Expert Systems With Applications 166 (2021) 113987

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Multi-view ensemble learning method for microblog sentiment classification

Xin Ye, Hongxia Dai, Lu-an Dong [*], Xinyue Wang

School of Economics and Management, Dalian University of Technology, No.2 Linggong Road, Dalian 116024, China

ARTICLE INFO

Keywords:
Microblog
Sentiment classification
Ensemble method
Features

ABSTRACT

With the rise of microblog services in recent years, microblog sentiment classification has been widely studied and applied in many fields like public opinion monitoring, commodity evaluation and market forecasting. Ensemble methods have been widely used in the feature construction and classification stages of microblog sentiment classification due to their excellent performance. In feature construction, most researchers use feature concatenation or ensemble methods to combine different features while the fusion of two methods is ignored. In classification, most ensemble classification methods combine classifiers based on majority voting or weighted averaging, and they do not fully consider the differences in the information contained in classifiers. In this paper, a novel multi-view ensemble learning method is proposed to fuse the information contained in different features for better microblog sentiment classification. This method consists of two stages: the local fusion stage and the global fusion stage. In the local fusion stage, the raw features and concatenation features are used to construct basic classifiers, and these basic classifiers are combined into five classifier groups to identify the microblog sentiment among all raw feature information. In the global fusion stage, these classifier groups with a global view are further integrated to make more accurate and comprehensive predictions. Two public microblog benchmark datasets provided by Sina Weibo are used in the experiment, and the experimental results show that our method outperforms other compared methods in identifying the polarities of microblog posts.

1. Introduction

Microblogs such as Sina Weibo and Twitter are popular social media services on which millions of users exchange their opinions, attitudes and emotions with others via posts (Jiang et al., 2015). A lot of posts with different sentiments are generated on microblogs. How to identify the actual sentiment of each post, which is regarded as microblog sentiment classification, is meaningful for analyzing the opinions and attitudes of users. Microblog sentiment classification has been widely used to improve decision-making in many domains, such as public opinion monitoring (D'Andrea, Ducange, Bechini, Renda, & Marcelloni, 2019; Xiang, Soon, Zhi, & James, 2015), commodity evaluation (Mostafa, 2013; Mou & Du, 2016) and market forecasting (Checkley, Higon, & Alles, 2017; Groß-Klußmann, König, & Ebner, 2019; Oliveira, Cortez, & Areal, 2017), and has achieved promising performances (Ibrahima & Wang, 2019; Oliveira et al., 2017).

Due to its enormous practical value, microblog sentiment classification has attracted increasingly more attention from researchers, companies and governments (Bermingham & Smeaton, 2010; Hu, Tang, Tang, & Liu, 2013; Liu, Li, & Guo, 2012; Yoo, Song, & Jeong, 2018). The lexicon-based methods were first used on sentiment classification, which uses the emotion lexicons as the main basis to identify the sentiment of sentences. In recent years, machine learning-based microblog sentiment classification methods have achieved excellent performance (Jiang et al., 2015); Tellez et al., 2017; Tripathy, Agrawal, & Rath, 2016; Wang, Sun, Ma, Xu, & Gu, 2014). Most of these methods mainly include two stages: the feature construction stage and the classification stage (Wang et al., 2014). In the feature construction stage, microblog posts are mapped into feature vectors using the text representation method. Various text representation methods have been proposed in previous studies, such as the bag-of-words (BOW) (da Silva, Hruschka, & Hruschka, 2014), n-grams (Bermingham & Smeaton, 2010; Kouloumpis, Wilson, & Moore, 2011; Park & Paroubek, 2010; Tripathy et al., 2016), parts of speech (Bermingham & Smeaton, 2010; Park & Paroubek, 2010) and word embedding techniques (Jiang et al., 2015; Tang et al., 2014). Since most of these methods only consider the partial information of

* Corresponding author.
E-mail addresses: yexin@dlut.edu.cn (X. Ye), dhx.dlut.edu.cn@mail.dlut.edu.cn (H. Dai), 1097992798@mail.dlut.edu.cn (L.-a. Dong), xinyuew@mail.dlut.edu.cn (X. Wang).

https://doi.org/10.1016/j.eswa.2020.113987
Received 26 February 2020; Received in revised form 3 September 2020; Accepted 7 September 2020
Available online 22 September 2020

microblog posts, a model constructed using a single feature may identify the sentiment of posts from a local view. Therefore, many researches proposed considering features created by different text representation methods during model construction, which is useful to comprehensively identify the sentiment from multiple views. These researches could be divided into two groups. In the first group, different features are concatenated into one new feature vector. The new feature vector provides multiple views to construct the microblog sentiment classification model (Mohammad, Kiritchenko, & Zhu, 2013; Tripathy et al., 2016). In the other group, different features are used to construct diverse basic classifiers, respectively, and these basic classifiers provide different local preferences for sentiment classification (Clark & Wicentwoski, 2013; Cotelo, Cruz, Enríquez, & Troyano, 2016). Compared to considering a single feature, both of them achieve more competitive performance. However, few researches consider the reasonable combination of the two groups of methods. In the classification stage, identifying the microblog sentiment is regarded as the classification problem, and many classification algorithms could be used to solve this problem. Bravo-Marquez, Mendoza, and Poblete (2013) compared the Twitter sentiment classification performance of the Decision Tree (DT), Naïve Bayes (NB), Logistic Regression (LR) and Support Vector Machine (SVM). The experimental results reveal that the SVM achieves the best results among all methods. Jiang et al. (2015) proposed an emoticon space model for microblog sentiment classification in which the SVM was used to identify the sentiment of microblog posts. In recent years, ensemble methods were introduced to microblog sentiment classification, and they have outperformed individual methods by combining a set of homogeneous or heterogeneous basic classifiers (Araque, Corcuera-Platas, Sánchez-Rada, & Iglesias, 2017; da Silva et al., 2014; Lin & Kolcz, 2012; Prusa, Khoshgoftaar, & Dittman, 2015; Hassan, Abbasi, & Zeng, 2013). However, most ensemble methods combine basic classifiers using the majority voting method or weighted averaging method, which do not fully consider the differences of the information contained in basic classifiers.

In this paper, a novel multi-view ensemble learning method is proposed for microblog sentiment classification. To consider more comprehensive information for microblog sentiment classification, two text representation methods, namely, the Chi-square statistic method and the emoticon space mapping method, are considered to construct feature vectors; and, moreover, the emotional element statistical method is proposed as a supplement to the two text representation methods from another view. The different raw features are combined in different ways to construct the basic classifiers. To better integrate different basic classifiers, a local fusion stage and a global fusion stage are considered, respectively. In the local fusion stage, to identify the microblog sentiment from all raw feature information, the basic classifiers are combined into five classifier groups. In the global fusion stage, these classifier groups with a global view are further integrated to make more accurate and comprehensive predictions. To evaluate the performance of the proposed method, two public Chinese microblog benchmark datasets are applied in this paper's experiments. The experimental results demonstrate that the proposed method enhances the performance of microblog sentiment classification by effectively fusing multi-view information. Compared with other compared methods, the proposed method is competitive on microblog sentiment classification. To ensure reproducibility and repeatability, the experiment code for our proposed model (TSEF) has been uploaded to Github.[1]

The remainder of this paper is organized as follows. Section 2 reviews the related works on microblog sentiment classification, and the details of the multi-view ensemble learning method for microblog sentiment classification are described in Section 3. The experimental results are illustrated and discussed in Section 4. Finally, the conclusion and future work directions are provided in Section 5.

[1] https://github.com/xinyuew-ec/Microblog-sentiment-classification.

2. Related work

In this paper, a summary of microblog sentiment classification is presented from two following aspects. Section 2.1 discusses the feature construction for microblog sentiment classification, and Section 2.2 outlines the classification methods for microblog sentiment classification.

2.1. Feature construction for microblog sentiment classification

Constructing a feature vector to effectively represent a piece of a microblog post is an important part of a machine learning-based microblog sentiment classification method. In the past years, researchers have proposed several text representation methods to construct features by considering different information contained in posts.

