# INTEGRATED ASSOCIATIVE CLASSIFICATION AND NEURAL NETWORK MODEL ENHANCED BY USING A STATISTICAL APPROACH

Linda Sara Mathew

Asst. Professor,
Department of Computer Science &Engineering,
MACE, Kothamangalam, Kerala, India

***ABSTRACT***

*Association rules is a novel data mining technique that has been mainly used for data description, exploration and prediction in knowledge discovery and decision support systems. The association rule mining algorithm is modified to handle the user-defined input constraints. Associative classification is provided with a large number of rules, from which a set of quality rules are chosen to develop an efficient classifier. Many attribute selection measures are used to reduce the number of generated rules. In this paper the pruning of rule sets is facilitated by chi squared analysis thereby only positively correlated rules are used in the classifier. Also the Neural Network Associative Classification system is used in order to improve the accuracy of the classifier. The trained network is then used to classify the future data. The performance of the Neural Network Associative Classification system is analyzed with the datasets from UCI machine learning repository.*

***KEYWORDS***

*Data mining, Associative Classification, Chi Square, gini index, Multilayer perceptron, Back propagation neural network.*

## 1. INTRODUCTION

Data mining techniques extract hidden predictive information from very large databases. The prediction of future trends and behaviors by data mining tools helps businesses to make proactive, knowledge-driven decisions. Data mining tools can answer time consuming business questions that were difficult to resolve [2] .They search for hidden patterns, finds the missing predictive information that lies outside the expectations of experts.

Big Data , the large collection of datasets in the areas of bioinformatics, social networks, sensor networks, Internet text and documents, biological, and other complex and often interdisciplinary fields, are collected daily and warehoused. In the year 2008 it has been claimed that Google processes 20 PB a day. Statistical Modeling and Knowledge discovery of Big data can be done with data mining and warehousing tools.

DOI : 10.5121/ijdkp.2013.3407 

Data mining techniques can be implemented rapidly on existing software and hardware platforms. When implemented on high performance client/server or parallel processing computers, data mining tools can analyze massive databases to answer questions like, "Which clients are most likely to respond to my next new product , and why?"

## 2. PROPOSED METHODOLOGY

The proposed face recognition method consists of different parts:

i. Feature transformation
ii. Associative Classification
iii. Use gini index as attribute selection measure. Further rule pruning is done by chi square test.
iv. Classification using neural network

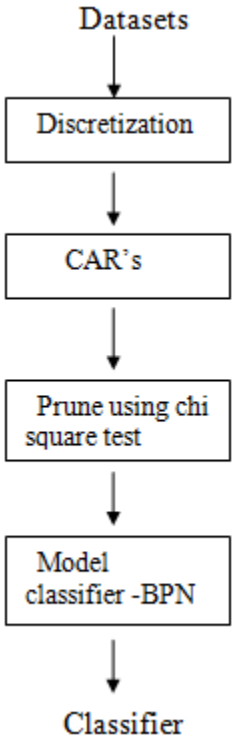


Figure 1: Architecture of the system

## 3. FEATURE TRANSFORMATION

The transformation methods enhance the richness of knowledge extracted from the data set. Discretization, filling missing data, feature bundling are the most common forms of data transformation [6]. Classification accuracy can be enhanced with these methods.

## 4. ASSOCIATIVE CLASSIFICATION

Associative classification is a technique that combines the methodology of association and classification [1].Class association rules are of the form P1 Λ P2 Λ…$P_l$ ➔ $A_{class}$ = C where P1,P2 are items and C is the class label.

The steps are :-

1. Mine the frequent itemsets

2.Generate association rules per class which satisfy confidence and support

3.Organize the rules to form rule-based classifier.

The ruleset is pruned ,ranked ,a model is constructed out of it and prediction is done. Large amount of computation is required for generating rich ruleset. The dataset is divided into training and test sets. This ruleset which is generated from the training dataset is used to build a model, which is used to predict test cases present in the test dataset [2].

