RESEARCH ARTICLE

A Study on Different Approaches on Software Review Mining

Anju
Student, M.Tech
Deptt. Of Computer Sc., Baba Mast Nath Engg. College
Rohtak, Haryana
anjuman.anju98@gmail.com

Sunita
Asstt. Professor,
Deptt. Of Computer Sc. & App., College,
Shri Baba Mastnath Engineering College,
Rohtak, Haryana
Kashisuni5@gmail.com

Abstract*:* Questionnaire based analysis is one of the traditional approach for software quality evaluation under different measurement vectors. But generally a questionnaire involves the technical terms and boundation so that the end user is not interested to answer the answer sequence. An improved form to obtain the user interest analysis in the software is the software review analysis. In this form, the user interest is accepted in the form of text review and an intelligent analysis over the review is performed to identify the software product context analysis and the review comment analysis. Based on this analysis, the overall user adaptive interest to the software product is identified. In this paper, a study on different approaches of software product review is defined.

Keywords: Software Review, Questionnaire, Text Mining, Context Analysis

I.   INTRODUCTION

Software review analysis is one the most effective technique to identify the software user interest in the software as well as to accept the user feedback. When a software product is being launched, it goes under different testing stages to identify the software product reliability. These testing techniques include the module level, integration level and system level testing. After all these testing approaches, when the software product is ready for sale, it is checked on the customer platform for the final test analysis. This kind of testing is called alpha testing or beta testing. Alpha testing is generally user specific in which the software product is installed on the customer environment and obtained the personal view of the customer[1][2]. A Registered software product is generally tested under the acceptance testing. Another kind of

testing performed for a general globalized product is the beta testing. In this testing type, before launching the software product, a demo version of the software is launched on trail bases. Along with this trail software, the software vendor also demands for the customer feedback so that the changes in the software product will be done. These feedbacks are accepted in the text form and based on the textual analysis the software quality is analyzed. The reliability of the software product is tested under different vectors such as software quality, user friendly, speed, accuracy etc. The feedback can also be taken in the form of suggestions or the features that user want in that particular product. Once these reviews or feedbacks are accepted, the analysis over these reviews is performed to attain the user interest in software product[4][5]. Different aspects of software products respective to which the software product review can be taken is shown in figure 1
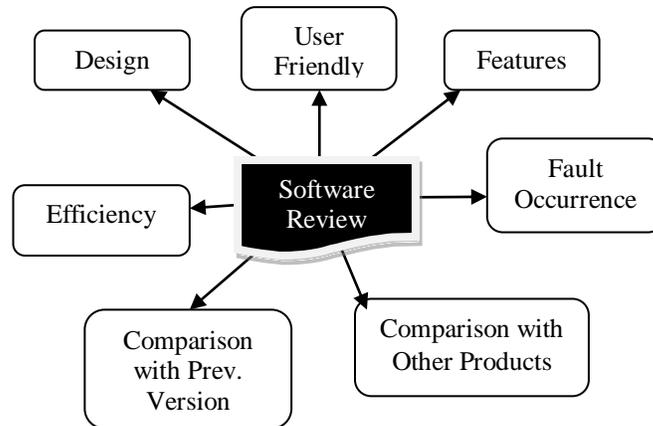

Figure 1 : Different Aspects of Software Review

Software review is collected in the form of simple textual information accepted without specification of any major rules. But the expectation of a software product vendor about this review about different user oriented software elements. These elements are based on software product context analysis. Most of the reviews are collected in the form of design constraint specific analysis. This analysis includes the appearance of the software product and the product features. The availability of these features, look and the user friendly access to the user. Another context based analysis is defined in terms of key feature analysis present in the software product. The expected options in the software product are analyzed by the product user[6][7].

The second main vector of user interest in software product is the efficiency and the reliability vector. The efficiency is here defined in terms of processing speed or the execution of some user task. If the software response time is acceptable then the review obtained will be positive otherwise it will be negative review. The reliability is about the analysis of the software system under the fault or the failure. It includes the analysis respective of fault frequency, fault criticality and the fault associated module analysis. If the software or the particular software module is having fault or the failure, the user review will be negative otherwise it will be positive[8].

The third main vector of software review analysis is the feature analysis in terms of existing software product. The existing software product can be the earlier version of same software product or so other related software product. Users generally perform the comparative analysis of the new product with earlier and based on the feature analysis the reviews are commented. If the feature is improved, the review will be positive otherwise the review will be negative.

639

Once the review is collected from multiple users the next work is to perform the extraction of key terms from this review. These key terms are shown in figure 2.
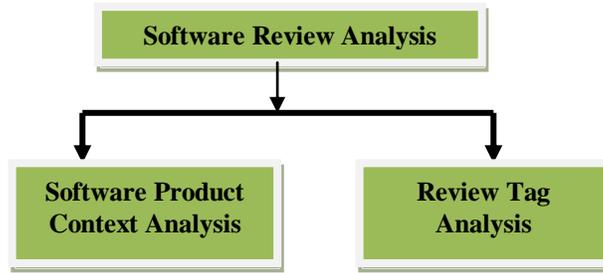

Figure 2 : Key Extraction of Software Review Analysis

Here the context analysis is the extraction of different features defined in figure 1. Each category of software feature is defined by different keywords. The context analysis will identify all these vectors. The second key analysis is the review tag extraction. These tags include the adjectives used by the customer to represent the review. The adjectives are further divided in two categories called positive or negative reviews.

