JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

A Sentiment Classification Approach of Sentences Clustering in Webcast Barrages

Jun Li*, Guimin Huang**, and Ya Zhou**

Abstract

Conducting sentiment analysis and opinion mining are challenging tasks in natural language processing. Many of the sentiment analysis and opinion mining applications focus on product reviews, social media reviews, forums and microblogs whose reviews are topic-similar and opinion-rich. In this paper, we try to analyze the sentiments of sentences from online webcast reviews that scroll across the screen, which we call live barrages. Contrary to social media comments or product reviews, the topics in live barrages are more fragmented, and there are plenty of invalid comments that we must remove in the preprocessing phase. To extract evaluative sentiment sentences, we proposed a novel approach that clusters the barrages from the same commenter to solve the problem of scattering the information for each barrage. The method developed in this paper contains two subtasks: in the data preprocessing phase, we cluster the sentences from the same commenter and remove unavailable sentences; and we use a semi-supervised machine learning approach, the naïve Bayes algorithm, to analyze the sentiment of the barrage. According to our experimental results, this method shows that it performs well in analyzing the sentiment of online webcast barrages.

Keywords

Semi-supervised, Sentences Clustering, Sentiment Analysis, Webcast Barrages

1. Introduction

As a new media, webcasts are increasingly favored by an increasing number of people, and timeliness and interactivity are the most obvious characteristics of webcasting. The viewers can share their sentiments and opinions on the screen while watching the program. This interaction is not only between the anchor and the audience but also among viewers, thus the live broadcast will produce a large number of short comments that scroll across the screen, and we call them barrages. Both the managers of the platform and the webcasters want to understand the audience's feedback on the webcast programs. Thus, the sentiment analysis and opinion mining of these barrages is a special and meaningful task, and such barrages usually focus on current issues and events. Our task was also motivated by an online webcast company where the viewers actively participated in the webcast and generated a large amount of evaluative opinions.

Corresponding Author: Guimin Huang (send_huang@126.com)
* School of Information and Communication, Guilin University of Electronic Technology, Guilin, China (lijun5221@126.com)
** Guangxi Key Lab of Trusted Software, School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin, China (send_huang@126.com, 570229495@qq.com)

Sentiment analysis and opinion mining involves the computational research of people's subjective information, such as the sentiments and opinions in text [1-3]. Most sentiment analysis and opinion mining tasks focus on social media reviews, product reviews, forums and microblogs [4,5]. Different from these product reviews and social media reviews that contain a similar or relevant topic in each comment, each text of a webcast barrage is shorter and sometimes contains just one word or number. Although it is difficult to extract the topic or emotional words from a single short barrage text, we found that the comments from the same commenter are relevant within a certain time, so we proposed clustering the sentences based on time. We will discuss this importance of the time parameter later because barrage topics change quickly.

In this paper, we aim to identify the sentiments in an online webcast barrage. At present, there are few relevant studies to our knowledge [6]. Although the webcast barrages and online comments may look too similar for subjective classification, the texts are rich in emotional opinions and they are entirely different as the emotional expressions of webcast barrages are quite random. At the same time, barrages are vulnerable to the interference of other comments, which we call "spam" barrage. "Spam" barrages are usually initiated by a bellwether and followed by other viewers. The sentiment orientation for these "spam" barrages often comes from the bellwether, and the polarity of the sentiment is explicit but influential because of the number of viewers that participate. "Spam" barrages do not always indicate a positive sentiment, in fact these barrages usually contain much abusive, defamatory or regionally discriminative language. Although some of these "spam" barrages are emotional opinions, we must filter out these barrages.

Like most sentiment analysis and opinion mining tasks, the sentiment analysis for the barrages is also a two-classification problem, positive and negative. The neutral class is not considered in this paper to achieve a strong contrasting effect of the webcast barrages. The datasets are an important part of sentiment analysis and opinion mining, and a large number of barrages are generated every day. In our task, we collected barrage reviews from the same studio over one month. We also chose the barrage reviews of 3 typical webcasts for datasets: the barrage reviews of a concert, a cross talk performance by GuoDegang and a Xiaomi product launch. The concert will have different singing stars in attendance and the audiences will have different preferences for these stars, so the barrages will include rich emotional perspectives. We found that the barrages of the cross-talk performance by GuoDegang contained two distinct emotional tendencies. However, in the barrages of the Xiaomi product launch, more sentiment comments came from different attributes of the product. Therefore, we chose these 3 datasets after preprocessing, and then we added these datasets to the HowNet sentiment corpus for training.

Clearly, the problem of sentiment analysis is a classification problem; supervised learning is the classic method, but this approach usually requires a large amount of manually labelled training data. In our task, we discarded this time-consuming supervised learning approach and we choose a small number of seed words and Internet pop words to incorporate in the training in a similarly weakly semi-supervised or unsupervised manner. We experimented with the NB (naïve Bayes), SVM (support vector machine) and CRF (conditional random field) methods in terms of machine learning, and the results demonstrated the feasibility and effectiveness of this method.

The remainder of this paper is organized as follows: Section 2 introduces related work, Section 3 describes our approach, includes the datasets collection and processing, Section 4 introduces experimental results, and conclusions are presented in Section 5.

