IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 12, 2020, accepted January 21, 2020, date of publication January 27, 2020, date of current version February 4, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2969473

Aspect-Level Drug Reviews Sentiment Analysis Based on Double BiGRU and Knowledge Transfer

YUE HAN[1], MEILING LIU[1,2], (Member, IEEE), AND WEIPENG JING[1], (Member, IEEE)

[1]School of Information and Computer Engineering, Northeast Forestry University, Harbin 150006, China
[2]College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China

Corresponding author: Meiling Liu (lmling2008@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61702091, and in part by the Fundamental Research Funds for the Central Universities under Grant 2572018BH06.

ABSTRACT Aspect-level sentiment analysis is a fine-grained sentiment analysis task designed to identify the sentiment polarity of specific target in a sentence. However, this task is rarely used in drug reviews. Some models for this task ignore the impact of target semantics, and others do not perform well because the datasets are relatively smaller. Therefore, we propose a Pretraining and Multi-task learning model based on Double BiGRU (PM-DBiGRU). In PM-DBiGRU, we first use the pretrained weight learned from short text-level drug review sentiment classification task to initialize related weight of our model. Then two BiGRU networks are applied to generate the bidirectional semantic representations of the target and drug review, and attention mechanism is used to obtain the target-specific representation for aspect-level drug review. The multi-task learning is further utilized to transfer the helpful domain knowledge from the short text-level drug review corpus. We also propose a dataset SentiDrugs for aspect-level drug review sentiment classification, in which each review may contain one or more targets. Experimental results on SentiDrugs demonstrate that our approach can improve the performance of aspect-level drug reviews sentiment classification compared with other state-of-the-art architectures.

INDEX TERMS Aspect-level, drug reviews, double BiGRU, pretraining, multi-task learning.

I. INTRODUCTION

Sentiment analysis is an important task in NLP which main purpose is to identify people's sentiments, opinions and attitudes of products, services, individuals, organizations and other entities [1]. Aspect-level sentiment analysis is a fine-grained sentiment analysis task that aims to analyze the sentiment polarity of specific aspect in its sentence [2]. For example, given a sentence "*This drug works great for **water retention**, but **its side effect** is severe*" the sentiment polarity for aspects "water retention" and "its side effect" are positive and negative respectively.

Aspect-level sentiment classification is usually provided with domain dependence, which main reason is that a word may have different sentiment polarity due to the different contexts it appears [3]. At present, this task has been widely used in film reviews, e-commerce and other fields. Therefore, it has received a rising concern of researchers [4], [5]. However, the studies of aspect-level sentiment analysis based on drug reviews are very limited. Text mining [6] for sentiment analysis in medical internet data center [7] has many practical application values, for example, in drug recommendation systems [8], post-marketing monitoring, understanding of patients' treatment opinions and sentiments, and finding adverse drug reactions [9]. Most of the sentiment analysis approaches in the field of medical social media are rule-based and machine learning. These conventional approaches often need to design rules and extract handcraft features (such as sentiment lexicon and bag-of-words features) to train a classifier [10]–[13]. However, it is often complicated to design rules and extract features. In addition, because different fields have different language rules and features, classifiers are somewhat susceptible to them.

As the development of neural networks in the field of NLP, the approaches based on neural networks(such as LSTM [14] and GRU [15]) have been applied in many fields of aspect-level sentiment analysis classification to obtain a promising result. Interaction between sentiment words, target, degree words and negative words is very important in aspect-level sentiment classification. Bidirectional neural networks

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wei.

(BiLSTM, BiGRU) can help bring more benefit in terms of better results to those domains where it is appropriate. There are also some problems with neural network models. Jiang *et al.* [16] find that 40% of the errors in aspect-level sentiment analysis are due to the fact that no aspects are considered. For example, the sentiment polarity of "*I had no* ***side effects****, however* ***the infection*** *did not clear up*" will be positive if the target is "side effects" but negative when considering the target "the infection". In this case, it is easy to cause sentiment classification error with ignoring the target words.

Many attention-based [17]–[25] methods have been proposed to improve the performance of aspect-level sentiment classification by generating target-specific representations. However some targets are relatively long and have certain sentimental feature in drug reviews. For instance in the sentence "*apparently lowered* ***cholesterol and blood pressure***", the target "cholesterol and blood pressure" is long and its semantics determines that the sentiment words "apparently lowered" is positive. On the contrary, "low" always represents negative sentiment in other fields. For instance, in the Laptop dataset [26], in sentence "*The wireless card is low* ***quality***", the polarity of the target "quality" is negative. Similarly, in the Restaurant dataset [26], "*If you don't mind pre-sliced low* ***quality fish****, ...*",the polarity of target "quality fish" is also negative. Therefore, the semantics of targets are crucial for aspect-level drug reviews sentiment classification. Generally, the attention-based approaches all need to be trained on a large-scale dataset to achieve better results. The targets and sentiment categories in the aspect-level dataset are always generated by manual annotating, but annotating large-scale data takes a lot of time and labors. Therefore, the existing public aspect-level datasets are all relatively small. Despite the lack of aspect-level annotated corpus, document level annotated datasets are larger and easier to obtain. He *et al.* [27] propose a method of transferring knowledge from document-level sentiment classification tasks to improve the aspect-level sentiment classification performance. However, the positive impact of the target semantics on sentiment classification results is ignored in [27]. Some datasets (such as in [13], [28] and [29]) are introduced for the sentiment analysis task to mine the sentiment and opinions in medical social media. Most of these datasets are document or sentence-level, making it impossible to conduct more fine-grained sentiment analysis.

Therefore, we introduce the dataset SentiDrugs for the task of aspect-level drug reviews sentiment analysis. We also propose a Pretraining and Multi-task learning model based on Double BiGRU(PM-DBiGRU) for this task. We use two independent BiGRU networks to learn the bidirectional semantics of target and sentence to emphasize the importance of target, and then utilize attention mechanism to capture the important information in review. Furthermore, we use pretraining and multi-task learning to transfer domain knowledge from short text-level task to address challenges because of the limited aspect-level data availability.

The main contributions of this work are summarized as follows:

1) We introduce the dataset SentiDrugs. Each review involves one or more targets with sentiment polarity, which can be divided into three categories: negative, neutral and positive. There are more than 4,000 manually annotated reviews, which lays a good foundation for fine-grained drug review sentiment analysis.

2) We propose a new model PM-DBiGRU. It firstly transfers the learned weights pretrained on the short text-level drug review corpus to the aspect-level task hierarchically. Then it uses two BiGRU networks to learn the hidden semantics of sentence and target to play the important role of target, and utilizes attention mechanism to generate powerful target-specific representations for sentences. Moreover, multi-task learning can bring the benefits for aspect-level task by incorporating knowledge.

3) Experiments results show that the SentiDrugs can carry out effective study on aspect-level drug reviews sentiment analysis. Besides, the PM-DBiGRU model can improve the performance compared with several baselines, indicating that our model can fully employ target's semantics representation. We also illustrate that pretraining and multi-task learning can transfer beneficial knowledge from short text-level tasks.

