LACI: Lazy Associative Classification Using Information Gain

S. P. Syed Ibrahim, K. R. Chandran, and C. J. Kabila Kanthasamy

Abstract—**Associative classification method applies association rule mining technique in classification and achieves higher classification accuracy. However, it is a known fact that associative classification typically yields a large number of rules, from which a set of high quality rules are chosen to construct an efficient classifier. Hence, generating, ranking and selecting a small subset of high-quality rules without jeopardizing the classification accuracy is of prime importance but indeed a challenging task. Lazy learning associative classification method eliminates the need of constructing the classifier but suffers with high computation cost.**

This paper proposes lazy associative classification using Information gain where, the system first chooses the Information gained attribute from the training sample and computes highest subset probability and then it directly predicts the class label. This proposed method not only reduces the computation cost but also improves the classification accuracy. Experimental result shows that the proposed system outperforms the traditional associative classification methods and the existing lazy associative classification method.

Index Terms—**Associative classification, information gain, lazy learning.**

I. Introduction

In the world where data is all around us, the need of the hour is to extract knowledge or interesting information, which is hidden in the available dataset. Data mining principally deals with extracting knowledge from data. Association rule mining and classification are data mining functionalities. Association rule mining is concerned with extracting set of highly correlated features shared among large number of records in the given database. It uses unsupervised learning where no class attribute is involved in finding the association rule. On the other hand, classification uses supervised learning where class attribute is involved in the construction of the classifier to predict the new instance. Both, association and classification are significant and efficient data mining techniques.

Associative classification is a recent and rewarding technique that applies the methodology of association into classification and achieves high classification accuracy.

Merschmann and Plastino [1] classified associative classification methods in two ways namely 1. Eager and 2. Lazy learning method. Eager Associative classification method constructs the generalized model to predict the class whereas Lazy learning Associative classification delays the processing of data until a new sample needs to be classified and does not build the generalized model to classify an instance.

Eager aassociative classifier [2], [3], [4] construction is of two phases. In the first phase, association rule mining is applied to discover class association rules. The important element in controlling the number of rules generated in associative rule mining is the support threshold. If the support value is high then number of rules generated is very less, but many high-confidence rules may get eliminated. On the other hand, if support value is set to minimum, then huge numbers of rules are generated. So in the next phase, the rules are ranked. After rule ranking, only the high-ranking rules are chosen to build the classifier and the rest are pruned.

So generating high quality rules, ranking the rules and building the classifier without drop in accuracy is of prime importance but a tedious job. This is achieved in lazy associative classification method, since it does not build a generalized classifier.

Merschmann and Plastino [1] proposed the lazy learning approach for protein classification. The authors proposed Highest Subset Probability (HiSP) algorithm, which is based on Bayes theorem. Here motif structures are evaluated based on training dataset and class protein class is predicted.

Recently, Merschmann and Plastino [5] proposed lazy approach for general classifier. Here HiSP-GC algorithm evaluates the subsets based on probabilistic analysis for the general classifier-datasets.

Syed et al., [6] proposed lazy learning associative classification method based on support and confidence measures.

These lazy learning associative classification method improves the classification accuracy but leads to high computation cost.

In [7] and [8] the authors proposed information gain attribute based approach for associative classification where high informative attribute is chosen for rule generation. Based on this idea the authors of [9] and [10] proposed information gain based weighted associative classification methods. In [11] the authors proposed associative classification method using genetic network programming. Here information gain attribute is used to construct the initial genetic network. The method proposed in [7] – [11] shows that, Information gain based approach reduces the computation cost.

This motivates us to propose a new computational technique in lazy learning associative classification. The rule generation phase in lazy learning associative classification is a hard step that requires a large amount of computation. To reduce the computation cost, this proposed method proposes

Manuscript received September 22, 2011, revised December 1, 2011.

S. P. Syed Ibrahim, K. R. Chandran and C. J. Kabila Kanthasamy are with Department of Computer Science and Engineering, PSG College of Technology, Coimbatore-641004, Tamilnadu, India (e-mail: sps_phd@yahoo.co.in; chandran_k_r@yahoo.co.in; cjkapila@gmail.com).

