International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 04 | Apr 2019 www.irjet.net p-ISSN: 2395-0072

Sentimental Analysis for Students' Feedback using Machine Learning Approach

Kousalya. L[1], Subhashini. R[2]

[1]*Sathyabama Institute of Science and technology, Dept. of Information Technology, Chennai, India.*
[2]*Sathyabama Institute of Science and technology, Dept. of Information Technology, Chennai, India*

Abstract - *As World Wide Web is growing at higher rate, this has resulted in enormous increase in online communications. The online communication data consist of feedbacks that are posted by students. Sentiment analysis system classifies text data into their respective sentiments of positive polarity, negative polarity or neutral. There are some other robust classifiers which have ability to provide comparable or better results. In this project, we try to focus our task of sentimental analysis for students' feedback collected through online. We examine the sentiments present in the text document for classification of students' feedback based on polarity (positive/ negative/ neutral) using machine learning and lexicon based approach. Also we have used the Random Forest classifier for the evaluation of performance and for finding the accuracy. By using Random Forest classification technique we have achieved the best accuracy of 90%.*

Key Words: Sentimental Analysis, NLP, Feature Extraction, Polarity.

1. INTRODUCTION

In the existing system students used to give feedbacks manually on paper feedback forms. It was a time consuming and very inefficient process. Then the forms are collected and the HOD's views the feedbacks of students and analyze the performance of that teacher of that particular department. Then to overcome the limitations of that system came the online feedback systems which takes the feedback of students online and automatically analyzes the feedback to analyze the performance of teachers. But the existing online feedback systems only analyzes the objective type questions it doesn't analyzes the descriptive type questions. In this project a sentiment analyzer is implemented to analyze the descriptive type questions so to increase the accuracy of the feedback system[1]. The sentiment analyzer is build using machine learning algorithms. There's an algorithm which analyzes the descriptive type questions. The algorithm is trained using the training data set which contains positive and negative words. And then the model is created using the trained algorithm. Then the test data is given to the algorithm to see the accuracy of the system. If it lacks in accuracy, then more training data and features are added to the analyser to increase the accuracy of the system and in such manner the machine learning is implemented to build the sentiment analyser. The taking of feedback plays a very significant role in the life of students as well as the teachers. The students give the feedback so to convey what is the difference between the actual teaching which is currently taking place in colleges and what type of teaching students really desire for. And these feedbacks show the teachers their overall performance in their particular subjects. They can improve their teaching accordingly. This system is a secured system. The identity of the students giving feedback is not disclosed to anyone not even the admin. And a single student can give only a single feedback to a particular teacher. The accounts of students are created by the admin so no one other than the students can give the feedback. Sentiment analysis has received much attention from research and industry communities recently. In this feedback system, a database is created which contains negative and positive words. Then it contains a java API which is used to parse and check the words present in the student's descriptive type answers if there positive or negative word by comparing it with the words present in the database. Then there's an API to set the database information. Such as which driver is used, which port number is database present in and the username and password for accessing the database. These API's are transformed into jar files and added to the libraries of the project.

1.1 RELATED WORK

Sentiment Analysis has been extensively studied during the past few years. The reported work can be broadly classified into three main approaches: (a) machine learning based, (b) lexicon-based and (c) hybrid.

1.2 MACHINE LEARNING BASED

Machine learning based approaches of sentiment analysis learn a predictive model using the provided training dataset and evaluate the performance of the learned model on the test dataset. It can be further classified into supervised learning and unsupervised learning methods.

1.3 LEXICON BASED

Lexicon based approach of sentiment analysis makes use of a sentiment lexicon to determine the polarity of a given textual content. A lexicon or dictionary represents a list of words with associated sentiment polarity. The lexicon can be

constructed either manually or automatically. They utilized an online lexical resource WordNet to predict the semantic orientation of an opinion word. Taboada et al. He proposed another lexicon-based approach that determines the polarity of a word by using the dictionaries constructed.

1.4 HYBRID APPROACH

Hybrid Approaches use sentiment lexicon in machine learning methods. Zhang et al. [10] proposed a hybrid approach for sentiment analysis of Twitter data. An opinion lexicon was used to label training dataset with sentiment polarities. The labeled dataset was then used to train a binary classifier to predict sentiment polarity on the evaluation dataset. Appel et al. [11] performed sentiment analysis at the sentence level using a hybrid approach. Their approach was based on a sentiment lexicon extended using SentiWordNet and fuzzy sets to determine sentiment polarity of a sentence. This paper also presents a hybrid approach that combines the use of sentiment dictionary and machine learning methods to determine the semantic orientation of a textual feed provided by students.

