Available online at www.isas.org.in
JOURNAL OF THE INDIAN SOCIETY OF
AGRICULTURAL STATISTICS 63(2) 2009 189-197

Discretization based Support Vector Machines for Classification

Anshu Bharadwaj[1*] and Sonajharia Minz[2]
[1]*Indian Agricultural Statistics Research Institute, New Delhi*
[2]*Jawaharlal Nehru University, New Delhi*

(Received: December 2008, Revised: August 2009, Accepted: August 2009)

SUMMARY

Discrete values have important roles in data mining and knowledge discovery. They are about intervals of numbers which are more concise to represent and specify, easier to use and comprehend as they are closer to the knowledge level representation than continuous ones. Discretization is the process of quantizing continuous attributes. It has been used for decision tree classifier. The success of discretization can significantly extend the borders of many learning algorithms. Support Vector Machines (SVM) are the new generation learning system based on the latest advances in statistical learning theory. SVM is the recent addition to the toolbox of data mining practitioners and are gaining popularity due to many attractive features, and promising empirical performance. In this paper, a new approach to classify data using SVM classifier, after discretization is looked into. The classification results achieved after discretization based SVM are much better than the classification results using simple SVM in terms of accuracy. To acquire the better accuracy, discretization has been instrumental This is an attempt to extend the boundaries of discretization and to evaluate its effect on other machine learning techniques for classification namely, support vector machines.

Keywords: Support vector machines, Discretization, Radial basis function, Confusion matrix, Boolean reasoning based method, Entropy based method.

1. INTRODUCTION

Support vector machine (SVM) is a novel learning method based on statistical learning theory. SVM is a powerful tool for solving classification problems with small samples, nonlinearities and local minima, and is of excellent performance. To address the discretization process of continuous-valued features in an efficient and proper manner has always been an important issue for any machine learning technique. SVM is a widely used method for classification in variety of applications. The results of the experiment conducted in this study clearly show that the classification results using SVM are better when discretization process is undertaken before the classification. However, various methods of discretization affect the classification accuracy. Therefore, it is important to decide a method to improve the performance of the SVM model. The points in the dataset that fall on the bounding planes of the hyperplane in a SVM are called support vectors. They play an important role in the theory as well as in the classification task at the prediction stage. Vapnik (1974, 1979, 1998) has shown that if the training vectors are separated without errors by an optimal hyperplane, the expected error rate on a test sample is bounded by the ratio of the expectation of the support vectors to the number of training vectors. Since this ratio is independent of the dimension of the problem, and, if one can find a good set of support vectors, good generalization is guaranteed. We aim at a good generalization from the classification task that we have carried out using SVM after discretization. Even though

Corresponding author : Anshu Bharadwaj
E-mail addresses : anshu@iasri.res.in, ans_dix@yahoo.com

SVMs can handle continuous attributes, its performance can be significantly improved by replacing a continuous attribute with its discretized values. Data discretization is defined as a process of concerting continuous data attribute values into a finite set of intervals and associating with each interval some specific data value. There are no restrictions on discrete values associated with a given data interval except that these values must induce some ordering on the discretized attribute domain. Discretization significantly improves the quality of discovered knowledge (Catlett 1991), (Pfahringer 1995) and also reduces the running time of various data mining tasks such as association rule discovery, classification, and prediction. In this study, we have also used two spatial datasets. These datasets have been used to examine the performance of the classification technique used for classical data mining task on it. Spatial datasets differ from non-spatial datasets as they have spatial aspects involved in them. Here the spatial datasets used are in the vector format. The spatial attributes in the spatial datasets used, are latitudes and longitudes. The datasets have been considered just to experiment with it using discretization based SVM classifier. In this paper, we describe discretization methods and compare them according to accuracy of the classification results. We focus our work to find out the significance of discretization before classification using SVM.

Section 2 of this paper gives the overview about the data preprocessing step of data mining along with the need of discretization and detailed description of the applied discretization methods. Section 3 deals with the basic concepts of support vector machines and its parameters in detail. Section 4 describes the confusion matrix as the performance evaluation measure for the classifier. Section 5 gives the detail about the experimental setup, summary of the data used and its analysis. Section 6 contains the results and Section 7 draws the conclusions.

