International Journal of Advances in Computer and Electronics Engineering
Volume: 1 Issue: 1, May 2016, pp. 18–22

Feature Extraction of Opinion Mining Using Ontology

[1]P.Baranikumar, [2]N.Gobi
[1]PG Scholar, Department of Computer Science and Engineering,
Dr. Mahalingam College of Engineering and Technology, Tamil Nadu, India
[2]Assistant Professor, Department of Computer Science and Engineering,
Dr. Mahalingam College of Engineering and Technology, Tamil Nadu, India
[1]suganya.jo92@gmail.com, [2]sekar.vrc@gmail.com

Abstract: *The growth in the web sites and the online businesses induces the need for a better opinion derivation system. To avoid the subjective or biased human opinions, it is always better to have an automatic opinion mining technique. Such techniques are expected to extract the various kinds of value added information from the user's opinion. Among the various opinion mining techniques available, the ontology based opinion mining has a significant contribution to the field. Most of the ontology based opinion mining system uses static ontologies for feature identification. Though there exists few systems using dynamic ontology, none seems to utilize the complete potential. The proposed system discusses a novel approach to dynamically build and update the ontology, based on the domain to which the input reviews belong to. The process of feature identification is the core part, which greatly impacts the effectiveness of the opinion mining system. Thus the above proposed methodology which is believed to be better in the feature identification process utilizing the maximum potential of dynamic ontologies helps in improving the effectiveness of the system.*

Keyword: *Opinion mining; Ontology; Sentimental analysis; Feature extraction;*

1. INTRODUCTION

With the norm of users present data in the era of Web 2.0, increasingly more people have submitted or retrieved individual perspective about products, organizations, or political issues via a variety of Web-based channels such as Blogs, forums, social networks, e-Commerce sites. The content that is produced as a direct consequence of this user participation on websites such as Amazon (http://www.amazon.com) or IMDB (http://www.imdb.com) Due to the problem of information overload manually browsing a large number of consumer reviews publish to the Web may not be feasible, if not totally impossible

Opinion mining (Sentiment Analysis) is a Natural Language Processing (NLP) and Information Extraction (IE) task that aims to obtain the feelings of the writer expressed as positive, negative or neutral opinions by analyzing large number of documents. Opinion mining system combines the techniques of computational linguistics and Information Retrieval (IR).Sentiment Analysis aims to decide the attitude of a speaker or writer with respect to some topic or the overall contextual polarity of a document.

The sentiment analysis can be performed at one of the three levels: the document level, sentence level, feature level. In document level sentiment analysis main challenge is to extract informative text for inferring sentiment of the whole document. The challenge faced by the sentence level sentiment classification is the identification features indicating whether sentences are related to a particular topic. Product features are defined as product attributes or components. Analysis of such features for identifying sentiment of the document is called as feature based sentiment analysis. In this approach positive or negative opinion is identified from the already extracted features.

In recent years, experts from various countries are studying how to obtain a simple and efficient method of building ontology. Through introducing the technology of natural language processing and classification and text mining to ontology construction, they have developed some methods and techniques by using some text corpus of some fields to construct domain ontology.

Ontologies are explicit specifications of conceptualizations, where conceptualizations are simplified summaries of the world we want to represent for some purpose. Domain ontologies that describe vocabularies in the Web Ontology Language (OWL).OWL is a stable specification developed by the Web Ontology Working Group. It is considered a Web standard for industry and academy. Ontology can reflect the nature of the objective things and its external performance and

International Journal of Advances in Computer and Electronics Engineering
Volume: 1 Issue: 1, May 2016, pp. 18–22

the field of artificial intelligence, ontology can help us to get the essential knowledge of things ontology can solve the problem that results of general search engine are not accurate.

2. RELATED WORK

Toutanova.K & Mamming [5] proposed preprocessing the review sentences include POS tagging The tagging involves assigning parts of speech tags to the sentences so that it can be used for easy identification of features and opinion words. The features will be the noun term and opinion words will be the adjective words. Zhao.L & Li.C [6] proposed movie ontology is used to find feature of movie and ontology construction based on Select the relevant sentences including conceptions and extract the conceptions from those sentences.