The bag-of-words (BOW) is a sample and traditional text representation method for feature construction. It uses an unordered set of words to express a text or a document (Catal & Nangir, 2016; Cotelo et al., 2016; da Silva et al., 2014; Pang & Lee, 2008), and thus there are two major limitations associated with its application. First the feature vectors constructed based on the BOW have high dimensionality, which may cause the curse of dimensionality in many classification algorithms. Therefore, some feature selection methods are used with the BOW, such as the document frequency (DF), information gain (IG), mutual information (MI), Chi-square statistic (CHI), and term strength (TS). Second, the BOW ignores some important information contained in microblog posts, such as the word order information, syntactic structures and semantic relationships between words (Xia & Zong, 2010). Thence, some other text representation methods are adopted for microblog sentiment classification such as n-grams (Mohammad, Kiritchenko, & Zhu, 2013; Park & Paroubek, 2010; Saleh, Martín-Valdivia, Montejo-Ráez, & Ureña-López, 2011; Zhang, Ghosh, Dekhil, Hsu, & Liu, 2011) and part-of-speech (POS) (Agarwal, Xie, Vovsha, Rambow, & Passonneau, 2011; Mohammad, Kiritchenko, & Zhu, 2013; Park & Paroubek, 2010). With the development of deep learning techniques, word embedding techniques are widely used for converting words into feature vectors for microblog sentiment classification (Kim, 2014; Ren, Zhang, Zhang, & Ji, 2016; Tang et al., 2014; Xiong, Lv, Zhao, & Ji, 2017; Zhang, Xu, Su, & Xu, 2015). Tang et al. (2014) proposed the sentiment specific word embedding (SSWE) for Twitter sentiment classification, which not only models the syntactic context of words but also encodes sentiment information in the continuous representation of words. Following SSWE, Ren et al. (2016) proposed the topic-enriched multi-prototype word embedding (TMWE), which incorporates the topic information into word embeddings. Xiong et al. (2017) proposed a multi-level sentiment-enriched word embedding model (MSWE) to integrate word-level sentiment information and sentence-level sentiment information in word embedding learning. Based on the traditional word embedding techniques, Jiang et al. (2015) projected the words and microblog posts into an emoticon space to construct the feature vector. Experimental results indicated that this method could effectively leverage emoticon signals and outperformed previous methods on microblog sentiment classification.

In general, the single text representation method only considers the partial information in microblog posts. Therefore, some researchers proposed combining the features constructed by different text representation methods into one feature vector for better microblog sentiment classification. Park and Paroubek (2010) used n-grams and POS tags as features and achieved better performance on Twitter sentiment classification. Zhang et al. (2011) combined unigram features, emoticon features and hashtag features into one feature vector for Twitter sentiment classification. Mohammad, Kiritchenko and Zhu (2013) combined n-gram features and POS features with other features constructed from emoticons, punctuations and negations for Twitter sentiment classification. The experimental results showed that the concatenation features

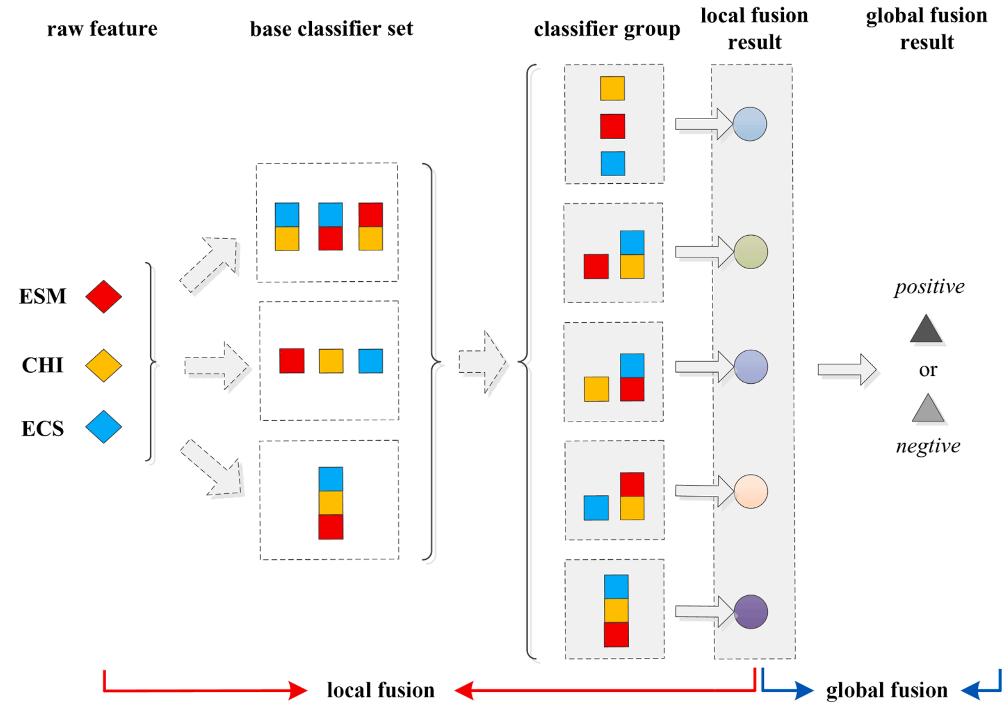

Fig. 1. The framework of the multi-view ensemble learning method.

led to the best prediction performance improvement. Nguyen and Nguyen (2019) introduced a novel approach named the freezing scheme to combine the CNN-f (Convolutional Neural Network) sentiment feature vectors and LSTM-f (Long Short-Term Memory) sentiment feature vectors generated from the CNN and LSTM models, respectively. The experimental results show that feature combining scheme achieves higher accuracy than those of both the CNN-f and LSTM-f. The feature concatenating method is widely used to construct a new feature vector in these methods. Besides concatenating features, some researches combine different features for sentiment classification via the ensemble method, where each basic classifier is constructed by different features, respectively. Clark and Wicentwoski (2013) constructed NB classifiers using different features, such as n-gram features, sentiment lexicon features, parts of speech and special token features, and combined them based on a confidence-weighted voting schema. Cotelo et al. (2016) proposed combining the textual content of tweets and the structural information of a social network for Twitter sentiment classification. The experimental results show that concatenation features do not always get very good results while an ensemble of classifiers trained with different features performs better.

The features constructed by different text representation methods provide valuable information from different views. Feature concatenating and ensemble methods provide two effective ways to consider more different information for microblog sentiment classification, which could improve the prediction performance (Xia, Zong, & Li, 2011). However, few researches consider the fusion of the two methods and discuss the effectiveness of existing methods.

2.2. Microblog sentiment classification methods

In the classification stage, many classification methods are adopted to identify the microblog sentiment. The existing classification methods could be divided into two groups: individual methods and ensemble methods.

The individual methods, such as the Naïve Bayes (NB) (Bermingham & Smeaton, 2010; Bravo-Marquez et al., 2013; Park & Paroubek, 2010; Sailunaz & Alhajj, 2019), Support Vector Machine (SVM) (Cortes & Vapnik, 1995; Dave, Lawrence, & Pennock, 2003; Mohammad, Kiritchenko, & Zhu, 2013; Pang, Lee, & Vaithyanathan, 2002; Zhang et al., 2011), k-nearest neighbor (KNN) (Davidov, Tsur, & Rappoport, 2010; Hand, Mannila, & Smyth, 2001) and so on, mainly employ a stand-alone classification algorithm to identify the sentiment of microblog posts. Davidov et al. (2010) proposed a supervised sentiment classification framework that used the KNN classifier for Twitter sentiment classification. Pak and Paroubek (2010) proposed using the Multinomial Naïve Bayes (MNB) to construct a sentiment classifier for Twitter sentiment classification and obtained better performance than previous methods. Bermingham and Smeaton (2010) employed the MNB and SVM for microblog sentiment classification, and their experiment results showed that the MNB outperforms the SVM in the short-form domains, but the opposite is true in the long-form domains. Bravo-Marquez et al. (2013) compared the Decision Tree (DT), NB, Logistic Regression (LR) and SVM on microblog sentiment classification, and their results showed that the SVM achieved the best performance with respect to the accuracy and F-measure.

Compared with individual methods, ensemble methods obtain more stable and accurate prediction results by combining a set of basic classifiers. Rodriguez-Penagos et al. (2013) proposed an ensemble system for sentiment classification. In the system, three individual classification methods, namely, the Conditional Random Fields, SVM and dictionary-based heuristic method, are combined into an ensemble classification model for sentiment classification. Compared with individual components, the ensemble system achieved good performances on Twitter sentiment classification. Khan, Bashir, and Qamar (2014) pipelined three heterogeneous classifiers to identify the sentiment of Twitter posts. Compared with other methods, the proposed ensemble method is more accurate. da Silva et al. (2014) employed MNB, SVM, Random Forest (RF), and LR to construct a novel ensemble method for Twitter sentiment classification. Similarly, Catal and Nangir (2016) proposed using bagging, NB and SVM as basic classifiers to construct the ensemble model for sentiment classification, and the proposed ensemble model achieved better performance than those of the NB and SVM. Siddiqua, Ahsan, and Chy (2016) used probabilistic NB and MNB, SVM, and sequential minimal optimization (SMO) as the basic classifiers to construct an ensemble model, and the ensemble model was further combined with a rule-based classifier. The proposed method showed significant improvements over other methods on Twitter sentiment classification. Saleena (2018) proposed a Twitter sentiment classification ensemble model, which uses NB, RF, SVM and LR as the basic classifiers; and the experiment results show that the proposed ensemble model performs better than individual classifiers.

