http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 13 (9): 1612-1621, 2014
ISSN 1812-5638 / DOI: 10.3923/itj.2014.1612.1621

Research of Method in Sentiment Orientation Analysis of Texts Based on Domain Sentiment Lexicon

Yun Pan, Xiaojun Li, Hanxiao Shi and Hong Liu
School of Computer and Information Engineering, Zhejiang Gongshang University,
Hangzhou, China

Abstract: To improve the performance of sentiment orientation analysis of texts, an analysis model which combines with ontology learning and domain sentiment lexicon is proposed. This model can enhance the ability to distinguish the sentiment words and improve efficiency of sentiment analysis. As domain sentiment lexicon has the problem of inadequate coverage, this paper introduces ontology learning model to solve the problem. The experimental results show that the method based on domain sentiment lexicon can improve the performance of sentiment orientation analysis of texts.

Key words: Sentiment orientation analysis, sentiment lexicon, domain sentiment lexicon, ontology

INTRODUCTION

In recent years, along with the rapid development of e-commerce, more and more subjectivity comment articles are appeared for product or business. It not only helps to make a choice for the potential customers but also gives effective feed-back for the providers. It's very important for the provider to receive the latest information from these comments, so that they can improve the quality of their products and services. However because there are too many comments for people to browse hardly by manual and at the same time there may be some comments which are intended to write, as a result it will be difficult for the customers and providers to get the wanting information comprehensively and correctly. And what's more, spread information via Internet has become a new mode and it plays more and more important role in social public opinion. To understand the social public opinion in time and actively guide is the government's important work. The government should capture the hot topics on the Internet, analysis and statistics everyone's opinion and mine more useful information, in order to take effective strategy which can induce the social public opinion to be more objective.

Sentiment orientation analysis of texts has become a research hotspot in the field of natural language processing. Through searching comment information on the internet automatically and analysis the information to figure out the sentiment orientation which includes happiness, sadness or to indicate whether it is approved or refused, even to find out the sentiment change rule along with the time evolution. The analysis results are very useful for government to understand the social public opinion, for producer to make market research and for customers to choose their suitable product.

Recently, the research about text sentiment orientation mainly includes three kinds of different granularity: word-level sentiment analysis, sentence-level sentiment analysis and chapter-level sentiment analysis. For word-level sentiment orientation analysis, Weibe used word collocations to find out the sentiment orientation words and the relationship from subjectivity text (Wiebe *et al*., 2001). Hu and Liu (2004) used the co-occurrence sentiment words and its candidate attributes to obtain the overall tendency of the customers for an product or an attribute of the product; comprehensively considered the spread, density and intensity of polar elements by the calculation of word semantic orientation, so as to get the public evaluation of celebrities (Tsou *et al*., 2005). The objects to be processed by sentence-level sentiment orientation analysis are the sentences in a particular context. The purpose of the analysis is to analyze and extract the subjectivity of the sentence as much as possible, which includes the evaluated subjects, sentiment polarity, sentiment orientation and the importance of the evaluation itself. Whitelaw Casey calculated the overall orientation of a sentence by the calculation of sentiment phrase of that sentence Whitelaw *et al*. (2005) and Wang and Zhao (2007)

Corresponding Author: Hanxiao Shi, School of Computer and Information Engineering, Zhejiang Gongshang University, Hangzhou, China

proposed a classification model to distinguish subjective or objective, positive or negative and sentiment orientation intensity of a sentence, which could determine the sentence-level sentiment orientation. Chapter-level sentiment orientation analysis mainly obtains the overall sentiment orientation from the perspective of the entire text (Riloff and Wiebe, 2003; Durant and Smith, 2006; Pang *et al*., 2002; Pang and Lee, 2005). Early-stage sentiment classification of positive and negative meanings originated from chapter level, while the purpose is to mine the overall attitude of the evaluation articles on a certain product or service. Recently, the analysis adopted machine learning methods, such as Pang Bo utilized respectively standard bag-of-words, Naive Bayes, maximum entropy and SVM classification to classify the orientation of movie comments on Usenet and the experimental results validated the superior performance of SVM classification with the accuracy rate of up to 80% (Pang *et al*., 2002; Pang and Lee, 2005).

All above methods for sentiment orientation analysis of texts are closely related with the sentiment lexicon. Due to the limitation of existed sentiment lexicon, it is difficult to distinguish the different sentiment orientation of the same sentiment words in different attributes and fields. Therefore, this paper constructs a sentiment lexicon based on ontology to improve the performance of sentiment orientation analysis of texts, which can effectively to distinguish the sentiment words in different attributes and fields by introducing general ontology and domain ontology. In this paper, we first adopt ontology technology to construct a domain sentiment lexicon which is being automatically expanded, quantitative and domain-oriented. Then, we analysis the sentiment orientation of texts based on the domain sentiment lexicon.

