ATLANTIS PRESS
Advances in Social Science, Education and Humanities Research, volume 341
5th International Conference on Arts, Design and Contemporary Education (ICADCE 2019)

Cross-cultural Learning Resource Recommendation Method and Corpus Construction Based on Online Comment Sentiment Analysis

Dongmei Li
School of Foreign Language
Fujian Jiangxia University
Fuzhou, China 350108

***Abstract*—Through the emotional analysis of cross-cultural learning resource lines, a data-oriented online commentary sentiment analysis system was established. The emotional bias of the reviewer's content is obtained from a large number of online comments, and the relationship between the various elements is comprehensively considered to realize the multi-classification of online commentary corpus emotions. According to the results of emotional assessment, dynamic and static sentiment lexicons are established to expand the existing emotional corpus, and provide reference for learners to choose cross-cultural communication resources. It is necessary to combine corpus research with computer science to achieve corpus research technology upgrade.**

Keywords—cross-cultural learning resources; online reviews; emotional analysis

I. INTRODUCTION

Corpus development tends to be applied to big data. In the era of big data, various new types of natural corpus are emerging, which is conducive to expanding the boundaries of corpus research and development. At the same time, with the progress of network technology and natural language processing technology, the graphical interface of various natural language processing software, the popularity of network application and the gradual rise of the fourth-generation corpus analysis tools are all conducive to reducing the difficulty of corpus processing and corpus construction. In the language service industry, big data technology is mainly applied in the following aspects: visualization analysis based on big data, prediction based on big data, and business transaction based on big data, among which visualization analysis and business transaction based on big data can be combined with corpus construction and research. Big data transactions include such services as corpus or translation memory database transactions and multilingual data processing services. As a language resource in a narrow sense, corpus not only has academic value in language research, but also has commercial value in natural language processing, such as lexicography, machine translation and software development. Therefore, the finished product and various links in the development process can also provide business opportunities for big data trading. Sentiment analysis as an important branch field corpus study, involving multiple disciplines, such as linguistics, sociology, management, economics, computer science, etc., but it is not a more complete and reliable theory system, this topic proposed from the perspective of different disciplines and to build a relatively perfect emotional analysis theory, as a basic framework of research, so as to improve the effectiveness of the selection decision.

II. LITERATURE REVIEW

Database design should be considered according to research purpose before corpus development. The significance of corpus in language research lies in that "through corpus, we can observe language patterns that we were not previously aware of or only vaguely aware of" (Johansson 2007:1). This also means that the capacity of corpus is generally larger, and especially for sporadic corpus research needs to be based on "corpus", leading to "corpus" and "corpus" confused with each other. Secondly, the corpus retrieval software can be used to obtain data without any explanation. It has the name of "corpus" but no corpus. Corpus is the starting point and core of corpus research. The problem of corpus design and editing is directly related to the validity and reliability of corpus research.

In the development of corpus, a uniform labeling system should be determined. With regard to the English word annotations, Lancaster university published the English word collections CLAWS5 and CLAWS7 and provided online annotations on the CLAWS WWW tagger; In terms of Chinese part of speech tagging, there are various Chinese part of speech tagging sets such as Chinese academy of sciences (ICTPOS), Beijing university (PKU), and Chinese communication (CUC), which are slightly different in the classification of part of speech and have coarse and fine segmentation granularity. Therefore, selection should be made according to the research purpose. The online Chinese word segmentation and part-of-speech tagging can be provided by the "corpus online" website and the national audio media center for language resources monitoring and research, affiliated to the computational linguistics research office of the language and text application research institute of the ministry of education. However, basic standards such as sampling and labeling have not been unified in corpus

development. For example, different labeling results in word segmentation process will affect the reliability and validity of subsequent studies. At the same time, single and multimodal corpus collection, retrieval and statistical tools for Chinese translation language analysis are not enough.

As an important branch of corpus research, sentiment analysis involves many disciplines, such as linguistics, sociology, management, economics, and computer science. However, there is no more complete and reliable theoretical system. Different disciplines and different perspectives are used to construct a relatively complete sentiment analysis theory as a basic framework for further research, thus improving the effectiveness of decision-making.

