VOL. 10, NO. 9, MAY 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

MULTI AGENT BASED ASPECT RANKING REVIEW SUMMARIZATION FOR CUSTOMER GUIDANCE

S. Sobitha Ahila[1] and K. L.Shunmuganathan[2]
[1]Department of Computer Science and Engineering, Easwari Engineering College, India
[2]Department of Computer Science and Engineering, R. M. K Engineering College, India
E-Mail: sobitha.ooviya@gmail.com

ABSTRACT

Generally, any product has various features (also known as aspects). For example, a mobile phone can have more than three hundred aspects, such as battery, design, multi-touch, and 3G network. It can be argued that some aspects are more important when compared to others, and have greater impact on the eventual consumers' decision making as well as firms' product development strategies. Identifying these important product aspects from the consumer reviews or feedbacks will increase the usability of reviews and is beneficial to both consumers and firms. It becomes easier for consumers to make wise purchasing decision by paying attention to the important aspects and can firm also concentrate on improving the quality of these aspects and this will enhance the product reputation effectively. However, it is tedious and time consuming for people to manually identify the important aspects of products from numerous reviews. Thus, this project proposes agent based product aspect ranking framework to identify the important aspects of products and rate them according to the consumers' opinion about them. A false detection technique is also employed to indicate those features for which an overall positive or negative opinion cannot be determined. The potential of aspect ranking can be shown in many real-world applications like e-commerce, retail, entertainment etc.

Keywords*:* agents, aspect ranking framework, review summarization, false detection.

1. INTRODUCTION

Before a customer can make any purchasing decisions, the first step that he/she takes is to check out the various reviews available online for that product. With the rapid expansion of e-commerce, millions of products are offered online, and these retail online websites encourage customers to write reviews to express their opinion on the various aspects of the products. Here the word aspect, which is synonymous with the word feature, refers to a component or attribute of a certain product. For example, a sample review "The user interface of iPhone 6 is awesome", expresses a positive opinion on the aspect "user interface".

The reviews contain rich and valuable information which can serve as an important resource not only for the customers but also for the manufacturing firms. The customers can rely on the reviews when they want to choose a product which will exactly satisfy their requirements based on the priorities they give for the features of the products and the firms use the information extracted from the reviews to help with their product development, marketing and consumer relationship management.

But there is a lot of difficulty in extracting the exact required information from the reviews. The reviews are numerous, diverse and not precise leading to difficulties in information navigation. And hence knowledge acquisition is tough. Also we can argue the some aspects of the product are more important than the others, as they have a greater impact on the eventual customer's decision making. Agent-based review summarization using aspect ranking algorithm has advantages of aspect ranking and Agent: it can extract feature efficiently and intelligently, so it is becoming more and more important in modern applications. Motivated by these reasons, we propose a product aspect ranking framework, which automatically identifies the important aspects of products from online consumer reviews and improves the usability of numerous reviews by representing the aspect ranking graphically.

Agent has a powerful ability to solve problems with the characters of autonomy, intelligence and interaction, it can help human to finish simple or complicated tasks. For example, the comparison shopping system can help users searching information and price of their desired goods from numerous malls automatically by using intelligent agent technology, then users would compare the searching results to optimize the purchase decision. User Information Retrieval Agent can help users to complete complex information retrieval tasks, and even users can quickly locate interested topics, to achieve the purpose of personalized retrieval and centralized view. Agent technology can search data purposely and experienced in the mining process, also it can track information of users' interest, notify the update of related information to users promptly. Agent encapsulates the search engine what enable users not have to pay attention to work process, thus users can interact with the system more conveniently.

