A Review on Class Association Rule Mining

Ms. Ayushi Biyawala[1], Mr. Mihir Patel[2], Ms. Himani Parekh[3]

[1] Dept. of Computer Engineering and Information Technology,Uka tarsadiya university,
[2] Dept. of Computer Engineering and Information Technology,Uka tarsadiya university,
[3]Dept. of Computer Engineering and Information Technology,Uka tarsadiya university,

Abstract— In a real world huge amounts of data are collected daily. Extracting useful and interesting knowledge from those data is called data mining. There are several techniques for extracting knowledge like clustering, classification, association mining, outlier detection etc. Association Classification (AC) is an important study, which supporting classification and prediction. There are some well-known AC algorithm: CBA, CMAR, MCAR, FACA, ECBA and WCBA. The AC techniques uses the association rules in the classification process to enhance hidden pattern for better accuracy of the classification process.

Keywords— Association classification, Classification, Prediction, Weighted approach, Harmonic Mean (HM) measure

I. INTRODUCTION

Data mining is all about to extract the knowledge form large amounts of data, that is known as Knowledge Discovery (KDD) process. An important task of data mining is to build accurate and efficient classifiers for huge datasets. Data mining focuses on three problems, mining association rules, classification, clustering and Association classification (AC). Association Classification (AC) is an important study, which supporting classification and prediction.

Classification techniques are used to predict the class label for any given instance. The goal of classification rule mining (CRM) is to mine a small set of rules that builds an accurate classifier *[1]*. The AC techniques uses the association rules in the classification process to enhance hidden pattern for better accuracy of the classification process *[1]*. Some AC algorithms mainly uses two estimated values for association rule generation: support and confidence, which are given by users.

II. ASSOCIATION RULE MINING

Association rules finds frequent patterns, associations, correlations or informal structures from sets of items or we can say objects in transactional databases. Mining association rules from massive amount of data in the database can help in many business decision making processes. There are techniques for discovering association rules from the data, which have focused on identifying relationships between items traditionally, and show some aspect of human behavior.

The two basic measures of association rules are support threshold and confidence threshold. If there are two items then support is defined as the ratio of occurrence of that two items and total number of transactions *[2]*. Minimum support is that if rules which have support greater than a user defined support. And the possibility of seeing the rule's consequent under the condition that the transactions also contain the antecedent is called confidence *[2]*. Also minimum confidence is one in which rules have a confidence greater than a user defined confidence.

III. CLASS ASSOCIATION RULE MINING

Associative Classification (AC) is a classification learning approach in data mining which contains association rule discovery methods and classification technique for building the accurate classifier. In the special case of the association rule, the target attribute is considered in the rule's right hand side. This approach depends on two important thresholds named minimum support (MinSupp) and minimum confidence (MinConf).

Class association rule mining is most important study for researchers now a days. It is supporting classification and prediction methods. A particular subset of association rules, in which right-hand side is indicate as class attribute is mined *[1]*. This subset of rules is denoted as class association rules(CARs) *[1]*. There are some methods available, which include classification based on predictive association rules as follows: classification based on multiple association rules, multi class classification based on association rules, multi-label association classification.

An AC algorithm works in three main phases. In the first phase, it searches for hidden correlations between the attribute values and the class attribute values in the training data set. Once all frequent ruleitems are found, the rules "Class Association Rule" (CARs) are generated from them in "if-then" format *[3]*. Ranking and pruning procedures start process in the second phase. At this stage, CARs are ranked according to a number of parameters like confidence and support values to ensure that the rules with high confidence are given higher priority for selected as part of the classifier *[3]*. However, since the number of generated rules run into several thousands, and many of them are redundant and not discriminative both, among the classes, rule pruning are needed to discard the contradicting and duplicating rules from

the complete set of CARs *[3]*. The output of the second phase is the set of CARs which represents the final classifier model *[3]*. At last, the classification model is going to use for predicting the class values on new unseen data set (test data).

IV. LITERATURE SURVEY

A number of research works has been done in the association rule mining field.

liu et al. *[4]* developed an algorithm based on AC technique, named Classification Based on Association Rules (CBA). CBA algorithm contains three stages : rule generation, rule pruning and prediction. In rule generation apriori algorithm is used to discover the frequent patterns that represent the class association rules (CARs), rule pruning is responsible for selecting the best rules from the generated rules and prediction is used to predict unknown instances. The most important phase of CBA algorithm is to build a classifier. The CBA-CB algorithm for building a classifier using CARs, which is known as M1 method *[4]*. Here a classifier built by CBA-CB algorithm is performs very well as compared to that built by C4.5.