Information gain based lazy learning associative classification method which computes subset probabilities only for the information gained attribute. This method reduces the number of rules generated and does not build a generalized classifier from training data to classify the new samples.

The rest of the paper is organized as follows: Section 2 gives an insight about the past work in this field and section 3 explains the proposed system. Section 4 provides sample computation. The last section presents the experimental results and observations followed by the conclusion.

II. Related Work

Associative classification was first introduced by Liu et al., [4] which focuses on integrating two known data mining tasks, association rule discovery and classification. The associative classification is aims at a special subset of association rules whose right hand side is restricted to the class attribute; for example, consider a rule R: X → Y, Y must be a class label.

Associative classification generally involves two stages. In the first stage, it adopts either Apriori candidate generation [12] or Frequent Pattern (FP) growth [13] association rule generation algorithm to generate the class association rules. For example Class Based Association (CBA) [4] algorithm employs Apriori candidate generation and other associative classification algorithms such as Classification based on Predictive Association Rules (CPAR) [14], Classification Based on Multiple Class-Association Rules (CMAR) [3] and Lazy rule pruning methods in associative classification [15]–[17] adopts FP growth algorithm for rule generation.

The rule generation step generates huge number of rules and it is a hard step that requires a large amount of computation. Experimental results reported by Baralis et al., [17] has shown that the CBA algorithm which follows Apriori association rule mining algorithm generates more than 80,000 rules for some datasets that leads to memory exceptions and other severe problems, such as over-fitting etc., If all the generated rules are used in the classifier then the accuracy of the classifier would be high but the process of classification will be slow and time-consuming.

So in the next stage, generated rules are ranked based on several parameters and interestingness measures such as confidence, support, lexicographical order of items etc. After rule ranking, only the high-ranking rules are chosen to build a classifier whereas the rest are pruned.

In CBA method, the rules are ranked based on their confidence value. If two rules have the same value for the confidence measure then the rules are sorted based on their support. If both confident and support values are same for two rules then the sorting is done based on their rule length. Even after considering confidence, support, and cardinality measures, if there exists rules with the same values for all three measures then the rules are sorted based on its lexicographic order as in Lazy pruning [17] approach.

After rule ranking, CBA method uses database coverage method to construct the classifier. CMAR applies the Chi-square test [18] which gives positively correlated rules that can be used in the classifier. CPAR uses the Laplace accuracy measure to estimate the expected error rate for each rule.

A recent approach for rule pruning is lazy pruning [17] where a rule is pruned only if it misclassifies the data. The entire ruleset is segregated into three sets namely, useful rules, harmful rules and spare rules. A rule which classifies atleast one data item correctly is said to be useful rule and misclassifies a data item is known as harmful rule and the leftovers are the spare rules which are not pruned but used when needed. Though lazy pruning strategy works well for small datasets, in the case of large datasets, there exist constraints in memory space and ruleset quality.

Evolutionary based associative classification method [19] randomly selects some rules from the generated rule pool to construct the classifier. Richness of the ruleset was improved over the generation.

The following section explains the proposed work.

III. Proposed Work

A. Attribute Selection based on Information Gain

Generally Apriori based rule generation algorithm generates $2^k - 1$ rules for the dataset with K items [20]. So it leads to high computation cost. To reduce the number of the rule generated, Information gain attribute can be used.

Information gain is a measure that is used in information theory to quantify the 'information content' of messages [20]. In ID3 decision tree algorithm [21] information gain is used to choose the best split attribute.

In the process of generating the class association rules, instead of considering all the attributes, information gained attribute is used to generate the class association rules. This method generates j*g rules for a single test case with j non-class attributes and g classes in the entire data set. Hence if x test cases are to be predicted the number of rules generated will be x*j*g. In this way, the proposed work generates minimal number of high quality rules.

Suppose an attribute A has n distinct values that partition the dataset D into subsets T_1, T_2... T_n. For a dataset, freq (C_k, D) / |D| represent the probability than a tuple in D belongs to class C_k. Then info (D) is defined as follows to measure the average amount of information needed to identify the class of a transaction in D:

$$\text{Info(D)} = -\sum_{k=1}^{g} \frac{freq(C_k, D)}{|D|} X \log_2\left(\frac{freq(C_k, D)}{|D|}\right) \quad (1)$$

where |D| is the number of transactions in database D and g is the number of classes.

