SVM and HMM Based Hybrid Approach of Sentiment Analysis for Teacher Feedback Assessment

Balaji Jagtap¹ and Virendrakumar Dhotre²

¹Research Scholar, Dept. of Information Technology, Maharashtra Institute of Technology, Pune
²Assistant Professor, Dept. of Information Technology, Maharashtra Institute of Technology, Pune

Abstract: Sentiment analysis view from broad range of field of natural language processing and opinion mining. Most of the researcher focus on the product and services. This paper design to automate teacher feedback assessment system. As know, teacher is important part of education. Therefore progress and performance monitoring of teacher is also important factor of education system. It could be measure by taking feedback from student for particular teacher. Data collect from student is larger in size to concluding result is difficult task. Here, sentiment analysis play the role with help of HMM and SVM base hybrid sentiment classification model with advent feature extraction method.

Keywords: Sentiment Analysis, HMM, SVM, Feature Extraction.

1. INTRODUCTION
The advent of digital world, noticeable change in lifestyle of people and decision making has been observed. Meanwhile, growth in opinion resource like review sites, online forums and web blogs, social networking, together called as user generated data. This data support to track customer decision and predictable to analyze the user view. However, this is laborious or impossible task for analyzer or reviewer to read such enormous data and derive the conclusion of data. The intension of sentiment analysis to analyze such large amount data, whereby review can be classified into positive and negative category by analyzing the content expressed in the review. Thus the automatic classification make task easier and make way for mine the valuable knowledge from the reviews.

Another application of sentiment analysis can also apply to manage user generated content. Many websites those work for news group, blogs and other online place where spontaneous discussion occurs from this large opinion data is generated. Worthwhile, the generated textual data is unstructured and without any score attachment or allocation. Here, Sentiment analysis play crucial role to classify the data into positive and negative category according to opinion express in reviews. This process is widely studied and applied to user who do not explicitly express their sentiment orientation in particular context. So the sentiment analysis is necessary to analyze such large unstructured data [12].

Noteworthy, sentiment analysis also called as opinion mining which is study of human behavior in which we extract user emotions and opinion from textual data. In business intelligence sentiment analysis play crucial role, for instance, it help industry people to analyze their product review given by customer to enhance the product quality, product services and sometimes new opportunities in business. In healthcare it has been incorporate to understand the user need [1]. In politics, it help public to identify their right politician by analyzing comments. In daily life user review about particular product gives more information while shopping particular product. This will give us better decision while buying the product. Regarding teacher feedback evaluation, sentiment analysis is applied to classify emotion and opinion about particular teacher expressed by student in their feedback. Sentiment analysis is applied to classify the feedback data into positive and negative category to evaluation of overall performance of teacher in given academic period. Student express real character of teacher in their feedback secretly. Student give feedback on different qualitative attribute of teacher, for instance, subject knowledge helpfulness, regularity, presentation skill etc. which can be positive or negative. Indirectly, student gives suggestion and improvement in various teaching attribute for teacher by giving positive and negative feedbacks.

Every institute always care about their quality of education. Teacher is crucial part of education system which should be always good performer. Worthwhile, assessing the teacher performance is also important to minimize the risk of quality education [2] [3] [4]. Considering institution feedback assessment, the count of data is collected in multiplication of number of teachers, number of courses, number subject taken by particular teacher and number of students are much larger than thousands. The evaluation of such large data is tedious and even difficult task. Therefore, it is necessary to automate the system to improve teacher feedback evaluation. Hence, sentiment analysis come to perform automated system for feedback evaluation of student opinion about teacher express in the feedback. This approach not only make system automated but also improve accuracy of result based on above stated motivation. This is not only assess the teacher feedback but also provide suggestion to teacher for particular weakness and good point to awarded teacher for their better performance.

This paper design to propose feedback assessment of teacher using sentiment analysis by classifying feedback into positive and negative classes. Here, SVM and HMM
based Hybrid approach of classification of student feedback into positive and negative class with statistical method for feature extraction.