## 5. RELATED WORKS

One of the earliest algorithms for associative classification is CBA. It uses a similar approach to apriori and the number of passes is equal to the length of the longest rule.CBA uses a heuristic method to construct the classifier and the rules are ordered according to the decreasing precedence based on their confidence and support. CMAR adopts a variant of the FP-growth algorithm to find the ruleset [7].CPAR uses a different approach which is based on FOIL [10].

Rule Support and Confidence are two measures of rule interestingness. They respectively reflect the usefulness and certainty of discovered rules. In general, association rules are considered interesting if they satisfy both a minimum support threshold and minimum confidence threshold. For example, a rule like "chest pain --> Heart attack" with sup=30% and conf=70% means that 30% of the patients who experienced a pain in the chest had a heart attack and that 70% of the patients who had chest pain also had a heart attack. That is, degree of support corresponds to statistical significance, while degree of confidence is a measure of the rule's strength [3].

## 6. ATTRIBUTE SELECTION

The popular attribute selection measures are Gini index, information gain and gain ratio. It provides a ranking for each attribute describing the training tuples. No measure is superior to the others. We can use it as a filter to reduce the number of itemsets ultimately generated. Only those attributes with minimum Gini index needs to be selected for rule generation [2].

$Gini(D) = 1-\sum_{i=1}^{m} pi2$

Table 1:Sample Training Dataset1

| Cust_id | Gender | Car_type | Credit_rating | Class |
|---|---|---|---|---|
| 1 | M | FAMILY | S | C0 |
| 2 | M | SPORTS | M | C0 |
| 3 | M | SPORTS | M | C0 |
| 4 | M | SPORTS | L | C0 |
| 5 | M | SPORTS | EL | C0 |
| 6 | M | SPORTS | EL | C0 |
| 7 | F | SPORTS | S | C0 |
| 8 | F | SPORTS | S | C0 |
| 9 | F | SPORTS | M | C0 |
| 10 | F | SPORTS | L | C0 |
| 11 | M | LUXURY | L | C1 |
| 12 | M | FAMILY | E | C1 |
| 13 | M | FAMILY | M | C1 |
| 14 | M | FAMILY | EL | C1 |
| 15 | F | LUXURY | S | C1 |
| 16 | F | LUXURY | S | C1 |
| 17 | F | LUXURY | M | C1 |
| 18 | F | LUXURY | M | C1 |
| 19 | F | LUXURY | M | C1 |
| 20 | F | LUXURY | M | C1 |
| 21 | F | LUXURY | L | C1 |

*Gini index for the Cust_id attribute = 0*
*Gini index for the Gender attribute = 0.48*
*Gini index for the Car_type attribute =0.163*

*Gini index for the Credit_rating attribute = 0.4915*

Out of the three attributes car_type attribute has the lowest gini index. When comparing with the other attributes car_type would be the better attribute.

## 7. RULE GENERATION

Each transaction in training dataset with k items can generate upto $2^k$ candidate itemsets. In our approach Gini index measure is first used to reduce the number of candidate itemsets. Only the candidate itemsets that includes the best split attribute value will be generated. In the dataset given below age is the attribute with minimum gini index.

Table 2: Sample Training dataset2

| Age | Car_type | Credit_rating | location | class |
|---|---|---|---|---|
| youth | sports | S | chicago | C0 |
| youth | sports | S | New york | C0 |
| Mid_aged | sports | S | chicago | C1 |
| old | family | S | chicago | C1 |
| old | Luxury | M | chicago | C1 |
| old | Luxury | M | New york | C0 |
| Mid_aged | Luxury | M | New york | C1 |
| youth | family | S | chicago | C0 |
| youth | Luxury | M | chicago | C1 |
| old | family | M | chicago | C1 |
| youth | family | M | New york | C1 |
| Mid_aged | family | S | New york | C1 |
| Mid_aged | sports | M | chicago | C1 |
| old | family | S | New york | C0 |

The following candidate itemsets can be generated :

youth → C0
youth Λ sports → C0
youth Λ S → C0
youth Λ chicago → C0
youth Λ sports Λ S → C0
youth Λ sports Λ chicago → C0
youth Λ sports Λ S Λ chicago → C0

Attribute selection based on gini index and the above approach for rule generation results in the generation of lesser number of rules and these rules are further pruned based on some measures on interestingness.

