

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSI*net*

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

Construction and Automatic Expansion of Domain-oriented Sentiments Lexicon

Hong Liu, Xiaojun Li and Hanxiao Shi

College of Computer and Information Engineering, Zhejiang Gongshang University Hangzhou,
310018, People's Republic of China

Abstract: Analyzing sentiments behind natural language processing is the focus of current research and the construction of sentiments lexicon is the foundation to analysis sentiments. This study was focused on domain features of sentiments tendency. By analyzing comprehensively available lexicons and introducing ontology technology, this resulted in a framework for the construction and automatic expansion of domain-oriented sentiment lexicon, which solved fundamental problems between the construction and automatic expansion of sentiments lexicon and the integration of lexicon application. The proposed framework reflected extremely fine characteristics of the domain feature. Experimental results showed that the proposed framework was effective in the automatic evaluation and layed a sound foundation for the domain-oriented sentiment tendency analysis.

Key words: Sentiment lexicon, ontology, sentiment tendency, corpus

INTRODUCTION

In recent years, with the rapid development of China's e-commerce, more and more subjective review articles are published for certain products or merchants, which to some extent helped potential customers make the right decision and provided feedback to the merchants. Manufacturers are in a great need of the first-hand customer evaluation, which provides information to improve product quality or services (Hanxiao and Jun, 2011). On the other hand, Internet has created a new communication channel and plays an increasingly important role in public opinion. Keeping abreast of and actively guiding the public opinion has become an important part of government work (Yang *et al.*, 2011).

Sentiment tendency analysis is based on such practical demands, which automatically extracts the information contained in the network comments. Effective analysis and data mining is executed to identify the sentiment tendency, such as "agree" or "disagree", which provides in-time and accurate decision support to various domains, including governmental monitoring and guidance on public opinion, business market research and customer shopping references (Shi *et al.*, 2009). Sentiment tendency analysis has become a research hotspot of natural language processing, while most text tendency analysis methods are based on sentiment lexicons, where sentiment words, namely the basic unit, is the foundation for analyzing sentence or chapter tendency. Therefore, the construction of sentiment lexicon will directly influence the effectiveness of text tendency analysis.

Some studies are carried out on the construction of sentiment lexicon, which directly adopts available sentiment lexicons and corpus resources or derives polarity lexicons for specific applications. Baccianella *et al.* (2009) adopted General Inquirer (GI) to construct viewpoint lexicon, which comprises of word categories and vocabularies. Each vocabulary carries appraised meaning, while for words with more than one meaning, the dictionary will separately list out different entries to distinguish the appraised meaning for certain definitions or categories. Wiebe and Mihalcea (2006), directly used the corpus WordNet and MPQA to process sentiment analysis. Gyamfi *et al.* (2009) used MPQA to construct the sentiment lexicon and combines it with WordNet to expand the sentiment lexicon. Based on SentiWordNet, Devitt and Ahmad (2007) and Esuli and Sebastiani (2007) performed the sentiment polarity identification and viewpoint mining. The method to generate sentiment lexicons has gradually shifted from relying solely on manual (Das and Chen, 2001) to the semi-automatic (Hu and Liu, 2004; Yu and Hatzivassiloglou, 2003) and for some systems, if accuracy is not fully taken into consideration, automatic generation (Turney and Littman, 2003; Zhi Ming *et al.*, 2005) cannot be realized.

In China, the construction of sentiment lexicon has also been initiated. Wei Ping *et al.* (2009) extracted the basic sentiment words from HowNet to construct the sentiment lexicon. Since, the underlying sentiment lexicon is based on HowNet, its accuracy strongly depends on the HowNet sentiment words set. For a word with

unknown sentiment tendency, the sentiment tendency is inferred from a word set of known polarity through searching for similar semantic words in the dictionary. The Students' Dictionary of positive and negative vocabulary compiled by Wei *et al.* (2004). Marks each word clearly with its sentiment tendencies. Furthermore, sentiment tendencies of the same word in different categories are also clearly indicated and synonyms with identical appraise properties are listed. The HowNet, developed by Zhendong *et al.* (2007), divides vocabulary into words with varying degrees i.e., negative evaluation words, negative sentiment words, positive evaluation words, positive sentiment words and declaration words. In addition, there are Dictionary of Positive Vocabulary edited by Jilin and Gui (2005) and Dictionary of negative vocabulary edited by Ling and Gui (2005). Traditional semantic dictionary, such as WordNet, HowNet, are mostly manually compiled, which requires a lot of time and manpower. The maintenance cost is extremely high and the vocabulary is unable to meet the specific requirements of certain professional field. On the basis of these semantic resources, automatic or semi-automatic semantic lexicon technology for different purposes emerged.