II. RELATED WORK

Aspect-level sentiment analysis is a branch of sentiment analysis, which has been applied in many domains. It is of great significance to explore the sentiment tendency of text in medical social media. Therefore, relevant datasets for text sentiment analysis are proposed. Ali *et al.* [28] introduce a sentiment analysis dataset in the domain of the hearing loss to analyze users' opinions expressed on medical forums. Jiménez-Zafra *et al.* [29] establish two different corpora about drug and doctor reviews, DOS and COPOS. The experiment results demonstrate that drug reviews are more difficult to classify than those about doctor because drug reviews contain more specific expression. Gräßer and Kallumadi [13] propose two datasets, Drugs.com and Druglib.com, by obtaining text from two web pages about drug reviews. The Drugs.com contains 161,297 training reviews and 53,766 testing reviews, each of which is a short text contained one or more short sentences. The sentiment polarity can be divided into three categories: negative, neutral, and positive. There are 6,214 training reviews and 2,072 testing reviews in Druglib.com, which includes two aspects, effectiveness and side effects, as well as three categories of sentiment polarity. Although the corpora in [13] and [29] are larger than those in [29], none of them can be used for fine-grained sentiment classification because the specific target and corresponding sentiment polarity are not annotated in the dataset. Therefore, it is impossible to perform aspect-level sentiment analysis according to specific aspects.

Traditional aspect-level sentiment classification methods try to solve this problems by making complex rules or manually extracting features, and then train a sentiment classifiers, such as Random Forest, HMM [30], DCU [31]. Unfortunately, texts in different domain often have different linguistic rules and features, and classifiers are susceptible to them, which in turn require expensive costs to improve the performance of classifiers.

Deep neural networks can automatically learn features of text and generate its powerful vector representation. Tang *et al*. [17] propose to split the sentence into two parts and concatenate target with contextual features as the representation for prediction. However, the neural networks based models only focus on modeling the sentences without consideration of targets which make great contributions for prediction. Attention mechanism can effectively generate target-specific representations for sentences. Wang *et al*. [18] propose attention-based LSTM to obtain different sentiment information for given target in sentence. Ma *et al*. [19] propose an interactive attention network(IAN), which learns the interactions between sentence and target. Similar to IAN, Huang *et al*. [20] propose the AOA module that can learn word-pairs interaction between sentences and targets. Tang *et al*. [21] develop the MemNet model, which applies attention mechanism over the word embeddings multiple times. RAM [22] is a multi-layer architecture where each layer consists of attention-based aggregation of word features and a GRU cell to learn the sentence representation. Cabasc [23] employs sentence-level content attention mechanism to capture the important information about given targets from a global perspective. Fan *et al*. [24] integrate fine-grained attentions to compose the multi-grained attention network. Li *et al*. [25] employ a CNN layer to extract salient features and propose a novel target-specific transformation component to better integrate target information into the word representations.

However, those models perform better on large-scale datasets. Insufficient training data limits the effectiveness of neural models. The experimental results in [32] show that the transferability of neural model in NLP is largely dependent on the semantic relatedness between source and target task. Therefore, He *et al*. [27] explore two approaches for aspect-level sentiment classification, pretraining and multi-task learning based on LSTM and attention, which transfer knowledge from document-level data obtained less expensive. However, in [27], the semantic of the target phrase is represented simply averaging the word embedding which ignores the influence of it on the classification results especially in the drug review, leading to suboptimal performance.

Target plays an important role in capturing important sentiment information in the sentences in [18]–[25], and these models have better classification results than the LSTM/GRU-only model. However, the effectiveness of neural model largely limits by the scale of dataset. He *et al*. [27] solve this problem by pretraining and multi-task learning, but this model only represents target as the average value of word embedding, ignoring the impact of target semantics on classification results. In addition, LSTM has more training parameters than GRU and cannot capture the bidirectional semantic representation of sentences and targets. In this paper, we establish an aspect-level drug review dataset, which can be used for more fine-grained sentiment analysis. We assume that aspect-level sentiment classification can learn helpful knowledge from short text-level sentiment classification. Different from [18]–[25], we use pretraining and multi-task learning to transfer domain knowledge from short text-level drug reviews sentiment classification. Different from model in [27], we utilize double BiGRU modeling for drug reviews and targets respectively to obtain bidirectional semantic information, which makes up for the missing information of targets represented as the average of word embedding. Besides, GRU is more efficient than LSTM in training.

III. DATASET

In this section, we describe the proposed dataset for aspect-level drug reviews sentiment analysis task, called SentiDrugs.

A. DATA SOURCES

SentiDrugs is an aspect-level sentiment analysis dataset based on the Druglib.com proposed by Gräßer and Kallumadi [13]. We randomly select 4,200 reviews on effectiveness and side effects in Druglib.com, in which longer than 200-word reviews are removed. Then one or more targets and corresponding sentiment polarities are manually annotated for each review by several annotators. Targets include side effects, effectiveness, degree of improvement, severity of symptoms, and changes in organs and mood. For each selected target, sentiment polarity may be one of positive, negative and neutral.

B. ANNOTATION GUIDELINES

1) TARGET TERM

(1) Nominal phrases explicitly mentioning targets in drug review. One or more targets may appear in a review.

The patient felt mild ***headaches****, in addition there were mild* ***symptoms of nausea****.*

Significantly decreased ***depression and suicidality****, improved sleep, more stable* ***moods*** *and increased* ***energy****.*

(2) Verbs or verbals (words formed from a verb, e.g., participles and gerunds) naming targets, like “thinking” and “rested” below.

I am ***thinking*** *more clearly and don’t feel so down.*

I could fall asleep and was usually well ***rested****.*

(3) Subjectivity indicators should not considered target terms or parts of target terms; annotate all targets in the review as much as possible (even if they are misspelled).

2) SENTIMENT CATEGORY

Referring to the classification criteria in the dataset Druglib.com [13], we classify sentiment polarity into three categories: negative (-1), neutral (0), and positive (1).

A target term should be classified positive or negative if the review expresses a positive or negative evaluation, emotion, experience, influence, change and degree, as shown in the following example.

skin *looked healthier with fewer break outs and was smoother. I noticed* ***pores*** *becoming smaller. ["****skin****": positive; "****pores****": positive]*

When the review uses "effective", "apparent" and "consistent" to describe the reduction, reduction or improvement of symptoms, or when the symptoms "stop", "clear", "curved" and "gone", the sentiment polarity is positive, for example:

Effective at ***controlling seizures*** *without dizziness. ["****controlling seizures****": positive]*

When "series", "extreme" and "horrific" are used to describe symptoms, organs, side effects, etc., the sentiment polarity is negative.

This medication plunged me into experiencing menopausal ***symptoms****.* ***My skin*** *became extremely dry. ["****symptoms****": negative; "****My skin****": negative]*

Target terms should be annotated as neutral in the following cases:

(1)When factual information (no sentiment) about the target term is provided.

My cholesterol *was lowered by about 15 points. ["****My cholesterol****": neutral]*

(2) When expression like "mild", "moderate", "normal", "slight", "some", etc. are used, for example:

There is a mild ***burning*** *in my eyes for a couple of minutes after the drops are out in. ["****burning****": neutral]*

C. PROCEDURES

First of all, we carry out the target annotating work. Three annotators are initially selected for the task. Each annotators is then required to annotate a small subset of the data and to annotate the target independently for each drugs review according to the annotation guide. After each round of annotation, the annotators discuss and reach a reasonable agreement until all targets of the drugs review are annotated.