In this section, the exploration of the software review extraction and analysis process is defined. The exploration includes the review keyword extraction and its classification under the context and tag analysis. In section II, the work defined by the earlier researchers is discussed. In section III, different approaches of software review extraction and analysis is defined. In section IV, the conclusion obtained from the work is presented.

II.   EXISTING WORK

In this section, the work defined by the earlier researchers in the area of opinion mining, text mining and software review mining is defined. This stage includes the keyword extraction approaches adapted by earlier researchers. In year 2009, Shishir K. Shandilya has defined a work on opinion mining for text extraction. Author presented the work in the form of a framework to extract the user review as well as analyzed them under the statistical methods. Author performed the review classification and clustering under knowledge based analysis so that the feature analysis will be performed over it[1]. Another work on the software reusability analysis and its prons and cons were discussed by Tim Menzies in year 2003. Author defined a expert opinion analysis and empirical data analysis based work under different vector for software organization. Author defined the comment analysis under the review implementation analysis[2]. Marek Reformat has defined a work on software maintenance analysis under rule based knowledge extraction. Author defined a issue analysis at different levels for rule based evaluation. Author defined a rule similarity analysis and rule inclusion analysis so that the diverse set based evaluation criteria will be performed[3].

In year 2008, Shishir K. Shandilya has presented a work on opinion extraction and classification. Author performed the extraction of information from these opnions or reviews and based on it defined a concrete information flow so that effective decision will be drawn[4]. Another work on software defect analysis based on data mining approaches was defined by Yuan Chen in year 2010. Author defined the software defect analysis and prediction so that the defect prediction and software quality analysis will be performed. Author defined model based approaches for software quality prediction so that the effectiveness of work will be drawn[5]. Another work video recommendation analysis and sentiment classification was defined. The work includes the rating information analysis and the sentiment classification approach so that the effectiveness evaluation will be performed[6]. A work on the change detection in software product under the user perspective was defined by Li-Chen Cheng in year 2011. Author defined an associative classification approach to perform the feature analysis and to identify the keyword extraction over the review. Author defined a framework to analyze the opinion and to identify the product effectiveness under customer point of view. Author presented the work ipad review analysis[7]. JunWei Liang has defined a work on through text mining under the code repository analysis. Author defined the work for an open source product called Chromium. Author presented work as a tool and generated a question set of 7 research question to perform the effective mining[8].

Another work on software quality analysis under the use characteristic mining on customer review. Author defined the ontology oriented analysis for rule based classification so that the software quality will be measured. Author defined a work on characteristic analysis so that the software quality analysis will be performed[9]. Arun Singh defined the work on software quality analysis under the defect prediction so that the effective quality estimation will be performed. Author presented a defect prediction model so that the defect density and the module based analysis will be performed on software product[10].

In year 2013, Warit Leopairote presented the work on software quality analysis under the user review analysis. It includes the software quality analysis under the ontology specification and defined as the characteristic model. The classification approach will include the review analysis in terms of positive and negative reviews. This analysis process is based on intelligent rule based classification approach[11]. Yasutaka Sakamoto has presented a visualizing metric based analysis on personal review process. Author defined supported value analysis for user interactive analysis. This metrics oriented evaluation includes the process improvement under different vectors as well as the source code visualization under different vectors[12]. Rahim Dkharghani has presented a work on tweet sentiment analysis. The software product opinion is here obtained on tweets and later on the tweet analysis is performed to derive the conclusion[13]. Mingxing Wu has presented a product usability analysis under the review mining and usability analysis. Author defined the customer review analysis includes the opinion mining along with feature opinion analysis[14]

III   REVIEW MINING APPROACHES

Software product analysis is defined under the product review
Sentiment Analysis is basically used to identify the conclusion of any review or the feedback by performing the sentiment encoding. This analysis is generally based on the single word, adjective or the phrase. The example of the sentiment analysis is listed here under

Review : The Movie screenplay was *awesome*.

Here "awesome" defines that the movie review is positive and the reviewer like it. But this single word phenomenon is always not accurate, because some of the technical reviews are multiline and having different aspects about different movie qualities. These qualities include the direction, music, screenplay, star cast, story etc. Each category is then analyzed under the respective review class. Once the initial pre-processing is done, there are number of approaches to perform the sentiment analysis. The most common categories of review analysis is shown in figure 3.
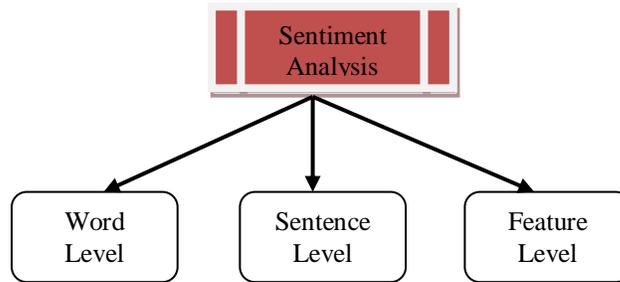

