



A Review of Lazy Learning Associative Classifications

¹Preeti Tamrakar and ²S.P. Syed Ibrahim

¹School of Computing Science and Engineering,
VIT Chennai, India.

preeti.tamrakar2014@vit.ac.in

²School of Computing Science and Engineering,
VIT Chennai, India.

syedibrahim.sp@vit.ac.in

Abstract

Associative classification (AC) is an approach in data mining that utilizes the technique of association rule discovery to learn classifier. Associative classification can be categorized into two ways; one is eager associative classification and second is lazy learning associative classification. Over the most recent couple of years, many number of lazy learning associative classification algorithms have been introduced by the different researchers, i.e. LAC, HiSP, HiSP-GC, LLAC, LACI, CHiSC and others. These algorithms utilize a few distinctive subset generation and subset evaluation methods. This survey paper focuses on comparing and studying the ultra-modern lazy learning associative classification techniques.

Key Words: Associative classification, attribute (feature) selection, lazy learning, information gain.

1. Introduction

All over the world a tremendous measure of information being gathered and put away in databases. There is invaluable data and learning covered up in such databases.

Data mining otherwise known as knowledge discovery process principally deals with extracting knowledge from data using algorithms or techniques. Classification, summarization, clustering, trend analysis and association rule discovery [1] are important and data mining tasks. Classification utilizes supervised machine learning where the class label is engaged with the development of the classification system to predict the unseen data. Whereas association rule mining (ARM) deals with the extraction of highly correlated features with reference to the huge database records. Unsupervised machine learning is utilized by association rule mining where class attribute is not available.

Associative classification [3] is a current as well as remunerating procedure which employs the philosophy of association rule mining into classification and accomplishes very high accurate classifiers. It is a special kind of association rule mining where class label is attached in the right hand side of the rule. Antecedent is composed of attributes and consequent is class label. The important advantage using associative classification over the normal classification is that the outcome of the associative classification is in the form of simple if-then rules which makes very easy to understand for the end users. Associative classification methods are characterized into two ways; first one is Eager Learning Method and second one is Lazy Learning Method.

Two phases are involved in the construction of eager associative classification method. Association rule mining such as a priori algorithm used by CBA[3] or frequent pattern (FP) growth algorithm are used by CMAR, CPAR to generate the classification association rules (CARs). Multiple numbers of rules are generated. Every rule is not very important to construct the classifier. So ranking and pruning is involved in second phase. For ranking, some interesting measures have been used such as support, confidence, rule generated first etc. highest ranked rule are used to build the classification system and rest rules are pruned.

Eager associative classification provides better accuracy but generating the large number of rules, ranking and pruning are very important and unavoidable jobs that takes too much of time.

To overcome these challenges, Lazy learning associative classification is introduced. It postpones the processing of data or construction of the classifier until the point when another new instance demands for classification and does not fabricate the model to classify a test sample.

These lazy associative classification methods provide higher accuracy of the classifier but leads to high computation cost. Various information gain based attribute selection methods are proposed to reduce the computation cost. The goal of this paper is to compare and study the different ultra-modern Lazy learning Associative Classification techniques. The following paper is arranged as follows: The Lazy learning Associative Classification problem and its main steps are given in Section 2. Different methods used to generate and evaluate the subsets are surveyed in Sections 3. Final section presents the conclusions of this study.

2. Lazy Learning Associative Classification

Eager associative classification consumes more time on generating the high quality rules, ranking and pruning. To avoid these problems lazy learning associative classification method came into existence. Lazy learning associative classification postpones building the classification model until a query is given. Lazy learning systems can simultaneously solve multiple problems and deal successfully with changes in the problem domain. It best suits for large datasets with few attributes. It focuses on the attributes of the given test tuple. Lazy learning associative systems are applicable where training datasets are frequently updated. Some lazy learning associative classifications are explained here:

Lazy Associative Classifier

Dissimilar to the eager associative classifier that concentrates on a set of ranked classification association rules from the training data, the lazy associative classifier induces [4] classification association rules particular to each test tuple. Lazy associative classification i.e. LAC delays the building of classifier till the test tuple comes. It projects the transactions of the training datasets which matches the test tuple. The final classification is based on only from the projected dataset. The lazy approach takes the training data, only on those features in the test instance. From this selected training data, the classification association rules are constructed and ranked, and the best Classification association rule is used. From the set of all training dataset, only the tuple sharing at least one attribute with the test tuple are used to form the small subset of dataset. Then, a rule-set is generated from the subset and accordingly class is assigned. Model is shown in fig 1. This method gave better accuracy when compared to the traditional classification system.