DOMAIN SENTIMENT LEXICON

Through the study of the characteristics of the domain knowledge, we can find that there are not only similarities but also difference between sentiment words which are used to describe different domain knowledge. For example, sentiment words such as "好", "美好", "漂亮", are clearly positive words or words such as "差", "难看", "丑陋" are clearly negative words and no matter in which field their sentiment orientation is firmed. These words can be classified to non-domain words. If some sentiment words appear in domain A but never appear in domain B, these words are called domain words; In addition, if some sentiment words appear in both A and B domain but they have different degree of sentiment orientation in different domain, these words are also called domain words and their domain characteristics are more conspicuous than former. What's more, considering degree adverbs and negative adverbs are important to sentiment expression, we should also aware that they play the role in the sentiment analysis. In order to reflect the domain characteristic of sentiment words and improve sharing and fusion of sentiment words between different domain knowledge, ontology technology is used to construct the sentiment lexicon in this paper. In this sentiment lexicon, the unique concept of one domain, the common concept in every domain and attributes of concepts, relations of concepts, relations of attributes are distinguished, so ontology in sentiment lexicon are classified in three part: genetic ontology, domain ontology and application ontology. The genetic ontology is used to represent the domain sentiment words which have formed sentiment polarity, degree adverbs and negative adverbs which express intensity. Domain ontology is used to represent the unique sentiment words in certain domain; Application ontology is used to model knowledge of specific application domain. The structure of ontology is shown in Fig. 1.

Sentiment words ontology can more fully represent semantic information of sentiment words, such as the sentiment orientation of words, similarity, progressive and turn relations between words, so they provide effective analysis basis for sentiment orientation analysis of texts. In this sentiment lexicon, the sentiment ontology is described as a quintuple $O = \{C, A^c, R, H, X\}$, where $C = \{c_1, c_2,...\}$ is the set of domain concepts, specifically the name of sentiment words; $A^c = \{A^c(c_1), A^c(c_2),...\}$ is the attributes set of concepts, representing the basic information about sentiment words, including the number, intensity and etc.; $R = \{r_1(c_1, c_2), r_2(c_2, c_3),...\}$ represents set of semantic relations between concepts, such as the synonymous and hyponymous relations; $H = \{(c_1, c_2), (c_2, c_3),...\}$ is the hierarchical relations of concept set C, describing the relationship between parent class and sub-class of sentiment words; $X = \{x_1, x_2,...\}$ is the axiom set, representing the constraints on concept properties, associated attributes and object concepts and being the basis for ontology learning.

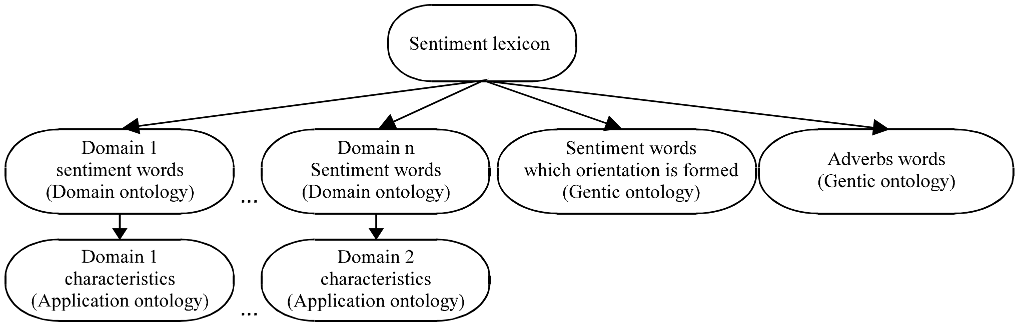

Fig. 1: The ontology structure of domain sentiment lexicon

Genetic ontology: Genetic ontology represents sentiment words whose sentiment orientation has nothing to do with its domain, degree adverbs or negative adverbs. Sentiment words such as "好", "美好" and "漂亮" are all positive words, while sentiment words such as "差", "难看", "丑陋" are all positive words. The attribute set of this type of sentiment words ontology can be represented as:

$$So^{sentiwordset} = \{SO^{sentiwordset}(sentiword_i)|i = 1,\ldots, n\},\quad sentiwordset = \{sentiword_i|i = 1,\ldots, n\}$$

where, sentiwordset indicates the set of sentiment words, $sentiword_i$ indicates the specific sentiment words and SO represents the sentiment value of certain word, whose value is manually scored and within a range of -1 to 1, without including the value of zero.

Considering the importance of degree adverbs in sentiment expressing, this paper adopts the classification model proposed by Lin and Guo (2003). The degree adverbs are divided into four hierarchies, namely extreme-degree, high-degree, middle-degree and low-degree and a coefficient is assigned respectively to every degree adverb. Once there appears a sentiment word decorated by degree adverbs, its polarity value becomes the polarity value of this sentiment word multiplied by the coefficient. Specific coefficients are shown as in Table 1. For negative adverb, the polarity value is multiplied by -1.

The attribute set of degree adverbs is defined as follows:

$$(Factor, Nature)^{adverbset} = \{(Factor, Nature)^{adverbset}(adverb_i)|i = 1,\ldots, n\},\ adverbset = \{adverb_i|i = 1,\ldots, n\}$$

where, adverbset represents the set of degree adverbs, adverb represents the specific adverb, factor is the coefficient of the degree adverb and nature represents the nature of degree adverbs, namely absolute or relative.

