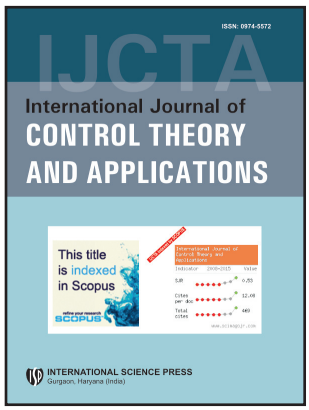

International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 9 • Number 44 • 2016

Feature Selection based on Correlation and Information Gain for Plant Leaf Classification

A.V. Senthil Kumar[a]

[a]*Professor, Department of Post Graduate and Research in Computer Applications, Hindusthan College of Arts and Science, Coimbatore-641 028, Tamil Nadu, India. Email: avsenthilkumar@yahoo.com*

Abstract: The major problem in plant classification is extraction and reduction of features. Feature subset selection is of great importance in the field of data mining. The attribute reduction is one of the key processes for knowledge acquisition. Some data set is multidimensional and larger in size. If that data set is used for classification it may end with wrong results and it may also occupy more resources especially in terms of time. Most of the features present are redundant and inconsistent and affect the classification. In order to improve the efficiency of classification these redundancy and inconsistency features must be eliminated. In this paper, a method of plant leaf classification is proposed and explores the efficacy of feature subset selection using correlation and information gain. Evaluation demonstrates that information gain helps select features that result in considerable improvements on MLP with Batch Back propagation algorithm classifier performance with an accuracy of 94.81% for 9 species.

Keywords: Multi-Layer Perceptron (MLP), Information Gain, Correlation based Feature Selection (CFS), Leaf Classification.

1. INTRODUCTION

About 7,000 species of plants have been cultivated for utilization in human history. Through photosynthesis, plants provide the oxygen we breathe and the food we eat and are thus the foundation of most life on Earth. The great diversity of varieties resulting from human and ecosystem interaction guaranteed food for the survival and development of human populations throughout the world in spite of pests, diseases, climate fluctuations, droughts and other unexpected environmental events. Presently, only about 30 crops provide 95% of human food energy needs [1]. It is also found that about 68 percent of evaluated plant species are threatened with extinction [2]. Hence it is the need of the hour to start a plant protection database.

Artificial Neural Networks (ANN) consists of many interconnected processing elements linked by weighted connections inspired by biological neurons. Learning in a biological system is through training or exposure to an input/output data set where a training algorithm adjusts weights iteratively. ANN are good pattern recognition

engines and robust classifiers deciding about imprecise input data (Master, 1993) trained using Back Propagation Algorithm. Figure 1 reveals a block diagram representing an ANN.

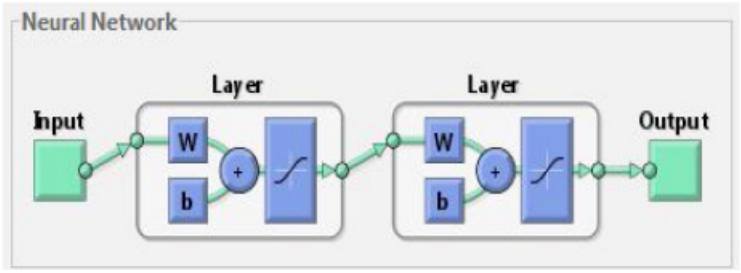

Figure 1: Artificial Neural Network

Relative experimental studies have consistently shown Information Gain based feature selection to result in good classifier performance [3]. The principal task of feature selection applications is to improve the performance criterion such as accuracy.

2. RELATED WORKS

Algorithms for feature selection fall into two broad categories: wrappers that use the learning algorithm itself to evaluate the usefulness of features and filters that evaluate features according to heuristics based on general characteristics of the data. For application to large databases, filters have proven to be more practical than wrappers because they are much faster. This paper describes a fast, correlation- based filter algorithms that can be applied to continuous and discrete problems [4].

Feature subset selection is of great importance in the field of data mining. The high dimension data makes testing and training of general classification methods difficult. In the present paper two filters approaches namely Gain ratio and Correlation based feature selection have been used to illustrate the significance of feature subset selection for classifying Pima Indian diabetic database (PIDD). The feature subset obtained is then tested using two supervised classification method namely, Back propagation neural network and Radial basis function network. Experimental results show that the feature subsets selected by CFS filter resulted in marginal improvement for both back propagation neural network and Radial basis function network classification accuracy when compared to feature subset selected by information gain filter [5].

