

IDENTIFYING E-LEARNER'S OPINION USING AUTOMATED SENTIMENT ANALYSIS IN E-LEARNING

P.Bharathisindhu¹, S. Selva Brunda²

¹Assistant professor, Department of computer science, Navarasam arts and Science College for women, Arachalur.

²Professor & Head, Department of CSE, Sasure College Of Engineering, Vijayamangalam

Abstract

E-learning is becoming more powerful and popular medium for learning through online. It is very important to identify the opinions of the users of E-learning. E-learning commonly refers to teaching efforts propagated through the use of computers to impart knowledge in a non traditional classroom environment. It is difficult to find the users mind whether they are satisfied with the E-courses. Sentiment analysis helps to find about the users who are visiting the E-learning portals. Sentiment analysis refers to the use of natural language processing text analysis and computational linguistics to identify and extract subjective information in source materials. It aims to determine the attitude of speaker or a writer with respect to some topic or overall contextual polarity of a document. Automated sentiment analysis helps to identify the sentiment of the users from the E-portal pages in which users generally browse on a topic or area they are interested. This information can be used by E-learning system to identify the user's emotional factors and can analyse the user's activity each time. The user's sentiments towards a course can serve feedback for E-learning portals in online learning environment. In the proposed E-learning system the automated sentiment analysis helps to analyse the E-portal visited pages by the user. This helps the developer and educator to know where the learner concentrating on specified areas with any manual work or any feed backs from the user. Here for automated sentiment analysis the Bayesian statistics classification, Naive Bayes classifier is used.

Keywords: Automated Sentiment analysis, Bayesian classification, Naive Bayesian classification Support Vector Machine.

1. INTRODUCTION

The E-learning systems have recorded a rapid growth in the last decade. The E-learning systems are very useful to students and teachers. In the said systems, the content of E-learning courses can be stored, arranged and distributed to the E-learners in meaningful ways, and assessments can be conducted and results presented to the learners. Some popular learning systems are Black board learning system CCNet and ALEKS.

Since the users of the E-learning portal are spread over the globe it is unlikely difficult to evaluate and rate the performance of E-learning Systems as compared with that of traditional classroom approach. The task of sentiment analysis is to classify the polarities like whether it is positive, negative or neutral to check the sentiment. Further that may vary such as happy, excited, joy, anger, sad, cry etc instead of single sentences.

In proposed paper the automated sentiment analysis itself collect the opinion of the E-learners how they are concentrating on their courses. The tasks to perform here is information extraction like discarding subjective information, recognizing opinion oriented questions, summarization. With the specified data the developers can concentrate on the specific areas where the users can concentrate more and spent

more time on the topics. The opinions and reaction of the learners of the proposed ideas are relevant in adopting the new ideas in E-learning portals.

2. RELATED WORKS

In educational institution and in corporate, E-learning plays a vital role. Recognition critical success factors (CSFs) and performance evaluation measurements of E-learning can assist educational institutions and organizations which apply these electronic solution to understand how they gain goals of adopting of E-learning and also perceive where need adaption, improvement and change[1]. The social networks are analyzed in this paper to represent the groups of students who have similar contacts and interact in similar social circles [2]. In many natural language processing tasks, Subjectivity and sentiment classification has been used as a first phase to generate more viable data. Research that done from these additional layering ranges from question answering [3], to conversation summarization and text semantic analysis [4]. In such system there is no need of classrooms and coaching system. For the last decade in E-learning courses evaluated and analysed and rated only through feedbacks like ,email and online questionnaire or manually at times. The sentiment analysis plays a role for getting the feedback on overall rating of the courses. Through sentiment analysis can assess how much the E-learning courses are effective, and we can arrive

an opinion on E-portal whether it is positive, negative or neutral.

3. PROPOSED APPROACH

In proposed paper it makes E-learning more comfortable for both learners and educators. For developers it makes very easy to know the opinions of the learners through automated sentiment analysis. To identify the sentiment in E-learning portals and through natural language processing automated sentiment analysis helps. There are so many parameters available for gathering the opinions and employ the different techniques and statistical methodology to evaluate sentiment across the web. Automated analysis can be almost as good as human analysis. In the proposed research the dataset is collected with the various instances Gender, Age, Visited pages, properties, news, groups, peer feed, labs and challenges. The dataset which helps to find the automated sentiment analysis of the E-learning portals. From the data analysis the classification is done to classify the relevant data which is useful for the developers. They can know the area of interest and interest shown on the topics. From that they can modify according to the users. Thus helps to increase the number of users to the specific E-learning portals.

