International Journal of Management, IT & Engineering
Vol. 9 Issue 4, April 2019,
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

APPLICATION OF NAÏVE BAYES CLASSIFICATION FOR DISEASE PREDICTION

Sharad Mathur*

Dr. Bhavesh Joshi**

Abstract

In today's era data mining is widely used for disease prediction in healthcare sector. Data mining is process of discovering information from large datasets or other repositories. It is a difficulty work to predict diseases from the available large medical dataset. To find solution of this problem researcher use apply various data mining technique. Classification is a method of data analysis that can be implemented for developing models to elaborate significant classes of data. One of the efficient and famous classification techniques of data mining is Naïve Bayesian (NB). Bayesian method is basically very significant and effective data mining technique. Providing the probability distribution, this classifier can provably obtain best result. It works on the basis of theory of probability. In this paper we applied Naïve Bayesian classifier on Asthma disease dataset and compared different Bayes Classifiers available in Weka data mining tool. We compared algorithms on the basis of three measures – Recall, Precision and F-Measure..

.

Keywords:

Disease Prediction;

Classification;

Naïve Bayesian;

Weka;

Precision.

***Research Scholar,Faculty of Computer Science, PAHER University, Udaipur (Rajastha) -India**

****Research Guide, Faculty of Computer Science, PAHER University, Udaipur, (Rajastha) -India**

1. Introduction

Now a day the main challenge which is faced by healthcare sector is the diagnosis of disease at reasonable cost and with accuracy. In the digital world larger and complex data is available in hospitals and diagnosis centers that are can be utilized to find useful information for diagnosis purpose [1]. Collection of medical data can be done by various methods like interviews with patient, Scanning many images, data from diagnosis labs and by physical examination of patient by doctors. This data can be utilized by data mining to perform future predications. This is a necessary step for knowledge discovery from dataset and suitable techniques are used to find patterns. The goal is to search valid and novel correlation of data via relating complex data sets to discover patterns those are very hard to identify by people. Data mining offers several algorithms for classification, association, clustering, that attempt to fit a model nearest to the characteristics of data under consideration. Naïve Bayes technique is one of the most effective, accurate and widely used technique. One major benefit of Naïve Bayes which is attractive to doctors is that all available information can be utilized to explain the decision. The explanation looks natural for medical diagnosis i.e. is close to pattern how physicians diagnosis patients [2]. In case of medical data, Naïve Bayes algorithm takes into account evidence from many attributes to make final decision and gives transparent explanation of its decisions and that why it is known as one of the very useful classifier to support physician's decisions.

Many organizations that are working on Respiratory disease have warned that large numbers of persons are ignorant about lung diseases and it is the reason of most of deaths than any other disease. More than 55 million populations in India suffering with chronic lung disease like asthma, chronic bronchitis and emphysema. Economic burden of asthma in India is calculated about 139.45 billion per year [3]. Asthma diagnosis is difficult as every person has different patterns of symptoms [4]. Usually peoples neglect common asthma symptoms such as cough, cold wheezing etc. before consulting doctor. Many times asthma disease can become dangerous for life and statistics also depicts that more than 180,000 people dies per year due to it [5].

So it is required to create a diagnosis system that can assist doctor to know that a patient is suffering from asthma or not at very early stage. To reduce time and cost to obtain the above goals, a proper computer based information decision support system is required that can perform

authentic and efficient diagnosis of asthma. Data mining is helpful to develop such diagnosis system. In over study we concentrated on Naïve Bayes method of classification. To know how well it works on medical data we applied it three variations on asthma disease dataset. We used Weka data mining tool to perform our experiment.

2. Related Work

Dhamodharan et.al [6] implemented Naïve Bayes algorithm to predict diseases like Hepatitis, Liver cancer, Cirrhosis using different symptoms. They implanted FT Tree and Naïve Bayes algorithm to develop prediction system. Both algorithms are analyzed on the basis of classification accuracy. The outcome of their research is that the Naïve Bayes algorithms predicted disease with high classification accuracy as compared with FT Tree.

Chaitrali S. Dangare et.al [7] has investigated prediction system of Heart disease with the help of large attribute count. They used many classification methods like Naïve Bayes, Neural Networks and decision tree to execute on Heart disease dataset. All there algorithms are analyzed on the basis of accuracy. The outcome of research is that Neural Network has produced high accuracy for heart disease prediction.

DevendraRatnaparkhi et.al [8] recommended a system for heart disease prediction with the help of Naïve Bayes algorithm and they analyzed its output with decision tree and neural network algorithms. Their result shows that Naïve Bayes method produces.

