IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 4, April 2018

Intelligent Feedback System Using Latent Semantic Analysis

Tejesh Gaikwad[1], Shantanu Mirashi[2], Rahul Shinde[3], Pradip Shinde[4], Prof. S. A. Deshpande[5]

Department of Information Technology, PVG's College of Engg. & Tech, Pune[1-5]

Abstract: Feedback is most important for any organization to analyse and improve. Getting accurate feedback within time is important for an organization to grow. The purpose of this project is to get online feedback from the students' comments. The comments will be analyzed using the latent semantic analysis and then show the summarized results in the text format. The system takes a student's comment given on the lecture and analyses it to provide better performance of the lectures. The feedback system uses the semantic analysis to understand the meanings of the comments, after finding the Sentiment of student's comments and summarizes the result. After the analysis, summarization is done and the result is shown. After the process is done system shows the quality of the lectures.

Keywords: Sentiment analysis, NLP

I. INTRODUCTION

The online feedback system takes students comments; these comments will be analyzed by using the latent semantic analysis and will show the summarized results in the text format. The system takes a student's comment given on the lecture and analysis of these comments can help to provide better lectures performs. The feedback system uses the semantic analysis to understand the meanings of the comments, after finding the Sentiment of student's comments and summarizes the result. After the analysis, summarization is done and the result is shown. After the process is done system shows the quality of the lectures.

Semantic analysis is the study of language; it is the process of relating syntactic structure. The idea of transformational generative grammar is introduced by the syntactic structures. This method uses some of the phrase rules which can break down the sentences into more than one part. Sentiment and semantic analysis use the natural language processing. NLP contains, first we take all comments text data from database by using the RSS and RSS Feeder, because comments are stored in the database. This text data is stored in one file which generated through the RSS feeder. After converting, divide text in some parts for understanding the meanings of words by using the auto tags algorithm. When the text is divides into small parts then system applies sentiment analysis to understand the meaning of word. After done these process systems summarize and show overall results in the text format.

II. LITERATURE SURVEY

The main goal is to create a web application which will provide the facility to submit their feedback and show feedback result. Some aspects of our model take references from these researches. Sentiment analysis has been studied in wide area of domain such as movie review, teaching review [1] [4], product review, e-learning [2], hotel review and many more. Most scholars focused to quantitative data analysis. However, some studies have been done on qualitative data using sentiment analysis, we found six works that mentioned the idea of using opinion mining and sentiment analysis in education.

El-Halees [1] proposed course evaluation model using students' attitudes posted on discussion forum written in Arabic language to improve course evaluation. Naive Bays, k-nearest and Support Vector Machine are used in opinion classification. To extract the features for each course, the author used Gate Information Extraction tool and RapidMiner data mining tool to classify the students' posts. The experimental result shows that Naive Bays method has better performance than the other two methods. Their system calculates the positive or negative orientation by computing the difference between the number of positive and negative adjective words. The author didn't consider the score of opinion words. This work suggests the effectiveness of user-generated content to improve course evaluation.

III. PROPOSED WORK

The web application which will be a gathered data for the analysis on comment portal. Learn a semantic analysis from a word and phrases enter by the students to giving a review for the teacher. To create a teacher panel where the teacher can able to add the attendance of the students as per the data and the lecture basis. Each teacher can mark the attendance of the students, after marking the attendance we are allowing to the students to enter the review of the

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 4, April 2018

teacher's lecture. That allowed for the limited time period. But student can be able to see the complete his given reviews as per the selected date wise. Also, we are analysing this data reviews. And analysing the teacher's performance, this can be save lots of paper work and lots of time also.

Analysing newly added comment texts one by one to get its' emotion. Mark it with the appropriate emotion (word) of identified emotion. Update overall analysis of that particular comments, an important note is, this research concentrated on developing a model to accomplish the objective of the project, but not to develop a program (code based) or web-based application. Further, the research assumes that there are no emoticons already inserted within the analysed texts and they have written in grammatical way according to the English language

1. Processing:

Here we are gathering the data for the analysis of the admin on that comment using semantic analysis. Now for the analysis and generation of report and summarize the report we will perform the sematic analysis with natural language processing tools.

2. Retrieval Response:

From that comments and reviews, using semantic analysis are conclude that the teacher teaching is good, bad, average or better. This analysis can be allowed to see to the only admin. No any teacher can saw the student's reviews, but admin can able to saw all the reviews student by student and Complete teacher and subject wise analysis.