2. RELATED WORKS

Zheng-Jun Zha *et al.* [1] proposed a product aspect ranking framework which first identifies product aspects by a shallow dependency parser and determine consumer opinions on these aspects via a sentiment classifier. They then developed a probabilistic aspect ranking algorithm to infer the importance of aspects by simultaneously considering aspect frequency and the

ARPN Journal of Engineering and Applied Sciences

influence of consumer opinions given to each aspect over their overall opinions

Jin and Ho [14] learned a lexicalized HMM (Hidden Markov Model) to extract aspects and opinion expressions. Different from previous approaches that have mostly relied on natural language processing techniques or statistic information, they proposed a novel machine learning framework using lexicalized HMMs. The approach naturally integrates linguistic features, such as part-of -speech and surrounding contextual clues of words into automatic learning. But this and such other related methods require sufficient labeled samples for training. And it is time-consuming and labor-intensive to label samples.

On the other hand, unsupervised methods have also emerged. The most notable unsupervised approach was proposed by Hu and Liu [7]. They assumed that product aspects are nouns and noun phrases. The approach first extracts nouns and noun phrases as candidate aspects. The occurrence frequencies of the nouns and noun phrases are counted, and only the frequent ones are kept as aspects.

Wu *et al.* [13] utilized a phrase dependency parser to extract noun phrases from reviews as aspect candidates. They then employed a language model to filter out those unlikely aspects. After identifying aspects in reviews, the next task is aspect sentiment classification, which determines the orientation of sentiment expressed on each aspect. Two major approaches for aspect sentiment classification include lexicon-based and supervised learning approaches. The lexicon-based methods are typically unsupervised. They rely on a sentiment lexicon containing a list of positive and negative sentiment words. To generate a high-quality lexicon, the bootstrapping strategy is usually employed.

Wang *et al.* [15] developed a latent aspect rating analysis model, which aims to infer reviewer's latent opinions on each aspect and the relative emphasis on different aspects. This work concentrates on aspect-level opinion estimation and reviewer rating behavior analysis, rather than on aspect ranking.

Liu Lizhen *et al.* [4] proposed a feature-based vector model and a novel weighting algorithm for sentiment analysis of Chinese product reviews. Specifically, an opinionated document is modeled by a set of feature-based vectors and corresponding weights. Their model considers modifying relationships between words and contains rich sentiment strength descriptions which are represented by adverbs of degree and punctuations. Dependency parsing is applied to construct the feature vectors. A novel feature weighting algorithm is proposed for supervised sentiment classification based on rich sentiment strength related information.

Xu Xueke *et al.* [6] proposed a paper which focuses on how to improve aspect-level opinion mining for online customer reviews. They first proposed a novel generative topic model, the Joint Aspect/Sentiment (JAS) model, to jointly extract aspects and aspect-dependent sentiment lexicons from online customer reviews. An aspect-dependent sentiment lexicon refers to the aspect-specific opinion words along with their aspect-aware sentiment polarities with respect to a specific aspect. They then apply the extracted aspect dependent sentiment lexicons to a series of aspect-level opinion mining tasks, including implicit aspect identification, aspect-based extractive opinion summarization, and aspect-level sentiment classification.

Victor C. Cheng *et al.* [3] developed a generative probabilistic aspect mining model (PAMM) for identifying the aspects/topics relating to class labels or categorical meta-information of a corpus. Unlike many other unsupervised approaches or supervised approaches, PAMM has a unique feature in that it focuses on finding aspects relating to one class only rather than finding aspects for all classes simultaneously in each execution. This reduces the chance of having aspects formed from mixing concepts of different classes; hence the identified aspects are easier to be interpreted by people. The aspects found also have the property that they are class distinguishing: They can be used to distinguish a class from other classes. An efficient EM-algorithm is developed for parameter estimation. Experimental results on reviews of four different drugs show that PAMM is able to find better aspects than other common approaches, when measured with mean point wise mutual information and classification accuracy. In addition, the derived aspects were also assessed by humans based on different specified perspectives, and PAMM was found to be rated highest.

