Research on Chinese Text Sentiment Classification Process

Ying Yu [a], Yanjun Yin [b]

College of computer science and technology Inner Mongolia normal university Hohhot, China

[a] 2683558457@qq.com, [b] ciecyyj@163.com

Abstract. In recent years, with the rapid development of computers and social networks, online text information such as product reviews, news events and articles have grown exponentially, and automatic emotional classification of text information has become a research hotspot. This paper describes the current research status of Chinese text sentiment classification, namely preprocessing, feature extraction and sentiment classification, and expounds the problems existing in Chinese text sentiment classification and future research directions.

Keywords: Chinese text; sentiment classification; classification process.

1. Introduction

With the rapid development of the Internet, especially the rise of social platforms such as microblog and forums, the Internet has become an important way for people to obtain information and create information. A large number of web users publish and disseminate massive amounts of information every day, and a large part of them (published data) express the user's opinion and emotional feelings about the events described. Judging the emotional bias of the published text not only can understand the subjective attitude and emotional feelings of the text publisher on the published content, but also the analysis of the text sentiment bias can also be used for information prediction, product word-of-mouth judgment and public opinion monitoring. For example, if a consumer comments on a purchased product on a large website such as Taobao, other potential users can refer to the comment as a basis for whether or not to purchase the product. More and more individual users and organizations use the emotional information on the network as a reference standard for decision-making, and the current subjective text information is growing exponentially every day. Therefore, the treatment of the emotions expressed by the network text information has become a research hotspot in the current academic circles.

Text sentiment classification is the basis of information filtering and information retrieval. Many scholars have carried out research on it. In this paper, the concept and process of text sentiment classification are briefly summarized. Then the domestic and international research status and future research directions of preprocessing, feature extraction and sentiment classification are elaborated and analyzed.

2. Introduction to Sentiment Classification

Emotional classification is the core of sentiment analysis technology, also known as sentiment orientation analysis. It refers to the process of excavating and analyzing subjective texts with emotions, and dividing the text into positive, neutral and negative through statistics, induction and reasoning, and it is a hot topic in the field of sentiment analysis.

Traditional text categorization, people pay more attention to how to deal with the objective content in the text. For example, the text is classified according to different themes such as politics, military, economy, and culture, etc.; or extracting specific events in the text by keywords. Different from objective texts, subjective texts are authors' views or thoughts about various things, with the author's affective tendencies. Compared to traditional text categorization, sentiment classification is to get information from the text whether it supports a certain viewpoint. It involves more complex emotional expressions, generally directed at subjective texts with emotional tendencies. At present, text sentiment classification is widely used in film reviews, product quality and service evaluation, public opinion analysis, news report review, book recommendation, stock review, and information prediction.

Throughout the current research work on emotional classification of subjective texts, the classification process generally includes the following steps: firstly, the text is subjected to pre-processing such as word segmentation and de-stopping words, and the text is normalized. Secondly, the text is vectorized. It is expressed as a language that the computer can understand; then select and extract features from the text, and reduce the time, space efficiency and accuracy of subsequent experiments by reducing the dimensions; finally selecting the classification algorithm for emotional classification. Among them, there are three main research ideas for classification methods: classification methods based on sentiment lexicon, classification methods based on machine learning, and classification methods based on deep learning.

3. Text Preprocessing

Text preprocessing is mainly to unify the format of the document for subsequent work, mainly including document segmentation, word segmentation, de-stopping words, text representation, and so on.

3.1 Document Segmentation

Document segmentation is optional when categorizing, depending on the form of the obtained document collection and the classification requirements. For example, if the resulting collection of documents itself is an article that is separate from each other, then this step can be omitted directly; On the other hand, if they obtained document collection is a single file, all the articles are stored in this file, so that the document segmentation is needed. And it is to extract the files and store them in separate files for subsequent experiments. Generally speaking, in a document collection of a single file, some tags are used to distinguish between different articles, such as blank lines or other specific symbols.