On microblog sentiment classification, ensemble methods are more competitive than individual methods. Most ensemble methods combine the basic classifiers using majority voting or weighted averaging, which do not comprehensively consider the difference of the information contained in basic classifiers. This limitation may affect the performance of the ensemble model on microblog sentiment classification.

In addition, some studies on multi-view learning have been conducted. Xu, Tao, and Xu (2013) reviewed several current trends of multi-view learning and classified these algorithms into three different settings, including co-training, multiple kernel learning, and subspace learning. Sun (2013) surveyed recent developments on the theories and methodologies of multi-view machine learning and gave a taxonomy of the approaches according to the machine learning mechanisms involved and the fashions in which multiple views are exploited. Figueroa and Atkinson (2019) proposed a subspace supervised learning model to generalize the feature selection task by producing two different views containing partitions of the feature space. The above studies also provide us with references.

3. Multi-view ensemble learning method

3.1. Framework

In this paper, a novel multi-view ensemble learning method is proposed for microblog sentiment classification. The framework of the proposed method is shown in Fig. 1.

In terms of the text representation, the emotional element statistical method, which is a supplement of the Chi-square statistic method and emoticon space mapping method, is proposed to transform five emotional elements contained in a microblog post into a feature vector. These three text representation methods are considered to obtain three different raw features (denoted as ECS, CHI, and ESM), which are used to construct the microblog sentiment classification model in this paper.

Generally, the proposed method mainly consists of two stages: the local fusion stage and the global fusion stage. In the local fusion stage, different raw features and concatenation features are used to construct diverse basic classifiers, where the feature concatenating method is used to combine different raw features. These basic classifiers could provide a multi-view basis for microblog sentiment classification. According to the complementarity of the information contained in the different basic classifiers, these basic classifiers are combined into five classifier groups. The aim is to identify the microblog sentiment when each classifier group contains all the raw feature information. This process is similar to combining the opinions of experts with different knowledge backgrounds in decision-making, and thus it is beneficial to improving the decision-making quality (Kacprzak, 2019; Lin, Lee, Chang, & Ting, 2008; Xu, Duan, & Sun, 2016). Due to the differences in the composition of the base classifiers of these five classifier groups, there may be differences in their decision preferences (Kacprzak, 2019; Lin et al., 2008; Xu et al., 2016). Drawing lessons from group decision-making, the predictive results of these five groups are further integrated based on the predictive performance of each group in the global fusion stage.

3.2. Features

Three text representation methods are considered in the proposed method. In Section 3.2.1, a novel text representation method, namely, the emotional element statistical method, is proposed, which considers the statistical information of five types of emotional elements in microblog posts. Then, the other two methods, including the Chi-square statistic and emoticon space mapping, are introduced in Section 3.2.2 and Section 3.2.3, respectively.

Table 1
The details of two emotion lexicons.

WD_k	category	grade	example	number
WD_1	positive	1	well (安好), meek (温顺), etc.	375
		3	rejoice (庆幸), prosperity (景气), etc.	1748
		5	sincere (诚挚), capable (能干), etc.	4840
		7	noble (高贵), wise (高明), etc.	2907
		9	virtuous (贤惠), perfect (完美), etc.	1138
WD_2	negative	1	timid (胆怯), dim (暗淡), etc.	430
		3	relax (松懈), sham (虚假), etc.	1825
		5	impulsive (浮躁), obstinate (固执), etc.	4657
		7	distasteful (厌恶), malicious (恶毒), etc.	2714
		9	treacherous (奸诈), hypocritical (虚伪), etc.	974

3.2.1. Emotional element statistical method

The emotional expression of each microblog post depends on many elements (Barbosa & Feng, 2010; Davidov et al., 2010; Jiang et al., 2015; Shiota & Keltner, 2005; Siddiqua et al., 2016), such as emotion words, emoticons, etc. In this paper, the emotional element statistical method is proposed to transform the emotional elements into a feature vector, which considers five elements, including the emotion words, emoticons, parts of speech, punctuation and special structures in microblog posts.

(1) Emotion words

Emotion words are the words with emotional colorings, such as “happy” and “angry”; and they can convey information about internal states, attitudes, beliefs, social contexts, elicitors, motivations, values, behaviors, and many other referents (Shiota & Keltner, 2005). In a microblog post, emotion words intuitively reflect the emotional propensity of the author. The sentiment of a microblog post largely depends on these emotion words.

To evaluate the performance of positive emotion words and negative emotion words, the positive emotion word feature WF_1 and the negative emotion word feature WF_2 are calculated as in Eq. (1).

$$WF_k = \sum_{w_i \in WD_k} count(w_i, M) \cdot grade(w_i) \tag{1}$$

where w_i is the emotion word in the emotion lexicons WD_k. WD_1 represents the positive emotion lexicons and WD_2 represents the negative emotion lexicons.$count(w_i, M)$ represents the frequency that emotion word w_i appears in microblog post M. $grade(w_i)$ denotes the emotional grade of the emotion word w_i. The emotion lexicons used in this paper are obtained by combining the Emotion Words Ontology Library provided by the Information Retrieval Laboratory of Dalian University of Technology (DUTIR)[2] and the emotion lexicons from HowNet[3]. Besides, some network words frequently used on Sina Weibo[4] are extended to the emotion lexicons. In the emotion lexicons, the sentiment intensities of emotion words are divided into five different grades, as shown in Table 1. Higher grades indicate greater intensity. WF_1 and WF_2 denote the positive intensity and negative intensity of a microblog post, respectively.

(2) Emoticons

There are a great number of emoticons in microblogs. Compared with emotion words, these emoticons are more emotional and more widely used, and they could be used as one of the most important signals

[2] http://ir.dlut.edu.cn/EmotionOntologyDownload.
[3] http://www.keenage.com/html/c_index.html.
[4] https://weibo.com/.

Table 2
The details of two emoticon lexicons.

ED_k	category	grade	example	number
ED_1	positive	1	 (hee hee),  (smile), etc.	25
		2	 (too happy),  (good), etc.	14
ED_2	negative	1	 (doubt),  (aggrieved), etc.	39
		2	 (crazy),  (scolded), etc.	22

Table 3
The description of parameter POS_k.

POS_k	part of speech	example
POS_1	degree adverbs	very (非常), more (更加), etc.
POS_2	other adverbs	never (从不), why(为什么), etc.
POS_3	turning conjunctions	but (但是), however (然而), etc.
POS_4	exclamation	yeah (耶), oh (哦), etc.
POS_5	modal particles	wow (哇哦), geez (天呐), etc.
POS_6	pronouns	what (什么), how (怎么), etc.
POS_7	auxiliary words	of (的), is (得), etc.
POS_8	adjectives	serious (严肃), beautiful (美丽), etc.
POS_9	verbs	hate (讨厌), criticize (批评), etc.
POS_{10}	nouns	happiness (幸福), health (健康), etc.

Table 4
The description of parameter PUN_k.

PUN_k	punctuation
PUN_1	3 question marks or more
PUN_2	3 exclamation marks or more
PUN_3	3 full stops or more
PUN_4	1 ellipsis or more

Table 5
The description of parameter STR_k.

STR_k	structure	example
STR_1	negative word/prefix + negative emotion word	not angry (不气愤), not mistake (没有犯错), etc.
STR_2	negative word/prefix + positive emotion word	unhappy (不开心), unsightly (不好看), etc.
STR_3	negative word/prefix + other word	unworthy (不值得), unsatisfactory (不满意), etc.
STR_4	degree adverb + negative emotion word	great distaste (非常厌恶), very annoying(非常讨厌), etc.
STR_5	degree adverb + positive emotion word	really fond of (很喜欢), very cute (非常可爱), etc.
STR_6	negative word/prefix + degree adverb + emotion word	not very nice (没有很好看), not particularly good (不是特别好看), etc.
STR_7	negative word/prefix + negative word + emotion word	not dislike (没有不喜欢), not unhappiness (没有不开心), etc.

