ABSTRACT

In education system, student's feedback is important to measure the quality of teaching. Students' feedback can be analyzed using lexicon-based approach to identify the student's positive or negative attitude. The main objective of this research is to analysis the Students feedback and obtains the opinion. In most of the existing teaching evaluation system, the intensifier words and blind negation words are not considered. The level of opinion result isn't displayed-whether positive or negative opinion. To address this problem, we propose to analyze the student's text feedback automatically using lexicon based approach to predict the level of teaching performance**.** The opinion result is represented as whether strongly positive, moderately positive, weakly positive, strongly negative, moderately negative, weakly negative or neutral.

Key words: Opinion mining, Sentiment Analysis, Teaching Evaluation, Lexicon based approach, corpus based approach, Dictionary based approach, Qualitative information, quantitative information, Afinn lexicon.

CONTENTS

LIST OF FIGURES

CHAPTER-1: INTRODUCTION

Sentimental analysis is a method for identifying the sentiment expressed in texts. The need of Sentiment Analysis of text has gained more importance in today's situations faced by the people of the world. Generally, there are three approaches in sentimental analysis. They are lexicon based, machine learning and hybrid approach. In machine learning technique, it uses unsupervised learning or supervised learning. Classification problem can be carried out using several algorithms like support vector machine, naive bayes, random forest. In lexicon based method sentiment polarity of the textual content is detected using sentiment lexicon. A lexicon is a list of words with associated sentiment polarity Sentiment analysis is a process for tracking the mood of the people about any particular topic by review. In general, opinion may be the result of people's personal feelings, beliefs, sentiments and desires etc. This research work focus on students comments Analyzing students comments using sentiment analysis approaches can classify the students positive or negative feelings. Student's feedback can highlight various issues students may have with a lecture. Sometimes students do not understand what the lecturer is trying to explain, thus by providing feedbacks, students can indicate this to the lecturer. The Input we take is qualitative data rather than quantitative data. The processing of qualitative data analysis is very important and it can enhance the teacher evaluation effectiveness. Evaluating performance of faculty members is becoming an essential component of an education management system. It not only helps in improving the course contents and quality but is also often used during the annual appraisal process of faculty members. The evaluation is typically collected at the end of each course on a set of question which are answered. The evaluation form, however, also provides room for open feedback which typically is not included in the performance evaluation/appraisal due to lack of automated text analytics methods. The textual data may contain useful insight about subject knowledge of the teacher, regularity, and presentation skills and may also provide suggestions to improve the teaching of quality. Students provide feedback in quantitative ratings and qualitative comments related to preparation, contents, delivery methods, punctual, skills, appreciation, and learning experience. The delivery methods and preparation component refers

to instructor's interaction, delivery style, ability to motivate students, out of class support, etc. The content refers to course details such as concepts, lecture notes, labs, exams, projects, etc. The preparation refers to student's learning experience such as understanding concepts, developing skills, applying acquired skills, etc. The paper correction refers to correction of mistakes and providing solutions to overcome it. The punctual refers to the class timing and assignment or record submission. The appreciation refers to the comments given when something is done perfectly. Analyzing and evaluating this qualitative data helps us to make better sense of student feedback on instruction and curriculum. Recent methods for analyzing student course evaluations are manual and it mainly focuses on the quantitative feedback. It does not support for deeper analysis. This paper focus on providing qualitative and quantitative feedback to analyze and provide better teaching to improve the student's performance.

1.1 Supervised Methods

Supervised Learning is one of the types of Sentiment Analysis. Supervised learning as the name indicates the presence of a supervisor as a teacher. Basically supervised learning is a learning in which we teach or train the machine using data which is well labeled that means some data is already tagged with the correct answer. After that, the machine is provided with a new set of examples (data) so that supervised learning algorithm analyses the training data(set of training examples) and produces a correct outcome from labeled data. The increasing availability of labelled data has played an important role in the application of supervised machine learning methods to sentiment analysis. These methods represent the labelled data in the form of a set of features. The features are then used to learn a function for classification of unseen data. In this dissertation, I approach the problem of sentiment analysis as a classification task.

Supervised learning can also be performed using multiple classifiers, particularly if the labeling scheme allows for hierarchical relations. As described earlier, one example of this is the work by Pang and Lee. They represented sentences in the given document as graph nodes and calculated the minimal cut on that graph to identify the subjective sentences. Afterwards, standard machine learning classification algorithms (NB and SVMs) were applied only on the extracted subjective sentences to predict their polarity. On a balanced polarity corpus of 2,000

reviews, the minimum-cut framework resulted in an accuracy of 86.4%, which represents a statistically significant improvement in polarity classification accuracy over previous attempts.