2. DATA PREPROCESSING

Data preprocessing describes any type of processing performed on raw data to prepare it for another processing procedure. Commonly used as a preliminary data mining practice, data preprocessing transforms the data into a format that will be more easily and effectively processed for the purpose of the user. It is the most critical step in data mining process that includes the preparation and transformation of the initial dataset. Raw data are seldom used for data mining. Many transformations may be needed to produce more useful features for selected data mining methods such as prediction or classification. Discretization of numerical attributes is one of the important data preprocessing techniques. In this paper we have discretized the data before classifying it using SVM, as the preprocessing step.

2.1 Why Discretization?

There are many advantages of using discrete values over continuous one. Discrete features are closer to knowledge level representation (Simon 1981) than continuous ones. Data is reduced and simplified using discretization. For both users and experts, discrete features are easier to understand, use and explain. As reported by Dougherty *et al*. (1995), discretization makes learning more accurate and faster. In general, obtained results using discrete features are usually more compact, shorter and more accurate than using continuous ones; hence the results can be more closely examined, compared, used and reused. In addition to the many advantages of having discrete data over continuous one, a suite of classification learning algorithms can only deal with discrete data.

2.2 Discretization Methods

A large number of machine learning and statistical techniques can only be applied to datasets composed entirely of nominal variables. However, a very large proportion of real datasets include continuous variables, that is variables measured at intervals or ratio level. One solution to this problem is to partition numeric variables into sub-ranges and treat each such sub-range as a category. This process of partitioning continuous variables into categories is usually termed as discretization. Transformation of a continuous attribute to a categorical attribute involves two subtasks, deciding how many categories to have and determining how to map the values of the continuous attribute. They are then divided into n intervals specifying n – 1 split points. In the second, rather trivial step, all the values in one interval are mapped to the same categorical value. Therefore, the problem of discretization is one of deciding how many split points to choose and where to place them. The result can be represented either as a set of intervals $\{(x_0, x_1), (x_1, x_2),..., (x_{n-1}, x_n)\}$, where

x_0 and x_n may be $+\infty$ or $-\infty$ respectively, or equivalently, as a series of inequalities $x_0 < x \le x_1, \ldots, x_{n-1} < x < x_n$. A variety of discretization methods have been developed along different lines due to different needs: supervised vs. unsupervised; dynamic vs. static; global vs. local; splitting (top-down) vs. merging (bottom-up), and direct vs. incremental.

2.2.1 Supervised and unsupervised discretization methods

Data can be supervised or unsupervised depending on whether it has class information. Likewise, supervised discretization considers class information while unsupervised discretization does not; unsupervised discretization is seen in earlier methods like equal-width and equal-frequency. In unsupervised methods, continuous ranges are divided into sub ranges by the user specified width (range of values) or frequency (number of instances in each interval). This may not give good results in cases where the distribution of the continuous values is not uniform. Furthermore, it is vulnerable to outliers as they affect the ranges significantly (Catlett 1991). To overcome this shortcoming, supervised discretization methods were introduced and class information is used to find the proper intervals by cut-points. In this study, we have used both the unsupervised and supervised methods of discretization to discretize the datasets before applying SVM. We have selected two of the most popular and widely used methods of supervised discretization and similarly one unsupervised method of discretization is also selected. The supervised discretization methods used are described briefly for a better understanding of the methods.

The methods used are

1. Unsupervised : Equal-frequency
2. Supervised: Entropy based and Boolean reasoning based methods

2.2.1.1 *Entropy-based discretization method*

Entropy based discretization method uses a minimal entropy heuristic for discretization of continuous attributes. This method tries to find a binary cut for each attribute. Following a method introduced by Fayyad and Irani (1993), the minimal entropy criteria can also be used to find multi-level cuts for each attribute. The algorithm uses the class information entropy of candidate partitions to select binary boundaries for discretization.

2.2.1.2 *Boolean reasoning/rough set based discretization method*

The method that we have discussed (entropy based) discretize only one attribute at a time. It may therefore introduce more cuts than is absolutely necessary for discerning between the decision classes. Nguyen and Nguyen (1996), and Nguyen and Skowron (1995, 1997) have introduced a supervised method that considers all of the attributes simultaneously and creates consequently fewer cuts. Their method is developed with basis in rough sets methods and Boolean reasoning.