Secondly, the sentiment polarity of each target is annotated. Before annotating the sentiment polarity of all reviews, we randomly select 400 sentences to be annotated by three annotators. Then, we use Cohen's kappa coefficient [33] to verify the effectiveness of annotating, which is usually used for measuring the pairwise agreement between two each two annotators for aspect-level sentiment analysis task [34], [35] and other tasks [36]. Generally, the Kappa value is (0, 1). The higher kappa value, the higher agreement. The kappa value is calculated over target-sentiment pairs. Pairwise inter-annotator agreement for target categories measured using kappa value is 0.69, 0.70 and 0.73, which is considered to be high quality.

Finally, the most consistent annotator is selected to annotate the whole dataset, which contains 4,028 target-sentiment pairs.

D. ANALYSIS

We take some examples in the SentiDrugs to get a more comprehensive and detailed understanding, as shown in Table 1. By analyzing these samples, we can find that:

1) There is only one target in sentence 1, and the sentence is relatively short. On the contrary, the sentences 2, 3, 6 are relatively longer and contain professional expression in the medical field.

2)There are multiple targets in sentences 2-6. The sentiment polarity of different target in a review is the same in sentence 2, 3, and 4, but different in sentence 5 and 6. In sentence 1, 2, 4, and 6, the length of the target is greater than 3 words. In this case, it is necessary to determine which word is more important for classification results.

3) For example, "this drug" in sentence 3, "my joy" in sentence 4 and two targets in sentence 5, their sentiment polarity is mainly expressed by verbs.

4) For judging the sentiment polarity of "this drug" in sentence 3, it is necessary to infer from context that "lowering cholesterol" is generally positive for patients. Similarly, when judging the sentiment polarity of "side effects" in sentence 4, we require to consider that "side effects"

TABLE 1. Drug reviews examples in SentiDrugs.

ID	Drug reviews examples
1	[**My side effects**]$_{neu}$ were not noticeable and very mild.
2	Noticeable [**improvement**]$_{pos}$ within 24 hours. Marked [**improvement**]$_{pos}$ within three days. Cleared [**sinus infection and bronchitis**]$_{pos}$ in the ten day prescription time period.
3	[**This drug**]$_{pos}$ lowered my cholesteral considerably. In addition, I have experienced NO [**side effects**]$_{pos}$ whatsoever, so I am pleased.
4	[**The effexor**]$_{neg}$ actually makes me feel worse. [**my emotions**]$_{neg}$ are blunted and flat. [**my joy for living**]$_{neg}$ is gone. i'm having some concentration problems, and short term memory loss. so i'm going back to dr. to see what else is out there.
5	[**Anxiety**]$_{neu}$ is suppressed but not gone. [**Depression**]$_{pos}$ is 100 percent gone.
6	This drug works great for [**water retention**]$_{pos}$ because it is a potassium-sparing diuretic. However, it had only a moderate [**impact on acne**]$_{neg}$. It did seem to help with premenstrual symptoms and cyclical acne, but not a complete cure by any means. No apparent [**impact on blackheads / hard plugs**]$_{pos}$.

have negative sentiment according to certain medical background knowledge, so we can know that when patients express "no side effects", the sentiment polarity conveyed is positive.

It can be seen that the dataset SentiDrugs proposed in this paper mainly has the following features: Drug reviews often contain expressions in some medical fields. Therefore, it is necessary to judge the sentiment polarity of a target with some common-sense knowledge, which is the most significant feature different from other aspect-level datasets; There are many such examples that the length of target is greater than 2, accounting for nearly half of the total dataset, as shown in Fig. 1. Sometimes, we need to consider the semantic of the target to judge the sentiment polarity, which requires the model to have a deeper understanding of the context; We found that some reviews often use verbs to express the patients' sentiment, not limited to adjectives, and the expression are diversified; Drug reviews often have a variety of words and sentence patterns. As shown in Fig. 2, there are many drug reviews with length greater than 30 (words). In this case, we need to combine the context to better judge the sentiment polarity of a specific target, which brings a challenge to achieve higher classification accuracy.

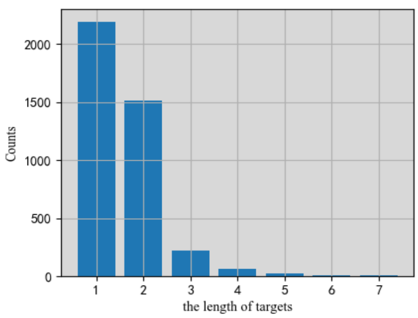


FIGURE 1. Distribution of the number of targets on different word length.

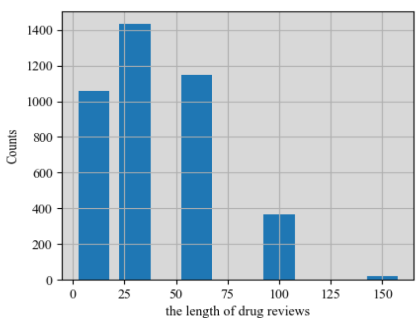


FIGURE 2. Distribution of the number of drug reviews on different word length.

SentiDrugs can be applied for more fine-grained sentiment analysis and other related research, laying a foundation for obtaining aspect-level sentiment information. We consider that SentiDrugs is a very important and challenging dataset.

IV. METHODOLOGY

In this aspect-level sentiment classification problem of drug reviews, we suppose that a sentence $s = \left[w_s^1, w_s^2, \cdots, w_s^n\right]$ consisting of n words and a target $x = \left[w_x^1, w_x^2, \cdots, w_x^m\right]$ which contains one or more consecutive words from s, the goal is classify the sentiment polarity towards the given target. We are given a sequence of words (usually, $l > n$) $t = \left[w_t^1, w_t^2, \cdots, w_t^l\right]$, aiming at classifying the whole sentiment polarity of short texts.

The overall architecture of PM-DBiGRU model is shown in Fig. 3, which is divided into embedding layer, DBiGRU layer, attention layer and softmax layer. The sentence s, the target x and the short text t share the same embedding layer to get their own word vectors, respectively. DBiGRU layer is composed of double BiGRU, one for learning the bidirectional semantic information of target x, the other for encoding contextual information of sentence s and short text t simultaneously. Attention layer is responsible for extracting the target information from the sentence and generating the target-specific representation. Finally, the sentiment polarity of aspect-level and short text-level are predicted separately by two softmax classifiers.The weights of DBiGRU layer and softmax layer are initialized by using the pretrained weights trained on the short text-level classification task in advance (the process of weight initialization is not shown in Fig.3). The pretrained model is the same as the task of short text-level classification in PM-DBiGRU model,which includes embedding layer, BiGRU layer and softmax layer.