Classification Based on Multiple Association Rules (CMAR) is developed by Li et al. *[5]*. It is based on the integration of the association rules and the classification techniques, like other AC algorithms. Here a new technique for the rule-generation and classification phases is implemented. At the rule-generation phase, to discover patterns CR-tree and FP-tree algorithms were used instead of Apriori algorithm *[5]*. The classification stage is depends on finding the class label for its unknown instances by finding all patterns that can be matched with these instances and analyzing all of these patterns for class prediction. From some experiments it is proved that CMAR is better than CBA and C4.5 algorithms.

The Multi-class Classification Based on Association Rule (MCAR) algorithm *[6]* is developed by Thabtah et al. to solve the multi-scanning problem of the CBA algorithm to generate association rules. They performed experiment with fifteen different data sets, the results indicated that the proposed MCAR method is more accurate and efficient classification technique than CBA. Here a single itemset was generated using the Tid-list method which stored the occurrence positions for each item for the next phase during the rule generation process without doing an extra scanning of the dataset. If compared with the classifier generated by CBA, the classifiers produced by the MCAR are highly competitive with regards to error rate and efficiency.

Hadi et al. *[7]* implemented a new Fast Associative Classification Algorithm (FACA). In the rule generation process the Diffset method has been used to enhance the speed of building the model. FACA sorts the discovered rules followed by the least number of attributes in the left hand side, with confidence, support and rule generated first respectively. For the prediction phase, it proposed a multiple rules method to increase the accuracy of the classification process. The bases of the this experiment are classification accuracy and the F1 evaluation measures. The results show that FACA is more accurate in terms of the F1 evaluation measure compared with the other four well-known algorithms (CBA, CMAR, MCAR and the Enhanced Class Association Rule (ECAR)).

Alwidian et al. *[8]* proposed the Enhanced CBA (ECBA) algorithm which is based on Apriori Optimization and the Statistical Ranking Measure *[8]*. To evaluate its accuracy and speed, algorithm was compared with CBA, CMAR and MCAR algorithms. Results showed better performance in terms of accuracy for ECBA compared with the other algorithms. ECBA algorithm enhanced the time complexity of building a classifier model of the original CBA algorithm and the rate of classification accuracy is based on optimizing Apriori algorithm with the use of set of clusters to aggregate the items that have the same features to generate new rules and eliminate any impossible combination and the accuracy of the classification process has been enhanced by using a new HM measure to make more fair enough rules prioritization process and generating new association rules that are used in the predication phase. In proposed algorithm HM measure is used to produces hidden strong correlation rules which are not discovered by above three considered AC algorithms.

Nguyen et al. *[1]* developed "A Weighted Approach for Class Association Rules" which supporting classification and prediction methods. Here presented a method for mining CARs, is modified equivalence class rules tree (MECR-tree) *[9]* to generate set of rules quickly. In weighted approach, each attribute is associated with a weighted value so it uses the interestingness measure combines with support and confidence to improve the accuracy than the CAR-Miner.

Alwidian et al. *[10]* enhanced the CBA algorithm and developed a new weighting algorithm known as WCBA(Weighted Classification Based on Association rules) *[10]*. Traditional rule mining algorithms follows the fact, whether generated rule is important or not, which is depends on the count of its itemsets in a dataset. Traditional rule mining algorithms consider support and confidence thresholds to generate frequent itemsets. But some business makers and other researchers put a different point, that whether generated rule is interesting or not, which can depends on the number of appearance of an item in dataset and human interpretation. So here not only based on the dataset, weights are also can be use to represent the knowledge and generate rules *[10]*. W.Wang defined a new approach, named Weighted Association Rule Mining (WARM) *[11]*, which modifies the traditional rule mining algorithms by assigning weights to the items in dataset. Weights reflect the importance of an item in given dataset. Here weights are used as input into WARM algorithm and applies the weights to the items and count weighted support to generate a list of interesting rules. Like the other AC algorithms, WCBA performs in three general stages: rule generation, rule pruning and prediction *[10]*. The basic purpose of WCBA is to replace the traditional estimated support-confidence method of the association rule mining model by the weighted model. Attributes that have major influence in datasets were assigned the highest priority by subject matter experts *[10]*. Each attribute is associated with a weighted value, which assigned by experts. A statistical harmonic mean (HM) measure is used to prioritize the association rules at the pruning and generation phases.