Supervised learning with dependency trees was also used by Joshi and Penstein-Rose (2009), who worked on solving the problem of identifying opinions from product reviews. Their method was to transform syntactic dependency relation triplets into features for classification. The motivation was to capture the general relationship between opinionated phrases by 'backing off' to the head word in the triplet. For instance, consider the phrases a great camera and a great mp3 player with the relations {amod, camera, great} and {amod,player,great}. Here, backing off the head words (camera and player) to their POS tags results in a more generalised form {amod,NN,great}, which makes a better indicator for opinion extraction. A collection of 2,200 reviews from the extended version of the Amazon.com/CNet.com product review corpus3 was used, 1,053 of which were subjective. With 11-fold cross-validation in an SVM learner, their method of backing off the head word to the POS achieved approximately 68% accuracy. For obtaining a balanced corpus, 700 documents of each label were selected. N-grams, part-of-speech (POS) tags and their combinations were used as features, and three-fold cross validation was used. Their best system achieved an accuracy of 82.9% when using the unigram presence feature set with SVMs. It should be noted that the corpus used in this work was balanced artificially, thus avoiding the problem of data sparsity for under-represented classes.

1.2 UnSupervised Methods

Next one, Unsupervised Learning, It is the training of machine using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance. Here the task of machine is to group unsorted information according to similarities, patterns and differences without any prior training of data.

Unlike supervised learning, no teacher is provided that means no training will be given to the machine. Therefore, machine is restricted to find the hidden structure in unlabeled data by our-self.

A similar approach was taken by Taboada and Grieve (2004), who used information about the position of the text in a document to expand the lexicon. They assigned more weight to certain parts of the text where they believed most subjective content was concentrated. Using the AltaVista search engine, this method achieved an overall accuracy of 65%.

Another way to increase the number of items in the lexicon is to use bootstrapping. The general approach is to start with a limited number of polar phrases in the lexicon and to extract similar phrases from unlabelled data. These extracted phrases are then added to the polar lexicon and the process is repeated until a stopping criterion is reached. Riloff and Wiebe's (2003) bootstrapping algorithm learns linguistically rich patterns for subjective expressions. For example, the pattern <subj> was satisfied will match all sentences with the passive form of the verb satisfied. High precision classifiers trained on known subjective vocabulary were used to automatically identify objective and subjective sentences in unannotated text. The labelled sentences from these classifiers were fed to an extraction pattern learner. The remaining unlabelled sentences were filtered through a separate pattern-based subjective sentence classifier that used the extraction pattern previously learned. To close the bootstrap loop, the output of the pattern-based classifier was returned back to the extraction pattern learner and the extracted patterns were used again in the initial high-precision classifiers. This system was used on a balanced corpus of roughly 34,000 sentences and achieved a precision of 0.902 and a recall of 0.401.

Exploring more lexical features in a later work, Wiebe and Riloff (2005) developed a Naive Bayes (NB) classifier using data extracted by a pattern learner. This pattern learner was seeded with known subjective data. Additional features for this NB classifier included strong and weak subjective clues from a pre-existing rule-based system, POS tags, pronouns, modals, adjectives, cardinal number data and adjectives. The classifier was used to classify the unlabelled text corpus, and the most confidently classified sentences were added to the training set for another cycle. They trained the system for two cycles. Using test corpus of 9,289 sentences, 5,104 of which were subjective, they reported up to 0.863 subjective recalls with a subjective precision of 0.713 (corresponding to 73.4% accuracy).

practice, algorithms formally designed for transduction or induction are often used interchangeably.

1.4 Problem Statement

This research shows the use of sentiment analysis to evaluate the student's narrative comment in the evaluation of their respective Faculty. The rest of the project is organized as follows: One Section presents a review of literature; proposed methodology is presented in another section. It also describes the construction of sentiment word database for teaching evaluation and also presents the architecture of our proposed system. It also presents the case study and results of the teaching evaluation and the final chapter presents the conclusion.

- R Mehana from Dr.Mahalingam school of Engineering and Technology Pollachi, Tamlinadu, India[6] developed Student feedback mining system adopting sentiment analysis in 2017.They projected a system to mine the feedback given by the students and acquire information from that and gift that info in qualitative method. they have known the frequency of each word and extract the topic that has the perfect frequency count. Similar comments in every topic are clustered then the clustered words are classified into positive or negative comments.

- S. MacKim and R. A. Calvo projected Sentiment analysis in student experiences of learning[7] in 2016.they Classify the text supported the presence of unambiguous have an effect on words. In their approach, a bit set of opinion words is collected manually as a seed. They have a sentiment lexicon contains a listing of words aboard their individual polarity. Many like corpora are developed and that they created freely out there.

- Other interesting approach of Sentiment Analysis was presented in the work of M. A. Ullah [5] wherein they extract sentiments with polarities of positive and negative for specific subjects from a document, instead of classifying if the document is positive or negative. In this paper, they applied semantic analysis with a syntactic parser and semantic lexicon which gave them a high precision of 75% to 95% in finding the sentiments within web pages and news articles.

- LI Caiqiang and Ma Junming [10] perform online education teacher evaluation model based on opinion mining. They collect the students' comments written in Chinese language published in the LMS (learning management system) by using web crawler. They didn't indicate the words polarity strength in their work. If sentiment words of subjective text are not in the polarity word dictionary, they used point wise mutual information (PMI) method to judge the polarity of it. This model gets an overall evaluation of each teacher. They collect the students' comments written in Chinese language published in the LMS (learning management system) by using web crawler. They didn't indicate the words polarity strength in their work. If sentiment words of subjective text are not in the polarity word dictionary, they used point wise mutual information (PMI) method to judge the polarity of it. This model gets an overall evaluation of each teacher.