3.1 Acquirement

The collection of 100 users review from the website Functionspace.org. The nature of our dataset is generalized. The dataset is collected based on the various keyword like visited pages, properties, discussions, news, groups, challenges, labs and messages. The sentiment analysis used for our dataset is

Table -1: Keywords of sentiment analysis used in the dataset

Positive	P
negative	N
Neutral	Ne

In discussion area the user can discuss about the topics and interested areas. We collected the sentiment of the discussion and it will help to motivate the developers to modify the courses according to the user’s interest. Depends on the visited pages of news the users can know the latest news available in the E-learning market. With the analysis of that area the developers can know how the users can concentrate on news.

4. CLASSIFYING SENTIMENT

Automated sentiment analysis reviews on-topic posts as the user comes in and determines the sentiment of the post at the sentence level, and aggregates positive, negative, or neutral designation at the post level based on specified sentiment keywords or phrases in a particular page or sharing of a page. If a particular document or post touches multiple topics, sentiment can be determined for each separate topic. Based on

the visited areas and on which subject the user can concentrate more, the sentiment can be classified. The classification of sentiment analysis can be done with gender, age, visited pages, properties and subject etc. The properties used in our dataset which shows about the users who are all visiting what type of topics, whether they concentrating on latest activity, advanced techniques or beginners topics.

Table 2: keywords of the properties used in the dataset

LA	LATEST ACTIVITY
KI	K-INDEX
AT	ADVANCED TOPICS
BT	BEGINNERS TOPICS
IT	INTERMEDIATE TOPICS

4.1 Machine Learning

Machine learning is a fundamental way that enables the computer to have the intelligence. Its application which has been used mainly the method of induction and the synthesis, rather than the deduction has already reached many fields of Artificial Intelligence [5]. The aim is to examine the sentiment classification of topic based categorization. For this classification we use Naïve Bayes classification is used. To implement such algorithms in our dataset, we have to follow the standard frame wok. Let {f1...fm} to be predefined set of m features. Let $n_i(d)$ be the number of times f_i occurs in document d [6]. Then the document vector $d = (n1(d), n2(d) \dots nm(d))$ is used.

4.2 Bayesian Classification (BSC)

BSC predicts the E-learning courses how the users use the courses according to the sentiment. Bayesian statistics offer a most forward for making inferences. The prior information about the process and the information about the process contained in the data are two sources of information. Bayesian classifiers use Bayes theorem, which says

$$p\left(\frac{c_j}{d}\right) \rightarrow \frac{p\left(\frac{d}{c_j}\right)p(c_j)}{p(d)}$$

Where $p(cj | d)$ is the probability of instance d being in class c_j , The $p(d | c_j)$ is the probability of generating instance d given class c_j . The $p(c_j)$ is the probability of occurrence of class c_j . The $p(d)$ is the probability of instance d occurs.

4.3 Naive Bayes Classification

Naive Bayes classification helps to assume the presence or absence of the particular feature related feature that is related to other feature in the given class. For example consider a fruit an orange if it is orange in colour, round, and about 4” in diameter. A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of the presence or absence of the other features.

The Naive Bayes classifiers is often represented as this type of graph

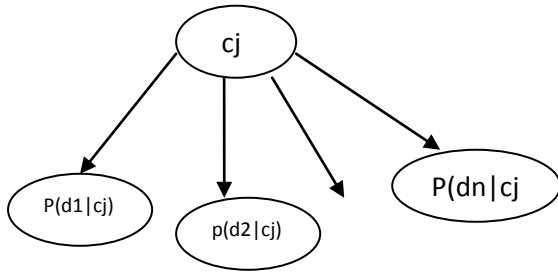


Fig -1: Arrow shows the probability of classes.

Naive Bayes text categorization tends to perform optimum for certain classes. A naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. For example, a fruit may be considered to be an apple if it is red, round, and about 4” in diameter. Even if these features depend on each other or upon the existence of the other features.

Table 3: Analysis of users concentrate on Properties

PROPERTIES	PHY	MATH	CS
AT	9.0	5.0	7.0
IT	4.0	11.0	10.0
KI	4.0	7.0	1.0
LA	14.0	4.0	9.0
BT	4.0	14.0	12.0
TOTAL	35.0	41.0	39.0

A naive Bayes classifier considers all of these properties to independently contribute to the probability that this fruit is an apple. Depending on the precise nature of the probability model. Naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood.

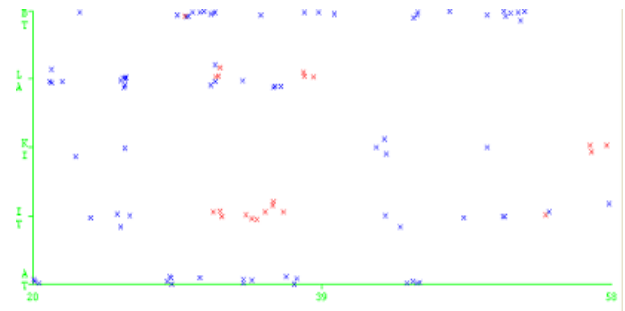


Fig -2: numbers of visited pages of properties.