3. Generative Response:

Select analysing method identification: Most of the reviewed SA approaches within the enhancements of regular SA that followed supervised learning approaches which used large dataset of sentences to train the models. Also, those sentences need to be annotated before input to the models. The activity of annotating the sentences is the most time-consuming function in SA. Process of LSA: Representing the text or phrases through a Term Document Matrix (TDM) is the first step of LSA procedure. A description of TDM could be find in next sub topic. Then the theory of Singular Value Decomposition (SVD) is applied to the TDM in order to determine patterns and relationships among terms found in text collection. Then the results used to determine similarities of given new text. Before starting any sort of emotion identification analysis, the texts needed to be clean. First, the texts will be lowercased. Then all special characters and stop words will be removed. To avoid conflicts among tenses and derivation of words, all words in the text will be stemmed.

Algorithms:

1] Ensemble SVM

To improve the limited classification performance of the real SVM, we propose to use the SVM ensemble with bagging (bootstrap aggregating) or boosting. In bagging, each individual SVM is trained independently using the randomly chosen training samples via a bootstrap technique. In boosting, each individual SVM is trained using the training samples chosen according to the sample's probability distribution that is updated in proportional to the errorless of the sample. In both bagging and boosting, the trained individual SVMs are aggregated to make a collective decision in several ways such as the majority voting, least-squares estimation-based weighting, and the double-layer hierarchical combining. Various simulation results for the IRIS data classification and the hand-written digit recognition, and the fraud detection show that the proposed SVM ensemble with bagging or boosting outperforms a single SVM in terms of classification accuracy greatly.

2] Boosting

Boosting is a machine learning ensemble meta-algorithm for reducing bias, and also variance in supervised learning, and a family of machine learning algorithms which convert weak learners to strong ones. Boosting is based on the question Can a set of weak learners create a single strong learner? A weak learner is defined to be a classifier which is only slightly correlated with the true classification (it can label examples better than random guessing). In contrast, a strong learner is a classifier that is arbitrarily well-correlated with the true classification.

3] Recursive Partitioning

Recursive partitioning is a statistical method for multivariable analysis. Recursive partitioning creates a decision tree that strives to correctly classify members of the population by splitting it into sub-populations based on several dichotomous independent variables. The process is termed recursive because each sub-population may in turn be split an indefinite number of times until the splitting process terminates after a particular stopping criterion is reached.

4] Random Forest based on Bagging Algorithm

Bagging constructs a large number of trees with bootstrap samples from a dataset. But now, as each tree is constructed, take a random sample of predictors before each node is split. For example, if there are twenty predictors, choose a

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 4, April 2018

random five as candidates for constructing the best split. Repeat this process for each node until the tree is large enough. And as in bagging, do not prune.

Random Forests Algorithm
The random forests algorithm is very much like the bagging algorithm. Let N be the number of observations and assume for now that the response variable is binary.

1. a random sample of size N with replacement from the data (bootstrap sample).
2. Take a random sample without replacement of the predictors.
3. Construct a split by using predictors selected in Step 2.
4. Repeat Steps 2 and 3 for each subsequent split until the tree is as large as desired. Do not prune. Each tree is produced from a random sample of cases, and at each split a random sample of predictors.
5. Drop the out-of-bag data down the tree. Store the class assigned to each observation along with each observation's predictor values.
6. Repeat Steps 1-5 a large number of times (e.g., 500).
7. For each observation in the dataset, count the number of trees that it is classified in one category over the number of trees.
8. Assign each observation to a final category by a majority vote over the set of trees. Thus, if 51% of the time over a large number of trees a given observation is classified as a "1", that becomes its classification.

Why Random Forests Work
Variance reduction: the trees are more independent because of the combination of bootstrap samples and random draws of predictors.

- It is apparent that random forests are a form of bagging, and the averaging over trees can substantially reduce instability that might otherwise result. Moreover, by working with a random sample of predictors at each possible split, the fitted values across trees are more independent. Consequently, the gains from averaging over a large number of trees (variance reduction) can be more dramatic.

Bias reduction: a very large number of predictors can be considered, and local feature predictors can play a role in the tree construction.

- Random forests are able to work with a very large number of predictors, even more predictors than there are observations. An obvious gain with random forests is that more information may be brought to reduce bias of fitted values and estimated splits.