Researchers have put a lot of energy and efforts into the construction of sentiment lexicons. Due to the lack of engineering methods and the complexity of sentiment analysis, complex analytical reasoning is difficult to be performed on the machines with no sentiment identification ability. Besides, the sentiment identification of the same word in different domains sometimes does not work well and may even result in misunderstanding. For example, the adjective "long" in "long battery life" is clearly positive and in many industries, such as mobile phones, computers and other digital industry, it represents superior product quality and advantageous performance in the product market competition. But on the contrary, a program with long response period, long re-boot time, actually means the opposite. Therefore, the conclusion can be drawn that "long" can be either positive or negative. Furthermore, this word can also be neutral in the following sentence, "After the winter solstice, day time becomes long and night time becomes short" (Hanxiao and Jun, 2011; Jiang *et al.*, 2011). So, the same word or words could hardly be interpreted properly through a simple and basic sentiment lexicon, such as HowNet, MindNet and etc., since these dictionaries can no longer apply to current complex sentiment changes as well as the appraisal tendency for a certain word or a paragraph. If the outdated simple and rough reasoning method is adopted, then the results would also be rather rough and unsatisfactory. Thus, the lexicon requires more manual input and the sentiment lexicon cannot function

practically thus, becoming useless. It is therefore, necessary to build a more detailed, more specific, more comprehensive and more humanized sentiment lexicon. A single word in the lexicon should not be limited only to its absolute appraised meaning; it should have more semantic space and a richer meaning, which will lead to higher accuracy and convenience (Zhi Ming *et al.*, 2005; Wei Ping *et al.*, 2009). A word will be fast and accurately judged as positive or negative according to its respective use in different industries or topics. For some emerging words, they can be divided into the proper categorizations after simply analyzing the popular trend of these terms.

Based on the analysis of existing construction methods of sentiment lexicon, this study introduces ontology technology to solve the fundamental problems of the construction and automatic expansion of sentiment lexicon. A framework for the construction and automatic expansion of domain-oriented sentiment lexicon is proposed, which would facilitate the analysis and understanding of the consumer sentiment and behavior changes under the network environment, contribute to the understanding of the sentiment expressed through consumer language and realize the intelligent personalized product recommendation.

FRAMEWORKS FOR THE CONSTRUCTION AND AUTOMATIC EXPANSION OF DOMAIN-ORIENTED SENTIMENT LEXICON

The proposed framework for the construction and automatic expansion of domain-oriented sentiment lexicon is shown in Fig. 1. Which includes pre-processing module, ontology construction module, ontology model, automatic expansion module and user interface.

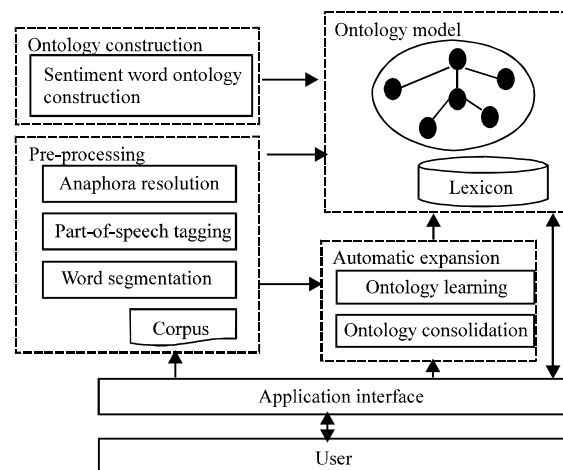


Fig. 1: The framework for the construction and automatic expansion of domain-oriented sentiment lexicon

The user submits the application request to the interface and then the application interface calls out different modules based on the types of the application. If the application is to identify sentiment words, then, the interface interacts directly with the ontology module for ontology matching. If the match is unsuccessful, the sentiment words would be submitted to the automatic expansion module. If the application is for sentiment analysis, then the pre-processed corpus would be combined with the lexicon for further sentiment tendency analysis and the results will be provided to the user through the user interface.

