Electronic Commerce Research (2020) 20:343–360
https://doi.org/10.1007/s10660-019-09354-7

Sentiment analysis for online reviews using conditional random fields and support vector machines

Huosong Xia[1,3] · Yitai Yang[1] · Xiaoting Pan[1] · Zuopeng Zhang[2] · Wuyue An[1]

Published online: 13 May 2019

Abstract

Sentiment analysis of online reviews is an important way of mining useful information from the Internet. Despite several advantages, the accuracy of sentiment analysis based on a domain dictionary relies on the comprehensiveness and accuracy of the dictionary. Instead of creating a domain dictionary, we propose an approach for online review sentiment classification, which uses a conditional random field algorithm to extract the emotional characteristics from fragments of the review. The characteristic (feature) words are then weighted asymmetrically before a support vector machine classifier is used to obtain the sentiment orientation of the review. In our experiments, the average accuracy reached 90%, showing that using sentiment feature fragments instead of whole reviews and weighting the characteristic words asymmetrically can improve the sentiment classification accuracy.

Keywords Sentiment analysis · Conditional random field · Online review · Support vector machine

✉ Zuopeng Zhang
justin.zhang@unf.edu

Huosong Xia
bxxhs@sina.com

Yitai Yang
994140279@qq.com

Xiaoting Pan
netswordof@163.com

Wuyue An
1664403370@qq.com

1 School of Management, Wuhan Textile University, Wuhan 430073, China

2 Coggin College of Business, University of North Florida, Jacksonville, FL 32224, USA

3 Research Center of Enterprise Decision Support, Key Research Institute of Humanities and Social Sciences in Universities of Hubei Province, Wuhan 430073, China

1 Introduction

With the growth of online shopping, consumers have become increasingly keen to express their views and emotions on the Internet. Mining users' emotions from online reviews is of great significance, which can not only attract potential users by helping them make purchasing decisions, but also enable organizations to get product feedback and even predict the outcomes of major events such as political elections. For instance, Devitt et al. [1] predicted the future trend of market transactions, stock prices, and the volatility of company earnings by judging the polarity of emotions expressed in financial texts. During the Italian parliamentary election in 2011 and the French presidential election in 2012, Ceron et al. [2] conducted sentiment analysis by calculating the Twitter support of political leadership candidates. Research reports show that in many industries such as computer hardware, sports and fitness products, and tourism, more than 50% of consumers will search for the reviews of related products and other relevant information to guide their purchasing decisions [3]. However, due to the large number of online reviews from consumers, it is impractical to manually deal with the opinions in these online reviews. Sentiment analysis is therefore an important approach to opinion mining that efficiently processes such reviews.

Sentiment analysis is a technique for automatically identifying the sentiments expressed in online reviews. In addition, sentiment analysis methods can be applied in many different fields [4–10]. An important prerequisite of sentiment analysis is how to select the words to use as features. Multiple studies have proposed methods for analyzing the sentiments expressed in reviews based on domain dictionaries. Despite the potential advantages, the results of sentiment analysis using a domain dictionary significantly depend on the comprehensiveness and accuracy of the dictionary [11]. The language of online reviews is different from formal language, which means it not only includes domain-specific words but also a lot of Internet catchwords, making it difficult to create a comprehensive and accurate domain dictionary [12].

There has been considerable research on the sentiment analysis of online reviews [13–17]. However, most of the research has focused on English text. English words are naturally separate, whereas the words in Chinese sentences need to be segmented before performing sentiment analysis. In addition, accurate word segmentation is important for obtaining good sentiment analysis results, particularly for Chinese sentences [18]; however, due to the complexity of the expressions in online reviews, it is difficult to obtain good word segmentation results.

Recent research has made some significant achievements in this field. Nevertheless, problems such as data scarcity, category imbalance, domain dependence, and language imbalance [19] still exit, which severely inhibits the classification accuracy and practicability. Hence, there is an urgent need to develop efficient and accurate sentiment analysis methods to excavate opinions and emotions hidden in mixed product reviews, in order to obtain user satisfaction and help enterprises and individuals make management decisions. Our research addresses this gap by proposing a sentiment analysis approach that integrates conditional random fields (CRFs) and support vector machines (SVMs) to analyze online reviews. The approach first extracts emotional feature fragments from the reviews using a CRF algorithm and then identifies the sentiment orientation of each review in terms of their emotional feature fragments, which are classified using an SVM. Applying the method to English and Chinese languages, we obtain high accuracy for both of them. Our approach only requires relatively simple processing to train the model, depends less on accurate word segmentation, and extracts emotional feature fragments instead of building a domain dictionary.

Based on the proposed sentiment analysis approach, this paper further studies the following research questions. First, *does the sentiment analysis method have the same classification effect in Chinese comments and English comments*? Second, *can the reduction of sentences irrelevant to emotional expression (i.e., extracting emotional feature fragments) improve the accuracy of the classifier*? Third, *will the unbalanced distribution of positive and negative emotion words affect the classification accuracy*? Fourth, *is there an algorithm that does not rely on word segmentation and feature combination to achieve high classification accuracy*?

The remainder of this study is structured as follows. Section 2 describes the recent related work on sentiment analysis. The proposed method is then presented in Sect. 3, including both the theory and practical implementation of sentiment analysis. In Sect. 4, we conduct experiments by applying our approach to real-world datasets and discuss the results. Finally, we present our conclusions in Sect. 5.

2 Related work

Sentiment classification refers to identifying the emotional meaning of online reviews and dividing them into positive, negative, or neutral categories [9]. Research on sentiment classification has mainly focused on the application of statistical language rules and machine learning algorithms.