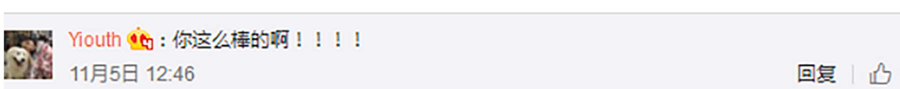

Fig. 2. An example of the continuous and repeated use of punctuation.

for microblog sentiment classification (Jiang et al., 2015; Muhammad, Wiratunga, & Lothian, 2016; Siddiqua et al., 2016; Zhang et al., 2013).

Similar to the evaluation method of emotion words, positive emoticon feature EF_1 and negative emoticon feature EF_2 are computed as follows.

$$EF_k = \sum_{e_i \in ED_k} count(e_i, M) \cdot grade(e_i) \tag{2}$$

where e_i is the emoticon in the emoticon lexicons ED_k. ED_1 represents the positive emoticon lexicons and ED_2 represents the negative emoticon lexicons. $count(e_i, M)$ denotes the frequency that emoticon e_i appears in microblog post M. $grade(e_i)$ is the emotional grade of emoticon e_i. The emoticon lexicons are constructed based on the emoticons collected from Sina Weibo. The emotional categories and grades of these emoticons are labeled according to emotions of the corresponding Chinese words and usage habits of netizens. In emoticon lexicons, the sentiment intensity of emoticons are divided into two different grades, as shown in Table 2. Higher grades indicate greater intensity. EF_1 and EF_2 reflect the degree of emotional propensity of a positive emoticon and negative emoticon in a microblog post, respectively.

(3) Parts of speech

Besides emotion words and emoticons, some words with special parts of speech influence the sentiment of microblog posts, such as degree adverbs, turning conjunctions, interjections, modal particles, etc. (Park & Paroubek, 2010). For example, authors always uses degree adverbs to weaken or strength the sentiment intensity of sentiment words. Besides, a microblog post is sometimes a complex sentence, which consists of some simple short sentences and conjunctions. Using different types of conjunctions may express different sentiment (Hatzivassiloglou & McKeown, 1997). For example, two sentences connected by progressive conjunctions usually have the same sentiment while two sentences connected by turning conjunctions are different.

According to (Barbosa & Feng, 2010; Siddiqua et al., 2016), 10 special parts of speech are considered in the proposed method, as shown in Table 3. The frequency TF_k at which the k-th part of speech POS_k appears in microblog post M is counted according to Eq. (3)

$$TF_k = \sum_{p_i \in POS_k} count(p_i, M) \tag{3}$$

where p_i is one word in the part of speech POS_k, and $count(p_i, M)$ represents the number of times that p_i appears in microblog post M. In this paper, a Chinese word segmentation utility named jieba[5] is used to tag the POS of words.

(4) Punctuation

The continuous and repeated use of punctuation is a common phenomenon on microblogs, which is usually a way for an author to express and emphasize his/her emotions (Bermingham & Smeaton, 2010; Davidov et al., 2010), such as shown in Fig. 2. It should be noted that even if there is no emotion word in the microblog post, the continuous and repeated use of punctuation could add a certain sentiment to the microblog post.

[5] https://pypi.org/project/jieba/.

$$ECS = [\underbrace{WF_1 \;\; WF_2}_{\text{Emotion word}} \;\; \underbrace{EF_1 \;\; EF_2}_{\text{Emoticon}} \;\; \underbrace{TF_1 \;\; TF_2 \;\; \cdots \;\; TF_{10}}_{\text{Part of speech}} \;\; \underbrace{PF_1 \;\; PF_2 \;\; PF_3 \;\; PF_4}_{\text{Punctuation}} \;\; \underbrace{SF_1 \;\; SF_2 \;\; \cdots \;\; SF_7}_{\text{Structure}}]$$

Fig. 3. The emotional element statistical feature.

Table 6
The relationship between words and sentiment categories.

	belong to c_j	do not belong to c_j
posts contain m_i	A	B
posts do not contain m_i	C	D

In this paper, four types of punctuation are considered, as described in Table 4. Using Eq. (4), PF_k is calculated to measure the frequency that the k-th type of punctuation PUN_k appears in microblog post M.

$$PF_k = count(PUN_k, M) \tag{4}$$

(5) Structure

In a microblog post, authors sometimes enhance an emotion or express an opposite emotion by combining several words, which is defined as "structure" in this paper. Compared to analyzing a single word in (1) and (3), these structures can help us identify the microblog sentiment from a high-level perspective. In this paper, seven types of structures are considered, as shown in Table 5. The measurement SF_k of each type of structure is defined as Eq. (5).

$$SF_k = \sum_{t_i \in STR_k} count(t_i, M) \tag{5}$$

where t_i is one instance in the k-th type of structure STR_k, $count(t_i, M)$ represents the number of times that t_i appears in microblog post M. SF_k indicates the degree of emotional propensity caused by structure STR_k.

(6) Emotional element statistical feature

The emotional element statistical feature vector considers the above-mentioned five elements, including the emotion words, emoticons, parts of speech, punctuation and special structures in a microblog post. These elements are further concatenated to generate the emotional element statistical feature ECS, as shown in Fig. 3.

3.2.2. Chi-square statistic

The Chi-square statistic method measures the degree of association between a word and the sentiment category in a microblog post (Yang & Pedersen, 1997). The Chi-square statistic value is formalized as Eq. (6).

$$CHI(m_i, c_j) = \frac{N \times (AD - CB)^2}{(A + C) \times (B + D) \times (A + B) \times (C + D)} \tag{6}$$

where m_i is the i-th word and c_j is the j-th sentiment category. A, B, C, and D represent the numbers of the four cases in Table 6. N denotes the total number of microblog posts. The words with high Chi-square values are selected since they are more important for microblog sentiment classification.

3.2.3. Emoticon space mapping

The emoticon space mapping method proposed in Jiang et al. (2015) constructs features through projecting the microblog posts into an emoticon space. In this method, the distributed representation vectors of all words are learned by the pre-trained *word2vec* (Mikolov, Chen, Corrado, & Dean, 2013; Mikolov, Sutskever, Chen, Corrado, & Dean, 2013). Two matrixes R_w and R_e are used, where R_w contains the representation vectors of words and R_e includes the representation vectors of emoticons. R_e is a sub-matrix of R_w. The cosine distance is used to identify the semantic similarity between word w_m and emoticon e_n, as shown in Eq. (7)

$$similarity(w_m, e_n) = \frac{\overrightarrow{w_m} \cdot \overrightarrow{e_n}}{|\overrightarrow{w_m}||\overrightarrow{e_n}|} \tag{7}$$

where $\overrightarrow{w_m}$ and $\overrightarrow{e_n}$ represent the distributed representation vectors of w_m and e_n, respectively. The coordinated matrix of all words and emoticons is denoted as C. The emoticon space mapping feature is constructed through the concatenation of the sum, maximum and minimum values of the coordinates of words and emoticons. More details of the emoticon space mapping method are given in Jiang et al. (2015).

3.3. Local fusion stage

3.3.1. Base classifier construction

In Section 3.2, the emotional element statistical method, Chi-square statistic method and emoticon space mapping method are used to construct three raw features (denoted as ECS, CHI, and ESM, respectively) from three different views, respectively.

Chi-square statistic method mainly focuses on the words with high Chi-square values that can effectively distinguish the sentiment categories. Although the Chi-square statistic method is simple and quite efficient on sentiment classification, it discards a great deal of information such as the word order information, syntactic structures and semantic relationships of microblog posts (Xia et al., 2011). Compared with the Chi-square statistic method, the emoticon space mapping method considers more about the word order and syntactic structures when using the *word2vec* pre-training model. Moreover, by projecting microblog posts into the emoticon space, the emoticon space mapping method makes full use of the rich emoticon signals and improves the sentiment classification performance greatly. However, both the Chi-square statistic method and emoticon space mapping method ignore the importance of emotional elements to some extent. The proposed emotional element statistical method can be seen as a supplement to the Chi-square statistic method and emoticon space mapping method from another view. Obviously, different text representation methods focus on different contents of microblog posts, and the information contained in different feature vectors has differences.

In the proposed method, three types of raw features are combined in different ways to construct the base classifier. The base classifier can be divided into three sets.

Set 1: The base classifiers are constructed using only one of the features among CHI, ECS and ESM.
Set 2: The base classifiers are constructed using two of features among CHI, ECS and ESM.
Set 3: The base classifiers are constructed using all three features, including the CHI, ECS and ESM.