2. PROPOSED SYSTEM

In this system, we are using machine learning based approach for sentiment classification. For this, we are constructing dataset of feedbacks got from students realtime. After obtaining the feedbacks, they are pre-processed to remove the noise. The feedbacks are labelled as either positive, negative or neutral. After pre-processing, useful and significant features are extracted from tweets. The machine learning classifiers are applied on the training dataset. The model obtained from training, is applied on unseen test dataset to check the accuracy of the model. A web application will be created which will display the results of the classification. The results are visualized and displayed on website for user convenience.

2.1 DATA COLLECTION

The real-time students' feedback is collected through an online student portal. Where a student has a separate login. Then a student can give single comment for per login id.

2.2 DATA PREPROCESSING

The text pre-processing techniques are divided into three subcategories:

Tokenization: The data present in the text document contains block of characters called tokens. These text documents are separated as tokens and used for further processing of data.

Removal of Stop Words: A web search tool or other natural language processing system may contain collection of stoprecords, or it may contain a solitary stop-list. Most of the more frequently used stop words in English are "an", "a", "of", "the", "you", "and" these are some words which do not carry any meaning. Hence, those words which appear too often that support no information for the task are removed.

Part of Speech Tagging: POS tagger parses a sentence or document and tags each term with its part of speech. For part-of-speech tagging we used the Stanford partof-speech tagger. This tagger used by splitting text data into sentences and to produce the POS tag for each word (whether the word is a noun, verb, adjective). Consider following example "Staffs are amazing".

In part-of-speech (POS tagging), each word in review is tagged with POS (such as noun NN, adjective JJ, verb RB). In tagged sentence, amazing is tagged with tag JJ which indicates 'amazing' is an adjective where as a 'movie' is tagged as NN which indicates noun.

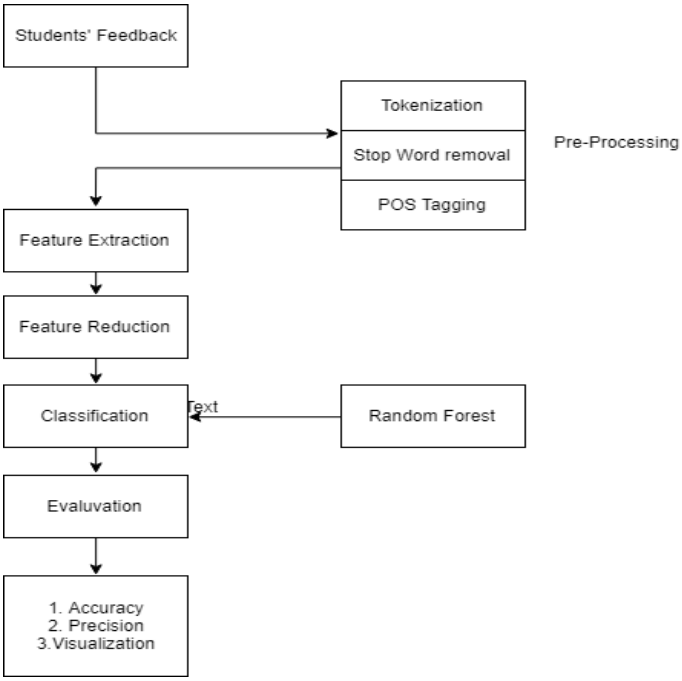

Fig -1: Work flow diagram

2.3. FEATURE EXTRACTION

In the process of feature extraction, movie features are extracted from every sentence. For finding the polarity of text document, it is necessary to understand the sentiment score with its usage as well as their relationship with all the nearby words. Following are some features that affect the polarity of the document.

1) Positive Sentiment Words: These are the words which are having a positive sentiment score according to

SentiWordNet. For example: Nice, Good, Fantastic, Pretty, Outstanding etc.

2) Negative Sentiment Words: These are the words which are having negative sentiment score according to SentiWordNet. For example: Bad, Awful, Disgusting, Pathetic etc.

2.4. FEATURE REDUCTION

One of the biggest problems of sentimental analysis is dealing with text data which are available in very high dimensions which may affect the performance of classifier. So, there is a need for such technique which will eliminate those features that are not relevant and keeping only those features which are much important and the techniques which will help to differentiate the sentences into class labels such as positive and negative. The Information Gain and Gain Ratio are the most popular techniques among number of feature reduction techniques.