Ghiassi.M & Skinner's [8] proposed basically set of co-occurring words with in the given window and when computing N-grams you typically move one word forward.
Pang .B & Lee L [9] proposed Identify object features that have been commented on and determine whether the opinions on the features are positive, negative or neutral. Baccianella.S&Esuli[10].A proposed opinion lexicons that assign to each synset of WordNet with any one of the three sentiment scores, positive, negative or objective. In the authors describe how positive, negative and objective are the words contained in the synset. SentiWordNet lexical resource uses various techniques to the problem of automatic sentiment classification of reviews.

Zhou and Chaovalit [11] proposed the use of ontologies has the potential to refine and improve the process of sentiment analysis by identifying properties and relations between concepts. In the architecture, reviews extracted from the Internet are preprocessed. After, each review is parsed in order to extract and map text segments according to the ontology. Finally, a polarity orientation is generated for each text segment and for the text as a whole. The advantage of the use of domain ontology is that it provides detailed topic-specific information.

Seven Steps method [7], developed by the college of medicine, Stanford University, was mainly used in the domain ontology construction. The main steps are as follows: (1) Determine ontology scope; (2) Consider to reuse the existing ontology; (3) List the terms of the field; (4) Define the classification, including the hierarchical relationship between classed and subclasses; (5) Define properties, including the domain and range of properties; (6) Define the facet of properties, which means some special values or features; (7) Instance the classes to complete the ontology

3. ONTOLOGY BASED OPINION MINING

3.1 Preprocessing

NLP tools including a sentence detection component, tokenizer POS taggers, lemmatizers and syntactic parsers has been developed using the Stanford Log-linear Part-Of-Speech Tagger framework.A Part-Of-Speech Tagger is a piece of software that reads text in a particular language and assigns parts of speech to each word (and other tokens), such as noun, verb, adjective, etc. The Stanford Log-linear Part-Of-Speech Tagger is a Java implementation for the log-linear part-of-speech (POS) taggers [5].The researchers have focused on improving its speed, performance, usability, and support for other languages. The Stanford log-linear part-Of-Speech Tagger framework contains trained tagger models for the English language.

3.2 Ontology based Feature Identification

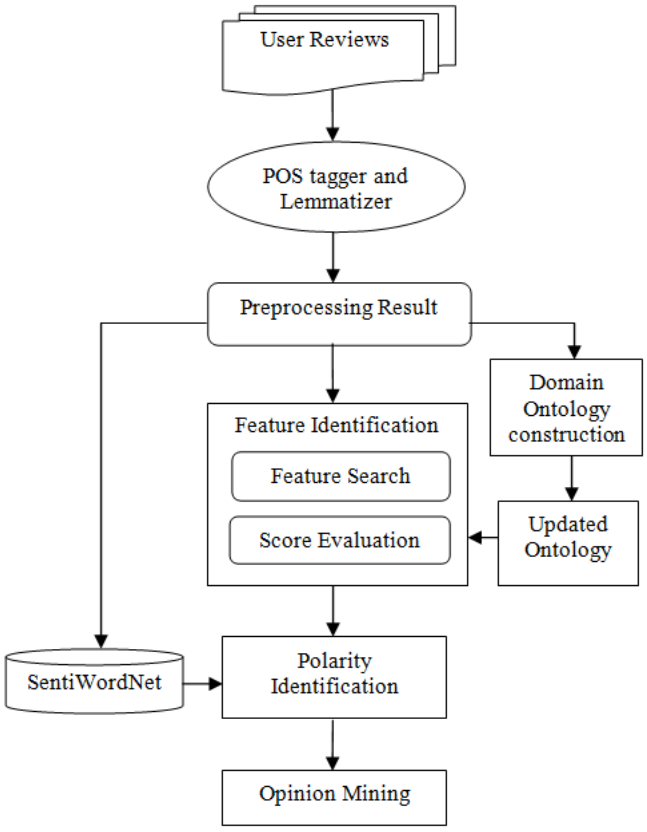

Figure 1 Proposed system architecture

Domain ontology is used in order to extricate the features included in the opinions expressed by users. Once the features in the opinions have been identified, a score that depicts the importance of a given feature is calculated for each of the features retrieved.

In this phase receives both user opinion and the domain ontology as input. The sentence in the user opinion that contains the classes, individuals, datatype and object properties of the domain ontology are then identified from this input. Once the features have been recognized, they are

grouped in following with their semantic distance and are then attached to a main concept of the ontology. The separate score is calculated for each retrieved feature based on following condition as follows:

1. All the features not related to the same class have the same relevance.
2. The features that are most often cited by users in their opinions are more relevant.

left3words-wsj-0-18.tagger

Trained on WSJ sections 0-18 using the left3words architecture, and includes word shape features.