Table 1: Coefficient table of degree adverbs

Degree	Relative degree adverb	Absolute degree adverb	Coefficient
Extreme	最, 最为,...	太, 极为,...	1.5
High	更, 更加, 越,...	很, 特别, 十分,...	1.3
Middle	较, 较为,...	不太, 太低,...	0.7
Low	稍, 略,...	有点, 有些,...	0.5

Domain ontology: The domain ontology of domain sentiment words are classified into two categories: positive and negative. Since an identical sentiment word expresses different sentiment orientation when describing different attributes within the same domain, such as for laptop domain, comments such as "电池的使用时间长" is positive while "程序响应时间长" is negative. For these sentiment words with dynamic polarity, this paper adopts the formalization processing of ordered pair in the nested definition of ontology, <domain attribute, sentiment orientation> is used to solve the above-mentioned problem, which is defined as:

$$So_{ij}^{domainwordset} = \{ So_{ij}^{domainwordset}(domainword_i, feature_j)|\ i = 1, 2,\ldots, n, j = 1, 2,\ldots, m\}\ domainwordset = \{domainword_i|i = 1,\ldots, n\}$$

where, $SO_{ij}^{domainwordset}$ represents the domain sentiment words, $domainword_i$ corresponds to the real-time attributes and $feature_j$ stands for the sentiment value.

Application ontology: Application ontology is used to describe the domain attributes, namely the domain attribute ontology, which adopts the hierarchical way to describe the classified attribute knowledge. Figure 2 depicts the ontology model for the attributes of the domain "clothing".

In many domains, there exists the hierarchical semantic inclusion and included relations. Taken the clothing domain as example, "service" can exist both as the attribute and as the attributes set for the clothing domain. When considered as the attributes set, it can include attributes such as "shop staff", "call center staff", "delivery staff" or "logistic".

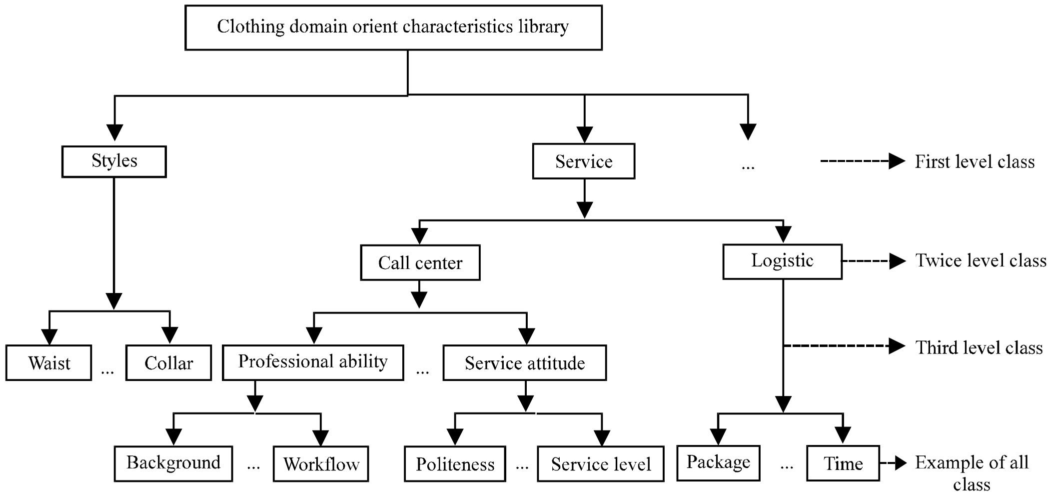

Fig. 2: The ontology model for the attributes of the domain "clothing"

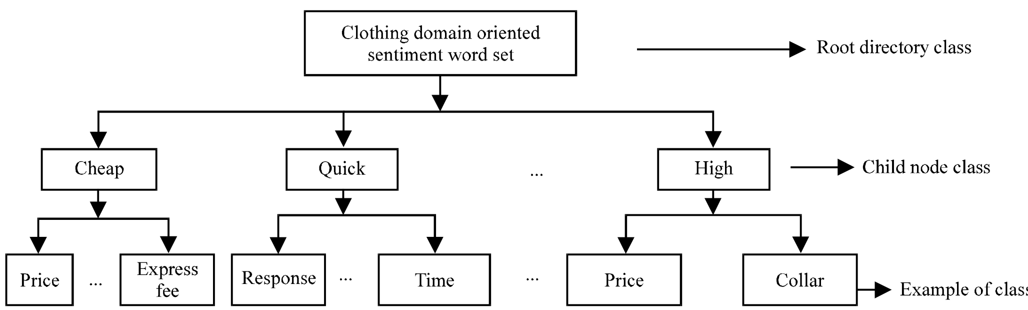

Fig. 3: The hierarchy structure of the ontology sub-library of domain "clothing"

From the angle of coarse-grained analysis, "service" could provide clients an initial understanding of clothing deal but for fine-grained analysis, various customers may have different requirements for clothing "service". For example, a customer who buys clothing from shop considering the "service" would focus more on the "shop staff", a customer who buy clothing from Internet may be more focused on the "call center staff" or "delivery staff".