With the rapid development of information technology, more and more learners share their learning experience through network platforms, thus forming online comments that take the Internet as the production and communication medium. Online comment information expresses learners' views and opinions on online learning resources, which is one of the important sources for learners to obtain information in the Internet age. Therefore, digging out valuable information resources from a large number of online comments becomes an important decision basis for online learning resources selection. Emotional analysis of online comments aims to obtain the reviewers' emotional bias to learning resources from a large number of online texts, so as to dig out the reviewers' interest and focus on learning resources. Emotional analysis can be divided into three levels: document, sentence and word. In the study of document-level emotion analysis, Yu H (2003) used the multi-naive bayesian model to train online texts and obtained better emotional output results. Pang B and Lee (2008) used PMI information and SVM method to learn corpus and build an emotional tendency classifier. Turney P d. (2002) proposed an unsupervised point mutual information method based on semantic analysis. He et al. He R, Gonzalez H. (2017) proposed a minimization optimization algorithm based on optimal control. In terms of sentence level, Meena A, Prabhakar T V. (2007) considered the influence of conjunctions on sentence structure, and then analyzed comments and classified emotions based on syntactic dependency tree. Zhang chenggong, liu peiyu and zhu zhenfang (2012) constructed a polarity dictionary, analyzed the influence of modifiers on polar words and formed polar phrases, and proposed an emotional analysis method based on the polarity dictionary. In the word-level fine-grained emotion analysis, Fu X, Liu G, Guo Y, (2013) combined the LDA theme model and HowNet Dictionary's unsupervised method to automatically implement emotion analysis. Kim s. M, Hovy e. (2006) established the emotion dictionary based on WordNet, and identified the emotion words and the polarity of emotion, then detected the resource characteristics and calculated the emotion tendency through the viewpoint words. Based on HowNet, yan LAN zhu et al. (2006) respectively used semantic similarity and semantic correlation field in dictionaries to calculate the semantic tendency of words. Carenini G, Ng R T, Zwart e. 9(2005) used word similarity to categorize the attributes of resources in specific fields, and calculated the affective tendency of categories based on categories. Yu J X, Zha Z J, and Wang M (2011) constructed the hierarchical tree of resource attribute concept with the help of existing resource attribute hierarchical tree and resource comment data, which improved the accuracy of feature classification to some extent. With the development of ontology technology, scholars began to try to apply ontology to emotion research. The mainstream thought is to extract features according to ontology and calculate the emotional tendency of features. Yin P, Wang H, Guo k. (2013) constructed a resource-oriented domain ontology, combined with the emotional dictionary, and realized the extraction of feature view pairs and emotional calculation in Chinese comments. Based on the domain ontology and emotion dictionary, li jinhai et al. (2016) realized the emotional intensity analysis from the text and sentence of online comments to the level of resource attributes.

Although scholars at home and abroad for network comments sentiment analysis research has made some achievements, but the study of sentiment analysis both in the document class levels, sentences, words, level, did not consider the properties between the hierarchical and subordinate relationship, combined with the ontology sentiment analysis only to extract features or characteristics of emotional words, emotional calculation for the attributes are often only for a single attribute itself or from the general property is divided into several categories, ignoring the logic relation between a single attribute, Characteristics of this kind of single particle size classification cannot meet the learners on the same resources to focus on the local characteristics of different even the same attribute existing bigger difference, and ignores the different class or same attribute classes under different child the importance of the attributes of the upper attribute is different, lead to emotional analysis of accuracy and practicality is not high.

Network comment statements are random and random, and many attributes or attribute characteristic words often appear at the same time. It is difficult to get a detailed and in-depth grasp of resources by calculating emotional tendency through general attribute classification. Only by unifying the normative concept of each attribute and understanding the subordination of each attribute word can we get an accurate attribute comment tendency. At the same time, the research paradigm of emotion analysis and evaluation mode of online evaluation corpus driven by big data is constructed. Through studying influence mechanism and factor analysis, the theoretical research of corpus is further improved and a systematic research framework is formed. Through collecting massive data, analyzing messy data and the correlation between various factors, establish a data-oriented online comment emotion analysis system, combine corpus research with computer science, and achieve the upgrading of corpus research technology. On the basis of a large number of data analysis, the online emotional evaluation dictionary is established. This paper proposes a multi-level network resources based on domain ontology word-of-mouth information granularity emotional mining method, construct domain ontology to reveal the hierarchical relationships between resource's attributes and dependencies between

attributes and attribute characteristics, on the basis of the characteristics of various attributes and the emotional analysis, further enhance the accuracy of information mining, on the one hand, help to the subsequent resource details improvement and user preferences recommend; On the other hand, it provides reference for learners to make resource selection decisions.

