International Journal of Computer Applications Technology and Research
Volume 6–Issue 8, 349-354, 2017, ISSN:-2319–8656

Effective Co-Extraction of Online Opinion Review Target-Word Pairs and Product Aspect Ranking

Chandana Oak
Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India

Prof. Manisha Patil
Department of Computer Engineering, Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India

Abstract: The popularity and use of e-commerce are increasing day by day. Recent trends have shown that many people are now buying their products online through different e-commerce sites such as Flipkart, Amazon, Snapdeal, etc. Customers review and rate the products they have bought over multiple independent review sites such as gsmarena, social networking sites like Facebook, blogs, etc. Customers can also comment on the quality of service they have received after buying their product. These online reviews are of immense help to potential buyers for that product in decision making and also to manufacturers/sellers to get an immediate feedback about the product quality, product performance, after sales service, etc. As the number of reviews for a product is usually large, it is next to impossible to go through all the reviews and form an unbiased opinion about the product. Also, there are multiple sources of these online reviews. Hence, online review mining is gaining importance.

A product can have many product features, wherein some features are more important than others. Review usefulness can also be increased by ranking the product aspects as per their importance and popularity. Ranking product aspects manually is very difficult since a product may have hundreds of features. So, an automated method to do this is needed.

This paper presents a methodology for co-extracting opinion targets and corresponding opinion words from online opinion reviews as well as for product aspect ranking.

Keywords: Online review mining, Opinion mining, Opinion Co-extraction, Opinion Target, Opinion Word, Product Aspect Ranking, Sentiment Analysis

1. INTRODUCTION

The importance and popularity of e-commerce is increasing day by day as a lot many people have started buying their products online. The convenience of online shopping, a wide variety of different product ranges, huge discounts given by the retailers, etc. are some of the reasons which have contributed to enhancement in e-commerce popularity. While the trend of online shopping is increasing, logging of opinion about the product bought, has also been gaining importance. Customers who have bought a particular product, say mobile phone, log their opinion about the phone either on the ecommerce sites such as Flipkart, eBay, etc., independent review sites such as gsmarena, blogs or social networking sites such as Facebook or Twitter, etc. These online reviews are of immense help to potential buyers for that product in decision making whether to buy the product or not, and also to manufacturers/sellers to get an immediate feedback about the product quality, product performance, after sales service, etc. For any product, the number of online reviews is usually pretty large. Also, most of the reviews are verbose. Any potential customer or a manufacturer can find it very difficult to go through all those reviews and form an unbiased opinion about the product. Hence, opinion mining or sentiment analysis is proving to be extremely useful.

Additionally, a product has numerous aspects associated with it. For example, iPhone 3GS has more than three hundred aspects, such as “usability," “design," “application," "3G network". Some aspects are more important than the others, and have greater impact on the potential consumers’ decision making as well as on product development strategies. For example, some aspects of iPhone 3GS, like, “usability" and “battery," are more important than the others such as “usb" and “button”. Online reviews might focus on some specific products rather than on the overall product itself. However, the reviews are often disorganized, leading to difficulties in information navigation and knowledge acquisition. Potential customers can make wise purchasing decisions by focusing on some important aspects rather than on some less important ones. Similarly, manufacturers/sellers can also focus on improving the quality of these more important products by enhancing the overall product reputation. But, considering the large number of features a product has, it is impractical to manually identify these important aspects from a large number of online reviews. Hence, a method to identify these important products is much necessary.

1.1 Opinion Mining

Opinion mining or sentiment analysis is related with mining and analysis of natural language for tracking the mood or feedback of people about a particular product. It can be treated in short as a system to collect and classify different opinions about a particular product or service. [1] It can help marketers evaluate the success of an ad campaign or new product launch, determine which versions of a product or service are popular and identify which demographics like or dislike particular product features. For example, a review on a website might be overall positive about a digital camera, but can be specifically negative about how heavy it is. Being able to identify this kind of information in a systematic way gives the vendor a much clearer picture of public opinion than surveys or focus groups do, because the data is created by the customer.

[2] gives the generic architecture for opinion mining.

