Volume 6, Issue 1, January 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Automated Course Feedback System Using Supervised and Unsupervised Machine Learning Approach

Sanjay Wankhade
VES Polytechnic Chembur, Mumbai, Maharashtra, India

Dr. Shalu Chopra
HOD (Info. Tech) VESIT Chembur Mumbai University, Maharashtra, India

Abstract: This paper presents an online automated course feedback system that collect data from student who appeared for a course in both structured and unstructured manner with anonymous approach ,to gain insight on academic course of Academic institution to improve institution or organization for their stake holders. This experimental work on designing an automated course feedback system. The automated course feedback system collect students responses as a set of binary and graded responses as well as unstructured responses are collected using text area field on which sentiment analysis or opinion mining are performed to compute positive or negative opinion about course feedback. The final feedback of course is overall score of computed value from binary and graded response and positive and negative opinion obtained from unstructured approach i.e. feedback collected from text area by performing sentiment analysis on sentences written in text area.We evaluated the experimental work on 10 courses with 200 review of each.

Keywords: Opinion mining, sentiment analysis, automated system, anonymous approach.

I. INTRODUCTION

Opinion mining also called as sentiment analysis has become area of exploration by linguistics and machine learning analyst. Lot of work has been done on opinion mining from used posted comment or data on web. Various experiment has been performed to identify opinion with product review ,movie review or post of social media, and prospects of candidate in election or political issue .Still much of the progress is remaining in the field of information retrieval, computational and machine learning linguistics and statistical and probabilistic text processing.

The goal of sentiment analysis or opinion mining is to mine the opinion and classify them as a positive or negative. Basically two types of approaches have been used by researchers in opinion mining

a) Supervised i.e. machine learning approach to classify each document in positive or negative category.
b) Unsupervised approach : to compute semantic orientation of text or post .Unsupervised approach use principle of Natural Language processing (NLP)technique such as POS (part of speech) tagging in documents which can be labeled as positive or negative.

Although Natural language processing has long history, opinion mining using NLP has become active research area .There are several reason for this because it has wide range of application almost in every area or domain. Nowadays we have seen that opinioned posting on social media has helped to build and reshaped the business. It has become important to collect and study opinion on the web .Customer feedback collected from mail or call center data or result from surveys conducted by organization can be used for opinion mining.

Similarly marketing and advertising agency may used opinion mining to decide about web pages where they should place their advertisement and where to marketwise the product.

Apart from real life application, many application oriented research paper have been published[2](Lie et al 2007a sentiments model was proposed to predict sales performance).Review were ranked for product and merchants. [2](McLuhan, Glance and Reiter, 2010). Twitter sentiment was linked with public opinion [3] (Timespan et al 2010) Twitter data ,movie review and blogs were used to predict box office revenue for movies [4] (Azure and Hubrman ,Joshi et al 2010) Some case studies also reported [1] (castellans et al 2011)

In General opinion mining can be carried out mainly at three level

i) Document level : the task at this level is to classify whether whole document express positive or negative opinion.
ii) Sentence level : The task at this level is to identify sentence polarity such as negative positive ot neutral.
iii) Entity or aspect level : Aspect level performed detail of fine grained analysis .It is also called as feature level.

In this experimental work we have used opinion mining to mine sentiments of student from course review.

We have developed and interactive automated feedback system where student may record his feedback in anonymous manner with binary and graded question .Additionally students are also encourage to write their overall experience about course and course parameter in text area provided to write their opinion about eighty words .

Responses entered by student are stored in database with their random handler which will offer them anonymity about course feedback. The responses are then used to answer some queries and generate various statistical report and comparisons.
Additionally data stored in in database collected as untrusted data have been processed by our custom sentiment analyzers with outcome as positive or negative opinion labeled based on semantics orientation of feedback. Semantic orientation is calculated by SentiwordNet[11]using natural language technique .We have developed system using C#.NET and Sql server as a backend.
We have carried out experimental work with good number of course at our Instituites with average course feedback of 60 for each course, Total number of feedback are more than 600. The key concept used in this experimental work is opinion mining.
The goal of opinion mining is mine the opinion from text and classify that opinion as positive or negative opinion.The feedback labeled as positive mean student liked the course and negative means student are not satisfied from course or not happy with course and its constituents or not having the good experience about the course.