Penn tagset Performance:

96.97% correct with WSJ 19-21

(88.85% correct with unknown words)

Figure 2 Stanford log linear part-of-speech tagger

3.3 Polarity Identification

In The feature polarity is calculated using SentiWordNet 3.0[4]. This tool provides the positive, negative and neutral values of nouns, adjectives and verbs. In our system it is necessary to retrieve this value for all the words that are located near to the linguistic expressions that represent a given feature in the opinion.

The words that are close to the feature can be obtained in a number of different ways. The following four methods have been implemented to evaluate our solution:

N_GRAM Before: this method obtains the N_GRAM words before the linguistic expression of the feature in the user's opinion.

N_GRAM After: this method obtains the N_GRAM words after the linguistic expression of the feature in the user's opinion.

N_GRAM Around: this method obtains the N_GRAM words before the linguistic expression of the feature in the user's opinion and the N_GRAM words after the linguistic expression of the feature in the user's opinion.

All Phrase: this method obtains all the words in the same sentence as the linguistic expression of the feature in the user's opinion.

These methods require the system parameter N_GRAM, which indicates the number of words near the feature that are to be considered in the polarity identification process.

3.4 Opinion mining

After analyzing the corpus of reviews by means of NLP techniques, extracting the relevant features [11] and identifying the features polarity, the framework proposed here provides an innovative opinion mining mechanism. The opinion mining module described in this section is based on vector analysis and enables an effective feature sentiment classification.

In this module determines the user given review is positive, negative or neutral. Since the origin point is always (0, 0, 0), the expression of the position vector is reduced to express the target point. A position vector is therefore expressed by V=(x, y, z). A feature with a strictly positive sense is determined by the position vector (N, 0, 0), where 'N' is a positive number. By analogy, vectors (0, N, 0) and (0, 0, N) represent the strictly negative and neutral senses, respectively.

3.5 Domain Ontology Construction

Dynamic ontology extension [3] in the paper employs text clustering algorithm and Bayesian Classification algorithm. The basic idea is as follows.

1. Perform the text clustering based on K means clustering algorithm and then choose the top five words in each cluster.
2. Classify the candidate set of words based on user supplied training set by using Bayesian classifier.
3. Discover the relationships to update the ontology.
4. Compute the probability for matching features.
5. If the features are not in ontology then add the features
6. Update the Ontology.

Therefore, the frame work uses automatic ontology extension in the paper based on text clustering algorithm and Bayesian classification algorithm will become more and more intelligent. In the subsequent data processing, it will be more accurate.

3.6 Algorithm

The algorithm explains proposed method of ontology construction and opinion mining.

Input: Document D, features = (f1, f2….fn)

1. Begin
2. Do preprocessing
3. Perform POS tagger and lemmatizer
4. PR←get preprocessing result // PR- Preprocessing results
5. For each input perform polarity identification
6. do feature identification {
7. feature search (fs)
8. score evaluation (se)
9. }while(f ≤ features)
10. Perform automatic domain ontology construction
11. Procedure : text clustering()
12. {
13. Apply k-means clustering
14. Select top frequent words
15. Cluster the text into k- clusters
16. }
17. Apply naïve bayes method
18. Compute conditional probability

19. P(d) = $\frac{P(c)P(\frac{d}{c})}{p(d)}$
20. Match the features()
21. Discover the relations
22. If(is, contains, consist of, association of, sub class of)
23. {
24. Updation is not required
25. Else
26. Update the features into ontology
27. Repeat }
28. Text←getSentiWordNetTexts
29. maxWordSimilarity ←0
30. for all text in words do
31. similarities ←getSimilarity
32. return polarity (positive, negative or neutral)
33. End

4 EXPERIMENT AND EVALUATION

In this section, we apply this method in the field of movie. The design goal of movie ontology is to capture the knowledge of the film industry, provide the understanding of movie knowledge, and define the vocabulary which has common recognition in movie field data Preparation.ontlogy based system is helping to map the features present in reviews.

