Volume 3, Issue 12, December 2013 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

Implementation of Bayesian Theory in Sentence Classification for Online Subjective Test

Arpit Trivedi*, Shreya Mahida

Smt.Chandaben Mohanbhai Patel Institute of Computer Applications,

CHARUSAT, India

Abstract---We described the experiment that classified a sentence (answer) entered by user was constructed by classifier with Bayesian theory used well by the text classification algorithm. It was confirmed to calculate the performance of the text classifier constructed by Bayesian theory and to show analysis ratio of correct answer and incorrect answer. Moreover, the text classifying method was built into Bayesian text classifier and the analysis ratio was able to be improved for online descriptive test.

Key words: Bayesian theorem, Naïve Bayes, Probability, Conditional Probability, Bayesian classifier, Result Analysis Ratio

I. INTRODUCTION

Recent years have been rapid growth in online examination. However, online examination conducted in the manner of objective test but the problem arise where subjective (Descriptive) online test. There have been many machines learning technique available for text classification. Naive Bayes, Maximum entropy classification and SVM are very efficient technique for classification of text. We applied Bayesian theorem based on information between documents. One document contains user answer and other document contains required information for correct answer (database). The classifier classify information (answer for particular question) by two domain either correct or incorrect answer. For classifying we calculate result analysis ratio. For Example, Determine whether answer submitted by user is correct or not.

In this paper, the system that classified document which contains answer either correct or incorrect was classified by classifier with Bayesian theory used well by the text classification task as a text classification algorithm. Bayesian approach is the statistical based text classification method which is strong algorithm for classification in which probability is calculated. Particular sentence has particular probability of occurring text. The probability of a certain cause when a certain event occurs can be calculated by the probability of all cause of event.

The goal of sentence classification is to classify sentence into predefined category. Example: to classify sentences either correct or incorrect in terms of right answer for particular question. Bayesian algorithm is easy to implement for probability based classifying a sentence and generate result analysis ratio of the given answer.

II. CLOSER LOOK AT THE PROBLEM

An Expert on using machine learning for text classification predicted relatively low performance for automatic methods. The other side, it seems that distinguishing correct from incorrect answer is relatively easy for humans who are expert for particular subjects. To test, we asked two students in computer application to give answer for particular questions. (Independently).The **figure1** shown the question asked to the students and correct answer for that questions.

Question	**What is Machine learning?**
Answer	Machine learning, a branch of artificial intelligence, concerns the construction and study of systems that can learn from data

Figure1: Question and answer from database

Human	**Answer**
User 1	It is a branch of artificial intelligence and level of algorithm.
User 2	Concerns the construction and study of systems that can learn from data.
User 3	Branch of construction and study of data those systems concerns.
User 4	Machine learning, a branch of artificial intelligence, concerns the construction and study of systems that can learn from data

Figure2: Answered given by different user.

The problem is how we can identify that which answer is correct and which one is incorrect relative online descriptive exam.

III. BAYESIAN CLASSIFIER

A Bayesian classifier assumes that the presence or absence of particular features is related to the presence or absence of any feature given the variable. For Example, A fruit may be considered to be an apple if it is red, round and other features. A Bayesian classifier considers each of these features to contribute independently to the probability that this fruit is apple regardless presence or absence of other features. For some types of probability models, Bayes classifier can be trained very efficiently in a machine learning technique. Bayes classifier has worked quite well in many complex real world situations. An advantage of Bayesian method is that it only requires a small amount of training data to estimate the parameters necessary for classification. Because independent variables are assumed, only the variances of the variable for each class need to be determined and not the entire matrix.

In this paper we are going to focus on text classification using Bayesian Theory. We classify answers given by multiple users for one question in terms of correct or incorrect answer. For the classification we have to focus on following steps:

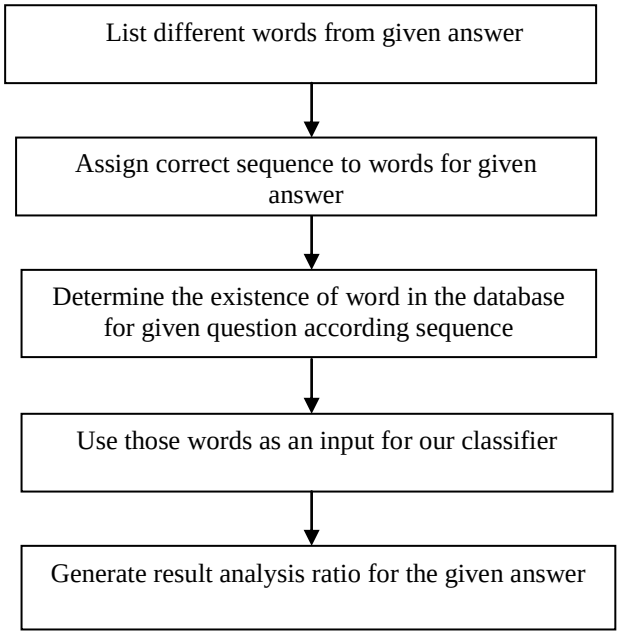

Figure 3: Answer Classification Process

Using Bayesian Theory we able to generate the probability of given answer and generate result analysis ratio to identify correct or incorrect answer given by different users as per flag indication. When word (flag) is inputted the existence probability that the answer is correct or incorrect is defined by following expression:

$$P(A) = \frac{P(incorrect\_words)/\ n_{incorrect\_words}}{(wc)\ .P(correct\_words)/\ n_{correct\_words} + P(incorrect\_words)/n_{incorrect\_words}}$$

In this Expression symbols are defined as follows:

- P(A) : Probability of answer is correct or incorrect with respect to flag generated.
- P(incorrect_words):Probability of incorrect words
- P(correct_words):Probability of correct words
- $n_{incorrect\_words}$:Number of incorrect words with compare to database
- $n_{correct\_words}$:Number of correct words with compare to database
- wc:Total number of words in answer given by user

Using formula we can identify the probability of given answer is correct or not. The Bayesian classifier is divided in to two phases

1. Preprocessing(Divide given answer in to words)
2. Classifying (Identification of given answer and find the result analysis ratio)

1. Preprocessing

a. assemble given answer from database and from users
b. Divide given answer in to words (flags) and calculate total number of words presented in given answer
c. Calculate the existence probability for each word (flag).