II. PROBLEM DEFINITION

The opinion mining can be defined as ,
Given set of docuemnts D, sentiment classifier classifies each document d ∈ D into one of the classes,positive or negative .Positive means that d expresses a positive opinion and negative means that d expresses negative opinion The various approaches are used for opinion mining in text as movie review and political campaigning and product ration .
These approaches are classified in two ways such as Supervised approach and unsupervised approach.Supervised approach takes machine learning approach to categorized document in positive and negative .In Unsupervised approach we compute semnatic orientation of document using POS tagging. The earlier research work can be found in [5][6][7][8[9][10][11].In our study we confined with document level classification .
We have implemented two supervised approach such as Naïve Bayes Machine Learining approach and Support Vector Machine approach and for unsuperivsed we c arried out sentiment analysis by semantic orientation with Pointwise Mutual Information (PMI).
Naïve Bayes Machine Learning Approach.
1) Naive Bayes Machine learning Approach.

It is an approach for text document classification that assigns the class c* = argmaxc P(c | d), to a given document d. A naive Bayes classifier is a simple probabilistic classifier based on Bayes' theorem and is particularly suited when the dimensionality of the inputs are high. Its underlying probability model can be described as an "independent feature model". The Naive Bayes (NB) classifier uses the Bayes' rule Eq.1

$$p(c|d) = \frac{P(c)P(d|c)}{P(d)} \qquad \text{Eq.-----------------(1)}$$

Where, P (d) plays no role in selecting c*. To estimate the term P(d|c), Naive Bayes decomposes it by assuming the fi's are conditionally independent given d's class as in Eq.(2)

$$P_{NB}(c|d) = \frac{P(C)(\prod_n^m (P(fi|c)^{xi(d)})}{P(d)} \qquad \text{Eq ---------------- (2)}$$

Where, m is the no of features and fi is the feature vector. Consider a training method consisting of a relative-frequency estimation P(c) and P (fi | c). Despite its simplicity and the fact that its conditional independence assumption clearly does not hold in real-world situations, Naive Bayes-based text categorization still tends to perform surprisingly well; indeed, Naive Bayes is optimal for certain problem classes with highly dependent features.
2) Support Vector Machines as a supervised Approach.

Support vector machines (SVMs) have been shown to be highly effective at traditional text categorization, generally outperforming Naive Bayes. They are large-margin, rather than probabilistic, classifiers, in contrast to Naive Bayes and Maximum Entropy. In the two-category case, the basic idea behind the training procedure is to find a maximum margin hyper plane, represented by vector w, that not only separates the document vectors in one class from those in the other, but for which the separation, or margin, is as large as possible. This corresponds to a constrained optimization problem; letting cj Ɛ {1, −1} (corresponding to positive and negative) be the correct class of document dj, the solution can be written as in Eq. (3)

$$\backslash\vec{w} := \sum_j a_{j\, c_j \vec{d} j = a_{j \geq 0}} \text{-} \qquad \text{Eq.----------------------(3)}$$

Where, the αj's are obtained by solving a dual optimization problem. That dj such that αj is greater than zero are called support vectors, since they are the only document vectors contributing to w. Classification of test instances consists simply of determining which side of w's hyper plane they fall on Support vector machines were introduced in [3] (Vapnik) and basically attempt to find the best possible surface to separate positive and negative training samples. Support Vector Machines (SVMs) are supervised learning methods used for classification. , SVM is used for sentiment classification. Support vector machines perform sentiment classification task on review data. The goal of a Support Vector Machine (SVM) classifier is to find a linear hyper plane (decision boundary) that separates the data in such a way that the margin is maximized. Look at a two class separation problem in two dimensions like the one illustrated in figure 1, observe that there are many possible boundary lines to separate the two classes. Each boundary has an associated margin. The rationale behind SVM's is that if we choose the one that maximizes the margin we are less likely to misclassify unknown items in the future

Fig 3. Shows Different Boundry decision of SVM..

