Product Feature-based Ratings forOpinionSummarization of E-Commerce Feedback Comments

Vijayshri Ramkrishna Ingale
PG Student, Department of Computer Engineering
JSPM's Imperial College of Engineering & Research
Wagholi, Pune, India

Rajesh Nandkumar Phursule
Professor, Department of Computer Engineering
JSPM's Imperial College of Engineering & Research
Wagholi, Pune, India

ABSTRACT
With the growth of internet, online social networking sites, blogs, discussion forums, etc have gained a tremendous importance. Consumers comment on net to express their views, feedbacks and opinions. The opinion of users is of great importance for mining useful information from the text which can be done through opinion mining techniques. Opinion mining or sentiment analysis is the computational field of study of people's opinions, emotions, and attitude towards particular Feature. When buying a new product buyer mostly refer the opinion of the other users who have bought the product. Hence, in this work a product Feature rating framework is being proposed. This dissertation comprises mainly of four modules preprocessing, Feature identification, review classification and Feature rating. Finally, the rating are been shown in the graph. For the analysis of the system, we have used Amazon review dataset which consists of customers reviews about product. In the system Apriori algorithm is used for Feature identification, Support Vector Machine algorithm for review classification and SentiWordNet lexicon for giving rating to each Feature of the product.

Keywords
Opinion Mining, Sentiment Analysis, Feature

1. INTRODUCTION
Customer's opinions represent a valuable unique type of information, which should not be mistreated or neglected by the researchers. Thus, this work emphasizes the need of special mechanisms that aims to provide the community better ways to take full advantage from this data. And, so for mining useful information from web Opinion Mining domain is being taken into consideration.

Mining important Features will improve the analysis of numerous reviews and is beneficial to both consumers and firms. Customers can conveniently make wise decision by paying more attentions to the important Features while buying a product, while organisations can focus on improving the quality of these Features and thus enhance product reputation effectively.

1.1 Motivation
Being a buyer, people mostly refer to the ratings of the product while buying a new product. The Existing systems are mostly focused on the number of users liking the product. Hence, this does not give a clear view about the product. Also, most of the systems are focused on give a rating to the complete product. But, sometimes a particular Feature matters more than the other Feature of the product. So, we need a look at all the Features of the product; regarding to whether what Feature we want has want rating for that product. Also, most of the systems are mostly focused on only giving the positive rating to the product, but in our system we would be giving positive as well as negative rating for the product. The positive rating tells how better the product is whereas; the negative rating tells how worst the product is. Thus, we aim to give ratings to each Feature of the product.

1.2 Study Of Existing Systems
Feature-based opinion summarization has two main characteristics. First, it captures the gist of opinions: opinion targets (entities and their aspects) and sentiments about them. Second, it is quantitative, which means that it gives the percent of people who hold positive or negative opinions about the entities and aspects. The quantitative side is crucial because of the subjective nature of opinions. The resulting opinion summary is a form of structured summary (Hu and Liu, 2004; Liu, 2010).[6]

Table 1. Study of existing systems

Sr. No.	Methodology	Appreciation	Limitations
1.	Semantic labeling and polarity computation	Effective method of feature lexicon construction has been made	Feature polarity is divided only to five levels which will not be sufficient for a large corpus of reviews
2.	feature extraction method based on polarity	Polarity is estimated not based on just the nature of objective but also based on the context in which the objective is used	Polarity changes with adjective position in sentence this needs to be analyzed.
3.	Feature based opinion mining	Co-occurrences of words are considered in analysis which increases the weightage for a particular feature. Frequency is calculated for each of the terms in the reviews feature list of candidate is created, reducing cost of feature.	Reviews are categorized as relevant and irrelevant on the basis of the domain they commented on. But this classification has been done on the basis of words present in the review

1.3 Proposed Methodology

The steps of whole process of mining features and rating them are described below.

