ISSN 2395-1621

Naive Bayes Implementation using Supervised Learning in Features Selection

#1Priyanka Konde, #2Prof. Aruna Verma

1piyu1818@gmail.com
2soni.aruna66@gmail.com

#12Department of Computer Engineering,
DPCOER, Wagholi, Pune.

ABSTRACT

Web Mining is searching useful data from the World Wide Web repository which is divided into Content Mining, Usage Mining and Structure Mining in which Content Mining uses text, images, Audio and Video to extract useful information which is Unstructured. Web Mining is sub process of Data Mining which involves Anomaly detection, Classification, Clustering, Association Rule Mining, Regression and Summarization. This discovers patterns in large data sets involving many disciplines such as Artificial Intelligence, Machine Learning, Statistics and Database Systems. Machine Learning is the emerging technology to make the machines to predict values for new data inputs according to the previous data inputs trained with some Algorithms. Among all, the Classification is in supervised Learning of Machine learning where a training set of correctly predicted observation is available. In this paper we study algorithms of Naïve Bayes.

ARTICLE INFO

Article History

Received: 9th May 2018

Received in revised form :

9th May 2018

Accepted: 11th May 2018

Published online :

12th May 2018

Keywords: Machine learning, classification, naïve bayes, precision, recall.

I. INTRODUCTION

In recent research, use of machine learning techniques in data mining has increased. This task of knowledge discovery with the help of a machine learning technique called as supervised learning. In supervised learning class labels are assigned to each and every tuple in training data, this labeled training data is used for deriving a function [1, 2]. This function further can be used for mapping new example. When feature selection is applied before supervised learning it increases the accuracy of classification. In feature selection, selection of most distinct feature is done. The goal of this technique is to remove redundant and irrelevant features [3, 4]. Due to this dimensionality of the original data set is reduced which results in the efficient performance of the classifier. When features are selected before applying data mining algorithm using some independent approach it is called as Filter method for feature selection. Impact of feature selection for supervised learning can be analyzed by comparing performance of different classification methods.

Naive classifier uses statistical as well as a supervised learning method for classification. It is based on application of Bayes theorem with naive independence assumptions.

II. LITERATURE REVIEW

Feature selection can improve the comprehensibility of the resulting classifier models and often build a model that generalizes better to unseen points. The problem of Online Feature Selection is aiming to resolve the feature selection problem in an online fashion by effectively exploring online learning techniques. In particular, two kinds of Online Feature Selection tasks are addressed in two different settings, learning with full inputs of all the dimensions or attributes and learning with partial inputs of the attributes [21].

Lei Yu and Huan Liu proposed a fast filter method to identify relevant features. Feature selection is preprocessing step to machine learning and effective in selecting relevant data by removing irrelevant data, increasing accuracy, and increasing result. However, the increase in dimensionality of data poses a severe challenge to many existing feature selection methods with respect to efficiency and effectiveness. A fast filter method can identify relevant features as well as redundancy among relevant features without pair-wise correlation

analysis. The efficiency and effectiveness of the method is demonstrated through extensive comparisons with other methods using real-world data of high dimensionality [19].

Rajdev Tiwari and Manu Pratap Singh developed a correlation based feature selection using genetic algorithm. Integrating data sources is referred to as the task of developing a common schema. It is also data transformation solutions for a number of data sources. The size of the data should fit to datawarehouse. The features are reduced using Attribute subset selection and correlation analysis and it detects the unwanted features [28].

Jasmina novakovic, perica strbac, Dusan bulatovic implemented a comparison between several feature ranking methods. This method is implemented on datasets. The six ranking methods is divided into two types. They are statistical and entropy based method. The supervised learning algorithms are used to build models. Naive Byes, C4.5 are used for classification. Based on the ranking methods, the classification accuracy is obtained. In this work, ranking methods with different supervised learning algorithms give different results for balanced accuracy [30]

III. MACHINE LEARNING OVERVIEW

Machine Learning is automatically learn to make predictions on current data based on past history. It is divided into Supervised and Unsupervised Learning. Supervised Learning is when for every observation i = 1, 2, 3, 4… n and a vector of measurement xi but not associated response yi. Unsupervised learning has inputs but no supervising Outputs to learn Relationships and structure of data. Predicting a continuous quantitative Output value is referred as Regression Problem. Predicting a non-numerical, Qualitative value or categorical Output value is Classification. Observing only Input Variables and No Output variables and grouping those input variables depending on their characteristics called Clustering. Input variables are referred as Predictors, Independent, Features or variables X. Output variables are referred as Response or Dependent variable Y.

The relationship between Y (Response) and X (Predictors) and it is written as,

$Y=f(x)+E(1)$…. [1]

‘f’ is fixed or unknown function and E is a random error term independent of X and mean zero. To estimate the function f apply a statistical learning method to the training data. Accuracy of f depends on two quantities Reducible error or Irreducible error. If the error can be reduced by increasing the accuracy then it is reducible error. If it cannot be reduced in any case then it is Irreducible error.