In addition, attribute words may have the same or similar semantic relations, such as "package" and "packet", "collar" and "neckwear", "fabric" and "cotton material" and etc. They have the same meaning but with similar or different form or just in a shorthand form but they refer to the same thing, namely they have the same superclass.

Based on the above-mentioned domain characteristics, when domain-oriented attributes set is designed, the principle is set as that it should proceed from large to small and from coarse to fine. The domain characteristics shall be manually divided according to the coverage of the concepts. Three principles for division are:

- The concept range of each hierarchy of sub-class should cover as much as possible the range of its direct parent class
- Each attribute class should avoid intersecting with other attribute classes or instances
- The order of division should proceed from large to small and from coarse to fine

The hierarchy structure corresponding to the sentiment word set of domain "clothing" (Fig. 2) is shown in Fig. 3.

In this structure, sentiment word is the sub-node of the root class directory, while the attribute properties are the instance of the domain sentiment words. For example, the sentiment word "high" is a sub-node in the structure of the "clothing" domain sentiment word set, while "price" and "collar" are considered as instances of this class and the attributes "price" and "collar" is considered as class attributes.

FRAMEWORKS FOR SENTIMENT ANALYSIS BASED ON DOMAIN-ORIENTED SENTIMENT LEXICON

The proposed framework for sentiment analysis based on domain-oriented sentiment lexicon is shown in Fig. 4. Which includes pre-processing module, identification module, ontology module, sentiment analysis module and automatic expansion module. The process is shown in Fig. 4:

- Store comment information crawled form the Internet to the comments database and establish the corpus for sentiment analysis
- Pre-process the comments, including noise handling, word segmentation, part-of-speech tagging and anaphora resolution
- Identify attributes and sentiment words from tagging result based on domain sentiment lexicon
- Do sentiment orientation analysis for recognition results and display visual results to user
- Extend no existed sentiment words to the sentiment lexicon by implementing the algorithm on the ontology expansion and consolidation of the sentiment words

Pre-processing module: Pre-processing the corpus extracted from Internet, including noise handling, word segmentation processing, part-of-speech tagging. Word segmentation processing is carried out by the ICTCLAS1 system developed by the Chinese Academy of Sciences. This word segmentation system adopts Cascading Hidden Markov Models (HMM) and performs the Chinese word segmentation, part-of-speech tagging, named entity identification and new words identification, while its segmentation identification accuracy is up to 98.45%.

The principles of Ontology mapping: To perform sentiment analysis, a set of sentiment words needs to be mapped into domain ontology, namely the set of sentiment words $T = \{t_1, t_2,\ldots, t_n\}$ is matched with the domain ontology to derive a concept set $C = \{c_1, c_2,\ldots, c_n\}$ and their attributes. The specific mapping principles are:

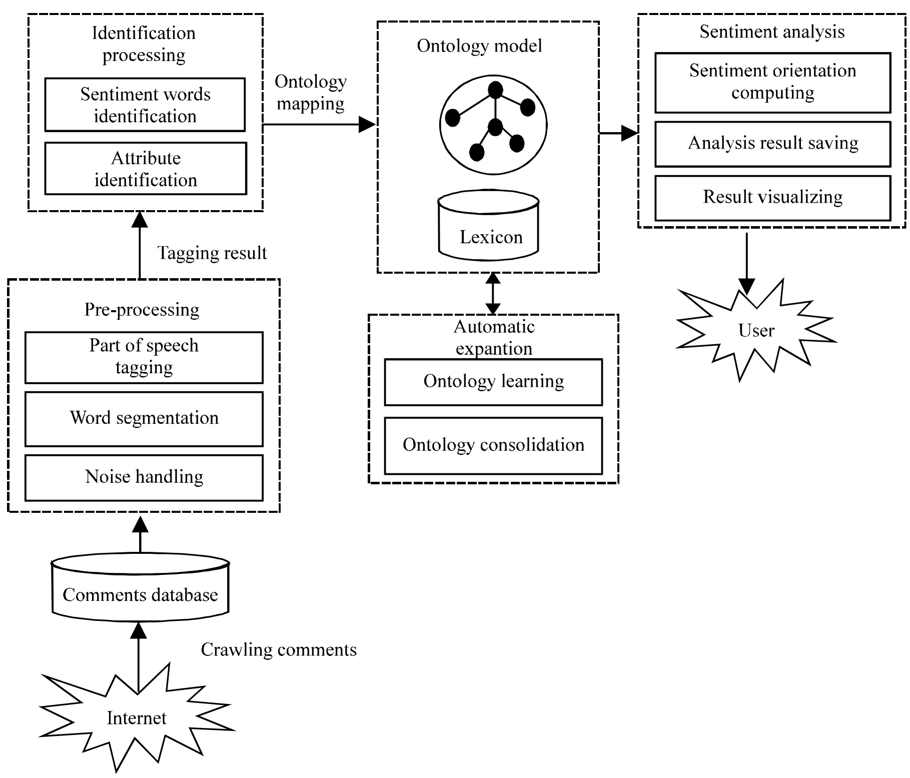

Fig. 4: The framework for sentiment analysis based on domain oriented sentiment lexicon