Language rules have been used to build domain-specific dictionaries and analyze syntactic dependencies [17–21]. Domain dictionaries comprise sets of words that appear in particular type of review, and the selected words reflect the features of the domain. The purpose of creating such dictionaries is to select words that affect the analysis or emotional orientation of the reviews. Khan et al. [11] propose a semi-supervised approach based on domain lexicons and incorporate information gain and cosine similarity to improve the performance of sentiment analysis and overcome issues such as data sparsity and domain dependence. Muhammad et al. [22] introduce SmartSA, a classification system for mining social media opinions. The system incorporates a lexicon that is based on the domain knowledge to improve the classification accuracy. Das and Chen [23] manually construct dictionaries of positive and negative emotions and use them to make statistics on the emotional words in the stock evaluation documents. Using a scoring strategy that marks a positive emotion word as 1, a negative emotion word as −1, and a medium-sized word as 0, they obtain the emotional tendency of the document and review the document and the relationship between stock movements. In addition, many researchers use the existing general dictionary WordNet or HowNet to expand the manually collected seed evaluation words to obtain a large number of polar words and polarity [17, 18].With the growing number of online words and spoken words used frequently in critical sentences, existing emotion dictionaries cannot recognize these words with appropriate emotions, resulting in unsatisfactory classification results. In general, the emotion dictionary-based classification method reflects the unstructured features of the text and can achieve the ideal classification effect under the condition of high coverage rate and labeling accuracy of the emotion words in the dictionary [24]. However, this kind of method relies on the background knowledge of domain, time, and language; the difficulty to catch new words and morphing words in time makes it challenging to construct a high-quality emotional dictionary.

Syntactic dependency analysis involves using the sequence of words in the sentence and word POS features to analyze the dependencies between words and uncover different semantic structures, such as the verb–object and verb–complement relations, based on certain rules [25–27]. The aim is to discover which language structures reflect different emotional tendencies. For instance, Cui et al. [28] propose an approach considering both the local and global contexts. Their approach incorporates local contexts using the "bag-of-n-gram" method and global contexts using the "W-S-F" and "POS-S-F" methods. Singh et al. [29] propose a method to translate the slang words in tweets. Their method uses n-grams to find the bindings and CRFs to check the significance of slang words. This can improve sentiment analysis accuracy for microblogs or other texts that include a number of slang words. Yan et al. [23] propose an approach based on rules and classification to analyze sentiment of Chinese microblogs related to finance by utilizing an Improved Label Propagation Algorithm (I-LPA) to construct the sentiment lexicon automatically. They use a rule-based sentiment analysis method to deal with multi-topic tweets, apply three-layer filtering rules to identify emotional agents of specific topics, and then calculate the emotion according to the syntactic dependence between the emotion word and the emotion agent. In fact, as the expression of emotion is closely related to a specific field, different fields have different ways of polar expression, or even completely opposite [30]. Although the rule-based emotion method can obtain classification results in a relatively short time and can add other rules such as prior causes to improve the accuracy of emotion classification, the maintenance of rules is complicated and not easy to expand when there is a large amount of data. In addition, the methods based on dictionaries or syntactic dependency analysis are not very portable.

Statistical and machine learning algorithms are used to classify the sentiments of reviews. Among these, SVMs, random forests, CRFs, and other machine learning algorithms are mainly used [31, 32]. Supervised learning algorithms are trained and tested using labeled samples and may involve the use of SVM, deep learning, or CRFs [31–36]. Sun et al. [37] present a model to solve the feature sparsity problem in Chinese microblogs. The shortness of microblog texts indicates that their feature spaces are sparse; however, the model proposed by Sun et al. [37] handles this problem using deep neural networks (DNNs) and a model called ConCAE, which are shown to be effective. Liao et al. [38] propose a deep learning approach to predict a user's behavior via sentiment analysis. Their approach involves the construction of a neural network to analyze the sentiments expressed in tweets and demonstrated better results than the traditional methods that employ SVM or naïve Bayes classifiers. Although the accuracy of unsupervised learning algorithms has improved, there is still a gap compared with supervised or semi-supervised algorithms [39]. Akhtar et al. [40] present a method for feature selection that is based on single-objective particle swarm optimization (PSO) and sentiment analysis with three different classifiers. Their method selects the main features based on the properties of the classifiers and domains and demonstrated good results for a variety of different domains.

CRF has been widely used in public opinion analysis, emotion classification, and other related research fields. For instance, Laddha and Mukherjee [41] study the problem of aspect-based sentiment analysis and propose a novel LDA-CRF hybrid model which employs discriminative conditional random field component for phrase extraction, a regression component for rating prediction, and a generative component for grouping aspect sentiment expressions (aspect-specific opinion expressions) into coherent topics. Liu and He [30] propose a choice method of emotional basic word based on graph. They apply emotional words to mine short sentence emotional features. Their CRF model has characteristics of feature constraint and is used to classify short sentences. Zhu et al. [42] combine CRF with genetic algorithm to select the best features randomly from the semantic feature set for training by using the idea of survival of the fittest of genetic algorithm. Integrating the characteristics required by conditional random field training, their method introduce novel features that are helpful to improve the recognition accuracy.

Because short comments contain more polarity words of emotion than long ones, they contain less noise and have clearer directions. So, making full use of "short" comments is more conducive to improve the overall classification effect; the classification effect of emotion decreases with the increase of the comment length [42]. Against this backdrop, Zhang et al. [43] propose that the major challenge of text sentiment classification modeling is to capture the intrinsic semantic, emotional dependence information, and the key part of the emotional expression of text. To solve this problem, they develop a Coordinated CNN-LSTM-Attention (CCLA) model. Akhtar et al. [44] propose a system for aspect-based sentiment analysis (ABSA), which is a sequence of processes including aspect term extraction, opinion target expression identification, and sentiment classification. They make use of Support Vector Machines (SVM) for sentiment classification and Conditional Random Fields (CRF) for aspect term and opinion target expression extraction tasks.