Pre-processing module: Pre-processing the corpus extracted from a web page, including word segmentation processing, noise handling, part-of-speech tagging and anaphora resolution. 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%. Anaphora resolution, including named entity identification, noun phrase identification and the noun phrase extraction, is carried out on the pre-processed materials. To improve the effectiveness of anaphora resolution, certain rules are set, such as the consistency of singular/plural forms and the uniformity of noun phrase gender, thus, the materials obviously not in conformity to these rules are filtered out and the scope of candidate word is narrowed. Then, the feature vector is extracted (the feature vector set is based on the twelve basic features proposed by Soon *et al.* (2001) and the items to be resolved are determined. Finally, the generated classifier based on machine learning methods predicts the items to be resolved, inferring the co-reference relations between these nouns. The results can be used to restore the pronoun in the sentence, thereby enhancing the sentiment words identification and the efficiency of sentiment analysis.

Ontology construction module: Performs the comprehensive analysis on the pre-processed corpus, combined with available sentiment corpus, such as the sentiment word set HowNet and the synonym lexicon compiled by Harbin Institute of Technology. The sentiment words with the highest frequency of occurrence and relative polarity of certain domain are collected as the seeds of sentiment lexicon. The lexicon ontology is constructed and the polarity of each seed of the sentiment words is determined and eventually stored into the lexicon.

Ontology model: Represents the model of the sentiment lexicon, including the ontology mapping algorithm.

Automatic expansion modules: Extends the sentiment lexicon with the support of ontology technology and implements the algorithm on the ontology expansion and consolidation of the sentiment words.

User interface: Different user requests are submitted through the interface, such as the accessing, browsing, querying of the tendency of the targeted objects and then the interface passes the request to different modules for processing and provides the results to the user.

DOMAIN-ORIENTED SENTIMENT LEXICON

Through analyzing the characteristics of the domain corpus, the sentiment words of various domains can both be different and overlapped. Some sentiment words with definite tendencies, such as the words “good”, “wonderful” and “beautiful” are clearly positive, while the words “poor”, “ugly” and “awful” are definitely negative. These words have the same sentiment tendencies even in different domains, so, they could be categorized as non-specific domain words. On the contrary, some sentiment words appear in domain A, but do not appear in domain B, so these sentiment words are categorized as domain-specific. Besides, some sentiment words appear both in domain A and B but have different degrees of sentiment tendencies and these sentiment words are considered as more significantly domain-specific. Taking into account the importance of varied degree adverbs and negative adverbs, sentiment analysis should pay much attention to the roles played by these adverbs.

Based on the above analysis, the structure of the proposed domain-oriented sentiment lexicon is shown in Fig. 2.

Definition 1: Sentiment Lexicon: Composed by the sentiment words with determined polarity (denoted by S), the domain sentiment words (denoted as $D = \{D_1, D_2, \dots, D_n\}$), modifiers (denoted as T) and corresponding domain characteristics. In this sentiment lexicon, domain-specific and domain-shared concepts, the properties and the relationship of these concepts and the relationship between these properties are definitely distinguished. The sentiment words with determined polarity and modifiers are designed as the general ontology, while domain-specific sentiment words are designed as the domain ontology and the domain characteristics are designed as application ontology. The design process fully embodies the characteristics of the knowledge consolidation and sharing.