- If $t_i(i = 1, 2,..., n)$ in the set T matches directly with the classes in the domain ontology, the corresponding concept $c_j(j = 1, 2,..., m)$ is directly added into the set C
- If $t_i(i = 1, 2,..., n)$ in the set T not only matches directly with the classes in the ontology but also matches with the other classes or instances of that class, then the principle is that class has the highest priority, instance the second, property the third. The class is the output of the matched concept; otherwise, the class with the property is added into the set C
- If $t_i(i = 1, 2,..., n)$ in the set T matches with a certain individual in the ontology but doesn't match with other classes or properties, then the corresponding class is added to the set C, otherwise the class name as a concept is added into the set C according to its priority
- If the element in set T doesn't match any objects, then the element is recorded and the ontology is extended

Principles of ontology expansion and consolidation: Considering the problem of insufficient coverage of domain sentiment lexicon proposed the expansion method for domain sentiment lexicon based on the principle of "two-way communication". This method accesses the attributes of a known sentiment word, with which to acquire other attribute words or other sentiment words modified by the relative attribute. The process is performed over and over, so as to extend the coverage of the sentiment lexicon. The specific principles for adding the new sentiment words into the existing sentiment lexicon ontology are:

- According to the ontology mapping principle, if the sentiment words could not be retrieved in any sentiment ontology library, namely non-existence in the lexicon, then the ontology for sentiment word construction is added directly into the corresponding location of the sentiment ontology library
- In accordance with the principle of ontology mapping, if the sentiment words could not be retrieved in certain domain sentiment ontology library but be retrieved in other domains, namely the existence of cross-domain retrieving, then the new polarity value derived would be compared with the polarity values of other sentiment words in this domain ontology library. If the difference is below the threshold limit value, the similarity of the sentiment tendency of this sentiment word within two or more domains is confirmed and the ontology of the sentiment word is merged and added into corresponding location in the general ontology library. If the difference exceeds the allowable threshold range, then this sentiment word has different sentiment tendency in various domains and the addition is processed as in step (1)

Ontology learning model: Since the scale of domain lexicon is limited, there is a shortage in the coverage of sentiment words. There are many words that is not included in the sentiment lexicon. Considering that using sentiment words usually have continuity, that is to say, if you have good impression of one thing, you will often continuously use positive words to describe it; if you have bad impression of one thing, you will often continuously use negative words to describe it. The continuity reflects the Co-occurrence attributes of sentiment words. Therefore, this paper proposes a machine learning model which the sentiment words set are automatically clustered. From the clustering results, we can get k clusters. There are co-occurrence attributes between sentiment words in the same cluster. Otherwise, there are no co-occurrence attributes between sentiment words in the different cluster. The pro-processing results from corpus are clustered into k clusters $C = \{C_1, C_2,... C_n\}$. The algorithm has three mainly steps: (1) pro-processing and obtaining sentiment word, (2) initialization and (3) clustering process.

Algorithm: Ontology learning model algorithm

Input: text corpus D for training.

Output: cluster result $U = \{U_1, U_2,...U_n\}$.

Detailed process:

- Pro-processing corpus
- Obtaining sentiment word set SW from corpus
- Counting the number of times that any two words appear at the same time in corpus and constructing vector space model. The model can be expressed as $Sw_i = \{w_{i,1}, w_{i,2},...w_{i,j}...w_{i,m}\}$. $w_{i,j}$ indicates the numbers of times that No.i sentiment and No.j sentiment appear at the same time in corpus
- Initializing cluster M: according to the means of every cluster information, calculate the distance between every web and means; then redistribute the related web according to the minimum distance and add these web to the nearest cluster, at last update the means of every cluster information
- Loop until there is no redistribution to clusters and obtain the clustering results $C = \{C_1, C_2,...C_n\}$

- For the sentiment word sw_i in the new corpus, according to the Eq. 1, we can calculate the distance between this word and the means C_k

$$\mathrm{sim}(SW_i, SW_j) = \frac{\sum_{k=1}^{n} w_k^i * C_k^j}{\sqrt{\sum_{k=1}^{n} (w_k^i)^2} \sqrt{\sum_{k=1}^{n} (C_k^j)^2}} \quad (1)$$

k = arg min $1_{\leq j \leq n}$ ($D(w_i, C_j)$), so we can set sw_i to U_k.

