



## Classification of Radar Returns from Ionosphere Using NB-Tree and CFS

Aung Nway Oo

University of Information Technology, Myanmar

### ABSTRACT

This paper present an experimental study of the different classifiers namely Naïve Bayes (NB) and NB-Tree for classification of radar returns from Ionosphere dataset. Correlation-based Feature Subset Selection (CFS) is also used for attribute selection. The purpose is to achieve the efficient result for classification. The comparison of NB classifier and NB-Tree is done based on Ionosphere dataset from UCI machine learning repository. NB-Tree classifier with CFS gives better accuracy for classification of radar returns from ionosphere.

**KEYWORD:** *classification, feature selection, NB, NB-Tree, CFS*

### 1. INTRODUCTION

Classification is one of the important decision making tasks for many real problems. Classification will be used when an object needs to be classified into a predefined class or group based on attributes of that object. Classification is a supervised procedure that learns to classify new instances based on the knowledge learnt from a previously classified training set of instances. It takes a set of data already divided into predefined groups and searches for patterns in the data that differentiate those groups supervised learning, pattern recognition and prediction. Typical Classification Algorithms are Decision trees, rule-based induction, neural networks, genetic algorithms and Naïve Bayes, etc.

Feature selection is one of the key topics in data mining; it improves classification performance by searching for the subset of features. In problem of high dimensional feature space, some of the features may be redundant or irrelevant. Removing these redundant or irrelevant features is very important;

hence they may deteriorate the performance of classifiers. Feature selection involves finding a subset of features to improve prediction accuracy or decrease the size of the structure without significantly decreasing prediction accuracy of the classifier built using only the selected features [7].

In this paper, we evaluate the classification of ionosphere dataset using WEKA data mining tool. The paper is organized as follows. Overview of Ionosphere is described in section 3. We outline overview of Naïve Bayes in section 3 and NB-Tree classifier in Section 4. Section 5 presents the Correlation-based Feature Subset Selection (CFS). The experimental results and conclusions are presented in Section 6 and 7 respectively.

### 2. OVERVIEW OF IONOSPHERE

The ionosphere is defined as the layer of the Earth's atmosphere that is world ionized by solar and cosmic radiation. It lies 75-1000 km (46-621 miles) above the Earth. (The Earth's radius is 6370 km, so the thickness of the ionosphere is quite tiny compared with the size of Earth.) Because of the high energy from the Sun and from cosmic rays, the atoms in this area have been stripped of one or more of their electrons, or "ionized," and are therefore positively charged. The ionized electrons behave as free particles. The Sun's upper atmosphere, the corona, is very hot and produces a constant stream of plasma and UV and X-rays that flow out from the Sun and affect, or ionize, the Earth's ionosphere. Only half the Earth's ionosphere is being ionized by the Sun at any time [8].

During the night, without interference from the Sun, cosmic rays ionize the ionosphere, though not nearly as strongly as the Sun. These high energy rays originate from sources throughout our own galaxy and the universe -- rotating neutron stars, supernovae, radio galaxies, quasars and black holes. Thus the ionosphere is much less charged at night-time, which is why a lot of ionospheric effects are easier to spot at night – it takes a smaller change to notice them.

The ionosphere has major importance to us because, among other functions, it influences radio propagation to distant places on the Earth, and between satellites and Earth. For the very low frequency (VLF) waves that the space weather monitors track, the ionosphere and the ground produce a “waveguide” through which radio signals can bounce and make their way around the curved Earth.

### 3. NAÏVE BAYES CLASSIFIER

Naïve Bayesian classifiers assume that there are no dependencies amongst attributes. The Bayesian Classification represents a supervised learning method as well as a statistical method for classification. Assumes an underlying probabilistic model and it allows us to capture uncertainty about the model in a principled way by determining probabilities of the outcomes. It is made to simplify the computations involved and, hence is called "naive" [2]. This classifier is also called idiot Bayes, simple Bayes, or independent Bayes [3]. NB is one of the 10 top algorithms in data mining as listed by Wu et al. (2008).

Let  $C$  denote the class of an observation  $X$ . To predict the class of the observation  $X$  by using the Bayes rule, the highest posterior probability of

$$P(C|X) = \frac{P(C) P(X|C)}{P(X)} \quad (1)$$

should be found. In the NB classifier, using the assumption that features  $X_1, X_2, \dots, X_n$  are conditionally independent of each other given the class, we get

$$P(C|X) = \frac{P(C) \prod_{j=1}^n P(X_j|C)}{P(X)} \quad (2)$$

In classification problems, Equation (2) is sufficient to predict the most probable class given a test observation.

### 4. NAÏVE BAYES TREE CLASSIFIER

The NB-Tree provides a simple and compact means to indexing high-dimensional data points of variable dimension, using a light mapping function that is computationally inexpensive. The basic idea of the NB-Tree is to use the Euclidean norm value as the index key for high-dimensional points. NBTree is a hybrid algorithm with Decision Tree and Naïve-Bayes. In this algorithm the basic concept of recursive partitioning of the schemes remains the same but here the difference is that the leaf nodes are naïve Bayes categorizers and will not have nodes predicting a single class [5].

