

Opinion based on Polarity and Clustering for Product Feature Extraction

Sanjoy Das

School of Computing Science and Engineering, Galgotias University, India
Email: sdas.jnu@gmail.com

Bharat Singh

School of Computing Science and Engineering, Galgotias University, India
Email: glabharat.mca@gmail.com

Saroj Kushwah, Prashant Johri

GLA University, India and Galgotias Institute of Management & Technology
Email: { sarojkushwahsiem , johri.prashant }@gmail.com

Abstract—In recent time, with the rapid development of web 2.0 the number of online user-generated review of product is increases very rapidly. It is very difficult for user to read all reviews and handle all websites to make a valuable decision at feature level. The feature level opinion mining has become very infeasible when people write same feature with contrary words or phrases. To produce a relevant feature based summary of domain synonyms words and phrase, need to be group into same feature group. In this work, we focus on feature based opinion mining and proposed a dynamic system for generate feature based summary of specific feature with specific polarity of opinion according to customer demand on periodic base and changed the summary after a span of period according to customer demand. First a method for feature (frequent and infrequent) extraction using the probabilistic approach at word-level. Second identify the corresponding opinion word and make feature-opinion pair. Third we designed an algorithm for final polarity detection of opinion. Finally, assigning the each feature-opinion pair into the respective feature based cluster (positive, negative or neutral) to generate the summary of specific feature with specific opinion on periodic base which are helpful for user. The experiment results show that our approach can achieves 96%accuracy in feature extraction and 92% accuracy in final polarity detection of feature-opinion pair in feature based summary generation task.

Index Terms—Clustering, infrequent feature, frequent feature, opinion mining, sentiment orientation, feature based analysis.

I. INTRODUCTION

Of late with the development of web 2.0 and increase of opinion-rich resources such as movie and product review. Social networking twitter and blogs are connecting more and more people with each other and sharing information

globally, one emerging field is opinion mining and sentiment analysis.

Individual and organization have been using public opinions for decision making. As social media has exploded with reviews, forum, blogs, micro blogs and social network on the web. But it has become very difficult task to monitor opinions sites on the web and distilled the information because of proliferation of different websites that imposes a great challenge for the average human reader to identify relevant site and accurately summarize the information and opinion content in them.

Online reviews play an important role in influencing purchase decision on certain products. They can play a vital role as an indicator of a product future sells performance.

Most of the online merchant uses online reviews to enhance customer level satisfaction and understand their shopping experiences. It has become very difficult for customer to make purchasing decision based on only short product description pictures available.

Some researcher mainly focus on the extracting objective information from review sentences in the document level opinion mining using the probabilistic approach [13].Some to analyze and extract opinions of sentiment information and present them in the form of opinion oriented summarization [3][4].

Hu and Liu [3] using the association rules mining for access the noun as frequent feature recognition. In [8,10,13,18] also use the association rule mining to extract the noun/noun phrases as frequent feature identification. In [5] uses the web PMI point wise mutual information to extracting the feature.

While for aspect based opinion mining, the main task is to extract the aspect and the corresponding from the user review.

Users use different words to represent same aspect. For example, connectivity: relation, communication, connection, transfer rate. Display: screen, touch pad.

Music player: audio, playlist, song. Battery: cell, backup, charging, long lasting etc.

So far many approaches have been proposed for building the aspect based opinion mining system to generate the feature based summary and further cluster the feature of product but almost each of them has some advantages and limitation as well.

Our work is organized in different phases are as follows:

First Phase- the collection of user generated review are extracted from the internet sources according to periodic based demand of customer and those customer review are analyzed as follows: we have done some preprocessing task (1) stop word removing (2) stemming (3)fuzzy machine technique.

Second Phase- Identify the feature: frequent feature which are directly define(Battery, Screen, picture etc.) and infrequent feature which are indirectly define (cheap, backup, durable etc.) from the extracted review sentences.