Since these base classifiers contain different information, different classifiers could analyze the microblog sentiment from different views. It should be noted that a simple integration of all base classifiers may be problematic. Since the base classifiers of Set 1 and Set 2 do not consider all information in the microblog posts, they may make one-sided predictions of the microblog sentiment, which may affect the final predictions. Therefore, a local fusion strategy that integrates the base classifiers of Set 1 and Set 2 in order to identify the sentiment when each classifier group contains all raw feature information is proposed in Section 3.3.2.

Table 7
The details of NLP&CC 2013 and NLP&CC 2014.

dataset	positive			negative						neutral
	like	happiness	total	sadness	anger	disgust	surprise	fear	total	
NLP&CC 2013	2153	1486	3639	1129	671	1360	349	151	3660	6701
NLP&CC 2014	3246	1900	5146	1362	797	1781	524	194	4658	10,194

3.3.2. Local fusion strategy

According to the principle of complementary information, the base classifiers of the three sets are integrated.

Through the local fusion strategy, five classifier groups are constructed, as shown in Fig. 1. Since each classifier group considers all the information of a microblog post through the complementary information, each classifier group will make a more accurate prediction. In each classifier group, the logistic regression is used as the meta-classifier to integrate the prediction results of the base classifiers (Liu, 2005). The meta-classifier trains on the new meta-data matrix Ω, which is constructed by the base classifier. The new meta-data matrix is described as Eq. (8).

$$\frac{\Omega = [I_1, I_2, \cdots, I_i, \cdots, I_{n-1}, I_n]}{I_i = [O_{i1}, O_{i2}, \cdots, O_{ij}, \cdots, O_{i,z-1}, O_{i,z}, c]} \tag{8}$$

where I_i is a vector that includes the prediction results of the base classifiers for the i-th microblog post and the actual sentiment c of the i-th microblog post. O_{ij} represents the prediction result of j-th base classifier for the i-th post.

3.4. Global fusion stage

In the local fusion stage, five classifier groups are constructed using the base classifiers. These five classifier groups are similar to five expert groups. Since the knowledge background of the experts included in each group may be different, the prediction results of the different expert groups for the same thing may also differ. Therefore, the prediction results of the five classifier groups need to be further integrated in the global fusion stage. In this paper, the accuracy-based weighted method (Ankit & Saleena, 2018) is used, as shown in Eq. (9).

$$prob = \sum_{q=1}^{5} acc_q \times O_q / \sum_{q=1}^{5} acc_q \tag{9}$$

where acc_q is the accuracy of the q-th classifier group, and O_q is the prediction result of the q-th classifier group for a microblog post. *prob* represents the weighted prediction result of the microblog post. The sentiment category of each microblog post is classified as in Eq. (10).

$$Sentiment = \begin{cases} positive & if\ prob \geqslant 0.5 \\ negative & if\ prob < 0.5 \end{cases} \tag{10}$$

4. Experimental results and discussion

4.1. Dataset

To evaluate the performance of the proposed method, two labeled Chinese microblog benchmarked datasets are used in all experiments, including NLP&CC 2013[6] and NLP&CC 2014[7]. These two datasets have been widely used in microblog sentiment classification. In the two datasets, each microblog post is collected from Sina Weibo and is annotated with one of the following eight emotional labels: neutral, like, happiness, sadness, disgust, anger, surprise and fear. The details of NLP&CC 2013 and NLP&CC 2014 are given in Table 7.

[6] http://tcci.ccf.org.cn/conference/2013/pages/page04_eva.html.
[7] http://tcci.ccf.org.cn/conference/2014/pages/page04_eva.html.

Table 8
The confusion matrix of the sentiment classification results.

Actual value	Prediction value	
	True	False
True	*TP*	*FN*
False	*FP*	*TN*

To obtain balanced data, the undersampling method is used to get the same number of positive microblog posts and negative microblog posts in the experiment.

4.2. Evaluation criteria

Four evaluation indicators are used to evaluate the performance of methods comprehensively. The confusion matrix of the sentiment classification results is discribed as Table 8.

(1) The accuracy refers to the ratio of the number of correctly predicted samples to the total number of predicted samples, as shown in Eq. (11).

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{11}$$

(2) The precision is the ratio of the number of correctly predicted positive samples to the number of predicted positive samples and is defined as Eq. (12).

$$precision = \frac{TP}{TP + FP} \tag{12}$$

(3) The recall refers to the ratio of the number of correctly predicted positive samples to the total number of actual positive is defined as Eq. (13).

$$recall = \frac{TP}{TP + FN} \tag{13}$$

(4) The F-measure is the harmonic mean of the precision and recall:

$$F - measure = \frac{2 \times precision \times recall}{precision + recall} \tag{14}$$

To measure the performances of the methods objectively, 10-fold cross-validation is used in the experiment.

4.3. Selection of the base model

The selection of the base model has a significant impact on the overall prediction performance of a method (Onan, Korukoğlu, & Bulut, 2016). To gain better prediction performance, eight methods are considered as the base model, including the following:

(1) The Decision Tree (DT) is a decision process that uses divide and conquer. The key of the decision tree algorithm is the optimal partition attribute selection. In this paper, the information

Table 9
The performance of the proposed method with different base models on NLP&CC 2013.

Algorithm	Positive			Negative			Accuracy
	Precision	Recall	F-measure	Precision	Recall	F-measure	
DT-E	86.17	83.18	84.58	83.87	86.56	85.13	84.87
DT-G	85.31	82.83	84.01	83.39	85.69	84.48	84.25
NB	75.35	86.26	80.42	83.96	71.72	77.32	78.99
KNN	82.64	81.01	81.79	81.42	82.96	82.17	81.99
SVM-R	85.21	83.21	84.17	83.64	85.52	84.54	84.36
SVM-P	84.92	85.05	84.97	85.08	84.88	84.96	84.97
SVM-S	73.79	73.61	73.65	73.69	73.73	73.66	73.67
SVM-L	87.25	85.30	86.26	85.63	87.52	86.56	86.41

Table 10
The performance of the proposed method with different base models on NLP&CC 2014.

Algorithm	Positive			Negative			Accuracy
	Precision	Recall	F-measure	Precision	Recall	F-measure	
DT-E	82.52	86.07	84.24	85.45	81.73	83.53	83.90
DT-G	83.18	86.24	84.67	85.75	82.55	84.10	84.39
NB	73.50	84.87	78.76	82.08	69.34	75.15	77.10
KNN	81.98	83.88	83.42	83.70	82.76	83.22	83.32
SVM-R	85.43	84.22	84.81	84.47	85.64	85.04	84.93
SVM-P	83.11	86.69	84.86	86.07	82.37	84.19	84.53
SVM-S	72.95	76.25	74.55	75.09	71.64	73.30	73.95
SVM-L	86.77	87.18	86.97	87.14	86.69	86.90	86.94

entropy (DT-E) and Gini index (DT-G) are selected to choose the optimal partition attributes.

(2) The Naïve Bayes (NB) is a classification method based on Bayes theorem and the independent hypothesis of feature conditions. In this paper, the multinomial Naïve Bayes (MNB) is selected to learn the model. For all MNBs used in this paper, the additive smoothing parameter is set to 1.0.

(3) The K-Nearest Neighbor (KNN) is a popular non-parametric classification method, which is simple but effective in many cases (Hand et al., 2001). The KNN determines the label of the sample according to the labels of its k nearest samples. The value of k is set to 5.

(4) The Support Vector Machine (SVM) is an effective machine learning method proposed by Cortes and Vapnik (1995). In this paper, four different kernel functions, namely, the linear kernel function (SVM-L), the RBF kernel function (SVM-R), the polynomial kernel function (SVM-P) and the Sigmoid kernel function (SVM-S), are selected to learn the models. For all the SVMs used in this paper, the regularization parameter C is set to 1.0.

Tables 9 and 10 show the prediction performance of the proposed method with different base models on NLP&CC 2013 and NLP&CC 2014, respectively. As shown in Tables 9 and 10, the performance of the proposed method has a significant difference when choosing different base models. Selecting a suitable base model could improve the performance of the proposed method. In Tables 9 and 10, the SVM-L shows a significant advantage over the other base models in terms of the accuracy and F-measure. Especially for the F-measure, the SVM-L achieves higher performance in both positive sentiment and negative sentiment on the two datasets, which indicates that the SVM-L has better overall predictive power on microblog sentiment. As for the recall and the precision, the SVM-L performs better than other base models as a whole. Based on the comprehensive consideration of the accuracy, recall, precision and F-measure, the SVM-L is selected as the base model for the proposed method.

