IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON ADVANCED DATA MINING METHODS FOR SOCIAL COMPUTING

Received December 31, 2019, accepted January 30, 2020, date of publication February 24, 2020, date of current version March 6, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2976196

The OL-DAWE Model: Tweet Polarity Sentiment Analysis With Data Augmentation

WENHUAN WANG[1], BOHAN LI[1,2,3], DING FENG[1], ANMAN ZHANG[1], AND SHUO WAN[1]

[1]College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China
[2]Key Laboratory of Safety-Critical Software, Ministry of Industry and Information Technology, Nanjing 211106, China
[3]Collaborative Innovation Center of Novel Software Technology and Industrialization, Jiangsu 210000, China

Corresponding author: Bohan Li (bhli@nuaa.edu.cn)

This work was supported by National Natural Science Foundation of China (61402225, 61728204), Innovation Funding (NJ20160028, NT2018028, NS2018057), Aeronautical Science Foundation of China (2016551500), State Key Laboratory for smart grid protection and operation control Foundation, and the Science and Technology Funds from National State Grid Ltd., China degree and Graduate Education Fund.

ABSTRACT Introducing negative items into sentences can shift the polarity of emotional words and leads to misclassification. Therefore, dealing with the negative item is indispensable to the analysis of the polarity of tweets. This paper first uses the combination of Conjunction Analysis (CA) technology and Punctuation Mark Identification (PMI) technology to detect negation cue and its scope. Besides, we propose the OL-DAWE model, which uses Data Augmentation(DA) technology to generate opposed tweets according to the original tweet. The model extends the training data set, and test data set and learns the original and opposed sides of the tweet in the training module. When predicting the polarity of tweets, the OL-DAWE model considers the positive degree (negative degree) of the original tweet and the negative degree (positive degree) of its opposed tweet. We conduct experiments on two real-world data sets. We prove the effectiveness of our combined technology in negation processing and show that the OL-DAWE model in the polarity sentiment analysis of tweets is better than the baseline for its simplicity and high efficiency.

INDEX TERMS Data augmentation, negation scope detection, polarity shift, sentiment analysis.

I. INTRODUCTION

Emerging of various social media and commercial websites has encouraged people to express their opinions on multiple platforms. New comments are generated every minute, and such massive amounts of data contribute to the generation of sentiment analysis (SA). SA is the computational analysis of the speaker's or writer's emotions, opinions and attitudes towards some topic, and identification of non-trivial and personal information from text repository. Sentiment analysis is often accompanied by opinion mining or text mining. The framework of SA mainly includes the following sub-tasks: obtaining text data, data cleaning and preprocessing, converting text into machine-readable forms, feature selection, and applying NLP and machine learning algorithms finally. Most studies focus on the positive, negative and objective classification of a given text at present. The research on the positive or negative tendency of emotions is also called polarity sentiment analysis. For SA, many studies (such as [1], [2], [5]) represent each word in the text as a real-valued, continuous and low-dimensional vector, which also known as Word Embedding. Although many efforts have been put into improving the accuracy of classification [3], [4], most of the methods have little effect because of the inherent difficulty of Word Embedding which called polarity shift. Besides, the randomness and irregularity of short texts on social networks also hinder the classification method. Given in this paper, we use the Stanford data set (TSCDS) [19] and Sanders Twitter Sentiment Corpus data set (TSDS) [27], which contains a large number of non-standard tweets. However, the simple Word-Embedding model [28] can not make full use of personal emotions contained in words. Therefore, polarity shift is more natural to appear (i.e., negative tweets are mistaken for positive, active tweets are misjudged as for negative).

Polarity shift refers to the reversal of text emotion due to some reasons. Negation cues (i.e., negative items) in the sentence is one of the most important causes of polarity shift. In SA, the detection of negative items and negation scope are essential. Sentiment classification accuracy can be improved

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

by efficient and automatic detection approaches [7]. In [8], the scope of a negation cue (a negative word or phrase) is hypothesized to be the negation cue to the end of the clause. This definition of negation scope ignores the complexity of language. For example, ''The package of this eye shadow tray is not elegant but its colors match my heart.'' In this sentence, the scope of negation is from the negation cue ''not'' to its next word ''elegant''. Still, if the definition of [9] is followed, the positivity of ''match'' will be affected, and the word's polarity will be shifted. The two data sets we used consisted of tweets, since [10] reports that the frequency of negation in spoken sentences was twice as frequent as in written texts. As an open communication platform, tweets generated by twitter tend to be colloquial. As in the case of [6], [9], [14], the scope of negation determined by punctuation marks will have a negative impact on the polarity sentiment analysis. Therefore, it is necessary to detect negation cues in texts and fully assess their negation scope.

Most of the sentiment polarity analysis methods on tweets are deficient in two aspects:

1) ignore the importance of negation cues and their scope;
2) ignore the emotional comparability between positive and negative tweets.

We proposed an opposed learning model based on Data Augmentation (DA) and Word Embedding (OL-DAWE model) in this paper. The OL-DAWE model uses Word Embedding technology to learn the two opposing sides of a tweet, which obtained through DA and utilizes the polarity comparison between the original tweets and the opposed tweets to improve the prediction accuracy and robustness.

A. RESEARCH CONTRIBUTIONS

- When using Data Augmentation technology to augment data set, we fully understand the complexity of the language in the most basic module, which contains detection negation cues and scope of negation. The scope of negation is not merely defined as all the words between the negation cue and the first punctuation mark following it. Instead, we consider the case of conjunctions in more complex negative sentences. The Punctuation Mark Identification (CPI) technology and Conjunction Analysis(CA) technology are combined to define the scope of negation. Also, we outline six rules to handle complex sentences that contain conjunctions.
- In this paper, tweets are divided into positive tweets and negative tweets. Each negative tweet (original tweet) in the original data set can be turned into a positive tweet (opposed tweet) by a series of operations such as detecting negative cues and their scope, reversing sentiment words, reversing polarity labels, etc., and vice versa. The data set composed of the opposed tweets is called opposed data set, and the technology used to obtain the opposed data set is called Data Augmentation(DA).
- We use DA to obtain the opposed training data set and opposed prediction data set. We use word2vec to convert tweets into word vectors. Three classifiers (Naïve Bayes, Logistic Regression, Support Vector Machines) are applied to the training module to learn the original data set and the opposed data set. When using the test data set to predict the polarity of a tweet, we consider the positivity (negativity) of the original tweet and the negativity (positivity) of the opposed tweet. Experiments show that our method can effectively improve the classification accuracy.