THE PROCESS OF SENTIMENT ANALYSIS

In the process of sentiment analysis, ICTCLAS (2008) performs word segmentation on the corpus first and then the part-of-speech tagging and anaphora resolution are carried out. Combined with the concept set of domain attribute, attribute identification is processed and the sentiment word retrieval is performed through ontology mapping in the lexicon. If some results are successfully retrieved, the direct or relevant polarity value of the sentiment word is adopted to calculate the sentiment tendency of the corpus; otherwise, the existing sentiment lexicon is automatically extended. The specific procedure is shown in Fig. 5.

While computing the sentiment orientation of a corpus, we use sentences as the processing unit and unified analyze and process the SRL result from comments information of each reviewer. Trough unified collect processing we can obtain sentiment orientation value of every attributes from all reviewers. For every different attribute of the combinations of attributes, sentiment words and adverbs, this paper designs five attribute based sentiment orientation calculation methods after experiments and analysis:

- Reduplicated word combined with public basic sentiment word, such as “高高兴兴”, “漂漂亮亮”. We can first find the basic sentiment word from reduplicated word and then get the sentiment orientation value of reduplicated word by computing the value of basic sentiment word. In general, reduplicated word has little effect on the sentiment value of the original basic word, so in order to simplify the problem, we can get the value of basic word as the value of reduplicated word
- The word combined with basic word and basic word, such as “小心谨慎”. The sentiment orientation value of this word can be gotten by computing the average of the sentiment orientation value of basic words
- The combination word combined with negative adverb and basic word, such as “不漂亮”. The sentiment orientation value of this word can be gotten by computing the negation operator to the sentiment orientation value of basic words
- The combination word combined with adverb of degree and basic word, such as “很 漂亮”. We can first get the sentiment orientation value of the basic word and then get the coefficient of adverb pre-defined (see Table 1), such as the coefficient of adverb “很” is 1.3 and adverb “比较” is 0.7. Finally, we can get the sentiment orientation value of the combination word by multiplying the value of basic word and the coefficient of adverb. If the value is beyond the range [-1, +1], we will take -1 or +1
- The combination word combined with negative word, adverb of degree and basic word or adverb of degree, negative word and basic word, such as “不太漂亮” or “太不漂亮”. The computing of this type is relatively complex and the position relation of negative word and adverb of degree directly affect the sentiment orientation value of the combination word. According to linguistic knowledge, this paper designs formula 2 to compute the sentiment orientation value of the combination word

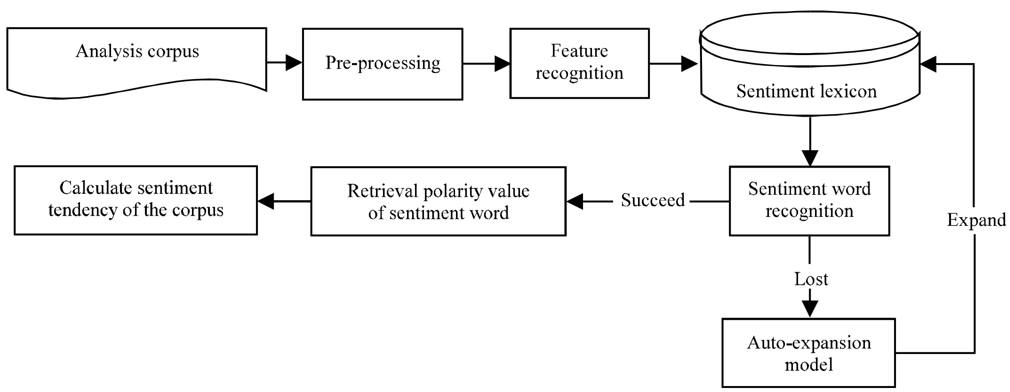

Fig. 5: The process of sentiment analyzing

$$S_w = \begin{cases} (\eta - D_{wi}) * S_{wi} & \text{cond1} \\ \text{sign}(S_{wi}) * (\text{abs}(S_{wi}) * D_{wi}) & \text{cond2} \end{cases} \quad (2)$$

where, cond1 represents that adverb of degree locates between negative word and basic word and cond2 represents that negative word locates between adverb of degree and basic word.

Where, S_{wi} is the sentiment orientation value of basic word, D_{wi} is the coefficient of adverb of degree. η is the coefficient of negative adverb and it is also the sum of the extreme value of D_{wi}, so η is set by 2. sign(S_{wi}) is symbolic value of S_{wi}, if S_{wi} is greater than zero, then sign(S_{wi}) is +1; if it is less than zero, then sign(S_{wi}) is -1. For example, the word "不太漂亮" combined with negative adverb "不", adverb of degree "太" and basic word "漂亮", the sentiment orientation value of "漂亮" is 0.8, the coefficient of "太" is 1.3, according to Eq. 2, the sentiment orientation value of "不太漂亮" is 0.56 but the sentiment orientation value of "太不漂亮" is -1. It is not difficult to see that this result suits with our subjective judgment.