1. Take online text or reviews of customers as input and perform pre-processing.
2. Split text into sentences and then tokenize each sentence.
3. (POS tagging) Part of speech tagging of all token, tag as /NN, /JJ, /VB, /RB for noun, adjective, verb & adverb.
4. The noun, noun phrases, adjectives, verb and adverb along with their word position are captured in the sentence.
5. The product features list from key noun phrases is prepared.
6. Select important features using frequency base selection method
7. Comparison among the important features.
8. Review Classification
9. Feature Rating

Noise Removal: The online text contains unnecessary tags and noise. In this work firstly all these noise are removed as a preprocessing step and then read the text for further processing.

Sentence Splitting and Tokenization process: For sentence level sentiment classification it is required to slip the whole document into sentences which have unique sentence ID. In this work the whole documents and reviews/online text are splitted into sentence by using "." as a sentence boundary. After getting a list of sentences, each sentence again is split into tokens along with their position in the sentence.

Part of Speech (POS) tagging: The standard standford-postagger POS tagger has been used for tagging each word. In this work consider only Noun, Adjective, Verb and Adverbs from the tokens and then assign following tag /NN /JJ /VB /RB respectively for the about part of speech.

Table 2. POS tagging.

S. No.	POS Name	POS Tag	SentiWordNet Tag
1	NOUN	NN	N
2	ADJECTIVE	JJ	A
3	VERB	VB	V
4	ADVERB	RB	R

Preparation of Product Features List: After tagging, the sentences have opinion words tags like Noun, Adjective, Verb and Adverbs. For product features selection we filter these tagged sentences and select those sentences that have noun or noun phrases. From the filtered tagged sentences the product features list are prepared and this list contains all the words/features of /NN tagged in the sentence.

Frequency based Important Product Features Selection Process: The product features list contains all features. We create the list by choosing the /NN tagged. Sometimes all the /NN tagged word are not important for consideration and in decision-making process. Therefore, there must be a list, which contains the important features selected from the product features list. For this purpose a threshold frequency has been defined. Apriori Algorithm is used for feature identification.

Review Classification: For review classification, we are giving an input of training dataset to SVM (Support Vector Machine), which should be compulsorily in the form of "review tab label" format.

Rating of product features: SentiWordNet Dictionary Lexicon is proposed for rating features, where aspect wise rating of each aspect of the product will be carried out and both positive score as well as negative score for each aspect would be given.

Comparison among various features: In sentiment analysis it is also important to check the Product features importance with respect to another product features.

2. SYSTEM ARCHITECTURE

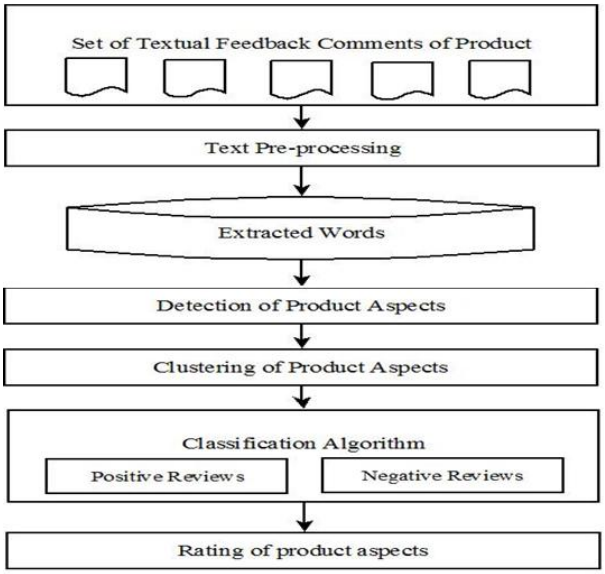


Figure 1. System Architectural Design

3. IMPLEMENTATION

Algorithm 1: Preprocessing ()

Input: Review dataset and stop words list

Output: Tagged file

Steps:

1. Input review file and stop words list
2. Process reviews and stop words list
3. Remove stop words from reviews
4. Tokenize reviews and create tokens
5. Select POS tagger
6. Tag each token in review file
7. Generate tagged file
8. Send tagged file to find nouns
9. Exit

Algorithm 2: Feature_Identification ()

Input: Tagged file

Output: Features

Steps:

1. Receive tagged file
2. Select all noun phrases from tagged file
3. Create nouns file
4. Remove tags of nouns and write them into nouns file
5. Select support value to find frequent items in nouns file
6. Input nouns file and support value to Apriori_Algorithm()
7. Call Apriori_Algorithm()
8. Receive frequent item set i.e. Features as output of Apriori_Algorithm()
9. Exit

Algorithm 3: Reviews_Classification ()

Input: Training dataset and review file

Output: Classified reviews

Steps:

1. Input training dataset and review file
2. Convert training dataset to LibSVM format
3. Convert reviews file to LibSVM format without labels
4. Training SVM model
5. Predicting reviews positive or negative
6. Classification of reviews file into positive and negative review set
7. If review is positive then write review to positive_review file
8. End If
9. If review is negative then write review to negative_review file
10. End If
11. Exit

Algorithm 4: Featurewise_Rating ()

Input: Features list and classified positive and negative reviews set

Output: Feature wise rating

Steps:

1. Input Features list and positive and negative review sets
2. Create SentiWordNet dictionary
3. For each Feature in Features list
4. Calculate positive and negative score
5. End For
6. Pass positive and negative rating values of Features to Graph_Generator()
7. Call Graph_Generator()
8. Rating chart is returned
9. Display rating chart
10. Exit

4. EXPERIMENT AND EVALUATION

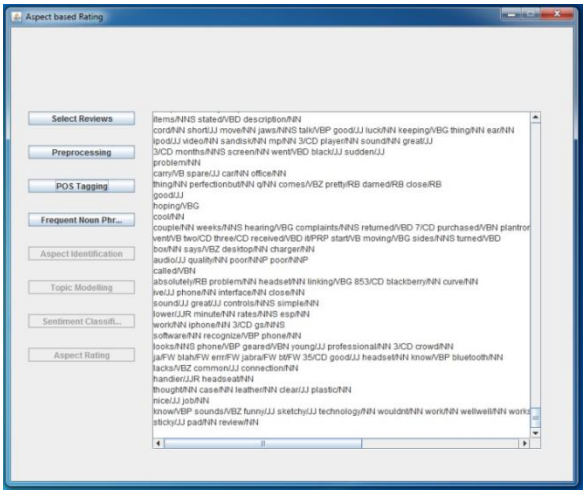
4.1 Dataset

The reviews or feedbacks of customers, were collected from Amazom.com in text format and processed.[6] The reviews contained 4000 sentences. Each dataset consisted of more than 260 sentences found to be opinionated reviews written by 325 customers. The format of the datasets is unstructured text files.

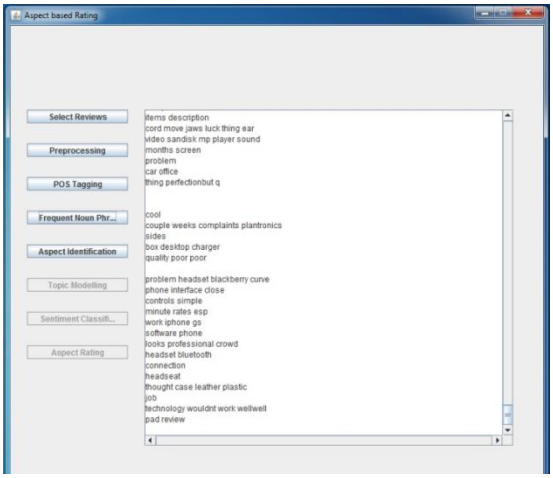
To evaluate the discovered features, a human tagger manually read all of the reviews and labeled aspects and associated opinions for each sentence as '0' or '1'. Before, we use the datasets, we pass the dataset to a pre-processing to remove all stopwords and get the original collected reviews.

