



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 7, Issue 6, June 2019

Hybrid Feedback Analysis system for Educational Institutions

Nikita Dipak Bachhav¹, Dr.Prashant Nitnaware²

Department of Computer Engineering, Pillai College of Engineering, New Panvel, Navi Mumbai, India¹

Department of Computer Engineering, Pillai College of Engineering, New Panvel, Navi Mumbai, India²

ABSTRACT: This paper proposes an academic feedback system that allows students to enter their opinions anonymously. This automated feedback system collects opinions as a set of normalized and un-normalized responses. Initially, sentiment analysis computes positive and negative opinions from given feedbacks followed by data mining algorithms like Naive Bayes and SVM classification to refine the analytical results. The conclusion is based on combined values from all these analysis. This proposed system intends to improve quality of service of educational institutions.

KEYWORDS: Supervised learning, unsupervised learning, Sentiment Analysis, Machine Learning.

I. INTRODUCTION

Feedback is an important insight to improve quality of product and amenities. Persistently increasing high use of internet has formed many ways to acquire these feedbacks from the consumers, employers or workforces. Various e-commerce sites, social sites and blogs collect opinion of their active users/ customers regarding their goods and services. Almost every organization is using social media, forums, and online platforms to allow its users to express their experience and advices regarding the commodities. These opinions can provide information about user's likes, choices, possible future purchases and many other things. Through feedbacks, companies can improve their quality of services as well as identify ways of new product recommendation for customers.

The feedbacks can be in any form such as reviews, polls, comments, and rating. These reviews have been categorized as normalized responses and un-normalized responses. Normalized responses are the responses given in the form of direct answer, for example, objective questionnaire or yes/no response. Un-normalized responses includes subjective expressions and complete description of experience. Reviews contain useful information but it's not easy to extract the actual sense of each review accurately through automated system. Many techniques have been developed to extract the emotions from the text or images to know what exactly user wants to say.

Data mining is a technique that deals with the detection of hidden features of data from a data set. It uses many machine learning algorithms for discovery of unknown features. Machine learning is a field that uses algorithms for data analysis and accurate prediction of features of data without external interference. The training data set predicts the feature of testing data set and concludes the result. Based on the availability of training data, the learning algorithms have been classified as supervised and unsupervised learning methods.

A. Supervised:

- System is trained using training data set. Features are already known and based on these features training data identifies the features of testing data and categories them accordingly. Examples, classification and regression.

B. Unsupervised learning:

- The unsupervised learning methods are useful when features are unknown. These algorithms identify some features and based on that they perform prediction on testing data set. Examples, clustering methods, anomaly detection, neural networks etc.



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 7, Issue 6, June 2019

The Naive Bayes classification technique is a supervised learning mechanism which uses very small set of training data for estimation of features required for classification. It also has semi-supervised learning approach. Similarly, SVM classifier is again a type of supervised learning model.

Sentiment Analysis is the process of identifying and categorizing thoughts expressed in a form of text to determine opinionant's response and categories as positive, negative, or neutral. Sentiment can be calculated at different level. Document Level - in this level, the whole document is given a single polarity positive, negative or neutral. Sentence Level - in this level, document is classified at sentence level. Each sentence is analysed separately and classified as negative, positive or neutral. Thus, overall document has a number of sentences where each sentence has its own polarity. Phrase Level – in this level, involves much deeper analysis of text and deals with identification of the phrases or aspects in a sentence and analysing the phrases and classify them as positive, negative or neutral. It is also called aspect based analysis.