To describe the common sentiment orientation of domain attributes, this paper takes an average of all attribute based sentiment orientation calculation methods and the averaged value is treated as the objective distribution of the sentiment orientation. The specific form is defined as follows:

$$\overline{SO_{fi}} = \frac{\sum_{j=1}^{n} SO_{fi}}{\sum_{j=1}^{n} \text{count}(fi)} \quad (3)$$

where, So_{fi} is the sentiment orientation value of attribute fi after considering the relevant levels of degree word and negative word of all the comments, $\overline{SO_{fi}}$ is the average So_{fi} of all the comments. The ultimate sentiment orientation values of all attributes would be $(\overline{SO_{f1}}, \overline{SO_{f2}}, \ldots, \overline{SO_{fn}})$.

EXPERIMENTS ADD ANASYSIS

Corpus used in this experiment included two aspects: (1) the corpus for the hotel domain, which have 3000 online positive or negative text comments mainly downloaded from Ctrip or individual hotel's official website (such as Hanting Inns & Hotels, Home Inn, etc.) and (2) the corpus for the clothing domain, which have 1000 positive or negative text comments mainly downloaded from Taobao, PaiPai, Amazon, 360buy and other large online shopping sites.

In order to measure the effect of the data processing, we used the comparative analysis with the experimental results and the results of manual processing method. The manual processing mainly refers to sentiment judgment and scoring by manual and getting a more objective sentiment orientation value through computing the average of multiple scoring results.

First, we randomly taken 100 positive comments and 100 negative comments from nearly 3000 hotel comments and then through a series of text pre-processing, natural language processing, extracting and identifying candidate attributes, sentiment words, adverbs of degree and negative adverbs, we ultimately identified and got 378 public basic sentiment words and their sentiment orientation values, 106 attribute-sentiment word pairs and their sentiment orientation values, 199 adverbs of degree and their coefficient, 72 negative adverbs and their coefficient. Through manual judgment, there are 76 public sentiment words, 13 domain sentiment words, 27 adverbs of degree and 4 negative adverbs which repeats one or more times in comments. The experiment results of hotel domain shown in Table 2.

Second, we also randomly taken 50 positive comments and 50 negative comments from nearly 1000 clothing comments and then through a series of text pre-processing, natural language processing, extracting and identifying candidate attributes, sentiment words, adverbs of degree and negative adverbs, we ultimately identified and got 187 public basic sentiment words and their sentiment orientation values, 71 attribute-sentiment word pairs and their sentiment orientation values, 96 adverbs of degree and their coefficient, 43 negative adverbs and their coefficient. Through manual judgment, there are 48 public sentiment words, 11 domain sentiment words, 11 adverbs of degree and 3 negative adverbs which repeats one or more times in comments. The experiment results of clothing domain shown in Table 3.

Consolidated sample data of two domains, there are 565 public sentiment words (including repeated

Table 2: Experiment results of hotel domain

	Words			
Result	Public sentiment words	Attribute-sentiment word pairs	Adverbs of degree	Negative adverbs
No.	378	106	199	72
Repeated No.	72	13	27	4

Table 3: Experiment results of clothing domain

	Words			
Result	Public sentiment words	Attribute-sentiment word pairs	Adverbs of degree	Negative adverbs
No.	187	71	96	43
Repeated No.	48	11	11	3

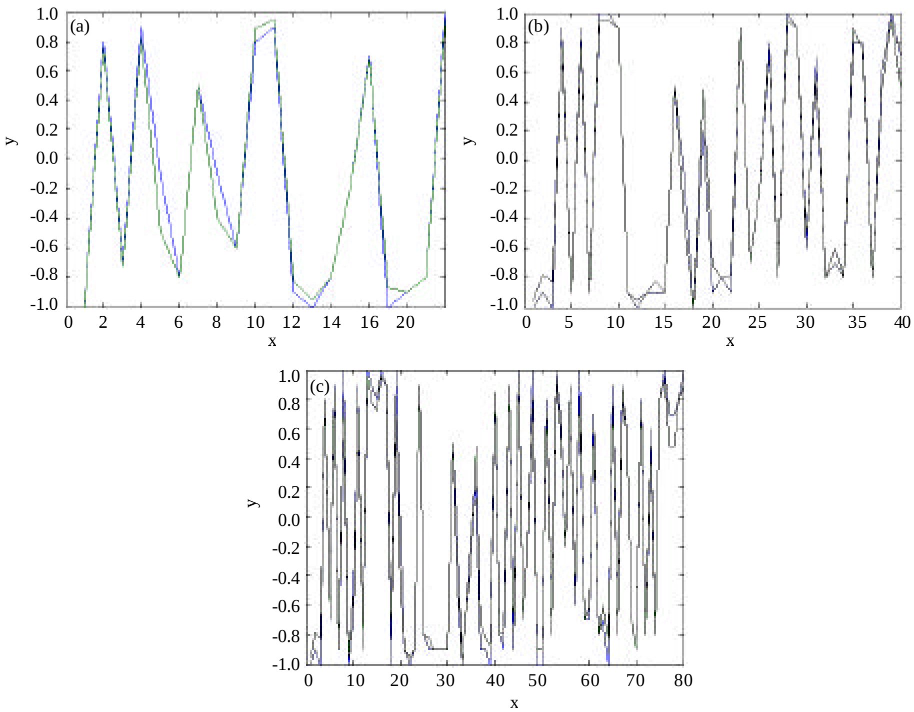

Fig. 6(a-c): Sentiment orientation value calculating results

Table 4: Experiment results for public domain

	Words			
Result	Public sentiment words	Attribute-sentiment word pairs	Adverbs of degree	Negative adverbs
No.	565	177	295	115

sentiment words), 177 attribute-sentiment word pairs, 295 adverbs of degree, 115 negative adverbs. The experiment results for the public domain shown in Table 4.