To the best of our knowledge, this paper uses DA for the first time to apply opposed training and prediction to tweets polarity sentiment analysis. We focus on the negative sentences on Twitter, overcoming the difficulties of previous works, which negated sentence processing, and ignored the complexity of negation cues and scope of negation. For the first time, a negation handling technology combining PMI and CA was proposed. In fact, because of the random patterns of tweets and their relevance to human descriptions, to apply DA in tweets scenarios is much harder for us than canonical text scenarios.

B. ORGANIZATION

The remainder of this paper is organized as follows: Section II discusses the related work in negation cues and scope detection, Data Augmentation, and tweet sentiment analysis. In the section III, we propose three modules to illustrate our methods. The experimental description and analysis of the results are given in section IV. Section V will give a summary of the work of this paper and describe our future development direction.

II. RELATED WORK

A. NEGATION CONTROL

A negative sentence is containing one or more negation cues, where a negation cue can be a word (e.g., not), or a multi-word (e.g., no longer) connotative expressing negation. A negation cue is given before a positive word or phrase that will change the sentence polarity from positive to negative. Negation forms have many kinds. Figure 1 is an intuitive observation of the negation forms. Depending upon certain language patterns, the form of negation may explicit (with clear negation cues such as not, no, etc.) or implicitly (with ambiguous negation cues such as little, hardly, etc.). At the grammatical structural level, the negation in negative sentences sometimes appear with contrast, syntactic, non-negative, compound and morphological negations. And non-negative negation, according to valence shifters, negation can be divided into intensifier negation and diminisher negation. In consideration of the negation distance, negation can be divided into local negation and long-distance negation. For example, ''I do not like this eye shadow tray.'' is a local negation because the negation cue ''not'' acts directly on the sentiment carrying word ''like'' that follows it. Conversely, a long-distance negation like ''This eye shadow tray does not have beautiful colors, nice opaque and elegant box.'', which

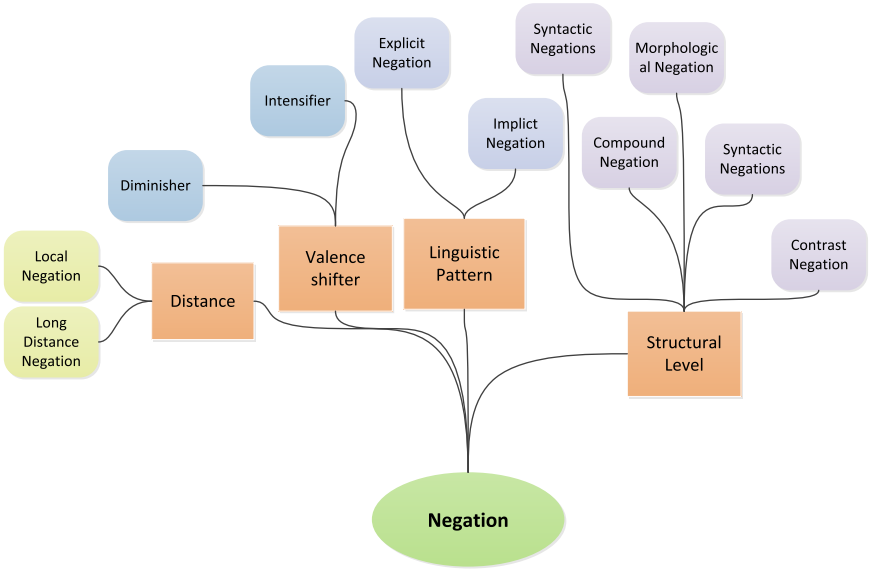

FIGURE 1. Classification of negation forms.

implies that negation does not directly apply to sentiment carrying words [11]. Our work focuses on the negative distance.

The hypothesis of the negation scope in [8], [9] is the same as [12] that only the first punctuation after the negation cue is used as the sign of the end of negation. They all treat the negation distance as a simple problem and do not distinguish the length of the negation distance. Reference [6], [13] deems that the scope of negation to be negation cue's next five words. Reference [14] supposes that the scope of a negation cue is several words of its right and expresses that the polarity of an emotional term can be flipped within the vicinity of negation. These negation scope definitions only consider local negation, while the texts they process all contain more or less long-distance negative sentences, which will affect the classification accuracy. Since long-distance negation is often associated with conjunctions, we propose a combination of Punctuation Mark Identification (PMI) technology and Conjunction Analysis (CA) technology to reduce the impact of negation on polarity shift.

B. DATA AUGMENTATION

Data Augmentation technology refers to a method of augmenting observation data to make it easier to analyze. The validation error must continue to decrease with the training error to build useful Deep Learning models. DA is a potent method to achieve this purpose. At present, DA is mostly applied for the field of image processing. Reference [15] covers the use of GAN image synthesis in medical imaging applications such as brain MRI synthesis and lung cancer diagnosis. Reference [16] creates new instances by interpolating new points from existing instances via k-Nearest Neighbors.

In Natural Language Processing (NLP), the application of DA is minimal, on account of the difficulty of obtaining universal data conversion rules that guarantee data quality and can be automatically applied to various fields. Currently, the common augmentation method is to replace synonyms selected from manual ontology in NLP [17]. In [18], DA is carried out through word similarity. In addition, different phoneme-font translation systems are adopted in [19]. For polarity sentiment analysis, however, the augmentation technology of synonym substitution does not consider the strong polarity comparison of emotional words. We first propose to use antonyms replacement to expand data sets with negation handling, using the original tweets data set to expand its opposed tweets data set. At the same time, we use these two data sets in training and prediction.

C. SENTIMENT ANALYSIS

Research on opinion mining and sentiment analysis of tweets have grown considerably in the last decade. Reference [20] demonstrates the linguistic function of micro-blogging for detecting sentiment of Twitter messages and the utility of existing vocabulary resources. Reference [22] uses a complex Bi-directional LSTM model to capture more context information. None of these approaches take into account the polar comparability of tweets and are less efficient due to their complex parameters and functions.