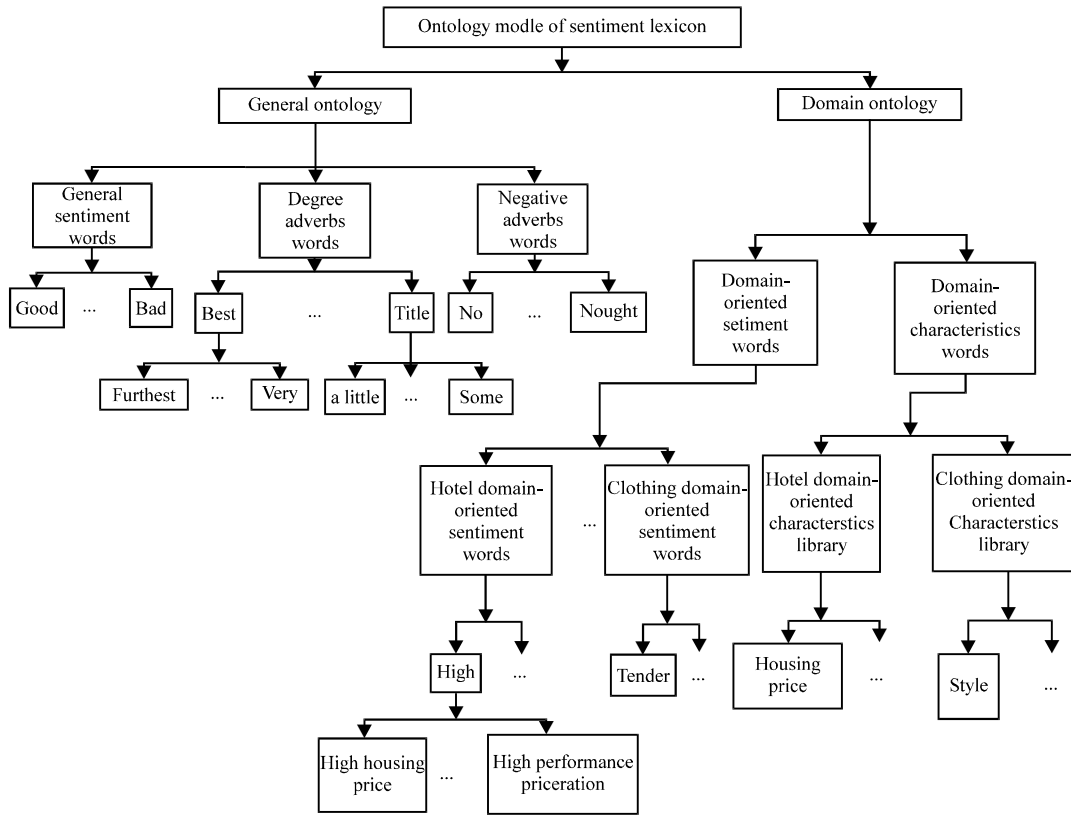


Fig. 2: The structure of the domain-oriented sentiment lexicon

Definition 2: Ontology structure: Described as a triple $O = \{C,P,R\}$, in which C is the concept set, P is the attribute set and R is the semantic relationship set of concepts.

Definition 3: Concept set of domain characteristics: Represents the set of domain characteristics, denoted as $A = \{I\}$, where I hierarchically describe the categorizing system of the domain characteristic knowledge, denoted as $I = \{characteristic_1, characteristic_2, \dots, characteristic_n\}$, in which $characteristic = \{chid, cname\}$. For example, $\{“1.2”, “ Delivery speed”\}$ means that in the domain characteristics system, the characteristics “delivery speed” is coded as 1.2. For the concept set of apparel industry, there are concepts such as $A = \{goods \{price, workmanship, size, material, style \{design, processing design, version, the details, waist type, length, sleeve length, pant length, collar type, style\}, color\}$ and business $\{packaging, delivery speed, service attitude\}$, logistics $\{service attitude of the courier, delivery speed, rate of the goods in good condition\}$. In the sentiment lexicon, initiated from the sentiment words’ perspectives, the link between sentiment words and domain characteristics is established, while the domain characteristics is considered as a sub-property of the

sentiment words. When domain sentiment words are used to perform sentiment analysis, the characteristics is distinguished at first and then the sentiment words are identified, finally the polarity of the sentiment words is jointly determined.

Definition 4: The subset of the sentiment words with polarity determination: $S = \{sentiment_1, sentiment_2, \dots, sentiment_n\}$, which includes polarity determined sentiment words from non-specific domain, such as the positive words “good”, “wonderful” and “beautiful”, or the negative words “poor”, “ugly” and “awful”. The characteristics of this set of sentiment words is denoted as $sentiment_i = \{ID, word, intensity|intensity \in (0,1)\}$, where ID represents the code of the sentiment words, $word$ represents the sentiment words and $intensity$ denotes the sentiment strength.

Definition 5: The subset of modifiers: $T = \{modifier_1, modifier_2, \dots, modifier_n\}$, which includes degree adverbs and negative adverbs. The property of the modifier is denoted as $modifier_i = \{ID, word, degree|degree \in (0,1) \text{ or } -1\}$. Since, degree adverbs are especially important for sentiment expressing, this paper adopts the classification method proposed by Lin Huang and classify the adverbs

into four degrees, namely the extreme volume, high volume, medium volume and low volume. Thus, every adverb is assigned a degree coefficient and once a sentiment word appears with a certain degree adverb, the polarity value is calculated as the polarity value of the sentiment word multiplied by the degree coefficient. For the processing of negative adverb, the polarity value will further be multiplied by -1 because of the negative adverb.

