Network review emotion mining based on Domain Ontology

XiaoYan Chen 1*, ShuFang He 1b

1School of management, Jiangsu University, Zhenjiang 212013, China

*Corresponding author e-mail: 20204568@qq.com, b1255118135@qq.com

Abstract: In the era of e-commerce, network reviews have become one of the most important means for people to exchange and obtain product information. It is an urgent need for consumers and enterprises to quickly and effectively obtain consumers' emotional attitude towards products from a large-scale network review. In order to better extract the product attributes involved in the network reviews and the emotional attitude of the consumer to the product attributes, this paper proposes a method of online comment sentiment analysis based on domain ontology. This method constructs the domain ontology, uses this ontology to extract product attributes, and constructs a domain sentiment dictionary to realize the emotional calculation of product attributes.

1 Introduction

The rapid development of information technology has laid a more solid foundation for online shopping. Online shopping is more convenient and quicker than traditional shopping. Consumers can complete the selection, payment, and distribution of goods without going out of the house which saves a lot of time and energy. However online shopping cannot personally experience and perceived , and consumers can not face to face determine the quality of the goods and whether it is suitable for themselves. At the same time, it is difficult to distinguish the authenticity of the network information, which has brought some drawbacks and limitations to its development. As a result, more consumers tend to browse, analyze product reviews, and share their shopping reviews when doing online shopping that makes the online commentary contains views and opinions of consumers on various methods of the product. Positive online commentary can help companies establish a good brand image and attract consumers' attention. On the contrary, negative online commentary will weaken or even reduce the willingness of consumers, aggravate product deterioration, and affect the normal business and even long-term development of the company. Therefore, filtering emotion recognition of the company's online review information has become a real demand for consumers and businesses.

Sentiment analysis, also known as opinion mining, refers to processing the text and analyzing the polarity and affective attitude of the user's viewpoint contained in the text. In document-level sentiment analysis, Pang et al.[1] used PMI information and SVM method to learn corpus, and constructed sentiment classifier. Turney [2] used the point mutual information method to calculate the emotional polarity of adjectives and adverb phrases, then make a weighted summation to judge the overall emotional inclination of the whole article. Taboada et al.[3] based on sentiment dictionary to identify and match, and combine inversion and intensity calculation to analyze the text's emotional tendencies. In terms of sentence level emotion research, Meena et al.[4] consider the effect of conjunctions on sentence structure based on the syntactic dependency tree, and analyzed the
commentary sentences in turn to achieve the emotional classification. Hu et al. [5] used emotional seed words to construct an emotional dictionary, match the words in the sentence, then calculate the emotional tendency according to the number of words matched. With the development of emotion analysis, scholars have studied word level based fine-grained sentiment analysis. Xianghua F and others [6] combined the LDA topic model and the HowNet dictionary's unsupervised method to automatically realize the multi-directional sentiment analysis. H.zhen et al. [7] used the association rule mining method to identify implicit attributes and display attributes respectively. Zhang et al. [8] used the method of hidden layer dependency tree to identify product characteristics based on parsing tree. Herunxin et al. [9][10] have present a theoretical formulation, and a corresponding optimal control numerical algorithm that achieved good results.

With the large-scale increase of online reviews and the diversification and individualization of product attributes, consumers often evaluate products from various aspects with various degrees of satisfaction when they make comments, so document level and sentence level emotion analysis can not meet the needs. Fine-grained emotional research has more practical value, but the extraction of product attributes in fine-grained sentiment analysis still has some incompleteness and inaccuracy. Therefore, this paper proposes a method of online comment sentiment analysis based on domain ontology, through domain ontology to extract product attributes, analysis the sentiment of each attribute to further improve the accuracy of texts.