A. EMBEDDING LAYER

Given an aspect-level drug review sentence $s = \left[w_s^1, w_s^2, \cdots, w_s^n\right]$, a targets $x = \left[w_x^1, w_x^2, \cdots, w_x^m\right]$, and a short text-level drug review context $t = \left[w_t^1, w_t^2, \cdots, w_t^l\right]$. w denotes a specific word. To represent a word, we map each word into d-dimensional vector $e^i \in \mathbb{R}^d$ (i is the word index in drug reviews or target) from an embedding matrix $E^{V\times d}$, such as glove [33]. And then the vector matrices of s, x and t are obtained respectively $\left[e_s^1; e_s^2; \cdots; e_s^n\right] \in \mathbb{R}^{n\times d}$,$\left[e_x^1; e_x^2; \cdots; e_x^m\right] \in \mathbb{R}^{m\times d}$ and $\left[e_t^1; e_t^2; \cdots; e_t^l\right] \in \mathbb{R}^{l\times d}$:

$$e^i = E(w) \tag{1}$$

B. DBiGRU LAYER

The main purpose of DBiGRU layer is to learn the hidden semantics of words in the aspect-level drug review sentence s, target x and short text-level drug review context t, and utilize the short text-level drug review sentiment classification task to learn the valuable domain knowledge of aspect-level drug review sentiment classification task. Therefore, compared to previous methods in [15]–[23] and [30], we not only use the separate BiGRU network to learn the hidden semantic representation of the target, but also use two methods to transfer the knowledge from the short text-level drug reviews which are the method of pretrained weight initializing the BiGRU weight and multi-task learning of sharing parameters of

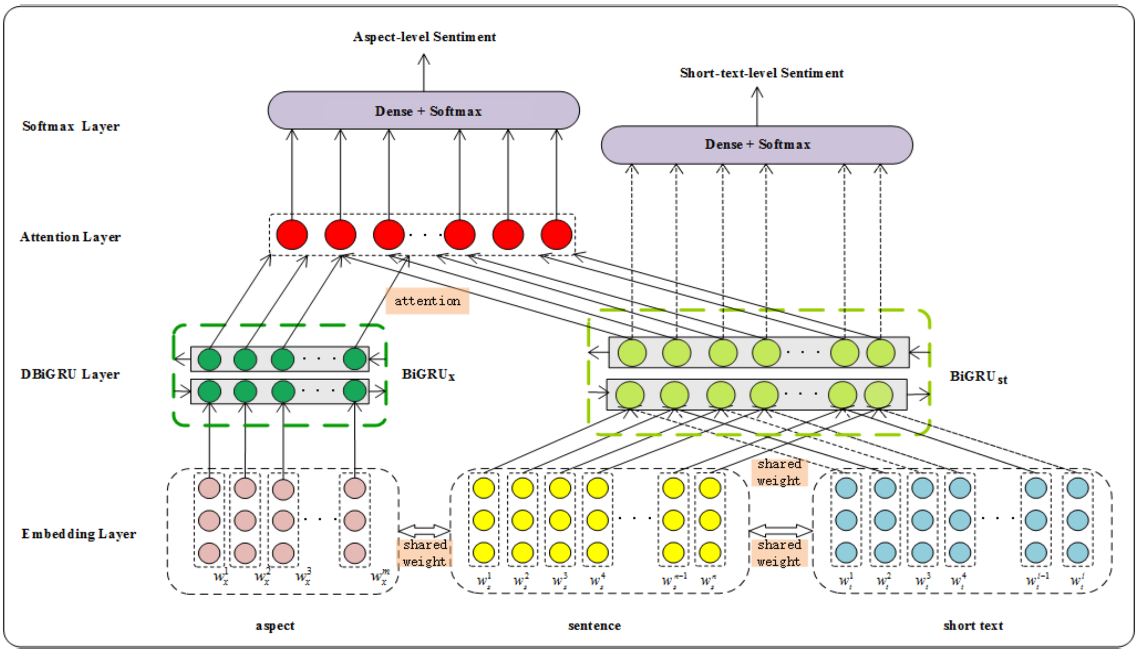


FIGURE 3. **The architecture of PM-DBiGRU Model.**

double BiGRU networks. In addition, BiGRU has less training parameters and higher training efficiency than BiLSTM.

We feed the word vectors of target $\left[e_x^1; e_x^2; \cdots; e_x^m\right]$ into a BiGRU network for learning the hidden semantics of words in the target x, which is denoted BiGRU_x. Similarly, we feed the word vectors $\left[e_s^1; e_s^2; \cdots; e_s^n\right]$ and $\left[e_t^1; e_t^2; \cdots; e_t^l\right]$ into another BiGRU network, and simultaneously capture the hidden semantics of words in the aspect-level drug review sentence s and short text-level drug review context t by sharing the weight. This BiGRU network is denoted as BiGRU_{st}. Each BiGRU is obtained by stacking two GRU networks.

With the word vectors of aspect-level drug review sentence $\left[e_s^1; e_s^2; \cdots; e_s^n\right]$ and a forward GRU_{st} network, we generate a sequence of hidden states $\overrightarrow{h_s} \in \mathbb{R}^{n\times d_h}$, where d_h is the dimension of hidden states. We generate another state sequence $\overleftarrow{h_s} \in \mathbb{R}^{n\times d_h}$ by feeding $\left[e_s^1; e_s^2; \cdots; e_s^n\right]$ into another backward GRU.

$$\overrightarrow{h_s} = [\overrightarrow{h_s^1}, \cdots, \overrightarrow{h_s^n}] = \overrightarrow{GRU_{st}}\left(\left[e_s^1; e_s^2, \cdots; e_s^n\right], \sigma_{st}\right) \quad (2)$$

$$\overleftarrow{h_s} = [\overleftarrow{h_s^1}, \cdots, \overleftarrow{h_s^n}] = \overleftarrow{GRU_{st}}\left(\left[e_s^1; e_s^2, \cdots; e_s^n\right], \sigma_{st}\right) \quad (3)$$

In the BiGRU network, the final output hidden state $h_s \in \mathbb{R}^{n\times d_h}$ is generated by summing $\overrightarrow{h_s}$ and $\overleftarrow{h_s}$:

$$h_s = \left[h_s^1, \cdots, h_s^n\right] = \left[\overrightarrow{h}_s^1 + \overleftarrow{h}_s^1, \cdots, \overrightarrow{h}_s^n + \overleftarrow{h}_s^n\right] \quad (4)$$

We simultaneously use the BiGRU_{st} to compute the hidden semantic states $h_t \in \mathbb{R}^{l\times d_h}$ for short text-level drug review context, where σ_{st} is the weight of the BiGRU_{st} network. Similarly, we use BiGRU_x to calculate the hidden semantic states $h_x \in \mathbb{R}^{m\times d_h}$ of target x, and the weight of BiGRU_x network is defined as σ_x. We use σ to represent the weight of BiGRU in the pretrained short text-level drug reviews sentiment analysis task. We initialize the patameters σ_{st} and σ_x with σ, and then train them on their respective examples to fine tune themselves.

Then the final representation z_t of short text-level drug reviews context is the average of the hidden vectors z_t output by the BiGRU_{st} network:

$$z_t = \frac{1}{l}\sum_{i=1}^{l} h_t^i \quad (5)$$

C. ATTENTION LAYER

The main purpose of attention layer is to understand the drug review sentences with respect to target and extract the important sentiment information to judging sentiment polarity. Given target representation h_x and aspect-level drug review sentence representation h_s, attention score $\alpha_i \in \mathbb{R}$ for each word representation h_s^i is computed as follows:

$$\tau = \frac{1}{m}\sum_{i=1}^{m} h_x^i \quad (6)$$

$$\beta_i = f_{\text{score}}\left(h_s^i, \tau\right) = \tanh\left(h_s^{iT} W_a \tau\right) \quad (7)$$

$$\alpha_i = \frac{\exp(\beta_i)}{\sum_{i=1}^{n}\exp(\beta_i)} \quad (8)$$

Different from [27], we calculate the attention score of each word by using the bidirectional hidden states vector h_x^i of the target, where τ is the average value of hidden states vector h_x, and f_{score} is a function to capture the semantic association between a word and target and its parameter matrix is $W_a \in \mathbb{R}^{d_h \times d_h}$. During the training, is randomly initialized.

