IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-5, Issue-8 E-ISSN: 2347-2693

Feature Extraction From Product Review Using Ontology

Drashti Naik[1*], Jitali Patel[2]

[1*]Dept. of CSE, Institute of Technology, Nirma University, Ahmedabad, , India
[2]Dept. of CSE, Institute of Technology, Nirma University, Ahmedabad, , India

[*]*Corresponding Author*: drashtinaik6893@gmail.com

Available online at: www.ijcseonline.org

Received: 12/Jul/2017, Revised: 23/Jul/2017, Accepted: 16/Aug/2017, Published: 30/Aug/2017

***Abstract*—** Opinion mining is accepting more attention because of the development of blogs, e-commerce, news, reports, forums and additional web sources where individuals tend to express their opinions. Different people have different opinions. People's thought may vary according to the domain and opinion may contain both positive and negative words. For a product, user may like or dislike some of its features. Filtering this review and extract domain related features is the important task of this paper. In this paper, ontology is used to extract the features and adjectives are used as the sentiment word. Sentiment Analysis is used to obtain positive or negative feature of the review.

***Keywords*—**Ontology,Natural-language-Processing,SentimentAnalysis.

I. INTRODUCTION

With the expanded utilization of the Internet around the world, the development of online business is very high, and purchasing items from the online websites has turned into another pattern in the present culture. There are a large no of items accessible in the market and its expanding day by day. Now, everything is online, people can also share their opinion online and make decisions. More and more people post reviews for any product, movies, event and political thought. From all of this review finding proper opinion is time consuming and burdensome. It is common for a person to learn what others like and dislike before buying anything and for a manufacturer to keep track of their opinion on its product to improve the satisfaction of the customer and provide them good services. So, now days there are so many researches available for sentiment analysis and recommendation system. Through the sentiment analysis customer can easily get the opinion for particular product. From research perspective, a product features can be express in the form of noun, adjective, verb or adverb. From this extract the feature and apply any machine learning algorithm for predict positive and negative review.

A. Sentiment Analysis:An Overview

Sentiment analysis, also called opinion mining, is the field of study that analyzes *people's* opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues, events, topics, and their attributes [1]. Its aim is to detect subjective information from any feedback and review given by the people. Opinions are the key influence of people's behaviours as they are important to any human activity online.

Sentiment analysis alludes to the utilization of natural language processing and text analysis to distinguish and separate subjective data from source data. There are so many applications for which sentiment analysis is connected to social media and reviews.

There are three levels for sentiment analysis:

Document Level:
It classifies whether the whole document expresses positive or negative opinion. It assumes that for a single product there is a document which expresses the opinion. So, when there is a comparison between two products this analysis is not applicable.

Sentence Level:
Its check the sentence one by one from the document and express positive, neutral or negative opinion.

Entity and Aspect Level:
It is also called feature level sentiment analysis. It checks for sentiment, whether it is negative or positive and target for which the opinion is given. For example, "Camera is not good, but i still like the phone" has a positive tone, but we can't say it's positive for the whole sentence.

Every review consist of opinion targets, sentiments, opinion holder and time, which is useful for predicting any review. Opinion target is for which product the opinion is given, Sentiments is the opinion given for the product is positive or negative, Opinion holder is the one who give the review of the product and Time is when the opinion holder given the review.

Types of Opinion:
There are two types of opinion regular and comparative opinion. Regular opinion is divided into two parts Direct and Indirect.

Direct Opinion: It expresses opinion directly on an entity and entity aspect."JBL headphone sound quality is great." express a positive opinion.

Indirect Opinion: It expresses opinion indirectly on an entity and entity aspect which affects other entity."After eating this dog food, my dog weight is decreased." express effect of food on dog health and give negative opinion.

Comparative opinion: It expresses the difference and similarity between the two or more entity and entity aspect."Iphone is better than Blackberry"

There are basically two types of review user can give. One format is user write the review in the form of advantages and disadvantages. Another format is user write the review in the form of full-text format.

B. Aspect Based Sentiment Analysis

For aspect based sentiment analysis we require two things, sentiment whether it is positive or negative and target for which the opinion is given. This process is divided into Aspect extraction and Aspect sentiment classification.

Aspect Extraction: It extracts the aspects which have been evaluated."JBL headphone sound quality is great." Here sound quality is the aspect of the entity JBL headphone. Whenever we have aspect we need to consider which entity aspect belongs to.

Aspect Sentiment Classification: Check if the opinion is positive or negative for given aspect. In above example, the sound quality is the positive opinion for entity JBL headphone.