3. SUPPORT VECTOR MACHINE

The foundations of SVM based on statistical learning theory were developed by Vapnik (1998) and Burges (1998) to solve the classification problem. The SVM is the recent addition to the toolbox of data mining practitioners and are gaining popularity due to many attractive features, and promising empirical performance. They are a new generation learning system based on the latest advances in statistical learning theory. The formulation embodies the Structural Risk Minimization (SRM) principle, which has been shown to be superior (Gunn *et al.* 1997), to traditional Empirical Risk Minimization (ERM) principle, employed by conventional neural networks. SRM minimizes an upper bound on the expected risk, as opposed to ERM that minimizes the error on the training data. It is this difference which equips SVM with a greater ability to generalize, which is the goal in statistical learning. SVM belongs to the class of supervised learning algorithms in which the learning machine is given a set of examples (or inputs) with the associated labels (or output values). Like in decision trees, the examples are in the form of attribute vectors, so that the input space is a subset of R^n. SVMs create a hyperplane that separates two classes (this can be extended to multi class problems). While doing so, SVM algorithm tries to achieve maximum separation between the classes. Separating the classes with a large margin minimizes a bound on the expected generalization error. By "minimum generalization error", it means that when new examples (data points with unknown class values) arrive for classification, the chance of making error in the prediction (of the class to which it belongs) based on the learned classifier (hyperplane) should be minimum. Intuitively, such a classifier is one which achieves maximum separation-margin between the classes. The two planes parallel to the plane are called bounding planes. The distance between these bounding planes is called margin and by SVM "learning", i.e. finding hyperplane which maximizes this margin. The points (in the dataset) falling on the bounding planes are called the support vectors. "Machine" in Support Vector Machine is nothing but the algorithm (Soman *et al.* 2006). SVM was designed initially as binary classifier i.e. it classifies the data into two classes but researchers have extended its boundaries to be a multi-class classifier. SVM was first introduced as a training algorithm (Boser *et al.* 1992) that automatically tunes the capacity of the classification function maximizing the margin between the training patterns and the decision boundary (Cristianini and Shaw-Taylor 2000). This algorithm operates with large class of decision functions that are linear in their parameters but not restricted to linear dependences in the input components. For the computational considerations, SVM works well on two important practical considerations of classification algorithms i.e. speed and convergence.

3.1 SVM and its Parameter

To construct an optimal hyperplane, SVM employees an iterative training algorithm, which is used to minimize an error function. According to the form of the error function, SVM models can be classified into two distinct groups

1. SVM for classification
2. SVM for regression

In this study we are dealing with classification problem, so the SVM for classification is described here.

For SVM, training involves the minimization of the error function

$$\frac{1}{2}w^T w + C\sum_{i=1}^{N}\xi_i$$

subject to the constraints

$$y_i(w^T\phi(x_i) + b) \geq 1 - \xi_i \text{ and } \xi_i \geq 0, \quad i = 1, \ldots, N$$

where C is the capacity constant or the model complexity, w is the vector of coefficients, b a constant and ξ_i are parameters for handling non-separable data (inputs). The index i labels the N training cases. Note that $y \in \pm 1$ is the class label and x_i is the independent variable. The kernel ϕ is used to transform data from the input (independent) to the feature space. It should be noted that larger the C, the more the error is penalized. Thus, C should be chosen with care to avoid over fitting.

3.2 Radial Basis Function

There are a number of kernels that can be used for support vector machine models. These include Linear, Polynomial, Radial Basis and Sigmoid.

A radial basis function (RBF) is a real-valued function whose value depends only on the distance from the origin, so that $\phi(x) = \phi(\|x\|)$; or alternatively on the distance from some other point c, called a *center*, so that $\phi(x, c) = \phi(\|x - c\|)$. Any function ϕ that satisfies the property $\phi(x) = \phi(\|x\|)$ is a radial function. The norm is usually to use RBF, although other distance functions are also possible. The following expression describes the RBF kernel for SVM

$$\phi = \exp\{-\gamma\, |x - c|^2\}, \text{ where } \gamma > 0$$

where γ is called the RBF kernel parameter. The RBF kernel is the most popular kernel type due to its localized and finite response across the entire range of real x-axis.