4.4. Comparison of the base classifier and classifier group

To further explore the influence of features constructed by different text representation methods and the fusion of different features on sentiment classification performance, we compare the performance of the base classifiers constructed using different features and the classifier groups formed by the local fusion stage. The experimental results are shown in Tables 11 and 12.

Table 11
The performance measures of the base classifiers and classifier groups on NLP&CC 2013.

Set	Features	Positive			Negative			Accuracy
		Precision	Recall	F-measure	Precision	Recall	F-measure	
SET 1	ECS	73.57	72.02	72.75	72.67	74.09	73.33	73.06
	CHI	85.25	81.43	83.28	82.25	85.90	84.03	83.66
	ESM	84.51	82.52	83.48	82.94	84.80	83.84	83.66
SET 2	CHI-ECS	85.66	82.38	83.97	83.07	86.18	84.58	84.28
	ESM-ECS	84.24	83.29	83.74	83.52	84.36	83.91	83.83
	CHI-ESM	86.07	83.76	84.89	84.19	86.43	85.28	85.09
SET 3	CHI-ESM-ECS	86.00	83.68	84.81	84.10	86.34	85.20	85.01
Group	ECS, CHI-ESM	86.16	84.17	85.15	84.52	86.45	85.47	85.31
	CHI, ESM-ECS	86.94	84.80	85.85	85.18	87.25	86.19	86.03
	ESM, CHI-ECS	87.03	84.69	85.83	85.12	87.36	86.21	86.03
	ECS, CHI, ESM	86.17	84.72	85.42	85.00	86.37	85.66	85.55

Table 12
The performance measures of the base classifiers and classifier groups on NLP&CC 2014.

Set	Features	Positive			Negative			Accuracy
		Precision	Recall	F-measure	Precision	Recall	F-measure	
SET 1	ECS	73.91	73.76	73.83	73.82	73.96	73.88	73.86
	CHI	83.89	84.24	84.05	84.20	83.81	84.00	84.03
	ESM	84.60	85.08	84.83	85.01	84.50	84.75	84.79
SET 2	CHI-ECS	84.61	85.06	84.82	85.01	84.52	84.75	84.79
	ESM-ECS	85.01	85.79	85.39	85.66	84.86	85.26	85.33
	CHI-ESM	85.02	87.51	86.24	87.13	84.56	85.82	86.03
SET 3	CHI-ESM-ECS	85.24	87.44	86.32	87.11	84.84	85.95	86.14
Group	ECS, CHI-ESM	85.36	87.38	86.35	87.08	84.99	86.01	86.18
	CHI, ESM-ECS	86.47	86.50	86.48	86.50	86.45	86.47	86.47
	ESM, CHI-ECS	86.35	86.82	86.58	86.76	86.26	86.50	86.54
	ECS, CHI, ESM	86.83	86.45	86.63	86.52	86.86	86.68	86.66

(1) Comparison of base classifiers

As shown in Tables 11 and 12, the base classifiers SET 2 and SET 3 perform better than SET 1, which indicates that it is valuable to build models with multi-view features. However, it should be noted that the performance of SET 3 is not higher than those of the base classifiers of SET 2 in both data sets as expected. The base classifier of CHI-ESM has better performance than that of CHI-ESM-ECS in NLP&CC 2013, which means that the concatenation feature commonly used in research may not always be effective (Abbasi, Chen, Thoms, & Fu, 2008; Cotelo et al., 2016). The reason is that there may be conflicts among the features in different views. These conflicts affect the performance of the model (Cotelo et al., 2016). Therefore, how to effectively fuse the multi-view features becomes increasingly more important. In addition, when the emotional element statistical feature is added to the base model of CHI and ECS, the performance of the base model is improved. The phenomenon proves that the emotional element statistical method is an effective supplement to the Chi-square statistic method and emoticon space mapping method.

(2) Comparison between base classifiers and classifier groups

Compared with the base classifiers of SET 1 and SET 2, the classifier groups achieve better performance. The information fusion based on complementary information enables the classifier to identify microblog sentiment from the global view, which is useful to improving the performance of the model. Furthermore, both SET 3 and the classifier groups identify the microblog sentiment from the global view. Furthermore, the classifier groups perform better than SET 3 in terms of the accuracy, F-measure, recall, and precision. This phenomenon further indicates that the simple concatenation of features may be problematic. Besides, what is more interesting is that classifier group (CHI, ESM-ECS) and classifier group (ESM, CHI-ECS) achieve better performance than classifier group (ECS, CHI, ESM) on NLP&CC 2013. The phenomenon indicates that the fusion of feature concatenating and ensemble methods may leverage distinct strengths and enhance the sentiment classification performance. Therefore, the efficient fusion of two methods is worthy of further study and application to microblog sentiment classification.

4.5. Performance evaluation

To evaluate the performance of the proposed method, five classical microblog sentiment classification methods and six classical ensemble approaches are compared, including the following:

- N-SVM: The linear Support Vector Machine with Unigrams, Bigrams and Trigrams is selected (Dave et al., 2003; Pang et al., 2002).
- N-MNB: The multinomial Naive Bayes with Unigrams, Bigrams and Trigrams is selected (Bermingham & Smeaton, 2010; Park & Paroubek, 2010).
- WN-SVM: The linear Support Vector Machine with TF-IDF and n-grams is selected (Unigrams, Bigrams and Trigrams) (Saleh et al., 2011).
- D-SVM: The linear Support Vector Machine with pre-trained vectors from *word2vec* is selected (Mikolov, Chen et al., 2013; Mikolov, Sutskever et al., 2013; Zhang et al., 2015). The pre-trained vectors have a dimensionality of 300 and were trained using the continuous bag-of-words architecture.
- CNN: The CNN (Convolutional Neural Network) with pre-trained vectors from *word2vec* is selected (Kim, 2014). The CNN has 3, 4, and 5 filter windows (h) with 128 feature maps each, a dropout rate (p) of 0.5, and a mini-batch size of 64. The pre-trained vectors have a dimensionality of 400 and were trained using the continuous bag-of-words architecture.

Table 13
The performance of the proposed method and compared methods on NLP&CC 2013.

Method	Positive			Negative			Accuracy
	Precision	Recall	F-measure	Precision	Recall	F-measure	
N-SVM	85.36	77.63	81.30	79.48	86.64	82.90	82.14
N-MNB	83.65	81.34	82.46	81.86	84.06	82.93	82.70
WN-SVM	84.93	78.87	81.77	80.31	85.98	83.03	82.43
D-SVM	82.53	81.56	82.03	81.79	82.72	82.24	82.14
CNN	82.83	85.19	83.97	84.81	82.33	83.53	83.76
ES-LR	86.24	84.28	85.24	84.67	86.53	85.57	85.41
ES-AW	87.25	84.61	85.91	85.08	87.63	86.33	86.12
AdaBoost	84.98	82.85	83.51	83.75	84.69	83.88	83.77
Bagging	86.79	84.04	85.36	84.57	87.17	85.82	85.60
RF	87.44	82.17	84.71	83.20	88.18	85.60	85.17
Stacking	86.81	84.89	85.82	85.23	87.06	86.12	85.97
TSEF	87.25	85.30	86.26	85.63	87.52	86.56	86.41

Table 14
The performance of the proposed method and compared methods on NLP&CC 2014.