Approach for Aspect Extraction:
Extraction based on frequent nouns and noun phrases: It finds aspect which is noun and noun phrase from the database for the particular domain. Use part-of-speech (POS) tagger to find noun and noun phrase. The disadvantage of this approach is that, it is not necessary that it find the aspect is domain related.

Extraction by exploiting opinion and target relations: It describes the grammatical relationship of sentences using the Stanford type dependencies. Dependencies are represented as directed graph, in which words are the nodes and grammatical relation are edges. Here the verb is selected as the root node. It uses part-of-speech tag and phrase label. The disadvantage of this approach is that it cannot identify the relation between semantic and syntactic structure.

Extraction using supervised learning: It uses machine learning algorithm like naive Bayes, support vector machine, maximum entropy, conditional random field etc to extract the feature and predict sentiment.

Extraction using topic modelling: The basic model is LDA which is Latent Dirichlet Allocation, is used. The disadvantage of LDA is that the topics are unlabeled, so it will not able to describe a direct relation between topics and a particular aspect of the entity.LDA is designed for the document level.

Section I contains the introduction of Sentiment Analysis and Aspect based Sentiment Analysis , Section II contain the related work of Aspect based sentiment analysis for different approach, Section III contain the proposed methodology with the flow chart and example, Section IV contain the experiment results for feature extraction and sentiment analysis, section V concludes research work with future directions.

II. RELATED WORK

Over the last several years, feature extraction and sentiment analysis have received increasing attention from the research. Ahmad and Najmud Doja [2] proposed an approach called candidate identification and frequent pattern generation. The system uses semantic analysis by using the decision tree classifier and use natural language processing for identifying feature and user FP-growth algorithm to extract feature. The disadvantage is that more memory is required for storing the transaction. Shruti Mishra and Sandeep Kumar [3] use fuzzy pattern tree approach. They compare the performance of original and fuzzified dataset for finding a frequent pattern. Using fuzzified data set it can capable of finding a large number of frequent pattern and have good running time capability. Hui Wang and Jiansheng Chen [4] use two-noun phrase approach for extraction of the feature. Two-noun phrases extract more specific features compared to one-noun phrases. They use three tagging method like CLAWS, NLProcessor, and Lingua::EN::Tagger for checking the accuracy and they find CLAWS tagger give a better result. Gulila Altenbek and Ruina Sun [5] use n-gram method and the experimental result shows that the phrase extraction accuracy is low the alternative for that is the use of basic noun phrase extraction.

A.Jeyapriya and C.S.Kanimozhi Selvi [6] use minimum support threshold to find the frequent aspect and used naive Bayes classifier for sentiment analysis. The disadvantage of the system is that it will not extract relevant feature all the time. Yamamoto, Yamasaki and Aizawa [7] proposed

approach service annotation and profiling. They apply IBMWaston relationship to extract POS and used TF-IDF algorithm for computing the relative score. They find the result that the precision without TF-IDF is better. Hamdan, Bellot and Bchet [8] use the conditional random field to extract feature and use naive Bayes classifier for predict whether the sentence is positive and negative. The disadvantage is that it will not work with unknown words.

Wanying Ding, Zunyan Xiong and Xiaohua Hu [9] proposed the hybrid HDP-LDA model. It uses Dirichlet process to find the aspect using part-of-speech. The advantage of this method is it can automatically determine the number of aspects and it will differentiate actual words from opinion words. Hai Son Le, Thanh Van Le and Tran Vu Pham [10] use GK-LDA for feature extraction and they prove that GK-LDA performs better than the LDA.

Lili Zhao and Chunping Li [11] use ontology to extract features and use sentiwordnet semi-automatic approach for polarity identification. The disadvantage of this method is they limit the sentiment analysis on the node of the model so its result gives general information about the movie. Mohd Ridzwan Yaakub [12] proposed ontology is too general and used very simple review for the manual extraction of the feature. Meleesa Alfonso and Razia Sardinha [13] use a fuzzy ontology to extract the features and sentiwordnet for polarity. Shein and Nyunt [14] use domain ontology and SVM classifier for sentiment analysis.

III. PROPOSED METHEDOLOGY

The main objective of the proposed system is to give a feature level opinion mining for a particular domain using ontology. Figure 1 describes the proposed method.

A. Ontology Creation

An ontology is a formal description of concepts in a domain of discourse (classes), properties of every concept describing various features and attributes of the concept, and restrictions on attributes.The concepts refer to various entities that may be any product or an organization. The use of ontology in feature level opinion mining is to distinguish the domain related features by defining the classes in the domain and giving the relationships between the classes and instances.