IV. RELATED WORK

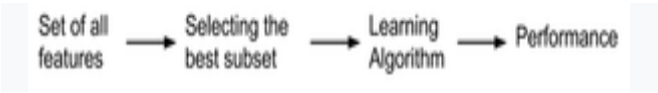
High-dimensional data analysis is a challenge for researchers and engineers in the fields of machine learning and data mining. Feature selection provides an effective way to solve this problem by removing irrelevant and redundant data, which can reduce computation time, improve learning accuracy, and facilitate a better understanding for the learning model or data. The central premise when using a feature selection technique is that the data contains many features that are either redundant or irrelevant and can thus be removed without incurring much loss of information. Redundant or irrelevant features are two distinct notions, since one relevant feature may be redundant in the presence of another relevant feature with which it is strongly correlated.

Feature selection techniques should be distinguished from feature extraction. Feature extraction creates new features from functions of the original features, whereas feature selection returns a subset of the features. Feature selection techniques are often used in domains where there are many features and comparatively few samples (or data points). Feature selection methods are presented in three ways

1.Filter Method

2.Wrapper Method

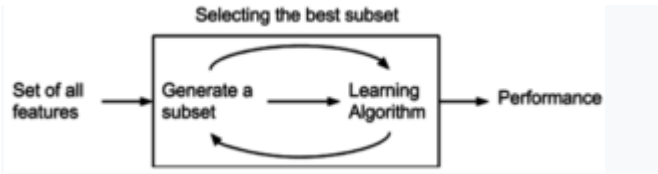
3.Embedded Method



Filter Method for feature selection

Filter type methods select variables regardless of the model. They are based only on general features like the correlation with the variable to predict. Filter methods suppress the least interesting variables. The other variables will be part of a classification or a regression model used to classify or to predict data. These methods are particularly effective in computation time and robust to overfitting. However, filter methods tend to select redundant variables because they do not consider the relationships between variables. Therefore, they are mainly used as a pre-process method.

Wrapper Method:



Wrapper Method for Feature selection

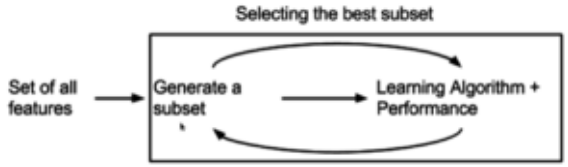
Wrapper methods evaluate subsets of variables which allows, unlike filter approaches, to detect the possible interactions between variables.

The two main disadvantages of these methods are:
1. The increasing over fitting risk when the number of observations is insufficient.
2. The significant computation time when the number of variables is large.

Embedded Method :

Embedded methods have been recently proposed that try to combine the advantages of both previous methods. A learning algorithm takes advantage of its own variable selection process and performs feature selection and

classification simultaneously.



Embedded Method for Feature selection

V. METHODOLOGY USED

Fig.1 illustrates the overall flow of the experiment. First, classification results are noted without doing any kind of feature selection techniques (Co-relation based Feature Selection, Wrapper, and Information Gain) on data sets. Then, using three feature selection techniques, separate feature subsets are chosen for each technique. The selected features are passed to the classifiers and results are noted. In this paper, Naïve Bayes (NB), classifiers are used for the classification of data sets. The data is preprocessed by using naïve bayes.

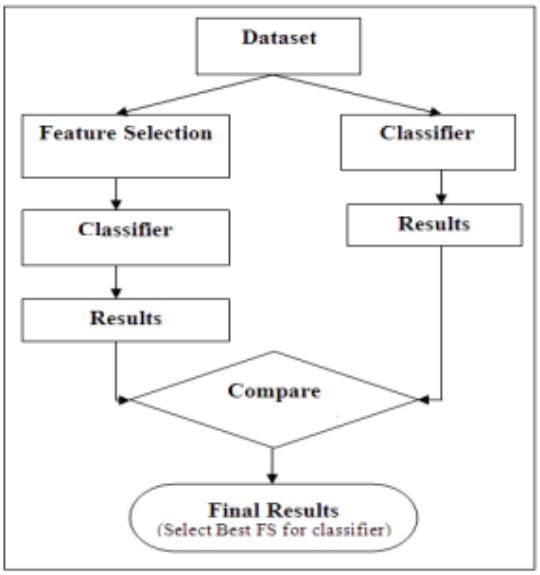


Fig 1. Overall flow of the experiment

5.1 Naive Bayes

The Naïve Bayes classifier is based on Bayesian probability model. If a class is provided, Naïve Bayes classifier assumes that the value of one feature is independent of any other feature. It is based on the mathematical principle of Conditional probability. If n attributes are given, independent assumptions made by the Naïve Bayes classifier is 2n!. A Conditional probability model for above classifier is given as:

P (Ci | x)

where, Ci is the ith class and x is input vector.