The final target-specific representation of the aspect-level drug review sentence s is then given by:

$$z_s = \sum_{i=1}^{n} \alpha_i h_s^i \tag{9}$$

D. SOFTMAX LAYERR

In the softmax layer, we use two independent softmax classifiers to predict the sentiment polarity classification labels of aspect-level and short text-level drug review sentiment classification tasks, respectively. The aspect-level drug review sentence representation z_s is fed into a softmax classifiers to predict the probability distribution $\hat{y}_s \in \mathbb{R}^c$ of given target:

$$\hat{y}_s = \text{softmax}\left(W_s \mathbf{z}_s + b_s\right) \tag{10}$$

where c is the number of sentiment categories, $c = 3$ in this paper, $W_s \in \mathbb{R}^{c \times d_h}$ and $b_s \in \mathbb{R}^c$ are the weight matrix and bias respectively. Similarly, the vector expression z_t of the short text-level drug review context is fed into another softmax classifier to obtain the sentiment polarity distribution $\hat{y}_s \in \mathbb{R}^c$, where $W_t \in \mathbb{R}^{c \times d_h}$ and $b_t \in \mathbb{R}^c$ are weight matrix and bias respectively.

We represent $W \in \mathbb{R}^{c \times d_h}$ and $b \in \mathbb{R}^c$ respectively as the weight matrix and bias of softmax classifier in the pretrained short text-level drug review sentiment classification task. W_s, W_t and b_s, b_t are initialized with W and b respectively and fine-tuned in their respective examples.

E. MODEL TRAINING

In PM-DBiGRU model, we need update all parameters above. Let $\hat{y}$ denote the predicted sentiment distribution, and y denote the ground truth. Cross entropy between y and $\hat{y}$ with L_2 regularization is used as the loss function for aspect-level drug review sentiment classification task, which is defined as:

$$loss_{aspect} = -\sum_{i \in D} \sum_{j \in c} y_i^j log \hat{y_i^j} + \lambda ||\theta||^2 \tag{11}$$

where D denotes the overall training examples, c is the same as above. λ is the L_2 regularization parameter and θ is a set of weight matrices that consists of $\{\sigma_x, W_a, W_s, b_s\}$ and the word embeddings. Similarly, the loss function of short text-level sentiment classification task is $loss_{short-text}$, which is calculated in the same way as $loss_{aspect}$. The corresponding parameter set is the parameters of $\{\sigma_{st}, W_t, b_t\}$ the and the word embeddings.

The overall loss function is then given by:

$$loss = loss_{aspect} + \eta loss_{short-text} \tag{12}$$

where $\eta \in (0, 1)$ is a hyperparameter that controls the weight of $loss_{short-text}$.

We further apply dropout strategy to avoid overfitting before DBiGRU layer and softmax layer. In our experiments, we use RMSProp as our optimization method to minimize the loss function with respect to the parameters in our model.

V. EXPERIMENT

A. EXPERIMENT SETTING

1) DATASET

We conduct our aspect-level drug reviews sentiment classification experiments on SentiDrugs. Distribution by sentiment polarity category of are given in Table 2. We run the short text-level drug review sentiment classification experiments on the public dataset Drugs.com[1] [13] because it is relatively large, which helps the aspect-level task to transfer knowledge from it.

TABLE 2. Statistics of the SentiDrugs.

Total	4028		
Train (75%)	Positive	Neutral	Negative
	1324 (43.89%)	1085 (35.94%)	607 (20.15%)
Test (25%)	Positive	Neutral	Negative
	454 (44.86%)	317 (31.32%)	241 (23.81%)

2) EVALUATION METRICS

We adopt the Accuracy metric to evaluate the performance of aspect-level sentiment classification, which measures the percentage of correct predicted samples in all samples and is defined as:

$$Acc = \frac{T}{N} \tag{13}$$

where is the number of correctly predicted samples, is the total number of samples. Generally, a well performed model has a higher accuracy.

Since our task is multi-class classification task, we also adopt $Macro - F_1$ as our evaluation measure to provide more indicative information. The calculation formula is as follows:

$$Macro - F_1 = \sum_{k \in C} \frac{F_{1,k}}{|C|} \tag{14}$$

where $F_{1,k}$ is the F_1 of the k-th category, $|C|$ represents the number of categories.

3) TRAINING DETAILS

In the experiment, we randomly select 20% of training data from SentiDrugs as verification sets to tune the hyperparameters and only use the rest 80% for training. The word embedding for the aspect-level drug review sentences, target and short text-level drug review context in both datasets are initialized with 300-dimensional Glove[2] word vector [37]. The dimension of BiGRU hidden vectors is set to 300 and the L_2 regularization coefficient is set to 0.001. We use dropout probability 0.5 on sentence/short text representations before the each BiGRU and softmax layer. The initial learning rate

[1] https://archive.ics.uci.edu/ml/datasets.php
[2] https://nlp.stanford.edu/projects/glove/

is 0.0005 for the RMSProp, the batch size is set to 32, and η is set to 0.1. Our experimental results are obtained as average value over 5 runs with random initialization to solve the problem that the performance fluctuates with different random initialization parameters.

B. COMPARED METHODS

We compare PM-DBiGRU model with the following baseline approaches on SentiDrugs.

LSTM: It only uses one LSTM network to model the sentence. The average value of all hidden states is treated as final sentence representation.

BiGRU: It applies BiGRU network to learn the bidirectional semantic representations of the sentence.

TD-LSTM [17]: It uses two LSTM networks to model the left context and the right context with target respectively. The left and right target-dependent representations are concatenated for prediction.

ATAE-LSTM [18]: It first models the sentence via a LSTM and appends the target embeddings with into each word embedding vector to represent the sentences.

IAN [19]: It interactively learns the coarse-grained attentions between the sentence and target, and concatenate the vectors to classify.

AOA [20]: AOA utilizes two bidirectional LSTM network to model sentences and targets. Then it calculate a pair-wise interaction matric to attend the important sentiment information in both sentence and target.

MemNet [21]: It applies multi-hop attentions on the word embedding, and the last attention's output is fed to softmax function for predictions.

RAM [22]: It uses multi-hop attentions on the output of BiLSTM and proposes to use GRU network to get the aggregated vector from the attentions.

MGAN [24]: It employs a fine-grained attention mechanism, which can capture the word-level interaction between target and context. And then it leverages the fine-grained and coarse-grained attention mechanisms for prediction.

TNet [25]: TNet applies a CNN layer to extract informative features from the transformed word representations originated from a bidirectional LSTM layer.

Cabasc [23]: Cabasc utilizes a context attention mechanism in which the word order information, the aspect information and the correlation between them are modeled into the calculated attention weight.

