Research on Emotional Orientation of Online Curriculum Review Text Based on Svm

Beilei Li

Chongqing Vocational Institute of Engineering University, Jiangjin, China

Keywords: Svm, Online courses, Reviews, Texts, Emotional orientation, Precision

Abstract: Emotion tendentiousness of users based on the review analysis refers to the use of machine learning or the method of semantic analysis, text information to user review the subjectivity as the research object, takes the comment text pre-processing, classification, clustering, feature extraction and sentiment analysis, in order to achieve the purpose of effectively excavate the potential value of information. Although some achievements have been made in the current research on the emotional tendency of review texts, most of them remain in the overall evaluation and criticism analysis of review texts, lacking the research on the emotional tendency of evaluation objects, and users need more fine-grained emotional analysis on evaluation objects, so the research has certain practical significance. In this paper, with the aid of predecessors on the basis of natural language processing and other related research, in view of the online network review the text puts forward a kind of emotional tendency analysis model for evaluation objects, this article will use the SVM based on the text to the text vector obtained using quantitative tools on the word graph model and combining the syntax of evaluation objects and interdependence between and evaluation feature extraction respectively, and the characteristics of evaluation after KU method to improve the emotional value of calculation, improve the evaluation characteristics of emotional measurement precision.

1. Introduction

Compared with news, periodicals, novels and other texts, online course comment texts are relatively short, and the phenomenon of online popular words, homophonic and wrong words is more common in comment essays. By effectively analysing these valuable essays, you can not only get many useful key information, relevant laws and business values from the commentary essays, but also further improve the adaptability of online courses. In addition, in the process of describing online courses, some users like to describe the advantages and disadvantages of online courses respectively, and then express their opinions on an online course as a whole, which makes the comment essays present a variety of emotional expressions.

Affective orientation analysis (text affective classification) is a hot topic in current research. Affective orientation analysis has been widely used in stock prediction, recommendation and other related fields. The main purpose of text affective orientation analysis is to effectively judge the category of subjective comment text affective orientation, which can be divided into positive category, negative category or multiple categories according to the actual situation of the research. In the research of text emotional orientation analysis, researchers conduct in-depth research from coarse-grained document-level emotional orientation analysis, sentence-level emotional orientation analysis to fine-grained emotional orientation analysis.

As the emotional tendency of text is mainly reflected by the emotional words in the text, the emotional dictionary has become a very basic work in the research of the emotional tendency of text, and it plays a very important role in the process of identifying the emotional tendency of text. Therefore, the construction of a high-quality and widely applicable affective dictionary has become an indispensable work in the research process of identifying the affective tendency of online comment texts. The following chapter will focus on the construction of affective lexicon and the research status of affective orientation analysis.

Since the affective dictionary is composed of a number of affective words, the core of

  DOI: 10.25236/icatpe.2019.014

constructing the affective dictionary is to distinguish the affective tendency of words. In the process of judging the propensity to emotional words, generally is by setting a worshiping value, then according to the method used to judge words emotions tend to score, and finally to the worshiping in the value of line, if the words emotions tend to score more than worshiping values were seen as positive emotional words, if the word emotion score is less than the value of worshiping is viewed as a negative emotional words.

Method based on semantic similarity discriminant method is commonly used words emotion tendency, the method in the process of distinguishing words emotion tendency, by using how net lexical semantic similarity method to calculate the candidate with all the positive emotional words benchmark (seed) the semantic similarity of emotional words and calculating the sum of the candidate with all the negative emotional words benchmark (seed) emotional words sum of semantic similarity, then its benchmark candidate with positive emotional words of word semantic similarity and minus the benchmark candidate with negative emotional words and word semantic similarity, which determine the emotional tendency of emotional words. Liu an equal word people calculate the process of emotional tendency weight calculation process is similar to that of a method based on semantic similarity and the difference is the method by calculating the candidate with positive emotional words benchmark average emotional word semantic similarity and calculating the candidate with a negative emotional words benchmark average emotional word semantic similarity, subtraction to get emotional words on both emotional tendency and emotional weight, finally on the basis of the construction of Chinese basic emotional lexicon. In the process of identifying the emotional tendency of words, the method based on semantic similarity needs to rely on the dictionary of the knowledge network and select relevant benchmark words. This method is difficult to effectively calculate the semantic similarity between the words such as network emotional words that are not included and the benchmark words, and thus fails to effectively judge the emotional tendency of these emotional words. In addition, the affective words with high similarity to the benchmark words are not necessarily consistent with the affective tendency of the benchmark words, and the method does not take into account the complex semantic structure and rules in the text.

