Aspect based Opinion Mining for Mobile Phones

Sonal Meenu Singh
Poornima University, Jaipur, Rajasthan, India
Singh.sonalmeenu@gmail.com

Nidhi Mishra
Poornima University, Jaipur, India
nidhi.mishra@poornima.edu.in

***Abstract*-With the explosion of Web 2.0, platforms such as blogs, e-commerce sites, peer-to-peer networks and social media, consumers have a broad platform and unlimited power to share their experiences in the form of reviews. With this, bulks of reviews are available for a single product, that needs to be analyzed, processed and mine. Opinion Mining or Sentiment Analysis is a Natural Language Processing and Information retrieval task that defines the customer's views or opinions through positive, negative or neutral sentiments. Aspect based Opinion Mining deals with aspects of the features. In this paper an aspect based opinion mining system is proposed that classify reviews as positive and negative. The system also deals with two aspects in a review. Experimental results using reviews of mobile phones show an accuracy of 75% as compared to other methods.**

Keywords: Opinion Mining; Semantic Orientation; polarity.

I. INTRODUCTION

Now a day, customers are showing their interest in online purchasing and for that they go for reading reviews. Reviews help users in decision making but instead of doing this, they are increasing the problem of customers due to availability of bulk of reviews. The customers get confuse and it becomes difficult to read all reviews in short time interval.

So, some work is needed to handle all these reviews. This task of classifying the reviews gives birth to Opinion Mining.

Opinion Mining (Semantic Analysis) is a sub part of data mining. Opinion Mining is a process of defining the experiences (by using polarities) of the users for particular product. Opinion Mining differentiates reviews into 3 polarities: positive, negative and neutral polarity.

Hu and Liu [4] define the components of the opinion mining such as:

Opinion holder: It is the one who writes the review.

Object: it is entity on which an opinion is expressed by user.

Opinion: it is an expression on an object by an opinion holder.

An **object** "O" is an entity which can be a product, person, event, organization, Each node represents a component and has set of attributes of the component. O is the root node. An opinion can be determined on any node or its attribute. Hu and Liu, use the features in representing both components and their attributes. They present the review of a model as follows:

1. An object O is represented with a set of features, $F = \{f1, f2 \ldots fn\}$.
2. Each feature fi in F can be expressed with a finite set of words or phrases Wi, which are synonyms. That is to say: we have a set of corresponding synonym sets $W = \{W1, W2, \ldots, Wn\}$ for the features.
3. An opinion holder j comments on a subset of the features $Sj \subseteq F$ of object O.
4. For each feature $fk \in Sj$ that j comments on, he/she, chooses a word or phrase from Wk to describe the feature, and expresses a positive, negative or neutral opinion on fk.

The Opinion Mining deals with two types of parameters. These two parameters are defined as:

- The objective parameter which includes number of reviews, ratings of the reviews, etc.
- The subjective parameters that deals with the emotions and feelings of the reviewers for a product.

The objective and subjective parameters can be used to calculate the score of the review.

Opinion Mining deals at 3 levels:

1. Document level
2. Sentence level
3. Aspect level

Document level: In this level of Opinion mining polarity of whole document is defined. Whole document is divided as of positive or negative polarity.

Sentence level: Each sentence from the review is defined as positive or negative sentiment.

Aspect level: In this level, each document/sentence is defined as positive or negative by considering the aspects of the products.

The Aspect based Opinion Mining deals with 2 types of reviews:

1. **Explicit aspect**: This review clearly defines the aspect for which it is written. For eg: Battery of Samsung is long-lasting.
2. **Implicit aspect**: It is the aspect which is not directly defined in the review

Proposed system focuses on extracting adverb-adjective pairs from the created database of reviews. The opinion words help in defining polarity of the aspects of different mobile phones. The system uses R programming for its working

which is a data analysis language. The rest of the paper is organized as follows: Section II describes the problem statement and objective of this paper. Section III describes related work. Section IV describes the architecture of proposed system. Comparison and experimental details are given in Section V and Section VI is about Future work and conclusion.