2. Related Works

In recent years, sentiment analysis and opinion mining have become the hot topic in natural language processing (NLP), and many institutes and researchers have participated in this domain. Many tasks focus on product reviews, social media reviews, forums and microblogs. Some researchers used machine learning algorithms to identify the sentiment tendencies in comments that laid the groundwork for this domain. Pang and Lee [7,8] carried out a large number of annotations on movie reviews and trained and tested the data through machine learning algorithms. In the follow-up work, they also designed a min-cut graph algorithm that performed well in emotion classification. Identifying the emotional tendencies of these online reviews is an interesting and meaningful effort. Hu and Liu [9] developed a method that extracts a feature-based summary that can provide guidance to ordinary consumers by extracting a large number of online product reviews. With the popularity of social media, research on the sentiment analysis of social media has been increasingly intensified. Since the text of twitter comments is short and the opinion is clearly stated, it has become the preferred data of many researchers. Jiang et al. [10] proposed an approach that was target-dependent and context-aware, and incorporated tweets into classification to avoid short and ambiguous tweet contexts. Documents in webcast barrages are different from traditional media because they are noisy and short, but there is a correlation among the barrages. Barbosa and Feng [11], instead of using traditional emotional features as inputs, use noise annotation to build models, which is a more abstract representation of information. Hu et al. [12] studied the impact of social relationships on the sentiment analysis tasks of short texts. Gong et al. [13] proposed the notion of shared model adaptation, where the diversity of human opinions is shaped by changing social norms and there is a tendency that opinions can be easily influenced by others. Like social media, the barrages are contagious. Usually, a large number of barrages are merely follow-up responses, and there is no change in the emotional opinion. For these same or similar barrages, we need to extract only one of them to perform a sentiment analysis and obtain better results. Dong et al. [14] proposed a new microblog sentiment analysis method that mines associated microblog emotions based on a popular microblog through user-building combined with spectral clustering to analyze the microblog content. Zhai et al. [6] proposed identifying evaluative sentences in online discussions. Hassan et al. [15] established a network of commentator interactions to identify whether the participants were positive or negative. Salehan and Kim [16] investigated the value of online reviews by means of a sentiment analysis, proving that comments with higher positive emotions in the title are more popular with other consumers. Machine learning is widely used in the task of sentiment analysis. Socher et al. [17] proposed an approach based on recursive auto encoders for predicting the sentiment orientation. Ramesh et al. [18] developed a weakly supervised joint probabilistic model; the model had improved generalization ability through the form of seed words and provided a weighting rule that captured the dependencies between attributes and sentiments. Fernandez-Gavilanes et al. [19] proposed an unsupervised dependency parsing-based method that reduced resource consumption in online texts. Li et al. [20] compared several popular classification algorithms through hierarchical filtering. Wu et al. [21] designed an approach based on the support vector machine and generalized autoregressive conditional heteroskedasticity modeling, and the method was used to perform a context-sensitive emotional analysis of the online market postings and achieved good results. Liu et al. [22] proposed a new sentiment classification algorithm to sort sentiment intensified product reviews.

The method of attribute extraction can achieve good results in product reviews and microblog sentiment analysis tasks [23-25]. Due to the real-time and randomness of live broadcasts, the sentiment

judgment results of the current methods will have hysteresis [26]. In this paper, we propose a method of clustering sentences for sentiment analysis that is similar to the sentiment analysis approach based on attribute extraction. Wu and Easter [27] proposed a probabilistic model by combining the advantages of collaborative filtering and aspect-based opinion mining. Wang et al. [28] proposed an attention-based long short-term memory (LSTM) network to identify the sentiment orientation. Amplayo and Hwang [29] proposed a model called the micro aspect sentiment model (MicroASM), which performs better on aspect-level and document-level tasks. Feng et al. [30] designed a new algorithm to identify implicit aspects by integrating a deep convolution neural network (CNN). Peng and Liu [31] designed a clustering-based multilabel classification model.

3. Proposed Approach

Fig. 1 gives the architectural overview of the proposed approach. Our method contains five steps to identify the sentiment polarity from the barrage, which is usually expressed as positive and negative:

- Barrage corpus collection: The size of the barrage depends on the number of viewers in a certain time. It is discussed in Section 3.1.
- Data preprocessing: This subtask removes the stop words, segments the words, and performs POS-tagging and sentiment word matching. The work ahead of the step is fairly common, so it will not be a focus of discussion in this paper. Sentiment word matching is important in this subtask because we can preliminarily judge the sentiment tendency of each sentence through the matching of the sentiment words, even though it is imprecise. We will discuss it in Section 3.2.
- Spam removed: The online barrages are full of violent, defamatory and discriminatory comments, and although these are not the focus of this paper, these comments have a negative impact on the overall emotional analysis task and must be removed, as discussed in Section 3.3.
- Sentences clustering: This step is discussed in Section 3.4.
- Sentiment identification: We used the naïve Bayes machine learning algorithm for this task, and it is discussed in Section 3.5.
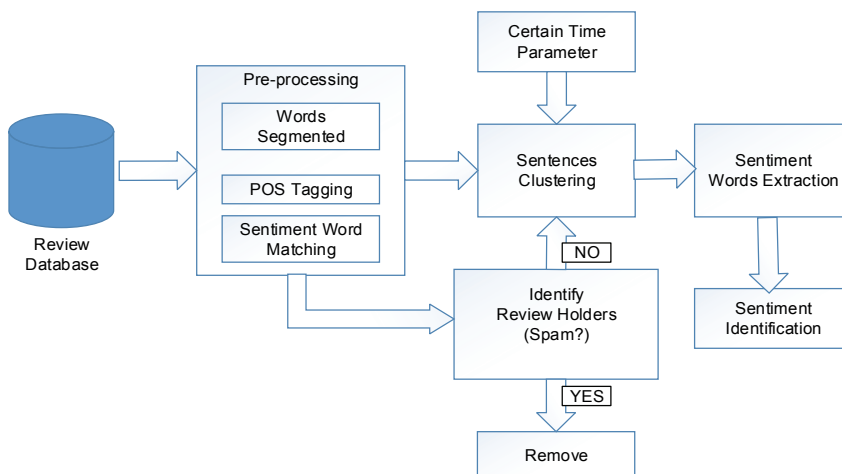

Fig. 1. Overview of the proposed approach.