Definition 6: The subset of the domain-specific sentiment words: $D = \{\text{sentiment}_1, \text{sentiment}_2, \dots, \text{sentiment}_n\}$, which includes the domain-specific sentiment words with undetermined sentiment tendency. The sentiment words are mainly classified into positive and negative categories. Since, the identical sentiment word expresses different sentiment tendency when describing different characteristics of the same domain, such as for laptop domain, comments such as “The long battery life” is positive and “The long response time of the program” is negative. For these sentiment words with dynamic polarity, id and ordered pairs are used to process, specifically expressed as:

$$\text{Sentiment} = \{ \text{id}, \text{word}, \langle \text{chid}_1, \text{intensity} \rangle, \langle \text{chid}_2, \text{intensity} \rangle, \dots, \langle \text{chid}_n, \text{intensity} \rangle \mid \langle \text{chid}_i, \text{intensity} \rangle \in I \}$$

where, chid_i corresponds to the code of application characteristics. The sentiment word “long” is depicted as $\{ \text{“002”}, \text{“long”}, \langle \text{“1.1”}, 0.4 \rangle, \langle \text{“2.1”}, -0.5 \rangle \}$, where the sentiment word “long” is coded as “002”. When the word describes the feature of the “battery” coded as 1.1, its polarity strength is 0.4, while it is used to describe the features of the “program” coded as 2.1 and the polarity strength becomes -0.5.

Definition 7: Property set of the lexicon: There are mainly two categories of properties: Object properties and datatype properties, such as the above-mentioned word “long”. For “battery”, “long” is used to describe the object, which is classified as object properties and the polarity value of the sentiment words is classified as numeric attributes.

Definition 8: Set of semantic relation between concepts: The semantic relation between concepts in the lexicon includes synonym relations and context relation.

THE MECHANISM OF AUTOMATIC LEXICON EXPANSION

In the process of sentiment analysis, ICTCLAS 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 characteristic, characteristic 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. 3.

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

- Step 1:** If $t_i = (i = 1, 2, \dots, n)$ in the set T matches directly with the classes in the ontology, the corresponding concept $c_j = \{c_1, c_2, \dots, c_n\}$ is right added into the set C
- Step 2:** If $t_i = (i = 1, 2, \dots, n)$ in the set T not only matches directly with the classes in the ontology, but also matches with the other classes or the instances of that class, then the principle is that class has the highest priority, instance the second, followed by property with the lowest priority. The class is the output of the matched concept; otherwise, the class with the property is added into the set C
- Step 3:** If $t_i = (i = 1, 2, \dots, n)$ in the set T matches with a certain individual in the ontology but doesn't match with other classes or properties, then the

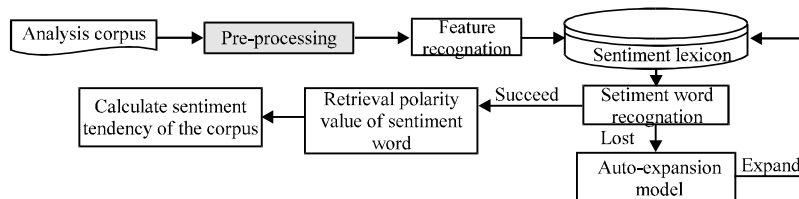


Fig. 3: The process of sentiment analyzing

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

Step 4: 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, Qiu *et al.* (2009) proposed the expansion method for domain sentiment lexicon based on the principle of "two-way communication". This method accesses the characteristics of a known sentiment word, with which to acquire other characteristics or other sentiment words modified by that characteristic. 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:

Step 1: 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

Step 2: 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

THE DESIGN OF APPLICATION INTERFACE

The application interface receives and parses the user's request, while the results are submitted to different modules for processing. The processing results are returned to the user. The main interface includes:

Step 1: Affective computing: This function calculates the sentiment tendency based on the characteristic level. The interface is defined as follows:

GetCalculationSentimentValue (String sentimentword, String domain, String d-adverbs, String n-adverbs, int position)

where, sentimentword represents the sentiment words, the domain represents the specific domains, d-adverbs is the degree adverbs, n-adverbs is the negative adverbs and position represents the specific position relations among sentiment words, degree adverbs and negative adverbs