III. ANALYSIS

Corpus development tends to apply big data in language services, application technology of data include but not limited to the following aspects: based on big data visualization analysis, based on the predictions of a large data, based on large data of business transactions, which based on the analysis of large data visualization and business transactions can be combined with corpus construction and research. So-called big data visualization analysis refers to a large number of abstract data into the visual representation, make readers intuitively grasp the spatial distribution pattern, the trend of the data, such as correlation description and inference statistics, and the statistics may be present mode is difficult to be found, in other word cloud based on word frequency analysis (word cloud) is one of the classic representative data visualization. The future development of corpus will not only focus on database construction or analysis, but also pay more attention to the presentation of analysis results, which should not be satisfied with the mechanical description such as index positioning, but present the statistical information of multi-modal data and the interaction between data in a more humane way. Big data transactions include such services as corpus or translation memory database transactions and multilingual data processing services. As a language resource in a narrow sense, corpus not only has academic value in language research, but also has commercial value in natural language processing, such as lexicography, machine translation and software development. Therefore, the finished product and various links in the development process can also provide business opportunities for big data trading. Corpus development will upgrade corpus technology by means of corpus cleaning, labeling and alignment.

The structure of the paper is as follows:

Data collection: Get the latest Internet data through web crawlers. Data processing mainly preprocesses data to reduce data volume. Through the study of relevant data cleaning, data reduction, data mining and other methods, remove the redundancy to get clean, tidy, valuable data. Emotional assessment is mainly to evaluate the emotional grade of online comment corpus. The domain ontology is constructed to reveal the hierarchical relationship between resource attributes and the dependency relationship between attributes and attribute characteristics. On this basis, the emotional analysis of each attribute and attribute characteristics is carried out.

Emotional rating: Combined with the emotional labeling results of corpus, the means of emotional expression were analyzed and the emotional evaluation was carried out. Through the analysis of various factors that affect learners' choice, and the analysis of the mutual relations among various factors by using the principle of interdependence, the domain ontology is constructed to reveal the hierarchical relations among various attributes of resources and the subordination relations between attributes and attribute characteristics. On this basis, the emotional evaluation of various attributes and attribute characteristics is carried out.

Corpus construction and emotion dictionary construction: According to the emotional rating level, cross-cultural learning resources are recommended, the emotional corpus is expanded, and the static and dynamic emotional words dictionary is built online. The static and dynamic emotional words dictionary includes special emotional dictionary, emoji dictionary and so on. Dynamic affective dictionary includes lexicon of modifiers and lexicon of special affective elements, enabling learners to quickly find satisfactory learning resources according to their own preference factors.

In addition to displaying the source of resource content, the source of information is also the embodiment of authority, professionalism, credibility and communicator's communication intention. This study selects overseas social networks and British and American media resources as research objects, such as the information of major newspapers and TV stations in Britain and America. The information of traditional media has the advantages of authenticity, authority and easy availability, etc., and such contents with high credibility are also trusted by learners, whose retweets and comments rank among the top among all kinds of information sources. Secondly, social network is only the second largest source of information, but communication power ranks the first among all kinds of information sources. Through reloading popular overseas social network contents through domestic social networks, it fills the structural hole between cross-cultural communication user groups at home and abroad and becomes the bridge of information sharing in domestic social networks. This communication path includes the dual choice of content by social network users and information bloggers, which to some extent makes up for the information fragmentation and uneven quality of the social network as a source of information, and also achieves high communication power by meeting the information needs of cross-cultural learners.

Getting big data information through web crawlers: Web Crawler (spider Web Crawler, also known as the network, the network robot) is a kind of request site and extract data automation procedures, initial Web crawlers from several of the uniform resource locator (URL) to obtain the initial Web site URL, the Web page text, images, video, etc., in the process of continuously new URL can be drawn from the current page in the queue, until meet system must stop condition, so as to realize batch data acquisition. 20,000 texts were randomly obtained from foreign media and websites, and the time interval was set as January 1, 2017 solstice May 1, 2019, which was used for the expansion of emotional dictionary. Text cleaning and sentence division are carried out on corpus. Some corpus resources are selected as corpus for emotional classification training, and emotional labeling is carried out on this part of text. According to different text granularity, emotion classification can be divided into word

level, phrase level, sentence level and text level. Due to the relatively complete semantic and emotional expression of sentences, this paper mainly identifies sentence-level emotions. Next, we preprocess the corpus used for emotional dictionary expansion.

Data preprocessing includes three parts: data cleaning, data integration and data selection. In the era of big data, data preprocessing is indispensable. Data preprocessing technology is used before data mining, which greatly improves the quality of data mining patterns and reduces the time needed for actual mining. Relatively tidy data can be obtained through data preprocessing. On the basis of these data analysis, further mining hidden behind the data potential, valuable information. Online evaluation language description is not normative, and the evaluation text is relatively short. It usually consists of incomplete, noisy, poorly structured sentences, irregular expressions, imperfect words, and non-dictionary terms. In order to make up the word count, some users will randomly write some information such as punctuation marks irrelevant to the content of the material. Such evaluation information is noise.