3.1 Barrage Corpus Collection

The size of the barrage depends on the number of viewers, and each barrage cannot exceed 50 characters, but we found that the webcast barrages are usually within 10 characters because the barrages appear very quickly, and a long barrage would take too much time to read so that viewers would not want to see it. We determine the distribution of the number of barrage reviews and viewers in one month via the Douyu live platform (https://www.douyu.com). Fig. 2(a) provides the barrage distribution, and Fig. 2(b) presents the viewer distribution for one month. In such data sets, there are usually many other types of barrages besides the comments on the topic, as the participants can interact with each other. In many cases, the discussions may raise emotional criticism and deviate from the topic. For example, in our dataset, more than half of the barrages are emotional, abusive, or numeric. Most of these barrages need to be removed, but two types of these numbers often appear in the barrage. The numbers "666" and "233" are fairly popular on the web and they come in many forms, such as "666" can use any number of 6s for the same kind of emotional meaning, and there are often at least two 3s in "233". The "666" and "233" often express positive emotions, such as convincing, amazing and lol.

These two types of numbers appear frequently, and they express the positive emotional orientation of the reviewers, so we characterize them as part of the sentiment analysis task.

Fig. 2(a) and 2(b) reflect the distribution of the barrages and viewers in the same live broadcast within one month. We found that they basically showed a positive correlation, which means that the number of viewers determines the number of comments, which is also easy to understand. The abscissa represents the date, and we use entire month of September for the collection dates. Some dates have no barrages because the days did not have a live show. After the dataset is collected, the next step is to preprocess the sparse data.
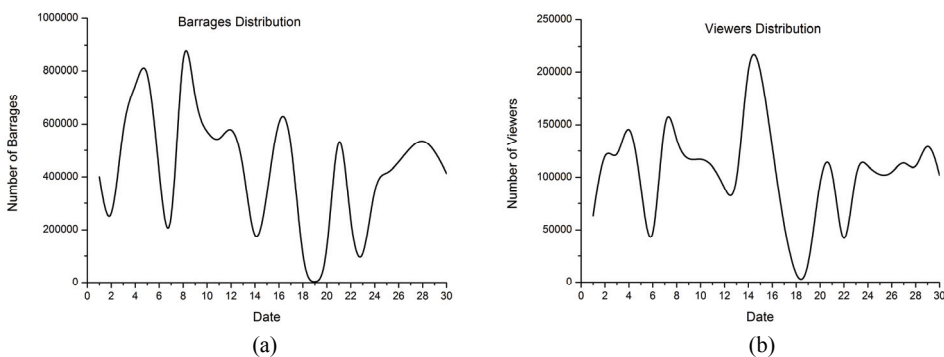

Fig. 2. (a) Barrage and (b) viewer distributions for one month.

3.2 Data Preprocessing

Data preprocessing includes the removal of stop words and word segmentation [32]—Jieba is Chinese text segmentation (http://www.oss.io/p/fxsjy/jieba), which will not be discussed in this article. We used the Stanford Parser software (https://nlp.stanford.edu/software/lex-parser.html) to preprocess these texts. Webcast barrages usually include large high-frequency sentiment words [33]. We will choose some of these words as our seed words and match each barrage. Then, the first simple classification is complete, but it is not accurate and requires further work. Table 1 presents some of the sentiment words. There are

two types of numbers in the seed word, "666" and "233". As we have already discussed above, there are usually a number of digital symbol comments present in the webcast barrages, which have particular meanings over a specific period. For example, if the anchor asked "How am I doing today? like the performance: please enter the number '1'; poor, please enter the number '2'" in the live interaction, there would be a large number of '1' and '2' barrages, and '1' expresses a positive sentiment and '2' expresses a negative sentiment. However, in other cases, '1' and '2' may express exactly the opposite sentiment, so the meaning of such numerical symbols is so random that such figures were removed in our preprocessing to improve the accuracy of the sentiment analysis, and just the "666" and "233" barrages that expressed a fixed sentiment were retained. The work also selected some current Internet buzzwords, such as "老铁/old friend", "么么哒/like", "凉/disappointed", "傻逼/sucker", and so on.

Table 1. Positive and negative sentiment seed words

Positive	老铁/old friend	暴富/parvenu	么么哒/like	666/amazing	233/laugh
	漂亮/beautiful	忠诚/devotion	干净/clean	加油/come on	实惠/substantial
	友善/friendly	优秀/excellent	欢喜/joyful	权威/authoritative	舒服/comfortable
	健康/healthy	天使/angel	安静/calm	专心/concentrate	准确/accurate
	完美/perfect	容易/easy	完整/intact	昌盛/prosperity	雄心/ambition
	和平/peace	活力/vigor	坚定/firm	亮点/light spot	赚/earn
	幸福/happiness	新/new	勤奋/industrious	开通/accomplish	稳了/confirmed
	诚信/faithful	文明/civilized	出名/famous	真实/real	成熟/mature
	聪明/clever	积极/active	精英/elite	捷报/good news	保险/assure
Negative	凉/disappointed	傻逼/sucker	垃圾/rubbish	智障/mentally retarded	滚/scat
	毒/poison	暴力/violence	非法/illegality	心机/craftiness	虚假/sham
	嘈杂/noisy	漏洞/leak	事故/accident	亏/deficit	变态/abnormal
	浪费/waste	花样/trick	失败/failure	冲突/conflict	陈旧/obsolete
	妒忌/jealous	谣言/rumor	病人/patient	恶势力/vicious power	残/incomplete
	色情/erotic	淫秽/bawdry	错误/error	失误/mistake	流氓/immoral
	疯狂/insane	缺少/lack	敌对/hostility	脆弱/weak	蠢/foolish
	黑客/hacker	陷阱/trap	脏/dirty	不合格/unqualified	不安/uneasy
	诈骗/cheat	自负/conceited	丑陋/ugly	恶意/malevolence	烦恼/upset