II PROBLEM STATEMENT AND OBJECTIVES

The **problem statement** of the system is "Aspect based Opinion Mining for Mobile Phones"

Objectives: The objectives of the system are listed as:

- To pre-processed the collected reviews
- To design different scenarios for experimentation.
- To design and implement one aspect based opinion mining.
- To design and implement two aspects based opinion mining.
- To compare the results with prior work.

III RELATED WORK

Sentiment Analysis is a hot topic among researchers. Many researches are going on aspect level of Opinion Mining. Venkata Rajeev P, et-al has proposed a system for comparing the aspects of a mobile phone [1]. R. Nithya, et-al used Naïve bayes supervised learning for classifying the mostly identified features and determined their polaity [2]. Lizhen Liu, et-al used LDA method and association rules to extract the feature and opinion words and cross validated them to remove unwanted feature and opinion pairs [3]. Minqing Hu, et-al proposed a system in which frequent features were determined using Apriori algorithm and their polarity was found by using bootstrapping. A summary was generated containing polarity of features that help users in easy decision making [4]. Sheng Huang, et-al extracted the features from the reviews and for that proposed system used CRF learning approach [5]. Yan Luo, et-al defined a framework for defining the opinion words. In the proposed system firstly web pages are extracted and manually label adjective opinion word set and identified comparative sentences and negation patterns. Finally, a summary was generated [6]. The proposed system finds the opinion words using Semantic role labeling and determined its semantic orientation and it was found that results were feasible [7]. Wei Wang proposed a Semi Supervised C0-LDA model to obtain the positive and negative opinion words to find the relationship between opinion and feature words [8]. Warih maharani used SentiWordNet based method to extract the opinion words was effective and correctly identified their polarities [9]. Zhixing Li et-al proposed a method which combines the associative classification methodology with the overall rating to find the relationship between opinion and feature words [16].

IV PROPOSED SYSTEM

The proposed system is based on the adjective-adverb pairing, where dictionary based approach is used to determine the orientation of the reviews (positive / negative) collected from e-commerce websites. Aspects from the reviews are manually created. The opinion words are extracted from the reviews and their polarity is determined. At last score of reviews is calculated. The score of aspect helps users in deciding which aspect of a mobile is best among all. The system works for both one and two aspects containing mobile reviews. The steps of processing are described below:

A. Proposed system for one aspect containing mobile reviews

To find the polarity of one aspect in a review, following steps are carried out:

1. Input

Reviews from e-commerce websites are provided as input to the system.

2. Pre-Processing

The reviews are pre-processed which include: stopword removal, white space removal, special symbol removal, POS tagging. For POS tagging Stanford parser is used.

3. Aspect Word List

All the aspects from the reviews are identified manually. There are eight aspects identified from the reviews, which are cost, size, battery, camera, OS, processor, storage and screen.

4. Opinion Word List

After POS tagging of reviews, opinion words (include all words of review excluding the aspect word) are identified and stored in database. The score of each word after pre-processing, excluding the aspects is identified through SentiWordNet 3.0. For calculating score of the word, the score of all synsets that have same part of speech as of the word are computed and averaged to give final score of the word.

5. Polarity Detection and Scoring

After finding score of each word, the score of mobile reviews (based on aspect) is calculated. If the calculated score of the review is greater than zero, than the aspect of mobile is commented as positive. If calculated score of the review is less than zero, than the aspect of mobile is commented as negative. Otherwise, the aspect of mobile is commented as neutral.

The system deals with three special cases, like:

- **Negation handling**: Some words like not, never, neither when attached with some positive or negative word they will change its meaning. The proposed system handled negation by reversing the polarity of opinion words.

- **Modifiers / Intensifier handling**: The words (like very, so, much etc) that provide more strength to positive words and weakness to negative words are called modifiers. The modifiers are handled in the system by adding 0.1 score to the positive modifier words and keeping negative modifier words unchanged.
- **Synonyms**: These are the words with same meaning. In the proposed, synonyms fro "cost" and "size" is handled.

