ISSN (Print) : 0974-6846
Indian Journal of Science and Technology, *Vol 8(33), DOI: 10.17485/ijst/2015/v8i33/71203, December 2015* ISSN (Online) : 0974-5645

A Classification Approach using Multi-Layer Perception with Back-Propagation Algorithm

K. Fathima Bibi[1*] and M. Nazreen Banu[2]

[1]Department of Computer Science, Bharathiar University, Coimbatore – 641046, Tamil Nadu, India; kfatima72@gmail.com
[2]Department of MCA, MAM College of Engineering, Tiruchirappalli – 621105, Tamil Nadu, India; nazreentech@gmail.com

Abstract

In data mining feature subset selection is a preprocessing step in classification that lessens dimensionality, eliminates unrelated data, increases accuracy and improves unambiguousness. The next step in classification is to produce enormous amount of rules from the reduced feature set from which high class rules are chosen to build effectual classifier. In this paper, Information Gain (IG) has been used to rank the features. Multi-Layer Perception (MLP) with back-propagation reduces features to achieve higher accuracy in classification. Artificial Neural Networks (ANN) classifier is used for classification. We handle the discretization of continuous valued features by dividing the series of values into a limited number of subsections. Wine Recognition data set taken from the UCI machine learning repository is used for testing. Original 13 features are drawn in classification. The thirteen features are reduced to five features. Experimental results show that the accuracy in training dataset is 98.62% and in the validation dataset is 96.06%. The accuracy difference between 13 features and 5 features in the training data is 5.54% and in validation data is 2.00%. We then build a Decision Tree and concentrate on discovering significant rules from the reduced data set that provide better classification.

Keywords: Back-Propagation, Classification, Decision Tree, Feature Subset Selection, Multi-Layer Perception

1. Introduction

Feature subset selection benefits social readers to know a learnt model. It can radically diminish the exploration space for a learner. Several readings have shown that a learner can ignore many features with little or no loss in classification accuracy[1-3].

1.1 Information Gain Feature Ranking

This is a stress-free and speedy technique for feature ranking[4,5]. This procedure deals with the entropy of the class before and after detecting a feature. The variation in the entropy gives a degree of the information gained as of that feature[6]. A final evaluation of this is used in feature selection.

1.2 Artificial Neural Networks

Artificial Neural Networks is cheered by efforts to simulate living neural structures. The human brain contains mostly of nerve cells named neurons, linked jointly with other neurons by means of a fiber named axons. Axons are used to send nerve impulses from one neuron to another when the neurons are moved. A neuron is linked to the axons of other neurons through dendrites. The point that makes connection between dendrite and an axon is named synapse. The Artificial Neural Networks are characterized in two clusters Feed-Forward and recurrent[6]. Feed-Forward networks cover many classes of Artificial Neural Networks the most notable of which is the MLP Artificial Neural Network[7]. Multi-Layer is Feed-Forward neural network trained with the standard back-propagation algorithm. It is a supervised network so they necessitate a looked-for response, and hence broadly used in pattern classification.

1.3 Decision Tree

Decision Tree learners recursively divided instances by ranking features allowing to how much they decline the

* *Author for correspondence*

variety of the classes in the split cells. As learning grows, lesser instances are presented to study the next-sub-tree.

This is an outline of the rest of the paper. Section 2 describes literature review. Section 3 the materials used. Classifier evaluation measures are illustrated in Section 4. Section 5 describes the experimentation and results. Finally, Section 6 concludes this work.

1.4 Literature Review

Proposed feature weighing technique[8]. The proposed technique was tested with wine dataset where 10 of 13 features were selected and an accuracy of 88% was obtained.

Introduced an association rule based methodology for feature reduction[9]. Experimental results showed that 12 of 13 were selected from the wine dataset which achieved 68.98% accuracy.

Described a harmony based feature subset selection[10]. The effort achieved an accuracy of 95.90 from 10 out of the 13 features of the wine dataset.

Proposed a cosine similarity measure for feature reduction[11]. The measure achieved an accuracy of 96.63 from the minimum set of 8 features.

2. Materials

The wine dataset is taken from the UCI machine learning repository. These data are the outcomes of a chemical investigation of wine grown-up in the similar region in Italy from three different cultivars. The learning determined 13 features found in each the three types of wines. The features are:- Flavanoids, OD280/OD315 of diluted wines, Alcohol, Magnesium, Total Phenols, Hue, Nonflavanoid phenols, Color Intensity, Alcalinity of ash, Proanthocyanins, Malic Acid, Ash and Proline. The class feature specifies the different cultivars. There are a sum 178 instances, 59 instances from Class 1, 71 instances from Class 2 and 48 instances from Class 3. All the features limited in the dataset are continuous.