2 Product attribute extraction based on domain ontology

2.1 Domain ontology construction

Domain ontology gives a set of domain-related concepts which expresses the semantic relations and hierarchical structure among concepts, and realizes knowledge sharing and reuse[11]. Therefore, this paper defines the basic concepts and attributes of the product field by constructing the domain ontology to provide the standard interpretation and semantic information for the description of the attribute features, and then makes a unified carding of the relation between the attributes and the concepts, provides a common understanding for the knowledge of the product field. The seven-step method is the most popular method of building the ontology. This article improves on the basis of it, taking the mobile phone as an example, from three aspects of the core concept, the determination of the class and the relationship between the concepts, the construction of the ontology to explain.

(1) Obtaining the core concepts of the domain

The concept terminology in the mobile phone domain is highly professional and contains a large number of domain proper nouns and terms, which requires the participation of domain experts, and the independence between terms is strong that reflects the relationship between categories to some extent. In addition, products are mainly faced with consumers so that part of the concept terminology mainly comes from consumers which from consumers to consumers. By summing up, the core concepts of this field come from three main aspects: product manuals, online reviews, and existing related ontologies. The product specification is a main introduction to the function and performance parameters of the product which covers the attributes and relationships of all aspects of the product with a strong professionalism. In the display pages of e-commerce website has given the detailed description of the product parameters, including the functions and attributes of the products. Network reviews are consumers' opinions on products, which embody the product attributes of consumers concerned and contain important product information. The concept words extracted from network reviews are closer to consumers and practical applications. There are some examples of research and construction for mobile phone ontology can be taken as reference.

These three aspects contain a large number of product-related concepts, which are important sources of core concepts in the field that can be obtained through mining, sorting and analysis.

(2) The determination of the class and the acquisition of the relationship between the concepts

According to the core concept set extracted from the product, the classification of product attributes and the relationship between concepts are defined according to the subordinate attributes of the parent
class and the sub class in the ontology. First, the domain attribute set C is classified in the first layer, 
C={T1, T2,...Tn-1, Tn}, as the highest level (first level) attribute under the entity, and at the two level
of the ontology. Secondly, each class of Ti is further classified as Ti={t1, T2,... TN-1, tn}, as the
subclass attribute under this class, and then the next round of class refinement. In this way, according
to the top-down method, the stratification is carried out in order to set up a hierarchical system in order,
until it is no longer classified. After determining the class, then sum up the relationship between the
semantic relations and the concepts in the ontology, and define the relationship in the ontology,
including the relationship of Subclass-Of, Part-Of, Instance-Of and Attribute-Of.

(3) The construction of Ontology

The Ontology is built by the tool Protégé developed by Stanford University which can provide the
construction of ontology concepts, relations and attributes. Then Select the visualization tool
Graphviz-2.20.2 to visualize the ontology, and automatically generate the representation of the OWL
language.

2.2 Product attribute extraction based on Ontology

(1) Extraction of explicit attributes

The preprocessed online comment text can be regarded as a set of words, matching the words in the
text with the ontology to extract the product attribute. However, because of the complexity and
colloquial of the language, it is impossible to fully agree with the standardized conceptual model in
the ontology. If we simply match the words in the text with the ontology, most of the conceptual words
will be ignored. For the concept that is not found in the domain ontology, use the semantic relation to
identify. There are three cases: (1) the synonym, An attribute word is a synonym for the concept of the
ontology. (2) hyponymy relation, attribute word is the hyponymy word of ontology concept. (3) part
relation, Candidate attribute words are related domain terms that fail to identify.

(2) Extraction of implicit attributes

Because of the casual and informality of network reviews, some online texts do not have clear
commentary objects, such as "It's too vague to see the video like that, no improvement at all but with
high price." does not indicate any product attributes, But according to semantic analysis and inference,
it can be known that "fuzzy" implies the property of "screen". Ontology can not identify such implicit
attributes. For the extraction of such implicit attributes, product features are mapped through a fixed
collocation relation between attribute words and emotional words. For example, in the field of mobile
phones, "clear" and "fuzzy" are generally used to describe the feature of "screen". "Expensive" and
"cheap" are usually used to describe the characteristics of "price". These are called special emotional
words or characteristic demonstrative words, so they can be used to indicate potential product
attributes.