Step 2: The addition of sentiment words and characteristics: Semi-automatic expansion of sentiment words and characteristics is performed. The specific interface is defined as follows:

AddSentimentWord (String sentimentword, String domain, boolean flag)
AddFeature (String featureword, String domain, boolean flag)

where, SentimentWord represents the new sentiment words; domain represents the domain that new sentiment words to be added; flag is the mark, true means that the new sentiment words is domain-specific and false represents the general sentiment words

Step 3: The deletion of sentiment words: When a domain-specific sentiment word is judged to be general sentiment word, this word would be removed from the domain-specific set and then added into the general sentiment set. The specific interface is defined as follows:

DeleteDomainWord (String domainword, String domain)

where, DomainWord represents the sentiment words judged to be general and domain indicates the domain it belongs to

INSTANCES

Based on the above-mentioned structure of sentiment lexicon, the sentiment ontology sub-library and the respective characteristics library is constructed for the domain of hotels. Fig. 4 shows the structure of the characteristics library of the corresponding sentiment set of domain "hotel".

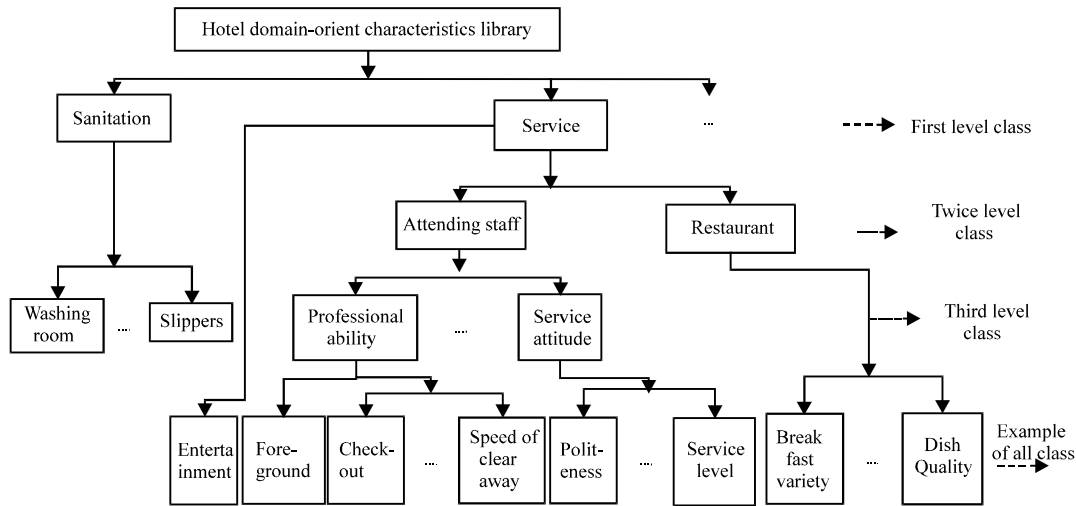


Fig. 4: Ontology hierarchy structure of the characteristics library of corresponding sentiment set of domain “hotel”

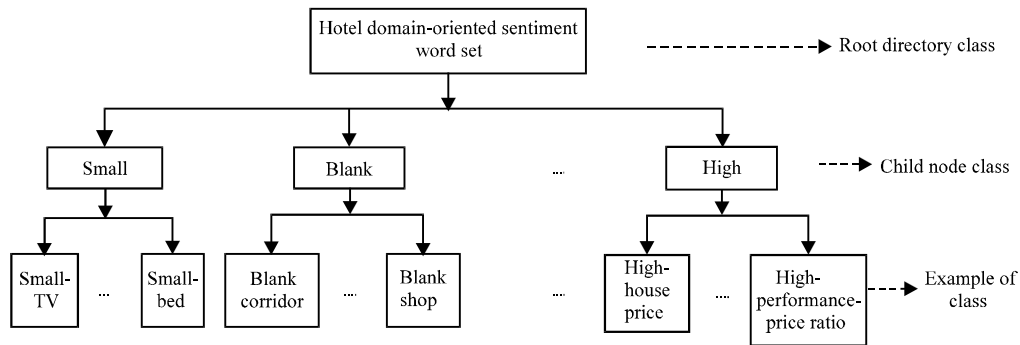


Fig. 5: The hierarchy structure of the ontology sub-library of “hotel” domain

In many domains, there exists the hierarchical semantic inclusion and included relations. Taken the hotel domain as example, “location” can exist both as the characteristics and the characteristics set for the hotel domain. When it is considered as the characteristics set, it can include characteristics such as “Hospital”, “School”, “Urban”, “Supermarket”, “Scenery” or “station”.