According to obtaining experimental data, three different groups were randomly selected:

Group 1: 15 public sentiment words, 5 pairs of attribute-sentiment word pairs and its related adverb

Group 2: 25 public sentiment words, 15 pairs of attribute-sentiment word pairs and its related adverb

Group 3: 50 public sentiment words, 30 pairs of attribute-sentiment word pairs and its related adverb

Each combination of these three groups of test sample was scored sentiment and scoring ranges from -1 to +1. Then, the scoring results were respectively stored in the three word sets. According to Eq. 1, 2 and the specific location of each combination, its sentiment orientation value was automatically calculated and the calculating results were stored in the word sets. The experimental results are shown in Fig. 6.

The abscissa x axis represents the number of sample sets, respectively, set of 20, 40, 80 discrete points. The ordinate y axis indicates the sentiment orientation value of each sample set. The green curve shows the experimental results, the blue curve represents the scoring results by manual. From the figure, the experimental comparison shows that there are more discrete points, there are better experimental results.

CONCLUSION

Sentiment computing is a crossover research area including natural language processing, psychology, linguistics and cause people more and more attention in recent years. As a result, it becomes a hot research problem practicality in the field of information retrieval and natural language processing. This paper proposed the method of sentiment orientation analysis based on sentiment lexicon oriented domain according to the domain characteristics of sentiment orientation analysis. The sentiment lexicon is constructed by ontology technology and automatically extended by learning model of ontology. Through experiments, the sentiment orientation analysis method proposed in this paper is conducive to control the scale of sentiment words in the lexicon to improve the sharing of the sentiment words, provide reliable support for the domain sentiment analysis.

ACKNOWLEDGMENT

This study is supported by Natural Science Fund of Zhejiang Province, P. R. China (No. Y1110995, Z1110551), Science and Technology Department of Zhejiang Province P. R. China (No. 2011C14018, 2011C21050, 2010R50041, 2012C21008), Education Fund of Zhejiang province, P.R.China (No. Y201223419), Humanity and Social Science on Young Fund of the Ministry of Education (No. 12YJC630170).

REFERENCES

Durant, K.T. and M.D. Smith, 2006. Mining sentiment classification from political web log. Proceedings of the Workshop on Web Mining and Web Usage Analysis of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 20, 2006, Philadelphia, PA.

Hu, M. and B. Liu, 2004. Mining opinion features in customer reviews. Proceedings of the 19th National Conference on Artificial Intelligence, July 25-29, 2004, San Jose, California, pp: 755-760.

ICTCLAS, 2008. Introduction. http://ictclas.org/index.html

Lin, H. and S.H. Guo, 2003. On the characteristics, range and classification of adverbs of degree. J. Shanxi Univ. (Philosophy Soc. Sci.), 26: 71-74.

Pang, B. and L. Lee, 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, June 17, 2005, Michigan, pp: 115-124.

Pang, B., L. Lee and S. Vaithyanathan, 2002. Thumbs up? Sentiment classification using machine learning techniques. Proceedings of the Conference on Empirical Methods on Natural Language Processing, July 6-7, 2002, Philadelphia PA., USA., pp: 79-86.

Riloff, E. and J. Wiebe, 2003. Learning extraction patterns for subjective expressions. Proceedings of the 2nd Conference on Empirical Methods in Natural Language Processing, July 11-12, 2003, Sapporo, Japan, pp: 105-112.

Tsou, B.K.Y., R.W.M. Yuen, O.Y. Kwong, T.B.Y. Lai and W.L. Wong, 2005. Polarity classification of celebrity coverage in the Chinese Press. Proceeding of the International Conference on Intelligence Analysis, May 2-6, 2005, McLean VA., pp: 1-6.

Wang, G. and J. Zhao, 2007. Sentence sentiment analysis based on multi-redundant-labeled CRFs. J. Chin. Inform. Process., 21: 51-58.

Whitelaw, C., N. Garg and S. Argamon, 2005. Using appraisal groups for sentiment analysis. Proceedings of the 14th ACM International Conference on Information and Knowledge Management, October 31-November 5, 2005, ACM Press, pp: 625-631.

Wiebe, J., R. Bruce, M. Bell, M. Martin and T. Wilson, 2001. A corpus study of evaluative and speculative language. Proceeding of the 2nd ACL SIG on Dialogue Workshop on Discourse and Di-alogue, Volume 16, September 2001, Aalborg, Denmark, pp: 1-10.