Although abundant research exists on sentiments classification, there are not many studies on sentiments analysis in Chinese. Most of the existing Chinese affective analysis methods are derived from English-based technologies. However, they still cannot achieve a good classification accuracy by applying the polarity discrimination method in English to Chinese reviews [16]. Convolutional neural network (CNN) has lately received great attentions because of its good performance in the field of computer vision and speech recognition. It has also been widely used in natural language processing, but such methods for English cannot be transplanted due to phrase segmentation in Chinese [45].

The analysis and summary of the prior research indicates the following problems in the existing research of sentiment analysis.

1. Prior research using emotional dictionary or semantic knowledge base relies on external resources of specific languages, but same words can express different emotions in different domain backgrounds, so the classification portability may be poor.
2. Prior research ignores the unbalanced distribution of positive and negative affective words in comments. Most of the existing work assumes that positive and negative samples are balanced, and the method suitable for equilibrium classification is often not ideal for unbalanced data. The unbalanced distribution of the sample data will make the machine learning method heavily biased towards the categories with multiple samples in the classification, which will affect the classification performance.
3. Most of the prior studies have processed the whole comment without paying attention to the interference of the information irrelevant to emotional polarity in the comment, so the proportion of polar words in the whole comment is relatively small, and the proportion of other characteristic words with noise characteristics is large, making the classification accuracy less than ideal.
4. Most of the existing sentiments analysis research is based on English, and it is difficult to transfer the research of English-based emotional analysis to other languages. In addition, non-English emotional analysis training set and test set are also relatively scarce, which greatly limits the non-English language sentiments analysis research.

To address some of these research challenges, the study proposes a supervised learning algorithm that integrates CRFs and SVMs, selects the key topics using CRFs, and then employs an SVM algorithm to classify the expressed emotions. Specifically, we make the following contributions to the literature with our algorithmic innovation. First, our approach only requires relatively simple processing to train the model and is less dependent on accurate word segmentation or building a domain dictionary. Second, in emotional feature fragments, the proportion of polar words is relatively high; our approach reduces the proportion of other characteristic words with noise characteristics. Third, in the vast majority of the comments, words expressing negative feelings do not appear in sentences with positive emotions, whereas words expressing positive feelings are likely to appear in sentences with negative emotions. To overcome the challenge of such unbalanced data effect, we design a method of asymmetric weights to improve the classification accuracy.

3 Emotion identification

Emotional feature fragments contain more polar words of emotions with less noise and clearer directions. Making full use of large-scale emotional feature fragments is more conducive to improve the overall classification effect. The sequence annotation of the chain CRF not only preserves the characteristic words in the short text, but also the sequential relationship between the words [41], so the emotional feature segment is extracted with CRF. Due to the limitations of dictionary classification method and poor portability, this paper adopts machine learning algorithm SVM for emotion classification. In order to solve the distribution unbalance problem, this paper uses the asymmetric weight assignment method. The next subsection introduces the relevant principles of CRF, the second subsection describes how to extract emotional feature fragments through CRF, the third subsection introduces the classification of emotional features by implanting asymmetric weights into SVM, and the last subsection summarizes the technical roadmap of the whole algorithm.

3.1 Emotional feature fragment extraction using conditional random fields

In a random field, each point in phase space has a random value, which is chosen according to a specific type of distribution. When ordered by time, random variable sequences satisfying the requirement that the distribution at time N + 1 is independent of the random variable values before time N satisfy the Markov property. Markov random fields are random fields with Markovian characteristics. Conditional random fields are based on the discriminant probability model. Let X and Y be random variables, where P(Y|X) is the conditional probability distribution of Y given X, and Y is composed of an undirected graph, G = (V, E). These variables obey the Markov property defined in Eq. (1):

$$\mathrm{P}\left(Y_V|X, Y_W, w \neq v\right) = P(Y_V|X, Y_W, w \sim v) \tag{1}$$

where $w \sim v$ denotes all the w nodes connected to v, $w \neq v$ denotes all the nodes w other than v, and Y_v and Y_w denote two random variables located at nodes v and w, respectively. The conditional probability distribution P(Y|X) for the nodes v constitutes a CRF. Since this paper uses linear-chain CRFs, we will now focus on these CRFs. Let $\mathrm{X} = \left(X_1, X_2, \ldots, X_n\right)$ and $\mathrm{Y} = \left(Y_1, Y_2, \ldots, Y_n\right)$ be two linear chains of random variables, where the sequence of Y of a given random variable is X. The conditional probability distribution P(Y|X) of this sequence of random variables constitutes a CRF, i.e., it satisfies the above Markov property defined in Eq. (2).

$$\mathrm{P}\left(Y_i|\mathrm{X}, Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_n\right) = \mathrm{P}\left(Y_i|\mathrm{X}, Y_{i-1}, Y_{i+1}\right) \tag{2}$$

where $i = 1, 2, \ldots, n$ (when $i = 1$ or n, only one other node is considered). The distribution P(Y|X) constitutes a linear-chain CRF, where X is the input observation sequence and Y is the corresponding output or state sequence.

3.2 Emotional feature fragment extraction

As online reviews are often filled with a large amount of irrelevant information, one or two sentences may not completely express users' emotions. We define statements that focus on expressing users' emotions as emotional feature fragments. In the affective feature segment, the proportion of polar words is relatively high, so the proportion of other characteristic words with noise characteristics is reduced [46].