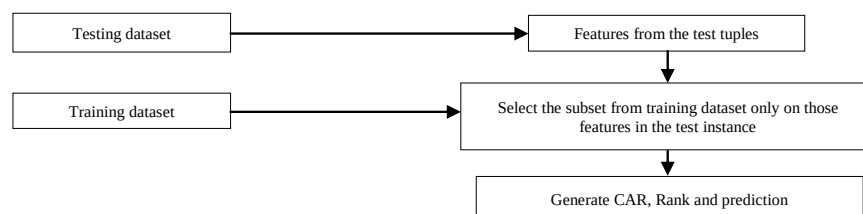


Fig. 1: Model for Lazy Associative Classifier

HiSP-GC(Lazy learning classification based on Highest Subset Probability–General Classifier)

LAC gave good accuracy but takes time for each and every test tuple to classify. To address this challenge, HiSP-GC: A Classification Method Based on Probabilistic Analysis of Patterns Highest Subset is proposed [6]. In the HiSP-GC approach a test tuple is classified based on posteriori probability of the subsets. All possible combination of subsets are generated and the generated subsets are evaluated based on posteriori probability given in equation (i), where the subsets of attribute values associated with larger probabilities of pertaining to some training class define the classification of the test instance X. Model is shown in fig 2.

$$P(C_a | S) = \frac{\text{Probability of a tuple having class } C_a}{\text{Probability of the subset } S} \quad (i)$$

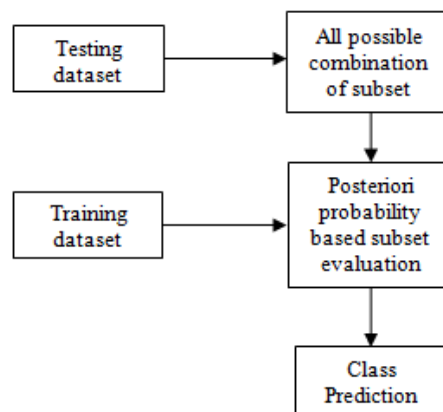


Fig. 2: Model for Highest Subset Probability–General Classifier

Finally this method uses two phases. First phase is subset generation and second phase is subset evaluation. In the first phase, all the possible combination of item set is generated based on the testing dataset from the training dataset. Second phase is involved with the calculation of posteriori probability for each of the item set generated from the first phase. Highest probability class is assigned to the test tuple.

LLAC (Lazy Learning in Associative Classification)

Association rules are created by using the threshold support and confidence to identify the most important relationships. Support tells how frequently the items arrives in the database whereas confidence tells the number of times the particular transactions have been found to be true. By introducing confidence and support measures, LLAC i.e. Lazy learning in Associative Classification [3] is proposed. This method is the modification over HiSP method. LLAC uses HiSC i.e. Highest subset confidence algorithm to evaluate the subset generated from the first phase.

In LLAC [6], test data is taken as input and generates the subset for each combination of class value. Support s and confidence c is calculated for each generated subsets using the equation (ii). The class label for the new unseen test tuple is assigned as per the highest support and confidence values. HiSC i.e. Highest Subset Confidence algorithm is used to predict the class label of unseen test instance. LLAC outperforms when compared with traditional classification systems. Model is shown in fig 3.

$$\text{Confidence of the rule} = \frac{\text{Occurance of instance } X \text{ and class } c \text{ together}}{\text{occurance of instance } X} \quad (\text{ii})$$

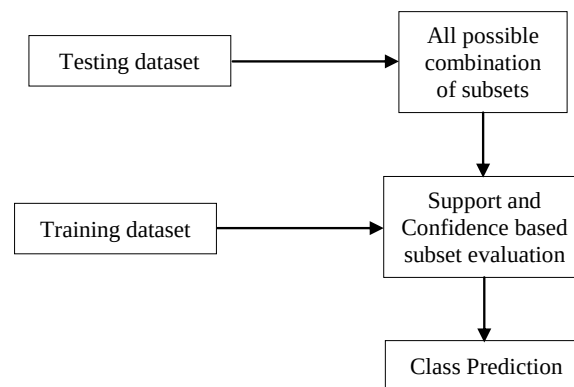


Fig. 3: Model for Lazy Learning in Associative Classification

LACI (Lazy Associative Classification using Information Gain)

HiSP-GC and LLAC methods generate all the combination of item set from the training set. Each and every time the entire attribute are not equally important. To choose best attribute among all the attributes, is known as feature selection. Information gain is one method for feature selection. To address the above problem LACI [5] and CHiSC [7] introduces information gain method. Information gain tells us how important a given attribute of the feature vectors is.