II. RELATED WORK

In [1] authors proposed a conceptual framework for analyzing student feedback. In this framework, natural language techniques are used for pre-processing and text mining techniques are used to extract the various dimensions from responses. In this framework, correlation is performed to link the various topics within qualitative feedback. This will help for deeper analysis of the collected data. In similar manner [4] proposed a system using combination of machine learning techniques and lexicon based approach for sentiment analysis of student's feedback which is trained using TF-IDF and lexicon based features to analyse sentiments expressed by student. In [7] authors have developed a system to analyse the students feedback which was collected from Facebook and applied a machine learning algorithm like support vector machine, Maximum entropy, naive Bayes and complement naive bayes to classify sentiments which define the optimal tool for feedback analysis. In [8] have proposed a framework for review analysis of online movie reviews of Box Office. In this system they have used both clustering and classification. They have mainly focuses on unsupervised clustering algorithms like K-Means and DBSCAN but system is not able to handle the double negation. In [11] authors have proposed a unified rating system to address the issue of opinion analysis and numeric rating by combining the numeric polarity of review and starred rating. In [9] authors have developed a framework to analyze the product review of e-commerce site. It has been filtered to remove noisy data and has been pre-processed to evaluate sentiment of the reviews using supervised algorithms like Naïve Bayes, Support Vector Machine (SVM) and Decision Tree. In [12] authors have proposed a framework on mining review from e-commerce websites uses Naïve Bayes text classifier, logistic regression and SentiWordNet to classify the review which concludes that based on their results Naïve Bayes is efficient among three algorithms. In [15], authors have proposed a feedback analysis system for e-commerce organization which will help buyer to select best product. It will collect data from e-commerce sites and done Pre-processing on collected data. Comments will be classified in three classes positive, negative and neutral and features will be identified to calculate the trust score. In [17], [20] authors have developed a system to classify product review in semantic meaning. This system works with different approach including spelling correction and implemented a hybrid algorithm using Decision tree and Naïve Bayes. Data is collected using crawlers and pre-processing is done to reduce ambiguity, redundancy and incomplete data. SentiWordNet is used to assign a polarity to a word. Naïve bayes is used to classify the data and decision tree calculates the final polarity after classification. Main aim of these systems is to remove anaphora's occurring in a review. To remove this occurring reference resolution is used. Stanford parser is used to extract the opinion and SentiWordNet is used to assign the scores. In [18], [19] authors have discussed the work on sentiment analysis based on machine learning approach. Major work is done on supervised learning algorithms and rule based classifiers are not used much. According to the authors, unsupervised techniques are useful when it comes to complicated training data set. Reviews are the most used form of data sources available widely. Opinion mining plays a very important role while making a decision towards a particular product.

III. PROPOSED ALGORITHM

This work proposes a simple algorithm for robust feedback analysis for an educational institute. Some of the limitations were found in the existing feedback analysis system. They are listed below:

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 7, Issue 6, June 2019

- Most of the works are based on unstructured data for feedback analysis. This data gives the polarity of positive and negative words.
- In some systems only correlation is performed to link the various topics in qualitative feedback analysis.
- Some of the works are limited to compute an overall feedback of students.

In order to overcome these drawbacks, a feedback analysis system is proposed. Architecture of feedback analysis system is shown in Fig. 1