Bo Pang *et al.* [8] considered the problem of classifying documents not by topic, but by overall sentiment,e.g., determining whether a review is positive or negative. Using movie reviews as data, we find that standard machine learning techniques definitively outperform human-produced baselines. However, the three machine learning methods we employed (Naive Bayes, maximum entropy classification, and support vector machines) do not perform as well on sentiment classification as on traditional topic-based categorization. They concluded by examining factors that make the sentiment classification problem more challenging.

Bo Pang *et al.* [9] proposed a novel machine-learning method to determine sentiment polarity(Sentiment analysis seeks to identify the viewpoints underlying a text span; an example application is classifying a movie review as thumbs up or thumbs down)that applies text-categorization techniques to just the subjective portions of the document. Extracting these portions can be implemented using efficient techniques for finding minimum cuts in graphs; this greatly facilitates incorporation of cross-sentence contextual constraints.

Longbing Cao *et al.* [15] discussed high-level overview of the agent-mining interaction from the perspective of an emerging area in the scientific family. To promote it as a newly emergent scientific field, multi agents are considered as key driving force to do review summarization.

Young joong Ko *et al*. [10] proposed a paper that focuses on how we can improve text classification by effectively applying class information to a term weighting scheme. We believe that text classification should utilize class information better than information retrieval because supervised learning based text classification has labeled training data. However, inverse document frequency, idf, is a measure of general importance of a term; there is no use of class information. This paper proposes new term weighting schemes for multiclass text classification, which include term weighting methods for test documents.

3. PROPOSED SYSTEM

The aspect ranking framework we have proposed works as follows. The front end consist of customer review collection by agents from 3 user websites, where the users can register, login , buy products and then give their feedback for the products. In each of the 3 websites the feedbacks or reviews are collected in different formats. In the 1st website the feedback can be given as a free text review. In the 2nd website, the users can give their ratings to the various features of each product which is already prelisted. In the 3rd website, the users can list the features either in the pros or cons text box and finally give their overall comment for the product. The comments and feedbacks given get stored in the database.

The pipeline of the aspect ranking framework consists of three main components:

1) Aspect identification
2) Sentiment classification
3) Ranking of aspects

Given any consumer review, we first identify the aspect mentioned in the review and then determine whether the consumer's opinion on that aspect is positive or negative. Then using an aspect ranking algorithm with the help of a Naïve Bayes classifier, we classify all the aspects mentioned in all three formats of the reviews as either pros or cons.

The front end also consists of an admin website, which the users can visit if they wish to know about the summarized feedback collected from all 3 websites. For any particular brand and model of a product selected by the user, the pros and cons will be listed and it can also be viewed in a graphical format. Another additional enhancement in the admin website is the false detection technique - aspects which appear for an equal or almost equal number in both pros and cons, will be listed in the undefined category.

The system design can be broadly divided into the following modules:

- Front end design and database creation
- Aspect identification and sentiment classification
- Aspect ranking with false detection

a) Front end design and database creation

The proclamation of the 1st module begins with initialization. It paves way for the user who pledges his views for the respected products to enroll him as a primary step in order to pacify the person as an active reviewer. The information of the user has been enhanced into the database. Separate storage on constructing the areas for the user and the place for stocking the reviews is congregated in the database. Once when the classifications of data are to be reviewed by the admin system, it can be partitioned according to the product and the final review can be out looked on the furnished model.

The 1st module can be segregated into 4 websites. Whereby 3 user websites are offered to the user or reviewer where they can advent their views towards the products. The user websites can be listed as free text, rating, pros and cons websites. The website 1, where the reviews can be given as free text initiates the users to create an account as a primary step after which it offers to share their views towards the products. The website 1 is more profoundly welcomed by the users as it offers them to share their views and ideas in their own context. These reviews will be automatically collected by agents along with the user details once it is submitted.