Before feature selection, a series of preprocessing is applied. Firstly, score, user ID and other information in the evaluation text are removed, evaluation information is retained, stop words are removed, abbreviations are extended, URL is removed, and negation is replaced to reduce noise amount. Text preprocessing can reduce the noise in the text, improve the performance of the classifier and speed up the classification process. In order to identify the emotional polarity in online comments, the existing methods are applied to text preprocessing. After preprocessing, the accuracy of emotional characteristics can be significantly improved by appropriate characteristics and representations. After preprocessing, corpus vocabulary decreases. When negative transformation and repetition normalization were used, affective classification accuracy increased. For evaluation information, negative evaluation is generally in the negative form. If words are used as the features of text information, ambiguity will appear. Based on this, combining these words to form new adjective phrases and adding them to the custom dictionary can eliminate most ambiguity problems. After precise word segmentation, stop words are removed according to the stop words list and part of speech, which can reduce the dimension of text information. Meanwhile, synonyms are normalized according to thesaurus, and the thesaurus is transformed into a specific word according to the mapping relationship of thesaurus.

Specific pre-processing contents include: Filtering user names: the existence of user names sometimes has a great impact on the judgment of emotions in the process of multi-affective classification. Therefore, user names in the form of "/ /@ user name" are needed for filtering. "#topic#": in the multi-affective classification, the existence of topics may also bring errors in the classification of emotions, and the emotional tendency of topics will affect the emotions expressed in sentences. Filtering urls in the text: this study categorizes sentences in the body emotionally, so link strings like "HTTP:/ /" need to be removed. Remove ratings, stop words, expand abbreviations, and replace negatives from the evaluation text. Perform regular matching according to the punctuation mark at the end of the sentence to complete sentence segmentation, such as using "...", "!", ".", ";", and other symbols for sentence segmentation.

Dimension reduction of data. To deal with the dimensional disaster, you need to dimensionalize the data. Dimensionality reduction essentially selects some features from the data set of a given value dimension. The core idea of principal component analysis (PCA) is to reduce the dimension of data sets composed of a large number of related variables, while preserving as many variables as possible in the data set. This is done by converting the original data into a new set of variables, the principal components, which are unrelated and ordered so that the first few eigenvectors retain most of the changes in all the original variables. In order to reduce the dimension of the original feature space, PCA is used to find the projection direction, which is the most effective representation of the minimum mean square error on the original data. Principal component analysis (PCA) was used to reduce the dimension of the data set to obtain the characteristic vector of the data. For any sample data, project the sample data onto the eigenvector, and get the projection coefficient as the characteristic representation of the sample. Support vector machine (SVM) is used to classify these different projection coefficient vectors for classification and recognition. Principal component features are obtained through mapreduce sorting, and the final classification is realized by SVM.

Online comment on emotional impact factor analysis Resource content. Dissemination content is the information entity to achieve the purpose of dissemination. As a kind of intercultural communication practice, intercultural learning resources successfully realize the function of intercultural communication by transmitting cultural information. Cross-cultural learning resources include news events, life anecdotes, culture, science and technology, study abroad, education, tourism, consumption, media (newspapers, television, etc.), enterprises and institutions and other issues. Cross-cultural learning resources can be divided into five types of issues. "Life/interesting stories", "news/events" and "culture and technology" related to Britain have a large proportion. Some resources focus on the delivery of new news to users, fun, with the topic of daily, grass-roots characteristics. It is helpful to improve the average amount of retweets and comments of cross-cultural communication learning resources by fully considering users' reading needs and interest setting topic types. It will also shorten the psychological distance of participants in cross-cultural communication and optimize the communication effect. The most retweeted and commented are "life/interesting news" and "news/events", and the number of other topics is positively correlated with the activity of user feedback.

Narrative mode: The narrative mode mainly includes three basic forms: graphic combination, short-chain link (i.e. video, music or web link) and pure text. Some learners prefer graphic combination, while some prefer video resources or pure text. Display information in the form of graphic

combination or graphic integration. While making the content interesting, they also have strong narrative expressive force. This form can be used to express the language and thinking mode of cross-cultural communication information. The content of text and pictures combined not only greatly expands the capacity of the article and overcomes the disadvantage of information fragmentation, but also is suitable for the supplement of cultural differences and backgrounds and the explanation of the context of events in cross-cultural communication.