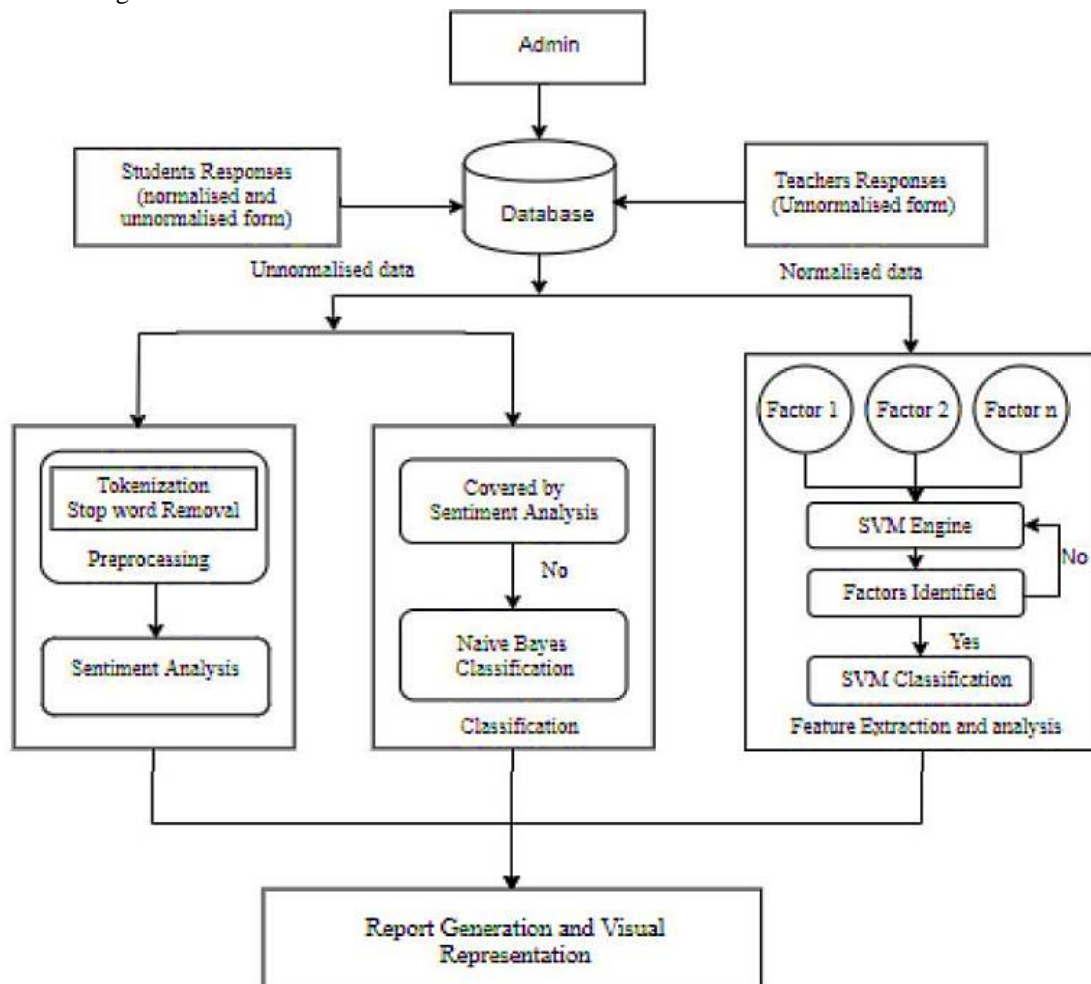


Fig.1. Proposed System Architecture

The proposed system is consist of 5 phases

- Data collection phase
- Sentiment Analysis
- Classification
- Feature extraction phase
- Report.



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 7, Issue 6, June 2019

Admin has full authority to add and delete the name of any faculty from the list of feedback analysis system as well as can add, delete the subjects and programmes offered and can download report in pdf format. The different phases of the system are as follows.

A. Data collection phase:

In this phase, students will select the programme they are enrolled in. After selecting a programme subject list will be displayed according to their courses. Students will add feedback for their respective courses and teachers. The feedback is in normalised as well as un-normalised form. Students will give their feedbacks marking their responses from the given options of the objective questionnaire. Along with this, a text box is given, so that, they can express their opinions about other factors which are not included in the questionnaire. Feedback will be anonymously submitted to the system. Similarly, there is a module where faculties can provide feedback regarding the courses and many other facilities [7].

B. Sentiment analysis:

In order to process un-normalised data sentiment analysis is performed. Input is programme and subject, once the admin will select a subject and program it will fetch data from database for that particular subject. This phase is consist of two parts. One is pre-processing and another is sentiment analysis which will calculate the polarity.

Step 1: Pre-processing:

Student feedback data represents an unstructured text. To extract useful information from the unstructured text, several pre-processing steps are applied to remove spelling errors, grammatical mistakes, URLs, etc. from the text.

Step 2: Tokenization

Tokenization is the process of breaking up a sequence of strings into pieces such as words, keywords, phrases, symbols and other elements called tokens. Tokens can be individual words, phrases or even whole sentences. In the process of tokenization, some characters like punctuation marks are discarded [7].

Step 3: Stop word removal

Stop words are used in every language. These words are eliminated to focus on important words. For example, search engine query is "Who is the Prime minister of India.", in this type of query search engine searches for the words like "who", "Prime", "minister", "India" then it retrieves more pages containing the words "India", "minister". So, by deleting or removing these stop word we can get proper data for analysis.

Step 4: Polarity measurement

In sentiment analysis, it will calculate the overall number of count for comments. After pre-processing unstructured data tokens will be collected. Sentiwordnet is lexical resource for opinion mining. Sentiwordnet assigns scores as positive, negative and neutral. Score for each token will be calculated with the help of Sentiwordnet.

C. Classification phase (Naïve Bayes classifier):

This classifier uses Bayes theorem to predict the probability, that a given set of features is a part of particular label. It uses bag of words (BOW) model for feature extraction. This model assumes that all the features are independent [17].

$$P(\text{label}/\text{features}) = P(\text{label}) * P(\text{features}/\text{label})/P(\text{features})$$

where $P(\text{label})$ = prior probability of label

$P(\text{features}/\text{label})$ = prior probability that feature set is classified as label

$P(\text{features})$ = prior probability that feature set will occur.