After the dataset D is partitioned into n values of attribute A, the expected information requirement could be defined as:

$$\text{info}_A\text{ (D)} = \sum_{i=1}^{n} \frac{|D_i|}{|D|} X \text{ info } (D_i) \quad (2)$$

The information gained by partitioning D according to attribute A is defined as:

$$\text{Gain (A)} = \text{info (D)} - \text{info A (D)} \quad (3)$$

The best split attribute is the one that maximized the information gain in the data set D. This best attributes is used to generate the subset.

B. Subset Evaluation

After identifying the information gained attribute, the subsets are generated. For each generated subset, probability values are calculated.

Let Tx be the set of transaction present in the dataset. Supposing that there are m classes, $\{C_1, C_2, \ldots, C_m\}$. Class values X is assigned to class C_k, 1 <= k <= m, if and only if

$$P(C_k / X) > P(C_j / X) \text{ For all j, 1<=j<=m, j not equal to k.} \tag{4}$$

$$P(Ci / t) = P(Ci \Lambda t)(P(t) \tag{5}$$

where $P(C_i \Lambda t)$ stands for the probability of a subset pertaining to class Ci and having the subset t. P(t) is the occurrence of the subset t.

They are estimated from the training dataset as in the following way:

$$P(C_i \Lambda t) = \frac{NC_i t}{N} \tag{6}$$

where NCit and N are the number of samples having the class Ci, and the total number of training samples. and

$$P(t) = \frac{N_t}{N_c} \tag{7}$$

where N_t is the number of sample having the subset t with class C_i, N_c is the number of class count.

The decision of which class will be assigned to the instance X is based on the analysis of the subsets of attribute values associated with higher posteriori probabilities P(Ci|t).

For this, all the posterior probabilities are sorted. Then, it uses lower limit to limit the number of possibilities. Lower limit can be calculated as

$$Lower\ Limit = \frac{Maximum\ probability}{Sqrt(M)}, \tag{8}$$

where m is number of class.

Maximum class posterior probability will be assigned to the testing samples.

IV. Sample Computation

Let us consider a sample dataset given in Table I, contains 14 transaction, and 2 class values and Table II consists of test dataset. The task is to predict the class label for new test data instance. Among the four attributes, attribute 'CD4 Cell Count' is selected as the information gained attribute as it has the maximum Information gain value.

Lazy learning algorithm calculates the probability value for each of subset of testing dataset using the equation (7). Then Class labels are assigned based on high probability of subset. Here equation (8) is used to find the lower limit.

The following are rules that are generated based on information gained attribute.

- {CD4 Cell Count =200 .. 500}
- {CD4 Cell Count =200 .. 500, Sweating at Night = Medium}
- {CD4 Cell Count =200 .. 500, Sweating at Night = Medium, Tuberculosis (TB) = No}
- {CD4 Cell Count =200 .. 500, Sweating at Night = Medium, Tuberculosis (TB) = No, Temperature = Normal}

TABLE I: Sample Dataset

CD4 Cell Count	**Sweating at Night**	**Tuberculosis (TB)**	**Temperature**	**AIDS**
>500	High	no	Normal	No
>500	High	no	High	No
<200	High	no	Normal	Yes
200 .. 500	Medium	no	Normal	Yes
200 .. 500	Nil	yes	Normal	Yes
200 .. 500	Nil	yes	High	No
<200	Nil	yes	High	Yes
200 .. 500	Medium	no	High	No
>500	Medium	no	Normal	No
>500	Nil	yes	Normal	Yes
200 .. 500	Medium	yes	Normal	Yes
>500	Medium	yes	High	Yes
<200	Medium	no	High	Yes
<200	High	Yes	Normal	Yes

TABLE II: Test Dataset

200 .. 500	Medium	no	Normal	?

