Vol. 5(18) Special Issue, Dec. 2015, PP. 2600-2607

Contents list available at IJMEC

International Journal of Mechatronics, Electrical and Computer Technology (IJMEC)

Journal Homepage: www.aeuso.org

PISSN: 2411-6173 - EISSN: 2305-0543

Classification with the Use of Association Rules with Local Weighting

Hanieh Mozhdeh[1] and Karim Faez[2]*

[1]*Faculty of Computer and Information Technology Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran*

[2]*Electrical Engineering Department, Amirkabir University of Technology Tehran, Iran*

***Corresponding Author's E-mail:** kfaez@aut.ac.ir*

Abstract

Classification is one of the best assets in the field of data mining and machine learning. In classification problems one can learn a model or a function by the use of training data, then he or she can use this function (or model) to classify other data that have never been seen before in the system. One of the classification techniques is to extract association rules from data. This technique tries to find meaningful relationships between the members of a dataset. These relationships can be defined by the use of some rules. Recently, the use of associative rules for the purpose of classification captured a lot of attention among scientists. Researches done in this area, showed the high potential of association rules for classification. This article tries to use local weighting of inputs (data) for extracting better rules and then by pruning of association rules enhance the accuracy of classification act. To estimate the performance of the proposed method, two different set of classification algorithms, classic classification and classification based on modern association rules, were compared to it. The acquired results, showed the high capabilities of proposed method in the field of classification.

Keywords: *Association rules, local weighting of inputs, classification.*

1. Introduction

Association rules show the relationships and dependencies between the items of a large set of data types. [1] Finding these types of rules are under consideration and they have a variety of applications. As an example, finding the association rules between large amounts of business interactions can be used for fraud detection; in addition to that, in the field of medicine [2], and data mining it is used for web using by users and personalization [3]. Association rules are like A⇒B, A and B are subsets of possible values and there is a meaningful relationship between A and B. To extract such rules, we have to look for frequent pattern that appeared together in datasets. For example in the products that customers bought in a supermarket in most cases, milk and cake were bought together and these two products allocated a large amount of sales to themselves. The main purpose of this article is to classification by using association rules. In classification the goal is to assign a label to data (like labeling an email as a span or in reverse). Usually the main goal of finding association rules is to find relationships and dependencies and it has nothing to do with targeted search. But in classification the goal (data label) is obvious. With some improvements we can use association rules extraction to solve classification problem.

2. Related Work

For the first time in [4], classification was done by the use of association rules and it was named as 'Classification based on association' (CBA). In [5] for improving the CBA method, a method called CMAR

Article History:

Received Date: Aug. 04, 2015

Accepted Date: Dec. 11, 2015

Available Online: Jan. 03, 2016

was presented, this method uses FP-growth to extract class rules. FP-growth is a fast method for association rules extraction [6]. In [5], after class rules extraction, some of these rules were attributed to a class. By the use of this method a notable improvement were achieved in comparison with CBA method. In [7], a method based on FOIL [8] which is a method based on learning rules from data and association rules of a class was used. The method proposed in [9], in contrast to previous methods, datasets are directly used for the extraction of Correlated Association Rules for classification. This method increases the accuracy of classification; moreover it reduces the training period.

The method proposed in [10] in addition to positive association rules benefits from negative association rules. These states, those relations that in them a customer did not bought milk and cake together are also important. In fact, in negative association rules, there is useful information for classification. Some of methods that are based on association rules are used in [11] to solve multi class problems. In HARMONY method [12], with the use of the center of data, rules will be produced. By benefiting from this mechanism, we can be sure that by each high confidence rule in rules base, a sample in the training set will be covered. So it will create a rule per each data.

The method proposed in [13] offered three new innovations for classification: 1) it uses "Information Gain" to create candidate items set. 2) The process of the creation of repetitive items produces "Integration Rules". 3) The combination of defense against the redundancy of rules and conflict strategies along with the process of rule extraction. Another one of the methods that was proposed for classification based on association rules is MMAC [14]. MMAC is used for multi label datasets.

In [15] the main concentration was on the overlapping of samples of training data in the process of association rules production. Due to the overlapping of samples in training datasets, a lot of rules have redundancy and are useless. This method tries to exclude the overlapped samples to produce stronger rules. To do that, it benefits from "Rank". Those produced rules can be used in Multi-label problems.

Another method that has been concentrated on the negative association rules is the method that was presented in [16]. One of the problems of negative association rule production is the high cost of creating these rules. In [16], Apiriori algorithm was manipulated so that negative association rules can be produced efficiently.