Third Phase- corresponding opinion word or opinion sentences are extracted and analyzed the polarity (positive, negative and neutral) for each sentence.

Fourth Phase- Make the feature-opinion pair. In case of multi opinion word in a review sentence when only the opinion word in the sentence not the feature word in the sentence existing technique have failed to solve this issue, we develop a new method to solve this kind of issue.

Fifth Phase- Grouping the different features which are domain synonyms are grouped under the same feature group.

Finally we assigned each feature opinion pair into their respective feature based cluster either positive cluster or negative cluster. The feature based summary of online product review is very relevant and highly useful. The proposed system is dynamic after each period (from launching date to current date, monthly, quarterly, daily or periodically according to customer demand) all the user generated review are collected. During that time period are extracted from the web and changed the feature based summary every time generated on the customer demand.

Our proposed work is different from the existing technique where the summary of all the feature generated from online user-generated review in the form of text based summarization because now days customer have demand to see the summary of specific feature with specific opinion or polarity (positive, negative, neutral) on of domain specific product shown in Table 1.

Suppose, a customer wants only to see the summary of screen, size, audio and battery with specific opinion (positive, negative, neutral).

However, it is not a trivial technique to extract the positive, negative and neutral polarity of review. In this research, we develop an automatically changing technique to extract the feature based opinion and generated the summary of customer on demand i.e. feature and polarity specific from the feature based review as shown in table 2.

As shown in table 1, customer demand or want to see the summary of specific feature with specific opinion (polarity) as: positive opinion of screen. Positive, negative and neutral opinion of size. Negative and neutral opinion of audio and both positive and negative opinion of battery

from the overall feature of domain specific product.

Table 1. Feature with specific opinion and Polarity

Feature name	Polarity		
	Positive(+)	Negative(-)	Neutral(*)
Screen	√		
Touchpad			
Size	√	√	√
Apps			
Audio		√	√
Video			
Picture			
Battery	√	√	
price			

Table 2. Summary of specific feature with specific opinion on customer demand.

Feature name	polarity		
	Positive(+)	Negative(-)	Neutral(*)
Screen	305+		
Size	416+	152-	84*
Audio		125-	78*
Battery	364+	185-	

II. RELATED WORK

Different types of approaches have been proposed for aspect based opinion mining.

In [1], authors describe the techniques for sentiment analysis for exploit sentiment topic information to generate context driven feature. They retrieved domain context feature is more effective than generally used feature in sentiment analysis. The domain specific sentiment analysis include prepare a domain corpus which contains relevant parameter a set of query in the domain and then combines with Bootstrapping algorithm. The experiment results show that the bootstrapping algorithm is able to commerce and aggregate new clues does not affect sentiment classification. The size of corpus does not affect the increased performance. In [6], authors widely elaborate the role of TREC (style give question answering system). In [7], author proposed aspect based opinion question answering (AQA) system for extract target product that have been commented on in the review. In [9] authors described a method for customer opinion polling from free from textual review, without requiring designing a set of question or assigning any rating .Multi aspect bootstrapping method is proposed to learn the ART of each aspect that are used to identify the aspect and aspect based segmentation model is proposed to divide a multi aspect sentence into a multiple single unit for opinion polling, then author generate the opinion poll. In [10] authors described a method for generate a summary by extracting the feature word and corresponding opinion form the customer review. Author uses the association rule mining for extracting the frequently occurring noun/ noun phrases. Which are considering as the feature of the given product? In [11] authors describe the feature