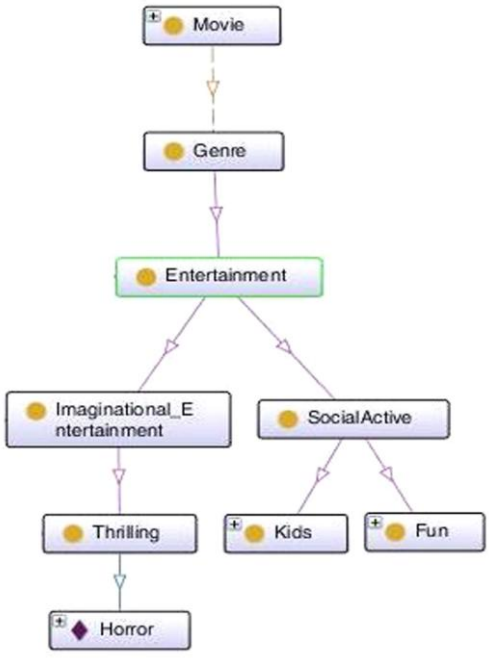

Figure 3 Structure movie of ontology

The movie review dataset is collected from http://ai.stanford.edu/~amaas/data/sentiment/ .It contains large number reviews and this is useful to evaluate the system. We employed the indicators of *precision*, *recall*, and *Fmeasure* to measure the performance on sentiment classification.

$$\text{Precision} = \frac{a}{a+b}$$

$$\text{Recall} = \frac{a}{a+c}$$

$$\text{Fmeasure} = \frac{2 \times precision \times recall}{precision \times recall}$$

With reference to a confusion matrix, *a*, *b*, *c* refer to the number of correctly classified positive (negative)reviews, the number of classified non-positive (negative) reviews, the number of non-classified positive (negative) reviews. The structure of movie ontology describes in Figure 3.

5 CONCLUSION

The Guided by the design science research methodology, we illustrate the design, construction, and evaluation of our ontology for opinion mining in this paper. In particular, we show that Domain Ontology can be automatically constructed to facilitate opinion mining, including the extraction of product features and sentiment words, extraction of feature relation. Our method can accurately predict the sentiment polarity of reviews. As a result, organizations can develop effective business strategies related to marketing, customer support, and product design functions in a timely fashion.

REFERENCES

[1] Isidro Peñalver-Martinez,Francisco Garcia-Sanchez,Rafael Valencia-Garcia, Miguel Ángel Rodríguez-García,ValentínMor AnabelFraga,Jose Luis Sánchez Cervantes(2014) Feature-based opinion mining through ontologies Expert Systems with Applications 5995–6008

[2] Balahur, A., &Montoyo, A. (2010). Semantic approaches to fine and coarse grained feature-based opinion mining.In Natural language processing and information systems (pp. 142–153). Berlin Heidelberg: Springer.

[3] Qiuxia Song, Jin Liu, Xiaofeng Wang, Jin Wang (2014), A Novel Automatic Ontology Construction Method Based on Web Data. In Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing. IEEE.

[4] Baccianella, S., Esuli, A., &Sebastiani, F. (2010). SentiWordNet 3.0: An enhance lexical resource for sentiment analysis and opinion mining. In Proceedings of the seventh conference on international language resources and evaluation (pp.2200– 2204). European Language Resources Association.

[5] Toutanova, K., Klein, D., Manning, C., & Singer, Y. (2003). Feature-rich part-of Speech tagging with a cyclic dependency network. In Proceedings of HLT-NAACL 2003(pp. 252–259).

[6] Zhao, L. & Li, C. (2009). Ontology based opinion mining for movie reviews. In Proceedings of the 3rd international conference on knowledge science, engineering and management (pp. 204–214)

[7] S. Li, Q. Yin, Y. Hu. Review on Ontology Research [J]. Computer Research and Development. 2004, 41(7)

[8] Ghiassi, M., Skinner, J., & Zimbra, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network Expert Systems with Applications: An International Journal, 40(16), 6266–6282.

[9] Pang, B., & Lee, L. (2004). A sentiment education: Sentiment analysis using subjectivity summarization based on minimum cuts. In Proceedings of the ACL (pp. 271–278).

[10] Baccianella, S., Esuli, A., & Sebastiani, F. (2009). Multi-facet rating of product reviews. In ECIR '09: proceedings of the 31th European conference on IR research on advances in information retrieval (pp. 461–472).
[11] Zhou, L., & Chaovalit, P. (2008). Ontology-supported polarity mining. Journal of the American Society for Information Science and Technology, 59(1), 98–110.