

## International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

*IJCSMC, Vol. 3, Issue. 6, June 2014, pg.465 – 470*

### **RESEARCH ARTICLE**

# A Statistical Approach for User Opinion Analysis on Software Review Analysis

**Anju**

Student, M.Tech, Deptt. of Computer Sc. & App., College, Shri Baba Mastnath Engineering College, Rohtak, Haryana  
[anjuman.anju98@gmail.com](mailto:anjuman.anju98@gmail.com)

**Sunita**

Asstt. Professor, Deptt. of Computer Sc. & App., College, Shri Baba Mastnath Engineering College, Rohtak, Haryana  
[Kashisuni5@gmail.com](mailto:Kashisuni5@gmail.com)

**Abstract :** Online Software reviews are collected by different agencies to predict or analyze the user response towards the Software. It is considered as the business plan to estimate the Software response and earning. But the review is collected in the form of textual feedback that requires some intelligent automated check called Opinion analysis. In this present work, a weighted statistical approach is suggested to perform the Software Opinion analysis. The work includes the analysis on overall Software review as well as feature level review analysis. The obtained results shows the clear identification of positive and negative Opinions at feature level so that the effective decision based on Opinion analysis can be taken

**Keywords :** Software Opinion Analysis, Weighted Approach, Feature Based

## I. INTRODUCTION

One of the effective, growing and leading terminology associated with text mining is opinion mining. Today most of the products and services are available online. To interact with customers, web is the direct and most effective way considered by different vendors. Customers can get the overview of available online products and services through the vendor's website. These sites provide the complete description of the available products along with the product features, limitations and advantages. Customer can buy these products or services online. Once, the products or services are taken by the customers, another responsibilities of the vendors is to take the feedback from customer about the product usage. The product usage consideration is required to identify the mood of customer for particular product or services. This feedback is the part of business plan to improve the reputation of the firm and to provide the most effective services to the customers. This kind of feedback is taken from the customer by asking for a set of relative queries or in can be taken in the form of free text. These options are taken through some social site, blog or some outsourcing agencies. But these kind of textual feedback is required to analyzed to conclude the customer mood or the response towards the product or the services. To analyze this mood, the Opinion analysis approach is used by the business firm. In the normal form,

Opinions are divided in two main categories called positive and negative Opinions. Positive Opinions are described by positive adjectives and negative Opinions are identified by negative adjectives as shown in figure 1.

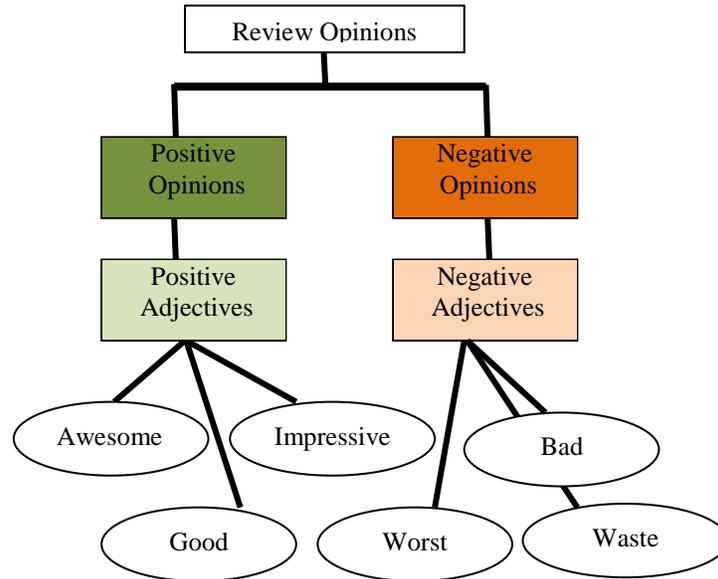


Figure 1 : Review Opinions Analysis Vectors

These adjectives are able to express the reviewer mood or the Opinions. All the adjectives used by the reviewers are also different in terms of adjective strength. The strength is here defined in terms of impact analysis of the adjective. Such as “awesome”, “impressive”, “worst” are high intensity words and are having the heavy score. Whereas the words like “average”, “ok” are the low scoring words. Another aspect of the review analysis is about the feature level analysis. It means, when some product or service is analyzed, it is not necessary it will be analyzed as the whole. In such case, the feature level analysis is required. The product features are analyzed in reviews respective to these adjectives and identified the positive and negative features associated in the product.