4. PERFORMANCE EVALUATION MEASURE: CONFUSION MATRIX

Evaluation of the performance of the classification model is based on the counts of the test records correctly or incorrectly predicted by the model. These counts are tabulated in a table called Confusion Matrix. Table 1 depicts the confusion matrix for a binary classification model. Each entry f_{ij} in this table denotes the number of records from class i predicted to be of class j. For instance f_{01} is the number of records from

Table 1. Confusion Matrix

		Predicted Class	
		Class = 1	Class = 0
Actual Class	Class = 1	f_{11}	f_{10}
	Class = 0	f_{01}	f_{00}

class 0 incorrectly predicted as of class 1. Based on the entries in the table the total number of correct prediction made by the model is $(f_{11}+f_{00})$ and the total number of incorrect predictions is $(f_{10} + f_{01})$.

5. EXPERIMENT AND ANALYSIS

Using the discretization methods before applying SVM, we clearly see that discretization simplifies data (continuous values are quantized into intervals) without sacrificing data consistency much (only a few inconsistencies occur after discretization). We have to evaluate the ultimate objective of discretization of the datasets before applying SVM—whether discretization helps improve the performance of learning and understanding of learning results. The kernel used for training is RBF. The improvement is measured in terms of the classification accuracy. The evaluation of the performance of the classification model is done using Confusion Matrix. As a general approach of solving classification problems, each dataset is split into two datasets training sample dataset and test sample dataset. Training dataset consists of the records having class labels and is used to build the classification model whereas the test dataset contains records without class labels and is used to validate the model, built by training dataset. Though discretization is usually a needless preprocess step for SVM, which can deal with continuous and hybrid attributes directly, it has been still attractive to use discretized datasets because it has improved the classification performance and reduced the training time.

5.1 Data Description

Four datasets are selected from various sources, with all numeric features and varying data sizes. The datasets used in the study are Boston2, CIMMYT and Hurricane. Boston2 and Hurricane datasets are from public domain i.e. UCI repository available online and the CIMMYT dataset is a live dataset. The live dataset used for this comparative study is Rice dataset. This dataset is in vector data format of spatial databases. Spatial attributes in the datasets are latitudes and longitudes. The data is obtained from Resource Conservation Technologies from Rice-Wheat Consortium, CIMMYT, India. Here only a small part of data with 50 observations has been used for illustration purpose. There are 4 classes in which the data has to be classified. Number of attributes in the dataset is 10 that includes the latitudes and longitudes being spatial attributes of the dataset. The CIMMYT dataset is modified as two different datasets, first by considering all the variables (latitudes and longitudes) as CIMMYT1, and secondly by ignoring the spatial variables, i.e. dropping the variables containing the spatial information, as CIMMYT2. The results may be different and the conclusions drawn here may change with the full set of data. The sample dataset is from different districts of Western Uttar Pradesh and contains different treatments (i.e. different types of seed cultivation), the spatial aspect of the location (longitudes and latitudes) with various biometrical characters of the rice plant. The task is to classify the varieties in different classes.

The second dataset is Boston2. This example illustrates an analysis of the Boston house price data (Harrison and Rubinfeld 1978) that was reported by (Lim *et al.* 1997). Median prices of housing tracts were classified as Low, Medium, or High on the dependent variable price. There was one categorical predictor, Cat1, and 12 ordered predictors, Ord1 through Ord12. The complete data set contains a total of 1012 cases.

The third data used in this study is the Hurricane data. This data was originally obtained from Atlantic tropical cyclone "best" track and intensity records managed by the Tropical Prediction Center (Formerly the National Hurricane Center, Jarvinen *et al.* 1984), where "best" refers to an accurate assessment of storm location based on a post analysis of available data. The dataset extends back to 1886 and includes all tropical cyclones, that reached tropical storm strength. A storm has latitude and longitude coordinates and maximum sustained winds every 6 hours during the storm's existence. Data are most reliable after 1944 when the US Air force began aircraft reconnaissance missions to investigate individual storm. The dataset has six independent and one dependent variable. Each storm contains the Julian day, the latitude and the longitude for initial depression and initial hurricane stages, that is the day and position for which the storm was first reported as a tropical depression and a hurricane, respectively. Day D and day H are the Julian days on which the storm first reached depression and hurricane strengths respectively, long D and long H are the initial depression and hurricane longitudes, respectively, lat D and long H are the initial depression and hurricane latitudes, respectively; TROP and BARO are tropical only and baroclinically influenced hurricanes, respectively. Summary of datasets can be found in Table 2.