3.2 Word Segmentation

For word segmentation techniques, Chinese is different from English. There is a space between the words of the English text as the boundary mark of the word, and there is no obvious separation between the Chinese words. Therefore, if the word is the feature object in the text sentiment analysis, the word segmentation is indispensable. And the word segmentation result has a great influence on the subsequent emotion classification. So, a fast and accurate word segmentation algorithm is very important. There are three main methods of word segmentation: string matching method, semantic analysis method and statistical method. In 2011, in order to solve the problem that the maximum word length of the maximum matching word segmentation algorithm is fixed and the word loss and the number of matches are too many, Ruixue Wang et al. [1] designed the dictionary structure to dynamically intercept the text length to reduce the number of matches and improve the speed and efficiency of word segmentation. Jun Zhou et al. [2] conducted Chinese word segmentation based on the principle of longest generalized word matching. They introduced generalized terms and induced word sets for word ambiguity recognition. In recent years, with the development of technology, some word segmentation tools such as Paoding Analyzer, IK Analyzer, Jieba participle and ICTCLAS have become popular.

3.3 De-Stopping Words

Stop words are words or characters that appear frequently but don’t have much effect on text classification. There are many stop words in English, Chinese or other languages. For example, in the English "the, of, and, for", these words appear in almost every English text, but there is less contribution to the emotion or content to be expressed in the text. Therefore, it is necessary to filter out these stop words in the original text. This process is called stop word filtering or stop words. De-stopping words involve the construction of the stop word list. In practical applications, the words appearing in the stop word list can be directly filtered. The method is simple and convenient for text classification.

3.4 Text Representation

Before you categorize your text, you need to convert the text into a format that the computer can handle. Most of the traditional sentiment classification methods use the Bag of Words (BOW), the Term Frequencv-inverse Document Frequency (TF-IDF), etc. [3] for text representation. Although these traditional text representation models are simple and convenient, they often have the following shortcomings:

1) Missing order information of words.
2) neglecting the semantic information of words.
3) The dimension of data is sparse.

Currently, most of the research on text classification is based on the Vector Space Model (VSM) [4]. The Vector Space Model achieves better results when dealing with long texts [5], but for the sparsity and the irregularities of short texts, this method will reduce the classification accuracy accordingly.

These methods usually require a large amount of external corpus, which increases the computational overhead and affects the classification accuracy. Therefore, the distributed vector representation of words has become a research hotspot in recent years, which can map words into high-dimensional space and can contain semantic information of words. The distributed representation of words in the deep neurolinguistic model was first proposed by Rumelhart et al. [6] in 1986, and Bengio et al. [7] used neural networks to train distributed vector representations of words in 2006. The Skip-gram model proposed by Mikolov et al. [9] can effectively learn high-quality distributed vector representations of words, and the vector representation can capture the grammatical and semantic relations between the words. Collobert et al. [8] applied deep learning techniques to the study of natural language processing earliest. In 2013, Google introduced Word2vec [10], a tool for training word vectors, which provides a way to represent text using distributed vectors. Compared with the traditional text vector space model, using the word2vec model to represent text can not only solve the high-dimensional sparse feature problem of the traditional vector space model, but also introduce the semantic features that the traditional model doesn't have, which is helpful for short text classification [11].

4. Feature Extraction

Feature extraction is to reduce the feature vector spatial dimension by selecting relevant initial feature subsets from the initial feature set of the test set according to certain feature extraction metrics. In the text, feature extraction performance is an important measure of text classification results, and the quality of feature selection can affect the accuracy of the final classification results.

The dimensionality and sparsity of the vector matrix formed by the text after preprocessing and representation are high, resulting in the computer being time consuming and less accurate in the learning and training process. In order to further achieve dimensionality reduction, it is necessary to select and extract text features. Zhongyang Xiong et al. [12] applied the frequency, concentration and dispersion of words to the statistical method of χ^2, and verified the validity and feasibility of the method through experimental comparison. In 2013, Tongsen Du et al. [13] proposed a desired cross-entropy algorithm based on inter-class concentration and intra-class dispersion, which organically combines the uniformity of feature items between the same classes and the different classes. The result shows that the algorithm has better feature selection effect in text classification. In the same year, Xueli Fan et al. [14] proposed a principal component analysis feature selection algorithm based on mutual information. Wei Dong et al. [15] proposed an adaptive method for the problem of information gain feature selection in text classification (Proportional correlation of positive correlation features and negative correlation features), which leads to the degradation of classification accuracy. This method shows a good classification effect on both the balanced data set and the unbalanced data set.

In the 1970s, Salton et al. proposed the Vector Space Model. Yuehua Tao et al. [16] introduced the representation of document vector, query vector, and the meaning of index word-document matrix.