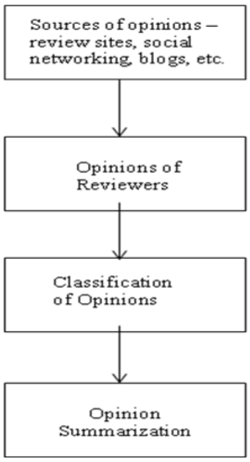

Fig 1: Generic architecture for Opinion Mining

People log their opinions on different products on online ecommerce sites, social networking sites, blogs, etc. Thus, there can be multiple sources which a potential buyer can visit for obtaining first hand information of the product which he/she is interested in buying. Information retrieval techniques like web crawling, etc. can be used to extract opinions from all these sources and can be stored into a database. These extracted opinions are analyzed and classified as per their polarity i.e. if the opinion is positive or negative.

1.2 Product Aspect Ranking

A potential customer is mainly interested in the most important and popular features or aspects of a product. E.g. while buying a smart phone, he/she will look for screen resolution rather than usb. Thus, some product aspects hold more importance than others and they can also influence the overall rating for the product. Ranking product features would help increase the usefulness and effectiveness of online reviews. But, it is next to impossible to identify and rank these product features as per their importance manually. Hence, an automated method to do so is much needed.

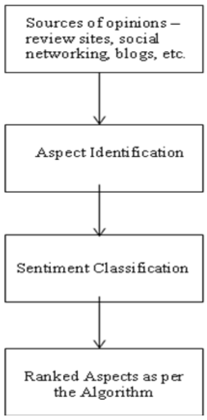

Fig 2: Generic architecture for Product Aspect Ranking

Fig 2 gives the generic architecture for product aspect ranking. Opinions can be extracted from different online ecommerce sites, social networking sites, blogs, etc. The next step is for aspect identification in which all the different product features are identified. The, these aspects are classified as either positive (i.e. good) or negative (i.e. bad) based on the comments that these features receive. This process is called as sentiment classification. Then these aspects are ranked based on some ranking algorithm.

2. MOTIVATION

Many online reviews are verbose, which reduces the readability and thereby the interest of any person reading the reviews. Also, there are many reviews for a product, thus making it difficult for a reader to go through all the reviews and derive meaningful information from them. Then the potential customer might go through just a few reviews making the opinion biased. Additionally, users can express their opinions either on e-commerce sites such as Snapdeal, Amazon or Flipkart, etc., or social networking sites, or blogs, etc. It makes it very difficult for a person to browse multiple sites to get all the reviews. Thus, opinion mining has become the need of the hour.

Product aspect ranking organizes the extracted features (such as processor speed, battery life, connectivity, screen resolution, etc.) of a product as per their importance to improve usefulness of reviews. But, identifying these features manually and then ranking them as per their importance is very difficult due to the large number of reviews. Thus, an automated method to extract and rank the important aspects of a product is much needed.

3. LITERATURE REVIEW

The importance and popularity of online review mining are increasing day by day. [2] gives an overview of different methods used in opinion mining and product aspect ranking. [3] proposes association mining technique to identify the most frequently occurring nouns and noun phrases in a review sentence. The ones which have a high occurrence frequency are then extracted as opinion targets. [4] proposes a Word based Translation Model (WTM) which is a graph based algorithm for extracting opinion targets. WTM is more effective than traditional opinion target extraction methods viz. adjacent methods and syntax based methods. This WTM method is independent of parsing performance, which plays a major role in syntax based methods and also of window size which is used in adjacent methods to find opinion relations with the corresponding opinion words. [5] proposes a method called Double Propagation which extracts opinion words or opinion targets iteratively from the words and targets already extracted during the previous iteration using syntactic relations. [6] highlights nearest neighbor rule which considers nearest adjective/verb to a noun as its opinion word. Thus, the span on which it works is limited and accuracy of results reduces.

[7] and [8] are based on sentence level extraction. In [7], Conditional Random Fields (CRFs) are used to jointly extract product aspects and positive/negative opinions from online review sentences. Phrase dependency parsing is used in [8] as many product features such as processor speed, screen resolution, wifi connectivity, etc. are noun phrases rather than just nouns.

