TEST Engineering & Management

May-June 2020
ISSN: 0193-4120 Page No. 4379-4381

Personality Analysis using Naïve Bayes Classifier

Rahul T N[1], Raj Chandan Goyal[2], Sagar HRN[3], Vikas Biradar[4], Geetha B[5]

[1,2,3,4,5]School of Computing and Information Technology, REVA University, Bengaluru, India
[1]rahullgowda.tn@gmail.com, [2]rajchandangoyal@gmail.com, [3]sagar20gandsi@gmail.com,
[4]vikasbiradar0802@gmail.com, [5]geetha.b@reva.edu.in

Article Info
Volume 83
Page Number: 4379-4381
Publication Issue:
May - June 2020

Article History
Article Received:*19 November 2019*
Revised: *27 January 2020*
Accepted: *24 February 2020*
Publication: *12 May 2020*

Abstract

In this project we aim to recognize editable text document and perform personality analysis. Words can be classified into different classes based on the relevance with topic searched. This project contains naïve Bayes classifier which uses movie review dataset for positive and Negative word classification. The paper aims to helps understanding the best classifier for text classification. Further to ensure real time computation .NLTK is used for language processing and naïve Bayes for text classification .The existing systems is based on predictions and helpful in personality analysis based on the words used by the user and predict person is positive minded or negative minded.

Keywords**:** *Naïve bayes classifier, Text Mining, Personality Analysis*

1. Introduction

In today's worlddigital platform is so enhanced that it has vital consequences on trade and marketing. People share theirviews, opinion, experiences on their social profiles so that others can get benefit out of these. Alsoin the field of business, education institutions and booking sites large amount of data is being uses. When this data is processed, the relatable information is used to rise the present business and economy. There are so many sites where individual post his/her opinions and reviews respectively and those can be used to segregate the reviews. Classification algorithms plays important role in processing text data of huge size. Sentiment analysis is the process of computationally identifying and categorizing emotions expressed especially in order to determine attitude of a person. For us it is easy to read the sentence and understand whether the sentence is positive or negative review But for computers, it is somewhat Harder than thatHere we are using movie review corpus and NLTK library for this process. This method categorizes the person based on the words used in a particular line and segregates the word positive or negative and alsogives the probability of the particular sentence to be positive or not. The overview of the data classification algorithms is to examine if the individual is positive or negative minded based on the data given by the user. In the model we use movie review corpus dataset comparing each word in the with the corpus and then we try to guess the probability of person being positive or negative

2. Literature Survey

1. Ye fei[1] Support Vector Selection and Parameter Optimization Using Support Vector Machines for Sentiment Classification. Bag of words was used for language processing Support vector machine algorithm used for classification The above paper mentioned important method that used logical algorithm to find unique support vector set while initilizing parameter K. Outputs show that the model has better predicts better than the later. Scalability challenge is not suitable for large dataset.

2. Sumandeep kauret.al [2] Prediction based on N-gram and KNN Classifier N-gram algorithm is used for the feature extraction and KNN classifier is used to classify input data into different classes. The tested input shows up to 7 percent improvement on prediction. N gram algorithm is inefficient and KNN algorithm is inaccurate as it is based only on predictions

3. Pavithra.B.Shettyet.al [3].An Integrated Approach for Personality Analysis using OCR and Text Mining. Optical recognition for is used in data extraction Text mining and machine learning for classification.. The personality analysis is very useful to various organizations to evaluate a person before giving the chance. Positive and negative word classification is based on frequency of occurrence of words which is inaccurate.

May-June 2020
ISSN: 0193-4120 Page No. 4379-4381

4. Ankur Goel et.al[4] Real time tweet analysis using naïve bayes classifier. Training data is sent that is tweets sent and data is preprocessed and sentiword net used for lexicalize the data and naïve bayes classifier is used for data classification. The performance of the model is not so accurate as its accuracy is only 70 percent

3. Objectives

Personality analysis of a person or a sentiment analysis of a review of a product movies etc can be analysed and can be classified as positive or negative can be done in this model. Our system will employee a naïve Bayes technique for classification and probability of person being positive or being negative. However the detection of word being positive or negative and analyse personality is a significant challenge. Classification can be done using various methods. To develop a system that detects sentence is for or against based on the occurrence of the words.