categorization of customer review sentences of product based on twice clustering technique with semantic association. The author mainly focused on the use of opinion word in the place of context word to assess the interrelationship of the any product features word. Group information and sharing information of opinion word extract automatically. The cluster results of active product feature are used as constraints and to cluster whole remaining feature word or phrases. For obtain better performance the twice clustering methodology has been preferred to single clustering. The experiment has been done at initial stage and has not been experimented on different language. In [12] authors proposed opinion based "scenario templates" to generate the summary of opinions expressed in the customer review of any product. The author used the information extraction question answering kind of system which is opinion orientation centric framework was created using the summary of opinions. In [13] authors proposed the probabilistic approach for the feature based summary of customer opinion. All the experiments have done at document level for all domains. There are some problem when grouped the same feature into the same feature group and when two feature indicate the same feature for movie domain and different feature for phone domain (for example. video and picture). In [14] describes the association rule mining method for extracting the feature, which are frequently occurring in the review sentences in the form of noun\ noun phrases but all frequently noun\ noun phrases are not the feature of product to generate the summary of customer review. In [18] authors proposed dynamic system for feature based summary of customer review using the combination of association mining and probabilistic approach. Association rule mining technique is used for extracting the frequently occurring noun\ noun phrases and probabilistic approach is used for remove the noun\ noun phrases which are not indicate the feature of product but there are some verb\ verb phrases, adverb and adjective also represent the feature of product .

III. PROPOSED METHOD

To deal with the issue and challenges of existing approach we proposed a dynamic system, which consider all the noun\ noun phrases, verb\ verb phrases, adverb and adjective for identification of feature word.

Association rule mining technique only work well for frequent or explicit feature where the feature word is directly define in the form of noun\ noun phrases. But some opinion word which modify the all feature word, e.g. (very good),(very bad),(not good),(cheap),(expensive) these are common in Chinese reviews, are implicitly expressed in the customer review sentences.

Our work is further enhancement of works proposed in [10,13]. In [10,13] proposed a generic feature based summarization approach, we have modified in order to make domain specific product feature classifier technique to create a feature based summary of customer review,

Our implementation focused the system dynamic (from launching date of product to current date, monthly,

daily, hourly according to customer demand) all the reviews update during the time of interval or period retrieved from the different internet sources(shopping websites) after a span of period.

Feature and corresponding opinion extraction

Here before retrieving the feature and opinion word some initial processing of all the user- generated review of online domain specific product are done on following steps:

Step-1: stop word removing process is used for remove the article, pronoun and preposition etc. which have little lexical meaning with other phases for sentence inside a sentence as are-the, on, of, with a, about. While, who, what, that their, where, who, be, why etc. to maximize the storage capacity.

Step-2: Stemming process reduces words by removing suffixes. This is much simpler heuristic process. The advantage or reducing word to a root is to increase the hit rate of identical terms and maximizing the storage capacity of a system.

Step-3: We use the fuzzy machine to remove or replace the misspelling word from the customer review sentences.

Feature extraction process has two phases:

Phase-I: Part of speech (POS) tagging. In this process, tagging all the word in the document with representing the user- generated review of product tagged with their respective part of speech (POS) and used stanford NLP parser [20] to tag all the document and generate the output file of POS tagged XML. doc.

Phase-II: Domain specific feature recognition task, here we extract the domain specific feature of product and the generic feature of product.

Earlier research on feature based opining mining represent the noun [21] as the feature, but each and every time noun not represent the feature of product for example name of person and place, positive and negative comments, limitation etc. not represent the feature of any product (camera, Nokia phone, computer etc.). Some researcher uses the association rule mining [8,10,22] for feature extraction, extract only the noun in the review sentences.

In contrast with [18] we used the verb, verb phrases, adverb and adjective for extract the corresponding feature of product. There is no matter what is the entity or feature of the entity, every adjective is used to modify the feature. Mostly in Chinese review feature is always left hand side of the opinion. For example, Price of MP3 player is little expensive instead of high price. All extracted noun and verb phrases not represent the feature of product.

To eliminating or remove the noisy data, means noun, verb or verb phrases which do not represent the feature of product can be solve by using the combination of probabilistic approach[13] and relationship between the feature and opinion words. The underlying principal of relationship between feature and corresponding opinions (noun, verb, verb phrases, adverb). However its generates

many feature which may not be represent the feature of product, but some frequently occurring words noun, verb, verb phrases, adverb for example. Nokia phone, camera, comments, advantage, problem, think, thought etc. To solve these most occurring words and phrases which are not represent the feature or opinion word by filter the low frequency of occurrence of feature and low frequency of occurrence of corresponding opinion words in addition to the probabilistic approach.