An OPINE algorithm is used in [9] which proposes syntactic parsing of reviews to find opinion relations among words. But it is very much dependent on parsing performance which is affected in case of informal writing style of online reviews.

[10] uses a partially supervised word alignment model (PSWAM) which co-extracts opinion targets and opinion words based on a graph-based co-ranking algorithm. This algorithm extracts opinion words and targets more accurately than nearest neighbor rules or syntax based methods.

[11], [12], [13], [14] and [15] focus on product aspect ranking and its different applications. [11] proposes double

International Journal of Computer Applications Technology and Research
Volume 6–Issue 8, 349-354, 2017, ISSN:-2319–8656

propagation for product feature extraction along with an algorithm for ranking of products based on feature importance. [12] and [13] use an aspect ranking algorithm which considers both the aspect frequency as well as their influence on overall opinion of the product.

Table 1: Summary of Related Work

Reference	Technique used for feature extraction	Is product feature ranking used?
Mining and summarizing customer reviews [3]	Association mining	No
Opinion target extraction using word based translation model [4]	Word Translation Model	No
Opinion word expansion and target extraction through double propagation [5]	Double Propagation	No
Mining opinion features in customer reviews [6]	Nearest neighbor rules	No
Structure-aware review mining and summarization [7]	CRF	No
Phrase dependency parsing for opinion mining [8]	Phrase dependency parsing	No
Extracting product features and opinions from reviews [9]	OPINE	No
Co-Extracting Opinion Targets and Opinion Words from Online Reviews Based on the Word Alignment Model [10]	PSWAM	No
Extracting and ranking product features in opinion documents [11]	Double propagation with part-whole relationship	Yes
Aspect ranking: Identifying important product aspects from online consumer reviews [12]	Stanford parser and SVM classifier	Yes
Product Aspect Ranking and Its Applications [13]	Stanford parser and SVM classifier	Yes

The proposed system uses the following - The product reviews are partially parsed using Stanford parser. The "Opinion TW Co-extraction Algorithm" is proposed to co-extract opinion targets and corresponding opinion words from classified and semi-supervised product reviews. In addition to this, product aspect ranking is also included as part of proposed work.

4. SYSTEM OVERVIEW

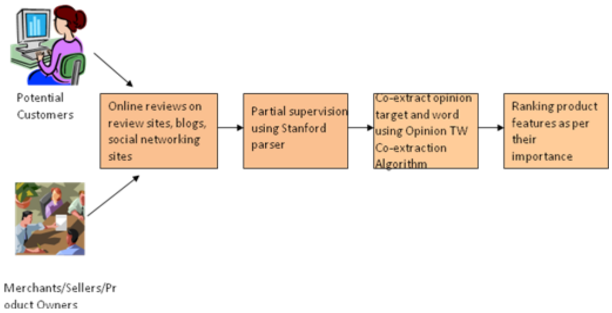

Fig 3: Architecture for Proposed System

4.1 Modules

Fig. 3 gives the overall architecture for the proposed system.

Millions of reviews of multiple products are available online on various sites such as different e-commerce sites like eBay, Flipkart, Amazon, etc., or also on independent review websites like gsmarena, rottentomatoes, etc. Users of products can also express their opinion through social networking sites like twitter, Facebook, etc. as well as over blogs. These online reviews are referred to by potential consumers in order to get review of the product. Also, merchants/sellers refer to these reviews in order to get a first hand feedback about their product/service. All these sources will act as the source of online reviews for co-extraction algorithm.

4.1.1 Data Processing

The Data Processing module is partial supervision using Stanford parser. Here, the input review data is fed to Stanford parser module to extract opinion targets (i.e. nouns or noun phrases) and opinion words (i.e. adjectives). Here, partial supervision technique is being used which is more advantageous over completely unsupervised technique. A completely unsupervised technique shows poor results since it does not have any training data to start with. Partial supervision is also not heavily dependent on parsing performance like a completely supervised parsing technique does. Here, partial parsing of review data would be done using Stanford parses. Of the product review data, 50% of the review data would be considered for training the parser and remaining 50% data would be test data.

4.1.2 TW Co-extraction

The second module is for Co-extraction of opinion target and opinion word. Co-extraction of opinion target and opinion words would be done using "Opinion TW Co-extraction Algorithm". This algorithm takes the partially supervised set of classified product reviews as input.