2. RELATED WORK:
As of interest, sentiment analysis has been studied in broad range of domain since last decade. There are different domain interestingly studied by many researcher such as movie review, product review, travel destination review, social networking, e-learning and many more. Most of the researcher attracted toward training machine learning classifier algorithm for opinion classification. The study of pang [5] have come up with a relevant work based on classic topic classification technique. Pang’s [5] proposed work show comparative study of different machine learning algorithm. Above work is intended to test whether a selected group of machine learning algorithm performance over categorization of review into positive and negative category. Pang’s [5] perform experiment on movie review data set use Naïve Byes, Support Vector Machine, Maximum Entropy. The performed experiment shown result accuracy between 71 to 85% depending on dataset and algorithm. Likely, E. Boiy [6] applied same classification techniques to classify the movie review and car blogs. E. Boiy [6] clears that maximum accuracy is achieved by Maximum Entropy with additional characteristic of unigram and subjectivity and also proved that Naïve Bayes is fastest classification algorithm than other machine learning algorithm, S. Dasgupta [7] Determine the polarity of movie review into positive and negative basis by using characteristic of unigram, bigram and trigram. It has find out that integration of bigrams and trigrams improve accuracy even more. H. Cui [8] applied different machine learning algorithm on large online product review data. The experiment concluded that combination of n-gram with machine learning algorithm increase the precision even more.

However, R. Prabowo [9] applied the combination of statistical methods for instance Mutual information, Information Gain, CHI with supervised learning support vector machine. This methods are tested on movie review dataset and show that hybrid methodology improve the accuracy more. Z. Zhang [11] applied different machine learning algorithm of support vector machine, Naïve Bayes and N-gram model to online review of travel destination agency. Jamshid beheshti [13] use supervised approach of HMM model for text categorization and shows that it surpasses the Naïve Bayes text categorization model. It determined that SVM model and characteristic based N-gram model achieved better performance that other methods. J. Kittler [14] and Josef Kittler [15] studied different combination rules and show that classifier combination method perform well. Emma Haddia [16] shows significant improvement by using proper preprocessing methods.

This paper is design by introducing the need of sentiment analysis in different domain of business, political, social networking and educational domain. As matter of automation sentiment analysis applied in context of teacher feedback assessment to classify the student feedback into positive and negative category.

3. PROPOSED METHODOLOGY:
In primary stage, data is collected from student is unstructured format. It contain noise and unwanted text which does not contribute any kind of sentiment. Also, it may divert sentiment analysis process. Therefore, preprocessing the data is necessary before the sentiment classification. Preprocessing deals with following task which will give proper data for classification of sentiment.

3.1 Preprocessing:
Stop word: Mainly stop word are grammatical word, conjunctions and phrases which does not contribute any sentiment value, for instance, is, are, have, also, many, ever, etc. In English language, it contains more than 600 stop words.
Stemming: same words are present in different form such as adjective, verb and adverb which may lead to different meaning for instance good, best, better, engineering, and engineer. It is process of removal of prefix and postfix of word and convert the word to its root word.
Tokenization: it is process of converting sentence into chain word so processing word by word can easily perform.

Further, preprocess data is applied to sentiment classification process. Here, Sentiment analysis aim to classify the student feedback into positive and negative classes by analyzing sentiment orientation of feedback content. Sentiment analysis based on machine learning approaches have large corpus size. Due to large corpus size machine learning task of classification may over fit the classifier and increase unnecessary computation or some time divert classification task. Henceforth, selection of required number of features is necessary to avoid the over fit problem. Feature selection task is to reduce the corpus size my removing irrelevance feature from the corpus. This lead to selection of suitable Feature selection method which give proper corpus size and relevance features. Proper selection of feature selection method able to determine more relevant feature about particular domain. Here, our aim is to apply new extraction method which describe as follow.

3.2 Feature Extraction:
Following formula calculate the weight of features which give more accuracy than other feature extraction methods. This formula work in principle of total term frequency in particular document with respect to number of document fall into particular category. This formula assign higher weight to attribute with less document occurrence. This method weighing attribute for two class classification which always seen in sentiment analysis.

\[
f(E,A) = \frac{\log(N_0 + 1) \frac{N_0}{N_0 + N_1}}{N_1 + 1} = \frac{\log(N_0 + 1) \frac{N_0}{N_0 + N_1}}{N_1 + 1}^{-1}
\]  

(1)
Where, \( N \) is total number of document in particular class and \( S \) number of document occurred particular term in particular category. If there are equal number of document in both the category then given formula (1) can be modified as follows.

\[
f(x, A) = \frac{\ln (S_x + 1)}{\ln (N + 1)}
\]

Above formula show that relation between weight of term and number of documents where this term appear in particular class. Afterword, this data is store according to the principle of vector space model for the convenience of further processing.

System framework is depicted in figure 1 which is explore in further discussion. Feature vector data is pass to the combination of SVM and HMM based model which will take care of sentiment classification of data. SVM and HMM individually perform the text classification on given data. The result derive from each classifier is combine by classifier combination rule.