On this basis, the feature extraction method of index formula weights is discussed. Xiaojun Yu et al. [17] proposed an improved N-Gram text feature extraction algorithm in 2012. The experiment shows that the algorithm can be applied to Chinese information processing such as text feature processing and data mining. Xiaobin Peng [18] used word2vec to vectorize the text, combined with the SVM classifier in SGD Classifier to conduct experiments. Compared with the combination of word bag model and classifier, the accuracy rate is improved. Qian Zhang [19] proposed a short text classification of microblog based on the word2vec model for the problem of feature dimension catastrophe and non-semantic features of traditional text classification model, and introduced TF-IDF to weight the word2vec word vector. The experimental result shows that the combined model classification's accuracy is higher than the weighted word2vec model and the traditional text classification model using TF-IDF. The researchers also used word2vec for the classification of emotions such as Internet goods and hotel reviews [20, 21], verifying the validity and accuracy of the method in emotional classification.

5. Classification Algorithm

At present, according to domestic and foreign research, the methods of text sentiment classification can be summarized into three categories: text sentiment classification based on sentiment dictionary, text sentiment classification based on machine learning and text sentiment classification based on deep learning.

5.1 Emotional Dictionary based Text Sentiment Classification

The sentiment classification method based on sentiment dictionary is the simplest simulation of human memory and judgment thinking. Emotional dictionary is the basis of text sentiment classification. With high-quality sentiment dictionary, the actual application system can take a simple and fast method to obtain better results. The model is shown in Figure 1.

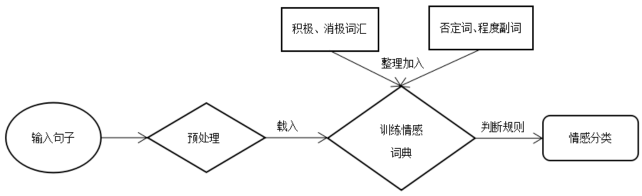

Fig.1 Text sentiment classification model based on sentiment dictionary

Foreign text sentiment analysis began in the late 1990s. Turney [22] proposed a method of word sentiment polarity recognition based on point mutual information, which has strong dependence on sentiment lexicon. With the expansion of sentiment lexicon, the time and space's consumption of point mutual information is larger, and the calculation process is slower. Yi [23] used emotional lexicon and custom rules to calculate the emotional polarity of words. Weibe [24, 25] studied the fixed relationship between emotional words in the text. Richardson [26] used Wordnet as a knowledge base to study the emotional polarity of words and achieved satisfactory results. Godbole et al. [27] constructed an emotional analysis system based on an English dictionary to evaluate the emotional tendencies of various entities (people, places, time) in the news text. Taboada et al. [28] proposed a more sophisticated model, considering the effect of degree adverbs and negative words on the sentiment orientation scores of sentences or phrases.

In Chinese, Linhong Xu and Hongfei Lin [29] extracted nine semantic features that affect sentence emotions from the vocabulary and structure of sentences. They construct an emotional vocabulary ontology library by combining manual and automatic acquisition, and conduct a preliminary study on sentiment analysis. Based on the emotional dictionary HowNet, Yanlan Zhu [30] proposed two methods of emotional orientation calculation. The experimental result shows that the word sentiment polarity classification method based on word frequency weighting has better practical value. Chun Li [31] extracted emotional words with strong emotional polarity from HowNet to build an emotional dictionary, and obtained the emotional polarity of common words by calculating the semantic similarity between common words and emotional words, and achieved better experimental results. In recent years, Bin Wen, Tingting He et al. [32] proposed a text sentiment classification method based on semantic understanding, assigning conceptual sentiment semantics to emotional words, and recalculating the emotional value of words. This method can effectively determine the sentiment orientation of words. Yanyan Zhao et al. [33] proposed an automatic recognition method for emotion evaluation unit based on syntactic path, and achieved good experimental results in the field of electronic products.

In general, the use of sentiment lexicon to classify text emotions can be applied to word feature level and sentence level, which is simple to implement, easy to determine and accurate. However, the accuracy of classification mainly depends on the construction of the emotional word dictionary, which is time-consuming and labor-intensive, and it is easy to lose important patterns hidden in the data set.