In this present work, the review Opinion analysis is defined for Softwares. When a Software is released, there are number of agencies that capture the user review for the particular Software under different aspects and conclude the overall Software rating or the feature based rating. To automate this rating process, the opinion mining is required. Opinion mining will analyze these Software features and provide the response in terms of aspect analysis.

In this paper, an effective weighted statistical approach is suggested to perform the Software Opinion analysis. The work includes the analysis over multiple reviews and to extract the Opinion features and the relative Opinions so that the conclusion about the complete Software review can be taken. The presented work is divided in two main phases. In first phase, individual review analysis is performed under feature specification. Once the feature level analysis will be completed, the next work is to perform the score analysis to obtain the complete review response.

In this section, an overview to the Opinion analysis and opinion mining is defined. The section also includes the description of analysis dependency on adjectives. The adjective analysis itself is capable to describe the Software Opinions. In section II, some of the work defined by earlier researchers is discussed. In section III, the proposed model is given and in section IV, the results obtained from the work are discussed. In section V, the conclusion obtained from the work is presented.

## II. EXISTING WORK

Lot of work is already presented by different researchers in the area of Opinion analysis. Some of the work defined by earlier researchers is discussed in this section. Author defined a work on polarity estimation on Opinion words and phrase analysis. Author presented a machine learning algorithm to analyze the lexicons associated with Software Opinions. Author presented the polarity analysis and Opinion analysis to derive the effective results[1]. Another work on adjective level analysis was proposed by the author[6]. Author has defined a pattern analysis approach over the subjective expression to identify the review Opinion. Author presented an automated approach to identify the binary emotion over the review. Author[12] presented a work on Opinion analysis and polarity assignment approach at document level. Author defined the document or text level analysis to obtain the Software review analysis and to provide the effective derivation from the Opinion orientation so that the effective propagation of Software Opinions is taken. Author[2] presented a machine learning algorithm based approach for Opinion classification. Author defined the entropy analysis based SVM approach to provide effective division of Software Opinions. Author[13] also presented SVM based approach to analyze the terms associated in a topic. Author provided the topic categorization and subjective portion analysis approach to reduce the user efforts so that effective Opinion classification will be done. Author defined the polarity based classification approach.

Another work on Opinion analysis from online reviews and submitted news articles was presented by[11]. Author presented a work on positive and negative Opinion analysis approach for sentence level analysis. Author designed a syntactic parser and Opinion lexicons based approach for polarity classification. Another work on content level analysis is defined by the author[10]. Author presented the work on word dependency analysis so that the document prediction will be done effectively. Author presented the sentence assignment approach along with weight analysis. Author presented the individual sentence analysis approach for aggregating the individual sentences and Opinions so that the effective lexicon derivation will be obtained. Another work on different kind of clause analysis and Opinion analysis was presented by the author[9]. Author categorize the features and Opinions so that the effective Opinion polarity will be derived. Author presented the tag based analysis under the score analysis. The sentence level analysis has improved the working and provided the effective results over the system.