Algorithm 1: Opinion TW Co-extraction Algorithm

Input: Partially supervised set of classified product reviews I = {p1, p2, …, pn}
Output: The probability of co-alignment for sentences based on optimal association score

```
1.  Initialization:
2.  Initialize R as review data of products
3.  Initialize review = i[data]
4.  Initialize o_word [ ] = review [opinion word]
5.  Initialize o_target [ ] = review [opinion target]
6.  for each review[]
7.      for each o_target[]
8.          Nt := Count (opinion targets)
9.      end for
```

```
    for each o_word[]
        Nw := Count (opinion words)
    end for
    Ntw := Count (collocated opinion target and opinion word)
    P (o_target | o_word) = Ntw / Nw

    P (o_word | o_target) = Ntw / Nt
    Optimal Association Score OAS = [P (o_target | o_word) + P (o_word | o_target)] / 2
end for
Co-extract the opinion target and opinion word with higher OAS as being aligned with each other
```

The input to Opinion TW Co-extraction Algorithm is the set of product reviews which is partially parsed. Consider n to be the total number of product reviews. R denotes the review collection per product. For each category of product review, assign "review" collection with the partially parsed review data for that product. Fetch the opinion words from review into o_word based on training data. Similarly, fetch the opinion targets from review into o_target based on training data.

From the entire corpus, find Nt to be the total number of opinion targets and Nw to be the total number of opinion words. Also find the count of collocated opinion target and opinion word as Ntw. Find the estimated alignment probability for a potential opinion target and word pair as P (o_target | o_word) and P (o_word | o_target). To find the optimal association score (OAS), we take a mean of the 2 probability values fetched earlier. Co-extract the opinion target and opinion word with higher OAS as being aligned with each other.

To find one-many relation between opinion targets and opinion words (E.g. – The ambience and food in this restaurant are very good – here, good indicates both ambience and food), we find 2 opinion targets separated by a conjunction and aligned to one opinion word. Similarly, find 2 opinion words separated by a conjunction and aligned to one opinion target. Co-extract these as opinion word and target having one-many relation.

4.1.3 Product Aspect Ranking

The third module is Product Aspect Ranking. A product has many features. E.g. – if we consider mobile phone as a product, it has features like processor speed, screen resolution, battery life, wifi connectivity, etc. Similarly, if we consider a camera, it has aspects like shutter speed, lenses, picture quality, etc. These product features can be treated as opinion targets. Of all the features that a product has, some can be more important than others. A potential customer gives more importance to these features rather than giving much focus on the less important ones while buying any product. Also, manufacturers can focus mainly on the more important features while deciding on their product development strategies. Ranking these features as per some parameters, like frequency at which the product features were commented on, etc. would help increase the usefulness of online reviews. But, since the number of product reviews and also the number of features a product has is large, it is next to impossible to rank these features manually. The proposed work build a ranking framework on top of co-extraction framework which will organize all the product features as per their popularity i.e. as per the number of reviews a product feature receives, and also as per the influence a feature has on the overall opinion of the product. This will improve the usability of the review summarization. We will use the same set of opinion targets extracted by the Opinion TW Co-extraction Algorithm as product aspects or features and Naïve Bayes classifier for identifying the polarity (i.e. whether the reviews are positive or negative) of the feature expressed in the review based on opinion word extracted for the feature.

Algorithm 2: Product Aspect Ranking Algorithm

```
Input: Product name and opinion target-word pairs
Output: Ranked product aspects
Initialization:
Initialize set of product reviews as R
Initialize Negative sentimental word set N[]
Initialize Positive sentimental word set P[]
Initialize set of product aspects as aspect[]
Initialize set of overall ranking of aspect Or[]
for each aspect[]
    Initialize array for sentiments as S[]
    Int i = 0
    For each i
        If S[i] is negative then
            N[i] = S[i]
        Else if S[i] is positive then
            P[i] = S[i]
        End if
    end for
end for
Overall rating of product as o[]
Calculate Importance weight for aspect wt[] as frequency of the aspect commented on
Or[] = o[] * wt[]
Rank the product aspects as per Or[] value
```