2.5. MODEL TRAINING

Model Training After the extraction of features from the train and test dataset, learning algorithms were applied for training model. The hybrid model for sentiment analysis was trained using unigrams, bigrams, TF-IDF and lexicon-based features. A brief description of the learning algorithms is given below:

1) Random Forest: Random Forest Algorithm was proposed by. In this study, scikit-learn implementation of Random Forest algorithm was used. The hyper parameters were tuned using three fold cross validation.

2) Support Vector Machines (SVM): The scikit-learn implementation of SVM with linear kernel was used to train model.

3. RESULT

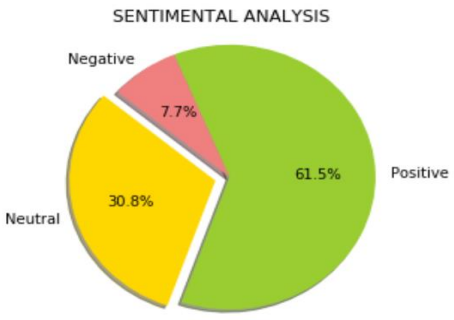

Fig -2: overall result

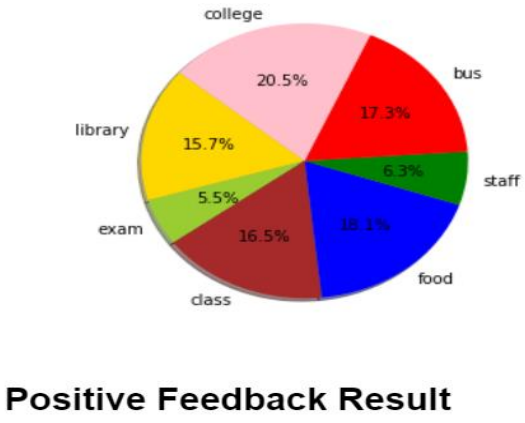

Fig -3: Positive feedback

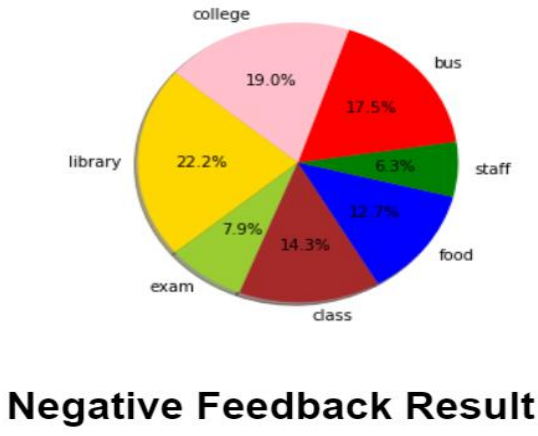

Fig -4: Negative Feedback

4. CONCLUSION

Sentimental analysis has become popular research area due to the increasing number of internet users, social media etc. In this work, we extracted new features that have a strong impact on finding the polarity of the movie reviews. We then perform the feature impact analysis by estimating the information gain for each feature in the feature set and used it to derive a reduced feature set. The main goal of this work is to classify the sentences according to its sentiment by using Random Forest classification technique. This process of extracting the text having sentiment deals with finding the sentiment feature set from the sentences. As final output is displayed graphically it becomes easier for user to understand the exact polarity result. In future work we would like to apply the concept of NLP in more detail for the better prediction of the polarity results. We would like to use the best classification technique for achieving the highest accuracy. This technique can also be implemented on other domains of opinion mining such as product reviews, political discussion forums, hotels, tourism etc.

REFERENCES

[1] P.Nagamma, Pruthvi H.R, Nisha K.K, Carlos Soares," An ImprovedSentiment Analysis of Online Movie Reviews", IEEE 2015, International conference on Computer and Inforamation Technology.

[2] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?:sentiment classification using machine learning techniques," in Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10. Association for Computational Linguistics, 2002, pp. 79–86.

[3] J. Erman, M. Arlitt, and A. Mahanti, "Traffic classification using clustering algorithms," in Proceedings of the 2006 SIGCOMM workshop on Mining network data. ACM, 2006, pp. 281–286 A. Baloglu, Mehmat A. Aktas, "An Automated Framework for Mining Reviews from Blogosphere," International Journal on Advances in Internet Technology, vol. 3, 2010.

[4] Turney, Peter, and Michael L. Littman. "Unsupervised learning of semantic orientation from a hundred-billionword corpus." (2002).

[5] Baccianella, Stefano, Andrea Esuli, and Fabrizio Sebastiani. "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. "LREC. Vol. 10. 2010.