The method proposed in [17] extracts association rules in two stages. Firstly, general rules with smaller lengths will produce. Afterwards, those more specific rules will be extracted, Those rules that are smaller in length and are more general, are often supported more than other rules in training data. To produce specific rules much lower min-support is required to produce association rules. In all of those previously proposed methods, the main concentration was on the improvement of association rules extraction, in contrast to their work, in this paper we tried to use local weighting of inputs (which relies on sampling) to improve the accuracy of classification.

3. The Proposed method

Our proposed method will be done in 4 stages. We begin with the weighting of input data, some data will be produced based on the distribution of dataset. Then these produced data will be added to the training dataset. Now by using the new dataset, association rules will be extracted. Finally, these obtained rules will be pruned and classification will be done.

3.1. Local weighting of inputs

In this paper our main concentration is on the weighting of inputs, this weighting will result in the production of rules with higher quality. The method we used for local weighting of inputs is based on the sampling of data according to the probability distribution function of inputs. Sampling means that new training data samples can be produced. With this the support for repetitive values will increase, in addition to that, association rules will be produced with higher reliability. The proposed sampling won't be done evenly. It means that the function for producing new samples won't be a uniform distribution function. By using different distribution functions, different weights can be allocated to each area of input data (Local weighting). For example, in (figure 1), two balanced classes with 100 two dimensional (having two features) samples have been shown. One can add new date with arbitrary distribution per each class and then new models (rule extraction) can be extracted from this class. In (figure 2) per each class 50 new samples with normal distribution are going to be added. As you can see in (figure 2), due to the increment in the number of samples in the center of each class, the resolution power between classes will increase and it is easier to imagine a separate boundary between two different classes. It can be shown that by weighting of data, usually stronger rules will be produced in the central areas of data. As data scatters from the center, weaker rules will be produced. These modifications, improve classification accuracy.

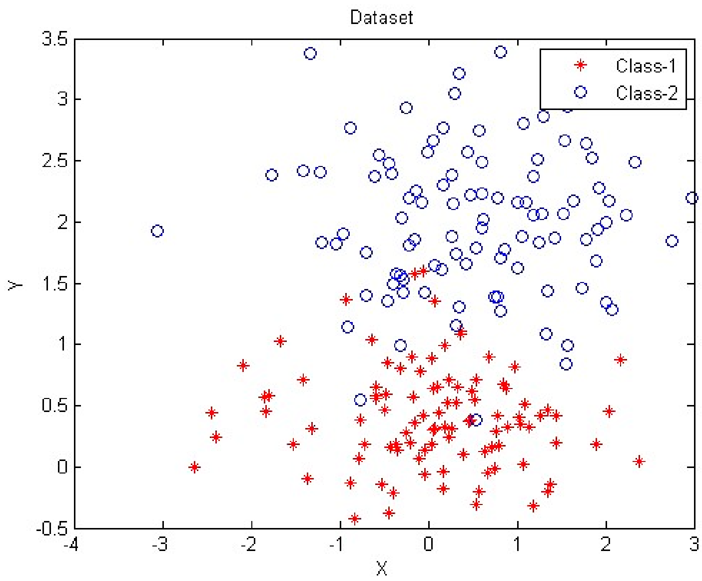

Figure 1: two dimensional dataset with two classes.

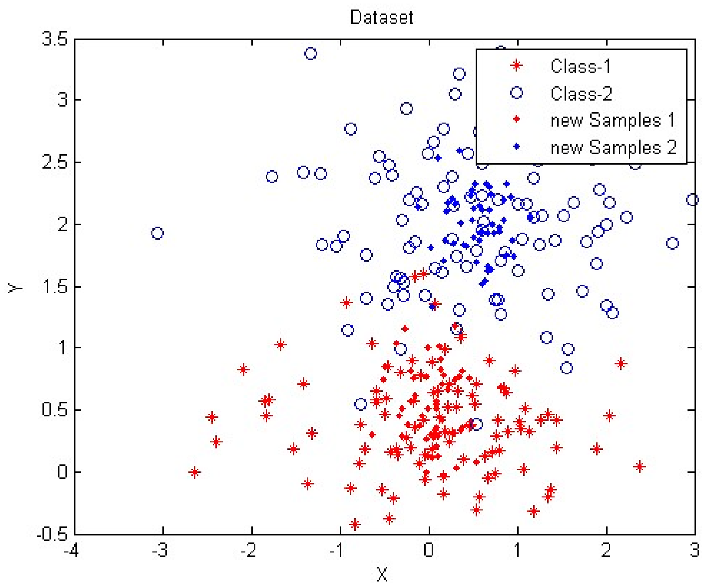

Figure 1: *A*dding samples to weighting the inputs.