The next website offered is one where the users can give their ratings for the aspects of the products. Here the user is offered a prefixed list of aspects which can be rated out of 10. This website is advertised and accepted by the users due to time conservation and since awarding the ratings is simple and easy. This information also collected automatically by agents. The final user website is where the users can give their reviews in the form of pros and cons wherein the merits and demerits of the aspects of the products can be submitted directly and collected by agents. In addition, an optional comment text box is also offered where the user can give their overall opinion of the product as well.

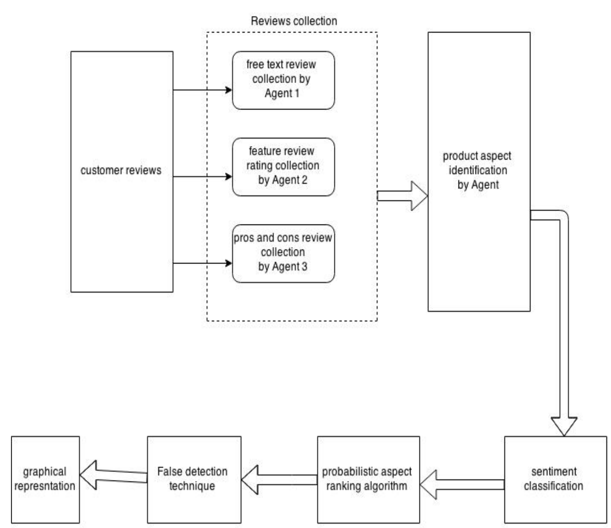

Figure-1. Flow chart for proposed aspect ranking framework.

The admin website also uses agents to oversee the overall process and can access the backend storage database where the reviews prompted by the user are stored.

To demonstrate the working of this framework we have initially taken 4 products under consideration- mobiles, laptops, cars and bikes. In the 1st website mobiles, cars and bikes are stagnated in order for the user to trade his reviews. In the 2nd website laptops, cars and bikes are implemented as products and the respective reviews can be prompted. The 3rd website contains all the 4 products i.e., laptops, mobiles, cars and bikes.

Thus the overall reviews with virtue of the user's comments are stored in the database which can be either viewed or plotted as graphs in the admin website in order to analyze and summarize the overall opinion.

b) Aspect identification and sentiment classification

The module 2 intrigues about the aspect identification with response to which the sentiment classification is also inferred. In the previous case (module 1), the user websites were discussed wherein the user forwards his reviews on the itinerary products. The reviews thus categorized are summarized in the aspect classes. For better understanding, if an opinion given by the users such as "Airbags are awesome. Headlamps are poor" for AUDI Q3 car, the "Airbags" is classified in the aspect "Safety" and "Headlamps" are categorized under the aspect "Lighting". Particulars as such are many and the product can be subdivided with various aspects. Another example that can be pictured is that "Battery life is poor. Picture clarity is excellent." Therefore, "Battery life" falls under the aspect "Battery" and the "Picture clarity can be classified as "Camera". Thus for every quality of the product that is recognized as pros and cons, the aspect with which that quality is categorized under is also identified with the help of agents. This entire process is termed as aspect identification.

Sentiment classification is simply to visualize the merits and demerits of a product summarized by the user as pros and cons. This can be extracted from the feedbacks given by the reviewer where the merit-components specified by the user is kept under pros and demerit parts are placed under cons. Thus when the overall feedback is complete, which is given in the form of a free text, the website acclaims the merits and demerits separately by searching the words in sentiment database. On calling the function with respect to the words entered, the website receives the final authentication whether it comes under the pros or cons ultimately. Thus the positively said aspects are aligned on one side and the opposed aspects are on the other side. With this, the aspects under each product that is pointed by the user is conceded within the sentiment. This silhouette is verified as sentiment classification.

c) Aspect ranking with false detection

The module 3 denunciates the overall depictions on the clustered feedbacks. It summarizes the number of users who have reviewed the various aspects in which the pros and cons towards a product are taken as the primary consideration. From the overall models showcased, the user selects the product that he/she wishes to view in order to know the summarized opinions of the users from all the three websites which is displayed as a list of aspects divided as pros and cons.