Table 2. Summary of datasets

S.No.	Dataset	Total number of instances	Number of features	Number of classes	source of data
1.	CIMMYT	50	10	4	CIMMYT, INDIA
2.	CIMMYT1	50	8	4	CIMMYT, INDIA
3.	BOSTON2	1012	13	3	STATISTICA
4.	HURRICANE	209	6	2	STATISTICA

5.2 Experimental Set-up

A total of three discretization methods (equal-frequency (unsupervised), entropy and Boolean reasoning (supervised)), have been used to study the effect of discretization on classification results. Experimental design is given in Table 3.

The datasets are split into train and test datasets, then the discretization algorithms (entropy based, Boolean reasoning and equal-frequency) are used to discretize the train dataset one by one. Once the train dataset is discretized using any of the algorithms, the same cuts points (Liu *et al.* 2002) or intervals generated for the train dataset using the particular discretization algorithm are saved in a file and the same cuts points are then used to discretize the test dataset, for test dataset the class labels are not used during discretization. Once the data has been split (into train and test datasets) and discretized, the original dataset (i.e. the undiscretized data) has not been used anywhere in the study. The experiment was conducted with 8 runs each for each dataset. Each run means, to classify the data at split of different seed value. Seed values used for splits are 1000, 900, 800, 750, 600, 500, 350, 100. The seed values were randomly selected. Classification using SVM was carried out on the discretized datasets so that the results can be compared and the effect of the discretization on SVM can be studied. CIMMYT and Hurricane datasets are spatial datasets in vector format with latitude and longitude as spatial attributes.

Table 3. Experimental design

S.No.	Experimental Steps
1.	Split each dataset into Training (70%) and Test Sample (30%) datasets of each complete data. Split of the dataset is carried out using simple random sampling
2.	Discretize the train and test datasets separately. (Use all the three methods, i.e. equal- frequency, entropy and boolean reasoning for discretizing the train dataset and then use the same cuts to discretize the test dataset without using class labels, for all datasets)
3.	Apply SVM for classification on the datasets (both train and test separately) using 10x10 fold cross-validation.
4.	Compare the classification results with the SVM classification results without discretization. Classification accuracy of train and test datasets is compared separately.

Table 4. Hurricane training data sample after discretization

Attribute	Continuous values	Intervals after discretization
DAYDEPR	224, 239, 285, 231, 266, 257, 237, 243, 245, 364	[*, 270), [279, 287), [288, *)
LONDEPR	45.7, 25.6, 78.2, 19, 62.2, 61.7, 74.9, 67.7, 56.4, 50.9	[*, 58.2), [78.1, 80.6), [62.1, 62.7), [60, 61.8), [70.4, 77.3), [66.8, 67.8)
LATDEPR	12.2, 12.3, 14.3, 14.6, 14.4, 16, 24, 19.3, 11.2, 22.1	[*, 14.2), [14.4, 14.9), [15.4, 16.5), [22.7, 24.1), [19.2, 19.8), [21.5, 22.3)
DAYHUR	228, 245, 287, 240, 269, 258, 239, 244, 250, 365	[*, 270), [282, 290), (292, *)
LONHURR	62.5, 58.8, 82.1, 64.7, 73.8, 67, 76.2, 69.5, 70.8, 55.2	[62.3, 62.9), [58.5, 59.3), [80.5, 82.2), [62.9, 65.0), [73.7, 74.0), [66.7, 67.5), [75.5, 76.4), [69.2, 69.7), [70.4, 70.9), [54.6, 55.3)
LATHURR	15.4, 14.1, 18.5, 21.9, 24.5, 20.9, 28.9, 24.8, 22.2, 20.6	[*, 24.7), [28.9, 29.0),[24.7, 25.2)

A sample (10 data points) of Hurricane dataset after discretization is given in Table 4 for better understanding. The table shows the discretized training data sample using Entropy method, it includes original continuous values and the intervals into which the data has been divided after discretization.

6. RESULTS AND DISCUSSION

The results are shown in Table 5. Each result consists of the classification accuracy of the SVM learning technique with and without discretization of the datasets.

SVM classification using discretization shows that the results obtained are improved and better classification accuracy is attained. The parameter of SVM decision function i.e. capacity or model complexity does not get affected by discretization as discretization process works on the dataset rather than the model. Similarly, the parameter of the RBF kernel i.e. γ also remain unaffected by the discretization of the datasets before applying SVM.