Narrative perspective: The selection of content is concise, interesting, and easy for users to understand and interact with information. It makes cross-cultural communicators have a sense of participation with the same identity and vision, and strengthens the communicators' enthusiasm for the discussion and dissemination of articles. For example, the information of British universities and students' life is released from the perspective of overseas students, which not only builds a picture of British youth's life for young users living in China, shortens the psychological distance, but also provides practical information and information. This narrative mode also further promotes the formation and development of intercultural social network communication mode.

Interaction frequency: Interaction types can be divided into three types: "retweeted and commented by fans", "responding to fans' comments" and "retweeting other website resources". Firstly, user forwarding is the most core link in the communication chain. Users' active forwarding enables the resource content to spread through multilevel transmission nodes, covering a wider range of users and smoothly contributing to the accumulation of popularity and expansion of influence. Secondly, the behavior of comments is the most intuitive intentional feedback of users to the content of resources. The growing number of comments gives publishers the opportunity to receive more information from users and refine the content based on the opinions of user groups such as some big V, such as "British tourism", "I was shocked at that time", and related to British cultural information.

Affective categorization refers to mapping the text information to be classified into the affective categorization system defined in advance. According to the different classification system of emotion, emotion classification can be divided into binary classification (subjective, objective), triple classification (positive, negative, neutral), or multivariate classification (that is, text is divided into more detailed categories of emotion). Among them, multivariate classification is able to classify the emotions expressed by users in a more detailed way. Compared with binary and ternary classification, it is more close to the description of human real emotions, which has increasingly attracted the attention of researchers. At present, the research is basically based on binary or ternary classification, which has certain limitations on the restoration of human real emotions. At the same time, in the research process, most of the existing emotional dictionaries are used, and there are few cases of expanding emotional dictionaries for online comments on cross-cultural communication resources. The scale of existing emotion dictionaries is often small, and the online comments tend to be colloquial, which makes the existing emotion dictionaries unable to accurately cover as many emotional phenomena as possible, thus reducing the performance of emotion classification.

Resource attributes based on domain ontology conceptual space model Ontology conceptual space model construction based on resource-oriented online comment information. The definition of ontology most cited by people is proposed by Gruber, which is a description of the standardization and clarity of conceptual system, so as to realize knowledge sharing and common understanding. To apply domain ontology to network text mining, first of all, a conceptual space model of domain ontology needs to be built to provide accurate and standard interpretation for domain attribute expression, establish logical connections between objects and attribute layers, standardize the commonness and opposability of each attribute node, and automatically realize multi-level information organization. The construction process of domain ontology conceptual space model mainly includes three parts: acquisition of domain core concepts, determination of classes and relations between concepts, and construction of ontology. This paper takes cross-cultural learning resources as an example to elaborate.

Acquisition of domain core concepts: Online comment information is the exchange of learners' opinions on resources, which reflects the resource attributes concerned by learners and contains important resource information. Through the collation and classification of cross-cultural communication resources on the websites of major foreign media and cultural institutions, the initial and main conceptual attribute sets of resources are obtained. At the same time, grab the comment text online, after preprocessing of the online text expression pattern analysis, using regular expressions to extract the noun, verb + noun, verb + adjective, noun + noun form of word combinations, as candidate resource attribute sets, filtered to get word set, and the initial concept and attribute set final core concept set of resources.

The determination of classes and the acquisition of relationships between concepts: According to the extracted core concept set of resources, the class classification of resources is carried out, and the attributes of the class are defined. First, the core concept set of resources is classified, and topic attribute words of each category of attribute sets are extracted, which are regarded as the highest level (level 1) attribute directly subordinate to entity and at the level of ontology. Secondly, further classify each class of attribute set, determine the relevant subclass of each class, extract subject attribute, as the subclass attribute under the corresponding category attribute. And so on, from top to bottom, until there is an attribute feature that is no longer classifiable, the lowest level.

Resource attribute extraction based on domain ontology conceptual space model explicit properties. As for the extraction of explicit attributes, domain ontology can be mapped directly. Domain ontology concept space model with rich semantic relationship and the level of the complete

system, and provides the concepts and relationships in the field of vocabulary, introducing ontology knowledge and method can better extraction of text feature vector, online text after pretreatment with ontology matching and transform to detect the word for the concept of ontology, properties, extraction research field contains the words may be synonymous with ontology concept of words, concepts, or under a word, need to be standardized.

Implicit properties: Due to the casual and informal nature of network comments, some online texts do not have explicit comment objects, which need to be obtained based on semantic analysis and inference. Ontology-based matching cannot identify such implicit attributes. For the extraction of implicit attributes, the existing researches mainly map resource characteristics through the fixed collocation relationship between attribute words and emotional words.

