orthogonally initialize the multi-attention queries' embeddings. pun denotes that we add orthogonal punishment during the model optimization. pol denotes that we use sentiment words with various orientations to initialize sentiment-polarity-specific attentions. If the attention queries' embedding can be updated during the training process, it is denoted as train; otherwise, it is denoted as untrain.

For BMA, BMA+pre_ort, and BMA+pre_ort+pun, imposing an orthogonality constraint on multi-attention can effectively preserve the differences among attentions during the model optimization. An orthogonal initialization can improve the performance by 0.6%(line 2-line 1). In terms of the sentiment polarity specific attentions, BMA+pol-train without the orthogonality constraint can realize improvements of 0.1%(line 5-line 1) compared to BiLSTM with the multi-attention model. The initial sentiment clue is powerful for the implicit sentiment analysis task.

(5) Visualization and instance analysis

To further evaluate the performance of our proposed model, we visualize the model's multi-polarity attentions in the experiment, as shown in Figs. 2 and 3. The first line corresponds to the real and predicted labels and the last three lines in each figure correspond to the neutral, positive, and negative attention information, respectively. We color the words in each sentence with various color depths according to the attention weight: the heavier the weight, the darker the color. In the attention visualization, the multi-polarity attention differentiates feature among sentiment polarities. As shown in Fig. 2, the positive attention is focused on "金光灿烂 (the light is shining)", while the neutral and negative attentions are associated with much lower weights. In Fig. 3, negative attention is focused on on "无缘" (no opportunity to participate)",



- 真实标签：2 | 预测标签：2
- 刘翔 摔倒 8 月 7 日 下午 的 奥运 男子 110 米栏 预赛 中， 刘翔 摔倒 无缘 半决赛。
- 刘翔 摔倒 8 月 7 日 下午 的 奥运 男子 110 米栏 预赛 中， 刘翔 摔倒 无缘 半决赛。
- 刘翔 摔倒 8 月 7 日 下午 的 奥运 男子 110 米栏 预赛 中， 刘翔 摔倒 无缘 半决赛。

Golden label: 2 | Predict label: 2

In the Olympic men's 110-meter hurdle preliminaries on the afternoon of August 7, Liu Xiang lost the semifinal due to falling.

Fig. 3. Attention Visualization Example2.

while the other attentions focus on other words. This is consistent with our intuition and supports our model hypothesis.

5. Conclusion

In this paper, we proposed a new BiLSTM model with multi-polarity orthogonal attention for implicit sentiment analysis.

For the expression of an implicit sentiments, which are obscure and contain no explicit sentimental words, the proposed model focuses on modeling the differences in the attentions' weights among polarities. The multi-polarity attention mechanism uses the sentiment words to initialize polarity-specific attention queries and orthogonal punishment is introduced into the model during the optimization process to preserve the differences among attentions. The experimental results on the 2019 SMP-ECISA Chinese implicit sentiment analysis evaluation dataset demonstrate that our method realized approximately 4.9% improvement over the baseline model. Further experiments on the COAE and SemEval datasets for explicit sentiment analysis task validate the generalization of our proposed model.

In the future, we will continue the research from the following aspects:

(1) We will introduce external knowledge bases (a commonsense knowledge base, for example) into the model. External knowledge is typically implicit; hence, it is not specified in the sentences or by the context directly. We must identify a suitable method for representing external knowledge.

(2) We will explore semantic inference mechanism-based implicit sentiment identification methods. This aspect is the combination of the former aspect and semantic modeling. Future research may be fruitful from the perspectives of commonsense-based inference and cognitive conflict.

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.

CRediT authorship contribution statement

Jiyao Wei: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization. **Jian Liao:** Conceptualization, Methodology, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition. **Zhenfei Yang:** Software, Validation, Investigation, Visualization. **Suge Wang:** Resources, Supervision, Project administration, Funding acquisition. **Qiang Zhao:** Software, Validation, Visualization.