4.2 Mathematical Model

1. Co-extraction of opinion target and opinion word –

   $$P(\text{o-target} \mid \text{o-word}) = Ntw / Nt \quad (1)$$

   $$P(\text{o-word} \mid \text{o-target}) = Ntw / Nw \quad (2)$$

   $$OAS = [P(\text{o-target} \mid \text{o-word}) + P(\text{o-word} \mid \text{o-target})] / 2 \quad (3)$$

   Where,

   Nt := Count (opinion targets)

   Nw := Count (opinion words)

   Ntw := Count (collocated opinion targets and words)

2. Product Aspect Ranking –

   Consider the opinion of various aspects as o[]

   Consider the importance weight for aspect as wt[]

   Then, product rank can be found as –

   $$Or[] = o[] * wt[] \quad (4)$$

   Rank the product aspects as per Or[] value.

4.3 User Interface of the Application

The graphical user interface of the application is simple and self-explanatory. The user has to provide the user id and password to the application as the system allows only authorized users to enter. After authentication, the user is asked to select the product which he/she wishes to get opinions on. The display window will have all the opinion target-word pairs displayed which are extracted from a set of online reviews along with their OAS, using which the user can form an opinion about the product. Then, on the Product Aspect Ranking screen, the user would see a list of product features ranked as per their importance along with their polarities. This summarization would assist the user in decision making.

5. RESULTS AND DISCUSSION

Place Co-extraction of opinion target and corresponding opinion word as well as product aspect ranking enhance the usability of online reviews and assist users in quick and unbiased decision making about the product.

The first dataset selected is extracted from Amazon. This dataset has more than 3800 review sentences (i.e. review ids) for phone and its accessories. The second dataset selected is from CRD database which depicts phone records. This dataset has 546 review sentences. In the application, the dataset is first loaded and all the review sentences are selected. The next step extracts all the opinion targets and opinion words list. Then we execute the co-extraction algorithm i.e. Opinion TW Co-extraction algorithm to give the count of each opinion target, count of each opinion word, count of co-located opinion target and word and also calculates the Opinion Association Score (OAS) for each pair.

The results for proposed Opinion TW Co-extraction algorithm are compared against the base system. The graphs shown below depict the results for the base and proposed systems.

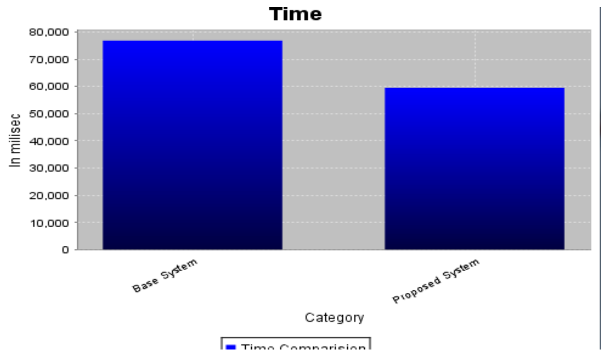

Fig 4: Execution time for Amazon dataset

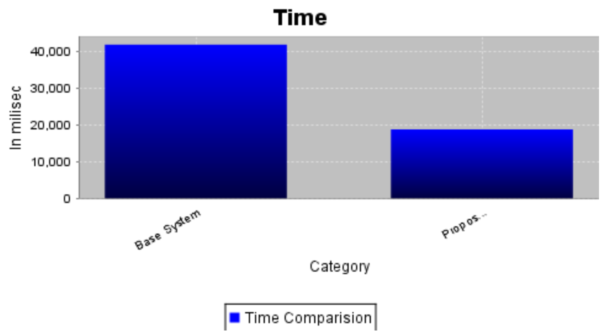

Fig 5: Execution time for CRD dataset

Fig 4 and Fig 5 show the execution time needed to execute the Amazon dataset and CRD dataset respectively. The execution times for base algorithm and proposed Opinion TW Co-extraction Algorithm are compared and the time needed to execute proposed algorithm is much lesser as compared to base algorithm. Thus, the proposed algorithm shows a better time complexity as compared to base algorithm.