Although the attribute independence assumption of naive Bayes is always violated on the whole training data, it could be expected that the dependencies within the local training data is weaker than that on the whole training data. Thus, NB-Tree [4] builds a naive Bayes classifier on each leaf node of the built decision tree, which just integrate the advantages of the decision tree classifiers and the Naive Bayes classifiers. Simply speaking, it firstly uses decision tree to segment the training data, in which each segment of the training data is represented by a leaf node of tree, and then builds a naive Bayes classifier on each segment. A fundamental issue in building decision trees is the attribute selection measure at each non-terminal node of the tree. Namely, the utility of each non-terminal node and a split needs to be measured in building decision trees. NB-Tree significantly outperforms Naïve Bayes in terms of classification performance indeed. However, it incurs the high time complexity, because it needs to build and evaluate Naive Bayes classifiers again and again in creating a split.

### 5. CORRELATION-BASED FEATURE SUBSET SELECTION (CFS)

CFS evaluates and ranks feature subsets rather than individual features. It prefers the set of attributes that are highly correlated with the class but with low intercorrelation [6]. With CFS various heuristic searching strategies such as hill climbing and best first are often applied to search the feature subsets space in reasonable time. CFS first calculates a matrix of feature-class and feature-feature correlations from the training data and then searches the feature subset space using a best first. Equation 1 (Ghiselli 1964) for CFS is

$$\text{Merit}_s = \frac{\overline{krcf}}{\sqrt{k+(k-1)\overline{rff}}} \quad (3)$$

Where  $Merit_s$  is the correlation between the summed feature subset  $S$ ,  $k$  is the number of subset feature,  $\overline{rcf}$  is the average of the correlation between the subsets feature and the class variable, and  $\overline{rff}$  is the average inter-correlation between subset features.

## 6. EXPERIMENTAL RESULTS

This radar data was collected by a system in Goose Bay, Labrador. This system consists of a phased array of 16 high-frequency antennas with a total transmitted power on the order of 6.4 kilowatts. The targets were free electrons in the ionosphere. "Good" radar returns are those showing evidence of some type of structure in the ionosphere. "Bad" returns are those that do not; their signals pass through the ionosphere. Received signals were processed using an autocorrelation function whose arguments are the 66% of dataset is used for training. The dataset is collected from UCI repository. The WEKA data mining tool is used for evaluation and testing of algorithm. The following tables show the experimental results of different classifiers time of a pulse and the pulse number. There were 17 pulse numbers for the Goose Bay system. Instances in this database are described by 2 attributes per pulse number, corresponding to the complex values returned by the function resulting from the complex electromagnetic signal. The dataset contains 351 instances and 35 attributes [1].

**Table1. Accuracy results of classifiers**

Naïve Bayes	NB Tree	NB-Tree+CFS
82.3529 %	88.2353 %	89.916 %

**Table2. Test results of Classifications**

	Naïve Bayes	NB-Tree	NB Tree+CFS
Correctly Classified Instances	98	105	107
Incorrectly Classified Instances	21	14	12
Kappa statistic	0.6474	0.7583	0.7936
Mean absolute error	0.1647	0.1294	0.1144
Root mean squared error	0.3866	0.3118	0.2978
Relative absolute error	34.3174 %	26.9712 %	23.8381 %
Root relative squared error	75.2862 %	60.7123 %	58.0024 %

## 7. CONCLUSION

The paper proposed the comparative analysis of Naïve Bayes and NB-Tree classifier. The experimental showed that the accuracy of Naïve Bayes classifier is 82.3529 % and NB Tree is 88.2353 %. According to the evaluation results, the highest accuracy 89.916% is found in NB-Tree classifier using Correlation-based Feature Subset Selection.

## REFERENCES

1. <http://archive.ics.uci.edu>
2. J. Han and M. Kamber, Data Mining: Concepts and Techniques. Morgan-Kaufmann Publishers, San Francisco, 2001.
3. T. Miquelez, E. Bengoetxea, P. Larranaga, "Evolutionary Computation based on Bayesian Classifier," Int. J. Appl. Math. Comput. Sci. vol. 14(3), pp. 335 – 349, 2004.
4. R. Kohavi, "Scaling up the accuracy of naive-bayes classifiers: A decision-tree hybrid," ser. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. AAAI Press, 1996, pp. 202–207.
5. R. Kohavi. "Scaling Up the Accuracy of Naive-Bayes Classifiers: a Decision-Tree Hybrid" Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 1996.
6. I.H. Witten, E. Frank, M.A. Hall "Data Mining Practical Machine Learning Tools & Techniques" Third edition, Pub. – Morgan kouffman.
7. D. Koller, and M. Sahami, "Toward optimal feature selection" In proceedings of international conference on machine learning, Bari, (Italy) pp. 284-92, 1996.
8. <http://solar-center.stanford.edu>