Method	Positive			Negative			Accuracy
	Precision	Recall	F-measure	Precision	Recall	F-measure	
N-SVM	81.12	86.91	83.90	85.89	79.73	82.69	83.32
N-MNB	83.25	83.68	83.46	83.61	83.15	83.37	83.42
WN-SVM	84.78	83.92	84.34	84.08	84.91	84.49	84.41
D-SVM	83.10	82.16	82.62	82.36	83.27	82.81	82.72
CNN	83.79	85.04	84.36	84.88	83.47	84.12	84.25
ES-LR	85.60	87.14	86.35	86.91	85.32	86.10	86.23
ES-AW	86.33	87.40	86.86	87.25	86.15	86.69	86.77
AdaBoost	84.25	85.36	84.58	85.41	83.51	84.20	84.43
Bagging	85.10	87.89	86.46	87.50	84.59	86.00	86.24
RF	84.05	86.45	85.22	86.05	83.56	84.78	85.00
Stacking	85.93	85.70	85.81	85.75	85.96	85.85	85.83
TSEF	86.77	87.18	86.97	87.14	86.69	86.90	86.94

- ES-LR: Three raw features (ECS, ESM and CHI) and four concatenation features (CHI-ECS, ESM-ECS, CHI-ESM and CHI-ESM-ECS) are used to train the diverse base classifiers, and the Logistic Regression is selected as a meta-classifier to integrate the base classifiers.
- ES-AW: Three raw features (ECS, ESM and CHI) and four concatenation features (CHI-ECS, ESM-ECS, CHI-ESM and CHI-ESM-ECS) are used to train the diverse base classifiers, and the accuracy-based weighted method is used to integrate the base classifiers.
- AdaBoost: ECS, ESM and CHI are concatenated and used to train the AdaBoost classifier. The linear Support Vector Machine is selected as the base estimator and the number of estimators is set to 7.
- Bagging: ECS, ESM and CHI are concatenated and used to train the Bagging classifier. The linear Support Vector Machine is selected as the base estimator and the number of estimators is set to 7.
- RF: ECS, ESM and CHI are concatenated and used to train the Random Forest classifier. The Decision Tree is selected as the base estimator and the number of estimators is set to 7.
- Stacking: ECS, ESM and CHI are concatenated and used to train the Stacking classifier. The Decision Tree, Support Vector Machine, K-Nearest Neighbor and Naïve Bayes are selected as base estimators and the Logistic Regression is used to combine the base estimators.

The experimental results of the proposed method (TSEF) and compared methods are shown in Tables 13 and 14. On the two datasets, the proposed method is significantly better than other classical microblog sentiment classification methods with respect to its accuracy, F-measure, recall and precision. As shown in Tables 13 and 14, the accuracy of the proposed method is 86.41% and 86.94% on NLP&CC 2013 and NLP&CC 2014, respectively. Compared with other classical methods, the accuracy of the proposed method is 2.65%–4.27% better on NLP&CC 2013 and 2.53%–4.22% better on NLP&CC 2014, respectively. As for the recall and the precision, the proposed method shows competitive performance on both positive sentiment and negative sentiment. The reason for the improvement is that the proposed method takes into account more information (features) contained in the microblog posts and fully fuses the information (features) from different views.

Compared with classical ensemble approaches, the proposed method achieves better performance with respect to the accuracy and F-measure on NLP&CC 2013 and NLP&CC 2014. The accuracy of the proposed method is 0.29%–2.64% better on NLP&CC 2013 and 0.17%–2.51% better on NLP&CC 2014, respectively. The phenomenon proves that the proposed multi-view ensemble learning method is an effective integration that achieves better microblog sentiment classification.

5. Conclusion and future work

In this paper, a novel multi-view ensemble learning method is proposed to achieve better microblog sentiment classification. We first compared the prediction performance of the proposed method with those of different base models, and the linear Support Vector Machine that is significantly better is selected as the base model for the proposed method framework. Then, we further explored the influence of the features constructed by different text representation methods. The experimental results show that the emotional element statistical method is an effective supplement to the Chi-square statistic method and the emoticon space mapping method from another view. We also compared the performance of the base classifiers constructed using different features and the classifier groups formed by the fusion stage. The experimental results prove that the fusion of the feature concatenating and ensemble methods may leverage distinct strengths and enhance the sentiment classification performance. Through comparing the performance of different sentiment classification methods, we found that the proposed method achieves better microblog sentiment classification performance. This phenomenon further proves that the proposed multi-view ensemble learning method is an effective integration that achieves better microblog sentiment classification.

In addition, there are only three text representation methods explored in this paper. In future work, more representative text representation methods will be considered and their microblog sentiment classification performances will be compared. Furthermore, some other techniques will be considered in the proposed framework to further improve the microblog sentiment classification performance, such as multiple kernel learning, subspaces, and related deep learning algorithms. Besides, we are going to apply this methodological framework to the emergency management, e-government and e-commerce fields to support corresponding decision-making, and multi-class microblog sentiment classification will be studied according to some other practical requirements.

CRediT authorship contribution statement

Xin Ye: Resources, Writing - review & editing, Supervision, Funding acquisition. **Hongxia Dai:** Conceptualization, Methodology, Software, Validation, Writing - original draft. **Lu-an Dong:** Methodology, Writing - original draft, Writing - review & editing. **Xinyue Wang:** Validation, Writing - review & editing.

Declaration of Competing Interest

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

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (Grant Nos. 71874020 and 71533001).

References

Abbasi, A., Chen, H., Thoms, S., & Fu, T. (2008). Affect analysis of web forums and blogs using correlation ensembles. *IEEE Transactions on Knowledge and Data Engineering, 20*(9), 1168–1180.

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of twitter data. In Proceedings of the Workshop on Languages in Social Media, LSM'11, Stroudsburg, PA, USA (pp. 30-38). Association for Computational Linguistics.

Ankit, & Saleena, N. (2018). An ensemble classification system for twitter sentiment analysis. *Procedia Computer Science, 132*, 937–946.

Araque, O., Corcuera-Platas, I., Sánchez-Rada, J. F., & Iglesias, C. A. (2017). Enhancing deep learning sentiment analysis with ensemble techniques in social applications. *Expert Systems with Applications, 77*, 236–246.

Barbosa, L., & Feng, J. (2010). Robust sentiment detection on twitter from biased and noisy data. In *Proceedings of the 23rd international conference on computational linguistics: Posters, Beijing, China* (pp. 36–44). Association for Computational Linguistics.

Bermingham, A., & Smeaton, A. F. (2010). Classifying sentiment in microblogs: Is brevity an advantage?. In *Proceedings of the 19th ACM international conference on information and knowledge management* (pp. 1833–1836).

Bravo-Marquez, F., Mendoza, M., & Poblete, B. (2013). Combining strengths, emotions and polarities for boosting twitter sentiment analysis. In *Proceedings of the second international workshop on issues of sentiment discovery and opinion mining* (pp. 1–9).

Catal, C., & Nangir, M. (2016). A sentiment classification model based on multiple classifiers. *Applied Soft Computing, 50*, 135–141.

Checkley, M. S., Higon, D. Arion, & Alles, H. (2017). The hasty wisdom of the mob: How market sentiment predicts stock market behavior. *Expert Systems with Applications, 77*, 256–263.

Clark, S., & Wicentwoski, R. (2013). SwatCS: Combining simple classifiers with estimated accuracy. In *Proceedings of Second Joint Conference on Lexical and Computational Semantics (*SEM), Seventh International Workshop on Semantic Evaluation (SemEval 2013), Atlanta, Georgia, USA* (pp. 425–429). Association for Computational Linguistics.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297.

Cotelo, J. M., Cruz, F. L., Enríquez, F., & Troyano, J. A. (2016). Tweet categorization by combining content and structural knowledge. *Information Fusion, 31*, 54–64.

D'Andrea, E., Ducange, P., Bechini, A., Renda, A., & Marcelloni, F. (2019). Monitoring the public opinion about the vaccination topic from tweets analysis. *Expert Systems with Applications, 116*, 209–226.

da Silva, Nádia F. F., Hruschka, E. R., & Hruschka, E. R. (2014). Tweet sentiment analysis with classifier ensembles. *Decision Support Systems, 66*, 170–179.

Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the International World Wide Web Conference (WWW)* (pp. 519–528).

Davidov, D., Tsur, O., & Rappoport, A. (2010). Enhanced sentiment learning using twitter hashtags and smileys. In *Proceedings of 23rd International Conference on Computational Linguistics: Posters* (pp. 241–249). Association for Computational Linguistics.

Figueroa, A., & Atkinson, J. (2019). Dual-view learning for detecting web query intents. *Computer, 52*(8), 34–42.

Groß-Klußmann, A., König, S., & Ebner, M. (2019). Buzzwords build momentum: Global financial twitter sentiment and the aggregate stock market. *Expert Systems with Applications, 136*, 171–186.

Hand, D., Mannila, H., & Smyth, P. (2001). In *Principles of Data Mining* (pp. 621–622). The MIT Press.

Hassan, A., Abbasi, A., & Zeng, D. (2013). Twitter sentiment analysis: A bootstrap ensemble framework. In *Proceedings of 2013 International Conference on Social Computing* (pp. 357–364).

Hatzivassiloglou, V., & McKeown, K. R. (1997). Predicting the semantic orientation of adjectives. In *Proceedings of the eighth conference on European chapter of the association for computational linguistics* (pp. 174–181). Association for Computational Linguistics.

Hu, X., Tang, L., Tang, J. L., & Liu, H. (2013). Exploiting social relations for sentiment analysis in microblogging. In *Proceedings of the sixth ACM international conference on web search and data mining* (pp. 537–546). ACM.