4.2 Results

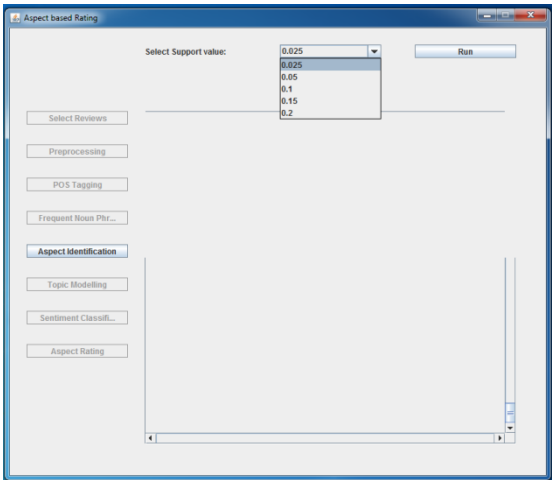
- POS Tagging of the pre-processed dataset



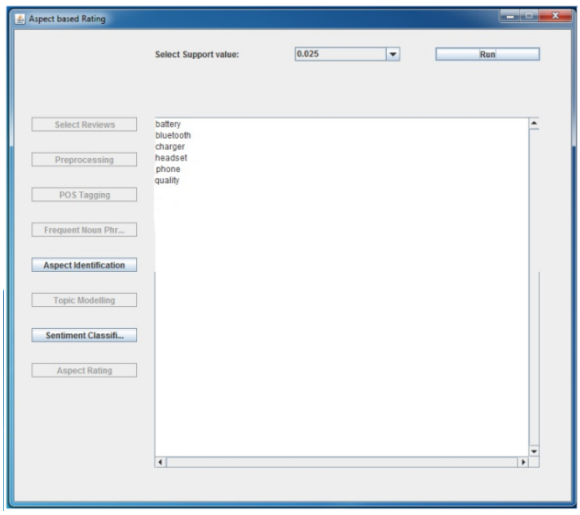
- Get the Noun phrases



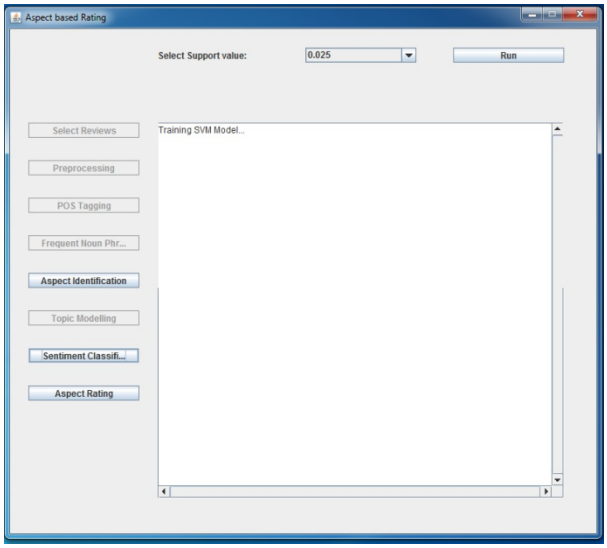
- Select the value of support



- Find the frequent Features



- Train SVM model



- Feature wise rating

5. CONCLUSION

In future, this work can be extended by adding timestamps of feedback for analyzing any opinion polarity change over time. Experiments are performed on amazon customer review dataset shows that our system performs efficiently. For Feature identification, Apriori algorithm is used, whereas for review classification Support Vector Machine algorithm is used and for giving rating to the products Features SentiWordNet Lexicon is used.

6. ACKNOWLEDGMENTS

I am thankful to respected Principal Dr. S. V. Admane and guide Prof. R. N. Phursule who have managed to get everything done on time and provided me with many pieces of valuable advice. Finally, my greatest gratitude goes to my own family who have helped in so many ways.