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 7, Issue 6, June 2019

A Naïve Bayes classifier implements Bayes theorem with simplified assumptions to classify inputs based upon the presence or absence of features. Here, the features are distinct words and the classifier characterizes each training data comment. Using this characterization, the classifier computes the most probable category of each new comment based on the presence or absence of words from the training data. The Naïve Bayes classifier implementation used here is *classify*, which is designed as a multipurpose classifier for the classification of text sources. It is designed to be quick and simple to use – and offers no further control other than input texts. No manual pre-processing of the text was done before loading into the classifier system. As before, set comments were used to train the classifier for positive and negative sentiment. Some individual test data comments will be loaded into the classifier, and the category with the highest probability will be recorded [17].

As the comments will be positive, negative and neutral. There are only two predefined classes that is positive and negative. As the new comments will come classifier will calculate its probability whether that comments is positive or negative.

D. Feature Extraction and Data analysis:

In this phase, Support vector machine classification engine is used to extract the features. Support vector machine is a supervised learning model which is used for classification. Its main aim to determine best linear separators for classification. It is a non-probabilistic classifier. For a given set of training data, each is labelled for belonging to one of the classes, SVM training algorithm create a model which assign new data into one or two classes. Hyper plane is used to separate two classes. In the diagram below, for example, to classify triangle and circle shapes we compute three hyper planes A, B and C. C is best separator as items on both sides are at maximum distance from Hyper plane and B is worst separator [4].

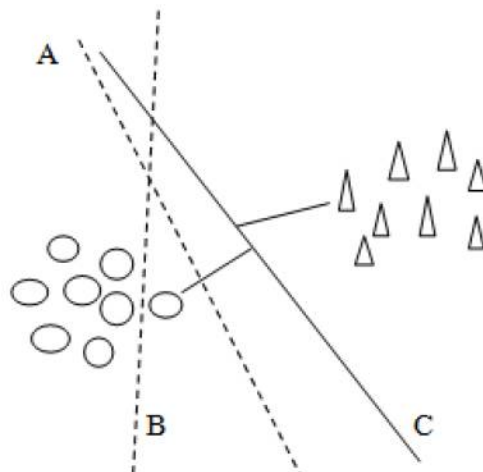


Fig.2. Classification using Support Vector Machine (source [20])

In student's feedback system, normalised data collected from students is processed with the help of support vector machine engine. Support vector machine will extract the feature from normalised data and calculate the sentiment value for that feature. List of questions will be given to students from that questions feature will be extracted [16]. For example,

- Did the course content meet your expectation?
- Feature- Course outcome
- Course increases your knowledge?
- Feature- Learning



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 6, June 2019

Where the lecture easy to understand?
Feature- Explanation methods
Did the teacher explain the material well?
Feature- Teaching skills
Course Completion in Time?
Feature- Time
Standard of Lab facilities?
Feature- facilities

In this way features will be extracted from the questionnaire. Now the count of positive and negative will be displayed for each feature after support vector classification. It will classify data linearly. In this phase it will give the output for features and their respective positive negative scores. Also analysis of feedback will be displayed in this, which will include session id, course, and sentiment value per comment wise. Also the detailed analysis of the feedback can be reviewed in the analysis section.

E. Report

This the last phase of feedback analysis system. In this the detailed report will displayed in the form of pdf which can be downloaded for further use. This report will include the positive and negative output of sentiment analysis, refined results of classification module and feature extraction and feedback analysis using support vector machine. Admin can download this pdf and forward it to higher authorities. Along with the analysis, results could be displayed in graphical form for better visual representations and comprehension.

IV. APPLICATIONS

A. Education institutes:

Educational Institutions are always concern about the quality of education provided to their students to maintain or achieve higher standard as per government or regulating bodies. Online Feedback System is a web application which provides platform to schools and colleges to conduct feedbacks online. This is easy and requires less time.

B. Review analysis:

Every online business is concern to grow more and provide customer satisfaction and develop customer loyalty. The online feedback system is always best way to know about consumers opinions. It is very common to give reviews after purchasing any items for ecommerce site, visiting newly opened restaurants and movie reviews. Analysis of such reviews helps to understand pros and cons of targeted organizations.

C. Economic forums:

Many online economic forums are active which are continuously updating the information about financial investments, stock market and fund related information. People who take advice or suggestion from such forums always post feedbacks. Analysis of such feedback will help new user to take correct decision before investing.

D. Social media opinion analysis:

Social media platforms are widely used now a days. People are using this platform to conduct the surveys for various topics and create awareness. Analysis of such surveys can be done with the help feedback analysis system.