Hierarchical division and classification of resource attribute sets. Resource attributes extracted according to domain ontology and implicit attribute strategy containing various aspects of the resource, and there is a logical relationship between these attributes. According to the whole and part, subclass and parent relationship in ontology, these attributes are classified and hierarchical, mapped to different attribute levels, the position of each attribute in ontology is defined, and its upper and lower levels are determined. Finally, the attribute hierarchy structure is constructed from top to bottom.

Multi-level and fine-grained analysis of affective tendency of attribute characteristics: Multi-level and fine-grained mining of emotional information. The overall emotional analysis of the comments on cross-cultural communication resources reflects learners' satisfaction with the resources as a whole, only reflects the general opinions of learners, and cannot better reflect learners' in-depth views on the characteristics of products from the whole to various attribute categories. Different attribute classes have different attribute characteristics, which are the most detailed characteristics. Fine-grained analysis of different attribute characteristics' emotional polarity under the same attribute class has important practical significance for cross-cultural communication and communication. Through the logical space domain ontology model, to comment on attribute mapping for different levels of concept, clear the dependencies between attributes, attribute hierarchical model construction, starting from the multi-level and logical relevance, learners are calculated respectively on the whole, properties and characteristics of each class of attribute emotional tendencies, to achieve different levels of each attribute fine-grained precise emotional mining.

Emotion classification system was designed, and labeling design was made for the basic lexical means of emotional expression (noun, verb, adjective and phrase collocation). The labeling system of this study was built based on TEL (Text Encoding Initiative), an international information Encoding scheme commonly used in the labeling of large corpus. Then mark the collected web comment materials based on the marking system.

The feature classification method based on support vector machine can improve the accuracy of judgment of online comments. Emotional feature items are the collocation of nouns, adjectives, verbs and phrases marked with emotional colors in online commentary corpus. The frequency of different emotional categories in the text is taken as the weight of features by means of emotional intensity of emotional expression, such as comparative degree, degree adverb on the impact of emotional expression.

Through the data statistics analysis of the use of emotional expression means in the network review characteristics. First of all, the classification of inspection in the comments of emotion, emotional expression means (nouns, verbs, adjectives, adverbs and phrases) statistical characteristics, and analysis the influence factors of other means of emotional expression, such as other parts of speech, negation, punctuation, emoticons, onomatopoeia expression, etc., found that the emotional evaluation of language represent characteristics, differences and universal law. This paper studies the use of irony, metaphor, metaphor and other rhetorical devices in corpus to further accurately judge the emotional tendency of text.

IV. DISCUSSION

The construction of emotion dictionary and the expansion of corpus: Different emotional words have different influences on emotional intensity and polarity. The field dictionaries constructed include emotional evaluation lexicon, degree adverb, negative lexicon and other modifiers. In addition, in e-commerce online platform, there are many pictures and symbols that appear together with the text. The appearance of ",", ".", etc. will change the meaning of the text. Therefore, pictures and symbols are also introduced into the emotional analysis to build a dictionary of special emotional elements. Compared with the vast text containing many colloquial words and emoticons, the existing emotion ontology library is still not large enough to accurately realize the multi-emotion classification of text. Emoticons are commonly used in online comments to express users' emotions instead of words. All the emoticons in the corpus are classified and prioritized to determine the emoticons that can represent emotions and mark the corresponding relationship between emotional categories and emoticons. Sentences containing only specific emojis are extracted from corpus and word segmentation is carried out. Words with high sentence coverage and words with synonyms in corpus are added to the emotional dictionary to realize the expansion of the emotional dictionary.

Identify emotional words: To extract the sentences of the designated emoticons from the corpus text that has been preprocessed by text and sentence segmentation, these sentences are considered to have obvious specific emotions. These sentences are participified, stop words are removed, and the remaining words are used as candidate emotional words. Select the first k words that cover 90% of the specified emotion sentences, and add them to the specified emotion in the ontology library of the emotion vocabulary. Rank the words related to a certain kind of emotion according to the degree of relevance. Then, the coverage of

the ordered words to this kind of emotional sentences is investigated, and the words with high coverage are extracted successively. According to this process, the selected words not only have a higher relevance to this kind of emotion, but also have a higher coverage of this kind of emotional sentence. These words can be added into the emotional dictionary as an effective representation of this kind of emotion.

Identify synonyms for emotional words: The synonyms of emotional words are generally consistent with the meaning and emotion expressed by the emotional words. The identified emotional words are searched for their corresponding synonyms in the forest of synonyms. The synonym forest classifies words and marks each word with a code in front of it. The last bit of code "=" means equal and synonymous; "#" means unequal but homogeneous; "@" stands for independent and closed, and there are no synonyms or related words in the dictionary. Therefore, only a group of words whose last digit is "=" are selected and added to the emotional lexicon.