It is also observed from the results given in the above table that the supervised discretization algorithms are better than the unsupervised discretization algorithm as the classification accuracies using the supervised discretization algorithms are better than the unsupervised discretization algorithm. Out of the supervised discretization algorithms, Boolean reasoning based algorithm is performing better in attaining better classification accuracy. It is known that supervised discretization is better than the unsupervised discretization but we have used one method of unsupervised discretization to compare the difference it brings to the classification accuracy if the data is classified after unsupervised discretization as compared to the supervised discretization. It is observed that for one of the datasets, Boston2, the classification accuracy attained after discretization using unsupervised method (equal frequency) is higher than the classification accuracy attained after supervised classification using entropy based method. Although this accuracy is less than the accuracy attained using the other supervised discretization method i.e. Boolean Reasoning.

Discretization yields the reduction in unique tuples by assigning the discretized value of the attribute to the objects whose numeric value lies in the corresponding discrete interval. Thus, we could observe that there had been a reduction in the number of support vectors per class during classification of the discretized dataset. The number of support vectors was reduced to give better classification accuracy.

Table 5. Results of classification using SVM

Dataset	SVM with discretization						Without discretization	
	Entropy		Boolean reasoning		Equal frequency		Original	
	Train	Test	Train	Test	Train	Test	Train	Test
Boston2	90.25	89.27	98.16	98.94	94.18	94.86	79.85	79.01
Hurricane	91.37	88.88	97.26	85.18	88.12	88.88	89.72	87.93
CIMMYT	84.57	80.00	92.14	53.15	76.57	69.00	61.85	76.85
CIMMYT1	78.23	56.89	85.71	60.00	68.73	63.00	57.33	74.00

Table 6. Comparison of classifiers in terms of classification accuracy

Dataset	ANN		SVM		Discretization based SVM					
					Entropy		Boolean reasoning		Equal frequency	
	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
Boston2	75.43	78.78	79.85	79.01	90.25	89.27	98.16	98.94	94.18	94.86
Hurricane	80.91	83.89	89.72	87.93	91.37	88.88	97.26	85.18	88.12	88.88
CIMMYT	36.13	55.44	61.85	76.85	84.57	80.00	92.14	53.15	76.57	69.00
CIMMYT 1	35.65	42.06	57.33	74.00	78.23	56.89	85.71	60.00	68.73	63.00

Hurricane dataset has earlier been classified using a method explained in (Elsner *et al.* 1996). The method used is Partially Adaptive Classification Trees (PACT) algorithm (Shih 1993) based on linear discriminant analysis (Mardia and Bibby 1979) and tree structured classification method (Brieman *et al.* 1984). This algorithm gives a classification accuracy of around 90% which is less than the accuracy attained by discretization based SVM classification. The classification accuracy attained by supervised discretization method based SVM for hurricane dataset are 91.37 and 97.14 respectively for entropy based method and Boolean reasoning based method. Similarly in Minz and Dixit (2007), these four datasets have been classified using Artificial Neural Network and SVM and it is seen that result obtained by using discretization based SVM are much better than the results obtained by the earlier two methods. Comparative results are shown in Table 6.

7. CONCLUSION

The study was undertaken with an aim to explore the effects of discretization on support vector machines. Although data discretization has been a step for applying machine learning technique of classification such as decision tree but it has not been tried for support vector machines classifier, the reason being its ability to handle continuous and hybrid data unlike the decision tree algorithm ID3, which can handle only discrete datasets for classification. Therefore, we tried to explore the effect of discretization of the datasets before applying SVM classifier. This was done with the aim of attaining better classification accuracy without disturbing or distorting the parameters (C and Gamma) of SVM. The results clearly indicate that the accuracies of discretization based SVM are better as compared to the classification accuracy without SVM of the same datasets when they were classified without getting discretized. We have also observed that the supervised discretization algorithm works better than the unsupervised discretization algorithm. Among supervised discretization algorithm also, Boolean Reasoning based method performs best. This study establishes that discretization can be used for SVM classifiers also.

ACKNOWLEDGEMENT

The authors thank the referee for the critical evaluation and suggestions that led to commendable improvement in this paper. The authors are also thankful to CIMMYT, India, for providing the rice data for this study.

REFERENCES

Boser, B.E., Guyon, I.M. and Vapnik, V.N. (1992). A training algorithm for optimal margin classifiers. *Proc. of 5th Annual Workshop on Computer Learning Theory,* Pittsburgh, PA: ACM, 144-152.