## 8. RULE PRUNING

Assessment of rules depends on several parameters and interestingness measures like confidence, support, lexicographical order of items etc. In CBA method, the rules are arranged based on their confidence value. In case of tie the rules are sorted based on their support and rule length. In lazy pruning a rule is pruned only if it misclassifies the data. The rule set are mainly classified into three, namely useful rules, harmful rules and spare rules. A rule that could segregate atleast one data item correctly from the data set is called useful rule and a rule which misclassifies a data item is a harmful rule and the rest of them are named spare rule which are not subjected to be pruned but can be used when needed. Though lazy pruning well suits for small datasets, but in large datasets, memory space and rule set quality issues becomes the major constraints[4].

The Apriori rule generation method and genetic algorithm were respectively used for generating the work rules and for selecting randomly a subset of rules[5].In this paper a statistical approach is used to prune the less significant rules [11].

## 9. RULE EVALUATION

Chi-square test is a well known discrete data hypothesis testing method. This method evaluates the correlation between two items i.e. it checks whether the two items are positively correlated or negatively correlated. Only positively correlated rules are used in the classifier [12]. Calculation of $\chi 2$ value of the rule $X \rightarrow Y$ is described as follows. Let support (X) = x, support (Y) = c, support(X U Y) = z and the number of transactions in the dataset equal N.

Table 2 shows the contingency of X and C: the upper parts are the expectation values under the assumption of their independence, and the lower parts are observational. Now, let E denote the value of the expectation under the assumption of independence and 0 the value of the observation. Then the $\chi 2$statistic is defined as follows:

Table 3: Chi Square Computation

| | C | ¬ C | Σmax |
|---|---|---|---|
| X | Nxc<br>Nz | N(x-xc)<br>N(x-z) | Nx |
| ¬ X | N(c-xc)<br>N(c-z) | N(1-x-c+xc)<br>N(1-x-c+z) | N(1-x) |
| Σ | Nc | N(1-c) | N |

$$\chi 2 = \frac{(Observed - Expected)^2}{Expected}$$

We can calculate $\chi 2$ using x, c, z and N of Table 3 as follows:

$$\chi 2 = \frac{N(z-xc)^2}{xc(1-x)(1-c)}$$

This has 1 degree of freedom. If it is higher than a cutoff value 3.84 at the 95% significance level, the rule is accepted or else the rule is rejected.

## 8. BACKPROPAGATION NEURAL NETWORK

A multilayer feed forward neural network consists of an input layer, one or more hidden layer and an output layer. The linear inseparability can be resolved using FFNN. The layered structure gets input from the previous layers [8].

The most popular neural network algorithm is backpropagation, a kind of gradient descent method. Backpropagation iteratively process the data set, comparing the network's prediction for each tuple with actual known target value to find out an acceptable local minimum in the NN weight space in turns achieves the least number of errors. Error can be defined as a root mean square of difference between real and desired outputs of NN [9].

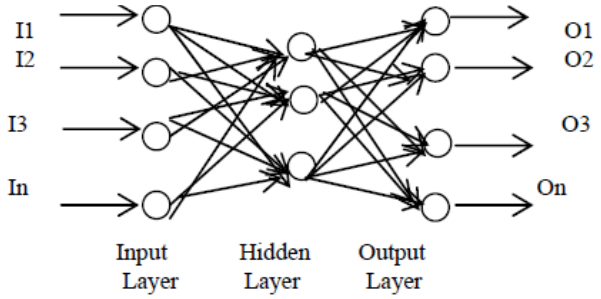


Figure 2: FFNN Architecture

The inputs are fed simultaneously into units making up the input layer. The number of units in the input layer mist be equal to the number of predicted attributes. Each node in the hidden layer is fully connected in such a way that each unit provides input to each unit in next forward layer. Output layer predicts the class and full connectivity with the hidden layer nodes is implemented. The weighted outputs of the last hidden layer are input to units making up the output layer which emits the network's prediction for given tuples. The nuber of output node should be same as that of number of classes of data. Generally Backpropagation Neural Network use sigmoid activation function.