In a review, only some parts may be involved in expressing sentiments. For example, in the Chinese sentence, "相当棒, 后排座放倒后相当于一张两人床, 每次出游孩子都躺在后面" ("Pretty good. The back seat works like a double bed, and the children can lie behind when we travel around."), the only emotional feature fragment is "相当棒" (pretty good), while the rest of the text does not express any direct emotion. We mark the irrelevant statement as 0 (referring to the annotation rules of the experiment below). Large-scale affective feature segments, when used appropriately, can improve the overall classification effect. Therefore, if we can extract the fragments that express sentiments as features and use these emotional feature fragments to determine the sentiments expressed in reviews, we may be able to improve the classification accuracy. CRF models are better in dealing with the dependence of words on their context; thus, herein, we use CRF++ tools to extract the sentiment feature fragments from online reviews.

3.3 SVM classification based on asymmetric weights

SVMs are used to map sample spaces onto feature spaces with a high (or even infinite) dimension through nonlinear mapping, transforming the problems that are not linearly separable in the original sample space into linearly separable problems in the feature space [33]. The SVM based text emotion analysis method has low generalization error rates and little computational overhead. In addition, it can achieve good emotion analysis results for texts with small training samples and obtain good processing effect for high-dimensional data. However, before the online reviews can be classified, they must first be divided into words and then converted into weighted word vectors, to measure the importance of a feature in a document or the strength of the ability to distinguish [24]. This is often done using the term frequency–inverse document frequency (TF–IDF) method, which fully considers the combination of the two, assigning higher weights to those features with high frequency and good discrimination to the classification [47]. However, this approach does not work well for some types of reviews. In Chinese online reviews, phrases expressing the negative and positive emotions are likely to be considerably similar, such as "喜欢车子的内饰" ("like the interior of the car") versus "不喜欢车子的内饰" ("do not like the interior of the car"). In most reviews, words that express negative emotions, such as "不" ("no") or "差评" ("bad"), do not appear in sentences expressing positive emotions. However, positive words often appear in phrases expressing negative emotions, such as "喜欢" ("like") and "不喜欢" ("dislike") or "漂亮" ("beautiful") and "不漂亮" ("not beautiful"). This study therefore proposes an asymmetric weighting approach that enhances the characteristics of the sentences expressing negative emotions. The approach can be expressed via Eqs. (3)–(6):

$$F^i_{post} = \frac{N^i_{post}}{N_{post}}, \tag{3}$$

$$F^i_{nega} = \frac{N^i_{nega}}{N_{nega}}, \tag{4}$$

$$W^i_{post} = \frac{F^i_{post} - F^i_{nega}}{\sum_{i=1}^{n} F^i_{post}}, \quad (F^i_{post} - F^i_{nega} > 0) \tag{5}$$

$$W^i_{nega} = \frac{F^i_{nega} - F^i_{post}}{\sum_{i=1}^{m} F^i_{nega}} * w, \quad (F^i_{post} - F^i_{nega} > 0) \tag{6}$$

here N^i_{post} and N^i_{nega} denote the frequencies of the i-th word in the positive and negative reviews, respectively. N_{post} and N_{nega} denote the total number of words in the positive and negative reviews, respectively. $\sum_{i=1}^{n} F^i_{post}$ and $\sum_{i=1}^{m} F^i_{nega}$ are the normalization factors for the positive and negative reviews. In addition, n and m denote the number of different words in the positive and negative reviews, respectively. W^i_{post} and W^i_{nega} denote the weights given to word i in the positive and negative reviews, respectively. w is an enhancement factor for negative reviews, which adjusts the weights assigned to words in negative reviews based on different contexts. Here in, we set the value of w to 10. In the training sample, the number of occurrences of positive and negative emotion words is unbalanced. According to the proportion of positive and negative emotion words in the comments, the negative weight is set as 10, which can make the sample balanced and improve the classification accuracy.

3.4 Sentiment analysis of online reviews using CRFs and SVMs

Our method involves two main steps. First, we extract the emotional feature fragments from online reviews using a CRF algorithm. As shown in Fig. 1, the first step is to download the online reviews from the Internet and store them in a database. This is followed by preprocessing of the reviews to remove any duplicate reviews, after which normalization and word segmentation are performed. Then, we tag the words with features (discussed further in Sect. 4) and use CRF++ to train the CRF model. Finally, we use the trained CRF model to remove the parts of reviews that have been preprocessed but not tagged, obtaining the sentiment feature fragments.

Second, we use an SVM classifier to classify the reviews. As shown in Fig. 2, the first step is to complete the acquired sentiment feature fragments by tagging them according to their sentiment tendencies, converting all words into their numeric

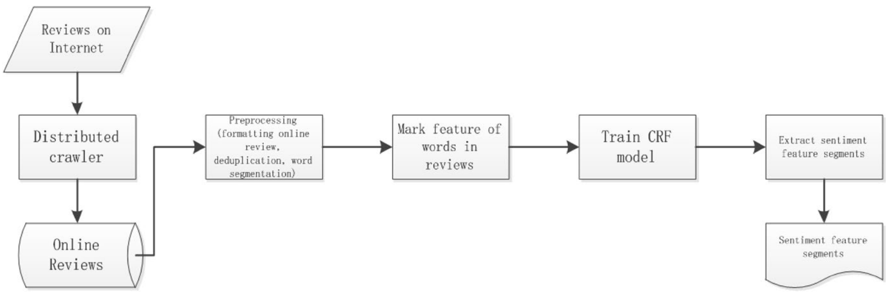

Fig. 1 Extracting sentiment feature fragments using a conditional random field (CRF)

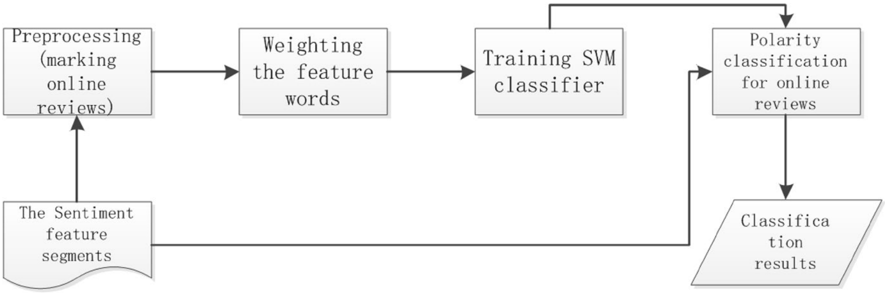

Fig. 2 Review classification using a support vector machine (SVM)

equivalents, and weighting the feature words according to Eqs. (3)–(6). The second step is to train the SVM classifier and then classify the remaining sentiment feature fragments (i.e., the test dataset).