The other itemsets that are commonly generated by the association rule mining procedure are eliminated. The excluded rules are

- {Sweating at Night == Medium }
- {Tuberculosis == No}
- {Temperature == Normal}
- {Sweating at Night == Medium, Tuberculosis == No}
- {Sweating at Night == High, Temperature == Normal}
- {Tuberculosis == No, Temperature == Normal}
- {CD4 Cell Count =200 .. 500 ,Tuberculosis == No}
- {CD4 Cell Count =200 .. 500 ,Temperature == Normal}
- {CD4 Cell Count =200 .. 500 , Sweating at Night == High, Temperature == Normal}
- {CD4 Cell Count =200 .. 500, Tuberculosis == No, Temperature == Normal}
- {Sweating at Night == Medium, Tuberculosis == No, Temperature == Normal}

This clearly shows this algorithm generates minimal number of rules. Table III shows the sample computation.

TABLE III: SAMPLE COMPUTATION

Rules	Class (AIDS)	Occurrence with Class	Posteriori Probability
{CD4 Cell Count =200 .. 500}	Yes	3	1.92
	No	2	0.72
{CD4 Cell Count =200 .. 500, Sweating at Night = Medium}	Yes	2	1.28
	No	1	0.36
	No	1	0.36
{CD4 Cell Count =200 .. 500, Sweating at Night = Medium, Tuberculosis (TB) = No, Temperature = Normal}	Yes	1	0.64
	No	0	0

Probability for YES Class: 9/ 14 = 0.64.

Probability for NO Class : 5/14 = 0.36.

$$\text{Lower Limit} = \frac{\text{Hightest probability}}{\text{Sqrt(M)}} \quad (9)$$

$$= 1.92/1.414 = 1.36$$

So number of Yes 1 → So yes class is assigned.

V. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed system was tested using benchmark datasets from the University of California at Irvine Repository (UCI Repository) [22]. The datasets were preprocessed to convert to a general format. A brief description about the datasets is presented in Table IV.

TABLE IV: DATASET DESCRIPTION

Dataset	Transactions	Classes
Anneal	998	5
Balance Scale	625	5
Breast Cancer	286	2
Breast -w	699	2
Car	1728	4
Credit – a	690	2
Diabetes	768	2
Ecoli	336	8
Flare	1389	3
Glass	214	7
Ionosphere	351	2
Iris	150	3
Mushroom	8124	2

The experiments were carried out on a PC with Intel Core 2 Duo CPU with a clock rate of 1.60 Ghz and 2 GB of main memory. Holdout approach [20] was used where 90% of the data was randomly chosen from the dataset and used as training dataset and remaining 10% was used as the testing dataset. The training dataset is used to generate subset and the test dataset is used to estimate the classifier performance.

A. Accuracy Computation

Accuracy measures the ability of the classifier to correctly classify unlabeled data. It is the ratio of the number of correctly classified data over the total number of given transactions in the test dataset.

$$Accuracy = \frac{Number\ of\ Correctly\ Predicted\ Test\ Data}{Total\ Number\ of\ Test\ data} \quad (10)$$

Table V shows the accuracy comparison. The first column describes the dataset name; next column describes the highest information gain attribute. The third column shows the accuracy for traditional associative classification method [4]. The forth column shows the accuracy for Lazy Learning associative classification which was proposed in [6] and last column shows accuracy values for various dataset for the proposed system. The overall accuracy was obtained by calculating the average of the accuracy values obtained from the ten different runs.

TABLE V: ACCURACY COMPUTATION

DATASET	Info Gain Attribute	CBA	LLAC	LACI
Anneal	12	**80.18**	76.11	76.11
Balance Scale	2	69.29	**71.43**	70.32
Breast	6	66.48	**76.55**	67.86
Breast -w	2	**93.7**	90.86	88.57
Car	6	67.33	**70.12**	**70.12**
Credit – a	9	76.48	**77.43**	76.81
Diabetes	2	**69.1**	68.31	68.83
Ecoli	6	73.53	71.76	**79.41**
Flare	1	81.58	**84.71**	**84.71**
Glass	8	57.94	62.73	59.09
Ionosphere	5	82.29	92.67	**94.44**
Iris	4	**96.67**	78.89	95.33
Mushroom	5	91.89	90.94	**98.52**
AVERAGE		**77.42**	**77.89**	**79.24**

Table V shows the accuracy comparison for thirteen datasets, the proposed algorithm has about +1.82 percent improvements against the traditional associative classification and about +1.35 percent improvements against the existing lazy learning method respectively.