3.3 Spam Removal

The online barrages usually are full of violent and defamatory comments and regional discrimination, and although these are not the focus of this paper, these comments have a negative impact on the overall emotional analysis task and must be filtered out. Pagani et al. [34] presented a comprehensive study of the advantages and disadvantages of using two not very popular spam filtering techniques: no-listing and gray-listing. In our task, we adopted an approach similar to gray-listing. Some commentators regularly post personal attacks, regional discrimination and other spam information in webcast barrages. We put these commenters' network ID on the gray list to filter out all their barrages. Although this approach may remove some useful emotional information, it can be very effective in many cases. After this step, our data set contains only 40% of the original data, and this new data set contains a wealth of emotion and opinion information.

3.4 Sentence Clustering

Liu [1] defined a five-tuple sentiment analysis and opinion mining model; the model contains five parts shaped as (e, a, s, h, t), wherein e represents the target entity of opinion evaluation, a is an entity attribute or aspect of a viewpoint evaluation in entity e, s represents the sentiment and opinion orientation of attribute a in entity e, h represents the reviews holder, and t is the comment publication time. Sentiment analysis and opinion mining in social media comments or product reviews is not too much research on reviews holders as each comment contains relatively rich information and represents the basic view orientation of current commentator. In contrast, each sentence of a webcast barrages is very short and it is difficult to extract more information. Che et al. [35] presented an approach for sentiment sentence compression. However, webcast barrages also have special characteristics because each comment contains a small amount of information; usually, the same commentator can publish a number of comments in a certain period of time to express his opinions and emotions. Based on this, in the paper, we first cluster the comments from the same reviewer. The comments from the same reviewer are relevant and similar within a certain period of time. In our task, we must remove the duplicate opinions and emotions and draw the current opinion of the sentiment. For example, a user with the audience ID "Shark88" posted six barrages in one minute while watching a live concert:

> Barrage 1: Shark88(20:08:12):刘德华(Andy Lau, a famous singer)要(is going to)出场(be present),
> 太(so)开心(happy)了.
> Barrage 2: Shark88(20:08:25):华仔(Andy Lau),么么哒(love).
> Barrage 3: Shark88(20:08:43):听到(hear)这(this)首歌(song)想(want to)哭(cry).
> Barrage 4: Shark88(20:08:56):666666(amazing).
> Barrage 5: Shark88(20:12:34):他的(his)歌(song)是(is)垃圾(rubbish).
> Barrage 6: Shark88(20:13:03):听(Do)不(not)懂(understand)鹿晗(LUHAN, another singer)的歌(song).

The commenter's ID and publication time are included at the beginning of each barrage. We can obtain all the current barrages of the reviewer via the same ID and noticed that the topics discussed in these reviews vary greatly from time to time, and the emotions and opinions tend to be different for different topics. Thus, a sentiment analysis of all reviews of the same ID would obviously not be very accurate. Therefore, we added a time parameter for the sentence clustering. The above example contains five barrages, and our task selection interval of one minute selects the first four comments to analyze. The preprocessing procedure for barrages includes filtering out stop words, Chinese word segmentation, POS-tagging, etc. In this paper, we choose ICTCLAS for Chinese word segmentation, POS-tagging, entity named recognition, etc. The latest version of ICTCLAS shows high accuracy.

In barrage 1, "刘德华(Andy Lau)" as a proper name, it is easy to identify, the word "开心(happy)" as a positive sentiment word can also be recognized. In barrage 2, "么么哒(love)" is an Internet buzzword to express love, which is a strong tendency of positive emotions. Both barrages expressed their like of the singer, Andy Lau. However, in barrage 3, the word "哭(cry)" appears in a comment that expresses negative emotions in the vast majority of cases. For example, in the sentence "someone sadly cried out", the word "cry" is an expression of negative emotions, however, in some cases, "cry" can also express a positive emotion, such as "someone is so moving to cry out". Although we know that the word "cry" expressed a positive emotion in this commentary, it has a high probability of being a negative emotion in