3 Emotion analysis of product attributes based on Domain Ontology

3.1 The construction of the emotional dictionary

The construction of emotion dictionary is the core issue of semantic emotion analysis that used to
extracts the evaluation words of Web text, including emotional evaluation words and modifiers and the
calculate the sentiment tendency based on it. Different emotional words have different effects on
the intensity of emotion and polarity. The completeness of emotion dictionary directly affects the
accuracy of emotion analysis results. There are three types of emotion words: Static emotion words,
dynamic emotion words and modifiers.

The emotional evaluation directly reflects the subjective attitude of the consumer which divided
into two categories, positive and negative, including static emotion words and dynamic emotional
words. The emotional polarity of static emotional words is absolute and does not change with
context. At present, there are general static emotion dictionaries such as Hownet dictionary. On the basis of these dictionaries to form a static emotion evaluation dictionary which more perfection and supplement by adding special emotional words in the field and rejecting the emotional words that are not conforming to the attributes of the dictionary. In addition, there are some special words in the field of product and some network new words, this kind of emotional word needs to be further extracted as a supplement.

However, emotional polarity of dynamic emotional words will change when different evaluation objects are modified, this paper combines network reviews and domain related knowledge to build dynamic affective dictionaries by hand, and marks their emotional polarity by combining their evaluation objects.

Besides emotional evaluation dictionaries, the influence of modifiers on emotional intensity is also very important, mainly including negative words and degree adverbs. The addition of negatives often affect the emotional tendencies of sentences such as “not”. Adverbs of degree mainly play an important role in strengthening and weakening emotion, for example the sentiment of “very beautiful” is more strong than “beautiful”.

To sum up, the emotional dictionary is composed of two parts: the dictionary of emotion evaluation words and the dictionary of modifiers.

3.2 Calculation of emotional tendencies of product attributes

Pre-processing the collected product review data, including putting off the unrelated sentences, sentences segmentation and so on. Then combining the domain ontology and emotional dictionaries to map the evaluation unit of pre-processing comments. The specific steps are as follows:

1) Match the pre-processed comment clause with the constructed domain ontology. If the noun in this clause can match the attribute in the ontology then extract the comment attribute of the sentence, if not, judge whether it is the synonym, the lower word or the new concept of the concept in the ontology, identify, extract and standardize and match of the emotion word in the (3) step. If it is a default clause, enter into the step(2) is to identify the implicit attribute first.

2) After processing the explicit attribute, the implicit attribute is identified by mapping the viewpoint word, and add the attribute to the sentences to make the implicit attribute explicit.

3) For all the sentences that have already identified the comment attribute, the emotion words, negative words, degree adverbs are identified by using the constructed domain sentiment dictionary.

After finished, stored as the evaluation unit of \{F, S (Sd/Sq) and P\}. F is a product attribute, S is an emotional word may be a static emotion word (Sq) or a dynamic emotion word (Sd), and P is a modifier, including the degree adverb and the negative word. According to the extraction evaluation unit, the emotion value of attribute F is calculated by the emotional value of the emotional word(seni) , the number of the negative word(n), the emotion adjustment of the degree adverb(r), as shown in formula (1):

$$\text{Sentiment (attribute } F) = (-1)^n \times \text{seni} \times r$$

4 Experiment and result analysis

4.1 experimental data

Taking JD mall as the source of experimental data and mobile phone as the research object, we collected more than 7000 comments by using Octopus data collector. Preprocessing these data, including data cleaning, segmentation and remove stop words, finally obtain 6213 effective data. The effective experimental data are divided into two parts, part1 is for ontology construction, the part2 is for emotional computing.
4.2 Experimental results and analysis

JD mall has detailed description of product which includes most of the product domain concepts. And extract product attributes from experimental data in part1, refer to the existing ontology, then merge and get the final attribute. Classifying these attributes and defining the relationship between classes and classes based on the properties of classes to construct a mobile phone ontology.