Support Vector Machines, one of the new techniques for pattern classification, have been widely used in many application areas. The kernel parameters setting for SVM in training process impacts on the classification accuracy. Feature selection is another factor that impacts classification accuracy. The objective of this research is to simultaneously optimize the parameters and feature subset without degrading the SVM classification accuracy. We present a genetic algorithm approach for feature selection and parameters optimization to solve this kind of problem [6].

This paper presents an application of Information Gain (IG) attribute evaluation to the classification of the sonar targets with C4.5 decision tree. C4.5 decision tree has inherited ability to focus on relevant features and ignore irrelevant ones, but such method may also benefit from independent feature selection. In our experiments, IG attribute evaluation significantly improves

C4.5 decision tree. This research also shows that feature selection helps increase computational efficiency while improving classification accuracy [7].

The attribute reduction is one of the key processes for knowledge acquisition. Some data set is multidimensional and larger in size. If that data set is used for classification it may end with wrong results and

it may also occupy more resources especially in terms of time. Most of the features present are redundant and inconsistent and affect the classification. In order to improve the efficiency of classification these redundancy and inconsistency features must be eliminated. This paper discusses an algorithm based on discernibility matrix and Information gain to reduce attributes [8].

3. MATERIALS AND METHODS

Feature Selection

Correlation based feature selection (CFS) measures the worth or the merit of the subset of features. The algorithm works on the basis "good feature subset contains highly correlated features but uncorrelated with each other". The CFS searches the features in greedy step wise. It either performs a greedy backward or forward search through the attribute subsets. Starting with nothing/all attributes or from an arbitrary point in the space the feature subset is formed. The algorithm stops when there is a decrease in evaluation on addition/deletion of any remaining attributes. CFS can also produce a ranked list of attributes on the basis of the order that attributes are selected [9].

Information gain is a well known feature selection method. It is a reasonable objective to use for selecting feature. Using information gain will help to reduce the noise which is due to irrelevant features for influencing classifier. Information gain (IG) measures amount of information in bits about class prediction, when the only information available is presence of a feature and corresponding class distribution [10].

Information gain measure selects test attribute at every node in the tree [11]. The attribute with highest information gain (greatest entropy reduction) is selected as test attribute for current node. This attribute minimizes information needed to classify samples in resulting partitions.

In classification setting, higher entropy (more disorder) corresponds to sample of mixed label collection. Lower entropy corresponds to a case where there are pure partitions. In information theory, sample D's entropy is defined as follows:

$$\mathrm{H(D)} = -\sum_{i=1}^{k} \mathrm{P}(c_i|\mathrm{D}) \log_2 \mathrm{P}(c_i|\mathrm{D})$$

where, $\mathrm{P}(c_i|\mathrm{D})$ is probability of a data point in D being labeled with class c_i, and k is number of classes. $\mathrm{P}(c_i|\mathrm{D})$ is estimated directly from the data as follows:

$$\mathrm{P}(c_i|\mathrm{D}) = \frac{|\{x_j \in \mathrm{D} | x_j \text{ has label } y_j = c_i\}|}{|\mathrm{D}|}$$

Also weighted entropy of a decision/split are defined as follows:

$$\mathrm{H}(\mathrm{D_L}, \mathrm{D_R}) = \frac{|\mathrm{D_L}|}{|\mathrm{D}|} \mathrm{H}(\mathrm{D_L}) + \frac{|\mathrm{D_R}|}{|\mathrm{D}|} \mathrm{H}(\mathrm{D_R})$$

where, D was partitioned into $\mathrm{D_L}$ and $\mathrm{D_R}$ due to split decision. Finally, information gain for a given split is defined as:

$$\mathrm{Gain(D, D_L, D_R)} = \mathrm{H(D)} - \mathrm{H(D_L, D_R)}$$

In other words, Gain is anticipated entropy reduction caused by knowing an attribute's value.