5.2 Text Sentiment Classification based on Machine Learning

The classification method based on machine learning needs to train the model. The model is shown in Figure 2. First, the text is labeled artificially as a training set, then extracts the text emotion features, constructs the emotion classifier through the machine learning method, and finally classifies the text through the classifier.

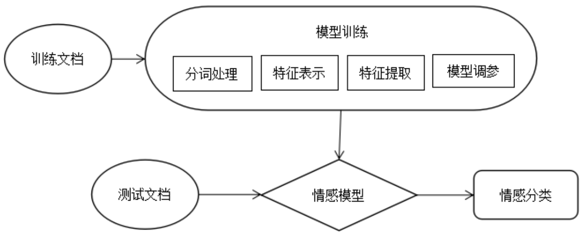

Fig.2 Machine learning text sentiment classification model

In foreign countries, Pang et al. used basic word features and a series of machine learning methods (Naive Bayesian, Maximum Entropy and Support Vector Machine, etc.) [34] to classify praise and derogation into three categories (positive, negative and neutral emotions). The method of Support Vector Machine (SVM) achieved higher accuracy. Cui et al. [35] demonstrated that the distinguishing classifier is more suitable for sentiment analysis tasks, which can better distinguish mixed sentences. Read et al. [36] proposed a semi-supervised machine learning algorithm for emotional analysis, which is independent of domain, topic and aging; Yu et al. [37] studied semi-supervised learning algorithms to solve the problem of sentiment analysis in sparse data and cross-domain. The experiments show that the semi-supervised learning algorithm of collaborative training can achieve the best classification performance. Zhu used a graph-based semi-supervised classification method to classify the IMDB film review dataset into four levels [38].

In China, Lixing Xie et al. used SVM to analyze the sentiment of microblog [39]. Huifeng Tang et al. [40] used different nouns, adverbs, adjectives, and verbs as different text representation features to compare experiments with different feature selection methods and text classification methods. Huosong Xia et al. [41] used the TF-IDF weight calculation and the RBF kernel-based support vector machine method to analyze the Ctrip customer comments and studied the impact of the stop word list in sentiment classification. Zhang et al. used the machine learning method based on String Kernel to classify Chinese documents [42]. Tan et al. classify Chinese documents through a combination of different feature selection methods and classification methods [43].

In general, the key to the machine learning-based sentiment classification method is the effective extraction of feature information. The advantage of this method is that the information acquisition is more objective, and the effect on the short text and vernacular sentences is outstanding; the disadvantage is that the dependence on the training corpus is strong, and the training period is relatively long. With the development of training corpus and the maturity of technologies such as deep learning, machine learning methods should have better development space.

5.3 Text Sentiment Classification based on Deep Learning

Deep learning experienced two lows from 1943 to 2006, but it has risen from 2006 to the rapid development period and the outbreak period since 2012. Since its birth, deep learning has achieved many research results in many fields such as speech recognition, character recognition and natural language processing. With the continuous improvement of text emotional classification algorithm based on deep learning, words, sentences and even paragraphs are distributed representation, and then it is more mature to be applied to text emotional classification tasks. The model is shown in Figure 3.

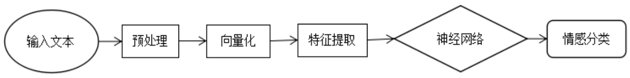

Fig.3 Deep learning text sentiment classification model

Socher et al. [44] experimented with the recurrent neural network model on Stanford emotional tree database, and the classification performance of the model was improved by 9.7% compared with the method based on feature bag. With the successful application of convolutional neural networks in the field of computer vision, researchers began to apply them to natural language processing. Kim used the convolutional neural network to make partial improvements for text sentiment analysis tasks and achieved good classification performance [45]. The convolutional neural network structure designed by Krichevsky [48] won the 2012 ImageNet Challenge, reducing the Top 5 error rate from 26% to 15% [48]. RAO et al. [46] applied a combination of LSTM neural network and word-embedding to classify social media information.