The main idea of this proposed approach is that each product field has specific language. For example Nokia phone have screen, touchpad, audio, video etc. Thus for any product have a higher probability of occurrence in the document belonging to than any other product field. Some words given above as: Nokia phone, camera, comments, advantage, problem, think, thought etc. Which are not represent the feature or opinion of any product field but have the similarity of probability of occurrence in the document of all the user-generated product review, therefore to solve this type of issue, generate a characteristics power value which would not be less than θ . and will not be considered as the feature or opinion, even though the frequency of occurrence in the document is high.

We use the probabilistic power equation to remove the noisy data which are occurs in terms of noun\noun phrases, verb\verb phrases, adverb.

$$X_i = \{cf \in C\} \cup \{co \in C\} \quad (1)$$

$$Y_i = \{cf \in G\} \cup \{co \in G\} \quad (2)$$

$$P(X_i) - P(Y_i) \geq \theta \quad (3)$$

$$P(X_i) = \frac{|cf \text{ in } C| + |co \text{ in } C|}{|W \text{ in } C|} \quad (4)$$

$$P(Y_i) = \frac{|cf \text{ in } G| + |co \text{ in } G|}{|W \text{ in } G|} \quad (5)$$

Where, cf represent the frequency of occurrence of feature word, co represent the frequency of occurrence of opinion word retrieved from the product review which we are trying to extract the feature and opinion for generate the feature\aspect based summary.

$$\{cf \in C\} \cup \{co\}$$

Here cf represent the frequency of occurrence of feature word, co represent the frequency of occurrence of corresponding opinion word in the specific product field.

$$\{cf \in G\} \cup \{co\}$$

Here cf represent the frequency of occurrence of feature word, co represent the frequency of occurrence of corresponding opinion word in the generic product field.

C represent the corpus of specific product.

G represent the corpus of Generic product.

The value of θ is adjusted based on the result. We choose a generic product in such a way that have completely repulsion or different feature with the product field. For example noun as video quality and opinion word as like, good etc. extracted from the specific product review (e.g. mobile phone), its characteristics calculated by the probability of feature word and opinion word in the specific product corpus C (mobile phone) minus that the generic product review G..

$P(X_i)$ represent as the frequency of occurrence of feature word and opinion word in a specific product review divided by the total number of words present in the specific product review.

$P(Y_i)$ represent as the frequency of occurrence of feature word and opinion word in a generic product review divided by the total number of words present in the generic product review.

In [14] methodology we have observed following shortcomings first for the mobile phone review using the noun "samsung" is the feature but "samsung" is not a feature of product, it is the name of product and the frequency of occurrence of "samsung" is less due to large, the characteristics power value of relatively depends upon both the specific product as well as the generic product [13], here the noun "samsung" is not present in the generic product review. The characteristic power of "samsung" become greater than θ , i.e. θ value is adjusted related to specific as well as generic product field. In contrast to this our proposed approach solve this problem by using the frequency of occurrence of feature and corresponding opinion, thereby remove the noisy data as all the feature and corresponding opinion words in the feature identification of the product. Secondly our probabilistic approach for extraction the domain specific feature at feature level or word level than at the document level [13]. The problem with probabilistic approach is, if a particular noun, accumulated in the domain specific few document and absent in all the document in generic product field. Hence the probability of occurrence of noun at document level always less than in the generic product field and high at the word level.

Infrequent feature extraction

People have different way of writing, sometime they will use phrases, idioms, and omission in dialogue process such as comments, we extract adjective using the part of speech tagged all the document for extract the infrequent or implicit feature but some adjective cannot find the corresponding feature. e.g. Not bad, you will buy it. The word (not bad) not the noun or verb phrases and cannot find the by the frequent feature extraction methodology to be modified. But this kind of opinion word can find the feature. For this type of opinion, we will use the context information of the review sentences. For example, the sentence (expensive, I cannot purchase it) the word (expensive) hinder the specific feature without know the context.