## III. RESEARCH METHODOLOGY

In this present work, a textual analysis on Software review is presented to analyze the user response and interest. To attain the user interest or to analyze the customer satisfaction for a particular product or the service, the most common approach adapted by most of the companies is the review system. In this review system, the users or the customer can register their view about the liking or disliking of that product or service. In the film industry, the review system is also implemented to obtain the user interest or the view for a Software just after the Software release. These reviews are collected using the social sites or different review companies. Based on these reviews, the Software quality is analyzed and it gives the recommendation to other users to watch that Software or not. If the review is positive the business for that Software will be increased but if the review is negative, the business for that Software will be decreased. Because of this there is a requirement to analyze these reviews effectively. Opinion Analysis is about to analyze these reviews and identify the Opinion associated with that Software. The reviews collected for a Software are very large in number. Because of this, there is requirement of an effective approach to provide the Opinion analysis accurately and fast. The presented work is in same direction. In this work, a word analysis based weighted statistical method is presented to perform Opinion analysis on Software reviews.

In this presented work, an effective weighted statistical approach is defined to perform the overall Software analysis as well as feature level analysis. The complete work is divided in number of stages. The first stage was to perform the adjective level analysis. This analysis includes the identification of adjectives over the Software review. These adjectives can be positive or negative reviews. The analysis process is shown in figure 2.

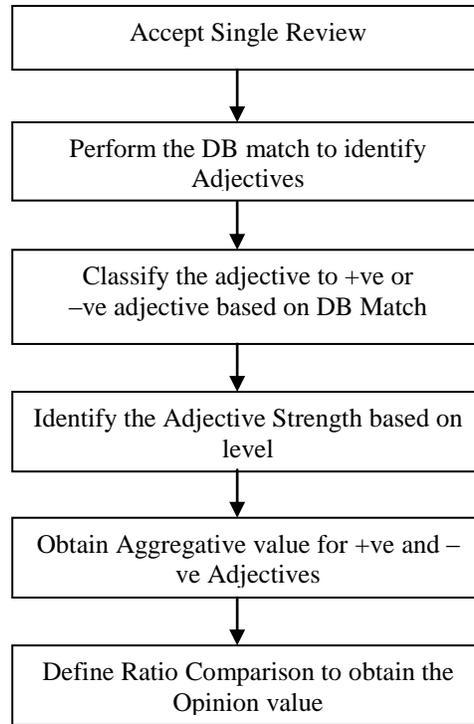


Figure 2 : Adjective Level Analysis

The work depends on the adjective database. This database contains the information about the available adjective, type and score. The type of adjective is identified as the positive and negative adjective. The adjective scores are defined to represent the strength of the adjective. The process includes the word separation of the review and to perform the filtration. After the filtration process, the main keywords from the review are retrieved on which the basic processing can be defined. Once the words are identified, the next work is to perform this word match with adjective database. As the adjective is identified, the next work is to identify the adjective type and score. Based on this analysis, the review is identified as the positive or negative review. The work not only includes the overall review analysis but also includes the feature level analysis shown in figure 3. This work is divided in two stages. In first stage the feature extraction is performed.

After obtaining the feature score of individual review, the next work is to assign the weightage to adjectives to identify the adjective strength and the polarity of the review as well polarity of feature review. At first, the weighted polarity analysis is performed on individual review of a Software but later on the same is applied on all the reviews. Based on this analysis, overall Opinion of Software review is identified. The presented work is implemented on realtime dataset available from the web repository. The result analysis obtained from the work is given in section IV.