2. Experimental Procedure

Dictionary rule method has been widely used in the field of text emotion analysis. This method mainly calculates the emotional tendency of text by using emotion dictionary, negative dictionary and other related dictionaries combined with manually defined rules. In most cases the method is based on the weight of emotional words centre (under normal circumstances, the positive emotion word weight of a set of positive and negative with the corresponding weight of negative emotional words), and then on the basis of combining with no dictionary definition and rules related to calculate the relevant documents of emotional propensity score, and worshiping at a set value of line, if the propensity score is greater than the set value of worshiping, said the text for positive text, if a propensity score less than worshiping value, has said that text is negative to the text .

This method has a simple calculation process and does not need to rely on the marked corpus, so it has a strong applicability in the field. When the rules set by the method are relatively simple, the method has a short running time and a fast operation speed. When the rules set by the method are complex, the running time of the method is increased. However, this method is susceptible to the influence of emotion dictionary and rule range. When the quality of constructing emotion dictionary is not high, this method is prone to misjudge the emotional polarity of the text. When the artificially defined rules are unreasonable, this method will also cause the wrong analysis of the emotional polarity of the text due to unreasonable rules.

Machine learning methods have been widely applied to text classification of emotion research, machine learning methods including supervised learning method, a semi-supervised learning methods, unsupervised learning methods, and in the case of labelling samples is enough for supervised learning method compared with the results obtained from the result of unsupervised learning and a semi-supervised learning method is more accurate. Support vector machine (SVM)

and naive bays method are two important classification methods in machine learning.

Word2vec, with three layers neural network consists of two kinds of Model of training, CBOW (Continuous Bag of Words Model) word Bag Model and Skip - “gram, and two kinds of accelerated algorithm (Negative Sample with Hierarchical Softback) to implement the vector, says Word2vec strong semantic relevance is a IDF amounted to less than the traditional TF, and high efficiency that Mikonos points out, An optimized stand-alone version can train hundreds of billions of words a day. Among them, the CBOW model predicts the probability distribution of the key words according to the context, while the Skip Gram model predicts the probability distribution of the context according to the key words, which is suitable for large corpus. Apart from the advantages and disadvantages of the two models, the main difference can be understood as the inverse three-layer model with incomplete symmetry to each other. Skim and Skim-Gram schematic models are shown in figure 1.

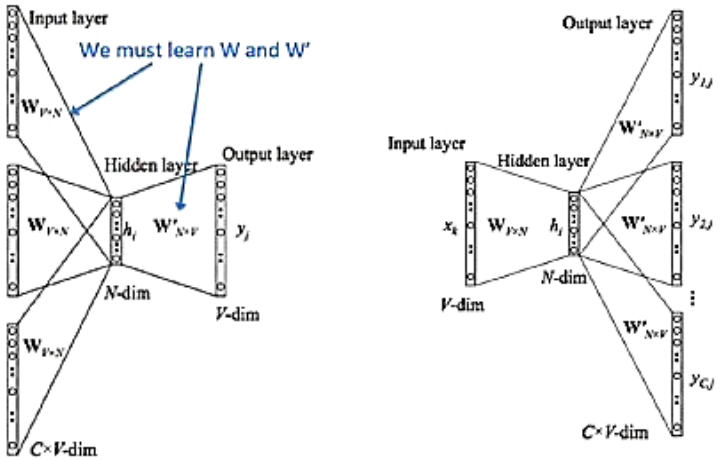

Fig.1 Skim-Gram Schematic of Cdow Model

In CBOW principle model, the basic idea is to use the window of the word vector used to predict the centre word, after an average model of the optimization goal is to predict the probability of one and the centre of the real word - hot vector is consistent, the ultimate goal is to the sum of all the vocabulary context vector to the target word by logistic regression probability of the path to maximize the product. W represents the words in the sample space, and Context (w) represents the Context content.

Text representation to quantify and text feature extraction for information retrieval, text mining research one of the most basic is the most important question, in the form of a key structure of the content of the text into a computer can process information, namely the mathematically to text information to set up a model, the model enables the computer to operating way to mining and analysis the content of the text. Usually directly using the text after word segmentation processing dimension to handle makes computer work is particularly big, not only inefficient, but also the accuracy of experimental results cannot be satisfactory, so the text further feature extraction can not only reduce the text of the original dimension can keep the original content of the text of greater degree is a very necessary work, the characteristics of the necessary work before feature extraction is the weight calculation, and then through specific worshiping values to filter the desired characteristics.