The Algorithm of Proposed System for One Aspect Containing Mobile Reviews:

Step 1: Extract the reviews from the database.
Step 2: Pre-processing like stopword removal, white space removal, special symbol removal of the extracted reviews.
Step 3: Perform Part Of Speech tagging of the pre-processed reviews.
Step 4: Create opinion word list from the POS tagged reviews.
Step 5: Compute the score of word through SentiWordNet dictionary, where the score of all synsets that have same part of speech as of the word are computed and averaged to give final score of the word.
Step 6: Calculate the score of the aspect based review by using step 5.
Step 7: if calculated total score is > 0 then
Polarity of aspect based review is positive.
Else if calculated total score is <0 then
Polarity of aspect based review is negative.
Else polarity of aspect based review is neutral.

B. Proposed system for two aspect containing mobile reviews

The system also handle reviews containing two aspects with coordinating conjunctions like but, and, or, while, rather. The reviews are extracted from database and split into two clauses when a coordinating conjunction is present in the review. After splitting, each clause is pre-processed (stopword removal, special symbol removal, white space removal). Pre-processed review is tagged and an opinion word list is generated through SentiWordNet. Afterwards, for calculating the score of reviews same processing is applied as used for one aspect containing mobile reviews.

Algorithm of the Proposed System for Two Aspects Containing Mobile Reviews

Step 1: Extract the reviews from the database.
Step 2: If coordinating conjunction is present in review, split it into two clauses (each having one aspect).
Else review containing one aspect.
Step 3: Pre-processing of the extracted reviews
Step 4: Perform Part Of Speech tagging (through Stanford Parser) of the pre-processed reviews.
Step 5: Create opinion word list from the POS tagged reviews.
Step 6: Compute the score of word through SentiWordNet dictionary, where the score of all synsets that have same part of speech as of the word are computed and averaged to give final score of the word.
Step 7: Calculate the score of the mobile review by using the score generated from step 6.
Step 8: if calculated score is > 0 then
Aspect of mobile is commented as positive
Else if calculated score is <0 then
Aspect of mobile is commented as negative.
Else aspect of mobile is commented as neutral.

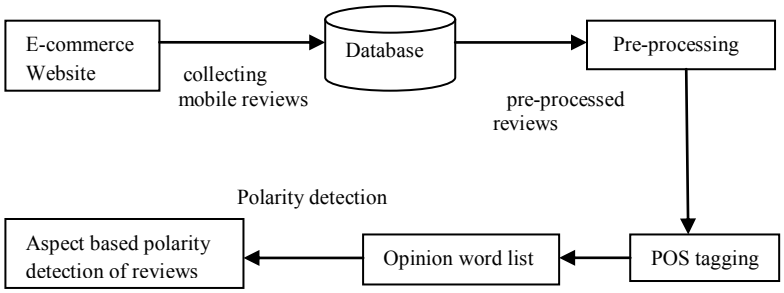
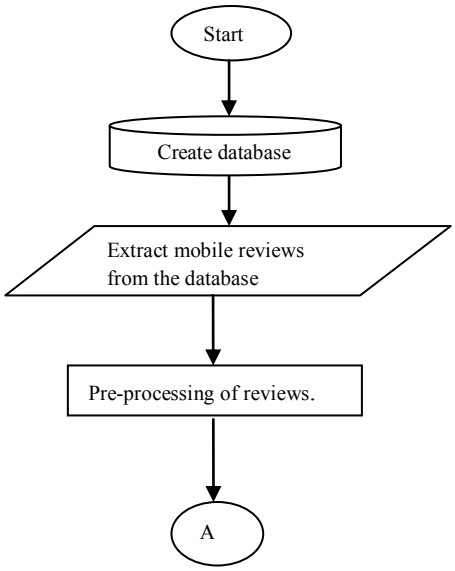


Fig. 1 Architecture of Proposed System

IV FLOWCHART

The flowchart of the proposed system is described as below:



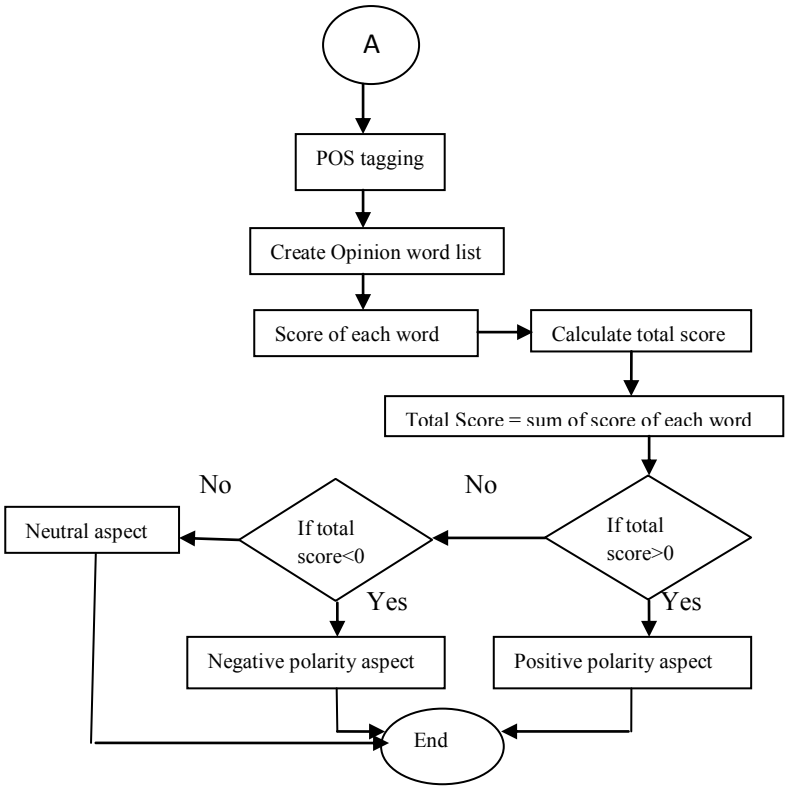


Fig. 2 Flowchart of Proposed System

V EXPERIMENTAL DETAILS

The experiment is conducted by collecting 80 mobile phone reviews from www.Amazon.com and www.CNET.com websites. The reviews are of 3 mobile phones: Iphone 6, Moto G3and blackberry Z10. The manually identified aspects are screen, size, battery, storage, camera, OS, processor.

Proposed system is compared with three online tools (WordNet, SentiWordNet and MPQA). [13]

In the table 1, two types of symbols are used to show the results of input reviews. The meaning of these symbols is described as:

✓ Means that all systems identified review correctly. That means, when a positive review is given as input than system detect it as of positive polarity. When negative review is given then the system detect it as of negative polarity.

X means that system failed to correctly define the sentimental orientation of the input.

Table 1 A Sample of Comparison Between Existing Online Tools and Proposed System

S No.	Reviews of different mobile phones	Word Net	Senti Word Net	MPQA	Proposed approach
1.	Battery life of iphone 6s is amazing.	✓	✓	✓	✓
2.	Extraordinary camera of iphone 6s mobile.	X	✓	✓	✓
3.	Excellent iOS operating system provided in iphone6.	X	✓	✓	✓
4.	The battery life of iphone 6 is not better.	✓	✓	X	X
5.	iphone 6 provides very high storage capacity of 128 GB	X	✓	X	✓
6.	The screen size of iphone 6 is very big can be squeezed to small.	✓	✓	X	✓
7.	The cost of iphone 6s is much.	✓	X	X	X
8.	Battery capacity of iphone 6 is endurance on a single charge.	✓	✓	X	✓
9.	Iphone 6 provides small storage space.	X	✓	X	✓
10.	The cost of iphone 6s is little.	X	X	X	X

The results obtained after comparison are summed up in the table 2 and a relevant bar graph of the obtained result is described in figure 2

Table 2 Results from Comparison of Proposed Approach with Existing Sentiment Analysis Tools