In China, Zhihong Zhao et al. [47] proposed a license plate character recognition method based on convolution neural network in 2010. The classical convolution neural network LeNet-5 was improved by increasing the number of convolution base units and output units. Xinhua Zhang et al. [49] applied BP algorithm based on neural network to text classification in 2015, and achieved good classification results. At present, convolutional neural networks are the most widely used deep learning structure. In 2015, Zhao Chen et al. [51] used the sentiment lexicon and convolutional neural network to classify the Chinese propensity analysis datasets. Compared with the current mainstream convolutional neural networks and naive Bayesian support vector machines, the classification performance of this method is better. In 2015, Jun Liang et al. [50] used the LSTM recursive network to conduct emotional analysis of Chinese text (movie film evaluation data) and achieved good results. Duo Feng et al. [52] proposed a microblog emotion classification based on convolutional neural network in 2017, aiming at the characteristics of short content, sparse features and rich in new words. Xiaoming Liu et al. [53] used convolutional neural network to classify short Internet texts in 2017, selected word vectors as original features, and then extracted features through convolutional neural networks, and finally trained Internet essays based on deep convolutional neural networks. The experimental result shows that the model improves the accuracy of emotional classification significantly.

The emotional classification of text based on deep learning can avoid the work of artificially extracting features, and it is transformed into word vector by word2vec technology, which makes the feature have semantic information, adds local feature abstraction and memory function, and has great advantages in sentiment classification. There will also be more space for development research in the future.

6. Conclusion

Text sentiment classification is a research direction of natural language processing and text mining. There are many topics worthy of further study. With the development of the information age, online buzzwords are constantly emerging. Traditional text categorization can no longer meet the needs of contemporary research in various fields. Automatic analysis of text emotions makes the current Chinese text sentiment classification research more meaningful. This paper mainly summarizes the overall process of Chinese text sentiment classification, and elaborates the preprocessing, feature extraction and sentiment classification methods in text classification.

- The amount of information on the current network text has increased rapidly. There are still many problems worthy of further study on the classification of text emotions. From the current research, the problems that need to be studied in the future are as follows:
- The amount of Chinese language data is large and complex, which directly affects the processing of some basic problems in natural language related technologies such as word segmentation, part-of-speech tagging, and syntactic analysis, which makes the effect of text sentiment analysis biased.
- Social media such as microblog, forums, blogs, etc. are colloquial and time-sensitive, and these media involve a wider range of fields, and the amount of data is larger than others. Therefore, the emotional classification of the field has certain challenges.
- Most of the existing studies are based on coarse-grained emotional classification, and the classification effect is deviated accordingly. Therefore, in the process of feature extraction, the text is processed with finer granularity and the classification results are compared.
- Emotional lexicon and standard corpus are the basis of text sentiment classification task. How to construct and improve the standard sentiment lexicon and corpus, how to more effectively combine text sentiment classification technology with practical application, how to improve the classification evaluation standard is also an important challenge to promote Chinese text sentiment classification theory and technology research.

Acknowledgments

Our thanks to the Inner Mongolia Natural Science Foundation: 2014MS0614 for our support.

References

[1]. Ruilei Wang, Jing Qi, Pan Xiaohua Pan, et al. An improved maximum forward matching algorithm for Chinese word segmentation[J]. Journal of Computer Applications and Software, 2011, 28(03): 195-197.

[2]. Jun Zhou, Zhonghua Zheng, Wei Zhang. Chinese word segmentation method based on improved maximum matching algorithm[J]. Computer Engineering and Applications, 2014, 50(02): 124-128.

[3]. OACHIMS T. Text Categorization with Support Vector MaChines: Learning with many ReleVant Features[M]. Berlin Heidelberg: Springer, 1998.

[4]. SALTON G, WONG A, YANG C S. A Vector Space Model for Automatic Indexing[J]. Communications of the ACM, 1975, 18(11): 613—620.

[5]. BERRY M W. Survey of Text Mining[J]Computing Reviews, 2004, 45(9): 548.

[6]. Rumelhart D E, Hinton G E, Williams R J. Learning Representations by Back-propagating Errors[J]. Cognitive Modeling, 1988, 5:533-536.

[7]. Bengio Y, Ducharme R, Vincent P, Jauvin C. A Neural Probabilistic Language Model[J]. The Journal of Machine Learning Research, 2003, 3: 1137-1155.

[8]. Collobert R, Weston J. A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning[C]// Proceedings of International Conference on Machine learning, 2008: 160-167.

[9]. Mikolov T, Chen K, Corrado G, Dean J. Efficient Estimation of Word Representations in Vector Space[J]. Computing Research Repository, 2013, abs/1301.3781.

[10]. MIKoLoV T, CHEN K, CORRADO G, et a1. Efficient Estimadon of Word Representations in Vector Space[J]. Computer Science, 2013(1): 28—36.

