Chinese Sentence Emotion Classification Method Based on Semantic Analysis

Yuqiansi He
Guangzhou Songtian Vocational College, Guangzhou, Guangdong, 511370, China

Keywords: Chinese Sentences, Emotional Classification, Emotional Dictionary, Support Vector Machines, Semantic Analysis


1. Introduction

Emotion is the Attitude Experienced by People When They Are Satisfied with Their Own Needs. This Concept Involves People's Opinions, Opinions and Judgments. Emotional Analysis is Also Called View Mining or Opinion Mining, Which Refers to the Process of Identifying and Extracting Relevant Tendentious Information from Subjective Texts. the Emotional Tendency of a Sentence is to Indicate the Positive or Negative Feelings of the Author in the Sentence. Sentence Emotional Tendency Analysis Method Based on Text Sentiment Analysis Research, through the Analysis of the Characteristics of Text Language Sentence, Draw Lessons from and Improves Text Emotional Analysis Related Technical Way for Research.

According to the Granularity of the Text, the Related Research of Text Emotion Analysis Can Be Divided into Three Categories: 1. Emotional Analysis of Sentence Level; Chapter Level Emotional Analysis. the Study of Emotional Orientation At the Word Level (Emotional Dictionary Construction) is the Basis and Precondition of Other Levels of Emotional Analysis. in Sentence Level, Emotional Tendency Analysis for a Particular Context of Sentences, the Sentence Tendentiousness Information Extraction and Analysis, the Analysis of Emotional Tendency and Tasks Include Judgment Sentence Extraction Emotional Tendencies Related Elements, Etc. in the Text Level, the Analysis of Emotional Orientation Refers to the Emotional Tendency of the Whole Text. If the Text as a Whole Processing Objects, There Are Greater Limitations, Because the Analysis Assumes That the Text is on the Objects of the Same Comments, and Real Text, Often Contain Several Comments Objects, and for Different Object's Point of View and Attitude
Orientation Information is Different. Therefore, There is a Tendency to Reduce the Trend of Discourse Level of Emotion Analysis. Sentence Affective Analysis Refers to the Analysis of the Opinions or Attitudes Expressed in the Sentence or the Subjective Information Expressed in the Sentence. Sentences Are: Emotional Tendency Analysis with Information Retrieval and Natural Language Processing, Information Extraction, Machine Learning and Computational Linguistics, Etc., Its Main Purpose is to Let the Computer Automatically Judge the Submitter is Reflected in the Sentence the Main Objective, the Positive and Negative Polarity and the Intensity of Emotional Tendency. According to the Different Goals, the Analysis of the Emotion Tendency of Sentences Can Be Divided into the Following Aspects: 1. Subjective and Objective Analysis: to Analyze Whether the Content of the Text Expresses Opinions and Emotions; Emotional Bias Analysis: Classifying the Emotional Polarity of Sentences, Such as Classifying Sentence Affective Categories into Positive and Negative Ones; Analysis of Emotional Intensity: the Analysis of the Emotional Intensity of the Sentence, Not Only the Emotional Inclination of the Sentence, But Also the Intensity of Emotional Inclination; (4) the Evaluation Object Extraction: for in the Sentence with the Vocabulary of Emotional Colour, Not Only for a Evaluation Objects, Mining Evaluation Objects, Emotional Words to Extract Effective Emotional Characteristics, Can Be Realized in View of the Evaluation Object Sentence Sentiment Analysis; (5) Research Views Holders: Sentences Are Copycat Campaign, Celebrity Endorsements and Other Characteristics, Analysis the Sentence the Holders of the Popular View, Emotional Tendency of the Masses Can Grab for a Certain Event, Conducive to the Public Opinion Supervision and Guidance; Abstract: in the Age of National Sentences, the Emotional Viewpoint of Massive Sentences Can Be Used for Commercial Service, Product Reputation and Public Opinion Survey of Government Departments. from the Current Study, the Text Affective Categories Are Usually Divided into Two Categories (Heads, Tails, and Neutral), and the Two Categories Are Mostly Considered. the emotional Analysis of Sentences Belongs to the Dichotomy Problem, Which is to Classify Sentences as Positive or Negative, Positive or Negative. Therefore, the Emotion Classification of the Large Number of Sentences is Carried out, and the Emotional Polarity Can Be Summed Up by Computer Automatic Statistics, Which Can Summarize the Attitude of People to an Objective Thing. At Present, the Academic Community Mainly Uses Machine Learning and Semantic Analysis to Analyze Sentence Emotion. the Machine Learning Methods Are Commonly Used in Svm, Nb, Me and K-Nearest Neighbour. in the General Method of Semantic Analysis is Commonly Used in Text Classification with Probabilistic Latent Semantic Analysis (Aim-Listed Probability Latent Semantic Analysis, Plsa), Lsa, and Based on These Two Methods to Improve and to Other Methods. There Are Two Main Categories of Emotion Classification Based on Semantic Analysis: First, the Method of Calculating the Emotional Value of Emotional Dictionary. the Second is to Construct the Semantic Pattern Library Based on Emotional Dictionary, and Use Pattern Matching Method to Calculate Emotional Tendency.