4 Experimental procedure and results

The datasets used in this study were obtained from two sources. We extracted 1488 Chinese online reviews of the Audi A4 sedan from www.autohome.com.cn and 1061 English online reviews of a screen protector for the Samsung Galaxy S7 from www.amazon.co.uk. After preprocessing the Chinese online reviews, the parts of the reviews that referenced "car interiors" were retained. For the English online reviews, the titles were removed.

We conducted three experiments. In Experiment 1, the CRF model was used to extract the emotional feature fragments from the reviews. Then, the positive and negative reviews were classified using an SVM with asymmetric weights. In Experiment 2, we first removed the stop words and then used TF–IDF to assign weights to the words. This was followed by classification using an SVM. In Experiment 3, we extracted the emotional feature fragments from the reviews using the CRF model, used TF–IDF to assign weights to the words, and performed classification using an SVM. All algorithms in this study were implemented in the Python language, and we used the jieba library to segment the reviews. CRF++0.58 was used to train the CRF model, and LibSVM was used for classification.

4.1 Experimental procedure

As CRF models are a type of supervised learning model, tagged samples were required. For each tagged sample, the first word in the sentiment signature was tagged as 11, the last word was tagged as 13, and the words in between the first and last words were tagged as 12, with the other unrelated words tagged as 2. Sentiment feature fragments that only comprised one word were tagged as 10. For punctuation marks, if removing them would not affect the meaning of the sentence (such as most commas and periods), they were tagged as 2. However, if removing them would affect the meaning, they were tagged with a number same as that of the previous word. For example, "Product as described. Very pleased with this product." was tagged as "2,2,2,2,11,12,12,12,12,12,13." Herein, we used the CRF++ open-source tool and the word features only comprised the words themselves. CRF++ requires a template file that lists the feature combinations used for a CRF. Table 1 lists the feature combinations used for these experiments.

The 1488 Chinese reviews were divided into two review sets, A and B, comprising 1000 and 488 reviews, respectively. Likewise, the English reviews were divided into two sets, C and D, comprising 661 and 400 reviews, respectively. In Experiment 1, the manually annotated CRF features were used for the A and C review sets, after which a CRF model trained using the A and C review sets was used to annotate the B and D review sets and extract the sentiment feature fragments. Finally, we employed asymmetric weights to classify the emotional feature fragments of the B and D review sets using an SVM. In Experiment 2, the B and D review sets were used as data sources. In Experiment 3, the B and D review sets obtained from Experiment 1 were used as data sources. The experimental results are shown in Fig. 3.

Figure 3 compares the results of all experiments. The horizontal axis shows the proportion of the dataset that was used for testing, while the vertical axis indicates the classification accuracy, comparing the results of Experiments 2 and 3. In Fig. 3a, TF–IDF was used to assign weights to the words, after which an SVM was used for the classification of the Chinese reviews, resulting in an average accuracy of only 66%. However, using the CRF model to extract the emotional feature fragments increased the average accuracy 78%. In Fig. 3b, the same process was applied to the English reviews, which resulted in an average accuracy of 80% without the CRF model and 91% when the CRF model was used. Comparing the results of Experiments 1 and 3, in Fig. 3a, the CRF model was used to extract the sentiment feature fragments, TF–IDF was used to assign weights to the words, and an SVM was used to classify the Chinese reviews, resulting in an average accuracy of only 78%. However, the use of asymmetric weights increased the average accuracy to 90%. In Fig. 3b, the same process was applied to the English reviews, which resulted in an average accuracy of 91%, irrespective of whether the asymmetric weights were used. The comparison of all three approaches shows that using a CRF model to extract sentiment feature fragments along with asymmetric weighting can considerably improve the classification accuracy compared with the conventional TF–IDF weight assignment method. This is because extracting the emotional feature fragments using a CRF model significantly reduces the number of words that are irrelevant to sentiment classification, thereby improving the quality of

Table 1 Characteristic unigram combinations used for the CRF models

Unigram Feature	Unigram feature
U00: %x[−2,0]	U05: %x[−2,0]/%x[−1,0]/%x[0,0]
U01: %x[−1,0]	U06: %x[−1,0]/%x[0,0]/%x[1,0]
U02: %x[0,0]	U07: %x[0,0]/%x[1,0]/%x[2,0]
U03: %x[1,0]	U08: %x[−1,0]/%x[0,0]
U04: %x[2,0]	U09: %x[0,0]/%x[1,0]