Reference [23] uses a noisy training set to reduce the labeling effort when developing classifiers, and designs a 2-step automatic sentiment analysis method for classifying tweets. Reference [25] proposes an influence probability model for twitter sentiment analysis. If a username is found in the body of a tweet, it is influencing action and it contributes to influencing probability. Reference [26] creates a twitter corpus by automatically collecting tweets using Twitter API and automatically annotating those using emoticons. However, the training set is also less efficient since it contains

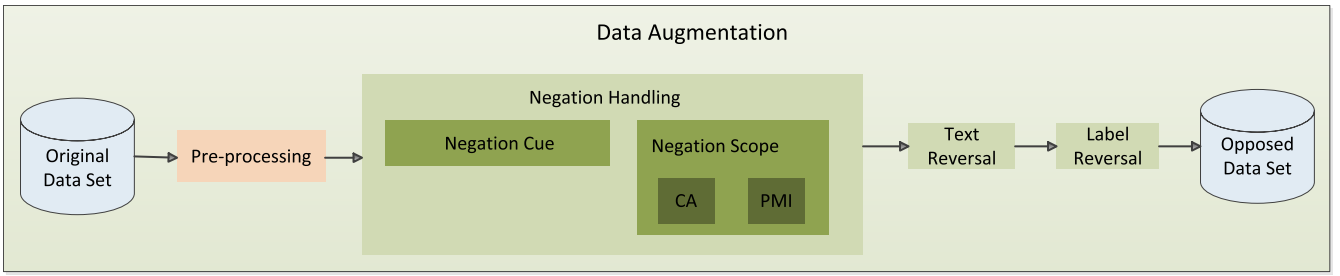

FIGURE 2. The process of Data Augmentation.

only tweets having emoticons. Considering the comparability of positive and negative polarity of sentiment, this paper analyzes two comparable data sets obtained by Data Augmentation technology in training and prediction. Due to the simplicity and high efficiency of the model, under certain conditions, the model we proposed performs better than the above method.

III. THE OL-DAWE MODEL

This paper proposes an OL-DAWE Model, which first uses Data Augmentation technology to reverse all original training tweets and test tweets to their opposed tweets. In the training module, we learn the original and opposing sides of a tweet at the same time. The degree of positivity or negativity of two opposing tweets is taken into account in polarity prediction. In section A, we detail the process of data Augmentation (DA), the flow chart of which is shown in Fig. 2. Two comparable data sets will be trained or predicted in sections B and C. In addition, the framework of our model is shown in Fig. 3.

A. DATA AUGMENTATION AND NEGATION HANDLING

In general, a more successful neural network requires millions of parameters to train data correctly. DA is an effective method to expand sample size. DA can

1) increase the amount of training data to improve the generalization ability of the model,
2) increase noise data and improve the robustness of the model.

However, when collecting data, it is often difficult to cover all the scenes. Data Augmentation technology is currently mainly used in image processing and target detection [20], and great achievements have been achieved. Less research was conducted on NLP and text mining. Considering the positive and negative polarity of tweets' short text data set, this paper, for the first time, adopts the antonym replacement data augmentation technology in the short text sentiment analysis. Based on the antonym dictionary(e.g., WordNet [24]), it constructs the opposed data set of tweets in the original data set in the way of one-to-one by reversing the sentiment words of tweets. Short texts generated by social platforms usually contain only a few words, more colloquial terms, and emojis, etc. compared to traditional regular texts (such as hotel reviews, movie reviews, news articles). For large-scale short text data sets, the text content is too sparse, so we first standard preprocess the tweet short text data sets. Note that the main difference between a tweet text and a traditional text is that it contains multiple emojis that can express both positive and negative emotions. For example, :) and :-) both express positive emotions. We will replace positive emoticons with token EMO_POS and negative emoticons with token EMO_NEG. our substitution methods and main emojis in tweets are shown in Table 1. Specific measures to extend the data set are as follows:

TABLE 1. Common emojis and our token replacement methods.

Token	Meaning	Examples
EMO_POS	EMO_SMILE	:), :), :-), (:, (-:, :')
	EMO_LAUGH	:D, : D, :-D, xD, x-D, XD, X-D, BD
	EMO_LOVE	<3, :*
	EMO_WINK	;-), : D, :-D, ;D, (;, (-;
EMO_NEG	EMO_SAD	:-(, : (, :(,):,)-:, </3, B(
	EMO_CRY	:,(, :'(, :"(
	EMO_ANGRY	X-(

1) NEGATION HANDLING

Negation handling in sentiment analysis includes two subtasks, namely negation cue, and scope detection. Negative cue detection is responsible for identifying negative words or phrases in sentences, such as no, not, rather than, etc. Negative scope detection is for defining the range involved in the negation cue.

- *Negation Cue Detection:* Negation cue is a negative word or part of speech reflected in the sentence. This paper uses rule-based keyword matching technology to perform negation cue detection and replace it with token "Negation". For example, the negative cue "do not" in S1 of Table 2 is replaced by "Negation" in S2.
- *Negation Scope Detection:* Negation scope detection technology study the range of language influences of negation cues in emotionally colored text.

This paper proposes a combination of Conjunction Analysis (CA) and Punctuation Mark Identification (PMI). PMI is sufficient to cope with local negation, while the combined

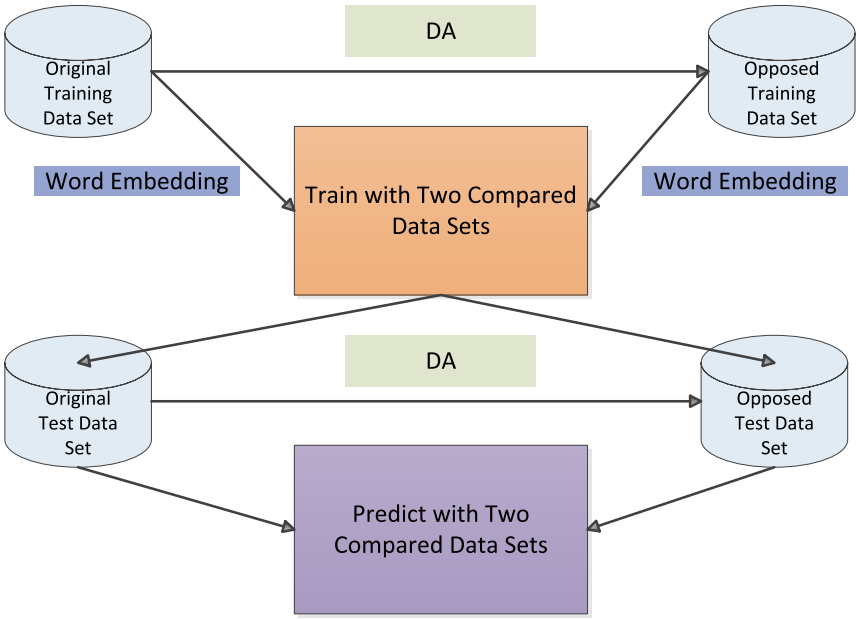

FIGURE 3. The process of polarity sentiment analysis by OL-DAWE model.

TABLE 2. Examples of negative sentences. The underlined word is the negation cue, and words in [] is in the scope of negation.