From the angle of coarse-grained analysis, “location” could provide a client an initial understanding of the hotel position, but for fine-grained analysis, various customers may have different requirements for hotel “location”. For example, a tourist considering the hotel “location” would focus more on the “Scenery”, a patient may be more focused on the “Hospital” and a general individual with no special requirements would just like to know the hotel “location”.

In addition, the characteristic words may have the same or similar semantic relations, such as “toilet” and “washing room”, “corridor” and “passage”, “standard” and “standard room” 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 characteristics set is designed, the principle is set as that it should proceed from large to small and from coarse to fine. The domain characteristics should 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 characteristic class should avoid intersecting with other characteristic classes or instances
- The order of division should proceed from large to small and from coarse to fine

The hierarchy structure corresponding the sentiment word set of domain “hotel” (Fig. 4) is shown in Fig. 5.

Table 1: The accuracy of the calculated sentiment orientation value

Data set	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$
Result of the method of this present study			
Commendatory set	43.4	69.8	87.2
Derogatory set	47.4	73.8	90.0
Average accuracy	45.4	71.8	88.6
Ku <i>et al.</i> (2006)			
Commendatory set	22.0	43.4	58.7
Derogatory set	17.9	34.9	52.5
Average accuracy	20.0	39.2	55.6

Where α is confidence interval, the sentiment orientation value ranges from -1 to +1

In this structure, sentiment word is the sub-node of the root class directory, while the characteristic properties are the instance of the domain sentiment words. For example, the sentiment word “high” is a sub-node in the structure of the “hotel” domain sentiment word set, while “house price” and “performance-price ratio” are considered as instances of this class and the characteristics “house price”, “performance-price ratio” are considered as class attributes.

Experiment and analysis: The constructed sentiment lexicon is totally included 5734 derogatory words and 1840 commendatory words. The proposed calculation method is applied to quantify these sentiment words after five times of cross-validation and compared with the method of Ku *et al.* (2006). For every sentiment word, three graduates are assigned to make the annotation independently and the averages are taken as the standard value. The quantification results would be compared with the standard value so as to have the corresponding accuracy rate. As each person judges the sentiment intensity subjectively, there would have inevitably differences, so, is allowed a certain amount of calculation error under the premise of consistent polarity. The experimental results are shown in Table 1.

From the results, it can be concluded that the results of derogatory set is better than the commendatory one, mainly because there are significantly more derogatory words than commendatory ones. During calculation, the coverage of derogatory words is also better than that of commendatory words, which results in higher calculation accuracy. With better calculation accuracy of sentiment polarity, the expansion of the sentiment words and the application of ontology will be solidly founded.

CONCLUSION

The interdisciplinary research field of natural language processing, psychology and linguistics, affective computing has increasingly aroused attention in recent years. It has become a research hotspot in the information retrieval and natural language processing. This study proposed the adoption of ontology technology in the construction of the sentiment lexicon,

fully taking into account the domain characteristics of the sentiment words. The results show that the proposed framework is good at controlling the scale of the sentiment words in the lexicon, improving the sharing of the sentiment words and providing a reliable support to the domain-oriented sentiment tendency analysis.

LIMITATION

The sentiment lexicon proposed in this study still has some limitations: The weights to calculate sentiment polarity needs further improvement and testing, the sentiment lexicon needs to be improved to extend the domain coverage and the sentiment lexicon should be adopted to further explore the sentiment tendency analysis method.

ACKNOWLEDGMENTS

This study is supported by Natural Science Fund of Zhejiang Province, People’s Republic of China (No. Y1080565, Z1110551), supported by Education Fund of Zhejiang province, People’s Republic of China (No. Y201017626) and also supported by Science and Technology Fund of Zhejiang province, People’s Republic of China (No.2011C23075).