Acknowledgements

The authors would like to thank all anonymous reviewers for their valuable comments and suggestions which have significantly improved the quality and presentation of this paper. The works described in this paper are supported by the National Natural Science Foundation of China (61906112, 61632011, 61573231, 61672331, 61603229), Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi (2019L0008), the Key Research and Development Program of Shanxi Province (201803D421024).

References

[1] B. Liu, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Cambridge University Press, 2015, doi:10.1017/CBO9781139084789.

[2] B. Liu, Sentiment Analysis and Opinion Mining, Morgan & Claypool Publishers, 2012.

[3] E. Cambria, S. Poria, A. Gelbukh, M. Thelwall, Sentiment analysis is a big suitcase, IEEE Intelligent Systems 32 (6) (2017) 74–80.

[4] D. S, K.H. K, Sentiment analysis for web-based big data: A survey, International Journal of Advanced Research in Computer Science 8 (5) (2017) 1996–1999.

[5] J. Liao, S. Wang, D. Li, Identification of fact-implied implicit sentiment based on multi-level semantic fused representation, Knowledge-Based Systems 165 (2019) 197–207, doi:10.1016/j.knosys.2018.11.023.

[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] H.-Y. Chen, H.-H. Chen, Implicit polarity and implicit aspect recognition in opinion mining, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 2: Short Papers, The Association for Computer Linguistics, 2016. http://aclweb.org/anthology/P/P16/P16-2004.pdf.

[8] L. Jian, L. Yang, W. Suge, The constitution of a fine-grained opinion annotated corpus on weibo, in: Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data, Springer, 2016, pp. 227–240.

[9] 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.

[10] Y. Wang, M. Huang, L. Zhao, et al., Attention-based lstm for aspect-level sentiment classification, in: Proceedings of the 2016 conference on empirical methods in natural language processing, 2016, pp. 606–615.

[11] Y. Yao, S. Wang, X.U. Ruifeng, B. Liu, L. Gui, L.U. Qin, X. Wang, The construction of an emotion annotated corpus on microblog text, Journal of Chinese Information Processing(In Chinese) 28 (5) (2014) 83–91.

[12] L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, K. Xu, Adaptive recursive neural network for target-dependent twitter sentiment classification, in: ACL, 2014.

[13] S. Kiritchenko, X. Zhu, S.M. Mohammad, Sentiment analysis of short informal texts, Journal of Artificial Intelligence Research 50 (2014) 723–762.

[14] L. Deng, J. Wiebe, Mpqa 3.0: An entity/event-level sentiment corpus, in: Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2015, pp. 1323–1328.

[15] L. Deng, J. Wiebe, Sentiment propagation via implicature constraints, in: Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, 2014, pp. 377–385.

[16] E. Shutova, S. Teufel, A. Korhonen, Statistical metaphor processing, Computational Linguistics 39 (2) (2013) 301–353.

[17] D. Zhang, H. Lin, L. Yang, S. Zhang, B. Xu, Construction of a chinese corpus for the analysis of the emotionality of metaphorical expressions, in: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 2018, pp. 144–150.

[18] M. Van de Kauter, D. Breesch, V. Hoste, Fine-grained analysis of explicit and implicit sentiment in financial news articles, Expert Systems with applications 42 (11) (2015) 4999–5010.