[11]. LILLEBERG J, ZHU Y, ZHANG Y. Support Vector Machines and Word2vec for Text C1assmcation with Semantic Features[C]// IEEE, International Conference on Cognitive Informatics & Cognitive Computing, July 6—8, 2015, Beijing, china. New York: IEEE, 2015: 136—140.

[12]. Zhongyang Xiong, Pengzhao Zhang, Yufang Zhang. Research on text classification feature selection method based on χ^2 statistic[J]. Journal of Computer Applications, 2008, 28(2):513-514.

[13]. Tongsen Du, Yajian Zhou. Research and improvement of expected cross entropy algorithm in feature selection [J]. 2013. http: www. paper. edu. Cn.

[14]. Xueli Fan, Haizhen Feng, Meng Yuan. Principal component analysis feature selection algorithm based on mutual information[J]. Control and Decision, 2013, 28(6): 915-919.

[15]. Wei Dong, Xue Liu, Hong Ni. Adaptive feature selection method based on information gain[J]. Computer Engineering and Design, 2014(8): 2856-2859.

[16]. Yuehua Tao, Xigang Wang, Yunai Wang. Research on feature extraction in information retrieval vector space model[J]. Journal of Yunnan Normal University: Natural Science Edition, 2000, 20(6): 18-20.

[17]. Xiaojun Yu, Feng Liu, Chun Zhang. An improved algorithm for text feature extraction based on N-Gram [J]. Modern Computer, 2012 (23): 3-7.

[18]. Xiaobin Peng. Research on text sentiment analysis method based on word2vec[J]. Network Security Technology & Application, 2016(7): 58-59.

[19]. Qian Zhang, Zhangmin Gao, Jiayong Liu. Research on Short Text Classification of Microblog Based on Word2vec[J]. Information Network Security, 2017(1).

[20]. Ren Huang, Wei Zhang. Research on Emotional Tendency of Internet Commodity Reviews Based on Word2vec[J]. Computer Science, 2016, 43(S1): 387-389.

[21]. Keyu Chen, Zhongshi He. Research on Emotion Classification of Hotel Reviews Based on Emotional Dictionary[J]. Modern Computer (Professional Edition), 2017(6).

[22]. Turney P D. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews[C]// Proceedings of Annual Conference of the Association for Computational Linguistics, 2002: 417-424.

[23]. Yi J. Nasukawa T, Bunescu R, ct al. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques[C]// Data Mining, 2003. ICDM 2003. Third IEEE International Conference on. IEEE, 2003: 427-434.

[24]. Wiebe J, Wilson T, Cardie C. Annotating expressions of opinions and emotions in language[J]. Language resources and evaluation, 2005, 39(2): 165-210.

[25]. Wilson T, Wicbe J, Hoffmann P. Recognizing contextual polarity in phrase-level semiment analysis[C]// Preceding's of the conference on human language technology and empirical methods in natural language processing. Association for Compulational Linguistics, 2005: 347-354.

[26]. Richardson R. Smeaton A, Murphy J. Using WordNet as a knowledge base for measuring semantic similarity between words[j], 1994: 83-98.

[27]. Godbole N, Srinivasaiah M, Skiena S. Large-scale Sentiment Analysis for News and Blogs[C]// Proceedings of International Conference on Weblogs and Social Media, 2007: 21.

[28]. Taboada M, Brooke J, Tofiloski M, Voll K, Stede M. Lexicon-based Methods for Sentiment Analysis[J]. Computational Linguistics, 2011, 37(2): 267-307.

[29]. Linhong Xu, Hongfei Lin. Discourse Emotional Computing Based on Semantic Features and Ontology. Computer Research and Development 2007, 44(3): 356~360.

[30]. Yanlan Zhu, Jin Yan, Yaqian Zhou, et al. Calculation of Semantic Tendency Based on HowNct. Chinese Journal of Information, 2006, 20(1): 16-22.

[31]. Chun Li, Baojun Qiao, Yuanda Cao, et al. Research on lexical tendency recognition based on semantic analysis [J]. Pattern Recognition and Artificial Intelligence, 2008, 21(4): 482-487.

[32]. Bin Wen, Tingting He, Le Luo et al. Research on text sentiment classification based on semantic understanding. Computer Science, 2010, 37(6): 261~264.