In Previous Research, the Researchers Mainly Used Machine Learning and Rule-Based Classification Methods to Analyze the Text. by Using High Quality Emotional Dictionaries, It is Easy to Achieve Better Emotional Feature Selection and Emotional Analysis. Based on Machine Learning, the Emotion Word is Generally Used as the Feature of Classification, and the Other Feature Training Classifier is Used to Complete the Text Emotion Classification Task. the Rule-Based Approach Determines the Emotional Tendency of the Sentence by Using the Polarity of the Emotional Word Contained in the Sentence. Liu Zhiming et al. Used Three Feature Selection Algorithms, Three Feature Weight Calculation Methods and Three Machine Learning Algorithms to Analyze the Sentence. the Experimental Results Show That the Svm and Nb Classification Algorithm Have Advantages in Different Feature Weight Calculation Methods, and Ig Feature Selection Method is Better Than Other Methods. Considering Three Different Factors, Using Svm and Ig, as the Feature Weight Value, the Combination of the Three Can Achieve Better Emotional Classification Effect. Lin Jianghao et al. Combined the Syntactic Path and Emotional Lexicon to Extract the Emotional Features, Constructed the Naive Bayesian Classifier, and Studied the Emotional Orientation of the Sentence Text of the Hot Topic. the Influence of Punctuation Mark
The Accuracy Improved to 67.283%. in the Rule-Based Method, the Recognition of Emotional Evaluation Unit Has a Great Influence on Improving Classification Accuracy. Since the Relationship between Emotional Evaluation Unit and Modifier is Ignored, the Method Can Only Be Used to Analyze Simple Sentence Patterns. in Order to Further Explore the Relationship between Evaluation Words and Evaluation Objects, Yao Tianfang et al. Summarized the Matching Rules of “Ascending Path” and “Down Path” Based on the Dependency Syntactic Analysis. Put Forward a Kind of Emotional Evaluation Unit Automatic Identification Method Based on Syntactic Path, the Method to Automatically Syntactic Path to Describe Evaluation Objects and the Evaluation of Relationship between Words. the Above Methods Are Widely Used in the Analysis Tasks of News, Film and Product Reviews. Pang Lei and Others Propose a Method of Unsupervised Emotional Classification Based on Emotional Knowledge. Use of Emotional Words and Expressions Picture Two Knowledge of Large-Scale Sentence Not Filtered Tagging Corpus and Automatic Tagging, with Automatic Tagging Good Corpora as a Training Set to Build Emotional Text Classifier, Emotional Polarity of Sentence Text Automatic Classification, the Highest Classification Accuracy of 83.6%.

\[
p(di) = \begin{cases} 
1 & \text{sumPos}(di) > \text{sumNeg}(di) \\
0 & \text{sumPos}(di) = \text{sumNeg}(di) \\
-1 & \text{sumPos}(di) < \text{sumNeg}(di)
\end{cases} 
\]

(1)

By using high quality emotional dictionaries, it is easy to achieve better emotional feature selection and emotional analysis. This method mainly determines the emotional inclination of the sentence through the polarity of the emotional word contained in the accumulative sentence, and the principle is shown in formula (1).

In the above formula, \( P(di) \) represents the emotional polarity value of document \( di \), and Sum Pos(\( di \)) represents the number of positive emotional words contained in document \( di \), and Sum Neg(\( di \)) indicates the number of negative emotional words contained in document \( di \).