3.1.1. Estimation of probability distribution function: As it was stated in previous example, artificial data (those data that were produced for the purpose of weighting) must be added to dataset according to the original data distribution. To do this we have to calculate the probability distribution function of each one of the classes of the problem. Since the probability function of original data is not clear we have to estimate the probability of data.

In statistics and probability, density estimation is the process of probability density function estimation of a random variable by using the visited samples of that variable. Usually it is supposed that the visited samples were distributed randomly and separately based on the probability distribution function [19].

The estimation of probability density can be done by using either parametric or non-parametric method [20]. In this article we used histogram method, to estimate the probability density of the data in each class. Based on the number of data in each bin, the same number of artificial data in the same bin will be produced. This strengthens the repetitive items and attenuation of non-repetitive items.

3.2. Association rules extraction

Since we locally weighted the dataset now can use any algorithm to extract association rules. In this article we used Appriori algorithm [23], this is one of the well-known algorithms for rule extraction, and Predictive Apriori algorithm [24] to extract association rules. Predictive Apriori provided a trade-off between support and confidence, so that there is a better chance of correct prediction for unseen data. To do this we use a criterion that is called predictive accuracy, this criterion is based on Bayesian framework and it shows the amount of the interference of support and confidence in the accuracy of prediction.

3.3. Pruning of extracted rules

After rule extraction, rules have to be pruned, so that the model accuracy and model speed for the classification of unvisited data enhances. In other words, our proposed method tries to increase generality by pruning. By pruning of rules according to attraction criterion, rules will be arranged and some of those rules with higher attraction will be selected, the classification will use this rules.

4. Experimental result

In this section, we are going to compare the results of experiments and then we are going to analyze them. In addition to that, our proposed method was compared to other common classification methods that are based on association rules.

4.1. Local weighting of inputs

To study the performance of our purposed method in the field of local weighting of inputs, an experiment was designed by the authors. This experiment was performed twice, one with common method (without weighting) and another time with allocating local weights to data, then in both cases the association rules on the IRIS dataset were obtained. The results of these experiments were presented in table 1 and table 2. IRIS dataset have four numeric characteristics (SW, SL, PW, PL). In this experiment, all of the configurations of Apriori algorithm are the same and their only difference is in weighting. The configuration we used here uses Irani [26] method for discretization. Min-support is selected as 0.1 and max-support of rules is chose to be 0.9. In addition to that 10 superior association rules have be chosen for classification.

Table 1: Extracted rules from IRIS dataset by Apriori algorithm with weighting.

Extracted rule			Class label	Confidence
PL='(-∞ 2.6]'			Setosa	1
PW='(-∞ 0.8]'			Setosa	1
PW='(-∞ 0.8]'	PL='(-∞ 2.6]'		Setosa	1
PL='(-∞ 2.6]'	SL='(-∞ 5.45]'		Setosa	1
PW='(-∞ 0.8]'	SL='(-∞ 5.45]'		Setosa	1
PW='(-∞ 0.8]'	PL='(-∞ 2.6]'	SL='(-∞ 5.45]'	Setosa	1
PL='(5.15 ∞)'			Virginica	1
PL='(5.15 ∞)'	PW='(1.75 ∞)'		Virginica	1
PL='(2.6 4.75]'			Versicolor	0.99
PL='(2.6 4.75]'	SW='(-∞ 3.05]'		Versicolor	0.99

Table 2: Extracted rules from IRIS dataset by Apriori algorithm without weighting.

Extracted rule			Class label	Confidence
PL='(-∞ 2.45]'			Setosa	1
PW='(-∞ 0.8]'			Setosa	1
PW='(-∞ 0.8]'	PL='(-∞ 2.45]'		Setosa	1
PL='(-∞ 2.45]'	SL='(-∞ 5.55]'		Setosa	1
PW='(-∞ 0.8]'	SL='(-∞ 5.55]'		Setosa	1
PW='(-∞ 0.8]'	PL='(-∞ 2.45]'	SL='(-∞ 5.55]'	Setosa	1
PW='(1.75 ∞)'			Virginica	0.98
PW='(1.75 ∞)'	PL='(4.75 ∞)'		Virginica	0.98
PL='(2.45 4.75]'			Versicolor	0.98
PL='(2.45 4.75]'	PW='(0.8 1.75]'		versicolor	0.98

As you can see in table 1 and table 2, weighting changed the distribution of data. As it was expected, stronger rules were extracted. Table 1 shows that the rules with higher confidence were produced. Moreover it shows that in classification without weighting SW property was not used however it is useful and it should be considered. According to table 1, after the weighting of data, this property was used.