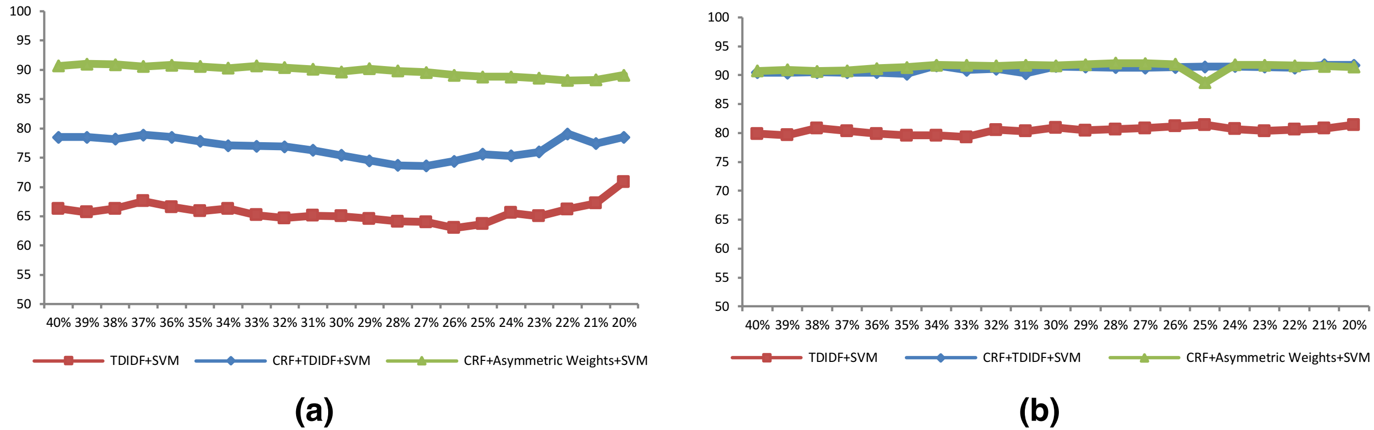

Fig. 3 Comparison of all three methods: **a** Chinese online reviews and **b** English online reviews

the features used for the classifier. Due to the different writing styles in Chinese and English, conventional sentiment analysis methods cannot achieve the same classification performance for these two languages. In Chinese online reviews, phrases that express both negative and positive emotions are likely to be considerably similar, but in English online reviews, these phrases are generally different. Our proposed approach of TF–IDF with asymmetric weights not only improves the classification performance on the Chinese online reviews, but also work equally well for the English reviews.

4.2 Experimental analysis

The experimental results show that the CRF model is used to extract the emotional feature fragments, and then the emotional tendency is extracted from the emotional feature fragments, instead of extracting the emotional tendency directly from the comments, which can improve the classification accuracy. Chinese online reviews can generally be divided into three parts; people generally first express their emotions (like or hate), explain the reasons, and then describe the characteristics of the goods. In the emotional feature segment, when the proportion of polar words is relatively high, the proportion of other feature words with noise characteristics can be reduced, which will improve the effect of the classifier (reference). Compared with the commonly used TF–IDF, the use of asymmetric weights for feature words can improve the accuracy of emotional classification of Chinese online reviews, but not for English online reviews. Since Chinese and English writing styles are different, the TF–IDF gives higher weights to features that are more frequent, resulting in a good discrimination to the classification. Besides, in most of the comments, words expressing negative feelings do not appear in sentences with positive emotions, whereas words expressing positive feelings are likely to appear in sentences with negative emotions. So, using asymmetric weight distribution achieves a better performance than TF–IDF.

We used the jieba package for Chinese word segmentation, after which a CRF model was used to extract the emotional feature fragments. Subsequently, an SVM was used for classification. This method is not sensitive to segmentation errors. This paper uses the word feature containing only the word itself, but not the use of lexical, syntactical, semantic and other characteristics. When classifying the sentiments expressed in online comments with words that are segmented incorrectly, the words in the comments will be segmented in the same way irrespective of the segmentation methods used. In other words, the weights assigned to the words in the online comments will be the same as those assigned to the same words used in the training model, which will not change the accuracy of segmentation. Therefore, this method is not sensitive to the accuracy of word segmentation.

5 Conclusion and discussion

In this research, we propose a method of classifying the sentiments expressed in online reviews. It mitigates the challenges of category imbalance, domain dependence, and language imbalance in sentiment analysis. We first extracted the emotional feature fragments from the reviews using a CRF algorithm and the sentiment expressed in each review was represented by the sentiment orientations of the emotional feature fragments. Next, when the sentiment orientation of the emotional feature fragments was classified using an SVM, the average accuracy was as high as 90%. This method was applied to two languages: both English and Chinese. The contributions of this study are summarized as follows: (a) sentiment classification accuracy can be improved using emotional feature fragments instead of whole reviews and (b) a sentiment analysis method, employing CRFs and SVMs, that can be applied to online reviews with a high level of accuracy (90%) has been proposed.

5.1 Managerial implications

Our research bears the following managerial implications for practitioners.

First, due to the different writing styles between Chinese and English, we find that weighting the words using the proposed asymmetric method was better than using TF–IDF for the Chinese online reviews. We therefore propose weighting the feature words using TF–IDF for English online reviews and using asymmetric weighting for Chinese online reviews. Therefore, we suggest that marketers use the asymmetric weighting approach when they weigh emotions in Chinese reviews. For English reviews, TF–IDF can be used to better understand consumers' emotional tendency and make correct decisions.

Second, we show that emotional feature fragments can improve the accuracy of classification. Chinese online reviews can generally be divided into three parts: expressing emotions (like or dislike), explaining reasons, and describing the characteristics of commodities. Therefore, managers should guide consumers to make standard comments. For example, the comments should be divided into emotional expression area and product feature description area, so that consumers can better express their emotions and describe the characteristics of the products. In this way, visitors can have a clearer understanding of the performance of the products. It is also beneficial for merchants to get more intuitive feedback from consumers, so as to improve the performance of their products and services.

Third, regarding the unbalanced distribution of positive and negative emotion words in Chinese reviews, managers should guide customers to write reviews with distinct emotional words. For instance, they can post a notice at the beginning of their review page to guide customers to use some common works to express their positive or negative sentiments. By using suggested positive and negative emotion words, the accuracy of emotion classification can be improved to help businesses make correct decisions.