Figure 3 : Type of Sentiment Analysis

A)   Word Level Sentiment Analysis
It is mostly used and effective sentiment analysis technique. An effective encoding is done between the sentiment words and the class. Such as

(Brillient, Awesom, Very Good) => Positive Sentiment

Sentiment Words   Sentiment   Class

There are number of databases that represents the adjective and the respective class. This adjective extraction comes under the lexical analysis of the review. The class generation is a kind of clustering. In the general form, two classes are formed to identify the positive and the negative reviews. The reliability of this approach depends on the adjective or the sentiment word set. The word set must include all the synonyms and antonyms relative to the word. Here the synonym represents the positive sentiments and the antonym represents the negative sentiments.

B) Sentence Level Sentiment Analysis

In this sentiment analysis approach, different levels of granularity are analyzed over the review. A rule based analysis is required to perform the sentence based sentiment identification. These rules include the negation rules extraction approach. It means the sentence or the review having the negative words such as no, not and never are used to represent the negative perspective of the sentiment. Some of the verbs that shows the negative sense also represents the negative reviews such as "stop", "problem" etc. These verbs are also analyzed in different verb forms and the combination. Some of the examples of the sentence level sentiments are given in table 2.

Table 2 : Sentence Level Sentiment

Type	Example	Sentiment
Negation Negative	No Problem	Positive
Negation Positive	Not Good	Negative
Negation Neutral	Will not work	Negative

C) Feature Level Sentiment Analysis

It is one of the most intelligent analysis over the movie review. This analysis process defines the feature identification from the review. This feature is compared from the review set and based on which the orientation score is identified. Each positive feature is assigned with positive weightage and the negative feature is assigned by the negative weightage. Once all the features are collected, the aggregation on the weightage is performed to identify the overall featured score. If the score is positive, the review is considered positive otherwise it is considered negative. The feature analysis approach is based on the statistical or the mathematical formula based on which overall prediction of the sentiment feature will be done.

IV. CONCLUSION

In this paper, a study oriented work is defined regarding the sentiment analysis. This analysis process is defined in terms of different approaches for sentiment analysis. The work also includes the exploration of the sentiment analysis.

REFERENCES

[1] Shishir K. Shandilya," Automatic Opinion Extraction from Web Documents", International Conference on Computer and Automation Engineering 978-0-7695-3569-2/09 © 2009 IEEE

[2] Tim Menzies," More Success and Failure Factors in Software Reuse", IEEE TRANSACTIONS ON SOFTWARE ENGINEERING 0098-5589/03@2003 IEEE

[3] Marek Reformat," Software Maintenance: Similarity and Inclusion of Rules in Knowledge Extraction", Proceedings of the 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'06) 0-7695-2728-0/06 © 2006 IEEE

[4] Shishir K. Shandilya," Opinion Extraction & Classification of Reviews from Web Documents", 2009 IEEE International Advance Computing Conference (IACC 2009) 978-1T-4244-1888-6/08© 2008 IEEE

[5] Yuan Chen," Research on Software Defect Prediction Based on Data Mining", 978-1-4244-5586-7/10@ 2010 IEEE

642

[6]   Weishi Zhang," Augmenting Chinese Online Video Recommendations by Using Virtual Ratings Predicted by Review Sentiment Classification", 2010 IEEE International Conference on Data Mining Workshops 978-0-7695-4257-7/10 © 2010 IEEE

[7]   Li-Chen Cheng," Detecting changes of opinion from customer reviews", 2011 Eighth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD) 978-1-61284-181-6/11 ©2011 IEEE

[8]   JunWei Liang," Analyzing Involvements of Reviewers Through Mining A Code Review Repository", 2011 Joint Conference of the 21st International Workshop on Software Measurement and the 6th International Conference on Software Process and Product Measurement 978-0-7695-4565-3/11 © 2011 IEEE

[9]   Warit Leopairote," Software Quality in Use Characteristic Mining from Customer Reviews", 978-1-4673-0734-5/12©2012 IEEE

[10]   Arun Singh," Assuring Software Quality using Data Mining Methodology: A Literature Study", 978-1-4673-5986-3/13©2013 IEEE

[11]   Warit Leopairote," Evaluating Software Quality in Use using User Reviews Mining", 2013 10th International Joint Conference on Computer Science and Software Engineering (JCSSE) 978-1-4799-0806-6/13©2013 IEEE

[12]   Yasutaka Sakamoto," Visualizing Software Metrics with Service-Oriented Mining Software Repository for Reviewing Personal Process", 2013 14th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing 978-0-7695-5005-3/13 © 2013 IEEE

[13]   Rahim Dehkharghani," Automatically Identifying a Software Product's Quality Attributes through Sentiment Analysis of Tweets", NaturaLiSE 2013 978-1-4673-6271-9/13@ 2013 IEEE

[14]   Mingxing Wu," A Novel Approach Based on Review Mining for Product Usability Analysis", 978-1-4673-5000-6/13©2013 IEEE