Example	Sentence
S1	I do not like this eye shadow tray, another is beautiful. *EMO_NEG*.
S2	I Negation like this eye shadow tray, another is beautiful.
S3	This eye shadow tray does not [have beautiful colors, nice opaque and elegant box].
S4	I does not [think this eye shadow tray is elegant] but its colors suit me well.
S5	This eye shadow tray does not [have beautiful colors and elegant box].
S6	I does not [like this eye shadow tray], and it is unpretty.
S7	I does not [like this eye shadow tray] and its colors looks unpretty.
S8	I does not [think this eye shadow tray is elegant], but I think its colors suit me well.

technology is mainly beneficial to the processing of long-distance negation.

- *Punctuation Mark Identification (PMI):* PMI technology defines the scope of negation is from "Negation" token to the first punctuation after it. For example, a simple local negation sentence S1 in Table 2, the first punctuation mark "," after the negation cue "not" is used to divide the two emotions of the speaker. But for long-distance negation, the punctuation mark sometimes fails to completely cover its negation scope, such as S3 in Table 2. The negation scope of S4 in Table 2 does not include items after "but". Therefore, we need to combine CA and PMI.
- *Conjunction Analysis (CA):* Conjunctions complicate the scope of negation. Simply defining the negation scope based on the punctuation reduces the accuracy of the classification. Therefore, consider the conjunction in negative sentences is necessary. For example, in S4 in Table 2, the speaker does not like the exquisiteness of the eye shadow tray but likes the color of the eye shadow. The conjunction "but" lead to a confrontation between the two clauses. Considering the general situation, the conjunctions are divided into adversative conjunction and coordinating conjunction. Through the analysis of the data set, when a tweet is reversed, we deal with the negative sentence involving a conjunction word according to certain grammatical rules.

The following is a brief description of the proposed rules based on some examples in Table 2 and the underlined word is the negation cue. Besides, words in [ ] are in the scope of negation.

1) If the clause after the first non-terminating symbol, which after the negation cue contains coordinating conjunction and the clause cannot be independently formed into a sentence, then the negation scope includes the clause. This rule applies to sentences like S3.
2) For the clause, the first non-terminating symbol after the negation cue contains coordinating conjunction

TABLE 3. The final reversal result. The original tweet with negative label in the original data set can be turned into a positive tweet by a series of operations such as detecting negative cues and their scope, reversing sentiment words, reversing polarity labels.

Data set	Class	Tweet
Original	Negative	I *Negation* like this eye shadow tray, and it is *unpretty EMO_NEG*.
Opposed	Positive	I like this eye shadow tray, and it is *elegant EMO_POS*.

TABLE 4. Examples of negative sentences. The underlined word is the negation cue, and words in [] is in the scope of negation.

Notation	Explanation
x	The original tweet sample
x'	The opposed tweet sample
y	The class label of the original tweet sample, $y \in \{0, 1\}$.
y'	The class label of the opposed tweet sample, $y' = 1 - y$.
X	The original data set,$X = \{x^i \| i = 1, \ldots, m\}$
X'	The opposed data set,$X' = \left\{x^{i'} \| i=1,\ldots,m\right\}$
θ	Weights of feature in a linear model
i	The ith tweet sample
m	The tital number of tweet sample
$L(\theta)$	Log-likelihood function of X
$l(\theta)$	Log-likelihood function of X'
$L_T(\theta)$	Simplified log-likelihood function

and the clause can be independent into a sentence, the negation cue does not include the clause, and the negation scope is from the negation cue to the first non-terminating punctuation. This rule applies to sentences like S6.

3) When a coordinating conjunction word is between the negation cue and the first punctuation mark, and no independent clause after this coordinating conjunction word, the scope of negation is the negation cue to the first punctuation mark. This rule applies to sentences like S5.
4) Suppose a coordinating conjunction word appears between the negation cue and its next first punctuation mark. The coordinating conjunction is followed by an independent clause, and then the negation scope ends with this coordinating conjunction. This rule applies to sentences like S7.
5) If adversative conjunction is between the negation cue and the first non-terminating symbol after it, the negation scope ranges from the negation cue to the adversative conjunction word. This rule applies to sentences like S4.
6) The case that no adversative conjunction between the negation cue and the first punctuation mark after it, the negative scope ends with this first punctuation. This rule applies to sentences like S8.

2) TEXT REVERSAL

If a tweet contains a negation token ''Negation'', the negation scope is first detected. All negation tokens are removed, the sentiment words in the negation scope are unchanged, and the sentiment words outside the negation scope are all reversed to their opposites. If the tweet contains the emoticon token ''EMO_POS'' (or ''EMO_NEG''), it is inverted to ''EMO_NEG'' (or ''EMO_POS''). Note that we are not looking for a full standard antonym word, because some words do not have antonyms, but instead, are replaced with words that express opposite feelings.

3) LABEL REVERSAL

For tweets in each training set, class labels are also reversed to their opposites (e.g., positive to negative, negative to positive) and added to the opposed data set. The final reversal result is as in Table 3.

To verify the effectiveness of the combination of PMI and CA technology used in the process of negation scope detection in mitigating the impact of polarity shift, we will compare the prediction accuracy and recall rate of the OL-DAWE Model with the combination technology (CA and PMI) and the Simple OL-DAWE Model without the CA technology.

B. TRAINING WITH TWO COMPARABLE DATA SETS

In the training stage, the original tweets used for training are inverted to generate opposed tweets using DA technology. The training tweets are labeled as the original training data set and the opposed training data set. The label of the opposed tweet is changed to the opposite of their corresponding original tweets. Training with two compared data sets(TTCDS) is performed with the combination of both original and opposed tweets. We first summarize some notations in Table 4 that will be used in the following descriptions before we proceed.

TABLE 5. Logistic regression classifier algorithm.

Input: the set of tweet vector x 'X' and x' 'X'', $X = \{x^i \mid i = 1, \ldots, m\}$, $X' = \left\{x^{i'} \mid i=1,\ldots,m\right\}$
Output: Improved classification accuracy and true positive rate of x being classified
Step 1:Begin
Step 2:For each tweet vector x and its opposed tweet vector x'
Step 3:Form Logistic Regression probability using (1)
Step 4:Evaluate Maximum Likelihood Estimates using (6)
Step 5:Measure probability of particular tweet using (8)
Step 6:End

For the sample points in multi-dimensional space, the Linear Regression model uses linear combination of features (feature weighting) to fit the distribution and track of space midpoint, which performs well in the regression problem. However, the effect is not ideal for the classification problem, for its uncertain range of output value. Even with the threshold value, it is difficult to become a classifier with good robustness. So we use the Logistic Regression model to deduce our algorithm, which shows in Tabel 5.