There are two main ways to construct an emotional dictionary: artificial collection and based on corpus statistics. The methods of manual collection and sorting through statistical analysis of a large number of relevant corpus, extract effective affective words, and adopt the method of manually tagging emotional weights to construct emotional dictionaries. It has the characteristics of high accuracy, simple and easy to do, but it is often time-consuming and inefficient. Method based on corpus statistics is a group for a given number of positive seeds seed word set and the negative word set, through the words from the corpus of calculation and the seed word set, the similarity between the size automatically recognize words according to the similarity of emotional polarity, added to the corresponding emotion in the dictionary. According to the method of corpus statistics, the advantages are short time and no manual intervention. Its disadvantage is that the quality of the set of seed words directly affects the affective values of words.

Sentiment analysis method based on dictionary has the characteristics of simple, efficient, fast, but the performance of classification depends on the quality of the build emotional dictionary, when text semantic structure more complex, the simple use of the word frequency statistics method to calculate the text emotional words emotional polarity, and cannot achieve good classification effect,
needs to consider the syntactic structure of the text, context scenarios, rhetorical devices and other factors on the impact of emotional classification. Therefore, this method is widely used in the analysis of emotional analysis of news, film and product reviews. However, the relationship between emotional evaluation unit and modifier is ignored because the emotional word polarity accumulative method is not ideal for the classification of sentence affective analysis. Support vector machine is a machine learning algorithm based on the theory of statistics. The support vector machine is widely used in the regression and solving of classification and other supervisory learning problems due to its good precision and excellent learning performance. Text emotion classification based on support vector machine (SVM), boils down to in n-dimensional feature space for maximum hyperplane segmentation problem, the hyperplane maximum points of text both positive and negative emotion attribution, two dimensional feature space is shown in figure 1.

![Fig.1 Linear Partitioning is the Most Hyperplane](image)

Figure 1 shows two types of data, where d represents the maximum margin; H is the optimal hyperplane. Let's say that in the m-dimensional eigenspace X is equal to \{x1, x2,..., x_m\}, there are n points that belong to two classes: c+ and c -. The support vector machine actually finds such a discriminant function: \( f(x) = w \cdot x + b \), which makes \( f(x_i) > 0 \), \( x_i \) is identified as \( c+ \) class; When \( f(x) \) is < 0, \( x_i \) is discriminated as \( c- \) class. That is:

\[
f(x) = \begin{cases} 
-1 & x_i + b < 0 \\
0 & x_i + b = 0 \\
1 & x_i + b > 0 
\end{cases}
\]  

(2)

In order to ensure maximum classification performance, select the hyperplane with the maximum edge distance between positive and negative classes. The above discussion is the case of positive and negative linearly separable, which is that the discriminant function is a linear function. In the case of nonlinear separable, the support vector machine provides the kernel function mechanism to transform the original data into higher dimensional feature space and solve the nonlinear decision boundary problem. The commonly used kernel functions are polynomial kernel functions, gaussian radial basis functions, s-shaped kernel functions, etc. Support vector machine (SVM) is to have the advantages of solid mathematics theory foundation, time and space complexity of the algorithm is only related to the number of vector, has nothing to do with the dimension of feature space, and the algorithm and kernel function is independent each other, according to different scenarios can switch different kernel function in the algorithm. The lack of support vector machine (SVM) is that it only supports the binary classification of the text, in considering multiple classification, must take a certain classification combination strategy, like for many methods of classification, the SVM decision tree classification, or combined with other classifiers.
2. Research on the Method of Emotional Dictionary Construction

The discrimination of word affective orientation is the basic work of sentence affective analysis. Many studies show that emotional dictionary based emotional analysis is a very effective method, and the method has been widely used. This chapter will analyze the existing methods of emotional dictionary building, and puts forward some ideas of improvement on the existing research methods, so as to realize the construction of Chinese basic emotional lexicon, network language emotional dictionary and emoticons dictionary.

