

CLASSIFICATION AND PREDICTION OF DERMATITIS DATASET USING NAÏVE BAYES AND VALUE WEIGHTED NAÏVE BAYES ALGORITHMS

V. Priya¹, N. Sathya²

¹Assistant Professor, Department. of Computer Science, Vellalar College for Women, Erode, Tamilnadu, India

²Research Scholar, Department of Computer Science, Vellalar College for Women, Erode, Tamilnadu, India

ABSTRACT - Data mining is the process of automatically discovering useful information in large data repositories. In medical data analysis feature selection is important task for diseases prediction process. This work reveals the chances of different dermatological disease and also finds the probabilities of occurrence of each disease. The Naive Bayes' (NB) algorithm are applied on medical data set to find out the dermatology disease prediction and analysis. The Automated medical analysis is helped to diagnosis and predict diseases by doctors. This work goes through different dermatological diseases with similar symptoms, data are gathered from UCI repository contains the 230 instances with 22 attributes. The experimental results show that performance of Naive Bayes and weighted Naive Bayes methods In order to compared the accuracy of both methods hence weighted Naive Bayes gives the better results for the prediction process.

Key Word: Classification, Naive Bayes, Value Weighted Naive Bayes.

I. INTRODUCTION

DERMATOLOGY DATASET

Dermatology is the medical science of disorders of the skin. Often skin lesions and disorders are markers of internal disease and mirror the processes occurring within the body. Skin is the largest organ in the body. Although skin diseases are easily detectable and diagnosing symptoms and deciding therapy are easier than other systemic diseases, many people ignore the importance of them. Skin is the major organ of the human body, with a range of purpose that supports survival.

There are six different categories of skin diseases they are,

1.Psoriasis.2. Seborrhea dermatitis.

3. Lichen planus.4. Pityriasis rosea.

5.Chronicdermatitis.6.Pityriasisrubra pilaris

1. PSORIASIS

A skin disease that causes scaling and swelling. These patches can itch or feel sore. They are often found on the elbows, knees, other parts of the legs, scalp, lower back, face, palms, and soles of the feet. But they can show up on other areas, as well. Psoriasis can be hard to diagnose because it can look like other skin diseases. Most psoriasis causes patches of thick, red skin with silvery scales. These patches can itch or feel sore.

2. SEBOREIC DERMATITIS

It can occur with or without reddened skin. Symptoms include red, scaly, greasy, itchy, and inflamed skin. Seboric dermatitis (SD) is a common, chronic, relapsing skin disease affecting the seboreic areas of the body including the scalp, face (nasolabial folds, ears, and eyebrows), and upper part of the trunk (chest/presternal region). Areas of the skin rich in oil-producing glands are often affected including the scalp, face, and chest.

3. LITCHEN PLANUS

Lichen planus is a chronic, inflammatory, autoimmune disease occurring in 0.1 to 4 percent of the general population, most often in premenopausal women. Lichen planus can appear at any age, but most cases occur between 30 and 60 years of age. Lichen planus is a chronic, inflammatory, autoimmune disease that affects the skin, oral mucosa, genital mucosa, scalp, and nails. Lichen planus has varied clinical forms in the oral mucosa and cutaneously that has different prognosis. This condition also arises in association with various other systemic conditions such as hypertension, diabetes mellitus.

4. PITYRIASIS ROSEA

This common skin disease causes patches on the skin. Pityriasis rosea is a common skin disease that causes a rash. This rash usually disappears on its own without treatment. You can expect to see the rash for about 6 to 8 weeks. Sometimes the rash lasts much longer.

5. CRONIC DERMATITIS

Dermatitis is also known as eczema, is a group of diseases that results in inflammation of the skin. These diseases are characterized by itchiness, red skin and a rash. In cases of short duration, there may be small blisters, while in long-term cases the skin may become thickened.

6. PITYRIASIS RUBRA PILARIS

Pityriasis rubra pilaris (PRP) refers to a group of skin conditions that cause constant inflammation and scaling of the skin. People with PRP have reddish, scaly patches that may occur everywhere on the body, or only on certain areas. Some people with PRP also develop thickened skin on the underside of the hands and feet (palm plantar), various nail abnormalities, and/or thinning of the hair. There are when symptoms begin, body areas involved, and whether other conditions are present. This condition occurs in adults (adult onset PRP) as well as children (juvenile onset PRP).

II. LITERATURE REVIEW

Chang-Hwan Lee et.al, [2017] has proposed assigning weights in features has been an important topic in some classification learning algorithms. The propose a new paradigm of assigning weights in classification learning, called value weighting method. While the current weighting methods assign a weight to each feature, the assign a different weight to the values of each feature. The performance of Naive Bayes learning with value weighting method is compared with that of some other traditional methods for a number of datasets. The experimental results show that the value weighting method could improve the performance of Naive Bayes significantly.