REFERENCES

- Baccianella, S., A. Esuli and F. Sebastiani, 2009. Multi-facet Rating of Product Reviews. Springer Press, France, pp: 461-472.
- Das, S. and M. Chen, 2001. Yahoo! for Amazon: Extracting market sentiment from stock message boards. Proceedings of the 8th Asia Pacific Finance Association Annual Conference, July 22-25, 2001, Bangkok, Thailand.
- Devitt, A. and K. Ahmad, 2007. Sentiment polarity identification in financial news: A cohesion-based approach. Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, June 23-30, 2007, Prague, Czech Republic, pp: 984-991.
- Esuli, A. and F. Sebastiani, 2007. Page-ranking wordnet synsets: An Application to opinion mining. Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, June 23-30, 2007, Prague, Czech Republic, pp: 424-431.
- Gyamfi, Y., J. Wiebe, R. Mihalcea and C. Akkaya, 2009. Integrating knowledge for subjectivity sense labeling. Proceedings of the Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, May 31-June 5, 2009, Boulder, Colorado, pp: 10-18.

- Hanxiao, S. and L.X. Jun, 2011. The research on the sentiment tendency analysis method on subjective sentences. *J. China Soc. Sci. Tech. Inform.*, 30: 522-529.
- Hu, M. and B. Liu, 2004. Mining and summarizing customer reviews. *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data*, August 22-25, 2004, ACM Press, Washington, USA., pp: 168-177.
- Jiang, Y., P.S. Yu and H. Min, 2011. Recognizing sentiment polarity in Chinese reviews based on topic sentiment sentences. *Appl. Res. Comp.*, 28: 569-572.
- Jilin, S. And Z.Y. Gui, 2005. *Dictionary of Commendatory Vocabulary*. 1st Edn., Sichuan Dictionary Publishing House, Chengdu, Sichuan, China.
- Ku, L.W., Y.T. Liang and H.H. Chen, 2006. Opinion extraction, summarization and tracking in news and blog corpora. *Proceedings of AAAI-2006 Spring Symposium on Computational Approaches to Analyzing Weblogs*, March 27-29, 2006, Boston, Massachusetts, pp: 100-107.
- Ling, Y. and Z.Y. Gui, 2005. *Dictionary of Derogatory Vocabulary*. 1st Edn., Sichuan Dictionary Publishing House, Chengdu, Sichuan, China, ISBN: 9787806821909.
- Qiu, G., B. Liu, J. Bu and C. Chen, 2009. Expanding domain sentiment lexicon through double propagation. *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, July 11-17, 2009, San Francisco, CA, USA, pp: 1199-1204.
- Shi, L., Y. Qiang and L.Y. Jun, 2009. Mining features of products from Chinese customer online reviews. *J. Manag. Sci. China*, 12: 142-152.
- Soon, W.M., H.T. Ng and D. Lim, 2001. A machine learning approach to coreference resolution of noun phrase. *Comput. Linguistics*, 27: 521-544.
- Turney, P.D. and M.L. Littman, 2003. Measuring praise and criticism: Inference of semantic orientation from association. *Proc. ACM Trans. Inform. Syst.*, 21: 315-346.
- Wei Ping, L., Z.Y. Hui and S.C. Liang, 2009. Research on building Chinese basic semantic lexicon. *J. Comp. Appl.*, 29: 2875-2877.
- Wei, Z., L. Jin and G.X. Zhen, 2004. *Students Dictionary of Commendatory and Derogatory Vocabulary*. 1st Edn., Encyclopedia of China, Beijing, China.
- Wiebe, J. and R. Mihalcea, 2006. Word sense and subjectivity. *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics*, July 17-21, 2006, Sydney, Australia, pp: 1065-1072.
- Yang, C., F. Shi and W.D. Ling, 2011. Analysis on web public opinion orientation based on extending sentiment lexicon. *J. Chinese Comp. Syst.*, 31: 691-695.
- Yu, H. and V. Hatzivassiloglou, 2003. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. *Proceedings of the Conference on Empirical Methods in Natural Language Processing, (CEMNL'03)*, USA., pp: 129-136.
- Zhendong, D., D. Qiang and H.C. Lin, 2007. Theory finding of HowNet. *J. Chinese inf.*, 21: 3-9.
- Zhi Ming, W., Z.X. Feng and Y.Y. Wen, 2005. Research on lexical emotional evaluation based on the grammatical knowledge-base of contemporary Chinese. *Comp. Linguis. Chinese Language Process.*, 10: 581-592.