Finally, the model proposed in this paper can effectively improve the accuracy of the classification of comment emotion and is expected to further promote the development of e-commerce. Our proposed algorithm is simple and efficient and can help excavate the user emotions in product reviews to enhance user satisfaction, which is valuable to the managerial decision of enterprise organizations and individuals.

5.2 Theoretical implications

Our research also makes the following theoretical contributions to the literature.

First, most of the existing sentiment analysis research ignores the differences between Chinese and English languages and uses the same sentiment analysis method. Based on the characteristics of Chinese comments, this paper proposes an emotion analysis method suitable for Chinese reviews. This method has good portability and uses machine learning algorithm for emotion analysis. It does not involve domain knowledge, so it is suitable for other fields of online comments.

Second, this paper defines sentiment feature fragments to extract sentences that can focus on the expression of emotional tendencies from the whole comment (different from short comments). The classification accuracy can be improved by extracting sentiment feature fragments.

Third, this research proposes methods to reduce the impact of word segmentation on experimental results, making sentiment analysis simple, efficient, and practical.

Finally, according to the characteristics of Chinese reviews, this paper designs an asymmetric weighting method, to enhance the characteristics of words in sentences expressing negative emotions. This asymmetric weight method is superior to TF–IDF.

5.3 Limitations and avenues for further research

Despite the significant contributions made by this research, there are some potential limitations with the methods discussed in the paper. First, our empirical analysis was conducted only for two product categories, automobile and screen protectors. Future research should conduct experiments on more commodity categories to further improve the portability of the model. In addition, when using a CRF model to extract the emotional feature fragments, each feature only considers the word itself, ignoring the other potential factors. While weighting the feature words, the weights may cause the accuracy to vary due to the differences between the current text and the training sample. In our future study, we plan to focus on how to extend the method to large-scale real-time analysis of online reviews using distributed processing while ensuring that the accuracy rate remains high.

Acknowledgements This research has been supported by the National Natural Science Foundation of China (71571139, Outlier Analytics and Model of Outlier Knowledge Management in the context of Big Data; 71871172, Model of Risk knowledge acquisition and Platform governance in FinTech based on deep learning).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Devitt, A., & Ahmad, K. (2007). Sentiment polarity identification in financial news: A cohesion-based approach. In *Proceedings of the 45th annual meeting of the association of computational linguistics* (pp. 984–991).
2. Ceron, A., Curini, L., Iacus, S. M., & Porro, G. (2014). Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. *New Media & Society, 16*(2), 340–358.
3. Double Click. (2005). Search before the purchase-understanding buyer search activity as it builds to online purchase. http://www.doubleclickcom/insight/pdfs/searchpurchase0502.pdf.
4. Swain, A. K., & Cao, R. Q. (2017). Using sentiment analysis to improve supply chain intelligence. *Information Systems Frontiers*, 1–16. https://doi.org/10.1007/s10796-017-9762-2.
5. Al-Radaideh, Q. A., & Al-Qudah, G. Y. (2017). Application of rough set-based feature selection for Arabic sentiment analysis. *Cognitive Computation, 9*(4), 436–445.
6. Tellez, E. S., Miranda-Jiménez, S., Graff, M., Moctezuma, D., Siordia, O. S., & Villaseñor, E. A. (2017). A case study of Spanish text transformations for twitter sentiment analysis. *Expert Systems with Applications, 81,* 457–471.
7. Sahu, S. K., Behera, P., Mohapatra, D. P., & Balabantaray, R. C. (2016). Sentiment analysis for Odia language using supervised classifier: an information retrieval in Indian language initiative. *CSI Transactions on ICT, 4*(2–4), 111–115.
8. Nasim, Z., & Haider, S. (2017). ABSA toolkit: An open source tool for aspect based sentiment analysis. *International Journal on Artificial Intelligence Tools, 26*(06), 1750023.
9. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal, 5*(4), 1093–1113.
10. Mohammad, A. S., Al-Ayyoub, M., Al-Sarhan, H., & Jararweh, Y. (2015). Using aspect-based sentiment analysis to evaluate arabic news affect on readers. In *2015 IEEE/ACM 8th international conference on utility and cloud computing (UCC)* (pp. 436–441). IEEE.
11. Khan, F. H., Qamar, U., & Bashir, S. (2017). A semi-supervised approach to sentiment analysis using revised sentiment strength based on SentiWordNet. *Knowledge and Information Systems, 51*(3), 851–872.
12. Li, R., Lin, Z., Lin, H., Wang, W., & Meng, D. (2018). Summary of text emotional analysis. *Journal of Computer Research and Development, 55,* 30–52.
13. Nguyen, T. H., Shirai, K., & Velcin, J. (2015). Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications, 42*(24), 9603–9611.
14. Hussein, D. M. E. D. M. (2018). A survey on sentiment analysis challenges. *Journal of King Saud University-Engineering Sciences, 30*(4), 330–338.
15. Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-Based Systems, 89,* 14–46.
16. Peng, H., Cambria, E., & Hussain, A. (2017). A review of sentiment analysis research in Chinese language. *Cognitive Computation, 9*(4), 423–435.
17. Serrano-Guerrero, J., Olivas, J. A., Romero, F. P., & Herrera-Viedma, E. (2015). Sentiment analysis: A review and comparative analysis of web services. *Information Sciences, 311,* 18–38.
18. Anand, D., & Naorem, D. (2016). Semi-supervised aspect based sentiment analysis for movies using review filtering. *Procedia Computer Science, 84,* 86–93.
19. Hogenboom, A., Heerschop, B., Frasincar, F., Kaymak, U., & de Jong, F. (2014). Multi-lingual support for lexicon-based sentiment analysis guided by semantics. *Decision Support Systems, 62,* 43–53.
20. Giatsoglou, M., Vozalis, M. G., Diamantaras, K., Vakali, A., Sarigiannidis, G., & Chatzisavvas, K. C. (2017). Sentiment analysis leveraging emotions and word embeddings. *Expert Systems with Applications, 69,* 214–224.
21. Appel, O., Chiclana, F., Carter, J., & Fujita, H. (2016). A hybrid approach to the sentiment analysis problem at the sentence level. *Knowledge-Based Systems, 108,* 110–124.