People use different word to express the same feature for example price, cost, rate, value and so on. Which one will be selected, everyone define the same feature, but

difference will disappear in the grouping of the feature in the clustering process.

That's why we don't use association rule mining technique used in [10,18]. Association rule mining approach has failed because it only extract the noun\ noun phrases as the feature indicator, but our proposed methodology uses the verb and adjective as the feature indicator and it also indicate the feature of domain specific product.

Prior research generally using the adjective as the opinion [10], we use the POS tagged XML document generated by the Stanford NLP parser to extract all the adjective (opinions) from the domain specific product review.

For the infrequent feature, the features which are corresponding to the same opinion group with implicit or infrequent feature are the candidate set. We select the feature with the highest valuable as the infrequent feature computed as follows:

$$\text{Weight}(z_i) = \sum_{f_i \in F_{i \in C}}(f_i) \sum_{f_i \in F_{i \in G}}(f_i) \quad (6)$$

$$\begin{aligned} \text{Hval}(\text{IF}) &= \text{Weight}(z_i) \\ \{ \text{sup}(P(X_i) - P(Y_i) \geq \theta) + (P(X_i) - P(Y_i) \geq \theta) \} \end{aligned} \quad (7)$$

After computing this we identify the semantic polarity of the adjective or opinions, the polarity of adjective has positive, negative and neutral, to find out the polarity of opinion we used the opinion lexicon [18,19]. This is the online dictionary which contain the large collection of positive, negative and neutral (+1,-1,0) adjective and for the certain non- familiar adjective, where opinion lexicon unable to detect the polarity, than we will use the another online service sentiwordnet [16][17].

The assignment of each opinion word to the feature word is achieved by computing distance of each opinion word to the feature in a sentence and assigned in opinion to the feature. This is the closest and less distance in case of two or more feature at the same distance assigned the opinion to the feature which is mentioned first.

When two or more opinion word at the same distance or in a sentence existing technique in [13] failed to deal such type of problem. This issue is solve by calculating the distance of each opinion word with respect to a feature than calculate all the opinion in a review sentence [18].

As there can be only one or more than one opinion word in sentence but feature word not mentioned, all the above mentioned existing technique[10,13,18] have failed to deal such type of issue. For example,(cheap, you can choose it.) and (expensive, short backup, sleek, not fit in my pocket). For that type of sentence where only opinion word "cheap" in 1st sentence and "expensive", "short backup", "sleek", "not fit in my pocket" in the second example not the feature word.

We developed a solution to this problem by considering the opinion word as a feature calculated according to equation-7. We computed the feature of product and then opinion is assigned to feature calculated.

Next we computed negation word in the neighborhood of each opinion (adjective), if any negation word is found, then the orientation of feature –opinion pair reversed..

There may be more than one opinion word in are view sentence. For example *I don't think camera is not bad*. Here, two negation word are (Not bad) the polarity is positive but when one more negation word before the (not bad) then final polarity of camera (feature) is negative.

The existing techniques have failed to deal this. Our algorithm helps in solve this problem,

The polarity of opinion at sentence level consisting of the following steps. We used one or multiple negation word (not, never, nothing, doesn't, don't, haven't, hadn't, can't, shalln't, isn't, willn't etc.), OW is opinion word (adjective, adverb, and verb) and opinion polarity (positive, negative and neutral).

Algorithm: Polarity of opinion at sentence level

Input: Review sentence. Here, OW represents the opinion word.

COUNT

(for multiple negation)

- Even number of negation and OW is negative than negative polarity.
- Even number of positive and OW is than negative polarity.
- Even number of negation and neutral than positive polarity.
- Odd number of negation and OW is negative than positive polarity.
- odd number of positive and OW is negative than negative polarity.
- odd number of negation and neutral than negative polarity.
- (for single negation)
- Negation and OW is negative than positive polarity.
- Negation and OW is positive than negative polarity.
- Negation and neutral than negative polarity.
- Positive and neutral than positive polarity.