Module 3 segregates the review of aspects as pros and cons which is partitioned in relevance to the words that are antagonized by the admin.

Finally on a further note, the website provides a clear view of the number of users who have imprinted good comments towards an individual aspect and the number of users who have passed negative feedbacks on the various aspects of the products. For this purpose we make use of a probabilistic aspect ranking algorithm which is implemented with the help of a Naïve Bayes classifier, which classifies the aspects as pros or cons based on the consumer's opinion for that particular aspect. Thus the aspect rating is done based only on the number of users who have rated the aspects.

The overall opinions received from the users can be visualized in the form of graphs. This can be done by picking up all the pros and cons reviews into consideration and selecting the "Make graph" option.

The false detection is an important trademark of our proposed system wherein the merits and demerits are compared. If any aspect comes under both pros and cons, then based on the majority of reviews under each category, the final classification is done in false detection. Also if any aspect on which an equal or almost equal number of users have given both a positive and negative opinion, then such aspects come under the undefined category i.e., they can neither be finalized as a merit or demerit for the product, and the final decision is left for the customer to make.

Thus the module first accepts the review of every user who initially creates a profile for themselves. After the reviews are accepted, the aspects are segregated into pros and cons depending on the reviews given whereby the type of review is detected as an advantage or disadvantage by checking the word given as input by the user with the database words which depicts the overall feedback as good or bad. The users can select either one of the 3 user websites according to their convenience or could also select all the 3 in order to share their reviews in a detailed proportion. Therefore, the overall reviews can be summarized with relevance to the respected aspect and the false detected on the aspects can also be stressed in the form of graphs. The graph images portray the overall feedbacks which can be visualized in a simple way by the admin.

4. RESULTS AND DISCUSSION

The various products with incorporated aspects and performance can be rated by the user on receiving the products on the merit basis. The main advantage that a user retrieves on creating a profile for him is to store his details in the database plot and later with that initial propagation; he receives the license to freely rate all the products contented within the user website. In website1,

VOL. 10, NO. 9, MAY 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

open source of feedback is given and in website2 products are just rated and website 3 serves as an induction that summarizes the overall individual aspects as pros or cons. In advancement to the pros and cons, the false detection gives an exact detailed view of analysis of users' feedback which creates a pole on individual component based function.

Thus this paper re-evaluates all the already available feedback methods and recreates the techniques with possible better innovative implementation that will easily bridge the gap between the user and the manufacturing department. The Research and Development team can rely on these methods directly to reach the people on complimenting their good products and offering suggestions for better improvement of their products.

Figure-2. Selecting product brand and model in admin website.

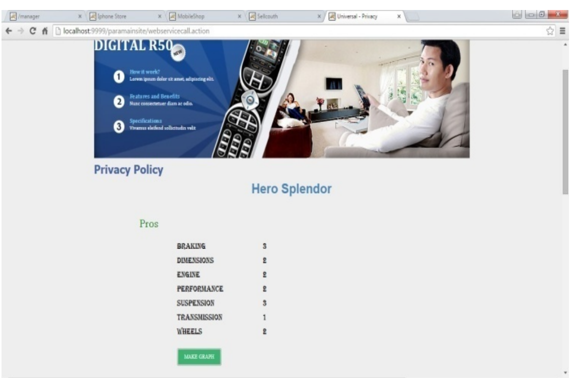

Figure-3. Aspect rating-pros.

Figure-4. Aspect rating-pros graph.