machine learning. To solve the problem of uncertainty about the emotional polarity of a barrage comment, we try to determine the emotional orientation of the adjacent contextual comments. Therefore, we analyze barrage 4 first: "666666" is also an Internet buzzword which means "amazing" to express a positive sentiment, and we have already explained this above. Then, we analyzed the four barrages that were clustered for one minute. The first, second and fourth barrages expressed strong a positive sentiment toward singer Andy Lau and his song, but is not clear whether the emotional polarity of the third comment is positive or negative. We hypothesize that the emotional polarity of any one barrage commentary is positively correlated with the emotional polarity of its adjacent contextual commentary in a clustered commentary set. In other words, "哭(cry)" expressed a positive sentiment in barrage 3. This method of clustering barrages by defining the time interval parameters worked well, and the classification accuracy in our experiments reached 90%. In the above example, some new information is obtained that the reviewer with the "Shark88" network ID is probably an Andy Lau fan and he could be introduced him to some related products and programs that he is interested in. When we set the time parameter to one minute, the classification result has a perfect performance, but when the time parameter is set to 5 minutes, the clustered barrage comments in the above example will include barrages 5 and 6, a total of 6 sentences. In barrage 5, the word "垃圾(rubbish)" can be identified as a strong negative sentiment word that modified the attribute "song", which is the exact opposite of the positive emotion expressed in barrage 3. Since the opposite sentiment polarity appears for the same commenter within five minutes, we need to confirm whether the opposite emotional commentary is for the same attribute. Although the words "song" appears in both comments, they express the opposite emotional tendency, and it is still not certain whether the same attribute is modified through all five barrages. We then try to analyze the sentiment expressed in barrage 6, which usually would have some relevance to the commentary in barrage 5. Using the same method to extract the sentiment words and named entities from barrage 6, the word "鹿晗(LU HAN)" as a proper name (who is also a singer), is easy to recognize and the aspect word "歌(song)" can be extracted via sophisticated POS-tagging and semantic analysis techniques. Thus, in barrage 6, the entity(e) word is "鹿晗(LU HAN)", the aspect(a) word is "歌(song)", and the sentiment word "理解(understand)" expressed a positive sentiment polarity, but in barrage 6, it contains the sentiment shifter "不(not)" so that this comment expressed negative emotions. Then, we think that the entity object in barrage 5 should be "鹿晗(LU HAN)" instead of "刘德华(Andy Lau)", and the set aggregated with 6 barrages should be divided into two parts; the first 4 barrages are aggregated as one set, and the rest of the barrages are aggregated in the other. One cluster expresses an emotional bias toward the entity "刘德华(Andy Lau)" and the other toward "鹿晗(LU HAN)". This result also proves that when the time parameter range is too large, the same commenter may publish different opinions on different entities or even totally different sentiment orientations for different attributes.

Specifically, we use the K-means clustering algorithm. We select different users as the initial centers to calculate the similarity distance of the bullet-curtain text produced by the same user. Here, we use the Euclidean distance to update the values of each clustering center one by one iteratively until the best clustering results are obtained.

3.5 Sentiment Classification

In this section, we will discuss the sentiment identification of the text barrage set. Our task used the naïve Bayes classifier, and the Bayes formula is as follows:

$$P(s \mid C) = \frac{P(C \mid s) \cdot P(s)}{P(C)} \tag{1}$$

where C is a barrage comment and s indicates the sentiment polarity. Because each sentence contains one of the sentiment polarities of positive, negative or neutral, we simplify the equation below:

$$P(s \mid C) = \frac{P(C \mid s)}{P(C)} \tag{2}$$

However, the probability distribution of the barrage comment is certain; that is, $P(C)$ is fixed. Therefore, the following relationship can be obtained by Eq. (2):

$$P(s \mid C) \sim P(C \mid s) \tag{3}$$

We used three different features, the POS-tagging distribution, $P(T \mid s)$; the appearance of an N-gram, $P(N \mid s)$; and sentiment words, $P(W \mid s)$, that were extracted from previously clustered sentences. In our task, we assume the conditional independence of N-gram features and part-of-speech information. We simplify the equation:

$$P(s \mid C) \sim P(T \mid s) \cdot P(N \mid s) \cdot P(W \mid s) \tag{4}$$

where T is the set for part-of-speech tagging, N represents the N-gram set and W represents the sentiment seed words. We also make an assumption that the POS-tags and n-grams are conditionally independent, so we can obtain the equation below:

$$P(s \mid C) \sim \prod_{t \in T} P(t \mid s) \prod_{n \in N} P(n \mid s) \cdot P(W \mid s) \tag{5}$$

Finally, we used the log-likelihood value of each sentiment:

$$L(s \mid C) = \sum_{t \in T} \log(P(t \mid s)) + \sum_{n \in N} \log(P(n \mid s)) + \log(P(W \mid s)) \tag{6}$$

We introduced some high-frequency network emotional words in the training datasets, but there are few such words; thus, our method is called semi-supervised learning or weak supervised learning. Webcast barrages contain large numbers of network buzzwords and the barrage texts are usually short but contain rich emotional tendencies. After joining the emotional words in training, the accuracy of the classification results improved.

4. Experiments and Results

4.1 Data Availability

We chose the live broadcast platform Douyu as the source for data collection, and we used the Chinese information gathering software, Bazhuayu, to collect the barrage information (the Bazhuayu software can be downloaded from http://www.bazhuayu.com/about). We also chose the barrage reviews of 3 typical webcasts as the datasets: the barrage reviews of a concert, a cross talk performance by *GuoDegang* and

a *Xiaomi* product launch. We put the original data on the Github website (https://github.com/mrlijun2017/barragesData). In addition, the sentiment seed words came from HowNet (the largest sentiment word lexicon in Chinese, http://www.keenage.com/html/c_bulletin_2007.htm).

4.2 Experiments

In this section, some of the experimental details of our current work will be discussed. We collected three datasets from webcast comment barrages: a concert, a cross talk performance by GuoDegang, and a Xiaomi product launch. After preprocessing and sentence clustering, the three barrage datasets included 1628, 2074 and 1988 postings. We also manually annotated all the sentences as positive or negative, and the statistics of the three barrage datasets are presented in Table 2.