Fig (3) Boundry decision of SVM

3) Semantic orientation approach Using Unsupervised Learning

Since sentiment words are often the dominating factor for sentiment classification, it is not hard to imagine that sentiment words and phrases maybe used for sentiment classification in an unsupervised manner. The method in (Turney, 2002) is such a technique. It performs classification based on some fixed syntactic patterns that are likely to be used to express opinions. The syntactic patterns are composed based on part-of-speech (POS) tags. The algorithm given in (Turney, 2002) consists of three steps:

Step 1:

Two consecutive words are extracted if their POS tags conform to any of the patterns For example, pattern 2 means that two consecutive words are extracted if the first word is an adverb, the second word is an adjective, and the third word (not extracted) is not a noun. As an example, in the sentence “This piano produces beautiful sounds”, “beautiful sounds” is extracted as it satisfies the first pattern. The reason these patterns are used is that JJ, RB, RBR and RBS words often express opinions. The nouns or verbs act as the contexts because in different contexts a JJ, RB, RBR and RBS word may express different sentiments. For example, the adjective (JJ)“unpredictable” may have a negative sentiment in a car review as in “unpredictable steering,” but it could have a positive sentiment in a movie review as in “unpredictable plot.”

Step 2:

It estimates the sentiment orientation (*SO*) of the extracted phrases using the point wise mutual information (PMI) measure

:PMI(term1,term2)=log2((pr(term1^term2)/((pr(term1)pr(term2)) Eq -------(4)

PMI measures the degree of statistical dependence between two terms. Here, Pr(term1 term2) is the actual co-occurrence probability of term1and term2, and Pr(term1)Pr(term2) is the co-occurrence probability of the two terms if they are statistically independent. The sentiment orientation(SO) of a phrase is computed based on its association with the positive reference word “excellent” and the negative reference word “poor”:

SO (phrase) = PMI (phrase, “excellent”) -PMI (phrase, “poor”) Eq. ------- (5)

The probabilities are calculated by issuing queries to a search engine and collecting the number of hits. For each search query, a search engine usually gives the number of relevant documents to the query, which is the number of hits. Thus, by searching the two terms together and separately, the probabilities in Equation (1) can be estimated. In (Turney, 2002). Let hits(query) be the number of hits Equation (5) can be rewritten as:

SO(phrase)=log2((hits(phraseNEAR”excellent”)hits(“poor”)/(hits(phrase(NEAR“poor”) hits(“excellent”)) Eq.---(6)

Step 3:

Given a review, the algorithm computes the average *SO* of all phrases in the review and classifies the review as positive if the average *SO* is positive and negative otherwise.

The following figure shows the method to receive student feedback through web form.