4.2. Pruning of rules (number of rules)

This experiment is related to choosing the number of rules. In this experiment we will investigate the impact of the number of rules on the accuracy of each model on the IRIS dataset. In this experiment we did not weight the rules so that we can study the impact of the number of rules, since our main purpose was to investigate this matter.

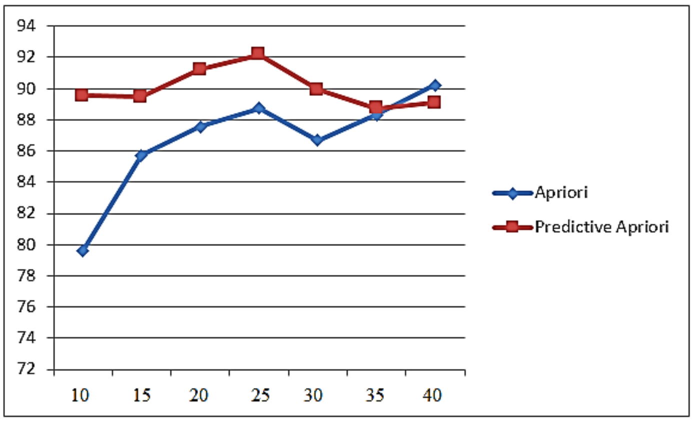

Figure 3: the impact of number of rules on the accuracy of two algorithms.

In the diagram of figure 9 we can see that the predicted Apriori algorithm with few rules functions better but with a large number of rules, Aprori algorithms performs better.

4.3. Comparison

For the evaluation of our proposed method, we compared our method with two different categories of classification algorithms. The first category is consisted of classic classification algorithms, and the second category is consisted of modern classification algorithms (these are based on association rules). In the first comparison, 5 runs of 10 fold cross validation per each class bands were used. All of the algorithms that we studied in this section benefits from the default values of Weka software [27]. The results have been presented in table 3.

Table 3: The comparison of accuracy obtained by our proposed method and other classification algorithms

	Proposed Method	Part	C4.5	3-NN	Naïve Bayes	MLP
Iris	**95.64**	92.27	93.1	92.04	94.65	93.15
tic-tac-toe	95.24	91.52	82.31	95.58	66.36	**96.67**
Wine	**96.96**	90.63	92.74	93.58	95.48	95.96
Glass	**78.63**	66.54	67.54	68.7	46.58	65.88
breast-cancer	**74.13**	66.09	71.98	69.79	71.53	66.51

In table 4, the proposed method was compared to the methods that were proposed in [4], [13], [28] and [29]. As it shows the accuracy of our proposed method was significantly improved in most of cases.

Table 4: The comparison of our proposed method with recent proposed methods

	CBC-fs	CBC	DA-AC	GARC	CBA	Proposed Method
Breast-Cancer	74.57	73.42	71.44	73.84	74.01	**75.84**
Glass	71.96	73.23	73.27	67.69	70.9	**79.81**
IRIS	92.41	92.82	93.27	93.23	91.38	**96.58**
Tic-Tac-Toe	98.06	98.38	85.54	**99.85**	99.26	95.81
Wine	90.25	90.05	88.68	82.61	91.25	**97.34**

Conclusion

One of the beneficial methods for classification is to use association rules. Association rules state a meaningful relationship between the items of a dataset and they can be shown with a number of rules. In this article we used local weighting for inputs (data) to extract higher quality rules. For data weighting, we used data production based on dataset distribution. Then by the pruning of the extracted association rules, the accuracy of classification improves. To investigate the performance of our proposed method, we compared the performance of two categories of classification algorithms, classic and modern classification algorithms, with the performance of our proposed method. Obtained results perfectly reflected the capabilities of our purposed methods in the field of classification. In fact, our proposed method improved the accuracy of classification by extracting the association rules more precisely and by utilizing the data properties in a more efficient manner.

References

[1] R. Agrawal, T. Imieliński, and A. Swami, “Mining association rules between sets of items in large databases,” in ACM SIGMOD Record, 1993, vol. 22, no. 2, pp. 207–216.

[2] K. Ali, S. Manganaris, and R. Srikant, “Partial Classification Using Association Rules.,” in KDD, 1997, vol. 97, pp. p115–118.

[3] M. S. Chen, J. S. Park, and P. S. Yu, “Data mining for path traversal patterns in a web environment,” in Distributed Computing Systems, Proceedings of the 16th International Conference on, 1996, pp. 385–392.