Ontology is used to find the domain related features from the domain review. For ontology creation, we can use the existing ontology and extend it as per our requirement or we can build our own ontology. We can use any ontology development tool or java API for ontology creation. Figure 2 shows the cell phone ontology.

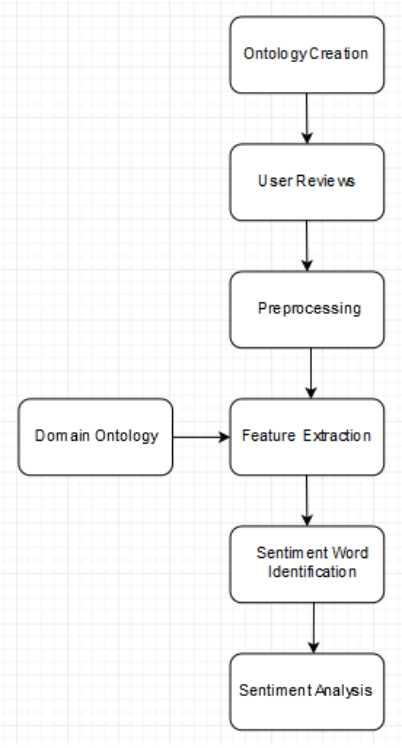


Figure 1. Proposed System

B. Preprocessing

Several Natural language pre-processing techniques are available for tokenization, stamming, stop word removal, part-of-speech tagging. In this paper, ontology is used so the only requirement is of POS tagging. POS tagging is used for identification of words as nouns, verbs, adjectives, adverbs, etc. NLTK POS tagger is used to identify the noun.

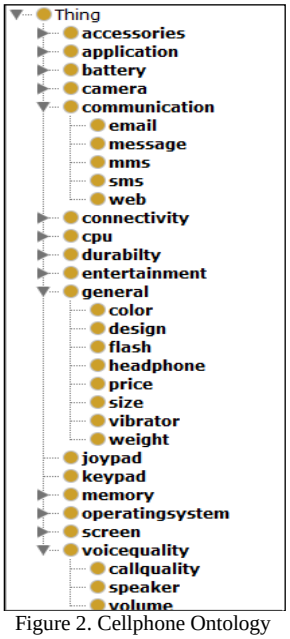


Figure 2. Cellphone Ontology

C. Feature Extraction

Domain ontology is used for feature extraction. Identified noun using POS tagger is compared with the concepts of the domain ontology. Here, cellphone ontology is used. Different users write the review and they use different words. For example, a consumer can write cost instead of price, to mention phone's rate. So, add synonyms in the domain ontology for better result of feature extraction. In Figure 3 user's review is available, on that POS tagging is applied and noun are compare with the ontology for feature extraction.

D. Sentiment word Identification

For the identification of positive or negative opinion, sentiment words are used. In this method, adjective words are extracted as a sentiment word. Using POS tagging adjective can be extracted.

Review	Awesome screen all the greatness of OLED perfect blacks high contrast looks good in direct sunlight it is a pleasure to look at from any distance The resolution is still 800x480 Very light Considering the size of this thing it is very hard to believe When you put the phone into someone else is hand for the first time they usually are confused because they expect it to feel more solid and not so feather weight.
POS tagging	[('Awesome', 'NNP'), ('screen', 'NN'), ('greatness', 'NN'), ('OLED', 'NNP'), ('perfect', 'NN'), ('blacks', 'NNS'), ('high', 'JJ'), ('contrast', 'NN'), ('good', 'JJ'), ('direct', 'JJ'), ('sunlight', 'JJ'), ('pleasure', 'NN'), ('distance', 'NN'), ('resolution', 'NN'), ('800x480', 'CD'), ('light', 'JJ'), ('size', 'NN'), ('hard', 'JJ'), ('put', 'VBD'), ('phone', 'NN'), ('hand', 'NN'), ('time', 'NN'), ('confused', 'VBN'), ('expect', 'VBP'), ('feel', 'NN'), ('solid', 'JJ'), ('feather', 'NN'), ('weight.', 'NN')]
Feature	screen, size

Figure 3. Feature Extraction Example

E. Sentiment Analysis

Sentiment Analysis means to identify whether the review is positive or negative for a particular feature. Pair the noun and adjective as a feature and its sentiment word. Use sentiment Analyzer to analyze whether it is positive or negative.

IV. EXPERIMENT AND EVALUATION

Apply this proposed method on cellphone review. Take 200 reviews of the cellphone from Amazon. Precision and recall are used to measure the performance of the extracted features.