John year wood et.al, [2014] has proposed is a popular classification technique for data mining and machine learning. Numerous methods a have been proposed in order to improve the performance of the Naive Bayes classifier by alleviating the attribute independence assumption. Another alternative is assigning the weights for attributes. The propose a novel attribute weighted Naive Bayes classifier by considering weights to the conditional probabilities. An objective function is model and taken into account, which is based on the structure of the Naive Bayes classifier and the attribute weights. The optimal weights are determined by local optimization method using the quasisecant method. In the proposed approach, the Naive Bayes classifier is taken as starting point. The report the results of numerical experiments on several real-world data sets in binary classification, which show the efficiency of the proposed method.

Chang-Hwan Lee et.al, [2011] has proposed Naive Bayesian learning has been popular in data mining applications. However, the performance of Naive Bayesian learning is sometimes poor due to the unrealistic assumption that all features are equally important and independent given the class value. Therefore, it is widely known that the performance of a naive Bayesian learning can be improved by mitigating this assumption, and many enhancements to the basic naive Bayesian learning have been proposed to a resolve this problem including feature selection and feature weighting. The propose a new method for calculating the weights of features in naive Bayesian learning using Kullback-Leibler measure. Empirical results are presented comparing this new feature weighting method with some other methods for a number of datasets.

Chang et.al.[2009] has proposed conducted five experiments focusing on six major skin diseases and used decision tree of data mining combining with neural network classification methods to construct best predictive model in dermatology. The study predicted and analyzed six commonly seen skin diseases namely psoriasis, seborrhea dermatitis, lichens planus, pityriasis rosea, chronic dermatitis and pityriasis rubra pilaris.

H. Zou and T. Hastie al., [2005] has proposed the stretchy net, a new regularization and changeable assortment technique. In adding, the stretchy net give confidence a group result, where powerfully connected predictors are inclined to be in or out of the replica together. The stretchy net is chiefly helpful when the number of predictors (p) is a great deal better than the figure of explanation (n). By difference, the noose is not an extremely acceptable changeable assortment practice in the pan case. An algorithm identify LARS-EN is future for figure stretchy net regularization pathway professionally, a lot like algorithm LARS do for the noose. In this paper a new regularization method which describe the stretchy net. Alike to the noose, the elastic net concurrently does routine changeable assortment and incessant reduction, and it can choose collection of connected variables. It like a stretchable fishing net that retains 'all the big fish'. Simulation studies and real data examples show that the elastic net often outperforms the lasso terms of prediction accuracy.

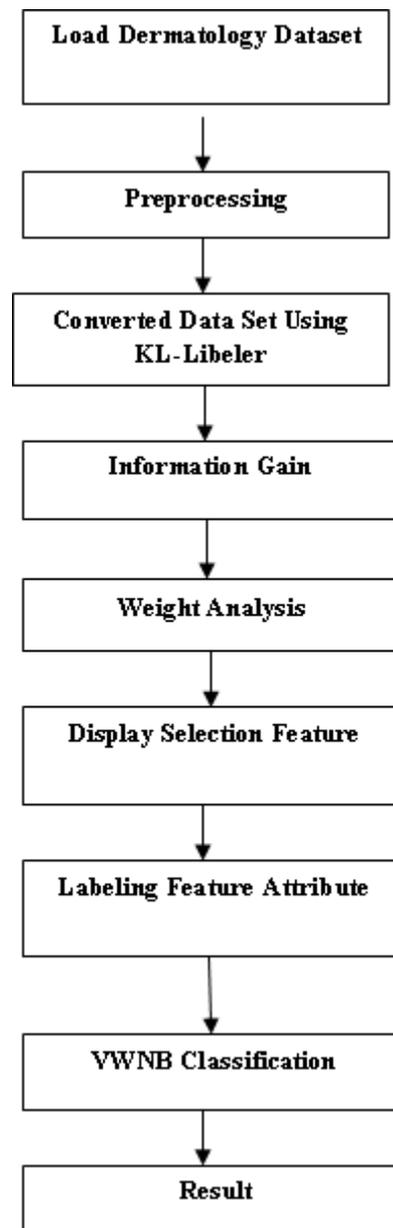
III. METHODOLOGY

NAIVE BAYES

The Naive Bayes classifier technique is based on Bayesian theorem and is especially used when the dimensionality of the inputs is high. The Bayesian classifier is adequate of calculating the most possible output based on the input. It is also possible to add new raw data at runtime and have improved probabilistic classifier. A Naive Bayes classifier regard that the presence (or absence) of a particular feature (attribute) of a class is unrelated to the presence (or absence) of any other feature when the class variable.

VALUE WEIGHTED NAIVE BAYES

To calculate probability and use probability to calculate information gain. Use feature selection method to remove the unnecessary feature based on information gain value. After that should calculate weight for each feature. They should find the missing value and also calculate weight using Value Weighted Naïve Bayes.



DERMATOLOGY DATASET

Dermatology is the medical science of disorders of the skin. Often skin lesions and disorders are markers of internal disease and mirror the processes occurring within the body

Skin is the largest organ in the body. Although skin diseases are easily detectable and diagnosing symptoms and deciding therapy are easier than other systemic diseases, many people ignore the importance of them. Skin is the major organ of the human body, with a range of purpose that supports survival.