TABLE VI: TIME TAKEN TO PREDICT SINGLE INSTANCE

Dataset	LLAC	LACI
Anneal	0.0603	0.0510
Balance Scale	0.0017	0.0015
Breast	0.0036	0.0027
Breast -w	0.0042	0.0026
Car	0.0024	0.0038
Credit – a	0.0201	0.0047
Diabetes	0.0051	0.0045
Ecoli	0.0017	0.0014
Flare	0.0202	0.0188
Glass	0.0030	0.0027
Ionosphere	0.0127	0.0125
Iris	0.1362	0.0185
Mushroom	0.681	0.065
Average	0.07325	0.01459

Table VI shows the average time taken to predict a single instance. The existing lazy learning approach predicts the class at an average of 0.07325 seconds and the proposed system takes only 0.01459 seconds to predict the single instance. For example the existing system takes 0.681 seconds to predict the single instance for mushroom dataset because mushroom dataset consists of 8124 records, 23 attributes and 2 class labels. So the existing system takes huge time to predict the single instance. On the other hand the proposed system choose information gain attribute so it takes only 0.065 seconds to predict the single instance of mushroom dataset.

B. Scalability Test

In order to analyze the scalability of proposed LACI method, the dataset is divided into various combinations of training and testing percentage. Figure 1 shows that the proposed LACI is scalable.

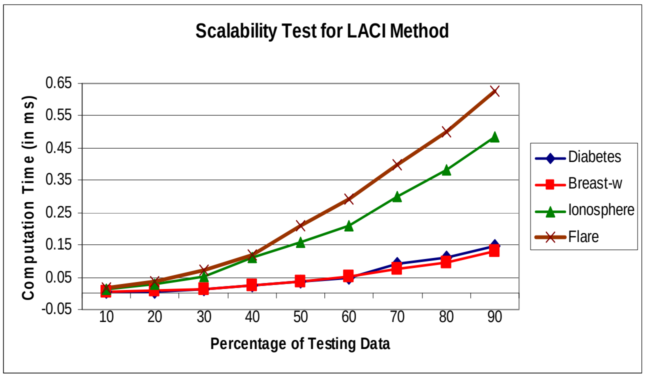

Fig.1. Scalability Test

VI. CONCLUSION

The main idea for LACI is to build computationally efficient classifier. LACI method predicts the class for each sample based on evaluation of its subsets of attribute values. This method uses information gain attribute to generate the subset and calculates the posteriori probabilities for each subset then it predicts the class based on that knowledge.

LACI was tested with thirteen datasets from the UCI Machine Learning Repository. The proposed system is compared with the traditional associative classifier and another variation of lazy learning method. The experimental results prove that the proposed system extremely better than the traditional and existing method and the proposed system takes average of 0.01459 seconds to predict the instance. Accuracy, computation time and scalability test shows that the proposed LACI is an efficient associative classifier.

REFERENCES

[1] Merschmann. L, Plastino. A, "A lazy data mining approach for protein classification". *IEEE Transactions on Nanobioscience,* vol. 6, no. 1, pp. 36-42, 2007.

[2] Dong. G, Zhang. X, Wong. L, and Li. J, "CAEP: Classification by Aggregating Emerging Patterns," in Proc of second International Conference on Discovery Science, 1999, pp .

[3] Li. W, Han. J, and Pei. J, "CMAR: Accurate and Efficient Classification Based on Multiple Class-Association Rules," Proceedings of IEEE International Conference on Data Mining (ICDM '01), Nov. 2001.

[4] Liu. B., Hsu. W and Ma. Y, "Integrating Classification and Association Rule Mining", Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, pp.80–86, 1998.

[5] Merschmann. L, Plastino. A "HiSP-GC: A Classification Method Based on Probabilistic Analysis of Patterns" Journal of Information and Data Management, Vol. 1, No. 3, October, Pages 423–438, 2010.

[6] Syed Ibrahim. S.P., Chandran.K.R, Nataraj.R.V, "LLAC: Lazy Learning in Associative Classification" in the Springer Lecture Series in Communications in Computer and Information Science (CCIS), Advances in Communication and Computers, 190, Part I, pp. 631–638, 2011.