Output: positive and negative polarity of opinion.

Clustering

In this phase first we grouped the word and phrases which are domain synonyms, because people uses different word and phrases to express the same feature by using the semi supervised COP-KMean [11] which is automatically partitioned the feature data into n feature groups. First we cluster the active product feature then transfer their cluster information as constraints (must-link, cann't-link) to cluster all the domain specific product feature.

$3n$ cluster are formed for n feature, one is the positive cluster, negative cluster and neutral cluster will store positive negative and neutral review respectively. Then extraction of feature- opinion pair (with polarity) and are placed into respective feature based cluster (positive, negative and neutral) based on their polarity whether the positive opinion, negative opinion about the feature or neutral one.

IV. RESULT ANALYSIS

We used Microsoft Visual Studio 2010 and SQL Server 2008 for implementation and result computation.

Overall methodology works in three different phases are: Phase-I: Extract the feature for each domain specific product then the feature word and the corresponding opinion word are selected from the product review sentences. Phase-II: Select the feature opinion pair. Phase-III: The feature opinion pair assigned or placed into their respective positive and negative cluster.

Our result of feature extraction compared with the association rule mining, probabilistic approach and combination of association rule mining and probabilistic approach on accuracy for different domain specific product and corpus.

Table 3. Comparison of different approach for feature extraction

Product name	Association mining[10]	Probabilistic approach [13]	Association mining+ probabilistic[18]	our
computer	81%	85%	87%	94%
phone	79%	82%	89%	93%
camera	80%	86%	90%	96%
MP3 Player	76%	80%	88%	91%

Table 3 show the result for camera, computer, phone and MP3 player on the feature detection and opinion polarity detection in terms of accuracy using the system. From the results presented in table-3 clearly shows that our methods of feature extraction outperforms other methods. Feature extraction using the our proposed approach achieves 94% accurate as compare to other methods. This achievement due to we used some pre-processing step (stop word removing, stemming and fuzzy machine technique). Second we use the frequent and infrequent feature extraction method. Third use the probabilistic approach at the word- level. Fourth grouping the domain synonyms feature than make 2n cluster (positive and negative) for each feature.

Table 4. comparison of different approach for opinion polarity detection

Product name	Probabilistic approach [13]	Association mining + probabilistic approach[18]	our
computer	76%	80%	89%
phone	70%	77%	86%
camera	79%	88%	92%
MP3	72%	75%	82%

Table.4 shows the comparison of results obtained for Opinion sentence polarity detection with our proposed methods with others. It is clearly evident from the results that polarity of 89% of opinion are correctly classified in the feature-opinion pair and 19% are wrongly classified whereas using the method of [13][18] only the 76% and 80% respectively correctly classify the polarity of opinion. Firstly the enhancement is due to the better way of find out the infrequent or implicit feature when only the opinion word not the feature word in the sentence, then assigned a correctly feature to each opinion word. Secondly by using the our proposed polarity detection

algorithm, it works for when not only the one negation word but also the two or more than two negation word present in the product review sentence.

Study of the effect of frequent changes in Review

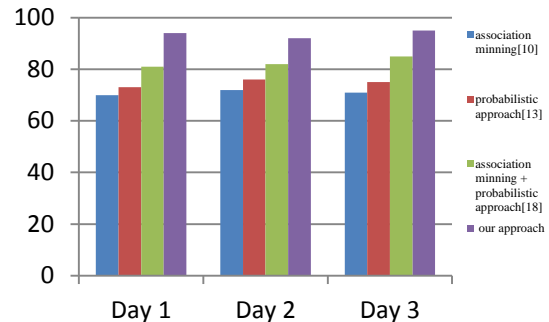


Fig.1. Change in accuracy with daily updation of customer review.