PRET+MULT [27]: It incorporates knowledge from document-level corpus for aspect-level prediction.

Among all the baseline models, LSTM, BiGRU and TD-LSTM belong to neural network methods; ATAE-LSTM, IAN, AOA,MemNet,RAM,MGAN,TNet and Cabasc can be classified into attention-based methods; PRET+MULT are part of methods based on knowledge transfer.

We also list the variations of the proposed model, which is BiGRU-ATT, DBiGRU-ATT,P-DBiGRU, M-DBiGRU and PM-BiGRU to better understand the influence of each part of PM-DBiGRU model on the results.

BiGRU-ATT: Based on BiGRU, it utilizes the attention mechanism associated with a target to get important information from the sentence.

DBiGRU-ATT: It uses two bidirectional GRU network to model sentence and target respectively, and generates the target-specific representation for sentence by the attention mechanism.

DBiGRU-ATT: It uses two bidirectional GRU network to model sentence and target respectively, and generates the target-specific representation for sentence by the attention mechanism.

P-DBiGRU: Based on DBiGRU-ATT, the parameters of this model are initialized by the corresponding parameters of the pretraining model.

M-DBiGRU: DBiGRU-ATT model is adopted for aspect-level classification model,and the multi-task learning method of aspect-level and short-text level is employed.

PM-BiGRU: The difference between it and PM-DBiGRU is that targets are only represented as the average of word vectors, not as the hidden state of BiGRU network output.

C. OVERALL PERFORMANCE COMPARISON

Our experimental comparison results of PM-DBiGRU with other baseline methods are illustrated in Table 3. We can find that LSTM performs worst since it only treats every word in sentence equally. BiGRU outperforms LSTM which can capture bidirectional semantics information. TD-LSTM is better than LSTM and BiGRU. One main reason may be TD-LSTM adds target information by concatenating the left and right contexts with targets. However, the neural network based models cannot highlight the important sentiment words of specific target. Attention-based models can solve this problem by assigning different attention weights to different words in the drug review.

TABLE 3. Experimental results in accuracy and macro-f1 (%).

Model	Accuracy	Macro- F1
LSTM	71.15	69.47
BiGRU	71.35	69.88
TD-LSTM	72.04	70.18
ATAE-LSTM	72.83	70.81
IAN	74.21	72.74
AOA	75.01	73.57
MemNet	72.96	71.24
RAM	75.52	74.43
MGAN	76.24	75.02
TNet	76.47	75.10
Cabasc	75.58	74.69
PRET+MULT	76.79	75.36
Ours: PM-DBiGRU	**78.26**	**77.75**

Further, attention-based models work better because these models propose different methods to model target representation. ATAE-LSTM performs better than TD-LSTM for reason that it appends the target embedding into word vector and then uses attention mechanism to capture the sentiment words in drug reviews. IAN outperforms ATAE-LSTM, which main reason may be that an interactive attention is implemented on the representation of drug reviews and targets. AOA models targets and drug reviews by two BiLSTMs and then calculates a pair-wise interaction matrix to captures the interaction between them. Therefore, the AOA model achieves better classification result. Cabasc employs sentence-level content attention mechanism to capture the important information about given aspects from a global perspective. MGAN achieves better results than IAN, AOA and Cabasc by introducing the interactive attention in coarse-grained and multi-grained ways respectively. Both MemNet and RAM apply multi-hop attended vector on the memory. RAM bring more remarkable improvements than MemNet models because RAM employs bidirectional LSTM network to generate contextual memory. TNet adopts a target specific transformation component to better integrate target information and relies on the context-preserving and position relevance mechanisms to maintain the advantages of LSTM-based models. However, the effectiveness of these models is limited because of insufficient aspect-level training data.

Knowledge-based transfer method can use one task to help another to improve the performance. PRET+MULT model achieves superior results among the baseline models. Because PRET+MULT model not only obtains specific-target representation through attention mechanism, but also transfers the domain knowledge from short text-level data through pretraining and multi-task learning, thereby improving the effect of aspect level sentiment classification in the case of limited data. However, the PRET+MULT represents the target as the average of word vectors, and treats each word equally. It cannot obtain powerful semantic representation of targets, which is very important for aspect-level drug reviews sentiment classification.

Our proposed PM-DBiGRU is more outstanding than AOA, MGAN, TNet and PRET+MULT. Specifically, the accuracy and the Macro-F1 of our proposed PM-DBiGRU are 1.47% and 2.39% higher than those of the PRET+MULT model, respectively. On one hand, PM-DBiGRU encodes contextual information for targets by a independent BiGRU network to strengthen target representation rather than using the averaged target vector to guide the attention, which will lose some information, especially on the targets with multiple words. On the other hand, PM-DBiGRU is able to transfer the domain knowledge from drug reviews short text-level datas through pretraining and multi-task learning.BiGRU is more conducive than LSTM to learning context information and acquiring more knowledge.

D. ANALYSIS OF PM-DBIGRU MODEL

To better understand the influence of each part of PM-DBiGRU model on the results, we design three variations of the proposed model. Experimental results are illustrated in Table 4. We can find that the performance of BiGRU is better than LSTM in Table 3, because BiGRU can better encode contextual information from both forward and backward directions. We adopt BiGRU as the basic model structure to judge the sentiment polarity of given target. For the BiGRU-ATT model, although it can extract more accurate sentiment information for given target, it also ignores the meaning of the target. Especially in the drug review, the semantics of target itself has a certain influence on the judgment of sentiment polarity, as shown in the previous examples. DBiGRU is better than BiGRU-ATT, which indicates a separate BiGRU for target could improve the performance, especially on the targets with multiple words in drug reviews.

TABLE 4. Experimental results of PM-DBiGRU and its variance (%).

Model	Accuracy	Macro- F1
BiGRU	71.35	69.88
BiGRU-ATT	71.84	70.28
DBiGRU-ATT	72.92	71.34
P-DBiGRU	76.16	75.24
M-DBiGRU	75.97	75.02
PM-BiGRU	77.37	76.51
PM-DBiGRU	**78.26**	**77.75**

P-DBiGRU model and M-DBiGRU model are better than DBiGRU-ATT model, which proves that aspect-level drug review sentiment classification task can incorporate knowledge from short text-level drug reviews corpus with certain semantic similarity through pretraining and multi-task learning. The classification effect is improved greatly, the accuracy and the Macro-F1 of the P-DBiGRU model are 3.24% and 3.9% higher than those of the DBiGRU-ATT model, respectively. Similarly, the accuracy and the Macro-F1 of the M-DBiGRU model are 3.05% and 3.68% higher than those of the DBiGRU-ATT model, respectively. This shows that pretraining may be more helpful to acquire valuable knowledge than multi-task learning. Furthermore, the accuracy and the Macro-F1 of the PM-DBiGRU model are 0.89% and 1.24% higher than those of the PM-BiGRU model, respectively. The accuracy and the Macro-F1 of DBiGRU-ATT model are 1.08% and 1.06% higher than those of the BiGRU-ATT model, which shows that the specific BiGRU network is built for the target to increase the Macro-F1 by more (1.24% > 1.06%) on the basis of pretraining and multi-task learning. In other words, whether or not to apply the pretraining and multi-task learning transfer knowledge, the BiGRU network for target can help to improve the aspect-level sentiment classification performance. However, the Macro-F1 is improved more obviously based on pretraining and multi-task learning.