Ibrahima, N. F., & Wang, X. J. (2019). A text analytics approach for online retailing service improvement: Evidence from twitter. *Decision Support Systems, 121*, 37–50.

Jiang, F., Liu, Y., Luan, H., Sun, J., Zhu, X., Zhang, M., & Ma, S. (2015). Microblog sentiment analysis with emotion space model. *Journal of Computer Science and Technology, 30*(5), 1120–1129.

Kacprzak, D. (2019). A doubly extended TOPSIS method for group decision making based on ordered fuzzy numbers. *Expert Systems with Applications, 116*, 243–254.

Khan, F. H., Bashir, S., & Qamar, U. (2014). TOM: Twitter opinion mining framework using hybrid classification scheme. *Decision Support Systems, 57*, 245–257.

Kim, Y. (2014). Convolutional neural networks for sentence classification. In *Proceedings of the 2014 conference on empirical methods on Natural Language Processing (EMNLP 2014), Doha, Qatar* (pp. 1746–1751).

Kouloumpis, E., Wilson, T., & Moore, J. (2011). Twitter sentiment analysis: The good the bad and the omg. In *Proceedings of the fifth international AAAI conference on weblogs and social media* (pp. 538–541).

Liu, C. (2005). Classifier combination based on confidence transformation. *Pattern Recognition, 38*(1), 11–28.

Lin, J., & Kolcz, A. (2012). Large-scale machine learning at twitter. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data* (pp. 793–804).

Lin, Y. H., Lee, P. C., Chang, T. P., & Ting, H. (2008). Multi-attribute group decision making model under the condition of uncertain information. *Automation in Construction, 17*(6), 792–797.

Liu, K. L., Li, W. J., & Guo, M. Y. (2012). Emoticon smoothed language models for twitter sentiment analysis. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence* (pp. 1678–1684).

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. In *Proceedings of workshop at ICLR* (pp. 1–12).

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Proceedings of neural information processing systems* (pp. 3111–3119).

Mohammad, S. M., Kiritchenko, S., & Zhu, X. (2013). NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets. Computer Science.

Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications, 40*(10), 4241–4251.

Mou, X., & Du, Y. (2016). Sentiment classification of Chinese movie reviews in micro-blog based on context. *2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*. IEEE.

Muhammad, A., Wiratunga, N., & Lothian, R. (2016). Contextual sentiment analysis for social media genres. *Knowledge-Based Systems, 108*, 92–101.

Nguyen, H. T., & Nguyen, M. L. (2019). An ensemble method with sentiment features and clustering support. *Neurocomputing, 370*, 155–165.

Oliveira, N., Cortez, P., & Areal, N. (2017). The impact of microblogging data for stock market prediction: Using twitter to predict returns, volatility, trading volume and survey sentiment indices. *Expert Systems with Applications, 73*(Complete), 125–144.

Onan, A., Korukoğlu, S., & Bulut, H. (2016). A multiobjective weighted voting ensemble classifier based on differential evolution algorithm for text sentiment classification. *Expert Systems with Applications, 62*, 1–16.

Park, A., & Paroubek, R. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the 7th international conference on language resources and evaluation* (pp. 1320–1326).

Prusa, J., Khoshgoftaar, T., & Dittman, D. (2015). Using ensemble learners to improve classifier performance on tweet sentiment data. In *Proceedings of 2015 IEEE international conference on Information reuse and integration (IRI), San Francisco, CA, USA* (pp. 252–257).

Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)* (pp. 1320–1326).

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*(1–2), 1–135.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: Sentiment classification using machine learning technique. In *Proceedings of the ACL-02 conference on empirical methods in natural language processing* (pp. 79–86).

Ren, Y., Zhang, Y., Zhang, M., & Ji, D. (2016). Improving twitter sentiment classification using topic-enriched multi-prototype word embeddings. In *Proceedings of thirtieth AAAI Conference on Artificial Intelligence (AAAI)* (pp. 3038–3044).

Rodriguez-Penagos, C., Atserias, J., Codina-Filba, J., Garcıa-Narbona, D., Grivolla, J., Lambert, P., & Saur, R. (2013). Fbm: Combining lexicon-based ml and heuristics for social media polarities. In *Proceedings of SemEval-2013—international workshop on semantic evaluation co-located with *Sem and NAACL, Atlanta, Georgia* (pp. 483–489).

Sailunaz, K., & Alhajj, R. (2019). Emotion and sentiment analysis from twitter text. *Journal of Computational Science, 36*.

Saleena, N. (2018). An ensemble classification system for Twitter sentiment analysis. *Procedia Computer Science, 132*, 937–946.

Saleh, M. R., Martín-Valdivia, M. T., Montejo-Ráez, A., & Ureña-López, L. A. (2011). Experiments with SVM to classify opinions in different domains. *Expert Systems with Applications, 38*(12), 14799–14804.

Shiota, M. N., & Keltner, D. (2005). What do emotion words represent? *Psychological Inquiry, 16*(1), 32–37.

Siddiqua, U. A., Ahsan, T., & Chy, A. N. (2016). Combining a rule-based classifier with ensemble of feature sets and machine learning techniques for sentiment analysis on microblog. In *Proceedings of 2016 19th International Conference on Computer and Information Technology (ICCIT)* (pp. 304–309). IEEE.

Sun, S. (2013). A survey of multi-view machine learning. *Neural Computing and Applications, 23*(7–8), 2031–2038.

Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., & Qin, B. (2014). Learning sentiment-specific word embedding for twitter sentiment classification. In *Proceedings of the 52nd Annual meeting of the association for computational linguistic, Baltimore, Maryland, USA* (pp. 1555–1565). Association for Computational Linguistics.

Tellez, E. S., Miranda-Jiménez, S., Graff, M., Moctezumaac, D., Siordia1, O. S., & Villaseñorb, E. A. (2017). A case study of Spanish text transformations for twitter sentiment analysis. *Expert Systems with Applications, 81*, 457–471.

Tripathy, A., Agrawal, A., & Rath, S. K. (2016). Classification of sentiment reviews using n-gram machine learning approach. *Expert Systems with Applications, 57*, 117–126.

Wang, G., Sun, J., Ma, J., Xu, K., & Gu, J. (2014). Sentiment classification: The contribution of ensemble learning. *Decision Support Systems, 57*, 77–93.

Xiang, J., Soon, A. C., Zhi, W., & James, G. (2015). Twitter sentiment classification for measuring public health concerns. *Social Network Analysis and Mining, 5*(1), 13.

Xia, R., & Zong, C. (2010). Exploring the use of word relation features for sentiment classification. In *Proceedings of the 23rd international conference on computational linguistics: Posters, Stroudsburg, PA, USA* (pp. 1336–1344). Association for Computational Linguistics.

Xia, R., Zong, C., & Li, S. (2011). Ensemble of feature sets and classification algorithms for sentiment classification. *Information Sciences, 181*(6), 1138–1152.

Xiong, S., Lv, H., Zhao, W., & Ji, D. (2017). Towards twitter sentiment classification by multi-level sentiment-enriched word embeddings. *Neurocomputing, 275*, 2459–2466.

Xu, C., Tao, D., & Xu, C. (2013). A survey on multi-view learning. *Computer Science*.
Xu, C., Duan, W., & Sun, Y. (2016). The recognition and integration method to composite structure preference of multi-attribute group decision making. *Chinese Journal of Management Science, 24*(8), 139–145.
Yang, Y., & Pedersen, J. O. (1997). A comparative study on feature selection in text categorization. In *Proceedings of International Conference on Machine Learning* (pp. 412–420).
Yoo, S. Y., Song, J. I., & Jeong, O. R. (2018). Social media contents based sentiment analysis and prediction system. *Expert Systems with Applications, 105*, 102–111.
Zhang, D., Xu, H., Su, Z., & Xu, Y. (2015). Chinese comments sentiment classification based on word2vec and SVMperf. *Expert Systems with Applications, 42*(4), 1857–1863.
Zhang, L., Ghosh, R., Dekhil, M., Hsu, M., & Liu, B. (2011). Combining lexicon-based and learning-based methods for twitter sentiment analysis. In *HP Laboratories (Technical Report)* (pp. 89–91).
Zhang, L., Pei, S., Deng, L., Han, Y., Zhao, J., & Hong, F. (2013). Microblog sentiment analysis based on emoticon networks model. In *Proceedings of the Fifth International Conference on Internet Multimedia Computing and Service* (pp. 134–138). ACM.