We discussed the naïve Bayes machine learning algorithm above and compared it to the SVM (support vector machine) and CRF (conditional random field) algorithms. We calculated the precision, recall and F1-score values of the classifier on each dataset to measure the performance:

$$F = \frac{(1+\beta^2)\,precision * recall}{\beta^2(precision + recall)} \tag{7}$$

In this paper, we set the parameter $\beta = 1$ and used the equation to calculate the F1-score:

$$F_1 = \frac{2\,precision * recall}{precision + recall} \tag{8}$$

Table 2. Statistics for the three barrage datasets

Dataset	Positive	Negative	Total
Concert (CC)	946	682	1628
Performance by GuoDegang (GD)	1520	554	2074
Xiaomi Product Launch (XM)	1224	764	1988

Table 3. Comparison results for the three datasets (*t* is the time parameter)

t=5 (min)	Precision (%)				Recall (%)				F1-score (%)			
	CC	GD	XM	Avg	CC	GD	XM	Avg	CC	GD	XM	Avg
NB	76	75	74	75	80	84	85	82	78	79	79	79
SVM	68	72	74	71	74	81	76	77	71	76	75	74
CRF	71	65	70	69	70	82	79	77	70	73	74	72

We set the time parameter to *t*=5 minutes to identify the sentiment tendency of the barrages within 5 minutes. The comparison results of the three datasets are shown in Table 3, where Avg represents the average result of the three datasets.

This paper focuses on the expression of sentiment in barrage reviews; therefore, there is a difference in the emotional tendency of different kinds of review datasets. In dataset CC, many of the barrage reviews are about different singers, and we can easily identify the named entities, so the precision value is very good. In the GD dataset, the attribute of the sentiment analysis is the cross-talk performance of GuoDegang, which is relatively fixed, and the F1-score of this emotional classification is the best. In

addition, the XM dataset, which is a typical product review dataset, has the best recall value of the three datasets. We believe that barrage sentences with the same emotional orientation from different attributes will be combined in sentence clustering, and the recall value will increase. However, it does not affect the classification of the previous two datasets because these two datasets do not conduct an emotional analysis according to the attributes. This paper aims to identify the sentiment of the barrages, and the sentence clustering method worked well in our task. We compared the precision, recall and F1-score values from Fig. 3; Fig. 3(a) compares the precision value, and Fig. 3(b) and 3(c) compare the recall and F1-score values, respectively.
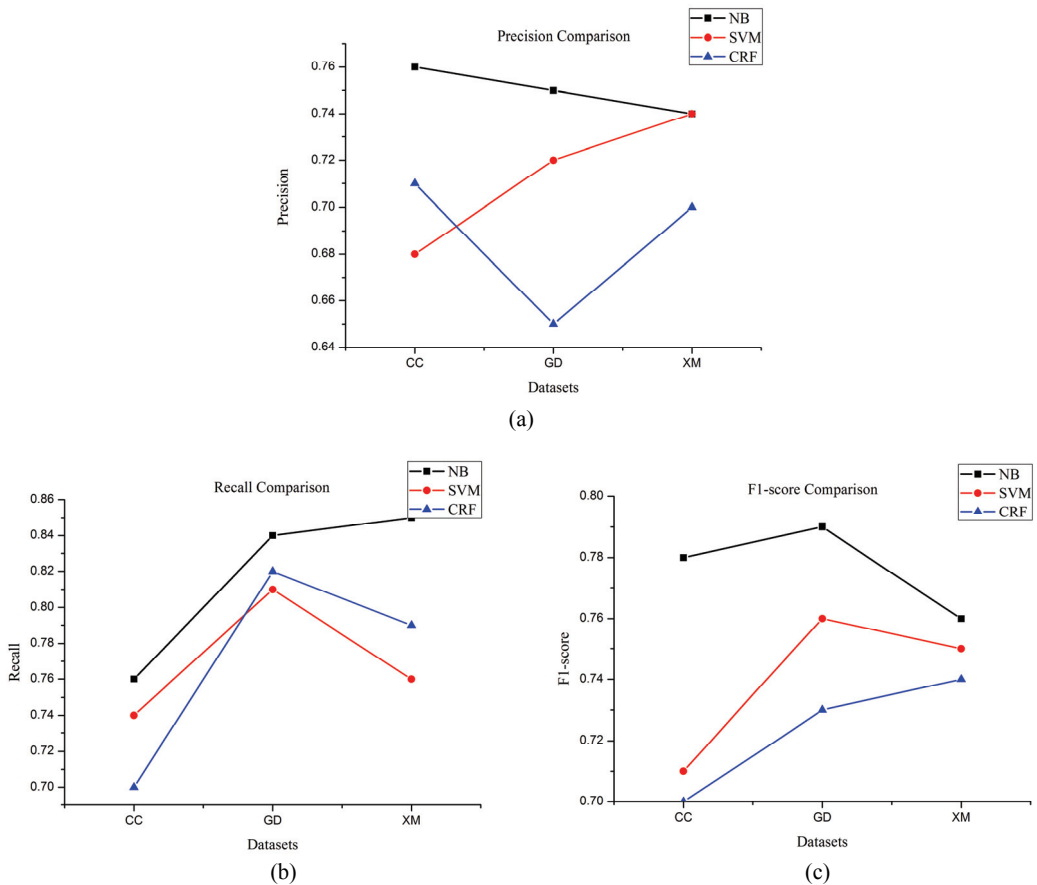

(a)

(b)

(c)

Fig. 3. Results of the comparison: (a) precision, (b) recall, and (c) F1-score.

We also compared the results of using this method to cluster sentences, and we used the results for the time parameter values of $t=1$, $t=5$ and $t=8$. We found that the F1-score value of not clustering is 67%, and the value climbed up to 76% after using the sentence clustering method with the time parameter $t=1$. Then, we further set the time parameter to $t=5$, and the result of the F1-score reached its highest value, but when we further increased the time parameter, the F1-score value dropped sharply. This result also verified the previous analysis. The sentiment views of barrages will be maintained for a certain period of time, but beyond that, the sentiment tendencies may be completely different. The comparison results are shown in Table 4, and we present the results of the comparison of different parameters in Fig. 4.