Backpropagation algorithm is used for error computation and also for calculating the new weights for each unit in neural network. The network undergoes the training process iteratively in which weights are modified so as to minimize the mean square error between the network prediction and actual target value. If it is approaching targets then training is considered done. It has been observed that as the number of attributes increases, the training time also increases. In this way network approaches the known correct outputs (targets) in order to be trained.

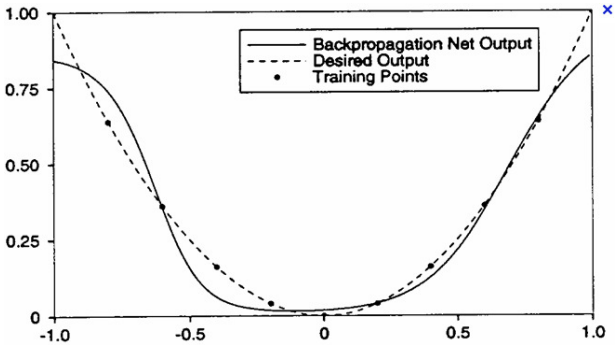


Figure 3: Backpropagation net output

The change in weights between hidden layer and input layer, is given by

$$\Delta w_{ij} = \alpha \delta_i h_j$$

where ά is the learning rate coefficient and

$$\delta_i = (t_i - o_i) o_i (1 - o_i)$$

The change in weights between hidden layer and input layer, is given by

$$\Delta w_{ij} = \beta \delta_{Hi} x_j$$

where $\beta$ is the learning rate coefficient and

$$\delta_{Hi} = x_i (1 - x_i) \sum_{j=1}^{k} \delta_j w_{ij}$$

## 9. DATABASE

For testing the proposed algorithm we will use data from UCI repository of Machine Learning Database. The details of the data are shown in Table 4.

Table 4: Dataset description

| Dataset | Transactions | Classes |
|---|---|---|
| iris | 150 | 3 |
| diabetes | 768 | 2 |
| glass | 214 | 7 |
| pima | 768 | 2 |

The proposed Neural network Associative Classification system is implemented using MATLAB 7.7.0. The experiments were performed on the Intel® Core duo CPU, 2.27 GHz system running Windows XP with 2.00 GB RAM

## 10. RESULTS

The present study indicates that the classification accuracy is enhanced with the adoption of a machine learning algorithm.

Table 5: Classification accuracy

| Dataset | CBA | Neural Network – Associative Classification system |
|---|---|---|
| iris | 88.34 | 91.87 |
| diabetes | 71.45 | 78.98 |
| glass | 68.98 | 76.56 |
| pima | 70.89 | 71.87 |

Table 6: Number of rules generated

| Dataset | CBA | Chi-square |
|---|---|---|
| iris | 184 | 90 |
| diabetes | 2554 | 267 |
| glass | 3516 | 860 |
| pima | 1520 | 285 |

For evaluating algorithms like learning time, execution time etc many metrics has been adopted. Evaluation measures brings out mainly two types of errors: false negatives and false positives.

$$False\ negative = \frac{Number\ of\ Missed\ Instances}{Total\ Number\ of\ Instances}$$

$$False\ positive = \frac{No:\ of\ Incorrectly\ classified\ Instances}{Total\ Number\ of\ Instances}$$

The following table and figures show the accuracy and output error at each epoch for the iris data set.

Table 7: Accuracy in different epochs

| Epoch | Accuracy | Output Error |
|---|---|---|
| 10 | 18 | 0.40 |
| 20 | 20 | 0.39 |
| 30 | 38 | 0.28 |
| 40 | 53 | 0.098 |

| | | |
|---|---|---|
| 50 | 60 | 0.096 |
| 60 | 70 | 0.075 |
| 70 | 78 | 0.067 |
| 80 | 88 | 0.059 |
| 90 | 90 | 0.046 |
| 100 | 92 | 0.035 |