Lamboder Jena et.al [9] investigated dataset of kidney disease that is available for free at repository of UCI machine learning. They have performed experiment on it with algorithms like Naïve Bayes, J48, SVM and Multilayer Perceptron and compared them on basis of classification accuracy. Experiments result demonstrated that multilayer perceptron algorithm shown higher classification accuracy and is the best for prediction of chronic kidney disease.

3. Naïve Bayes Classifier

Naïve Bayes classifier is a very effective and powerful classification technique. It is applied to answer diagnostic and predictive problems. An exhaustive comparison of Naïve Bayes algorithm

with other classification techniques in 2006 exhibited that it performs outstandingly in comparison of random forests or boosted trees [10]. This classifier significantly made learning simple by taking assumption that features are independent given class. Naïve Bayes is basically a simple probabilistic classifier that works on Bayes theorem. It records how frequently a target field value come out jointly with a value of input field. Naïve Bayes algorithm handles every symptom to work independently to the probability that patient has asthma. Naïve Bayes method calculates probability of observing a specific value in a particular class by the ration of its frequency in the class of interest over the prior frequency of that class. Category C1 represent a person suffering with asthma and C2 represent a person not suffering from asthma. x1,x2,uptoxn represents symptoms of asthma which in this research we included in our study.

$$P\ (x1, x2, x3\ldots.xd\ |Cj) = P(xi|Cj)$$

$$P\ (c|X) = P\ (x1|c) * P(x2|c) * \ldots.P(xn|c) * P(c)$$

If a person suffering from x1 that is breathing problem, x2 that is cold or flu,x3 that is allergy, x4 that is wheezing then following process will compute probability of a person is having asthma or not –

Step 1 - A person having asthma its probability can be compute.

P(x1 | C1)= Number of persons having breathing problem and have asthma / number of patients having asthma.

P (C 1)= Number of persons suffering from asthma / total number of patients.

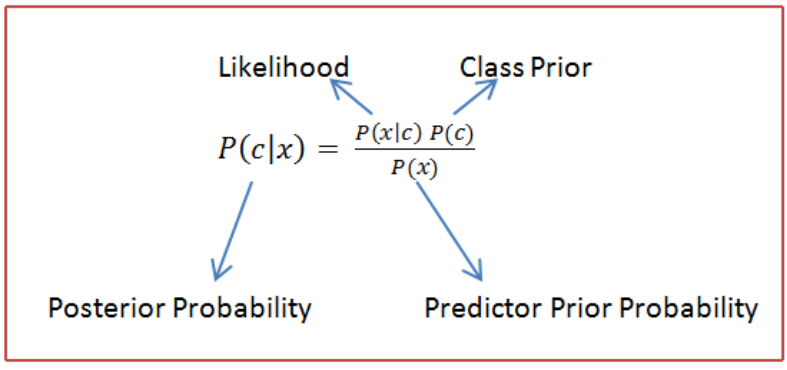

Fig – Naïve Bayes Formula

P (xn │C1)= P (x1 │C1)* P (x2 │C1)* P (x3 │C1)* P (x4 │C1)* P (C1)

Step 2- A person not having asthma, its probability can be compute.

P (xn|C2) = P (x1|C2) * P (x2|C2) * P (x3|C2) * P (x4|C2) * P (C2)

Step 3- Finally probability of a person having asthma or not compared. If P (xn|C1) higher it means having asthma or if P (xn|C2) larger it means person not having asthma.

4. Experimental Work

WEKA (Waikato Environment for knowledge analysis) is a powerful and popular data mining tool. It has several features like graphical interface, command line interface, open source, Java API and documentation. WEKA can be used by persons who do not have any programming language. This is the main reason of success of WEKA. WEKA provides six variations of Bays classifications that are Bayes Net, Naïve Bayes, Naïve Bayes Multinomial, Naïve Bayes Multinomial Text, Naïve Bayes Multinomial Updateable and Naïve Bayes Updateable. We used three of them to perform our experiment on asthma.arff dataset that are Naïve Bayes, Naïve Bayes Multinomial Text and Naïve Bayes Updateable.

We first performed our experiment on Naïve Bayes classifier of WEKA learning tool. Table 1 shows the output of this algorithm. The output of performance is given in tables in terms of Precision, Recall, and F-Measure. The last row of first three tables shows the average values of the whole classes.

TABLE 1: NAÏVE BAYES

Classes	Precision	Recall	F-Measure
Asthma	0.739	0.773	0.756
No Asthma	0.935	0.923	0.929
Average	0.892	0.890	0.891

Second experiment is with Naïve Bayes Multinomial Text algorithm. Table 2 represents the result this WEKA classifier.