In this case, class variable C is conditional on several features
variable x = x1,…….., xn

Using Bayes theorem equation (2) can be written as:

$$P(C_i \mid x) = \frac{P(C_i) * P(x \mid C_i)}{P(x)} \quad \text{....[2]}$$

VI. EXPERIMENTAL RESULTS

As per the methodology discussed earlier, experiments are performed on two data sets with and without feature selection. Analysis of results is done using following evaluation metrics.

a. Evaluation Measure

There various parameters to measure the performance, among them only Accuracy, True Positive Rate (TPR), False Positive Rate (FPR) are considered in this paper. Accuracy is the proportion of the total number of predictions that were correct. TPR gives the proportion of correctly classified instances out of total classified. FPR shows the proportion of negative cases that were incorrectly classified as positive.

$$AC = \frac{TN+TP}{TP+TN+FP+FN} \quad (3)$$

$$TPR = \frac{TP}{TP+FN} \quad (4)$$

$$FPR = \frac{FP}{FP+TN} \quad (5)$$

To compute these metrics, first confusion matrix for the data set is then computed using these values into above equations to find Accuracy, TPR, and FPR.

VII. CONCLUSION AND FUTURE WORK

In this paper, effect of feature selection on supervised learning based classifiers is compared. Accuracy, TPR and FPR are used as an evaluation metric for comparison. From the experimental results it has been observed that used method improves accuracy and True Positive Rate and minimizes False Positive Rate.

The future work will include combining different classifier using ensemble method and applying feature selection technique before classification.

REFERENCES

[1] Vipin Kumar and Sonajharia Minz, “Feature Selection: A literature Review,” Smart Computing Review, vol. 4, no. 3, 2014.

[2] Lei Yu leiyu and Huan Liu, “Feature Selection for HighDimensional Data: A Fast Correlation-Based Filter Solution,” ICML, 2003.

[3] Rajdev Tiwari, Manu Pratap Singh, “Correlation-based Attribute Selection using Genetic Algorithm,” International Journal of

Computer Applications, pp (0975 – 8887), Volume 4– No.8, August 2010.

[4] Jasmina novakovic, Perica strbac, Dusan bulatovic, "Toward Optimal Feature Selection using Ranking Methods and Classification
Algorithms," pp 119-135, 2011

[5] Shengyan Zhou, Iagnemma K., "Self-supervised learning method for unstructured road detection using Fuzzy Support Vector Machines",International Conference on Intelligent Robots and Systems, pp. 1183–1189, IEEE, 2010.

[6] Techo, J. Nattee, C. Theeramunkong, T., "A CorpusBased Approach For Keyword Identification Using Supervised Learning Techniques", 5th International Conference On Electrical Engineering/ Electronics, Computer, Telecommunications And Information Technology, Ecti-Con, Vol.1, pp. 33 – 36, 2008.

[7] Cecille Freeman, Dana Kuli´Cand OtmanBasir,"FeatureSelected Tree-Based Classification", IEEE Transactions on Cybernetics, Vol. 43, No. 6, pp. 1990-2004, December 2013.

[8] Wald R., Khoshgoftaar T.M., Napolitano A., "Stability Of Filter- And Wrapper-Based Feature Subset Selection", IEEE 25th International Conference On Tools With Artificial Intelligence, pp. 374 – 380, 2013.

[9] Dr.Saurabh Mukherjee, Neelam Sharma, "Intrusion Detection Using Naive Bayes Classifier With Feature Reduction", Elsevier, pp. 119 – 128, 2012.

[10] Amirasayed A. Aziz, Ahmad Taherazar, Mostafa A. Salamaand Sanaa El-Ola Hanafy, "Genetic Algorithm With Different Feature Selectiontechniques For Anomaly Detectors Generation", IEEE Federated Conference On computer Science And Information Systems, pp. 769– 774, 2013.

[11] Altidor W.,Khoshgoftaar T.M., Van Hulse J., "An Empirical Study On Wrapper-Based Feature Ranking", 21st International Conference On tools With Artificial Intelligence, pp.75-82, IEEE 2009.

[12] Yuguang Huang,Lei Li, "Naive Bayes classification algorithm based on small sample set", IEEE International Conference on Cloud Computing and Intelligence Systems, pp. 34 – 39, 2011.

[13] GuoQiang,"An Effective Algorithm for Improving the Performance of Naive Bayes for Text Classification", IEEE Second International Conference on Computer Research and Development, pp. 699–701, 2010.

[14] Menkovski, V.,Efremidis, S., "Oblique Decision Trees using embedded Support Vector Machines in classifier ensembles", 7th IEEE International Conference on Cybernetic Intelligent Systems (CIS), pp. 1-6, 2008.