The Logistic Regression(LR) model assumes that the data obey Bernoulli distribution. By introducing sigmoid function into the LR model, the continuous output of the uncertain range of linear regression is mapped to the range of (0, 1), which becomes a probability prediction problem. In this paper, we study a binary classification problem. Then, the probability expression that LR uses the sigmoid function to predict the positive class of vector x is:

$$p(y=1|x,\theta) = h_\theta(x) = g\left(\theta^T x\right) = \frac{1}{1+e^{-\theta^T x}}. \quad (1)$$

The probability that the vector x belongs to the negative class is:

$$p(y=0|x,\theta) = 1 - h_\theta(x). \quad (2)$$

Since the binary classification problem is studied, that is, the label is not 0 or 1, then the above two equations can be combined into:

$$p(y|x,\theta) = h_\theta(x)^y \left(1-h_\theta(x)\right)^{1-y}. \quad (3)$$

We follow the maximum likelihood criterion, then the loss function is

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{m} p\left(y^i|x^i; w\right) \\ &= \prod_{i=1}^{m} \left(h_\theta\left(x^i\right)\right)^{y^i} \left(1-h_\theta\left(x^i\right)\right)^{1-y^i}. \end{aligned} \quad (4)$$

In order to simplify the calculation, the standard LR simplifies the maximum likelihood to log-likelihood:

$$\begin{aligned} l(\theta) &= \log L(\theta) \\ &= \sum_{i=1}^{m} y^i \log h_\theta\left(x^i\right) + \left(1-y^i\right) \log\left(1-h_\theta\left(x^i\right)\right) \end{aligned} \quad (5)$$

In addition, the data set used in the training phase is the combination of the original data set and the opposite data set. Therefore, the loss function includes not only the original data set part, but also the reverse data set part. Since $y^{i'} = 1 - y^i$, the log-likelihood function can be further simplified as follows:

$$\begin{aligned} L_T(\theta) &= \sum_{i=1}^{m} \left[y^i \log h_\theta\left(x^i\right) + \left(1-y^i\right) \log\left(1-h_\theta\left(x^i\right)\right)\right] \\ &\quad + \left[\left(1-y^i\right) \log h_\theta\left(x^i\right) + y^i \log\left(1-h_\theta\left(x^i\right)\right)\right] \\ &= \sum_{i=1}^{m} y^i \log\left[h_\theta\left(x^i\right)\left(1-h\left(x^i\right)\right) h_\theta\left(x^i\right)\right] \\ &\quad + \left(1-y^i\right) \log\left[\left(1-h\left(x^i\right)\right) h\left(x^i\right)\right] \end{aligned} \quad (6)$$

We will illustrate the effectiveness of TTCDS in solving the polarity shift problem through a sample of tweets:

- *Original training sample:* I do not like this eye shadow tray, and it is unpretty. Label: Negative.
- *Opposed training sample:* I like this eye shadow tray, and it is elegant. Label: Positive.

In general, the word "like" is considered to be a word with strong positive sentiment. Still, due to the negation cue "not", the polarity is shifted, and the word "like" is misconnected to the negative label in the original tweets data set. Hence, the weight of it will be added by a negative score in estimation. As a result, the weight of "like" will be updated by mistake. While in TTCDS, because of the removal of "Negation" token in the opposed tweet, "like" is (correctly) associated with the positive label, and a positive score will add its weight. Based on this, we can conclude that the errors in the learning stage caused by negation cue can be partly amended in TTCDS.

C. PREDICTION WITH TWO COMPARABLE DATA SETS

We first invert the test sample one-to-one into their opposed samples, forming the opposed test data set. And the two sample data sets are combined to make predictions. In the prediction process, we take emoticons into consideration. Predict with Two Data Sets (PTCDS) means that we consider two opposed tweets of x (original tweet) and x' (opposed tweet) to aid in predicting the class of the original sample x. Our main task is not to predict the class of x' in the opposed test data set, but to predict the class of x with the help of x'. As shown before, $p(\cdot|x)$ and $p(\cdot|x')$ represent the posterior probabilities of the original tweet x and the opposed tweet x', respectively. '$\cdot$' represents positive mark (+) or negative mark (−). Two sides of the tweet are considered in PTCDS.

- The positive sentiment degree of a tweet is found using two components.
  1) How much positive is the original tweet x, $p(+|x)$.
  2) How much negative is the opposed tweet x', $p(-|x')$.
- The positive or negative sentiment degree of a tweet is found using two components.

1) How much negative is the original tweet x, $p(-|x)$.
2) How much positive is the opposed tweet x', $p(+|x')$.

The opposed tweets we created during the DA phase might not be as good as the human-generated tweets. Since there is no accidental introduction to some noise data, and the requirement to maintain grammatical quality in tweets is lower than human languages. Therefore, we will use a tradeoff parameter in the PTCDS to leverage both original and opposed tweets. Assigning a relatively small weight to the opposite tweet can protect the model from being corrupted by combining low-quality tweets, that is, assigning a tradeoff parameter α $(0 < \alpha < 1)$ to the posterior probability $p(\cdot|x')$ of the opposed tweet. Therefore, the prediction score of a weighted combination of two-component predictions as follows:

$$\begin{cases} p(+|x, x') = (1-\alpha)\cdot p(+|x) + \alpha \cdot p(-|x') \\ p(-|x, x') = (1-\alpha)\cdot p(-|x) + \alpha \cdot p(+|x') \end{cases} \tag{7}$$

We define $y \in \{0, 1\}$, where 0 is the negative class $(-)$ and 1 is the positive class $(+)$. Then (7) can be expressed as a compact form:

$$\begin{aligned} p(y|x, x') &= (1-\alpha)\cdot p(y|x) + \alpha \cdot p(1-y|x') \\ &= (1-\alpha)\cdot p(y|x) + \alpha \cdot [1 - p(y|x')] \end{aligned} \tag{8}$$

D. INSTANCE SPECIFICATION

In this subsection, we try to extract a practical example to explain why our OL-DAWE model can address the problem of polarity shift. Let us take a look at a real test sample extracted from the TSCDS (We use DA to get the opposed tweet from the original tweet and use italics to indicate the parts that have changed):

- *Original Tweet:* I think the color matching of this eyeshadow tray is *not very beautiful*, and the plastic packaging is *botchy*. I *do not think it is worth* my money *EMO_NEG*.
- *Opposed Tweet:* I think the color matching of this eyeshadow tray is *very beautiful*, and the plastic packaging is *exquisite*. I *think it is worth* my money *EMO_POS*.