Multi-Layer Perceptron

An important class of artificial neural network (ANN) is multilayer perceptron (MLP) which schematically depicted in Figure 2. This network consists of an input layer, one or more hidden layers of computation nodes and an output layer of computation nodes. MLP neural network uses error back-propagation algorithm in order to learn its parameters (weights)[12]. This algorithm consists of two passes through the different layers of the network: forward and backward pass. In forward pass, an input pattern is applied to the network to produce a set of outputs while the weights are all fixed. In l'th layer the output is computed as follows:

$$x_j^l = \sum_{i=1}^{m_{l-1}} y_i^{l-1} w_{ij}^{l-1,l}, \; y_j^l = f(x_j^l), \tag{6}$$

$$j = 1, 2, ..., n_l, \; l = 1, 2, ..., \mathrm{N}$$

where, x_j^l is the input to the j'th neuron in l'th layer, y_j^l is the output of the j'th neuron in l'th layer, is the weight of the connection between i'th neuron in layer and j'th neuron in l'th layer, f is an activation function, n_l is the number neurons in l'th layer and N is the number of layers.

During backward pass the error (the difference between the network outputs and the true values) is propagated back from the output to the connection weights and updates the weights to minimize the prediction error. The most common error function is mean squared error (MSE):

$$\mathrm{E} = \frac{1}{2} \sum_{i=1}^{n_\mathrm{N}} (t_i - y_i^{n_\mathrm{N}})^2 \tag{7}$$

where, t_i is the target value for the i'th output and n_N is the number of neurons in the output layer. Then the following algorithm is used to update the weights:

$$w_{ij}^{l-1,l}(k+1) = w_{ij}^{l-1,l}(k) - \frac{\lambda \cdot \Delta \mathrm{E}}{\nabla w_{ij}^{l-1,l}}(k) + \beta[w_{ij}^{l-1,l}(k) - w_{ij}^{l-1,l}(k+1)] \tag{8}$$

where, λ is the learning rate, β is momentum factor which helps to improve the performance of back-propagation algorithm.

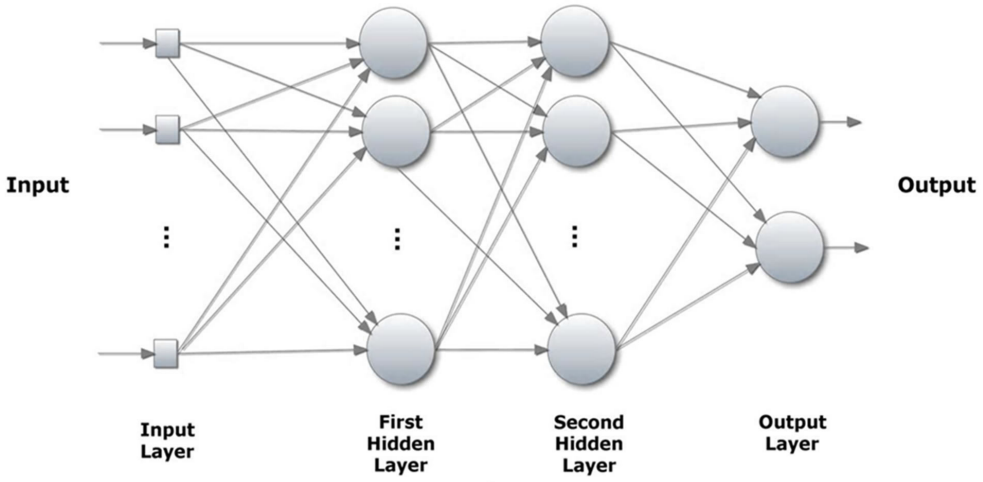

Figure 2: Multilayer Perceptron

MLP with Batch Backpropagation Algorithm based Learning

Patterns are presented to a network before learning in batch training. In almost all cases, several passes must be made through training data. In batch training protocol, all training patterns are presented first and corresponding weight updates summed up; only then are actual network weights updated. This process is iterated till a stopping criterion is met [13]. Patterns need not be selected randomly, as weights are updated only after patterns are presented. In contrast, every weight change during continuous training reduces error for that instance, but decrease or increases error on training set as a whole. Hence, batch training ensures more accuracy.