3. Classifier Evaluation Measures

The inspiration for this learning arises from the supremacy of ANN classification algorithm and it uses IG as the principle to first-rate a feature. We compute IG for every feature. IG provide feature A with respect to the class feature B drop in ambiguity about the value of B when we know the value of A, I (B; A). The ambiguity about the value of B when we distinguish the value of A is given by the entropy of B given A, H (B/A). IG (B; A) = H (B) – H (B/A). The grade of the features is then done with respect to the values of IG in a descending order. In the existing effort, we handle the discretization of continuous valued features by dividing the series of values into a finite number of subsets. Figure 1 depicts dividing the series of values into a limited number of subsets and Table 1 represents the subsets of each feature.

Table 1. Subsets of each feature

S. No.	Feature	Subset	Range
1	X1	S1	{11.0-11.9}
		S2	{12.0-12.9}
		S3	{13.0-13.9}
		S4	{14.0-14.9}
2	X2	S1	{1.0-1.9}
		S2	{2.0-2.9}
		S3	{3.0-3.9}
		S4	{4.0-4.9}
3	X3	S1	{1.0-1.9}
		S2	{2.0-2.9}
		S3	{3.0-3.9}
4	X4	S1	{11.0-15.0}
		S2	{16.0-20.0}
		S3	{21.0-25.0}
		S4	{26.0-30.0}
5	X5	S1	{51.0-75.0}
		S2	{76.0-100.0}
		S3	{101.0-125.0}
		S4	{126.0-150.0}
6	X6	S1	{1.0-1.9}
		S2	{2.0-2.9}
		S3	{3.0-3.9}
7	X7	S1	{0.1-0.9}
		S2	{1.0-1.9}
		S3	{2.0-2.9}
		S4	{3.0-3.9}
8	X8	S1	{0.1-0.5}
		S2	{0.6-1.0}
9	X9	S1	{0.1-0.9}
		S2	{1.0-1.9}
		S3	{2.0-2.9}
10	X10	S1	{1.0-5.0}
		S2	{6.0-10.0}
11	X11	S1	{0.0-0.9}
		S2	{1.0-1.9}
12	X12	S1	{0.0-0.9}
		S2	{1.0-1.9}
		S3	{2.0-2.9}
		S4	{3.0-3.9}
13	X13	S1	{500-1000}
		S2	{1001-1500}
		S3	{1501-2000}

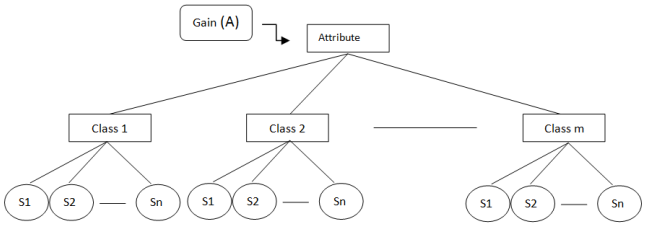

Figure 1. Dividing the series of values in a feature into subsets.

4. Experiment and Result

We adopt IG to sort the features according to their importance for classification in the following test. The ranking IGs of 13 features are shown in Table 2. We firstly used ANN without IG. Accuracy of the sets is made known in Table 3, Table 4 displays the confusion matrix and Table 5 is the thorough accuracy by Class. The accuracy deprived of IG, in the training dataset is 93.08% and in validation dataset is 94.67%. Then, we left the feature which has the lowest IG and use ANN to classify. If the accuracy is higher or equal than the accuracy with IG, we left the next feature which has the second lowermost IG and use ANN to classify. We complete the circlet when the accuracy is less than the accuracy without IG. We deduct the features until it has five features. The accuracy in training dataset is 98.62% and in the validation dataset is 96.06%. The accuracy difference between 13 features and 5 features in the training data is 5.54% and in validation data is 2.00%.

Table 2. Ranking of features

S. No.	Feature	Information Gain
1	X7- Flavanoids	0.9227
2	X12- OD280/OD315 of diluted wines	0.7415
3	X1- Alcohol	0.6161
4	X5- Magnesium	0.6101
5	X6- Total phenols	0.6043
6	X11- Hue	0.3442
7	X8- Nonflavanoid phenols	0.3032
8	X10- Color intensity	0.2753
9	X4- Alcalinity of ash	0.1994
10	X9- Proanthocyanins	0.1911
11	X2- Malic acid	0.1724
12	X3- Ash	0.1586
13	X13- Proline	0.1536

Table 3. Accuracy of the sets

Features	Training	Validation
13	93.08	94.67
5	98.62	96.06

Table 4. Confusion matrix

A	B	C	<-- classified as
11	0	0	A=1
2	7	1	B=2
0	0	12	C=3

Table 5. Thorough accuracy by class

Class	True Positives Rate	False Positives Rate
1	1	0.091
2	0.7	0
3	1.0	0.048

4.1 Ensembles of Decision Trees

Decision Trees (DTs) are unique thorough going non-linear classifier used in data mining. In a Decision Tree the feature space is fragmented into unique areas, consistent to the classes. Test on a feature relates to each internal node, ending of the test represents each branch and a class label for each class.

In our tests, the DTs were qualified on the subsets created by the feature selection algorithm as shown in Figure 2.