Static emotional dictionary: The emotional evaluation dictionary of this paper is constructed in combination with the field expertise. The specific situation is as follows:

First, remove emotional words that do not conform to attribute modification, such as rare words.

Second, extract emotion evaluation words based on online comment corpus, pay special attention to emotion words in the field.

Third, collect new Internet words. Due to the dynamic and diverse nature of Internet words, new Internet words emerge in an endless stream. Although the words are not standard, they are known and widely used by the public. The sentences of special symbols and emoticons are mainly short sentences. Users often add emoticons and special punctuation marks in the sentences to express their emotions, such as "[haha]", which means the expression of a smiling face; "! ! !" Express surprise or anger. In addition, users often USES the popular performing words express emotions, such as "o (* ≧ del ≦) ツ" on behalf of the smiling face, etc. Screening out these expressions as features is important for emotional classification.

The features of negative words in sentences often play a turning role in the negative words in the expression of emotion will lead to the reversal of emotion. The occurrence of "not" turns the expression of emotion in the whole sentence from positive to negative. Select common negative words to construct a negative dictionary, including "not", "no", "can't", "none", "seldom" and so on. Next, on the basis of the extracted text features, the multi-classification of text sentence-level emotion is realized. The static emotion evaluation dictionary was obtained by combining and deweighting the emotion evaluation words collected from the above methods.

Modifiers play a key role in the influence of emotional polarity, and the polarity expressed by the same emotional evaluator with different modifiers will also change. For example, the polarity expressed by the emotional evaluator "beautiful" is completely opposite to that expressed by "not beautiful". Combining the characteristics of network popular terms and network comments, the paper constructs the degree adverb word list and negative word list, adjusts the degree adverb's emotional polarity, gives each modifier an emotional intensity value, and divides it into different grades.

Dictionary of special emotional elements: When making online comments, users often attach pictures related to the comments, which is also a way to convey information. The data analysis and knowledge discovery research papers of pictures make the text more vivid and vivid, and intensify the expression and understanding of emotions. In addition, some symbols existing in sentences will also have an impact on the polarity of emotions. For example, for emotional words marked with double quotation marks (" "), the original meaning is generally expressed in reverse. Exclamation mark (!) Often have the effect of aggravating emotional coloring. Therefore, images and symbols are considered as a special emotional element in the process of emotional calculation, and the processing rules of this kind of emotional element are summarized by analyzing the text.

Emotional element processing rules: If pictures are used to assist the expression of critical text, the emotional tendency of the attributes expressed in the text should be adjusted. Depending on the semantic situation, it is possible to keep the semantic unchanged and reverse it. If the emotion word has the double quotation mark, the emotion polarity processing reverse. If there is an exclamation mark in the comment text, the emotional tendency expressed will be aggravated to the same degree as the picture, and the total emotional tendency will be multiplied by 1.25 times.

Multi-level and fine-grained calculation of affective tendency: According to this article to build the domain ontology, domain dictionary, evaluate the comments sentence after pretreatment unit mapping, including properties, negative number, the matching of degree adverbs, such as emotional words recognition, for the default sentences according to the implicit attribute identification, identify hidden attributes, for each attribute is stored as {attribute words, emotional evaluation, negative, adverbs of degree, special emotional elements} the attributes of emotional units, and with the help of ontology semantic relations between the various properties and hierarchical structure, according to the field of emotion dictionary set by emotional value, the level of emotional tendency to attribute values. In the hierarchical model, attributes no longer exist independently among each other, and the whole contains parts, and the child nodes are subordinate to the parent nodes, and their emotional tendency influences each other, and the importance of different child attributes varies, and the emotional influence on the parent attribute is different. Therefore, this paper starts from a single attribute node and assigns different emotional weights to sub-attributes at different levels. For the emotional tendency of a single attribute class at a certain level, its emotional tendency is the sum of its own independent emotional tendency value and the emotional tendency value of its sub-nodes. Through the hierarchical assessment of the emotional tendency of

attributes, the feelings of the lower attributes are transferred to the upper layer, and each layer is optimized separately, which improves the classification accuracy. In addition, the emotional dictionary in this paper is relatively perfect. It not only adds dynamic emotional dictionary, taking into account the influence of different situations on emotional polarity, but also considers the influence of pictures, symbols and other special emotional elements on emotional results, which further improves the accuracy of emotional calculation.