Table 4. Comparison of results for different *t* parameter values

NB	Precision (%)				Recall (%)				F1-score (%)			
	CC	GD	XM	Avg	CC	GD	XM	Avg	CC	GD	XM	Avg
Non-clustering	63	67	62	64	71	75	68	71	67	70	65	67
Clustering												
t=1	70	78	72	73	77	79	81	79	73	78	76	76
t=5	76	75	74	75	80	84	85	82	78	79	79	79
t=8	52	48	50	50	56	51	48	52	54	49	49	51
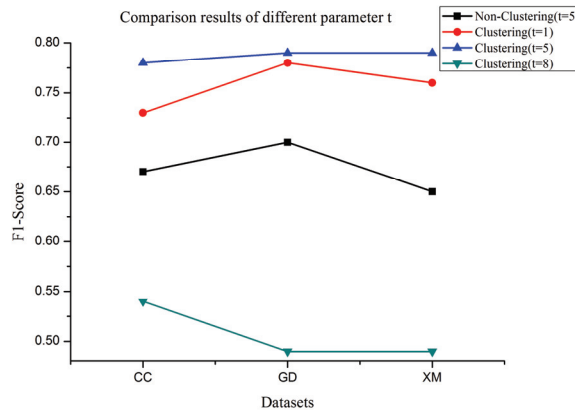
Fig. 4. Comparison results of the F1-score with different time parameter values.

In addition to the comparison between the three datasets, we also compared the machine learning algorithms. The SVM and CRF are good classification algorithms, but the naïve Bayes algorithm performed well in our experiments. We believe the main reason is that the NB algorithm assumed conditional independence; we also selected some popular network emotional words to simplify the classification task. The NB algorithm has a good effect on relatively simple classification tasks.

We also increased the comparison by including other current sentiment analysis methods, the aspect-based sentiment classification method proposed by Amplayo and Hwang [29] and the CNN-based method proposed by Feng et al. [30]. We used the public Yelp dataset (https://www.yelp.com/dataset challenge) as the experimental data, which was also used by Amplayo and Hwang [29], and we choose two categories of reviews: restaurant and shopping. The experimental results are shown in Table 5.

Table 5. Comparison of the results for different methods on the Yelp dataset

Yelp	Aspect-based [29]	CNN-based [30]	This study
Restaurant (%)	83.3	84.2	82.8
Shopping (%)	86.7	82.6	84.4

Table 5 shows the accuracy of different methods on the dataset. We can see that the latest deep learning method can achieve better results, but our traditional NB method performance is not too poor. To verify the efficiency of these methods, we compare the training time of the three methods. The results are shown in Table 6. It is obvious that our method is significantly less time consuming than the other two methods, which also reflects that our method is certainly efficient.

Table 6. Comparison of the training time results

Yelp	Aspect-based [29]	CNN-based [30]	This study
Training time (/s)	46	65	18

5. Conclusions

In this paper, we propose sentiment identification from webcast barrage reviews. This problem has not been studied to our knowledge. Sentiment analysis and opinion mining for webcast barrages is very important for practical applications. We conducted the data collection and preprocessing, and proposed a novel method of sentence clustering. We chose some of the popular network words as the seed words in our sentiment classification task. We compared the results using the sentence clustering method and analyzed the impact of the time parameters on the recognition of barrage sentiment tendencies. In our experiments, the highest F1-score of 79% was obtained when the time parameter was set as t=5. Compared to the three machine learning algorithms, NB, SVM and CRF, the NB algorithm performed the best. Extensive experiments show that our method can effectively categorize the sentiment of the webcast barrages to provide a reference for further data processing.

Acknowledgement

References

[1] B. Liu, *Sentiment Analysis: Mining Sentiments, Opinions, And Emotions*. Cambridge, UK: Cambridge University, 2015.

[2] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: tasks, approaches and applications," *Knowledge-Based Systems*, vol. 89, pp. 14-46, 2015.

[3] E. Cambria, "Affective computing and sentiment analysis," *IEEE Intelligent Systems*, vol. 31, no. 2, pp. 102-107, 2016.

[4] E. Cambria and B. White, "Jumping NLP curves: a review of natural language processing research," *IEEE Computational Intelligence Magazine*, vol. 9, no. 2, pp. 48-57, 2014.

[5] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, Philadelphia, PA, 2002, pp. 79-86.

[6] Z. Zhai, B. Liu, L. Zhang, H. Xu, and P. Jia, "Identifying evaluative sentences in online discussions," in *Proceedings of the 25th AAAI Conference on Artificial Intelligence*, San Francisco, CA, 2011, pp. 933-938.

[7] B. Pang and L. Lee, "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts," in *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, Barcelona, Spain, 2004, pp. 271-278.

[8] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and Trends in Information Retrieval*, vol. 2, no. 1-2, pp. 1-135, 2008.

[9] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Seattle, WA, 2004, pp. 168-177.

[10] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao, "Target-dependent twitter sentiment classification," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Portland, OR, 2011, pp. 151-160.

[11] L. Barbosa and J. Feng, "Robust sentiment detection on twitter from biased and noisy data," in *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, Cordoba, Spain, 2020, pp. 36-44.