- There are often a few predictors that dominate the decision tree fitting process because on the average they consistently perform just a bit better than their competitors. Consequently, many other predictors, which could be useful for very local features of the data, are rarely selected as splitting variables. With random forests computed for a large enough number of trees, each predictor will have at least several opportunities to be the predictor defining a split. In those opportunities, it will have very few competitors. Much of the time a dominant predictor will not be included. Therefore, local feature predictors will have the opportunity to define a split.

IV. METHODOLOGY

Here we are first designing the web panel for the teacher admin and student. in that web panel admin will perform all the major functions.Admin can add the teacher course and the years and subjects. Then he will allow authority to the teacher after registration of the teacher. Also, he will add the time table for the teachers. And he will perform the analysis of the data of comments gathered from the students after the lecture.Teacher can be mark the attendance of the students after the lectures. He will register himself for the using this system. Also, he will add the student's as per the roll number.

Students perform the registration part. he will able to see timetable as per his branch. He is able to mark comments after the marking attendance by the teacher. This comment can be visible as per the date to that student.
Here we are gathering the data for the analysis of the admin on that comment using semantic analysis. Now for the analysis and generation of report and summarize the report we will perform the sematic analysis with natural language processing tools.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 7, Issue 4, April 2018

SYSTEM DESIGN:

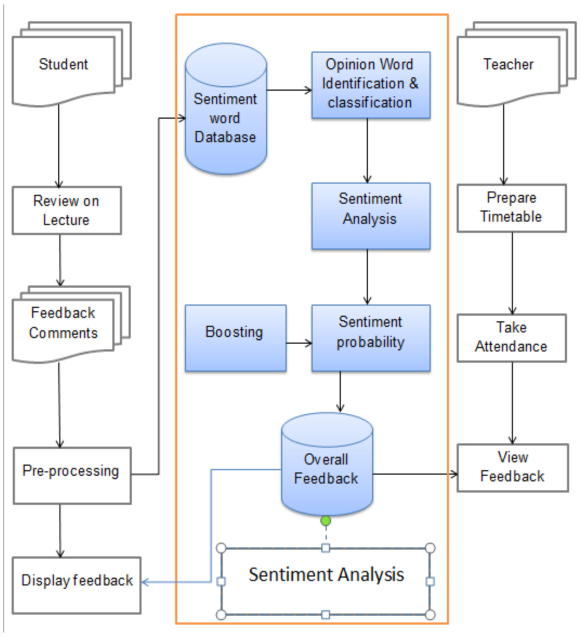

Fig 5.1.1 Detailed Architecture of Analysis

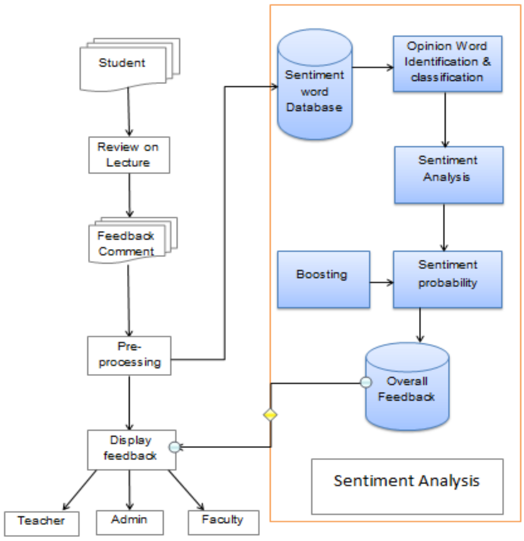

Fig 5.1.2 Overall Architecture of Analysis

V. IMPLEMENTATION

Sentiment Analysis of Twitter Data

Step 1: Gather Tweets
Step 2: Perform Sentiment Analysis on Tweets
Maximum Entropy Modeling Algorithm. Basic Idea is to calculate the probability distribution: p(a, b).Choose p with maximum entropy (or "uncertainty") suject to the constraints (or "evidence").A mathematical measure of the uniformity of a conditional distribution p(y|x) is provided by the conditional entropy

Sentiment Classification

Step 1: Generate Scores for Each Document. Let's say you have a 100-word blog post with the word "JavaScript" in it 5 times. The calculation for the Term Frequency would be:

TF = 5/100 = 0.05

Next, assume your entire collection of blog posts has 10,000 documents and the word "JavaScript" appears at least once in 100 of these. The Inverse Document Frequency calculation would look like this:

IDF = log(10,000/100) = 2

To calculate the TF-IDF, we multiply the previous two values. This gives us the final score:

TF-IDF = 0.05 * 2 = 0.1

Step 2: Decide a Threshold to Tag. After running this algorithm against all 100 of the blog posts with the word "JavaScript", you end up with a score for each. This is where you will have a chance to exercise the creative aspect of being a data scientist. Let's assume that you have a wide range of scores, ranging from 0.05 to 0.5. Continuing a simple example from this collection, 0.05 would be a 100-word document with 1 instance of "JavaScript" and 0.5 would be a 100-word document with "JavaScript" appearing 25 times. To determine if the document will be tagged with "JavaScript", you need to decide on a threshold score. The score you choose will vary depending on your data set. A document with only one instance of "JavaScript" (score 0.05) is unlikely to be focused on JavaScript, but obviously the high score of 0.5 is probably on topic.

VI. CONCLUSION

The system takes the students' comments and analyzes it. Also, it helps in analysis of student's attendance. This will make easy to calculate the student's monthly attendance. Also, the admin will be able to check it. And the admin can perform the students comment analysis and find the performance of the teacher based on the student's comment analysis. The semantic analysis makes this analysis of student's comment make easy and will help to reduce the admin work to gather the comments and read it and analysis of teacher. This will help to reduce the lots of paper and hard work and make the college system efficient and faster.

REFERENCES

[1] Sentiment Analysis of Students' Comment Using Lexicon Based Approach, Khin Zezawar Aung University of Computer Studies,MandalayMandalay,Myanmar,zezawar12@gmail.comUniversity of Computer Studies, Mandalay Mandalay, Myanmar zezawar12@gmail.com

[2] Nyein Nyein Myo University of Computer Studies, Mandalay Mandalay, Myanmar vicky.mdy@gmail.com

[3] M. Michael, :Sentiment And Semantic Analysis (2014), accessed 2 May 2015, http://www.digital-mr.comlbloglview/sentiment-&semantic-analysis

[4] A. El-Halees, "Mining opinions in user- generated contents to improve course evaluation," Software Engineering and Computer Systems, pp. 107115, 2011.

[5] A. Ortigosa, J.M. Martín and R.M. Carro, "Sentiment analysis in Facebook and its application to e-learning," Computers in Human Behavior, 31(1), pp. 527-541, 2014.

[6] A. Rashid, "Feature level opinion mining of educational student feedback data using sequential pattern mining and association rule mining," International Journal of Computer Applications (0975-8887). 81 (10), pp. 31-38, 2013.

[7] A. Sagum, Jessiree Grace M. de Vera, Prince John S. Lansang, Danica Sharon R. Narciso and Jahren K. Respeto, "Application of Language Modelling in Sentiment Analysis for Faculty Comment Evaluation," Proceedings of the International Multi Conference of Engineers and Computer Scientists, vol.1, March 2015.

[8] AFINN word database an affective lexicon by Finn Årup Nielsen, http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/6010/zip/imm6010.zip.

[9] B. 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). 3 (3), pp. 229-232, 2014.

[10] B. Liu, "Sentiment analysis and subjectivity," Handbook of natural language processing, pp. 627-666, 2010.

[11]] C. Pong-inwong and W. Rungworawut, "Teaching Evaluation Using Data Mining on Moodle LMS Forum" , 6th International Conference on New Trends in Information Science, Service Science and Data Mining (Issdm2012), Taipei: IEEE, 2012, pp. 550–555. [9] C. Pong-inwong, & W. Songpan (Rungworawut), "TeachingSenti-Lexicon for Automated Sentiment Polarity Definition in Teaching Evaluation," 10th International Conference on Semantics, Knowledge and Grids (SKG), Beijing: IEEE, pp. 84-91, 2014. [12] L. Caiqiang, and M. Junming, "Research on online education teacher evaluation model based on opinion mining," National Conference on Information Technology and Computer Science (CITCS), 2012.

[12] L. V. Avanco, M.G.V.Nunes, "Lexicon-based Sentiment Analysis for Reviews of Products in Brazillian Portugues," 2014 IEEE Brazillian Conference on Intelligent Systems, 2014, pp. 277-281, 2014.

[13] M. Taboada, J. Brooke, M. Tofiloski, K. Voll and M. Stede, "Lexiconbased methods for sentiment analysis," Association for Computational Linguistics, vol. 37, Issue 2, pp. 267-307, June 2011.

[14] N. Altrabsheh, M. Cocea, S. Fallahkhair, "Sentiment analysis: towards a tool for analyzing real-time students