For the 200 Review, a total number of actual extracted features are 934. Table 1 show it and it will also show the result for sentiment.

Table 1. Results

	Tp	Fp	Fn
Features	934	115	391
Sentiments	382	187	291

$$Precision = \frac{TP}{Tp + Fp}$$

$$Recall = \frac{Tp}{Tp + Fn}$$

$$Fmeasure = 2 * \frac{Precision * Recall}{Precision + Recall}$$

Here, Tp is true positive means the extracted feature and actual features are same.

Fp is false positive means the actual feature is not extracted from the system.

Fn is false negative means the feature is extracted but it is not the actual feature.

For the feature extraction, overall Precision is 89.03%, Recall is 70.49% and F-measure value is 78.68%.

For sentiment analysis, overall Precision is 67.13%, Recall is 56.76% and F-measure value is 61.51%.

Here, Tp means the sentiment for the feature is positive and its prediction is also positive.

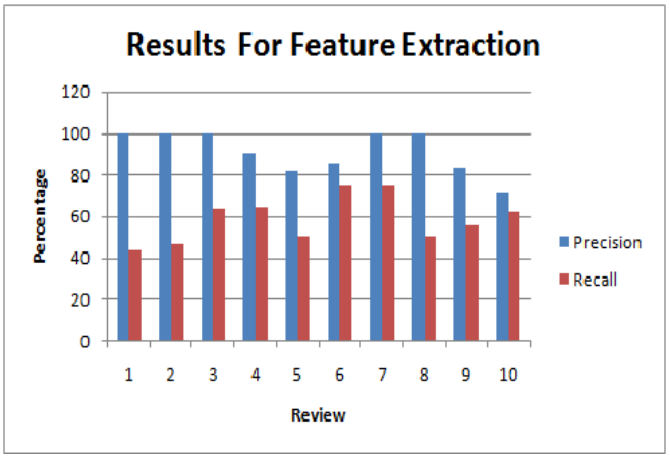
Fp means the sentiment for the feature is negative but the prediction is positive.

Fn means the sentiment for the feature is positive but the prediction is negative.

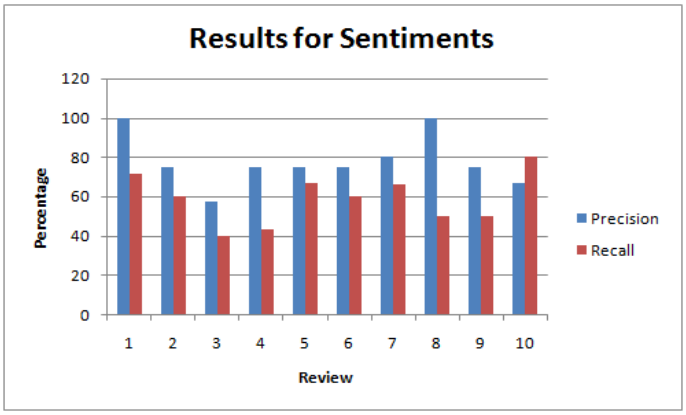
```
['camera', 'phone']
(camera, perfect)
(phone, loud)
{'neg': 0.0, 'neu': 0.0, 'pos': 1.0, 'compound': 0.5719}
{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
['photos', 'camera', 'callquality']
(photos, past)
(camera, outstanding)
(callquality, excellent)
{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound': 0.0}
{'neg': 0.0, 'neu': 0.0, 'pos': 1.0, 'compound': 0.6124}
{'neg': 0.0, 'neu': 0.0, 'pos': 1.0, 'compound': 0.5719}
```

Figure 4. Final Output

Figure 4 describe the final output. In that, the first line is of extracted features for the cellphone first review. The second line shows the pair of extracted feature and sentiment word. The fourth line shows the sentiment whether it is positive or negative. For all the reviews the results are shown in this format, first is the extracted feature next pair of feature and sentiment word and sentiments for extracted features.



First graph is of results for feature extraction. It will display precision and recall of extracted features for individual review.
Second graph is of result for sentiment analysis. It will display precision and recall of extracted feature's sentiment analysis for individual review.



V. CONCLUSION

Ontology is an idea which describes system in the form of knowledge and semantics. It expects to get information and provides an understanding of information in the domain. Opinion mining is a task, which analyzes opinions given by users for different features, and determines whether these opinions are positive, neutral or negative.