Extracting product attributes from preprocessed part2 comments based on domain ontology, the product attributes based on domain ontology may be synonyms or upper and lower terms of the concept of ontology node, for example, the attribute “system” is the synonyms of “operating system” in related concept of ontology, therefore, this kind of attribute synonyms should be unified as the corresponding concept words in the ontology. For the attribute word which is not in the ontology, it is added to the correct level and position of ontology to make the ontology is more perfected and supplemented.

After extracting the attribute, compute the emotion of attribute based on emotion dictionary and emotion formula, get the emotion value of each attribute, the part of results are shown in Figure1.

![Emotion score of feature](image)

Fig.1 Emotion score of feature

As you can see from the Figure1, users are most satisfied with the attribute of appearance, and more satisfied with the two attributes of logistics and battery, have a moderate attitude towards the two attributes of screen and storage. On the contrary, for the two attributes of Bluetooth and GPS, users are dissatisfied with them and have negative feelings. Through the emotional mining of the product attributes involved in the network review text, we can clearly get the advantages and disadvantages of the products, so as to better understand the the products and assist in decision-making.

5 Conclusion

In view of the low accuracy of attribute extraction in current sentiment analysis, this paper proposes a sentiment analysis method based on domain ontology for web reviews that using the domain ontology to extract product attribute and calculate the attribute emotion by constructing the domain sentiment dictionary. In this paper, the extraction of product attributes depends on the domain ontology constructed which the integrity and accuracy of ontology construction directly affect the research results in this paper. Due to the universality and dynamism of ontology knowledge, the efficiency, standardization and comprehensiveness of ontology construction still need further improvement.
Acknowledgements

This work was partially supported by a grant from Jiangsu Six Talent Peak Projects (No. JXQC-037), and a grand from Jiangsu Research Innovation Plan for Postgraduates (No. CXZZ11_0594).

References

[1] Pang B, Lee L. Opinion Mining and Sentiment Analysis[J]. Foundations and Trends in Information Retrieval, 2008, 2(1-2): 1-135.
[2] Turney PD. Thumbs up or thumbs down: Semantic orientation applied to unsupervised classification of reviews[C]. Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Stroudsburg, PA, USA: Association for Computational Linguistics, 2002: 417-424.
[3] Taboada M, Brooke J, Tofiloski M, et al. Lexicon-Based Methods for Sentiment Analysis[J]. Computational Linguistics, 2011, 37(2):267-307.
[4] Meena A, Prabakar TV. Sentence level sentiment analysis in the presence of conjuncts using linguistic analysis[C]. Proc.of the European Conf.on IR Research, 2007: 573-580.
[5] Hu Minqing, Liu Bing. Mining and summarizing customer reviews[C]/Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, Seattle, WA, USA, August 22-25 2004. New York, NY, USA: ACM, 2004: 168-177.
[6] Xianghua F, Guo L, Yanyan G, et al. Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon[J]. Knowledge Based Systems, 2013, 37: 186-195.
[7] ZHEN H, CHANG K, KIM J. Implicit feature identification via co-occurrence association rule mining[C] / Proceedings of the 12th international conference on computational linguistics and intelligent text processing.Berlin: Springer-Verlag, 2011: 393-404.
[8] Zhang J, Ackerman M S, Adamic L. Expertise networks in online communities:conferenceconstructure and algorithms[C]. In Proceedings of the 16th international world wide web.Banff,Alberta,Canada,2007:221-230.
[9] He Runxin, Gonzalez Humberto. Zoned HVAC Control via PDE-Constrained Optimization[C]/Proceedings of the American Control Conference(ACC). Boston USA: ACC,2016(7):587-592.
[10] He Runxin, Gonzalez Humberto. Numerical Synthesis of Pontryagin Optimal Control Minimizers Using Sampling-Based Methods[C]/ IEEE 56th Annual Conference on Decision and Control (CDC). Melbourne, AUSTRALIA: IEEE CDC,2017:733-738.
[11] Gruber T R. Toward principles for the design of ontologies used for knowledge sharing[J]. International journal of human-computer studies, 1995, 43(5): 907-928.