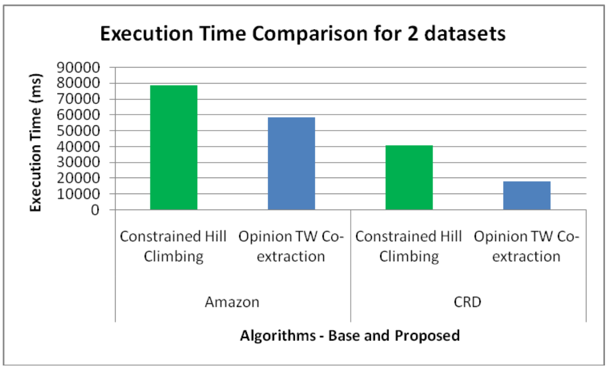

Fig 6: Execution time comparison for two datasets

As can be seen from fig. 6, since the number of records (or number of review sentences) for Amazon dataset is larger as compared to CRD dataset, the execution time for Amazon dataset is also higher than CRD dataset.

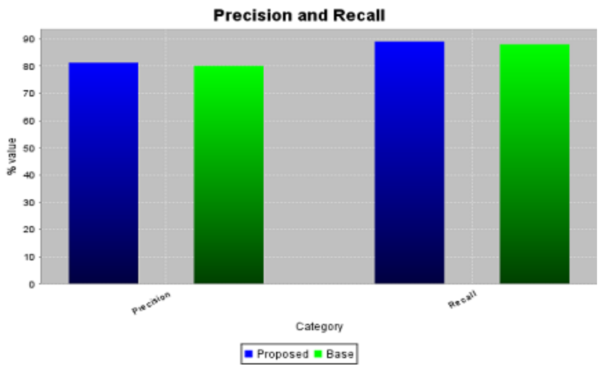

Fig 7: Precision and Recall for Amazon dataset

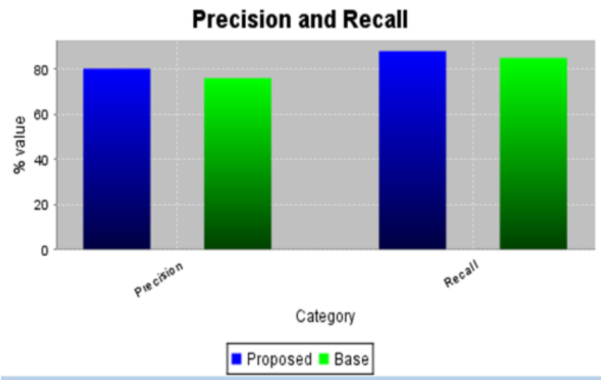

Fig 8: Precision and Recall for CRD dataset

Fig. 7 and fig 8 show the comparison between precision and recall parameters for Amazon and CRD datasets respectively. The values for precision and recall for base system and for proposed system are comparable, with those for proposed system being slightly higher.

6. CONCLUSION AND FUTURE WORK

This paper presents an overview of the proposed system along with the mathematical model and results, which will be used for opinion target and corresponding opinion word coextraction from online reviews. As can be seen from the results, the time taken to execute the proposed algorithm is lesser as compared to that for base algorithm. Thus, proposed algorithm shows better time complexity as compared to base algorithm. Also, precision and recall for base system and proposed system are comparable with each other. It also gives an overview of the Product Aspect Ranking feature that would be built on top of the co-extraction framework.

As future work, this system can be integrated with recommendation system and product recommendations can be made based on the product category selected and the product aspect rankings seen.

7. ACKNOWLEDGEMENT

This work would not have been possible without the support and help of many individuals. I would like to extend my sincere thanks to all of them. I am indebted to Prof. Manisha Patil, Department of Computer Engineering of Smt. Kashibai Navale College of Engineering affiliated to Savitribai Phule Pune University for her guidance and constant supervision as well as all the other staff members for providing important information regarding the work.

8. REFERENCES

[1] http://searchbusinessanalytics.techtarget.com/definition/opinion-mining-sentiment-mining

[2] C. Oak, M. Patil, "Opinion Mining and Product Features Ranking: A Survey", International Journal for Modern Trends in Science and Technology, Volume: 02, Issue No: 10, October 2016

[3] A M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proc. 10th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Seattle, WA, USA, 2004, pp. 168–177.