In a large number of experiments, we observe that the traditional sentiment classification method (such as the Simple Word-Embedding Model [28]), when it comes to the prediction of polarity sentiment of long-distance negative sentences containing negative words, is usually strongly influenced by emotional words, and then leads to the wrong prediction. For example, the traditional method gives a wrong result when predicting the above original tweet with $p(+|x) = 0.63$. The mistakes may occur for two reasons: 1) the words of ''beautiful'' and ''worth'' contribute very high positive scores; 2) the model deals with negation improperly. In comparison, the two predicted values given by the OL-DAWE model are $p(+|x) = 0.41$ (the score of original tweet) and $p(-|x') = 0.35$ (the score of opposed tweet). Considering these two prediction scores, the final prediction score obtained from (8) is as follows(set $\alpha = 0.5$ according to the experimental results):

$$p(y|x, x') = 0.5\, p(+|x) + 0.5\, p(-|x') = 0.38$$

Therefore, the prediction results of the OL-DAWE model are more robust for the reason that our model largely eliminates the negation cue's influence on the polarity shift of the original tweet.

IV. EXPERIMENTS

In this section, we discuss the implementation of polarity sentiment analysis with Data Augmentation and conduct comparative experiments on different tweets data set.

A. DATA SETS

Two data sets are scraped by twitter API i.e. Stanford data set (TSCDS) [19] and Sanders Twitter Sentiment Corpus data set (TSDS) [27]. Stanford data set contains 160,000 training tweets accompanied by 80,000 both positive and negative tweets. Whereas Sanders Twitter Sentiment data set contains 570 positive and 654 negative tweets.

B. EXPERIMENTAL SETTINGS

In our experiment, tweets on each category in two data sets are randomly divided into five folds (with four as training data and one as test data). All of the following results are reported and analyzed with an averaged accuracy of five-fold cross validation. We implement the Naive Bayes (NB) Classifier based on a multinomial event model with Laplace smoothing and Support Vector Machines (SVMs) Classifier based on the LibSVM toolkit. The kernel function in the SVM model is linear kernel, the penalty parameter is set to the default value, and the Platt's probability output is applied to approximate the posterior probability. The LibLinear toolkit is used as the Logistic Regression (LR) model with all parameters set to be the default value.

C. EVALUATION

In this part, we report and analyze the experimental results of the proposed OL-DAWE model. In order to verify the efficiency of our model in the field of sentiment polarity analysis of short texts, we will start from the following four models to evaluate.

- Baseline:Simple Word-Embedding Model (SWEM) [28] studied the original modeling ability of word embedding in the field of sentiment polarity analysis. This Model has no additional component parameters for encoding natural language sequences.
- State of the art: Bidirectional LSTM (Bi-LSTM) is proposed by [29], which uses a complex forward and backward recurrent neural networks to capture more contextual information. As far as we know, bidirectional LSTM technology is unique in the field of sentiment analysis.

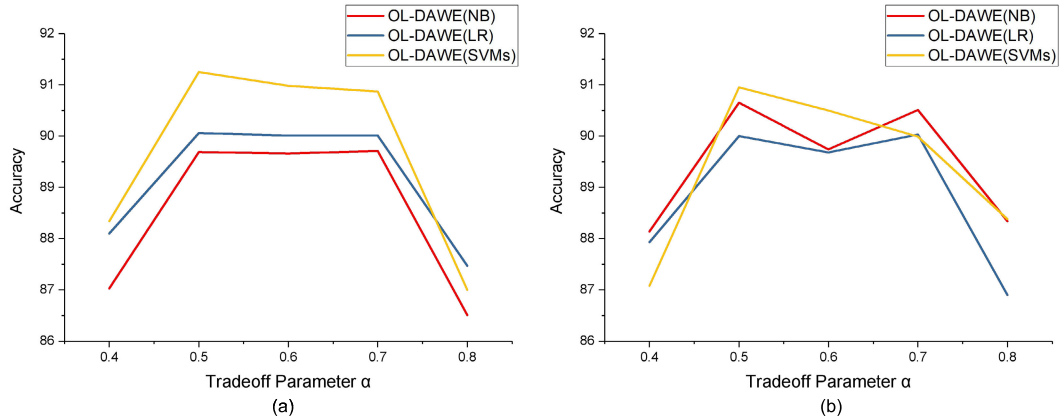

FIGURE 4. (a) and (b) represent accuracy variation for different tradeoff parameters when the OL-DAWE model is conducted on TSCDS and TSDS respectively.

- Simple OL-DAWE Model: we proposed a comparative learning model based on Data Augmentation and Word Embedding but without the combination of PMI and CA.
- OL-DAWE Model: We proposed a comparative learning model based on Data Augmentation and Word Embedding, which includes the combination of PMI and CA technology and the six rules we defined.

1) DYNAMIC RELATIONSHIP BETWEEN TRADEOFF PARAMETER α AND ACCURACY

Fig. 4 indicates the relationship between the tradeoff parameter and accuracy. Since our ultimate goal is to obtain a more robust prediction for the original tweet, the tradeoff parameter is to prevent the cart before the horse. Firstly, we do experiment based on TSCDS and find that the accuracy is higher when the tradeoff parameter value is between 0.4 and 0.8 compare with other values. Then, we run the model on TSDS with the tradeoff parameter ($0 < \alpha < 1$). Both the results give high accuracy with the values between 0.4 and 0.8. The above two experiments are based on three classifiers (NB, LR, and SVMs).

In Fig. 4, the subgraph (a) shows the influence of tradeoff parameter on prediction accuracy when the data set is extensive (160,000 tweets for TSCDS), and the subgraph (b) represents the influence of tradeoff parameters on prediction accuracy when the data set is small (1224 tweets for TSDS). When the tradeoff parameter values are 0.5 (the average accuracy on the data sets of TSCDS and TSDS are 90.31% and 90.53% respectively), 0.6 (the average accuracy on the data sets of TSCDS and TSDS are 90.22% and 89.97% respectively) and 0.7 (the average accuracy on the data sets of TSCDS and TSDS are 90.20% and 90.18% respectively), the prediction accuracy of the model remains at a relatively stable level, and the SVMs classifier performs well on these two data sets. However, when the tradeoff parameter is too large or too small, the prediction accuracy decreases by about 1.5 percentage points. The drop in accuracy maybe because the tradeoff parameter is the trade-off utilization of the original and opposed tweets. Assigning an appropriate tradeoff parameter can protect the model from the damage of low-quality tweet data set. We analyze the results and the images based on the results to reach a conclusion: when the value of the tradeoff parameter α is 0.5, the model can obtain the best and most stable performance. Therefore, in the following experiments, the tradeoff parameter $\alpha = 0.5$ is selected by default.