2. Classifying

a. Divide given answer into word(flag)
b. Query Existence Probability (EP) of each word (flag).
c. If EP is greater than zero the sentence is incorrect.
d. If EP is equal to zero the given answer is correct.

Based on EP we can generate the result analysis ratio for given answer using Bayesian classifier.
To generate Result analysis ratio the following formula is defined:

Result_Analysis_Ratio= total_no_of_correct_words / total_no_of_words_in _database

Bayesian classifier is used to classify answer using following process:

Process 1.

- We have to split given answer into words and count total number of words presented in given answer from database.

Question	**What is Machine learning?**
Answer	Machine learning, a branch of artificial intelligence, concerns the construction and study of systems that can learn from data
Words	Machine, learning, a, branch, of, artificial, intelligence, concerns, the, construction, and study, of systems, that, can , learn, from, data
Total number of Words	19

Process 2.

- Assemble different answer from multiple user and split into words and count total number of words present in given answer.

USER	**Total number of words in database**	**Total number of matching words**	**Existence Probability (EP)**	**Result Analysis ratio (%)**
User 1	11	05	0.023	26%
User 2	12	12	0.076	63%
User 3	10	09	0.074	47%
User 4	19	19	0.00	100%

From above test data we can classify that the answer of user is correct if the EP is equal to zero and Result analysis ratio is 100 %.Here in our experiment User 1, User 2 and User 3 submitted wrong answers and User 4 submitted the correct one.

V. CONCLUSION

Using Bayesian Theorem we can classify the sentences in many forms. We can apply Bayesian theorem for assessing online subjective test and add conditional probability where there is more than one occurrence for word entered by user. We can add synonyms and thesaurus in our database to get effective result. There are many machine learning techniques available for text classification.

REFERENCES

[1] Y., Hasegawa, T., Watanabe, I., Sato Ichimura, "Text Mining case Study," *Journal Japanese Society for Artificic Intelligence*, vol. 16, no. 2, pp. 192-200, 2001.

[2] M., Taira Nagata, "Text Classification - Showcase of learing Theories," *IPSJ Magazine*, vol. 42, no. 1, pp. 32-37 2001.

[3] Sanjay S. Gharde Shailedra Kumar Shrivastava, "Support Vector Machine for Handwritten Devanagari Numera Recognition," *International Journal of Computer Applications (*, vol. 7, no. 11, pp. 9-14, October 2010.

[4] Shivakumar Vaithyanathan Bo Pang and Lillian Lee, "Sentiment Classification using Machine Learning Techniques".

[5] Abdul Rahim Ahmad Christian Viard-Gaudin Marzuki Khalid Rubiyah Yusof, "Online Handwriting Recognitio using Support Vector Machine," *Proceedings of the Second International Conference on Artificial Intelligence i Engineering & Technology*, pp. 250-256, August 2004.

[6] Hirofumi Inomata, Masaki Miyamoto and Osamu Konishi Ayahiko Niimi, "Evaluation of Bysian Spam filter an SVM spam filter".

[7] I. Šileikienė, J. Lieponienė and R. Kulvietienė J. Mamčenko, "Evaluating the Data of an e-Examination System Usin a Descriptive Model in Order to Identify Hidden Patterns in Students Answers," *The Online Journal on Compute Science and Information Technology*, vol. 1, no. 2, pp. 45-49.

[8] E.Binaghi,B.Carminati,M.Carullo and E.Ferrari M.Vianetti, "Content Based Filtering in online socical network".

[9] Sathiyakumari.K Padma Priya.B, "Classification of Unwanted Messages in Online Social Network Using Machin Learning Algorithms," *International Journal of Computer Trends and Technology (IJCTT)*, vol. 4, no. 8, pp. 2697 2701, Aug 2013.

[10] John B. Carlin, Hal S. Stern, and Donald B. Rubin (2003) Andrew Gelman, *Bayesian Data Analysis*, 2nd ed.: CR(press, 2003.

[11] Kawano , H., Arimura,H Naskawa T., "Base Technology for Text mining," *Journal Of Japanese Society fo Artificical Intelligence*, vol. 16, no. 2, pp. 201-211, 2001.

[12] Durga Torshniwal Arun Rjput, "Adaptive Spam filtering based on Bayesian Algorithm," *International Journal o Advanced Computer Engineering and Communication Technology*, vol. 1, no. 1, pp. 8-11.

[13] nabeken. bsfilter/ bayesian spam filte. [Online]. HYPERLINK " http://www.h2.dion.ne.jp/ nabeken/bsfilter/ http://www.h2.dion.ne.jp/ nabeken/bsfilter/

[14] Arpit Trivedi Shreya Mahida, "Evaluation of Social Networking Site Conversation Filtering Based on Bayesia Theory," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 11 pp. 163-165, November 2013.