TABLE 2: NAÏVE BAYES MULTINOMINAL TEXT

Classes	Precision	Recall	F-Measure
Asthma	0.000	0.000	0.000
No Asthma	0.780	1.000	0.876
Average	0.608	0.780	0.684

At last Weka Classifier that is Naïve Bayes Updateable implemented on asthma disease dataset and following table is showing result of that.

TABLE 3: NAÏVE BAYES UPDATEABLE

Classes	Precision	Recall	F-Measure
Asthma	0.739	0.773	0.756
No Asthma	0.935	0.923	0.929
Average	0.892	0.890	0.891

5. Results and Discussion

The output of Naïve Bayes algorithm shows best resulted precision is with Asthma class and best recall was with No Asthma class.Results of Naïve Bayes Multinomial Text classier are very low especially in case of class Asthma.Average values of Precision, Recall and F-Measure of this algorithm are lower in comparison of Naïve Bayes Algorithm. For Naïve Bayes Updateable the best resulted precision obtained of No Asthma class and best recall value was also for No Asthma class. All Obtained Results of Naïve Bayes algorithm and Naïve Bayes Updateable algorithm matched exactly.

Following tables can be used for comparison of all three classifiers and describe the best classifier among them. Table 4 summarizes all of the above tables from table 1 to table3. It takes values of last rows of all tables that contain average Precision, average Recall and average F-Measure. Table 5 is showing classification accuracy of all thee classifiers.

TABLE 4: CLASSIFIERS RESULT ANALYSIS

Classifier	Average Precision	Average Recall	Average F-Measure
Naïve Bayes	0.892	0.890	0.891
Naïve Bayes Multinomial Text	0.608	0.780	0.684
Naïve Bayes Updateable	0.892	0.890	0.891

TABLE 5: ACCURACY OF DIFFERENT CLASSIFIERS

Classifier	Accuracy
Naïve Bayes	89%
Naïve Bayes Multinomial Text	78%
Naïve Bayes Updateable	89%

6. Conclusion

It is observed that good results are given by Naïve Bayes and Naïve Bayes Updateable. On other hand Naïve Bayes Multinomial Text has given worst result especially for class Asthma. In terms of classification accuracy Naïve Bayes and Naïve Bayes Updateable both have given results with 89% of accuracy but Naïve Bayes Multinomial Text produce result with accuracy of 78%. So we can conclude with words that Naïve Bayes and Naïve Bayes Updateable are better classifier than Naïve Bayes Multinomial Text for our asthma disease dataset.

7. Future work

For future work we will improve Naïve Bayes algorithm by combining features of all algorithms to implement Hybrid Algorithm to solve classification problems and to produce best classification algorithm. This Hybrid algorithm will produce best result especially for asthma disease dataset.

References

[1] S. G. Milan Kumari, "Comparative study of data mining classification methods in cardiovascular disease prediction," in International Journal on Computer Science and Technology (IJCST) Vol. 2, Issue 2, June 2011.

[2] Zelic, I., I. Kononenko, N. Lavrac and V. Vuga, 1997. Induction of decision trees and bayesian classification applied to diagnosis of sport injuries. J. Med. Syst., 21: 429-444

[3] Surendran Aneeshkumar, Raj B Singh, "Economic burden of asthma among patients visiting a private hospital in South India" in Lung India Journal, 35(4), July- Aug, 2018, 312-315.

[4] http://www.getasthmahelp.org/infant-public.aspx

[5] http://articles.cnn.com/2005-05-05/world/asthma

[6] Dhamodharan. S, Liver Disease Prediction Using Bayesian Classification, Special Issue, 4th National Conference on Advanced Computing, Applications & Technologies, May 2014, page no 1-3.

[7] Chaitrali S. Dangare, Sulabha S. Apte, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques", International Journal of Computer Applications (0975 – 888), Volume 47– No.10, June 2012, page no 44-48

[8] Devendra Ratnaparkhi, Tushar Mahajan, Vishal Jadhav, ―Heart Disease Prediction System Using Data Mining Technique‖, International Research Journal of Engineering and Technology (IRJET), Volume: 02 Issue: 08, e-ISSN: 2395 -0056, p-ISSN: 2395-0072, Nov-2015.

[9] Lambodar Jena, Narendra Ku. Kamila, "Distributed Data Mining Classification Algorithms for Prediction of ChronicKidney-Disease", International Journal of Emerging Research in Management &Technology, Volume-4, Issue-11, and ISSN: 2278-9359, November 2015.

[10] D. M. S. S. A. Miss Chaitrali S. Dangare, "A data mining approach for prediction of heart disease using neural networks," in International Journal Of Computer Engineering Technology.