[4] B. L. W. H. Y. Ma, “Integrating classification and association rule mining,” in Proceedings of the fourth international conference on knowledge discovery and data mining, 1998.

[5] W. Li, J. Han, and J. Pei, “CMAR: Accurate and efficient classification based on multiple class-association rules,” in Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on, 2001, pp. 369–376.

[6] J. Han, J. Pei, and Y. Yin, “Mining frequent patterns without candidate generation,” in ACM SIGMOD Record, 2000, vol. 29, no. 2, pp. 1–12.

[7] X. Yin and J. Han, “CPAR: Classification based on Predictive Association Rules.,” in SDM, 2003, vol. 3, pp. 369–376.

[8] J. R. Quinlan and R. M. Cameron-Jones, “FOIL: A midterm report,” in Machine Learning: ECML-93, 1993, pp. 1–20.

[9] A. Zimmermann and L. De Raedt, “Corclass: Correlated association rule mining for classification,” in Discovery Science, 2004, pp. 60–72.

[10] M.-L. Antonie and O. R. Zaïane, “An associative classifier based on positive and negative rules,” in Proceedings of the 9th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery, 2004, pp. 64–69.

[11] F. Thabtah, P. Cowling, and Y. Peng, “MCAR: multi-class classification based on association rule,” in Computer Systems and Applications, 2005. The 3rd ACS/IEEE International Conference on, 2005, p. 33.

[12] J. Wang and G. Karypis, “HARMONY: Efficiently mining the best rules for classification,” in SDM, 2005, vol. 5, pp. 205–216.

[13] G. Chen, H. Liu, L. Yu, Q. Wei, and X. Zhang, “A new approach to classification based on association rule mining,” Decis. Support Syst., vol. 42, no. 2, pp. 674–689, 2006.

[14] F. A. Thabtah, P. Cowling, and Y. Peng, “Multiple labels associative classification,” Knowl. Inf. Syst., vol. 9, no. 1, pp. 109–129, 2006.

[15] F. A. Thabtah and P. I. Cowling, “A greedy classification algorithm based on association rule,” Appl. Soft Comput., vol. 7, no. 3, pp. 1102–1111, 2007.

[16] G. Kundu, M. M. Islam, S. Munir, and M. F. Bari, “ACN: An associative classifier with negative rules,” in Computational Science and Engineering, 2008. CSE’08. 11th IEEE International Conference on, 2008, pp. 369–375.

[17] G. Kundu, S. Munir, M. F. Bari, M. M. Islam, and K. Murase, “A novel algorithm for associative classification,” in Neural Information Processing, 2008, pp. 453–459.

[18] Y.-Z. Liu, Y.-C. Jiang, X. Liu, and S.-L. Yang, “CSMC: A combination strategy for multi-class classification based on multiple association rules,” Knowledge-based Syst., vol. 21, no. 8, pp. 786–793, 2008.

[19] J. Friedman, T. Hastie, and R. Tibshirani, The elements of statistical learning, vol. 1. Springer series in statistics Springer, Berlin, 2001.

[20] B. W. Silverman, Density estimation for statistics and data analysis, vol. 26. CRC press, 1986.

[21] M. K. Varanasi and B. Aazhang, “Parametric generalized Gaussian density estimation,” J. Acoust. Soc. Am., vol. 86, no. 4, pp. 1404–1415, 1989.

[22] V. A. Epanechnikov, "Non-parametric estimation of a multivariate probability density," Theory Probab. Its Appl., vol. 14, no. 1, pp. 153–158, 1969.

[23] J. Han, M. Kamber, and J. Pei, Data mining: concepts and techniques: concepts and techniques. Elsevier, 2011.

[24] T. Scheffer, "Finding association rules that trade support optimally against confidence," Intell. Data Anal., vol. 9, no. 4, pp. 381–395, 2005.

[25] V. N. Vapnik and V. Vapnik, Statistical learning theory, vol. 1. Wiley New York, 1998.

[26] K. B. Irani, "Multi-interval discretization of continuous-valued attributes for classification learning," 1993.

[27] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," ACM SIGKDD Explor. Newsl., vol. 11, no. 1, pp. 10–18, 2009.

[28] H. Deng, G. Runger, E. Tuv, and W. Bannister, "CBC: An associative classifier with a small number of rules," Decis. Support Syst., vol. 59, pp. 163–170, 2014.

[29] V. Mangat and R. Vig, "Novel associative classifier based on dynamic adaptive PSO: Application to determining candidates for thoracic surgery," Expert Syst. Appl., vol. 41, no. 18, pp. 8234–8244, 2014.