[33]. Yanyan Zhao, Bing Qin, Wanxiang Che et al. Emotional Evaluation Unit Recognition Based on Syntactic Path. Computer Research and Development, 2011, 22(5): 887~898.

[34]. Pang B, Lee L, Vaithyanathan S. Thumbs up? Sentiment Classification using Machine Learning Techniques[C]// Proceedings of Annual Conference of the Association for Computational Linguistics, 2002:79-86.

[35]. Cui H, Mittal V, Datar M. Comparative Experiments on Sentiment Classification for Online Product Reviews[C]// Proceedings of American Association of Artificial Intelligence, 2006: 1265-1270.

[36]. Read J, Carroll J. Weakly Supervised Techniques for Domain-independent Sentiment Classification[C]// Proceedings of International Conference on Information and Knowledge Management Workshop on Topic-sentiment Analysis for Mass Opinion, 2009: 45-52.

[37]. Yu N, Kübler S. Filling the Gap: Semi-supervised Learning for Opinion Detection Across Domains[C]// Proceedings of Annual Conference of the Association for Computational Linguistics, 2011: 200-209.

[38]. Zhu X, Goldberg A. Introduction to Semi-Supervised Learning[M]. Morgan & Claypool, 2014.

[39]. Lixing Xie, MingZhou, Maosong Sun. Analysis and Feature Extraction of Multi-Strategy Chinese Microblog Based on Hierarchical Structure[J]. Journal of Chinese Information Processing, 2012, 26(1): 73-83.

[40]. Huifeng Tang, Songbo Tan, Xueqi Cheng. Comparative Study of Chinese Sentiment Classification Techniques Based on Supervised Learning. Chinese Journal of Information, 2007, 21(6): 88~94.

[41]. Huosong Xia, Min Tao, Yi Wang et al. The effect of stop word list on Chinese text sentiment classification based on SVM. Journal of Information, 2011, 30(4): 347~352.

[42]. Zhang C, Zuo W, Peng T, et al. Sentiment Classification for Reviewa Using Machine Learning Methods Based on String Kernel[C]. International Conference on Convergence and Hybrid Information Technology. IEEE, 2008: 909-914.

[43]. Tan C, Chen G, Song Q. Sentiment mining analysis based on network comments[J]. Metallurgical & Mining Tndustry, 2015, 7(5): 261-265.

[44]. Socher R, Perelygin A, Wu J Y, Chuang J, Manning C D, Ng A Y, Potts C. Recursive Deep Models for Semantic Compositionality over a Sentiment Treebank[C]// Proceedings of Annual Conference on Empirical Methods in Natural Language Processing, 2013: 1631-1642.

[45]. Kim Y. Convolutional Neural Networks for Sentence Classification[C]// Proceedings of Empirical Methods in Natural Language Processing, 2014: 1746-1751.

[46]. RAO A, SPASOJEVIC N. Actionable and political text classification using word embeddings and LSTM[J]. arXiv:1607.02501vl[cs.CL], 2016.

[47]. Zhihong Zhao, Shaopu Yang, Qiang Ma. Research on License Plate Character Recognition Based on Convolutional Neural Network LeNet-5[J]. Journal of System Simulation, 2010, 22(03):638-641.

[48]. Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012:1097-1105.

[49]. Xinhua Zhang. Text Classification Based on Artificial Neural Network[J]. Journal of Hubei Second Normal University, 2015, 32(2): 28-30.

[50]. Jun Liang, Yumei Chai, Huibin Yuan, Minglei Gao, Hongying Yan. Sentiment Analysis Based on Polar Transfer and LSTM Recursive Network[J].Journal of Chinese Information Processing,1003 - 0077(2015) 05—0152—08.

[51]. Zhao Chen, Ruifeng Xu, Lin Gui, et al. Chinese sentiment analysis based on convolutional neural network and lexical sequence features[J]. Chinese Journal of Information Science, 2015, 29(06): 172-178.

[52]. Duo Feng, Zheng Lin, Peng Fu, et al. Chinese microblog emotion classification based on convolutional neural network[J]. Journal of Computer Applications and Software, 2017, 34(04): 157-164.

[53]. Xiaoming Liu, Ying Zhang, g Qiusheng Zhen. Internet Short Text Sentiment Classification Based on Convolutional Neural Network Model [J]. Computer and Modernization, 2017(04): 73-77.