Batch learning proceeds as follows:

begin initialize nH, w, criterion q, h, $r \neg 0$

do $r \neg r + 1$ (increment epoch)

$m \neg 0$; D$wji \neg 0$; D$wkj \neg 0$

do $m \neg m + 1$

$xm \neg$ select pattern

D$wji \neg$ D$wji + \eta x_i \delta_j$; D$wkj \neg$ D$wkj + \eta y_j \delta_k$

until $m = n$

$wji \neg wji$ + Dwji; $wkj \neg wkj$ + Dwkj

until $\| \nabla J(w) \| < \theta$

return w

end

4. RESULTS AND DISCUSSION

Nine species of plant leaves were selected with 15 samples each species. Sample image of the plant leaves used is shown in Figure 3. In the first part Correlation based features were extracted and classified using MLP. Secondly Correlation based Features and Information Gain based features were extracted and classified using MLP with Batch Back propagation.

Figure 3: Sample leaves

Table 1
Classification Accuracy

Technique Used	*Classification accuracy*
CFS and MLP	91.85%
CFS and MLP with Batch Backpropagation learning	93.33%
IG and MLP with Batch Backpropagation learning	94.81%

Table 2
Precision, recall and f Measure

Technique Used	*Precision*	*Recall*	*f Measure*
CFS and MLP	0.92	0.919	0.919
CFS and MLP with Batch Backpropagation learning	0.935	0.933	0.933
IG and MLP with Batch Backpropagation learning	0.949	0.948	0.947

The classification accuracy obtained is given in Table 1. Table 2 tabulates the precision, recall and fMeasure for various technique used.

Experimental results show that the feature subsets selected by Information Gain (IG) resulted in marginal improvement for MLP back propagation neural network when compared to feature subset selected by content based (CFS).

5. CONCLUSION

In this study, correlation and information gain based features were extracted and classified using MLP with Batch Back propagation. Nine species of plant leaves were selected with 15 samples each species. Experimental results show that the proposed information gain based features and MLP with Batch Back Propagation learning method as classifier achieved a better performance of classification accuracy, precision, recall and fmeasures when compared to other MLP learning methods. The classification accuracy achieved by the proposed method is 94.81% which is the best accuracy when compared to other learning methods. The output obtained is promising with low relative error for a nine class problem. Further work needs to be done to improve the classification accuracy by proposing feature reduction techniques and for a larger dataset.

REFERENCES

[1] http://www.fao.org/biodiversity/components/plants/en/

[2] http://www.biologicaldiversity.org/programs/biodiversity/elements_of_biodiversity/extinction_crisis/

[3] Cover, T. M., & Thomas, J. A. (2012). *Elements of information theory*. John Wiley & Sons.

[4] Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning, Mark A. Hall,

[5] Comparative study of attribute selection using gain ratio and correlation based feature selection, Asha Gowda Karegowda, A.S. Manjunath & M.A. Jayaram, July-December 2010, Volume 2, No. 2, pp 271-277.

[6] A GA-based feature selection and paramenters optimization for support vector machines, cheng-lung Huang, chieh-Jen Wang, Expert Systems with Applications, 31 (2006) 231-240.

[7] Using Information Gain Attribute Evaluation to classify Sonar Targets, 17th Telecommunications forum TELFOR 2009, Serbia, Belgrade, November 24-26, 2009.

[8] Feature Selection based on Information Gain, International Journal of Innovative Technology and Exploring Engineering, ISSN: 2278-3075, Volume-2, Issue-2, January 2013.

[9] International Journal of Future Computer and Communication, Vol. 2, No. 3, June 2013 Plant Leaf Classification Using Soft Computing Techniques, C. S. Sumathi and A. V. Senthil Kumar.

[10] T. M. Cover and J. A. Thomas, Elements of information theory. John Wiley & Sons. 2012.

[11] R. C. Barros, M. P. Basgalupp, A. C. de Carvalho, and A. A. Freitas, "Towards the automatic design of decision tree induction algorithms," in Proc.13th annual conference companion on Genetic and evolutionary computation (ACM), July 2011 pp. 567-574.

[12] Application of Artificial Intelligence (AI) Modeling in Kinetics of Methane Hydrate Growth American Journal of Analytical Chemistry Vol.4 No.11(2013), Article ID:39386,7 pages DOI:10.4236/ajac.2013.411073.

[13] L.Fu, H. H.Hsu, and J.C. Principle, "Incremental backpropagation learning networks. Neural Networks," IEEE Transactions 7(3), pp. 757-761, 1996.