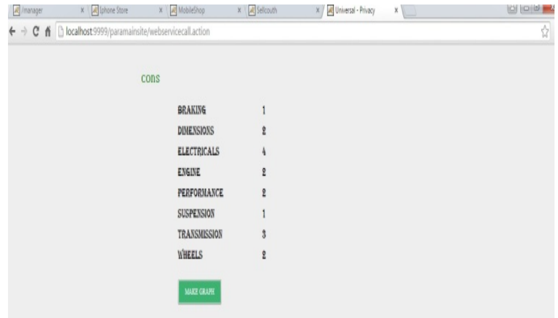

Figure-5. Aspect rating-cons.

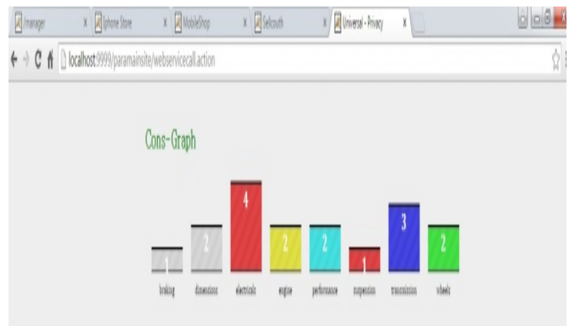

Figure-6. Aspect rating-cons graph.

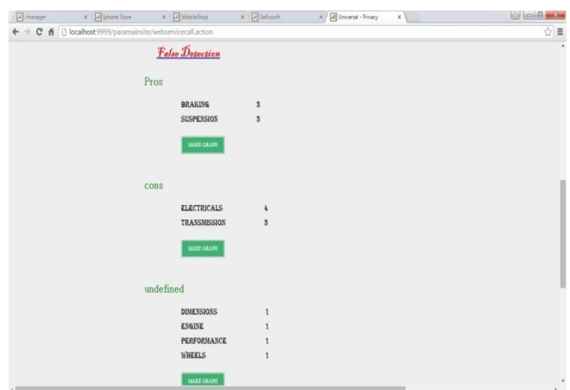

Figure-7. False detection.

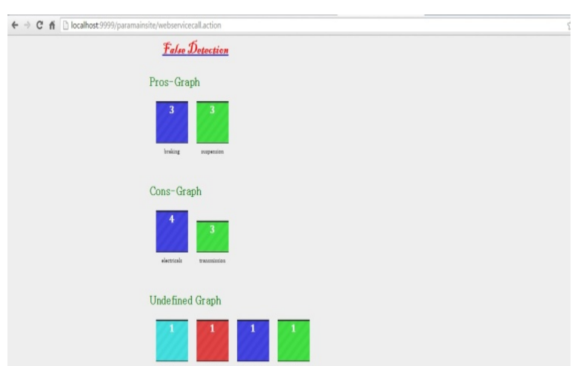

Figure-8. False detection graph.

The final carry by the people to the admin is easy whereby the admin and the firm is offered various comments on all aspects with a lot of people being attracted to simple methods of creating an own profile and partitioning the methods of feedback.

With all this done, the main aim of our innovations is thus achieved(i.e.)the motive of promoting effective feedback by a customer who enables the interest on forwarding his comments for betterment of individual product and complimenting the products whose quality are crowned best in the market. By this the welfare of the firm is further increased and these feedbacks serves as a direct conversing comments in a friendly way that could be implemented in further production for a better future.

5. CONCLUSIONS AND FUTURE WORK

We recreated better exposure of a feedback system, with an environment to attract a sheer number of users who involve themselves in rating the products with an intention of receiving back better products with the changes being implemented on the same. The combination of Aspect ranking technique and multi-Agent technique can help to reduce the amount of data transmission, lighten the network load, improve the mining performance and ensure data security. With response from the users, the admin team creates graphs to denote the uplift or downstream flow of quality in the manufacturing sector. This therefore rekindles the method of production or improvising the quality of performance in order to receive a warm welcome of the overall accomplishment in the changes done aftermath from the comments pictured. Thus the enhancement has been made providing a wide range of user websites and a user profile for creating a database for himself and offering him to share any number of feedbacks until the desired results are observed. Therefore the feedbacks given by the users are initially taken into account by the admin team. This paper thus initiates better techniques by using agents in offering feedbacks and the various promotions that can be complimented. These can be viewed either in text forms or as simple graph formation. In future the framework can be made more efficient in collecting reviews from online websites dynamically and integrate them and present a collective ranking and also as a recommender system to a specific user based on his selective criteria and priority.