Brieman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1984). *Classification and Regression Trees.* Wadsworth, 358.

Burges, J.C. (1998). *A Tutorial on Support Vector Machines for Pattern Recognition.* Data Mining and Knowledge Discovery, **2**, 121-167.

Catlett, J. (1991). On changing continuous attributes into ordered discrete attributes. *Proc. Fifth European Working Session on Learning,* Springer-Verlag, Berlin, 164-177.

Cristianini, N. and Shaw-Taylor, J. (2000). *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods.* Cambridge University Press, Cambridge.

Dougherty, J., Kohavi, R. and Sahami, M. (1995). Supervised and unsupervised discretization of continuous features. *Proc. Twelfth International Conference on Machine Learning,* Morgan Kaufmann, Los Altos, CA, 194-202.

Elsner, J.B., Lehmiller, G.S. and Kimberlain, T.B. (1996). Objective classification of atlantic hurricanes. *J. Climate,* **9**, 2880-2889.

Fayyad, U.M. and Irani, K.B. (1993). Multi-interval discretization of continuous-valued attributes for classification learning. In: *Machine Learning. 13th IJCAI,* **2**, Morgan Kaufmann, Chambery, France, 1022-1027.

Gunn, S.R., Brown, M. and Bossely, K.M. (1997). Network Performance Assessment for Neurofuzzy Data Modelling. In: *Intelligent Data Analysis,* **1208**, Lecture Notes in Computer Science (X. Liu, P. Cohen and M. Berthold (ed.), 313-323.

Harrison, D. and Rubinfeld, D.L. (1978). Hedonic prices and the demand for clean air. *J. Environ. Econ. Manage.,* **5**, 81-102.

Jarvinen, B.R., Neumann, C.J. and Davis, M.A.S. (1984). A Tropical cyclone data tape for the North Atlantic Basin.

1886-1983: Contents, Limitations and Uses. *NOAA Tech. Report. NWH NHC 22*, 21.

Lim, T.S., Loh, W.Y. and Shih, Y.S. (1997). An empirical comparison of decision trees and other classification methods. *Technical Report 979, Department of Statistics, University of Wisconsin, Madison,* Wisconsin, http://www.stat.wisc.edu/~limt/compare.ps

Mardia, K.V. and Bibby, J.T.M. (1979). *Multivariate Analysis.* Hartcourt Brace and Company, 521.

Minz, S. and Dixit, A. (2007). Neural networks and support vector machines: A comparative study for spatial data classification. *Proc. of the National Conference on Methods and Models in Computing 2007, New Delhi,* December 13-14, 2007.

Nguyen, S.H. and Skowron, A. (1995). Quantization of real value attributes: Rough set and boolean reasoning approach. *Proc. Int.Workshop Rough Sets and Soft Computing* at *2nd Joint Conf. Information Sciences (JCIS'95)*, Durham, NC, 34-37.

Nguyen, S.H. and Nguyen, S.H. (1996). Some efficient methods for rough set. *Proc. of the Sixth International Conference, Information Processing and Management of Uncertainity in Knowledge-Based Systems (IPMU-96),* 2, Granada, Spain, July 1-5.

Nguyen, S.H. and Skowron, A. (1997). Quantization of real value attributes: Rough set and boolean reasoning approach. *Bulletin of International Rough Set Society,* **1(1),** 5-16.

Pfahringer, B. (1995). Compression-based discretization of continuous attributes. *Proc. Twelfth International Conference on Machine Learning,* Morgan Kaufmann, *San Francisco,* CA, 456-563.

Shih, Y.S. (1993). Tree structured classification. *Ph.D. Dissertation, University of Wisconsin-Madison.*

Simon. H.A. (1981). *The Sciences of the Artificial.* **2**, Cambridge, MA, MIT.

Soman, K.P., Diwakar, S. and Ajay, V. (2006). *Insight into Data Mining: Theory and Practice.* Prentice Hall of India Pvt. Ltd.

Vapnik, V. (1974). *Theory of Pattern Recognition.* Nauka, Moscow.

Vapnik, V. and Chervonenkis, A. (1979). *Theory of Pattern Recognition.* Nauka, Moscow.

Vapnik, V. (1998). *Statistical Learning Theory.* Wiley, New York.