TABLE 6. Comparative analysis of Simple OL-DAWE model and OL-DAWE model in accuracy.

Classifier	Technology	TSCDS	TSDS
NB	PMI	81.53%	83.74%
NB	PMI+CA	89.69%	90.65%
LR	PMI	82.10%	80.20%
LR	PMI+CA	90.06%	90.00%
SVMs	PMI	82.98%	84.71%
SVMs	PMI+CA	91.25%	90.95%

2) THE EFFECTIVENESS OF THE COMBINED TECHNOLOGY

Table 6 describes the performance evaluation of the OL-DAWE Model with the combination of PMI and CA and the Simple OL-DAWE Model with PMI technology only for the sentiment classifier with negation control. The performance of the classifier is improved by adding the combined technology into negative tweets. The classifiers (SVMs, Naïve Bayes, Logistic Regression) without CA for negation control only achieve approximate 80% -85% accuracy rate over twitter as shown in Table 6. Besides, the SVMs classifier gains better performance over twitter data set and lead by approximate 1.5% in TSCDS and 4.5% in TSDS over other classifiers.

The performance of the OL-DAWE Model is significantly boosted up after incorporating CA and PMI techniques for sentiment analysis. In OL-DAWE Model, for incorporating CA and PMI, NB (89.69%, 90.65%), SVM

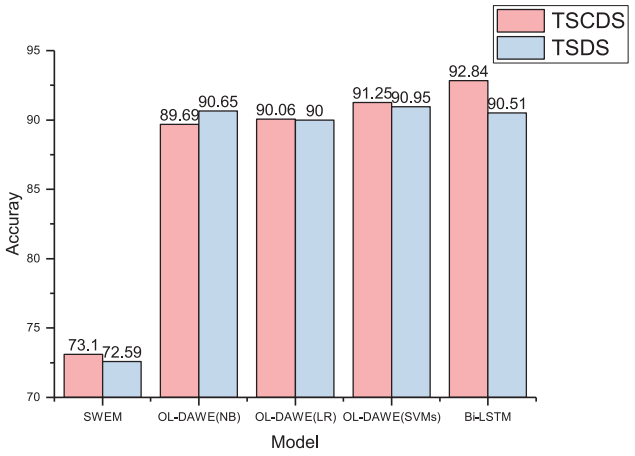

FIGURE 5. The comparison between the OL-DAWE Model based on three classifiers, the baseline model and the most advanced model.

(90.06%, 90.00%), and LR (91.25%, 90.95%) significantly boost the performance by approximately 6.91% - 8.16%, 7.96% - 9.80% and 6.24% - 8.27% respectively over two different Twitter data sets as shown in Table 6.

Significant improvement gained with classifiers over tweets data set results from the presence of a higher number of negative tweets i.e., approximate 50% and 53.43% in TSCDS and TSDS, respectively. With different angles of evaluating, the performance of classifier over combined technology and uncombined technology. It is observed that classifiers give better performance with PMI and CA.

3) THE EFFECTIVENESS OF OL-DAWE MODEL

The OL-DAWE model is based on three classifiers, i.e., SVMs, Naïve Bayes, and Logistic Regression. Fig. 5 shows that our model has a significant improvement over the baseline system in TSDCS and TSDS (increased 16.59% and 18.06% in the case of OL-DAWE model based on Naïve Bayes classifiers). This visible improvement stems from the negation handling of tweets in our model. Since the tweet data set contains a large number of negative sentences, the polarity shift caused by it misleads the baseline model. In our model, the SVMs achieve the best results (the accuracy was improved by 1.56% compared to the Naïve Bayes classifier in TSCDS), which may because our negation processing further increases the interpretability of SVMs.

However, Fig. 5 also demonstrates that in SA, the classifier using BiLSTM is about 1.59% more accurate than the OL-DAWE model in TSCDS. This finding is consistent with [8], where they hypothesize that the positional information of a word in text sequences may be beneficial to predict sentiment. This is reasonable since, for instance, the phrase ''not really good'' and ''really not good'' convey different levels of negative sentiment, while being different only by their word orderings. Surprisingly, on the TSDS, the classification accuracy of the OL-DAWE model based on the SVMs classifier and Naïve Bayes classifier is higher than that of the BiLSTM model (0.44% and 0.14%). The BiLSTM model requires a large number of compositional parameters, and the calculation is time-consuming with lower efficiency. In contrast, our model has fewer parameters and significantly improved computational efficiency. According to Occam's razor, simple models are preferred, if all else are the same. It shows that our model is better than the Bidirectional LSTM model when the data set is small.

V. CONCLUSION AND FUTURE WORK

In this work, we first use the combination of Conjunction Analysis technology and Punctuation Mark Identification technology to detect negation cue and its scope. In addition, we propose a novel Data Augmentation approach, which creates opposed tweets that are sentiment-opposite to the original tweets one-to-one. The proposed OL-DAWE model makes use of the original and opposed tweets which gained by DA in pairs to train a sentiment classifier and make predictions. Experiments demonstrate that the OL-DAWE model is effective for the polarity classification of tweets. In the future, we intend to combine the two aspects of emergency detection and polar sentiment analysis and apply the research results to Chinese short text. We hope that our future research will help detect public health emergencies such as 2019-nCoV.