7. REFERENCES

[1] Xiuzhen Zhang, Lishan Cui and Yan Wang, "CommTrust: Computing Multi- Dimensional Trust By Mining E-Commerce Feedback Comments," IEEE Transactions On Knowledge And Data Engineering, vol. 26, no. 7, pp. 1631-1643, 2014.

[2] Zheng-Jun Zha, Jianxing Yu, Meng Wang, Tat-Seng Chua," Product Aspect Ranking and Its Applications", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 5, MAY 2014

[3] Abdul Wahab," IMPORTANT FEATURES SELECTION DURING SENTIMENT ANALYSIS", Sci.Int(Lahore),26(2),961-966,2014.

[4] Bing Liu. Sentiment Analysis and Opinion Mining, Morgan &

Claypool Publishers, May 2012

[5] M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 2004, pp. 168-177.

[6] Wouter Bancken, Daniele Alfarone and Jesse Davis, "Automatically Detecting And Rating Product Features From Textual Customer Reviews," Proceedings of DMNLP Workshop At ECML/PKDD, pp. 1-16, 2014.

[7] Bo Pang, Lillian Lee and Shivakumar Vaithyanathan, "Thumbs Up? Sentiment Classification Using Machine Learning Techniques," Proceedings Of The Conference On Empirical Methods In Natural Language Processing (EMNLP), pp. 79-86, 2002.

[8] Zheng-Jun Zha, Jianxing Yu, Jinhui Tang, Meng Wang and Tat-Seng Chua, "Product Feature Ranking And Its Applications," IEEE Transactions On Knowledge And Data Engineering, vol. 26, no. 5, pp. 1211-1224, 2014.

[9] Minqing Hu and Bing Liu, "Mining and Summarizing Customer Reviews," 10th Proceeding ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), pp. 168-177, 2004.

[10] Jianxing Yu, Zheng-Jun Zha, MengWang and Tat-Seng Chua, "Feature ranking: Identifying important product Features from online consumer reviews," Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pp. 1496-1505, 2011.

[11] Kumar, Ravi V. and K. Raghuveer, "Web User Opinion Analysis for Product Features Extraction and Opinion Summarization," International Journal of Web and Semantic Technology, vol. 3, no. 4, pp. 69-82, 2012.

[12] Chien-Liang Liu, Wen-Hoar Hsaio, Chia-Hoang Lee, Gen-Chi Lu and Emery Jou, "Movie Rating And

Review Summarization In Mobile Environment," IEEE Transactions On Systems, Man, And Cybernetics, vol. 42, no. 3, pp. 397-407, 2012.

[13] Yuanbinwu, Qi Zhang, Xuanjing Huang and Lidewu, "Phrase Dependency Parsing For Opinion Mining," Proceedings of the 2009 Conference On Empirical Methods in Natural Language Processing, pp. 1533-1541, 2009.

[14] Shenghua Bao, Shengliang Xu, Li Zhang, Rong Yan, Zhong Su, Dingyi Han and Yong Yu, "Mining Social Emotions from Affective Text," IEEE transactions on knowledge and data engineering, vol. 24, no. 9, pp. 1658-1670, 2012.

[15] Mily Lal and Kavita Asnani, "Feature Extraction and Segmentation In Opinion Mining," International Journal Of Engineering And Computer Science, vol. 3, no. 5, pp. 5873-5878, 2014.

[16] Mikalai Tsytsarau and Themis Palpanas, "Survey On Mining Subjective Data On The Web," Data Mining Knowledge Discovery, Springer, pp. 478-514, 2012.

[17] Esuli and Sebastiani, "SentiWordNet: A publicly available resource for opinion mining," In Proceedings of the 6th international conference on Language Resources and Evaluation (LREC06), pp. 417-422, 2006.

[18] Amani K Samha,Yuefeng Li and Jinglan Zhang , "Feature-Based Opinion Extraction From Customer Reviews," arXiv preprint arXiv:1404.1982, pp. 149-160, 2014.