[12] X. Hu, L. Tang, J. Tang, and H. Liu, "Exploiting social relations for sentiment analysis in microblogging," in *Proceedings of the 6th ACM International Conference on Web Search and Data Mining*, Rome, Italy, 2013, pp. 537-546.

[13] L. Gong, M. Al Boni, and H. Wang, "Modeling social norms evolution for personalized sentiment classification," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Berlin, Germany, 2016, pp. 855-865.

[14] S. Dong, X. Zhang, and Y. Li, "Microblog sentiment analysis method based on spectral clustering," *Journal of Information Processing Systems*, vol. 14, no. 3, pp. 727-739, 2018.

[15] A. Hassan, V. Qazvinian, and D. Radev, "What's with the attitude? Identifying sentences with attitude in online discussions," in *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, Cambridge, MA, 2010, pp. 1245-1255.

[16] M. Salehan and D. J. Kim, "Predicting the performance of online consumer reviews: a sentiment mining approach to big data analytics," *Decision Support Systems*, vol. 81, pp. 30-40, 2016.

[17] R. Socher, J. Pennington, E. H. Huang, A. Y. Ng, and C. D. Manning, "Semi-supervised recursive autoencoders for predicting sentiment distributions," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Edinburgh, UK, 2011, pp. 151-161.

[18] A. Ramesh, S. H. Kumar, J. Foulds, and L. Getoor, "Weakly supervised models of aspect-sentiment for online course discussion forums," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Beijing, China, 2015, pp. 74-83.

[19] M. Fernandez-Gavilanes, T. Alvarez-Lopez, J. Juncal-Martinez, E. Costa-Montenegro, and F. J. Gonzalez-Castano, "Unsupervised method for sentiment analysis in online texts," *Expert Systems with Applications*, vol. 58, pp. 57-75, 2016.

[20] J. Li, S. Fong, Y. Zhuang, and R. Khoury, "Hierarchical classification in text mining for sentiment analysis of online news," *Soft Computing*, vol. 20, no. 9, pp. 3411-3420, 2016.

[21] D. D. Wu, L. Zheng, and D. L. Olson, "A decision support approach for online stock forum sentiment analysis," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 44, no. 8, pp. 1077-1087, 2014.

[22] Y. Liu, J. W. Bi, and Z. P. Fan, "Ranking products through online reviews: a method based on sentiment analysis technique and intuitionistic fuzzy set theory," *Information Fusion*, vol. 36, pp. 149-161, 2017.

[23] D. T. Vo and Y. Zhang, "Don't count, predict! an automatic approach to learning sentiment lexicons for short text," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Berlin, Germany, 2016, pp. 219-224.

[24] H. Saif, Y. He, and H. Alani, "Semantic sentiment analysis of twitter," in *The Semantic Web – ISWC 2012*. Heidelberg: Germany, 2012, pp. 508-524.

[25] K. Schouten and F. Frasincar, "Survey on aspect-level sentiment analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 3, pp. 813-830, 2015.

[26] A. Hussain and E. Cambria, "Semi-supervised learning for big social data analysis," *Neurocomputing*, vol. 275, pp. 1662-1673, 2018.

[27] Y. Wu and M. Ester, "Flame: a probabilistic model combining aspect based opinion mining and collaborative filtering," in *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, Shanghai, China, 2015, pp. 199-208.

[28] Y. Wang, M. Huang, L. Zhao, and X. Zhu, "Attention-based LSTM for aspect-level sentiment classification," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, Austin, TX, 2016, pp. 606-615.

[29] R. K. Amplayo and S. W. Hwang, "Aspect sentiment model for micro reviews," in *Proceedings of 2017 IEEE International Conference on Data Mining (ICDM)*, New Orleans, LA, 2017, pp. 727-732.

[30] J. Feng, S. Cai, and X. Ma, "Enhanced sentiment labeling and implicit aspect identification by integration of deep convolution neural network and sequential algorithm," *Cluster Computing*, vol. 22, no. 3, pp. 5839-5857, 2019.

[31] L. Peng and Y. Liu, "Feature selection and overlapping clustering-based multilabel classification model," *Mathematical Problems in Engineering*, vol. 2018, article no. 2814897, 2018.

[32] N. W. Xue, "Chinese word segmentation as character tagging," *Computational Linguistics and Chinese Language Processing*, vol. 8, no. 1, pp. 29-48, 2003.

[33] C. J. Hutto and E. Gilbert, "Vader: a parsimonious rule-based model for sentiment analysis of social media text," in *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media*, Ann Arbor, MI, 2014, pp. 216-225.

[34] F. Pagani, M. De Astis, M. Graziano, A. Lanzi, and D. Balzarotti, "Measuring the role of greylisting and nolisting in fighting spam," in *Proceedings of 2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, Toulouse, France, 2016, pp. 562-571.

[35] W. Che, Y. Zhao, H. Guo, Z. Su, and T. Liu, "Sentence compression for aspect-based sentiment analysis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2111-2124, 2015.

Jun Li https://orcid.org/0000-0001-5591-721X
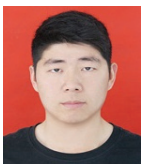
He is currently a PhD candidate at the School of Information and Communication, Guilin University of Electronic Technology. His research interests include natural language processing and sentiment analysis.

Guimin Huang https://orcid.org/0000-0003-3015-5639

He received Ph.D. degree in School of Computer Science from East China University of Science and Technology in 2005. He is now a full professor at Guilin University of Electronic Technology in China. His research interests include natural language processing and text mining.

Ya Zhou https://orcid.org/0000-0001-7294-7293

She is now a full professor at Guilin University of Electronic Technology in China. Her research interests include natural language processing and computer network.