![Figure 1: Sentiment classification framework](image)

### 3.3 Support Vector Machine (SVM)

As above mention, data obtain from feature extraction process is further applied to the sentiment classification. SVM is still date best classifier base on machine learning which have accuracy more than 90% for linear data classification.

SVM is mathematical model which plot the data set value into infinite multidimensional space. In this space it draw hyper plane which classify the data set into proper classes. When given dataset not in linear dimension then it maps to high dimension space. Due to that we get better optimal plane which gives better result and also classify data with less classification error. Firstly, SVM train with predefined label data through which it seek for hyper plane which classify data into two categories. This hyper plane is represented by notation \( \mathbf{w} \) that classify the data into negative and positive class with maximum margin to get optimal result with less classification error.

Finding such hyper plane problem solution belong to constraint optimization problem. Consider, \( y_i \) is label variable of value -1 and +1 and \( d_i \) is document index. The formula for SVM can be written as,

\[
\mathbf{w} = \sum_{i=1}^{N} \alpha_i y_i d_i
\]

(3)

Where, resultant solution of dual optimization problem \( \mathbf{w} \). Hyper plane \( \mathbf{w} \) constructed as combination of document \( d_i \) vector. This is applied with feature extraction method for proper classification.

#### 3.4 Hidden Markova Model:

This is related to automata theory which deals with transitions, states and time. Similarly, HMM is like probabilistic automata. Probabilistic automata which also deals with structure of states and transitions. Here, HMM is applied for text classification. HMM for sentiment text classification consist of some parameter which are describe as follows. It has mainly four measures represented in vector observable form: I initial probabilities, E output symbol emission probabilities, T states transition probabilities, O set of output symbols and \( S \) set of states. So vector representation of HMM.

\[ M = (I, E, T, O, S) \]

There are two dummy states ‘Start’ and ‘End’, emission probability associated with current state, current state only depend of previous state instead of all previous methods.

### 3.5 Classifier Combination Rules:

As above mention our two classifier, one is SVM and HMM are combine with classification rule as describe in following section. There are many possible method for combination of two different classifier. Consider, text classification problem where text is assign to one of the possible class label \( L_k \) \((L_k = f_1, f_2, \ldots, f_m)\). Let assume that we have \( N \) classifiers denoted as \( C_k \) \((k=1, 2, \ldots, N)\). Input sample \( X_k \) is assign to every classifier \( C_k \) and output measure in form of posterior probability vector represent as follows.

\[ p_k = [ p(f_1|X_k), \ldots, p(f_m|X_k) ] \]

Where, \( p(f_j|X_k) \) denote the probability of classifier when \( X \) labeled as \( f_j \). Here, consider \( j \) type of label is assign to \( z \) test data in given combination rules.

**Sum Rule:**

Label \( f_j \) is assign to \( z \) when following condition occur.

\[ \text{assign } z = \arg \max_j \sum_k p(f_j|X_k) \]

**Major Voting Rule:**

Shortly, each classifier directly assign label to test data then major voting rule assign the final label to test data by selecting major vote given by all the classifier for that test data. When different label got the same number of maximum counts, then randomly class label is selected among them. The mathematical formula for major voting rule is given as follow.
assign $Z \rightarrow j_k$

$$f = \max_{j_k} \sum_{k=1}^{K} A_{jk} P(j_k | x_i)$$

Where,

$$L_k = \frac{f_k}{L_k}$$

Max Rule:
Max rule is applied over information provided by probability of $P_j(f_k | x_i)$. Max rule is always winner in major voting class. It is expressed by following function.

assign $Z \rightarrow j_k$

$$f = \arg \max_{j_k} \{ \max_{j_k} P(j_k | x_i) \}$$

Min and mean rule also use for combination of two classifier. Min rule can be derive by the last result of max rule. Mean rule can be obtain by max rule also by taking mean of probability function.

Based on this classifier combination rules HMM and SVM model greatly outperform. Sum rule give the more accuracy than any other combination rule.

4. CONCLUSION:
Applying sentiment analysis to evaluation of teacher feedback turn out important fact. Such analysis helps to provide better understanding of student opinion about teacher which also important to improve the quality of education.

This study has demonstrate the effective method for sentiment analysis for teacher feedback assessment. On applying advance feature selection method with hybrid approach of sentiment classification. We have come to conclusion that hybrid approach work well for complex data. One of the challenge in student feedback which is complexly written and harder to find the sentiments.

References