[7] Chen. G, Hongyan Liu, Lan Yu, Qiang Wei, Xing Zhang, 'A new approach to classification based on association rule mining', Science Direct, Decision Support Systems, 42, pp. 674– 689, 2006.

[8] Zhang. X, Chen.G, Wei. Q, "Building a highly-compact and accurate associative classifier, Applied Intelligence, Volume 34, Number 1, pp.74-86, 2011.

[9] Syed Ibrahim. S. P., Chandran. K.R, Compact Weighted Class Association Rule Mining using Information Gain, accepted to publish in the International Journal of Data Mining and Knowledge Management Process (IJDKP), vol 1, no 6, pp 1-13, 2011.

[10] Syed Ibrahim S.P, Chandran K.R, Abinaya.M.S, "Compact Weighted Associative Classification" in the IEEE International Conference on Recent Trends in Information Technology (ICRTIT 2011), MIT, Anna University. Chennai, June 3-5, pp. 1099 – 1104., 2011.

[11] Syed Ibrahim.S.P, Chandran. K.R., "Efficient Associative Classification Using Genetic Network Programming" in the International Journal of Computer Applications ", ISSN 0975-8887, Volume 29– No.6, pp. 1-8, September 2011.

[12] Agrawal, R. and Srikant, R. "Fast algorithms for mining association rule." Proceedings of the 20th International Conference on Very Large Data Bases. pp. 487-499, 1994.

[13] Han, J., Pei, J. and Yin, Y. "Mining frequent patterns without candidate generation". In Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data. Dallas, TX: ACM Press, pp.1–12, 2000.

[14] Yin.X and Han. J, "CPAR: Classification Based on Predictive Association Rules," Proceedings of Third SIAM International Conference on Data Mining (SDM '03), May 2003

[15] Baralis, E. and Torino, P. A lazy approach to pruning classification rules. Proceedings of the 2002 IEEE International Conference on Data Mining (ICDM'02), Maebashi City, Japan, pp. 35-42, 2002.

[16] Baralis, E., Chiusano, S. and Graza, P. On support thresholds in associative classification. In Proceedings of the 2004 ACM Symposium on Applied Computing. Nicosia, Cyprus: ACM Press, pp. 553–558, 2004.

[17] Baralis. E, Chiusano. S, Garza.P, A Lazy Approach to Associative Classification, IEEE Transactions on Knowledge and Data Engineering, vol.20 no.2, pp.156-171, February, 2008.

[18] Snedecor.W, and Cochran. W, "Statistical Methods", Eighth Edition, Iowa State University Press, 1989.

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

[20] Syed Ibrahim.S.P, Chandran. K.R, Jabez Christopher. J, "An Evolutionary Approach for Ruleset Selection in a Class Based Associative Classifier" in the European Journal of Scientific Research, ISSN 1450-216X, Vol.50 No.3, pp.422-429, 2011.

[21] Quinlan. J, "Induction of decision trees". Machine Learning, pp.81–106, 1986.

[22] Blake. C.L. and Merz. C.J., UCI Repository of machine learning databases, 1998.

S. P. Syed Ibrahim received his BE in 2004 from Bharathidasan university and ME in 2006 from Anna University, Chennai. He has been a research scholar of Anna Univerisy – Coimbatore. He is currently an Assistant professor in the department of Computer Science and Engineering at PSG College of Technology, Coimbatore. He has published more than 20 research papers in various journals and conferences. His research interest includes data mining, storage management.

K. R. Chandran received his BE in 1982 and ME in 1986 from PSG College of Technology and PhD in 2006 from Bharathiar university, Coimbatore. He is currently a professor in the department of Information Technology and head of Computer and Information Sciences at PSG College of Technology, Coimbatore. He is a Fellow member in Institution of Engineers (India) and life member of ISTE, ACS, and The Textile Association (INDIA). He has published more than 60 research papers in various journals and conferences. His research interest includes software engineering, computer networks, data mining, system modeling and simulation.

C. J. Kabila Kanthasamy received her B.Sc in Information Technology in 2009 from Avinashilingam University, Coimbatore and is now perusing her MSc in Computer Technology at PSG College of Technology. Her current research interests include data mining and database systems.