In this paper, Ontology is used for the feature level sentiment analysis for the cell phone review. It is used for both producer and consumer. The producer can improve their services and product feature and it will help the consumer to make the decision about what to buy based on the features they like. First cellphone ontology is created for feature extraction. Feature and sentiment word pair is used for sentiment analysis.

For the future work develop the system which will automatically update the ontology and take two different domain ontology and try to extract features related to the particular domain.

REFERENCES

[1] B. Liu,"*Sentiment analysis and opinion mining*", Synthesis lectures on human language technologies, vol.5, no.1, pp.1-167, 2012.

[2] T.Ahmad and M.N.Doja,"*Opinion mining using frequent pattern growth method from unstructured text*", pp. 92-95, 2013.

[3] S.Mishra, D.Mishra, and S.K.Satapathy,"*Fuzzy pattern tree approach for mining frequent patterns from gene expression data*", vol.2, pp.359-363,2011.

[4] H.Wang and J.Chen,"*Extracting two-noun phrases from customer reviews*", pp.1-4, 2009.

[5] G.Altenbek and R.Sun,"*Kazakh noun phrase extraction based on n-gram and rules",* pp.305-308, 2010.

[6] A.Jeyapriya and C.K.Selvi,"*Extracting aspects and mining opinions in product reviews using supervised learning algorithm*", pp. 548-552, 2015.

[7] M.Yamamoto, T.Yamasaki, and K.Aizawa,"*Service annotation and profling by review analysis*", pp.357-364, 2016.

[8] H.Hamdan, P.Bellot, and F.Bechet,"*Supervised methods for aspect-based sentiment analysis*", pp. 596-600, 2014.

[9] W.Ding, X.Song, L.Guo, Z.Xiong, and X.Hu,"*A novel hybrid hdp-lda model for sentiment analysis*", pp. 329-336, 2013.

[10] H.S.Le, T.Van Le, and T.V.Pham,"*Aspect analysis for opinion mining of Vietnamese text*", pp. 118-123, 2015.

[11] S.Thirumaran and M.Sangeetha, "*Ontology Founded Mesh Flatterer Aimed at Removal Facilities info Result*", International Journal of Computer Sciences and Engineering, Vol.3, Issue.1, pp.224-227, 2015.

[12] Yaakub, R.M, Li and Feng Y,"*Integration of Opinion into Customer Analysis Model*", IEEE International Conference on e-Business Engineering, pp. 90-95, 2011.

[13] Priyanka Sharma, R.K. Gupta, "*A Novel Web Usage Mining Technique Analyzing Users Behaviour Using Dynamic Web log*", International Journal of Computer Sciences and Engineering, Vol.5, Issue.6, pp.106-111, 2017.

[14] Shein, Khin Phyu Phyu and Thi Thi Soe Nyunt,"*Sentiment classification based on Ontology and SVM Classifier*", IEEE Communication Software and Networks, pp. 169-172 ,2010.

[15] Freitas, Larissa A, and Renata Vieira,"*Ontology based feature level opinion mining for portuguese reviews*", 22nd International Conference on World Wide Web. ACM, pp. 367-370 ,2013.

[16] Hazman, Maryam, Samhaa R.El-Beltagy and Ahmed Rafea,"*A survey of ontology learning approaches*", vol. 22, No. 9,2011.

[17] Penalver-Martinez, Isidro, et al,"*Feature-based opinion mining through ontologies*", Expert Systems with Application, pp.5995-6008 ,2014.

[18] Ali, Farman, Kyung-Sup Kwak and Yong-Gi Kim,"*Opinion mining based on fuzzy domain ontology and Support Vector Machine: A proposal to automate online review classification*", Applied Soft Computing, pp. 235- 250, 2016.

[19] Kontopoulos, Efstratios, et al,"*Ontology-based sentiment analysis of twitter posts*", Expert systems with applications, pp.4065-4074, 2013.

Authors Profile

Drashti Naik pursed Bachelor in Information Technology from Uka Tarsadia University, Gujarat, India in 2015 and Master in Computer Science and Engineering from Nirma University, Gujarat, India in year 2017. She is currently working as Assistant Professor in Department of Computer Engineering, Uka Tarsadia University, Gujarat, India since 1St July 2017.

Jitali Patel pursed Bachelor in Information Technology from Saurashtra University, Wadhwan, India in 2006 and Master in Computer Engineering from DDU University, Gujarat, India in year 2010. She is currently working as Assistant Professor in Department of Computer Science and Engineering, Nirma University, Gujarat, India since 2011. Her main research work focuses on Big Data Analytics, Data Mining, Web Mining. She has 9 years of teaching experience.