E. Organizations:

Organizations give appraisal to employees based on their performance. Performance evaluation depends on various factors. In this case feedback analysis system will help to evaluate the performance of employees to make the appraisal process easy and informed decisions can be taken.

V. CONCLUSION

In proposed system data mining methodology for feedback analysis helps to understand the performance of the faculties of an educational institution based on certain characteristics. This rating includes qualitative as well as quantitative feedback. Processing the feedbacks submitted by the students using sentiment analysis, Naive Bayes and support vector classifier gives accurate and efficient result. The graphical representation improves the visual



International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijirccce.com

Vol. 7, Issue 6, June 2019

representation and understanding about the feedback. This helps to maintain or improve higher educational standards. This system can be modified to collect feedbacks for other organizations or companies with different set of questionnaire.

REFERENCES

1. Sandy Gan, Swapna Gottipati, Venky Shankararaman, "A conceptual framework for analysing student's feedback," 2017.
2. M. Fernandez-Gavilanes, T. Alvarez-López, J. Juncal-Martínez, E. Costa-Montenegro, and F. J. González-Castano, "Unsupervised method for sentiment analysis in online texts," *Expert Systems with Applications*, vol. 58, pp. 57–75, 2016.
3. P. D. Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews," in *Proceedings of the 40th annual meeting on association for computational linguistics*. Association for Computational Linguistics, 2002, pp. 417–424.
4. Zarmeen Nasim, Quratulain Rajput and Sajjad Haider, "Sentiment Analysis of Student Feedback Using Machine Learning and Lexicon Based Approaches," 2017.
5. L. Zhang, R. Ghosh, M. Dekhil, M. Hsu, and B. Liu, "Combining lexicon-based and learning-based methods for twitter sentiment analysis," Technical Report, 2011.
6. O. Appel, F. Chiclana, J. Carter, and H. Fujita, "A hybrid approach to the sentiment analysis problem at the sentence level," *Knowledge-Based Systems*, vol. 108, pp. 110–124, 2016.
7. Mohammad Aman Ullah, "Sentiment Analysis of Student Feedback: A case study towards optimal tool," 2016.
8. Nagamma P., Pruthvi H.R, Nisha K, "An improved sentiment analysis of online movie reviews based on clustering for Box office," 2015.
9. Zeenia Singla, Sukhchandan Randhawa, Sushma Jain, "Sentiment Analysis of 23 Customer Product Reviews Using Machine Learning," 2017.
10. N. Altrabsheh, M. Cocea, and S. Fallahkhair, "Learning sentiment from student's feedback for real-time interventions in classrooms," in *Adaptive and Intelligent Systems*. Springer, 2014, pp. 40–49.
11. Mir Riyanul Islam, "Numeric rating app for google play store by sentiment analysis on user reviews," 2014.
12. Santhosh Kumar K L, Jayanti Desai, Jharna Majumdar, "Opinion mining and sentiment analysis on online customer review," 2016.
13. Q. Rajput, S. Haider, and S. Ghani, "Lexicon-based sentiment analysis of teacher's evaluation," *Applied Computational Intelligence and Soft Computing*, vol. 2016, p. 1, 2016.
14. Han, H.J., Mankad, S., Gavirmeni, N., & Verma, R. (2016). What guests really think of your hotel: Text analytics of online customer reviews. *Cornell Hospitality Report*, 16(2), 3-17.
15. Sonali Bagul, Rakhi Wajgi, "Design Feedback Analysis System for e-commerce Organization." 2016.
16. Balaji. Jagtap and V. Dhotre, "SVM and HMM Based Hybrid Approach of Sentiment Analysis for Teacher Feedback Assessment" *International Journal of Emerging Trends of Technology in Computer Science (IJETCS)*. vol. 3, no. 3, pp. 229–232.
17. Gurneet Kaur, Abhinash Singla, "Sentiment Analysis of Flipkart reviews using Naïve Bayes and Decision Tree algorithm", 2016.
18. Neetika Bansal, Ashima Singh, "A Review on Opinionated Sentiment Analysis based upon Machine Learning Approach", 2015.
19. Shivprasad T K, Jyothi Shetty, "Sentiment Analysis of Product Reviews: A Review", 2017.
20. Hari Krishna, Ali Akbar, "A Feature Based Approach for Sentiment Analysis using SVM and Co-Reference Resolution", 2017.