E. CASE STUDY

We select some drug reviews in SentiDrugs to explore which word contributes the most to the sentiment polarity of given target and which word in the target is more important. We visualize the attention weights. The depth of color indicates the importance of a word, the darker the more important. In Table 5, we observe that the common words "*of*", "*the*", "*and*",and punctuation "," are rarely noticed by our model, because some common words and punctuations often make little contribution for prediction. Obviously, in the first example, "*received*" and "*consistent*" play an important

TABLE 5. Visualization of attention weights for aspect-level drug review.

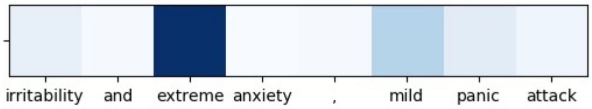
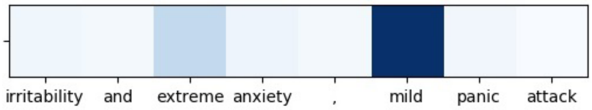
Aspect	Drug review sentences	Ans./Pre.
the pain from the ulcer		1/1
anxiety		-1/-1
panic attack		0/0

TABLE 6. Examples for each error category in SentiDrugs.

Error category	Drug reviews examples
no direct sentiment expression	1. I got a **rush on the back of my neck** that last for 1 year even thought I only took Prozac for 2 months.
lack of certain background knowledge	2. A slowing **in androgenic alopecia**, and some potential re-growth.
	3. ecreased **eczema patches**, however did not fully heal them.
complex sentence structure and expression	4. **cholesterol** was lowered but would have needed large dose to meet desirable levels.
	5. hopefully **this drug** got rid of many actinic keratoses permanently, but hard to say.

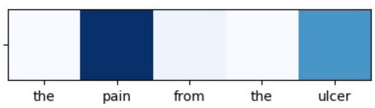


FIGURE 4. Visualization of attention weights for target "the pain from the ulcer".

role in sentiment classification of the target "*the pain from the ulcer*", and our model pays great attention to them as expected. As shown in Fig. 4, our model can learn that "*pain*" is the most important word in the target "*the pain from the ulcer*". The correct classification results show that our model can judge the sentiment polarity according to the semantics of the target, because in some cases, "*received*" does not mean positive sentiment polarity. However, our model can predict correctly when target is the "*pain*". In the last two examples, there are two targets "*economy*" and "*panic attack*" in the drug review sentence "*irritability and extreme **anxiety**, mild **panic attack***". We find that our model can automatically point to the correct sentiment words of each target in this case, where the target "*panic attack*" belongs to the expression of the medical field and our model can correctly classify sentiment polarity indicating that aspect-level classification indeed benefits from short text-level knowledge.

F. ERROR ANALYSIS

We carefully analyze the wrongly classification samples in the test set of the SentiDrugs to find out the limitations of our method. We classify these errors into three cases as shown in Table 6. The first type of errors is that there is no direct sentiment expression towards the target, which has also appeared in previous work [22]. For example, in the sentence 1, there is no obvious sentiment information for the target "rash on the back of my neck". Second type of errors is that sometimes certain background knowledge is required to judge the sentiment polarity, such as in the sentence 2, the target word "in androgenic alopecia" may not be correctly represented by BiGRU. Therefore, our model can correctly pay attention to the sentiment word "slowing", but it cannot be understood "slowing" is positive. Experimental results show that our model can transfer knowledge from short text-level sentiment classification tasks, but it may also lead to wrong classification due to the lack of such examples in short text-level drug reviews.The same can happen in sentence 3. The third case is the complex sentence structure and sentiment expression, such as in the sentence 4, the sentiment polarity of "cholesterol" should be neutral according to the meaning of the whole sentence and some common-sense knowledge. In this case, it is difficult for our model to judge the sentiment category of target words by detecting explicit opinion words.The same can happen in sentence 5.

VI. CONCLUSION

In this paper, we propose a new aspect-level sentiment analysis dataset SentiDrugs, which lays a foundation for the research of fine-grained sentiment analysis based on drug reviews. We also propose a pretraining and multi-task learning model based on double BiGRU for aspect-level drug reviews sentiment classification. Experiments on SentiDrugs verify that PM-DBiGRU can fully play the important role of bidirectional semantic information of target for judging sentiment polarity in drug reviews. And domain knowledge incorporated from short text-level corpus helps to improve the performance of aspect-level sentiment classification. Furthermore, we have built several powerful benchmark models for the SentiDrugs. In the future work, we plan to explore a more effective model to solve the problem of aspect-level sentiment classification in the medical background.

REFERENCES

[1] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, nos. 1–2, pp. 1–135, 2008.

[2] K. Schouten and F. Frasincar, "Survey on aspect-level sentiment analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 3, pp. 813–830, Mar. 2016.

[3] X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach," In *Proc. 28th Int. Conf. Mach. Learn. (ICML)*, 2011, pp. 513–520.

[4] H. H. Lek and D. C. Poo, "Aspect-based Twitter sentiment classification," in *Proc. IEEE 25th Int. Conf. Tools Artif. Intell.*, Herndon, VA, USA, Nov. 2013, pp. 366–373, doi: 10.1109/ictai.2013.62.

[5] M. Pontiki and D. Galanis. (Jun. 2016). *Semeval-2016 Task 5: Aspect Based Sentiment Analysis*. [Online]. Available: https://www.aclweb.org/anthology/S16-1002.pdf

[6] Q. Ke, J. Zhang, H. Song, and Y. Wan, "Big data analytics enabled by feature extraction based on partial independence," *Neurocomputing*, vol. 288, pp. 3–10, May 2018.

[7] H. Dou, Y. Qi, W. Wei, and H. Song, "A two-time-scale load balancing framework for minimizing electricity bills of Internet data centers," *Pers. Ubiquitous Comput.*, vol. 20, no. 5, pp. 681–693, Oct. 2016.

[8] L. Jiang and C. C. Yang, "User recommendation in healthcare social media by assessing user similarity in heterogeneous network," *Artif. Intell. Med.*, vol. 81, pp. 63–77, Sep. 2017.

[9] H. Sharif, F. Zaffar, A. Abbasi, and D. Zimbra, "Detecting adverse drug reactions using a sentiment classification framework," in *Proc. 6th ASE Int. Conf. Social Comput. (SocialCom)*, Stanford, CA, USA, May 2014.

[10] J. C. Na, W. Y. M. Kyaing, C. S. Khoo, S. Foo, Y. K. Chang, and Y. L. Theng. (Nov. 2012). *Sentiment Classification of Drug Reviews Using a Rule-Based Linguistic Approach*. [Online]. Available: https://www.ntu.edu.sg/home/sfoo/publications/2012/2012-ICADL-Na.pdf

[11] D. Cavalcanti and R. Prudêncio, "Aspect-based opinion mining in drug reviews," in *Proc. EPIA Conf. Artif. Intell.* Cham, Switzerland: Springer, Sep. 2017, pp. 815–827.