V. CONCLUSION AND SUGGESTION

Aiming at the one-dimension and fuzziness of single-level emotion analysis method, this paper proposes a multi-level and fine-grained emotion analysis method for online comments based on domain ontology. The constructed domain ontology is used to realize the extraction and hierarchical division of attributes, clarify the local and overall logical relations among attributes, and analyze the emotional tendency of attributes at different levels, effectively enhancing the accuracy of emotional analysis. The construction of the attribute hierarchy model in this paper depends on the domain ontology constructed, and the completeness and accuracy of ontology construction directly affect the research results. Due to the universality and dynamics of ontology knowledge, the efficiency, standardization and comprehensiveness of ontology construction still need to be further improved. At the same time, in the aspect of attribute emotion evaluation, the setting of attribute emotion weight can be further discussed and improved.

ACKNOWLEDGMENT

This paper belongs to the project of the Fund Project. Fund Project Type: "Fujian Jiangxia University Education and Teaching Reform Project"; Fund project number: J2018B003; Fund Project Name: Exploration on the Construction and Teaching Mode of College English Learning Website Based on MVC Framework.

This paper also belongs to the project of the Fund Project. Fund Project Type: "Fujian Young and Middle-aged Teacher Education Research Project (Social Science Class B)"; Fund project number: JBS14221; Fund Project Name: Research on cross-cultural communication application teaching.

This paper also belongs to the project of the Fund Project. Fund Project Type:“Fujian Young and Middle-aged Teacher Education Research Project”; Fund project number: JAT170624; Fund Project Name: Research on Decision-making Method and Application of Ciic set.

REFERENCES

[1] Johansson, S. 2007. Seeing through Multilingual Corpora: On the Use of Corpora in Contrastive Studies [M]. Amsterdam/Philadelphia: John Benjamins Publishing Company, 1.

[2] Johansson, S. 2007. Seeing through Multilingual Corpora: On the Use of Corpora in Contrastive Studies [M]. Amster dam/Philadelphia: John Benjamins Publishing Company, 1.Leech, G. 2014. The Pragmatics of Politeness [M] . Oxford: Oxford University Press.

[3] Yu H, Hatzivassiloglou V. Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences [C]//Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing. 2003: 129-136.

[4] Yu H, Hatzivassiloglou V. Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences [C]//Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing. 2003: 129-136.

[5] Pang B, Lee L. Opinion Mining and Sentiment Analysis [J]. Foundations and Trends in Information Retrieval, 2008, 2(1-2): 1-135.

[6] Turney P D. Thumbs up or Thumbs down?: Semantic Orientation Applied to Unsupervised Classification of Reviews [C]//Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics, 2002: 417-424.

[7] He R, Gonzalez H. Numerical Synthesis of Pontryagin Optimal Control Minimizers Using Sampling-Based Methods [C]//Proceedings of the IEEE 56th Annual Conference on Decision and Control (CDC). Melbourne, Australia: IEEE CDC, 2017:733-738.

[8] Meena A, Prabhakar T V. Sentence Level Sentiment Analysis in the Presence of Conjuncts Using Linguistic Analysis [C]// Proceedings of the European Conference on Information Retrieval. 2007: 573-580.

[9] Zhang Chenggong, Liu Peiyu, Zhu Zhenfang, et al. A Sentiment Analysis Method Based on a Polarity Lexion [J]. Journal of Shandong University: Natural Science, 2012, 47(3): 47-50.

[10] Fu X, Liu G, Guo Y, et al. Multi-aspect Sentiment Analysis for Chinese Online Social Reviews Based on Topic Modeling and HowNet Lexicon [J]. Knowledge Based Systems, 2013, 37: 186-195.

[11] Kim S M, Hovy E. Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text [C]// Proceedings of the Workshop on Sentiment & Subjectivity in Text at the International Conference on Computational Linguistics/the Annual Meeting of the Association for Computational Linguistics Sentiment and Subject. 2006: 101-108.

[12] Carenini G, Ng R T, Zwart E. Extracting Knowledge from Evaluative Text [C]//Proceedings of the 3rd International Conference on Knowledge Capture. Edmonton: ACM, 2005: 11-18.

[13] Yu J X, Zha Z J, Wang M, et al. Domain-Assisted Product Aspect Hierarchy Generation: Towards Hierarchical Organization of Unstructured Consumer Reviews [C]// Proceedings of 2011 Conference on Empirical Methods in Natural Language Processing. Edinburgh: ACL, 2011: 140-150.

[14] Yin P, Wang H, Guo K. Feature-Opinion Pair Identification of Product Reviews in Chinese: A Domain Ontology Modeling Method [J]. New Review of Hypermedia and Multimedia, 2013, 19(1): 3-24.