Most of the researchers involved the Chinese affective dictionary type mainly in the polar lexicon. At present, the typical research method for text emotion analysis is to combine the emotion dictionary with the feature extraction of the text, then combine the machine learning implementation to classify the emotional inclination of the text. But because the network comments tend to have spoken of the ideas expressed in, colour dense and non normative characteristics such as language, the researchers based on machine learning is introduced to analyze the semantic rules, to improve the text emotion classification system. These improvements have improved the system classification performance. But there is no fundamental solution to the diversity of Chinese texts in complex contexts. In addition to the polarity of reactive words, the Chinese emotional vocabulary ontology library, such as xu linhong, also includes the strength value of the emotional tendency of words, and also includes a small number of polysemous words. These polysemous words contain two or more polarity annotations and corresponding affective intensity values. However, the construction of the emotional vocabulary ontology library does not have a multi-directional analysis on the emotional tendency of each word. This makes the emotional polarity and the corresponding intensity of the emotion dictionary a one-sided and not universal applicability. Aiming at this problem, this paper proposes a comprehensive consideration of words in different semantic emotional tendency under the circumstance of the calculation method of weighting, to build a strength of both positive and negative emotions tend to value the emotion in Chinese dictionary. In this paper, the dictionary is applied to the extraction of the emotion feature of the text, and a better classification effect is obtained in the text emotion tendency classification experiment, and the validity of this dictionary is verified. Collection of words collection is the basic work of emotional dictionary construction. This paper collected How.net, NTUSD, Chinese emotional words praise or blame righteousness dictionary and Yang Ding build emotional lexicon, adopt the method of merging to heavy word set, and will set them as the basis of SCSL emotional words. The word collection contains 2,4130 words, with 10,269 positive words and 13,861 negative words. In How Net, words tend to contain multiple meanings, which are used to represent different meanings reflected by words in different contexts. Define a Chinese word meaning element as W, then W= {M, Mz,... N denotes the number of meanings contained in words, MN (N =1, 2,... N) is the NTH element of the word. By programming to call How Net's search API, you can get the semantic collection of words, for example, the word “beauty” is set to {beautiful, comely, handsome, and pretty}. If there is no input word in How Net, the API provided by youdao dictionary is used to obtain the corresponding English explanation of the word as the meaning element of the word. However, in the process of dictionary construction, there are still a few words that can't obtain the collection through the above methods. At this time, the article USES the method of human intervention to generate the corresponding collection of words.

WPos is the positive emotional strength value of words, and WNneg is the negative emotional intensity value of words

\[
WPos = \frac{1}{n} \sum_{i=1}^{n} MPos_i
\]

\[
WNeg = \frac{1}{n} \sum_{i=1}^{n} MNeg_i
\]
Among them, $nMPos$ is the positive emotional intensity value of the NTH, and $nMNeg$ is the negative emotional intensity value of NTH.

Synonyms collections are Sent Word.net component unit, a synonym set consists of multiple words meaning close, to a semantic concept. At the same time, Sent Word Net in each set of synonyms are corresponding with the intensity of emotional tendency value annotation, as shown in table 1, Pos said positive emotional intensity values, Neg said negative emotional intensity values. An English word may exist in many synonyms, indicating that it contains many different meanings. Definition set meaning Yuan synonymous with M package collection of S (M), S (M) = (S, S, 2., Sk}, where K means, set the number of synonyms Sk (K = 1, 2, K,) contains input righteousness Yuan for the first K M set of synonyms. Programming using Sent Word Net database to retrieve input righteousness Yuan can be synonymous with corresponding collection of positive and negative emotional intensity values.

### Table 1 Sent Word Net of Synonym Set Example

<table>
<thead>
<tr>
<th>Synonyms and annotations</th>
<th>Pos</th>
<th>Neg</th>
</tr>
</thead>
<tbody>
<tr>
<td>{pride#1,plume#2,congratulate#3};be proud of; “He prides himself on making it into law school”</td>
<td>0.25</td>
<td>0.5</td>
</tr>
<tr>
<td>{pride#1,p pridefulness#1};a feeling of self-respect and personal worth</td>
<td>0.63</td>
<td>0</td>
</tr>
<tr>
<td>{pride#2};satisfaction with your (or another's) achievements; “he takes pride in his son's success”</td>
<td>0.38</td>
<td>0.25</td>
</tr>
<tr>
<td>{pride#3};the trait of being spurred on by a dislike of falling below your standards</td>
<td>0.5</td>
<td>0.25</td>
</tr>
<tr>
<td>{pride#4};a group of lions</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>{pride#5,superbia#1};unreasonable and inordinate self-esteem (personified as one of the deadly sins)</td>
<td>0.63</td>
<td>0.25</td>
</tr>
</tbody>
</table>

### 3. Results and Discussion

LPT provides rich and efficient natural language processing modules including participles, lexical annotations, dependency syntactic analysis, naming entity recognition, and semantic role labeling. The dependency grammar reveals its syntactic structure by analyzing the dependencies between the components in the language unit. Intuitively, the syntactic analysis identifies the grammatical components of “subject-verb,” “definite complement”, and analyzes the relationship among the components. There are 14 relationships in the analysis of the dependency syntax of LTP3.1.0. The analysis of dependency syntax helps to dig deeper into the relationship between emotional evaluation unit and evaluation word. Sentence “I like their intelligence and the spirit of hard work.” The core of the predicate is “like”, the subject is “I”, “like” object is “spirit”.