Approaches	Retrieved Positive reviews	Actual positive reviews	Retrieved negative reviews	Actual negative reviews
WordNet [13]	15	50	13	30
SentiWordNet [13]	38	50	13	30
MPQA [13]	33	50	15	30
Proposed Approach	41	50	19	30

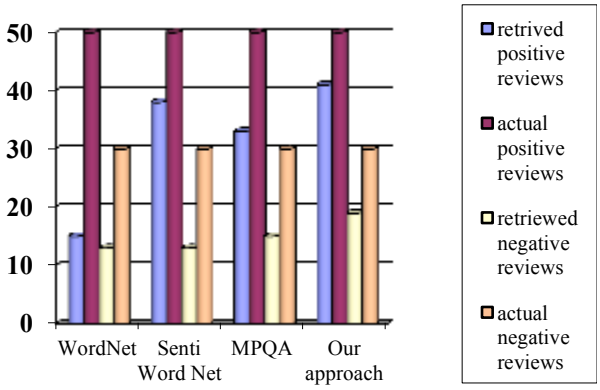


Fig. 3 Comparison of Results of Different Approaches

The above bar graph shows the effective performance of proposed system with other three approaches.
The proposed system is compared with some of the existing online tools and the results of accuracy is pin up in table 3

Table 3 Accuracies of Different Approaches

S No.	Applied approaches	Accuracy %
1.	WordNet [13]	35%
2.	SentiWordNet [13]	64%
3.	MPQA [13]	60%
4.	Proposed system	75%

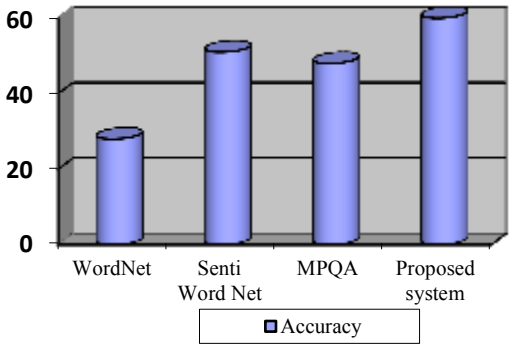


Fig. 4 Accuracies of Different Systems

In the figure 3, proposed system shows an accuracy of 75% which is very effective and highest among the all other approaches.
Table 4 below, describes the precision and recall value for identifying the positive reviews by the four systems.

Table 4 Performance Evaluation of Different Approaches using Precision and Recall Value of Positive Review

S No.	Applied approaches	Precision value for positive reviews	Recall value for positive reviews
1.	WordNet[13]	0.3	0.681
2.	SentiWordNet[13]	0.76	0.691
3.	MPQA[13]	0.66	0.687
4.	Proposed system	0.82	0.788

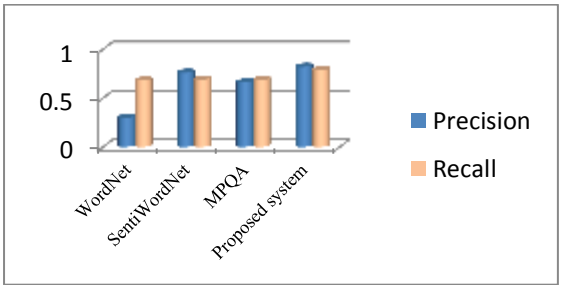


Fig. 5 Precision and Recall Value for Positive Reviews by all Approaches

From the figure 4, it can be concluded that proposed system shown highest precision and recall value for positive reviews.
The precision and recall value for correctly identifying negative reviews is shown below in Table 5

Table 5 Performance Evaluation of Different Approaches using Precision and Recall value of Negative Review

S No.	Applied approaches	Precision value for negative reviews	Recall value for negative reviews
1.	WordNet[13]	0.433	0.271
2.	SentiWordNet [13]	0.433	0.52
3.	MPQA [13]	0.5	0.468
4.	Proposed system	0.633	0.678