[19] E. Shutova, S. Lin, E.D. Gutiérrez, P. Lichtenstein, S. Narayanan, Multilingual metaphor processing: Experiments with semi-supervised and unsupervised learning, Computational Linguistics 43 (1) (2017) 1–88.
[20] Y. Kim, Convolutional neural networks for sentence classification, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, 2014.
[21] N. Kalchbrenner, E. Grefenstette, P. Blunsom, A convolutional neural network for modelling sentences, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics, 2014.
[22] R. Socher, J. Pennington, E.H. Huang, A.Y. Ng, C.D. Manning, Semi-supervised recursive autoencoders for predicting sentiment distributions, in: Proceedings of the conference on empirical methods in natural language processing, Association for Computational Linguistics, 2011, pp. 151–161.
[23] R. Socher, A. Karpathy, Q.V. Le, C.D. Manning, A.Y. Ng, Grounded compositional semantics for finding and describing images with sentences, in: Transactions of the Association for Computational Linguistics, 2, 2014, pp. 207–218.
[24] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (1997) 1735–1780, doi:10.1162/neco.1997.9.8.1735.
[25] J. Chung, C. Gulcehre, K. Cho, Y. Bengio, Empirical evaluation of gated recurrent neural networks on sequence modeling, in: NIPS 2014 Workshop on Deep Learning, December 2014, 2014.
[26] A. Graves, N. Jaitly, A.-r. Mohamed, Hybrid speech recognition with deep bidirectional lstm, in: 2013 IEEE workshop on automatic speech recognition and understanding, IEEE, 2013, pp. 273–278.
[27] J. Wang, L.-C. Yu, K.R. Lai, X. Zhang, Dimensional sentiment analysis using a regional CNN-LSTM model, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Association for Computational Linguistics, Berlin, Germany, 2016a, pp. 225–230, doi:10.18653/v1/P16-2037.
[28] Y. Wang, M. Huang, X. Zhu, L. Zhao, Attention-based LSTM for aspect-level sentiment classification, in: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Austin, Texas, 2016b, pp. 606–615, doi:10.18653/v1/D16-1058.
[29] 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, Association for Computational Linguistics, Austin, Texas, 2016, pp. 247–256, doi:10.18653/v1/D16-1024.
[30] M.T. Luong, H. Pham, C.D. Manning, Effective approaches to attention-based neural machine translation, Computer Science (2015).
[31] 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 International Conference on Learning Representations, 2017.
[32] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Advances in neural information processing systems, 2017, pp. 5998–6008.
[33] C. Peng, Z. Sun, L. Bing, Y. Wei, Recurrent attention network on memory for aspect sentiment analysis, in: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, 2017.
[34] Q. Qian, M. Huang, X. Zhu, Linguistically regularized lstms for sentiment classification, in: Proceedings of the Annual Meeting of the Association for Computational Linguistics, 2017.
[35] M. Sun, Y. Yuan, F. Zhou, E. Ding, Multi-attention multi-class constraint for fine-grained image recognition, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 805–821.
[36] S. Song, C. Lan, J. Xing, W. Zeng, J. Liu, An end-to-end spatio-temporal attention model for human action recognition from skeleton data, in: Thirty-first AAAI conference on artificial intelligence, 2017.
[37] J. Liu, G. Wang, L.-Y. Duan, K. Abdiyeva, A. Kot, Skeleton-based human action recognition with global context-aware attention lstm networks, IEEE Transactions on Image Processing 27 (2018) 1586–1599, doi:10.1109/TIP.2017.2785279.
[38] M.E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, Deep contextualized word representations, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2018.
[39] J. Devlin, M. Chang, K. Lee, K. Toutanova, BERT: pre-training of deep bidirectional transformers for language understanding, 2018 CoRR abs/1810.04805.
[40] L. Xu, H. Lin, Y. Pan, H. Ren, J. Chen, Constructing the affective lexicon ontology, Journal of the China Society for Scientific and Technical Information(In Chinese) 27 (2) (2008) 180–185.
[41] W. Zhang, Y. Li, S. Wang, Learning document representation via topic-enhanced lstm model, Knowledge-Based Systems 174 (2019) 194–204.
[42] G. Gao, E. Choi, Y. Choi, L. Zettlemoyer, Neural metaphor detection in context, in: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 2018.

Jiyao Wei is an undergraduate of School of Computer and Information Technology, Shanxi University, China. His research interests include sentiment analysis and knowledge graph.

Jian Liao, Ph.D., lecturer in School of Computer and Information Technology, Shanxi University, China. His research interests include implicit sentiment analysis and representation learning.

Zhenfei Yang is an undergraduate of School of Computer and Information Technology, Shanxi University, China. His research interests include implicit sentiment analysis and text classification.

Wang Suge, Ph.D., professor and Ph.D. supervisor in School of Computer and Information Technology, Shanxi University, China. Her main research interests include natural language processing and intelligent retrieval.

Qiang Zhao is an undergraduate of School of Computer and Information Technology, Shanxi University, China. His research interests include sentiment analysis and text summary.