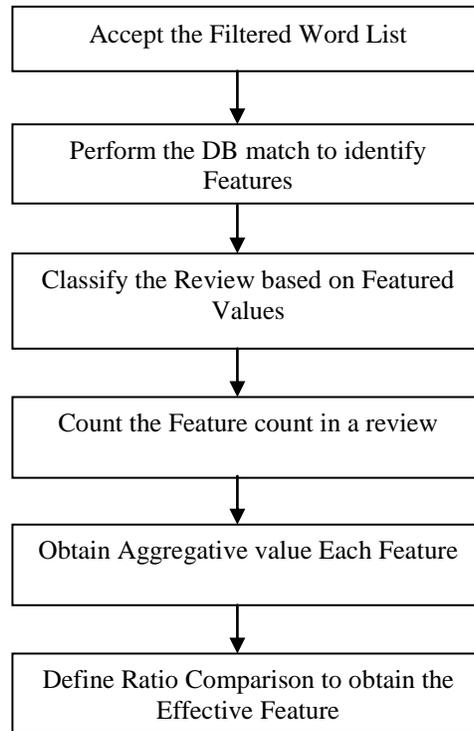


Figure 3 : Feature Level Analysis

#### IV. RESULTS

The presented work is about to identify the Software Opinion based on multiple review analysis. The work is applied on real time datasets obtained from the secondary source. The dataset collected here are the Software review dataset and adjective dataset. The analysis is here defined in terms of Software review analysis so that the review response is collected and Software Opinions are identified. The analysis performed on different Softwares under different aspects. The overall analysis on Software is shown in table 1.

Table 1: Review Analysis

Software Id	Positive Score	Negative Score	Score
1	61.75	53.25	6.50
2	42.25	49.25	-7
3	73.0	55.25	17.75
4	65.25	45.25	20
5	35.25	50.50	-15.25

Here the Software review of 5 Softwares is taken under the defined approach. The positive and negative scores are collected. The overall score is obtained from the Software. The results shows that Software 2 and Software 5 get the negative response from user whereas the Software 1,3 and 4 got the positive response. The favorite Software among users is Software 4 that gets the maximum response from users.

## V. CONCLUSION

In this paper, a statistical weighted approach is defined to perform the Software Opinion analysis. The analysis is here defined for overall Software Opinions as well as feature level Opinion analysis. The paper defined the algorithm approach adopted and the results shows the clear derivation of Software score based on review Opinion analysis.

## REFERENCES

- [1] Tun Thura Thet, Jin-Cheon Na and Christopher S.G. Khoo Nanyang Technological University, Singapore Aspect-based Opinion analysis of Software reviews on discussion boards
- [2] Bing Liu. Opinion Analysis and Opinion Mining, Morgan & Claypool Publishers, May 2012.
- [3] Opinion Analysis of Software review Comments Kvat Yessenov kuat , Sa'sa Misailovi'c May 17, 2009.
- [4] R. Quirk, S. Greenbaum, G. Leech and J. Svartvik, A comprehensive grammar of the English language (Longman,London,1985).
- [5] P. Chaovalit and L. Zhou, Software review mining: a comparison between supervised and unsupervised classification approaches, Proceedings of the 38th Annual Hawaii International Conference on System Sciences (2005)
- [6] P.D. Turney, Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews, Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics ACL (2002)
- [7] T. Wilson, J. Wiebe and P. Hoffmann, Recognizing contextual polarity in phrase-level Opinion analysis, Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing (2005)
- [8] P.J. Stone, D.C. Dunphy, M.S. Smith and D.M. Ogilvie, General Inquirer: a Computer Approach to Content Analysis (The MIT Press, Cambridge, MA, 1966).
- [9] J. Wiebe, Learning subjective adjectives from corpora, Proceedings of the 17th National Conference on Artificial Intelligence (2000)
- [10] E. Riloff, J. Wiebe and T. Wilson, Learning subjective nouns using extraction pattern bootstrapping, Proceeding of the 7th Conference on Natural Language Learning (2003)
- [11] E. Riloff and J. Wiebe, Learning extraction patterns for subjective expressions, Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (2003)
- [12] G. Qiu, B. Liu, J. Bu and C. Chen, Expanding domain Opinion lexicon through double propagation, Proceedings of the 21st International Joint Conference on Artificial Intelligence (Morgan Kaufmann, San Francisco, 2009)
- [13] T. Mullen and N. Collier, Opinion analysis using support vector machines with diverse information sources, Proceedings of the 9th Conference on Empirical Methods in Natural Language Processing (2004)