Support vector machine (SVM) is one of the common machine learning methods in the field of text emotion classification. The basic principle of the support vector machine method is to map the text data into a higher dimensional space, and establish an optimal classification hyper plane in this space to meet the classification requirements, and then use this optimal hyper plane to classify the samples of different category attributes. Naive bays method is also a simple and commonly used machine learning method, which is widely used in the research of text emotion classification. Naive bays classification method assumes that each attribute is mutually conditional independent, that is, the conditional probability of each feature word in the category of the text is mutually independent

with the conditional probability of other features in the category. It finally determines the category of the test text according to the probability value between the category and the word. Its main idea is to map the vector space formed by the input data into the high order feature space by means of nonlinear transformation. The nonlinear problem is transformed into a linear problem by kernel function to find the optimal separation plane in the high dimensional space, and then realize classification. See figure 2.

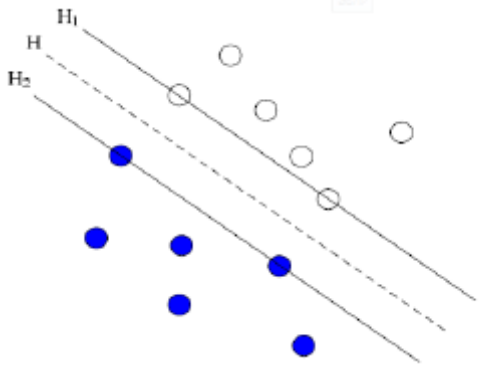

Fig.2 Svm Schematic Diagrams

The two types of points in the figure respectively represent two categories. The purpose of SVM is to find the optimal classification line between these two categories, as shown in figure H, forming the maximum classification interval between H1 and H2. Kernel function plays an important role in this process, which greatly reduces the difficulty of projection from low-dimensional space to higher-dimensional space. SVM has an excellent effect in the pattern recognition problem with small data size and high dimension and the nonlinear problem. In addition, in the process of seeking the optimal separation plane, it increases the generalization ability and avoids the over fitting problem in the learning process. Because of the excellent classification effect, it has been applied in many aspects.

3. Results and Discussion

The purpose of this experiment was to compare SVM with Baseline method to verify the feasibility of proposing a method to judge the emotional tendency of words. The second group of experiments was to compare the results of method classification in the absence of a single category of emotional relevancy when both kinds of emotional relevancy were considered. The purpose was to respectively verify the influence of the emotional relevancy of candidate emotional words and text categories and the emotional relevancy of benchmark emotional words on the discrimination of word emotional tendency. Since the affective words in the basic affective lexicon are the ones marked with affective tendency, the first and second groups of experiments used the affective words in the basic affective lexicon and existing corpus for verification.

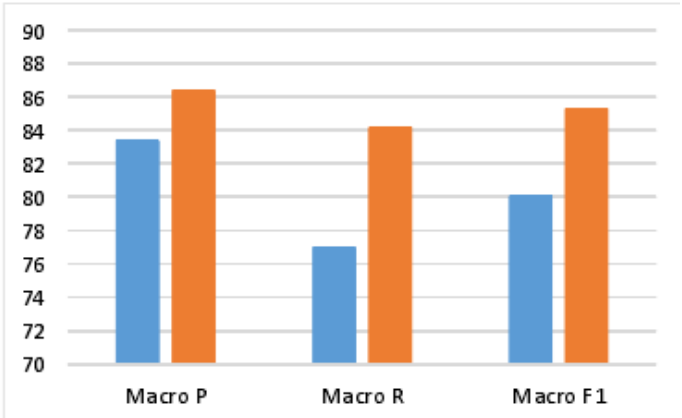

Fig.3 Results of Word Affective Tendency Discrimination by Different Methods (%)

The experiment of this group is to verify the feasibility of the method for judging the emotional tendency of words. For this group of experiments, PMI method was used as Baseline. During the experiment, when using the two kinds of method to calculate the candidate emotions tend to emotional words when the value is 0, the words are likely to exist in the number of positive emotional words more than the number of negative emotional words or a number of negative emotional words than positive emotional words, the number of random can be divided into positive or negative emotional words are emotional words is not reasonable. Therefore, in this group of experiments, the above correlation methods were used to calculate the situation that the affective tendency value of candidate affective words was 0, which could not be divided into positive affective words or negative affective words. Figure 3 shows the word emotional tendency discrimination results of the two methods. It can be seen from the experimental results that the overall effect of SVM is better than that of PMI method. The main reason is that PMI does not take into account the influence of negative words and word frequency on the affective tendency of candidate affective words, and it also ignores the influence of the affective correlation between candidate affective words and benchmark affective words on the judgment of the affective tendency of words. However, SVM takes into account the situation that negative words modify candidate emotion words in the text and the frequency of candidate emotion words in the category, and also considers the influence of the emotional correlation between candidate emotion words and benchmark emotion words on the discrimination of word emotional tendency. To some extent, therefore, the SVM not only can accurately identify negative word modify candidate emotions tend to emotional words, and can effectively identify the candidate of the emotional words contained in the two categories of document frequency the same situation candidate emotions tend to emotional words, but also can effectively determine the candidate with the benchmark emotional words related to emotional words case candidate emotions tend to emotional words. Therefore, compared with PMI method, SVM has a certain improvement in performance.