V. COMPARATIVE ANALYSIS OF DIFFERENT AC ALGORITHMS

Sr. No.	Algorithm	Rule generation method	Prediction method	Advantages
1	CBA [4]	Apriori candidate generation	Maximum likelihood	Build an accurate classifier and solve a number of problems in existing classification systems, such as C4.5
2	CMAR [5]	FP-growth approach	CMAR multiple label	Consistent, highly effective at classification and has better average classification accuracy in comparison with CBA and C4.5
3	MCAR [6]	Tid-list intersection method	Maximum likelihood	Discover and generate frequent items and rules in onne phase only and introduce a rule ranking technique that minimizes the use of randomisation
4	FACA [7]	Diffset	All exact match prediction method	Gives a better accuracy compare with CBA, CMAR, MCAR and ECBA.
5	ECBA [8]	Optimized apriori	Maximum likelihood	It outperformed the CBA, MCAR and CMAR in terms of accuracy, precision, recall and F1 measures
6	WCBA [9]	Weights by experts, apriori candidate generation by weights	Multiple rules based on HM measure	Overcome the problem of estimated values by assigning the weights and calculating HM values and gives better accuracy to compare with all other AC algorithms

VI. CONCLUSIONS

The problem of estimated support and confidence can be overcome by the concept of harmonic mean (HM) measures and weighted approach shown in WCBA algorithm. Because of these features WCBA gives more accuracy than other AC algorithms.

REFERENCES

[1] Loan T. T. Nguyen, Bay Vo, Thang Mai, Thanh-Long Nguyen, "A weighted Approach for Class Association Rules",Springer International Publishing, Morden Approaches for Intelligent Information and Database Systems, pp. 213-222, 2018
[2] Surbhi K. Solanki, Jalpa T. Patel, "A Survey on Association Rule Mining", Fifth IEEE International Conference on Advanced Computing & Communication Technologies, PP. 212-216, 2015.J.
[3] Suzan Wedyan, "Review and Comparison of Associative Classification Data Mining Approaches", International Scholarly and Scientific Research & Innovation, Vol:8, No:1, pp. 34-45, 2014.
[4] Bing Liu, Wynne Hsu, Yiming Ma, "Integrating Classification and Association Rule Mining", KDD'98 Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, pp. 80-86, 1998.
[5] W. Li, J. Han, and J. Pei, "CMAR: accurate and efficient classification based on multiple class-association rules", Proceedings 2001 IEEE International Conference on Data Mining, pp. 369-376, 2001.
[6] F. Thabtah, P. Cowling, and Y. Peng, "MCAR: multi-class classification based on association rule", The 3rd ACS/IEEE International Conference on Computer Systems and Applications, pp. 33, 2005.
[7] W. E. Hadi, F. Aburub, and S. Alhawari, "A new fast associative classification algorithm for detecting phishing websites", Applied Soft Computing, vol. 48, pp. 729-734, 2016.

[8]   Jaber Alwidian, Bassam Hammo, Nadim Obeid, "Enhanced CBA algorithm Based on Apriori Optimization and Statistical Ranking Measure", Conference: The 28th IBIMA conference on Vision 2020: Innovation Management, Development Sustainability, and Competitive Economic Growth, At Seville, Spain, 2016.

[9]   L. T. Nguyen and N. T. Nguyen, "An improved algorithm for mining class association rules using the difference of Obidsets", Expert Systems with Applications, vol. 42, no. 9, pp. 4361-4369, 2015.

[10] Alwidian, B. H. Hammo, and N. Obeid, "WCBA: Weighted classification based on association rules algorithm for breast cancer disease", Applied Soft Computing, vol. 62, pp. 536-549, 2018.

[11] W. Wang, J. Yang, P.S. Yu, "Efficient mining of weighted association rules(WAR)", Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 270–274, 2000.

[12] Y. J. Wang, Q. Xin, and F. Coenen, "A Novel Rule Weighting Approach in Classification Association Rule Mining", Seventh IEEE International Conference on Data Mining Workshops (ICDMW 2007), pp. 271-276, 2007.

[13] B. Sunil Kumar, B. Raja Sekhar, L. Swathi Reddy, "A class of association rule for pre-assigned weights", Asian Journal of Computer Science And Information Technology, pp. 213-216, 2012.

[14] C. M. Oliveira and D. A. Pereira, "An association rules based method for classifying product offers from e-shopping", Intelligent Data Analysis, vol. 21, no. 3, pp. 637-660, 2017.

[15] Huong Bui, Bay Vo, Ham Nguyen, Tu-Anh Nguyen-Hoang, Tzung-Pei Hong, "A weighted N-list-based method for mining frequent weighted itemsets", Expert Systems with Applications: An International Journal archive, vol. 96, pp. 388-405, 2018.

[16] Loan T.T. Nguyen, Bay Vo, Tzung-Pei Hong, Hoang Chi Thanh, "CAR-Miner: An efficient algorithm for mining class-association rules", Expert Systems with Applications, vol. 40, pp. 2305-2311, 2013.

[17] Jaber Alwidian, Bassam Hammo, Nadim Obeid, "FCBA: Fast Classification Based on Association Rules Algorithm", IJCSNS International Journal of Computer Science and Network Security, Vol. 16 No. 12, pp. 117-127, 2016.