[4] K. Liu, L. Xu, and J. Zhao, "Opinion target extraction using word based translation model," in Proc. Joint Conf. Empirical Methods Natural Lang. Process. Comput. Natural Lang. Learn., Jeju, Korea, Jul. 2012, pp. 1346–1356.

[5] G. Qiu, L. Bing, J. Bu, and C. Chen, "Opinion word expansion and target extraction through double propagation," Comput. Linguistics, vol. 37, no. 1, pp. 9–27, 2011

[6] M. Hu and B. Liu, "Mining opinion features in customer reviews," in Proc. 19th Nat. Conf. Artif. Intell., San Jose, CA, USA, 2004, pp. 755–760.

[7] F. Li, C. Han, M. Huang, X. Zhu, Y. Xia, S. Zhang, and H. Yu, "Structure-aware review mining and summarization." in Proc. 23th Int. Conf. Comput. Linguistics, Beijing, China, 2010, pp. 653–661.

[8] Y. Wu, Q. Zhang, X. Huang, and L. Wu, "Phrase dependency parsing for opinion mining," in Proc. Conf. Empirical Methods Natural Lang. Process., Singapore, 2009, pp. 1533–1541.

[9] A.-M. Popescu and O. Etzioni, "Extracting product features and opinions from reviews," in Proc. Conf. Human Lang. Technol. Empirical Methods Natural Lang. Process., Vancouver, BC, Canada, 2005, pp. 339–346.

[10] Kang Liu, Liheng Xu, and Jun Zhao, "Co-Extracting Opinion Targets and Opinion Words from Online Reviews Based on the Word Alignment Model", IEEE transactions on Knowledge and Data Engineering, Vol. 27, No.3, March 2015

[11] L. Zhang, B. Liu, S. H. Lim, and E. O'Brien-Strain, "Extracting and ranking product features in opinion documents," in Proc. 23th Int. Conf. Comput. Linguistics, Beijing, China, 2010, pp. 1462–1470.

[12] J. Yu, Z.-J. Zha, M. Wang, and T. S. Chua, "Aspect ranking: Identifying important product aspects from online consumer reviews," in Proc. ACL, Portland, OR, USA, 2011, pp. 1496–1505.

[13] Zheng-Jun Zha, Jianxing Yu, Jinhui Tang, Meng Wang, and Tat-Seng Chua, "Product Aspect Ranking and Its Applications", IEEE transactions on Knowledge and Data Engineering, Vol. 26, No.5, May 2014.

[14] O. Etzioni, M. Cafarella, D. Downey, S. Kok, A. Popescu, T. Shaked, S. Soderland, D. Weld, and A. Yates. 2005. Unsupervised named-entity extraction from the web: An experimental study. Artificial Intelligence, 165(1):91–134.

[15] K. Zhang, R. Narayanan, and A. Choudhary. Voice of the Customers: Mining Online Customer Reviews for Product Feature-based Ranking. WOSN, 2010.

[16] K. Zhang, R. Narayanan, and A. Choudhary, Mining Online Customer Reviews for Ranking Products, Technical Report, EECS department, Northwestern University, 2009.

[17] http://ijiset.com/vol2/v2s6/IJISET_V2_I6_106.pdf

[18] Rajesh Sharadanand Prasad, U. V. Kulkarni, "Implementation and Evaluation of Evolutionary Connectionist Approaches to Automated Text Summarization ", Journal of Computer Science 6 (11), @ 2010 Science Publication, ISSN: 1549-3636, Page - 366-76, Sep-2010.

[19] Rajesh Sharadanand Prasad, Dr. U. V. Kulkarni, "An Automated Approach to Text Summarization Using Fuzzy Logic", International Journal on Computer Engineering & Information Technology, IJCEIT, Volume 23, Issue – 01 ISSN: 0974-2034, Page - 07-21, May -2010.

[20] Rajesh Prasad, Dr. U.V. Kulkarni, "Two Approaches to Automatic Text Summarization: Extractive Methods and Evaluation", International Journal of Computer Engineering and Computer Applications, IJCECA, ISSN: 0974-4983, Ma -2010