Fig.4 Influence of Affective Correlation on Word Affective Tendency Judgment (%)

To verify whether emotional relatedness, respectively, to influence the outcome of words emotions tend to judge, this chapter by comparing the following three conditions: the SR SWR (regardless of the candidate and text categories of emotional words of words emotion tend to judge the effect of emotional relatedness), SR SWT (regardless of the benchmark candidate emotional words and feelings of emotional words emotional tend to judge the influence of the relative degree of words), SO the SRC (considering both emotional tend to judge the influence of the relative degree of words). Figure 4 shows the experimental results of the discrimination of word emotional tendency by the degree of emotional correlation between the candidate emotion words and the category. It can be seen from the experimental results that the effect is the best when the degree of emotional correlation of the two categories is considered. The reason is that some of the candidate emotion words that are not modified by negative words will appear in the text of the opposite category of their emotional tendency. Therefore, when the emotional correlation between candidate

words and benchmark emotion words is not considered in the calculation process, the method is likely to wrongly identify the emotional tendency of such emotional words. However, the emotional relevance between candidate emotion words and text categories is not considered, which makes the method overly dependent on the benchmark emotion words, unable to identify the emotional tendency of words that are not related to the benchmark emotion words, and unable to effectively identify the emotional tendency of candidate emotion words that are strongly related to text categories. When the two categories of affective relevancy are taken into account, the problem of inaccurate judgment of the affective tendency of words is further avoided when only the affective relevancy between candidate affective words and text categories is taken into account, or when only the affective relevancy between candidate affective words and benchmark affective words is taken into account. Therefore, the influence of the affective relevance degree between candidate affective words and text categories and benchmark affective words on the result of the discrimination of affective tendency of words cannot be ignored.

4. Conclusion

The appearance of short text has attracted more and more researchers' attention, and the analysis of short text's emotional orientation has become a very important research work. In this paper, the author builds an extended affective dictionary and proposes a method to analyse the affective orientation of short essays by integrating the contribution of emotion. However, there are still some problems in the analysis of emotional orientation, which need to be further studied. In the future, in-depth research will be conducted on the analysis of emotional orientation in the review essay from the following aspects:

1) The affective neologisms and implicit affective characteristics contained in the short text are the factors influencing the analysis of the affective orientation of the short text. Therefore, the next major research work is how to more accurately extract the affective neologisms and implicit affective characteristics contained in the short text, and further analyse the affective tendency of the short text by integrating the affective neologisms and implicit affective characteristics into the method.

2) Some of the emotional words in the short text may not express the same emotional tendency in different short text as they do in the emotional dictionary. Therefore, how to more effectively analyse the semantic emotion information expressed by each emotion word feature in different short texts will become the next major research work.

References

[1] Xu Z, Huang C, Pan H. Online comments of multi-category commodities based on emotional tendency analysis[J]. Cluster Computing, 2018:1-13.

[2] Qiu Z, Zhou L L. Apply Emotional Tendency Analysis to Sina Microblog[J]. Applied Mechanics & Materials, 2014, 668-669:1407-1410.

[3] Tan C F. The Chinese Microblog Emotional Tendency Analysis Based on Sentiment Dictionary [J]. Applied Mechanics & Materials, 2013, 333-335:795-798.

[4] Thompson D, Mackenzie I G, Leuthold H, et al. Emotional responses to irony and emoticons in written language: Evidence from EDA and facial EMG[J]. Psychophysiology, 2016, 53(7):1054-1062.

[5] Sun W, Pan M, Ye Q. Comparative Study on Objective and Subjective Emotional Tendencies of Online Reviews from Different Sources[C]// International Conference on Internet Technology & Applications. 2011.

[6] Lowe R, Ziemke T. The Feeling of Action Tendencies: On the Emotional Regulation of Goal-Directed Behavior[J]. Frontiers in Psychology, 2011, 2(4):346.

[7] Wang W, Yu Y, Zhang J. A new SVM based emotional classification of image[J]. Journal of Electronics, 2005, 22(1):98-104.

[8] Chiang W C, Wang J S, Hsu Y L. A Music Emotion Recognition Algorithm with Hierarchical SVM Based Classifiers[M]. 2014.

[9] Yang N, Yuan J, Zhou Y, et al. Enhanced multiclass SVM with thresholding fusion for speech-based emotion classification[J]. International Journal of Speech Technology, 2016, 20(1):1-15.