22. Muhammad, A., Wiratunga, N., & Lothian, R. (2016). Contextual sentiment analysis for social media genres. *Knowledge-Based Systems, 108,* 92–101.
23. Das, S., & Chen, M. (2001). Yahoo! for Amazon: Extracting market sentiment from stock message boards. In *Proceedings of the Asia Pacific finance association annual conference (APFA)* (Vol. 35, p. 43).
24. Yan, D., Hu, B., & Qin, J. (2018). Sentiment analysis for microblog related to finance based on rules and classification. In *2018 IEEE international conference on big data and smart computing (BigComp)* (pp. 119–126). IEEE.
25. Da Silva, N. F., Hruschka, E. R., & Hruschka, E. R., Jr. (2014). Tweet sentiment analysis with classifier ensembles. *Decision Support Systems, 66,* 170–179.
26. Saidani, F. R., & Rassoul, I. (2017). A weighted genetic approach for feature selection in sentiment analysis. *International Journal of Computational Intelligence and Applications, 16*(02), 1750013.
27. 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.
28. Cui, Z., Shi, X., & Chen, Y. (2016). Sentiment analysis via integrating distributed representations of variable-length word sequence. *Neurocomputing, 187,* 126–132.
29. Singh, T., & Kumari, M. (2016). Role of text pre-processing in twitter sentiment analysis. *Procedia Computer Science, 89,* 549–554.
30. Liu, X., & He, Y. (2017). Research on microblog sentiment analysis based on multi-feature. *Computer Engineering, 43,* 160–164.
31. Thakor, P., & Sasi, S. (2015). Ontology-based sentiment analysis process for social media content. *Procedia Computer Science, 53,* 199–207.
32. Kontopoulos, E., Berberidis, C., Dergiades, T., & Bassiliades, N. (2013). Ontology-based sentiment analysis of twitter posts. *Expert Systems with Applications, 40*(10), 4065–4074.
33. Mao, Y., & Lebanon, G. (2007). Isotonic conditional random fields and local sentiment flow. In *Advances in neural information processing systems* (pp. 961–968).
34. Nair, D. S., Jayan, J. P., Rajeev, R. R., & Sherly, E. (2015). Sentiment analysis of Malayalam film review using machine learning techniques. In *2015 international conference on advances in computing, communications and informatics (ICACCI)* (pp. 2381–2384). IEEE.
35. Shariaty, S., & Moghaddam, S. (2011). Fine-grained opinion mining using conditional random fields. In *2011 IEEE 11th international conference on data mining workshops* (pp. 109–114). IEEE.
36. Jin, Y., Zhang, H., & Du, D. (2016). Improving deep belief networks via delta rule for sentiment classification. In *2016 IEEE 28th international conference on tools with artificial intelligence (ICTAI)* (pp. 410–414). IEEE.
37. Sun, X., Li, C., & Ren, F. (2016). Sentiment analysis for Chinese microblog based on deep neural networks with convolutional extension features. *Neurocomputing, 210,* 227–236.
38. Liao, S., Wang, J., Yu, R., Sato, K., & Cheng, Z. (2017). CNN for situations understanding based on sentiment analysis of twitter data. *Procedia computer science, 111,* 376–381.
39. Fernández-Gavilanes, M., Álvarez-López, T., Juncal-Martínez, J., Costa-Montenegro, E., & González-Castaño, F. J. (2016). Unsupervised method for sentiment analysis in online texts. *Expert Systems with Applications, 58,* 57–75.
40. Akhtar, M. S., Gupta, D., Ekbal, A., & Bhattacharyya, P. (2017). Feature selection and ensemble construction: A two-step method for aspect based sentiment analysis. *Knowledge-Based Systems, 125,* 116–135.
41. Laddha, A., & Mukherjee, A. (2018). Aspect opinion expression and rating prediction via LDA–CRF hybrid. *Natural Language Engineering, 24*(4), 611–639.
42. Zhu, J., Wang, H., & Mao, J. (2010). Sentiment classification using genetic algorithm and conditional random fields. In *2010 2nd IEEE international conference on information management and engineering* (pp. 193–196). IEEE.
43. Zhang, Y., Zheng, J., Jiang, Y., Huang, G., & Chen, R. (2019). A text sentiment classification modeling method based on coordinated CNN-LSTM-attention model. *Chinese Journal of Electronics, 28*(1), 120–126.
44. Akhtar, M. S., Kohail, S., Kumar, A., Ekbal, A., & Biemann, C. (2017). Feature selection using multi-objective optimization for aspect based sentiment analysis. In *International conference on applications of natural language to information systems* (pp. 15–27). Cham: Springer.
45. Xiao, Z., Li, X., Wang, L., Yang, Q., Du, J., & Sangaiah, A. K. (2018). Using convolution control block for Chinese sentiment analysis. *Journal of Parallel and Distributed Computing, 116,* 18–26.

46. Zhang, L., Qian, G. Q., Fan, W. G., Hua, K., & Zhang, L. (2014). Sentiment analysis based on light reviews. *Ruan Jian Xue Bao/Journal of Software, 25*(12), 2790–2807.
47. Ye, J., Jing, X., & Li, J. (2017). Sentiment analysis using modified LDA. In *International conference on signal and information processing, networking and computers* (pp. 205-212). Singapore: Springer.