The fig.1, shows that our method obtain the better performance in terms of accuracy than the other method and more useful.

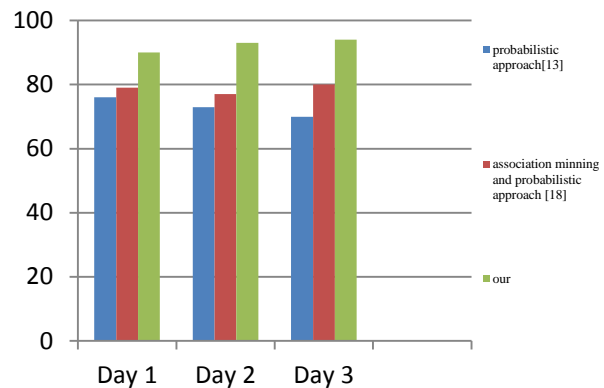


Fig 2. Change in accuracy with daily updation of review in polarity detection approach.

Fig.2, shows that our method obtain the better performance as compare to other methods and more useful.

From the above evaluation and comparison analysis, it can be seen that the proposed methodology provides better result than the other aspect based opinion mining used for the clustering and summarizing the product feature.

V. CONCLUSION

In this paper, we have proposed feature based summarization of online customer review in order to produce a relevant summary of domain specific product by extracting the frequent and infrequent feature using the probabilistic approach and also developed a final polarity detection algorithm when more than one negation word present in the review sentences. In this dynamic system we uses the stop word removing and stemming method and fuzzy machine technique for improve the accuracy and maximize the storage capacity of the system

and also retrieving rich and high quality of information with increase in accuracy for feature extraction and polarity detection. The experiment demonstrate that our approach achieve better performance for feature extraction and final polarity detection .it also very effective and more efficient for feature – based summary generation task.