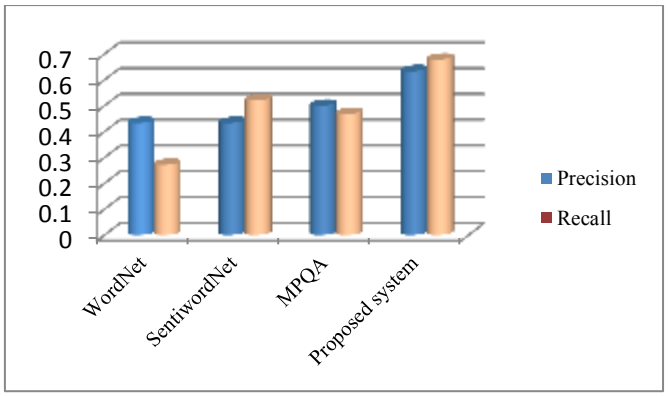


Fig. 6 Precision and Recall Value of Negative Reviews for All Approaches

From the above graph it is clear that proposed system shown highest precision and recall value for negative reviews.

Table 6 Comparison Among Aspects of Different Mobile Phones for Proposed Approach

S. no	Aspects	Mobile phones		
		Iphone 6	*Moto G3*	*Blackberry Z10*
1.	Camera	0.60	0.66	0.83
2.	OS	1.00	0.66	0.67
3.	Battery	0.80	0.66	0.33
4.	Process or	1.00	1.00	0.67
5.	Screen	0.75	0.66	1.00
6.	Size	0.75	0.83	1.00
7.	Cost	0.00	0.50	0.00
8.	Storage	1.00	1.00	0.40

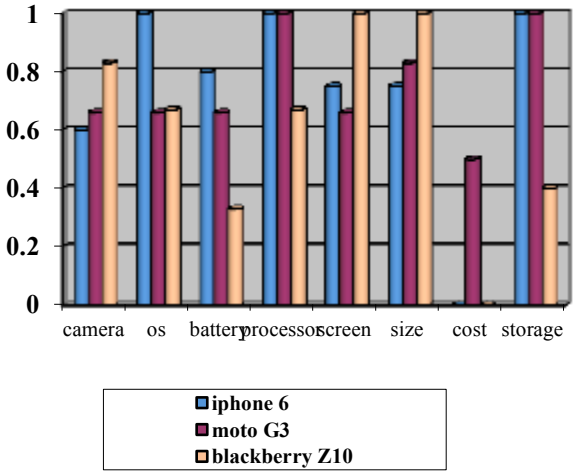


Fig. 7 Comparison Among Aspects of Different Mobiles.

This comparison among different aspects of mobiles helps user in deciding which aspect of particular mobile is best and useful.

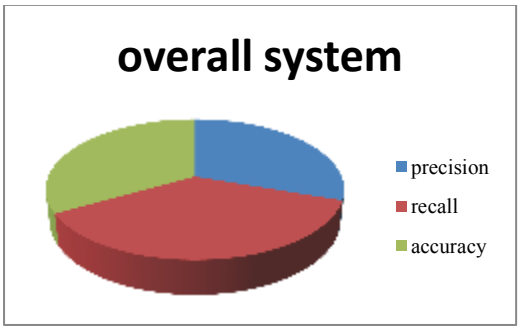


Fig. 8 Overall Performance of the Proposed System

The overall results of the "Aspect based Opinion Mining for Mobile Phones" system are effective with accuracy of 75%. Hence, it achieved its target in correctly defining the Semantic Orientation of Aspect based reviews.

VI CONCLUSION

The objective of this paper is to determine the Semantic orientation of aspect based reviews. The system achieved the target with efficiency of 75%. Our proposed system is compared with existing online tools to measure the performance of proposed system. Proposed system successfully handled negation, intensifiers and synonyms which were not effectively deal by the other system and also the proposed system is effective in defining the semantic polarity of input mobile reviews. After achieving such high efficiency the system has a drawback that it failed to correctly define polarity of the "cost" aspect. The system will be made more effective to handle this drawback. In future, we will work on implicit feature, clustering more than two aspects without using coordinating conjunctions, handling synonyms in better way.