Here the properties are taken as sample for classification in regards with subjects. The graphical representation of the diagram shows that which age group were concentrating on which properties of the topics. For example here the age group between 29-39 were more concentrated on Intermediate level topics, this shows the users of this age group concentrated more on the specific topics of intermediate level. So the developers come to know from the sentiment analysis of these areas helps to concentrate on the users concentrated areas.

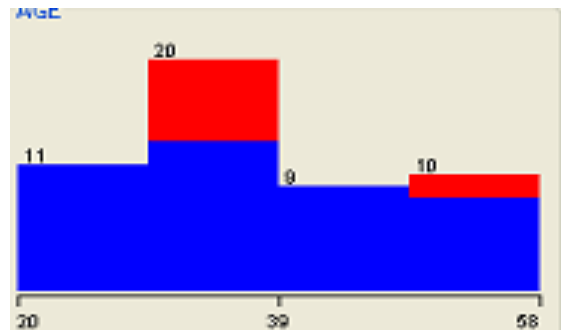


Fig 3: Numbers of users classified on their age

The histogram describes about the users on their age specifies the age group between 20-40 were concentrate more on the E-portal.

An advantage of the naive Bayes classifier is that it requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix.

5. SUPPORT VECTOR MACHINE

To analyze data and recognize pattern used for classification and regression analysis

$$W = \sum_i \alpha_i y_i x_i$$

Support vector machine (SVM) helps to find the similar hyper plane between the classes. It helps the developers to enhance the E-learning courses accordingly that make a updating regularly on the same.

6. RESULTS

The result of the proposed system will helps the developer and educator to identify the most concentrated pages in E-learning portals. It determines the sentiment of the E-courses in phrases. Also we can judge the E-learning portals .All the sentiments are classified with Bayes rule. With the classified data we can access the most accessed pages by the user. Finally automated sentiment analysis shows the E-courses highlighted the most frequently accesses courses. This motivates the users that in which area it highly focussed by the other user and also helps to know the current trends. It mainly helps the developers to develop the further ideas. And it helps to modify the courses accordingly. Give the fruit full information on most visited pages and remove the untapped pages and help to enhance E-portal developments. And also it helps the developers to maintain their E-portals regularly.

REFERENCES

- [1]. Introduce critical success factors (CSFs) of elearning for evaluating e-learning implementation success. Published in: Educational and Information Technology (ICEIT), 2010 International Conference on (Volume:1)Page(s):V1-224 - V1-228 E-ISBN : 978-1-4244-8035-7 Print ISBN:978-1-4244-8033-3
- [2]. H. Yu and V. Hatzivassiloglou. Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In EMNLP, pages 129{136, Sapporo, Japan, 2003.
- [3]. J. Wiebe and R. Mihalcea. Word sense and subjectivity. In Proceedings of the Annual Meeting of the Association for Computational Linguistics, Sydney, Australia, 2006.
- [4]. A Study and Application on Machine Learning of Artificial Intelligence, Published in:Artificial Intelligence, 2009. JCAI '09. International Joint Conference Page(s): 272 – 274
- [5]. Thumbs up? Sentiment Classification using Machine Learning Techniques Bo Pang and Lillian Lee Department of Computer Science ,Cornell University Ithaca, NY 14853 USA , Shivakumar Vaithyanathan ,IBM Almaden Research Center,650 Harry Rd. San Jose, CA 95120 USA.
- [6]. Stanley Chen and Ronald Rosenfeld. 2000. A survey of smoothing techniques for ME models. IEEE Trans. Speech and Audio Processing, 8(1):37–50.
- [7]. Web Data Mining - Exploring Hyperlinks, Contents and Usage Data Chap 11: Opinion Mining, 2nd Edition, July 2011.
- [8]. Sentiment Analysis and Opinion Mining (Introduction and Survey), Morgan & Claypool, May 2012.
- [9]. Online Web Sentiment Analysis on Campus Network Published in Computational Intelligence and Design (ISCID), 2011 Fourth International Symposium on (Volume:

2).Exploiting Combined Multi-level Model for Document Sentiment Analysis Published in: Pattern Recognition (ICPR), 2010 20th International Conference.

BIOGRAPHIES



Assistant professor, Department of computer science, Navarsam arts and science college, Erode



Professor & Head, Department of CSE, Sasure College Of Engineering, Vijayamangalam.