REFERENCES

[1] Zheng-Jun Zha, Jianxing Yu, Jinhui Tang, Meng Wang and Tat-Seng Chua. 2014. “Product Aspect Ranking and Its Applications”, IEEE Transactions on Knowledge and Data Engineering,Vol. 26, No. 5, pp. 1211-1224.

[2] V. Gupta and G. S. Lehal. 2010. “A survey of text summarization extractive techniques,” J. Emerg. Technol. Web Intell., Vol. 2, No. 3, pp. 258–268.

[3] Cheng V. C., Leung, C. H. C., Jiming Liu and Milani A. 2013. “Probabilistic Aspect Mining Model for Drug Reviews”, IEEE Transactions on Knowledge Engineering,Vol. 26, No. 8, pp. 2002 – 2013.

[4] LIU Lizhen, Song Wei, Wang Hanshi, LI Chuchu and LU Jingli. 2014. “A novel feature-based method for sentiment analysis of Chinese product reviews”, IEEE Transactions on Knowledge Engineering,Vol. 11, No. 3, March.

[5] H. Wang, Y. Lu and C. X. Zhai. 2010. “Latent aspect rating analysis on review text data: A rating regression approach,” in Proc. 16th ACM SIGKDD, San Diego, CA, USA, pp. 168–176.

[6] Xu Xueke,Cheng Xueqi,Tan Songbo, Liu Yue and Shen Huawei. 2013. “Aspect-level opinion mining of online customer reviews” IEEE Transactions on Knowledge Engineering, Vol. 10, No. 3, March.

[7] M. Hu and B. Liu. 2004. “Mining and summarizing customer reviews,”in Proc. SIGKDD, Seattle, WA, USA, pp. 168–177.

[8] Bo Pang and Lillian Lee and Shivakumar Vaithyanathan. “Thumbs up? Sentiment Classification using Machine Learning Techniques”, EMNLP '02 Proceedings of the ACL-02 conference on Empirical methods in natural language processing – Vol. 10 pp. 79-86.

[9] Bo Pang and Lillian Lee. “A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts”, ACL '04 Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics.

[10] Youngjoong Ko. “A Study of Term Weighting Schemes Using Class Information for Text Classification”, SIGIR '12 Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval pp. 1029-1030.

[11] B. Liu. 2012. “Sentiment Analysis and Opinion Mining” Mogarn & Claypool Publishers, San Rafael, CA, USA.

[12] B. Ohana and B. Tierney. 2009. “Sentiment classification of reviews using SentiWordNet,” in Proc. IT&T Conf., Dublin, Ireland.

[13] Y. Wu, Q. Zhang, X. Huang and L. Wu. 2009. “Phrase dependency parsing for opinion mining,” in Proc. ACL, Singapore, pp. 1533–1541.

[14] W. Jin and H. H. Ho. 2009. “A novel lexicalized HMM-based learning framework for web opinion

mining," in Proc. 26th Annu. ICML,Montreal, QC, Canada, pp. 465–472.

[15] Longbing Cao, Chao Luo and Chengqi Zhang. 2007. "Agent-Mining Interaction: An Emerging Area" Autonomous Intelligent Systems: Multi-Agents and Data Mining Lecture Notes in Computer Science Vol. 4476, pp 60-73.

[16] H. Wang, Y. Lu and C. X. Zhai. 2010. "Latent aspect rating analysis on review text data: A rating regression approach," in Proc. 16th ACM SIGKDD, San Diego, CA, USA, pp. 168–176.