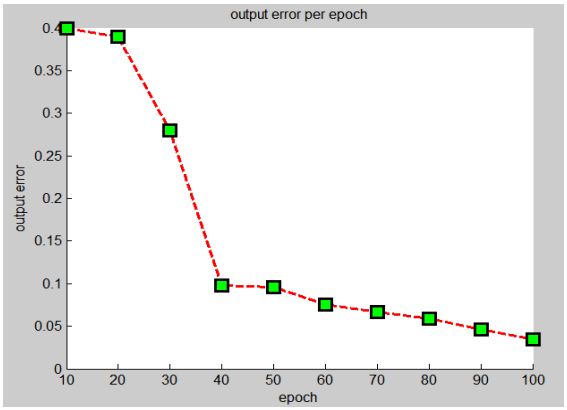


Figure 4: Output error per epoch

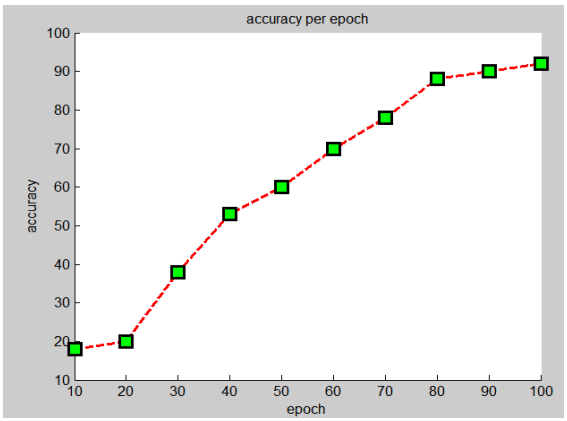


Figure 5: Accuracy per epoch

# 11. CONCLUSION AND FUTURE WORK

Associative classification is an important data mining task, which extracts high quality classifiers. In this paper, we have enhanced the method by selecting the rules which involve the attribute with minimum gini index. The process of rule generation and rule evaluation can be further enhanced by adopting a statistical approach. The significance of association via the $\chi^2$ test makes the ruleset more rich and informative. From the experimental results it can be inferred that Neural Network Classification system outperforms CBA in accuracy because it learns to adjust weights from the input that is class association rules and build a more accurate and efficient classifier. In future work, I intend to apply an optical neural network as a classifier model.

## REFERENCES


[1] Liu, B., Hsu, W. & Ma, Y. 1998 "Integrating classification and association rule mining", In Proceedings of the International Conference on Knowledge Discovery and Data Mining. New York, NY: AAAI Press,pp. 80-86.

[2] J. Han and M. Kamber, Data Mining: Concepts and Techniques, New York: Morgan Kaufmann Publishers, 2001.

[3] R. Agrawal, T. Imielinski, "A. Swami, Mining association rules between sets of items in large databases", Proceeding of 1993 ACM-SIGMOD International Conference on Management of Data, ACM Press, Washington, D.C., 1993, pp. 207- 216.

[4] Baralis, E., Chiusano, S. & Graza, P. A lazy approach to associative classification. IEEE transactions on knowledge and data engineering,VOL. 20, NO. 2,February 2008.

[5] S.P. Syed Ibrahim, K.R.Chandran, "Efficient Associative Classification using Genetic Network programming", International Journal of Computer Applications (0975 – 8887) Volume 29– No.6, September 2011

[6] Andrew Kusiak, "Feature Transformation Methods in Data Mining," IEEE Transactions On Electronics Packaging Manufacturing, Vol. 24, No. 3, July 2001, pp. 214–221.

[7] Li, W., Han, J. & Pei, J. CMAR: Accurate and efficient classification based on multiple-class association rule. In Proceedings of the International Conference on Data Mining (ICDM'01),San Jose,CA, pp. 369–376.2001.

[8] Schalkoff, R.J., Artificial Neural networks, McGraw Hill publication, ISE.

[9] J. Mao and K. Jain. ,"Artificial neural networks for feature extraction and multivariate data projection".IEEE Trans. Neural Networks, 6(2):296{317, 1995

[10] X. Yin and J. Han, "CPAR: Classification Based on Predictive Association Rules," Proc. Third SIAM Int'l Conf. Data Mining (SDM'03), May 2003.

[11] Snedecor, W., and Cochran, W. (1989) Statistical Methods, Eighth Edition, Iowa State University Press.

[12] www.highered.mcgraw-hill.com