In the previous works, all the possible subsets are generated. Generating all the subsets takes huge computation time. To address this challenge, LACI is proposed. To reduce the numbers of generated subsets, Information gain is calculated for each and every attribute from the training dataset and the highest information gain attribute is chosen to generate the subsets. Information gain is calculated using the following equation:

$$\inf o(D) = \sum_{i=1}^c -P_i \log_2 P_i \quad (\text{iii})$$

Where $|D|$ is the total count of rows in transactional database D and c is the total count of class labels. After the dataset D is partitioned into n values of attribute A , the expected information requirement could be defined as:

$$\text{Info}_A(D) = \sum_{j=1}^v \frac{|D_j|}{|D|} * \inf o(D_j) \quad (\text{iv})$$

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

$$\text{InfoGain}(D, A) = \text{info}(D) - \text{info}_A(D) \quad (\text{v})$$

After generating the compact subset, subsets are evaluated based on the highest posteriori probability $P(C_a|S)$. Subset posteriori probability is calculated using the equation no (i).

Then lower limit is calculated to reduce the number of likelihood to ease the classification task using the following equation:

$$\text{lower lim it} = \frac{\text{Highest prob}}{\sqrt{n}} \quad (\text{vi})$$

Where n is the number of class attribute.

Model is presented in fig 4.

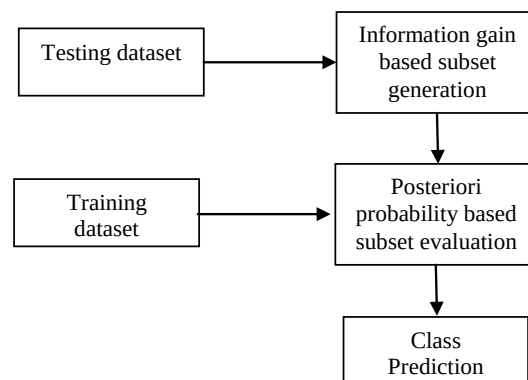


Fig. 4: Model for Lazy Associative Classification Using Information Gain

This method reduces the computation cost as well as also improves the prediction accuracy. The basic aim for LACI is to construct a precise and efficient classifier computationally.

CHiSC (Lazy learning classification based on Compact Highest Subset Confidence)

CHiSC method calculates the information gain for each of the attributes using the equation no (iii), (iv) and (v) and highest information gained attribute is chosen to generate the subsets. Support and confidence is calculated for each of the generated subsets using the equation (ii).

After analyzing the subsets of attributes values, those are associated with the highest support and confidence, that class is assigned to the test tuple. Model for Lazy learning classification based on Compact Highest Subset Confidence is presented in fig 5.

LACI i.e. Lazy Associative Classification Using Information Gain and CHiSC i.e. Compact Highest Subset Confidence-Based Associative Classification use information gained attribute to generate the subsets in the first phase. Number of generated items is less in these methods as compared to HiSP-GC and LLAC. So they take less time to generate the rules.

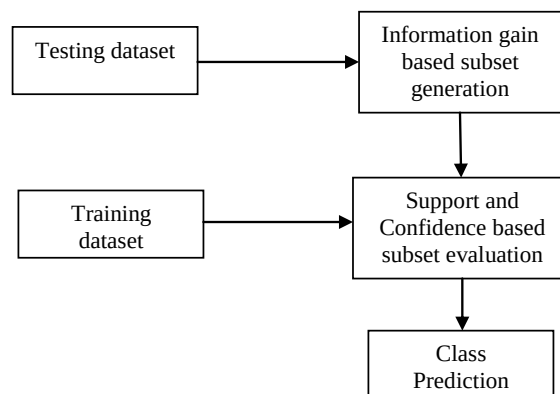


Fig. 5: Model for CHiSC

LACI uses Posteriori probability to evaluate the subsets whereas CHiSC uses support and confidence measure to evaluate the subsets and according class is assigned to the test tuple.

3. Result and Discussion

All the different lazy learning associative systems are tested and validated with the help of the standard datasets taken from the University of California at Irvine Repository (UCI Repository) [12]. The short illustration of datasets is given in Table no 1.