REFRENCES

[1] Venkata Rajeev P, Smrithi Rekha V, "Recommending Products To Customers using Opinion Mining of online Product Reviews and Features" in Proceedings of IEEE International conference circuit, power and computer technologies, pp. 1-5, Nagercoil 2015

[2] R Nitya, D Maheshwari, "Sentimental Analysis on Unstructured Reviews", in Proceedings of IEEE International conference on Intelligent Computing Application, pp. 367-371, Coimbatore , 2014.

[3] LiZhen Liu, WenTao Wang, Hangshi Wang, "Summarizing Customer reviews based on Product Features", in Proceedings of 5th international congress on Image and Signal Processing, pp. 1615-1619, Chongquing 2012.

[4] Minqing Hu, Bing Liu, "Mining opinion features in Customer reviews", in the proceedings of 19th National Conference on Artificial Intelligence pp. 755-760, San Jose, California 2008.

[5] Sheng Huang, Xinlan liu, Xuepang Peng, "Fine-grained Product features Extraction and Categorization in Reviews Opinion Mining", in Proceedings of 12th IEEE International Conference on Data Mining Workshops, pp. 680-686, Brussels 2012

[6] Yan Luo, Wei Huang, "Product Review Information Extraction Based on Adjective Opinion Words", in Proceedings of 4th IEEE International Joint Conference on Computational Sciences and Optimization, 2011.

[7] Wei Wang," Sentiment Analysis of Online Product reviews with Semi-Supervised Topic Sentiment Mixture Model", in Proceedings of 7th IEEE International Conference on Fuzzy Systems and Knowledge Discovery, pp. 680-686, Brussels, 2010.

[8] Weishu Hu, Zhiguo Gong, "Mining Product Features from Online Reviews" in Proceedings of IEEE International Conference on E-Business Engineering, pp. 24-29, Shanghai, 2010.

[9] Warih maharani, Dwi h.Widyantoro, Masayu L. Khodra, "SAE: Syntactic - based Aspect and Opinion Extraction from Product Reviews", ", in the proceeding of 2nd IEEE International Conference on Advanced Informatics Concepts, Theory and Applications, pp. 1-6, Chonburi , 2015

[10] Warih Maharani, Dwi h.Widyantoro, Masayu L. Khodra, "Learning-based aspect identification in customer Review Products", in Proceedings of 5th IEEE International Conference on Electrical Engineering and Informatics, pp. 71-76, Denpasar, August 10-11, 2010.

[11] Jian Liu, Jianxin Yao, "Opinion Searching in Multi-product Reviews", in Proceedings of the 6th IEEE International Conference on Computer and Information Technology, pp. 25-28, Seoul 2006.

[12] Jung-Yeon Yang, Jaeseok Myaung, "The Method for Summarization of Product Reviews using User's Opinions", in Proceedings of the IEEE International Conference on Informatics, Process, and Knowledge Management, pp. 84-89, Cancun, 2009.

[13] http://sentiment.christopherpotts.net/textscores.

[14] Zhixing Li, "Product Feature Extraction with a

Combined Approach", in Proceedings of the 3rd IEEE International Symposium on Intelligent Information Technology and SecurityInformatics, pp. 686-690, Jinggangsgan, 2015.

[15] Weiping Wang, Yuanzhuang Zhou, "E-Business Websites Evaluation Based on Opinion Mining", in Proceedings of the IEEE International Conference on Electronic Commerce and Business Intelligence, pp. 87-90, Beijing, 2009.

[16] Weishu Hu, Zhiguo Gong, "Mining Product Features from Online Reviews" in Proceedings of IEEE International Conference on E-Business Engineering, pp. 24-29, Shanghai, 2010.

[17] Xinghua Hu, Bin Wu, "Classification and Summarization of Pros and Cons for Customer Reviews", in Proceedings of IEEE International Conference on Web Intelligence and Intelligent Agent Technology, , pp. 73-76, Milan, Italy, 2009.