[12] N. Ofek, C. Caragea, and L. Rokach. (May 2013). *Improving Sentiment Analysis in an Online Cancer Survivor Community Using Dynamic Sentiment Lexicon*. [Online]. Available: http://people.cs.ksu.edu/~ccaragea/papers/society13.pdf

[13] F. Gräßer and S. Kallumadi. (Apr. 2018). *Aspect-Based Sentiment Analysis of Drug Reviews Applying Cross-Domain and Cross-Data Learning*. [Online]. Available: http://kdd.cs.ksu.edu/Publications/Student/kallumadi2018aspect.pdf

[14] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[15] K. Cho, B. van Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," Sep. 2014, *arXiv:1406.1078*. [Online]. Available: https://arxiv.org/pdf/1406.1078.pdf

[16] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, "Target-dependent Twitter sentiment classification," in *Proc. 49th Annu. Meeting Assoc. Comput. Linguistics, Hum. Lang. Technol.*, vol. 1, Jun. 2011, pp. 151–160.

[17] D. Tang, B. Qin, X. Feng, and T. Liu, "Effective LSTMs for target-dependent sentiment classification," in *Proc. 26th Int. Conf. Comput. Linguistics, Tech. Papers*, 2016, pp. 3298–3307.

[18] Y. Wang, M. Huang, X. Zhu, and L. Zhao, "Attention-based LSTM for aspect-level sentiment classification," in *Proc. EMNLP*, 2016, pp. 606–615.

[19] D. Ma, S. Li, X. Zhang, and H. Wang, "Interactive attention networks for aspect-level sentiment classification," Sep. 2017, *arXiv:1709.00893*. [Online]. Available: https://arxiv.org/pdf/1709.00893.pdf

[20] B. Huang, Y. Ou, and K. M. Carley, "Aspect level sentiment classification with attention-over-attention neural networks," Jul. 2018, *arXiv:1804.06536*. [Online]. Available: https://arxiv.org/pdf/1804.06536.pdf

[21] D. Tang, B. Qin, and T. Liu, "Aspect level sentiment classification with deep memory network," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 214–224.

[22] P. Chen, Z. Sun, L. Bing, and W. Yang, "Recurrent attention network on memory for aspect sentiment analysis," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 452–461.

[23] Q. Liu, H. Zhang, Y. Zeng, Z. Huang, and Z. Wu, "Content attention model for aspect based sentiment analysis," in *Proc. World Wide Web Conf. World Wide Web (WWW)*, 2018, pp. 1023–1032.

[24] F. Fan, Y. Feng, and D. Zhao. (Oct. 2018). *Multi-Grained Attention Network for Aspect-Level Sentiment Classification*. [Online]. Available: https://www.aclweb.org/anthology/D18-1380.pdf

[25] X. Li, L. Bing, W. Bing, and L. Bing, "Transformation networks for target-oriented sentiment classification," May 2018, *arXiv:1805.01086*. [Online]. Available: https://arxiv.org/pdf/1805.01086.pdf

[26] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 task 4: Aspect based sentiment analysis," in *Proc. 8th Int. Workshop Semantic Eval. (SemEval)*, Dublin, Republic of Ireland, 2014, pp. 27–35.

[27] R. He, W. S. Lee, H. T. Ng, and D. Dahlmeier, "Exploiting document knowledge for aspect-level sentiment classification," Jun. 2018, *arXiv:1806.04346*. [Online]. Available: https://arxiv.org/pdf/1806.04346.pdf

[28] T. Ali, D. Schramm, M. Sokolova, and D. Inkpen. (Oct. 2013). *Can I Hear You? Sentiment Analysis on Medical Forums*. [Online]. Available: https://www.aclweb.org/anthology/I13-1077.pdf

[29] S. M. Jiménez-Zafra, M. T. Martín-Valdivia, M. D. Molina-González, and L. A. Ureña-López, "How do we talk about doctors and drugs? Sentiment analysis in forums expressing opinions for medical domain," *Artif. Intell. Med.*, vol. 93, pp. 50–57, Jan. 2019.

[30] W. Wei, X. Fan, H. Song, X. Fan, and J. Yang, "Imperfect information dynamic stackelberg game based resource allocation using hidden Markov for cloud computing," *IEEE Trans. Serv. Comput.*, vol. 11, no. 1, pp. 78–89, Jan. 2018.

[31] J. Wagner, P. Arora, S. Cortes, U. Barman, D. Bogdanova, J. Foster, and L. Tounsi, "DCU: Aspect-based polarity classification for SemEval task 4," in *Proc. 8th Int. Workshop Semantic Eval. (SemEval)*, Dublin, Republic of Ireland, Aug. 2014, pp. 223–229.

[32] L. Mou, Z. Meng, R. Yan, G. Li, Y. Xu, L. Zhang, and Z. Jin, "How transferable are neural networks in NLP applications?" Oct. 2016, *arXiv:1603.06111*. [Online]. Available: https://arxiv.org/pdf/1603.06111.pdf

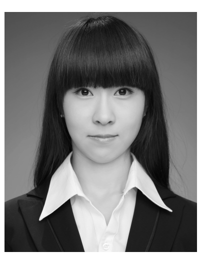
[33] J. Cohen, "A coefficient of agreement for nominal scales," *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, Apr. 1960.

[34] G. Ganu, N. Elhadad, and A. Marian, "Beyond the stars: Improving rating predictions using review text content," in *WebDB*, vol. 9, pp. 1–6, Jun. 2009.

[35] M. Saeidi, G. Bouchard, M. Liakata, and S. Riedel, "SentiHood: Targeted aspect based sentiment analysis dataset for urban neighbourhoods," Oct. 2016, *arXiv:1610.03771*. [Online]. Available: https://arxiv.org/pdf/1610.03771.pdf

[36] W. Jing, S. Huo, Q. Miao, and X. Chen, "A model of parallel mosaicking for massive remote sensing images based on spark," *IEEE Access*, vol. 5, pp. 18229–18237, 2017.

[37] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1532–1543.



YUE HAN received the B.S. degree from the Heilongjiang University of Science and Technology, Harbin, China, in 2017. She is currently pursuing the M.Sc. degree with the School of Information and Computer Engineering, Northeast Forestry University, Harbin. Her research interests include deep learning, natural language processing, and sentiment analysis, especially on aspect-level sentiment analysis.

MEILING LIU (Member, IEEE) received the B.S. and M.S. degrees in computer science and technology from the Northeast Forestry University, Harbin, China, in 2006, and the Ph.D. degree in computer application technology from Harbin Industrial University, Harbin, in 2012. From 2006 to 2019, she worked at Northeast Forestry University, as a Teacher. From 2014 to 2015, she was a Visiting Scholar with Virginia Tech University, USA. Her research interests include natural language processing, social media data mining, intelligent city transportation, and artificial intelligence technology application. She has been a member of the IEEE CS, since 2017, a member of ACM, since 2017, and a member of CCF, since 2012.

WEIPENG JING (Member, IEEE) received the Ph.D. degree from the Harbin Institute of Technology, China. He is currently an Associate Professor with Northeast Forestry University, China. He has published over 50 research articles in refereed journals and conference proceedings, such as CPC, PUC, and FGCS. His research interests include modeling and scheduling for distributed computing systems, fault tolerant computing and system reliability, cloud computing, and spatial data mining.