Specific Objectives

1.1 To be able to accurately to detect sentence is positive or negative.

1.2 To be able to detect the required word which helps in classifying.

1.3 To accurately classify the word using the most accurate algorithm.

1.4 To provide a personality analysis based on the sentence.

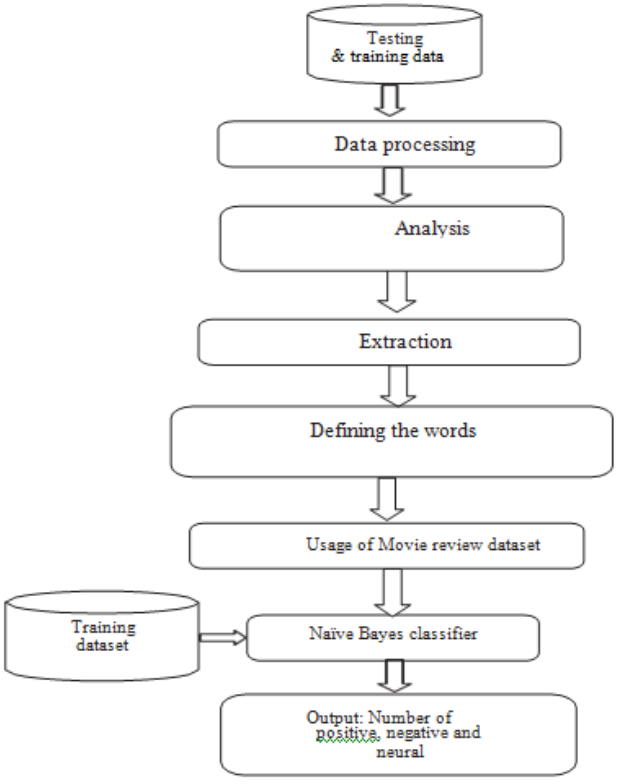

Figure 1: Flowchart for personality analysis using NLTK

4. Methodology

Training Data

The training data is the initial set of data that helps the computer learn how to process the information. The actual data the ongoing development process models learn with various algorithms to train the machine to work automatically. The individual opinions and views noted on different platforms were basically used to train the algorithm at the first stage. These training datasets are the key input that determine the output and memorize the information for future prediction.

Pre Processing

By using stop word algorithm we remove unwanted words such as articles which may slowdown the process.

Naïve Bayes classifier

This classifying algorithm produces competitive classification with great accuracy.it renders the solution by classifying the selected class labels associated with largest probability. The motive for the model algorithm is Bayes theorem. Bayes theorem gives the probability of an event occurring provides the probability of other event already occurred

$$P\left(\frac{A}{B}\right) = \frac{P\left(\frac{B}{A}\right)P(A)}{P(B)}$$

Bayes Theorem (1)

Where:

P(A|B): It tells the probability of occurring event A given event B has already occurred.

P(B|A): It tells the probability of occurring event B given event A has already occurred.

P(A): Probability of event A

P(B): Probability of event B

It is more accurate compare to other classifier

Algorithm : Personality analysis

Input: editable text document

Output :Classification of the document into positive and negative

Working:

```
While(movie_review dataset){
if(text= = positive)
        Positive←positive word
    else
        Negative←negative word
    }do(train ←unrecognized data)
```

Proposed Work Flow

1. Initially train the model using a good and suitable movie review corpus.

2. Further use the movie_review testing corpus one word a time.

May-June 2020
ISSN: 0193-4120 Page No. 4379-4381

3. Split the word and process it.

4. Considering all distributed words calculate the probability.

5. Calculate probability of all the words in stated sentence and compliment with the probability.

6. Calculate the conditional probability using Naive Bayes classifier mathematically for each word in the sentence, and then probability score for sentence being positive or negative

5. Results

Personality analysis: The last step of the algorithm is to find whether the person's personality is for or against based on the words used in sentence by a Person

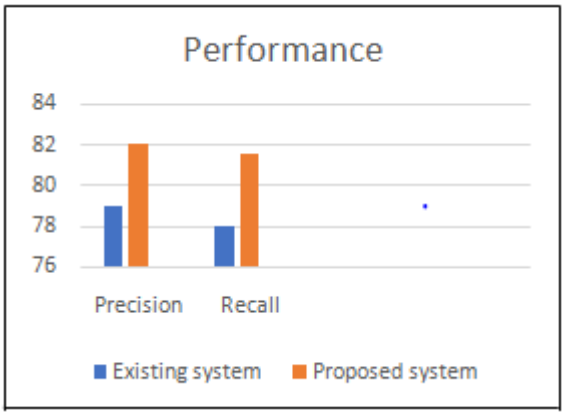

Figure 2: Performance analysis

Table 1: Performance comparison

Performance	Existing system	Proposed system
Precision	79%	82 %
Recall	78%	81.5%

6. Application

Since personality analysis allows us to detect the state of person particular mind, it can be used in various fields like,

- Sentiment analysis of positive and negative sentence is very helpful to any review based platforms or sites. Ex: items reviews in amazon, movie reviews on book my show etc.
- Detect the different kind of mails in your mail list.
- Determining user's personality becomes technically easier apparently.

7. Conclusion

A real time method for determining the personality analysis of a person has been described. Analysing data can be helped in various fields so there can be future changes based on the review of a person. The experimental result shows that there is a 7% percent improvement in the system

References

[1] Ye Fei, "Simultaneous Support Vector selection and parameter optimization using Support Vector Machines for sentiment classification," 2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, 2016, pp. 59-62.

[2] S. Kaur, G. Sikka and L. K. Awasthi, "Sentiment Analysis Approach Based on N-gram and KNN Classifier," *2018* First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 2018, pp. 1-4.

[3] Pavithra B Shetty, Sanath R Kashyap,Sneha V Karanth, Bhushan S.N "An Integrated approach for personality analysis using OCR and Text mining", Department of Computer Science and Engineering, Sahyadri College of Engineering & Management, Mangaluru-5750072018.2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC).

[4] A. Goel, J. Gautam and S. Kumar, "Real time sentiment analysis of tweets using Naive Bayes," 2016 2nd International Conference on Next Generation Computing Technologies (NGCT), Dehradun, 2016, pp. 257-261.