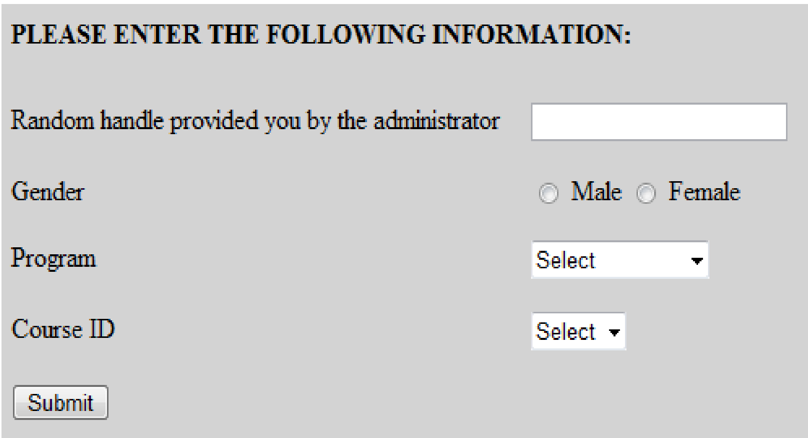
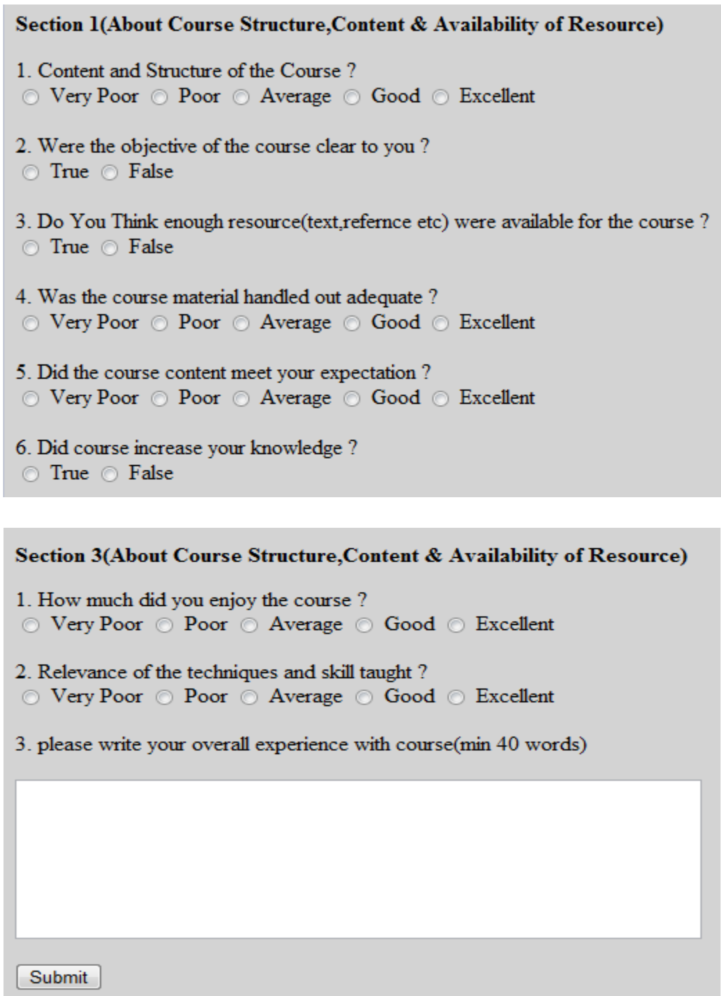


Fig 1.Feedback Form1.



Section 1(About Course Structure,Content & Availability of Resource)

1. Content and Structure of the Course ?
Very Poor Poor Average Good Excellent

2. Were the objective of the course clear to you ?
True False

3. Do You Think enough resource(text,refernce etc) were available for the course ?
True False

4. Was the course material handled out adequate ?
Very Poor Poor Average Good Excellent

5. Did the course content meet your expectation ?
Very Poor Poor Average Good Excellent

6. Did course increase your knowledge ?
True False

Section 3(About Course Structure,Content & Availability of Resource)

1. How much did you enjoy the course ?
Very Poor Poor Average Good Excellent

2. Relevance of the techniques and skill taught ?
Very Poor Poor Average Good Excellent

3. please write your overall experience with course(min 40 words)