In order to accurately calculate the phrases and sentences of emotional intensity values, the need for different syntactic relationship type a reasonable calculation rules, makes the calculating process to reasonably transfer and change process of emotional expression. The following are the emotional calculation rules for each syntactic pattern. Rule 1: if the control word phrases and modifier (SBV or vow emotion mode, depending on the different collocation condition to calculate the phrase emotional value. The word said modifier, S (word) means the emotional intensity of the word value, if the corresponding word in the dictionary does not exist the emotional intensity value is 0. With words, the emotional computing expression of the SBV and vow mode is S(words)@S (word), where @ represents the calculation method of these two affective modes. Rule 2: if the phrase syntactic relationship is the “the” mode, then judge whether the word in the pattern exists in the adverb dictionary, and if there exists, the emotion word should be revised in strength. The degree adverb has six levels of correction strength. Using S(degree) to indicate the degree of correction of adverbs, S(word) means an emotion word, and the equation of the equation is S (degree), which is the operator of the pattern. Rule 3: if the mode of an emotional expression is at or CMP, then the polarity of the emotional word is calculated. Word1 and word2 represent the words in the phrase respectively. The emotional computing expression of art and CMP emotions is S (word1) circle of S (word2), which represents the calculation method of these two affective modes. After syntactic analysis of the input sentence, the emotional expression tree needs to be constructed in order to calculate the emotional value of the emotional pattern combination. After facial expression detection and text segmentation, input sentence D will be divided into several clauses. Set the set of
clause S={s1, s2, s3,... The initial set of words corresponding to the clause is WSi, and the expression tree is a binary tree constructed by the following algorithm, set to T.

Expression tree construction algorithm:
Input: D, S, WSi
Output: T
Algorithm:
One of the clauses in the FOR D
IF si is not emoji
An emotional pattern in FOR si
Read the emotional pattern; Vocabulary w1 and w2;
Build tree T1; T1 - > Left Child = w1; T1 - > Right Child = w2; T1->Data= operator;
IF w1 ∈ WSi && w2 ∉ WSi
Set the subtree of the word to T0; T1 - > Left Child = T0;
T1 - > Right Child = w2; T1-> Data = operator; T0 = T1; Add w2 to WSi;
ELSE IF w1 ∉ WSi && w2 ∈ WSi set in the word subtree to T0; Child = T0 T1 - > Right;
T1 - > Left Child = w2; T1-> Data = operator; T0 = T1; W1 join WSi;
ELSE IF w1 ∉ WSi && w2 ∉ WSi
Add w1 and w2 to WSi; T - > leftchild = T1;
T - > Right Child = T; T - > Data = “+”; T = T;
ELSE delete the schema and the words at both ends of the schema;
END FOR ELSE IF T==NULL
T->Data= emojis
The ELSE
T - > Left Child = T;
T - > Data = “+”; 
T’-> Right Child= emojis;
T = T;
END the FOR

4. Conclusion

Research in basic emotional lexicon building method, this paper proposes a construct basic emotions in Chinese dictionary based on semantic analysis method, and use an emotional dictionary correction methods to modify words emotions tend to value; In the experiment of sentence emotion classification based on SVM, the classification effect of feature extraction using this dictionary is better than that of general polar emotion dictionary.

In terms of network language and emoticon emotive dictionary construction, different methods are used to construct. Based on semantic analysis of sentences emotion classification experiment is introduced into the network language and emoticons emotional dictionary to experiment, the results show that the construction of network language emotion dictionary and emoticons can effectively improve the performance of the classifier. In the study of sentence affective classification, a semantic analysis based on semantic analysis is presented. This method USES the emotional dictionary constructed to use the dependent syntax of the sentence text, and constructs the emotion expression tree according to the proposed algorithm. Then, according to the established rules, the emotional intensity of sentences is calculated according to the emotional intensity value. This method has good effect in sentence emotion classification experiment. In addition, this method can not only determine the emotional category of sentence, but also can calculate its intensity, which provides a better research foundation for the analysis of the emotional tendency of the mass sentences. In the research of experimental system, TSAES was designed and implemented. TSAES includes text data collection and management system, emotional dictionary building system, text emotion analysis system and text emotion analysis visualization system. TSAES has a friendly
interface and high stability, which provides a convenient and effective experimental tool for the study of text emotion classification.

References