REFERENCES

[1] S. Wan, B. Li, A. Zhang, K. Wang, and X. Li, ''Vertical and sequential sentiment analysis of micro-blog topic,'' in *Proc. Int. Conf. Adv. Data Mining Appl.* Cham, Switzerland: Springer, 2018, pp. 353–363.
[2] A. Zhang, B. Li, S. Wan, and K. Wang, ''Cyberbullying detection with birnn and attention mechanism,'' in *Proc. Int. Conf. Mach. Learn. Intell. Commun.* Cham, Switzerland: Springer, 2019, pp. 623–635.
[3] S. Shubha and P. Suresh, ''An efficient machine learning Bayes sentiment classification method based on review comments,'' in *Proc. IEEE Int. Conf. Current Trends Adv. Comput. (ICCTAC)*, Mar. 2017, pp. 1–6.
[4] I. G. Councill, R. McDonald, and L. Velikovich, ''What's great and what's not: Learning to classify the scope of negation for improved sentiment analysis,'' in *Proc. Workshop Negation Speculation Natural Lang. Process.*, 2010, pp. 51–59.
[5] D. Tang, F. Wei, B. Qin, N. Yang, T. Liu, and M. Zhou, ''Sentiment embeddings with applications to sentiment analysis,'' *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 2, pp. 496–509, Feb. 2016.
[6] G. Grefenstette, Y. Qu, J. Shanahan, and D. Evans, ''Coupling niche browsers and affect analysis for an opinion mining application,'' in *Proc. Recherche D'Inf. Assistée Ordinateur*, Jan. 2004, pp. 186–194.
[7] N. Pröllochs, S. Feuerriegel, and D. Neumann, ''Learning interpretable negation rules via weak supervision at document level: A reinforcement learning approach,'' in *Proc. Conf. North*, 2019, pp. 407–413.
[8] B. Pang, L. Lee, and S. Vaithyanathan, ''Thumbs up? Sentiment classification using machine learning techniques,'' in *Proc. EMNLP*, vol. 10, Feb. 2002, pp. 79–86.
[9] S. Li and C.-R. Huang, ''Sentiment classification considering negation and contrast transition,'' in *Proc. 23rd Pacific Asia Conf. Lang., Inf. Comput.*, Dec. 2009, pp. 307–316.
[10] M. Yaeger-Dror and G. Tottie, ''Negation in English speech and writing: A study in variation,'' *Language*, vol. 69, no. 3, p. 590, Sep. 1993.
[11] T. I. Jain and D. Nemade, ''Recognizing contextual polarity in phrase-level sentiment analysis,'' *Int. J. Comput. Appl.*, vol. 7, no. 5, pp. 12–21, Sep. 2010.
[12] R. Xia, T. Wang, X. Hu, S. Li, and C. Zong, ''Dual training and dual prediction for polarity classification,'' in *Proc. 51st Annu. Meeting Assoc. Comput. Linguistics*, Aug. 2013, pp. 521–525.
[13] M. Hu and B. Liu, ''Mining and summarizing customer reviews,'' in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2004, pp. 168–177.

[14] K. Yang, "WIDIT in TREC 2008 blog track: Leveraging multiple sources of opinion evidence," in *Proc. TREC*, Jan. 2008, pp. 1–13.
[15] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," *Med. Image Anal.*, vol. 58, Dec. 2019, Art. no. 101552.
[16] H. Han, W.-Y. Wang, and B.-H. Mao, "Borderline-smote: A new over-sampling method in imbalanced data sets learning," in *Proc. Int. Conf. Intell. Comput.*, vol. 3644, Sep. 2005, pp. 878–887.
[17] X. Zhang, J. Zhao, and Y. Lecun, "Character-level convolutional networks for text classification," in *Proc. Adv. Neural Inf. Process. Syst.*, Sep. 2015, pp. 649–657.
[18] W. Y. Wang and D. Yang, "That's so annoying!!!: A lexical and frame-semantic embedding based data augmentation approach to automatic categorization of annoying behaviors using #petpeeve tweets," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 2557–2563.
[19] G. Nicolai, B. Hauer, A. St Arnaud, and G. Kondrak, "Morphological reinflection via discriminative string transduction," in *Proc. 14th SIGMORPHON Workshop Comput. Res. Phonetics, Phonol., Morphol.*, 2016, pp. 31–35.
[20] E. Kouloumpis, T. Wilson, and J. Moore, "Twitter sentiment analysis: The good the bad and the omg!," in *Proc. 5th Int. AAAI Conf. Weblogs Social Media*, Jan. 2011, pp. 538–541.
[21] A. Hassan and A. Mahmood, "Efficient deep learning model for text classification based on recurrent and convolutional layers," in *Proc. 16th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2017, pp. 1108–1113.
[22] A. Williams, N. Nangia, and S. Bowman, "A broad-coverage challenge corpus for sentence understanding through inference," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics, Hum. Lang. Technol.*, vol. 1, 2018, pp. 1112–1122.
[23] L. Barbosa and J. Feng, "Robust sentiment detection on Twitter from biased and noisy data," in *Proc. 23rd Int. Conf. Comput. Linguistics, Posters*, vol. 2, Jan. 2010, pp. 36–44.
[24] G. A. Miller, "WordNet: A lexical database for English," *Commun. ACM*, vol. 38, no. 11, p. 39–41, 1995.
[25] Y. Wu and F. Ren, "Learning sentimental influence in Twitter," in *Proc. Int. Conf. Future Comput. Sci. Appl.*, Jun. 2011.
[26] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 5, no. 12, pp. 320–322, Dec. 2016.
[27] D. Ziegelmayer and R. Schrader, "Sentiment polarity classification using statistical data compression models," in *Proc. IEEE 12th Int. Conf. Data Mining Workshops*, Dec. 2012, pp. 731–738.
[28] J. Wieting, M. Bansal, K. Gimpel, and K. Livescu, "Towards universal paraphrastic sentence embeddings," Nov. 2015, *arXiv:1511.08198*. [Online]. Available: https://arxiv.org/abs/1511.08198
[29] A. Williams, N. Nangia, and S. R. Bowman, "A broad-coverage challenge corpus for sentence understanding through inference," 2017, *arXiv:1704.05426*. [Online]. Available: http://arxiv.org/abs/1704.05426

WENHUAN WANG received the B.E. degree in information management and information system from the Jiangsu University of Technology, Changzhou, in 2018. She is currently pursuing the master's degree with the Nanjing University of Aeronautics and Astronautics, Nanjing. Her research is about social network text sentiment analysis and knowledge graph. She is a Student Member of CCF.

BOHAN LI received the Ph.D. degree in computer application from the Harbin University of Science and Technology, Harbin, in 2009. He is currently an Associate Professor with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing. His current research interests include spatio-temporal database, recommendation systems, and sentiment analysis. He is a member of CCF and ACM.

DING FENG received the B.E. degree in computer science and technology from Jilin Jianzhu University, Changchun, in 2018. He is currently pursuing the master's degree with the Nanjing University of Aeronautics and Astronautics, Nanjing. His research interests are sentiment analysis and spatio-temporal data mining.

ANMAN ZHANG received the B.E. degree in information management and information system from the Jiangsu University of Technology, Changzhou, in 2017. She is currently pursuing the master's degree with the Nanjing University of Aeronautics and Astronautics, Nanjing. Her research is about social network text sentiment analysis and crowdsourcing. She is a Student Member of CCF.

SHUO WAN received the B.E. degree in computer science and technology from East China Jiaotong University, Nanchang, in 2017. He is currently pursuing the master's degree with the Nanjing University of Aeronautics and Astronautics, Nanjing. His research interests are sentiment analysis and spatio-temporal data mining. He is a Student Member of CCF.