Submit

Fig 2. Feedback Form

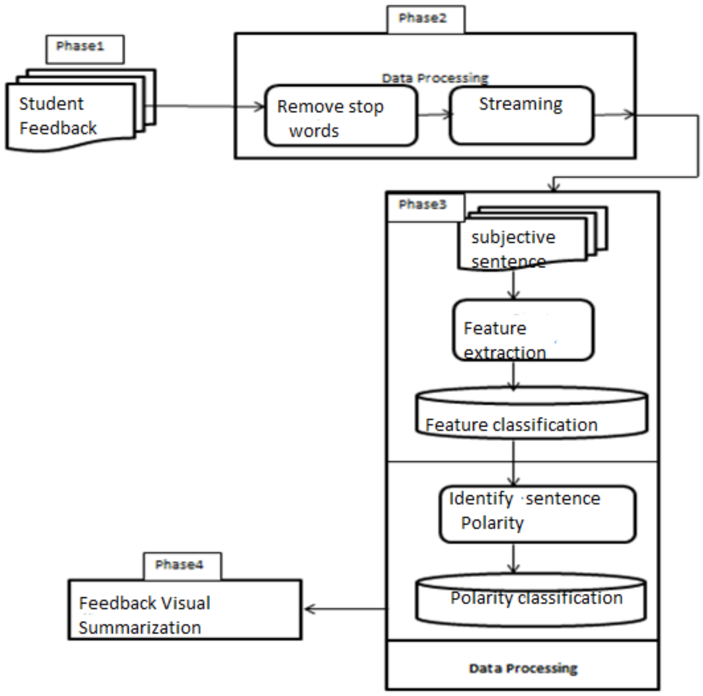


Fig 3: Proposed Architecture for Feedback system

III. RESULT

As per experimental setup we obtained various result from feedback data in two ways. First way use traditional queries to produce important result .we implemented web based form to submit certain queries to database .We obtained various result such as course e not completed in time ,course that require faculty to be changed ,course like by male students, course like by female students. By using machine learning algorithm we obtained individual student opinion and student opinion value has been calculated by using PMI-SO which is mentioned above which calculate positive and negative value of feedback .Sentiment orientation result have been verified by using Naive Bayes and support vector machine learning algorithm.

This automated system provides various statistical result analysis to feedback administrator ,where administrator can check individual anonymous response and free form text opinion value with sentiment orientation and compare feedback of student with supervised and unsupervised approach. This system generate report with sentiment analysis ,Naïve Bayes and SVM in .rdlc format[12]. Administrator can also add new program and new course. and then make the program available for entering the feedback for student on user end.

IV. CONCLUSION

We have implemented and automated student feedback system that works both on tradinional structured responses and employs an opinion mining phase to compute aggregate opinion of students of course .The data for student feedback has been collected through web based form .The form consists of both structured question and free form text box where student can express their overall opinion in 40 words ,because student may feel free to express their opnion rather than just answering the structured question. The result obtained from opion mining are largely accurate and also matches with the result result of traditional database queries. We have successfully implemented and evaluated our system with opinion mining in our institutes .this system explores new idea of using opinion mining and sentiment analysis and result shows the effectiveness of the system.

ACKNOWLEDGEMENT

The authors would like to thank his student for providing their reviews and feedback for course which provides valuable inputs to this system.

REFERENCES

[1] Lu YueMalu Castellanos ,Umeshwar Dalal " Automated construction of context aware sentiments lexicon :an optimized approach inProceeding of 20th international conference on worls wide web 2011

[2] Opinion mining to predict sales performance McGlohon,Glance and Reiter 2010.

[3] Twitter sentiment was linked with public opinion [3](Timespan et al 2010)

[4] K. Dave, S. Lawerence and D. Pennock, "Mining the Peanut Gallery- Opinion Extraction and Semantic Classification of Product Reviews", Proceedings of the 12th International World Wide Web Conference, 2003, pp. 519-528.

[5] P. Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews", Proceedings of ACL-02, 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, US., 2002, pp. 417-424.

[6] A. Esuli and F. Sebastiani, "Determing the Semantic Orientation of Terms through Gloss Analysis", Proceedings of CIKM-05, 14th ACM International Conference on Information and Knowledge Management, Bremen, DE, 2005, pp. 617-624

[7] B. Pang, L. Lee and S. Vaithyanathan, "Thumbs up? Sentiment classification using machine learning techniques", Proceedings of the Conference on Emperical Methods in Natural Language Processing, Philadelphia, US., 2002, pp. 79-86.

[8] S. M. Kim and E. Hovy, "Determining sentiment of opinions", Proceedings of the COLING Conference, Geneva, 2004.

[9] K.T. Durant and M.D. Smith, "Mining Sentiment Classification from Political Web Logs", Proceedings of WEBKDD'06, ACM, 2006

[10] F. Sebastiani, "Machine Learning in Automated Text Categorization", ACM Computing Surveys 34(1), pp. 1-47, 2002.

[11] SentiWordNet, http://patty.isti.cnr.it/~esuli/software/SentiWordNet

[12] Create a Drillthrough (RDLC) Report with Parameters using ReportViewer "https://msdn.microsoft.com/en-us/library/dn154771.asp"