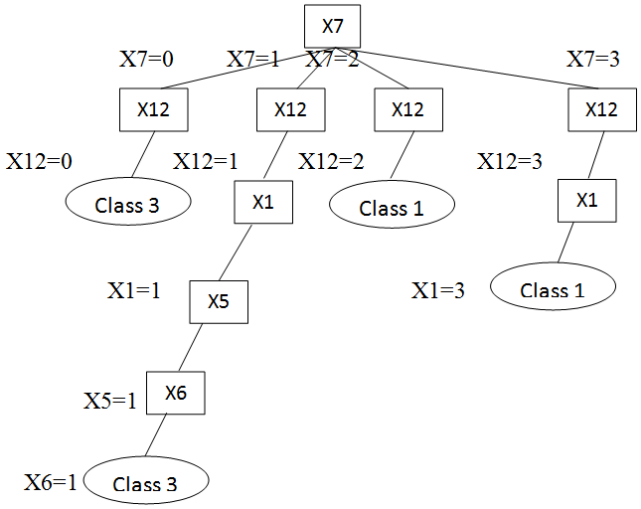

Figure 2. The Decision Tree.

4.2 Rule Extraction from Decision Tree

Afterward emerging the Decision Tree we have mined the rules from the tree. For our purpose the Decision Tree has been used to pull out initial rules from the dataset. Four noteworthy rules have been pruned.

- If X7 = 0 and X12 = 0 then = >Class 3.
- If X7 = 1 and X12 = 1 and X1 = 1 and X5 = 1 and X6 = 1 then = >Class 3.
- If X7 = 2 and X12 = 2 then = > Class 1.
- If X7 = 3 and X12 = 3 and X1 = 3 then = >Class 1.

5. Conclusion and Future Work

In this paper, we have concentrated on creating and applying effectual feature subset selection methods and mining rules from the reduced data set. The key objective of the study is to attain higher accuracy in classification and determining high quality rules. IG measure has been used to rank the features of wine data set. Multi-layer perception with back-propagation is used to remove the features. Artificial Neural Networks classifier is used for classification. We handle the discretization of continuous valued features by dividing the series of values into a limited number of subsets. We, then, built a Decision Tree and trimmed rules for prediction. Table 6 shows the improved accuracy of our work. This research shows that feature selection helps increase computational effectiveness while improving classification accuracy. Moreover, they lessen the complexity of the system by reducing the dataset. The experimental section conferred some decision rules predicting the different cultivars.

Table 6. Accuracy of the wine data set

S. No.	Previous Work	Reduced features	Accuracy of prediction	Error Rate
1	Feature Weighing Technique7	10(13)	88.00	--
2	Association Rule Methodology8	12(13)	68.98	--
3	Harmony Based Methodology9	10(13)	95.90	--
4	Cos. Sim. using J4810	8(13)	99.44	0.0606
	Cos. Sim. using Naive Bayes10	8(13)	96.63	0.1332
	Proposed Work			
5	Information Gain using ANN	5(13)	98.62	--

6. References

1. Holte RC. Very simple classification rules perform well on most commonly used datasets. Machine Learning. 1993 Apr; 11(1):63–90.
2. Kohavi R, John GH. Wrappers for feature subset selection. Artificial Intelligence. 1997 Dec; 92(1-2):273–324.
3. Hall M, Holmes G. Benchmarking feature selection techniques for discrete class data mining. IEEE Transactions on Knowledge and Data Engineering. 2003 Nov; 15(6):1437–47.
4. Dumais S, Platt J, Heckerman D, Sahami M. Inductive learning algorithms and representations for text categorization. Proceedings of the Seventh International Conference on Information and Knowledge Management, CIKM'98; 1998. p. 148–55.
5. Yang Y, Pedersen JO. A comparative study on feature selection in text categorization. International Conference on Machine Learning; 1997. p. 412–20.
6. Prochazka AP. Feed-forward and recurrent neural networks in signal prediction. IEEE. 4th International Corference on Computational Cybernetics; 2007. p. 93–6.
7. Zebardast B, Maleki I, Maroufi A. A novel multilayer perceptron Artificial Neural Network based recognition for Kurdish Manuscript. Indian Journal of Science and Technology. 2014 Mar; 7(3):343–51.
8. Quinlan JR. C4.5: Programs for machine learning. Morgan Kaufmann; San Mateo, CA. 1993. p. 235–40.
9. Ahmad W, Narayanan A. Feature weighing for efficient clustering. IEEE Sixth International Conference on Advanced Information Management and Services; Seoul. 2010 Nov 30-Dec 2. p. 236–42.
10. Shaharanee IZM, Jamil J. Features selection and rule removal for frequent association rule based classification. Proceedings of the 4th International Conference on Computing and Informatics, ICOCI 2013; 2013. p. 377–82.
11. Krishnaveni V, Arumugam G. Harmony search based wrapper feature subset method for 1_nearest neighbor classifier. 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering (PRIME); Salem. 2013 Feb 21-22. p. 24–9.
12. Bibi KF, Banu MN. Feature subset selection based on filter technique. IEEE International Conference on Computing and Communications Technologies; Chennai. 2015 Feb 26-27. p. 1–6.