REFERENCES

- [1] Y. Choi, Y. Kim, and S. Myaeng, "Domain-specific sentiment analysis using contextual feature generation," Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion, ACM, 2009.
- [2] C. L. Fermín, "A knowledge-rich approach to feature-based opinion extraction from product reviews," Proceedings of the 2nd international workshop on Search and mining user-generated contents, ACM, 2010.
- [3] M.Hu and B.Liu,"mining Opinion Feature in Customer Review," Proceedings of the 9th National Conference on artificial intelligence,2004.
- [4] [4]M.Hu and B.Liu,"Mining and Summarizing Customer Review," Proceedings of the international Conference on Knowledge Discovery and Data mining, , pp. 168-177,2004.
- [5] A.Popescu and O. Etzioni,"Extarcting Product Feature and Opinions form Reviews ," Proceedings of the Conference on Empirical Methods on Natural Language Processing ,pp.339-346,2006
- [6] S. Montazi,S. Kazalski, D. Klakow, "A Combined Query Expansion Technique for Retrieving Opinions from Blogs," Intelligent Systems Design and Applications, ISDA, Ninth International Conference on, pp. 791-796, 30 Nov 2009.
- [7] S. Moghaddam, M. Ester, "AQA: Aspect-based Opinion Question Answering," Data Mining Workshops (ICDMW),IEEE 11th International Conference on , pp.89-96,11 Dec. 2011.
- [8] B.pang and L.lee "Opinion Mining and Sentiment Analysis" Foundation and Trends in Information Retrieval,vol.2, pp1-i35, Jan2008.
- [9] J. Zhu, H. Wang, M. Zhu, B.K. Tsou, M. Ma, "Aspect-Based Opinion Polling from Customer Reviews," Affective Computing, IEEE Transactions on , vol.2, no.1, pp. 37-49, 2011.
- [10] M. Hiu and B.liu, "Miming and Summarizing Customer Review," International Conference on Knowledge Discovery and Data mining(KDD) USA,ACM,2004.
- [11] W. J. Jia, S. Zhang, Y.J. Xia, J. Zhang, H. Yu," A Novel Product Features Categorize Method Based on Twice-Clustering," Web Information Systems and Mining (WISM), 2010 International Conference on , vol.1, pp.281,284, 23-24 Oct. 2010.
- [12] Cardie, J. Wilson,et al. "Combining Low level and Summary Representation of Opinions for Multi Perspective Question answering" Spring symposium on New Direction in Question Answering, 2003.
- [13] S. Homocanu, M.Loster, Cristophlofi, W. balke , "Will I Like It-Providing Product Overview Based on Opinion Excerpts," Conference on Commerce and Enterprise Computing (CSE),Luxembourg,IEEE,2011.
- [14] S. Morinaga, K. Yamanishi, K.Tateshi, T. Fukushima, " Mining Product Reputation on the Web" 8th International conference on Knowledge Discovery and Date mining (KDD),Edmonton,Canada,ACM,2002.
- [15] Z. Zhai, B. Liu, H. Xu, & P. Jia, Clustering product features for opinion mining ,"Proceedings of the fourth ACM international conference on Web search and data mining", ACM, 2011.
- [16] A.Esuli, F. Sebastini, " SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining" 5th Conference on International Language Resource and Evaluation(LREC),Genoa,Italy,2006.
- [17] S. Baccianella, A.Esuli, F. Sebastini " SENTIWORDNET 3.0: An Enhanced Lexical Resource for and Opinion Mining" 7th Conference on International Language Resource and Evaluation(LREC), Marrakech Morocco European Language Resource Association (ELRA),2008.
- [18] K.bafana, D. Toshniwal "Feature based Summarization of Customers' Review of Online Products"17th International Conference in Knowledge Based and Intelligent Informationand Engineering System,ELSEVIER,PP 142-151,2013.
- [19] <http://aliasi.com/lingpipe/demos/tutorial/sentiment/readme.html>. LingPipe- tool Kit for Processing text using computational Linguistics.
- [20] <http://nlp.stanford.edu/software/tagger.shtml>. The Stanford Natural language processing group.
- [21] H. nakagaba, T. Mori, " A Simple but Powerful Automatic Term Extraction Method," International Workshop on Computational Terminology,Morristown,NJ,USA,2002.
- [22] R. Agrawal, R. Shrikant, " Fast Algorithm for Mining Association Rules" 20th International Conference on Very Large Database (VLDB), Santiago de Chile.

Authors' Profiles



Sanjoy Das did his B. E. and M.Tech, Ph.D in Computer Science. Presently, he is working as Assistant Professor, School of Computing Science and Engineering, Galgotias University, India since September 2012. Before joining Galgotias University he has worked as Assistant Professor at Computer Science and Engineering Department, G. B. Pant Engineering College, Uttarakhand, and Assam University, Silchar, from 2001-2008. His current research interest includes Mobile Ad hoc Networks and Vehicular Ad hoc Networks, Distributed Systems, Data Mining.



Bharat Singh, did his MCA from GLA University, Mathura, India and M.Tech in Computer Science from Galgotias University, India in 2011 and 2015 respectively. His current research area includes Data Mining and Opinion mining.



Ms. Saroj Kushwah, did her B.Tech and M.Tech in Computer Science from Uttar Pradesh Technical University, India in 2011 and 2014 respectively. Her current research area includes Data Mining and Opinion mining.



Prashant Johri is Professor & Director GIMT(MCA), Galgotias Institute of Management & Technology UP Technical University, Greater Noida, India. He has published several papers in International/National Journals and Proceedings. His areas of research are Software Reliability, Data Mining and Warehousing, Big Data

Security & Privacy , Big Data Open Platform and Pattern Recognition.

How to cite this paper: Sanjoy Das, Bharat Singh, Saroj Kushwah, Prashant Johri, "Opinion based on Polarity and Clustering for Product Feature Extraction", *International Journal of Information Engineering and Electronic Business*(IJIEEB), Vol.8, No.5, pp.36-43, 2016. DOI: 10.5815/ijieeb.2016.05.05