Preprocessing

Data Preprocessing for dermatology Data directly taken from the source will likely have inconsistencies, errors or most importantly, it is not ready to be considered for a data mining process. Furthermore, the increasing amount of data in recent science, industry and business applications, calls to the requirement of more complex tools to analyze it. Data preprocessing includes the data reduction techniques, which aim at reducing the complexity of the data, detecting or removing irrelevant and noisy elements from the data. The real-world large datasets are obtained from many sources and contain data that tend to be incomplete, noisy and inconsistent. In this context, it is important to prepare raw data to meet the requirements of data mining algorithms. This is the role of data pre-processing stage, in which data cleaning, transformation and integration, or data dimensionality reduction are performed.

INFORMATION GAIN

Information gain is a widely used method for calculating the importance of features, including in decision tree algorithms. It is quite an intuitive argument that a feature with higher information gain deserves higher weight. Uses the information gain (or gain ratio) that underpins the criterion to construct the decision tree for classifying objects. It calculates the difference between the entropy of a priori distribution and that of a posteriori distribution of class, and uses the value as the metric for deciding branching node.

WEIGHT ANALYSIS

Current feature weighting methods have some limitations that each feature value has the same significance with respect to the target concept. A feature takes on a number of discrete values, and each value might have different importance with respect to the target values. Since the current feature weighting methods assign a weight to each feature, all values of the given feature are given the same weight.

V. ALUE WEIGHT NAÏVE BAYES CLASSIFICATION

A feature takes on a number of discrete values, and each feature value has different importance with respect to the target value. In this work, they propose a new method for calculating the weights for feature values using filter approach. Information theoretic filter method for assigning weights to feature values, it has strong theoretical background for deciding and calculating weight values.

IV. RESULT AND DISCUSSION

The empirical results on several benchmark datasets and compare the Value Weighted Naïve Bayes(VWNB) algorithm against other state-of-the-art supervised algorithm. Skin data set collected from the UCI repository. The continuous features in datasets were discredited. The characteristics of the datasets the used are omitted due to space limitations. To evaluate the performance used to 10-fold cross validation method and used Weka software to run the program.

Table 5.2.2 Overall accuracy

Dataset	VWNB	NB
Dermatology	* 98.0	77.8

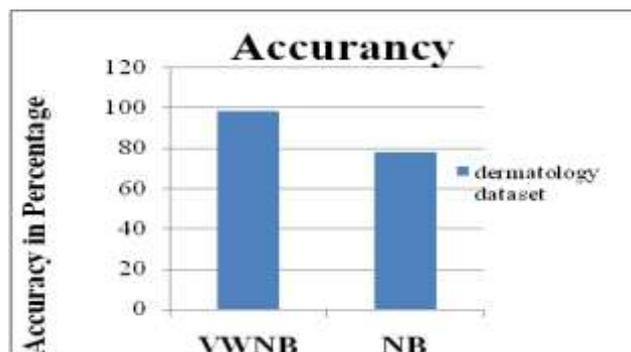


Fig: 5.2 Overall accuracy comparisons

V CONCLUSION

A new paradigm of weighting method, called value weighting method, is proposed. Unlike current feature weighting methods, a more fine-grained weighting methods, where we assign a weight to each feature value. An information-theoretic filter method for calculating value weights was developed. The proposed weighting method is implemented and tested in the context of Naive Bayes. In light of these evidences, the could improve the performance of classification learning even further by using a value weighting approach. As future work, the will combine the value weighting method with other classification algorithm, and study whether the value weighting method could show the better performance regardless of algorithms. In feature work the focus the area is to use pattern disease to mine the vast amount of unstructured data available in healthcare databases. Another challenge would be to integrate data mining and text mining it also to extend traditional Naive Bayes, and to improve its performance, weighted concept is incorporated. Exploration of Domain knowledge based weight assignment on UCI machine learning repository dataset of all dermatology data analysis is performed.

REFERENCES

- 1) Chang-Hwan Lee, An information-theoretic filter approach for value weighted classification learning in naive Bayes Received 2 August 2016; Received in revised form 1 November 2017; Accepted 11 November 2017.
- 2) John Yearwood, sattar, Attribute weighted naive bayes classifier using a local optimization. Article in Neural Computing and Applications · April 2014
- 3) Sona Taheri1 Musa Mammadov Improving Naive Bayes Classifier Using Conditional Probabilities CRPIT Volume 121 - Data Mining and Analytics 2011 .
- 4) Chang-Hwan Lee , Calculating Feature Weights in Naive Bayes with Kullback-Leibler Measure December 2011 with 312 Reads DOI:
- 5) Chang, Chih-Hao Chen 2009. Applying decision tree and neural network to increase quality of dermatological diagnosis, Expert System with Applications 3, 4035 – 4041