Table 1: Dataset Description

Dataset	Transactions	No of attributes	Classes
Balance Scale	625	5	3
Breast Cancer	286	10	2
Breast Wisconsin	699	10	2
Car	1728	7	4
Credit Approval	690	16	2
Diabetes	768	9	2
Ecoli	336	8	8
Flare	1389	13	6
Glass	214	10	6
Ionosphere	351	35	2
Iris	150	5	3
Mushroom	8124	23	2

Accuracy computation: The accuracy is calculated from the given below formula. Accuracy comparison is shown in Table 2.

$$\text{Accuracy} = \frac{\text{Number of correctly predicted test data}}{\text{Total no of test data}} \quad (\text{vii})$$

Table 2: Accuracy Comparison

Dataset	Traditional Eager AC Method	Lazy Learning A C methods			
	CBA	HiSP-GC	LLAC	LACI	CHiSP
Balance scale	69.29	75.80	71.43	70.32	71.27
Breast cancer	66.48	70.85	76.55	67.86	85.22
Breast Wisconsin	93.70	89.71	90.86	88.57	90.63
Car	67.33	69.53	70.12	70.12	70.12
Credit approval	76.48	73.34	77.43	76.81	83.33
Diabetes	69.10	67.10	68.31	68.83	76.75
Ecoli	73.53	70.60	71.76	79.41	80.00
Flare	81.58	87.05	84.71	84.71	84.71
Glass	57.94	71.42	62.73	59.09	60.91
Ionosphere	82.29	86.90	92.67	94.44	96.29
Iris	96.67	94.33	78.89	95.33	87.22
Mushroom	91.89	87.65	90.94	98.52	98.52
Average	78.21	78.69	78.03	79.50	82.08

Table 3: Comparison of different Lazy Learning Associative Classification

Sr. no.	Name	Subset Generation	Subset Evaluation
1.	Lazy Associative Classification	Projected Training data	Maximum likelihood
2.	Highest Subset Probability based Classification	All possible combination	Posteriori probability
3.	Lazy learning in Associative classification	All possible combination	Support and confidence
4.	Lazy associative classification using information gain	Information gain	Posteriori probability
5.	Compact Highest subset confidence based associative classification	Information gain	Support and confidence

4. Conclusion

This paper provides survey of different lazy learning associative classification methods. Four lazy learning methods namely HiSP-GC, LLAC, LACI and CHiSP are compared and results are tabulated. Result shows that CHiSP outperforms the traditional eager associative classification and other lazy learning associative classifications.

References

- [1] Agrawal R., Srikant R., Fast algorithms for mining association rule, Proceedings of the 20th International Conference on Very Large Data Bases, Morgan Kaufmann, Santiago, Chile (1994), 487–499.
- [2] Baldonado M., Chang C.C.K., Gravano L., Paepcke A., The Stanford digital library metadata architecture, International Journal on Digital Libraries 1(2) (1997), 108-121.
- [3] Syed Ibrahim S.P., Chandran K.R., Nataraj R.V., LLAC: Lazy learning in Associative Classification, the Springer Lecture Series in Communications in Computer and Information Science (CCIS), Advances in Communication and Computers (2011), 631–638.
- [4] Adriano Veloso, Wagner Meira Jr., Mohammed J. Zaki, Lazy Associative Classification, Sixth International Conference on Data Mining (2006), 645–654.

- [5] Syed Ibrahim S.P., Chandran K.R., Kabila Kanthasamy C.J., LACI: Lazy Associative Classification Using Information Gain, IACSIT International Journal of Engineering and Technology 4(1) (2012).
- [6] Merschmann L., Plastino A., HiSP-GC: A Classification Method Based on Probabilistic Analysis of Patterns, Journal of Information and Data Management 1(3) (2010), 423–438.
- [7] Syed Ibrahim S.P., Chandran K.R., Kabila Kanthasamy C.J., CHISC-AC: Compact Highest Subset Confidence-Based Associative Classification, Data Science Journal 13 (2014).
- [8] Liu B., Hsu W., Ma Y., CBA: Integrating Classification and Association Rule Mining, Proc. Fourth Int'l Conf. Knowledge Discovery and Data Mining (1998).
- [9] Yongjian Fu, Data Mining: Tasks, Techniques and Applications, Department of Computer science, University of Missouri, Rolla (2002).
- [10] Baralis E., Chiusano S., Garza P., A Lazy Approach to Associative Classification, IEEE Transactions on Knowledge and Data Engineering 20(2) (2008), 156-171.
- [11] Han J., Kamber M., Data Mining: Concepts and Techniques, New York: Morgan Kaufmann Publishers (2001).
- [12] Blake C.L., Merz C.J., UCI Repository of machine learning databases (1998).

