ISSN 23499842(Online), Volume 1, Special Issue 2(ICITET 15), March 2015
International Journal of Innovative Trends and Emerging Technologies

A SCRUTINY OF ASSOCIATION RULE MINING ALGORITHMS

M. MADHANKUMAR[A] AND Dr. C. SURESHGANADHAS[B]
[a] Research Scholar, Department of Computer Science and Engineering, Manonmaiam Sundaranar University, Tamil Nadu, India
[b] Professor & Head, Department of Computer Science and Engineering, Vivekanandhan College of Engineering for Women, Elayampalayam, Thiruchencode, Tamil Nadu
manomadhanbe@gmail.com

ABSTRACT - Association Rule Mining (AM) is one of the most popular data mining techniques. This paper provides a brief review about the existing Association Rules Data Mining. Association rule mining is normally performed in generation of frequent itemsets and rule generation in which many researchers presented several efficient algorithms. Association rule mining has been extensively studied in the data mining community. The intention of this review is to present a broad classification of existing association rule mining algorithms and to motivate for the research.

Keywords – Data Mining, Association Rules, ARM, Algorithms.

I. INTRODUCTION

Association rule learning is a popular and well researched method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using different measures of interestingness.[1] Based on the concept of strong rules, Rakesh Agrawal et al.[2] introduced association rules for discovering regularities between products in large-scale transaction data recorded by point-of-sale (POS) systems in supermarkets.
Following the original definition by Agrawal et al.[2] the problem of association rule mining is defined as: Let $I = \{i_1, i_2, \ldots, i_n\}$ be a set of n binary attributes called items. Let $D = \{t_1, t_2, \ldots t_m\}$ be a set of transactions called the database. Each transaction in D has a unique transaction ID and contains a subset of the items in I. A rule is defined as an implication of the form $X \Rightarrow Y$ where $X, Y \subseteq I$ and $X \cap Y = \emptyset$. The sets of items X and Y are called antecedent (left-hand-side or LHS) and consequent (right-hand-side or RHS) of the rule respectively.

Association rules are usually required to satisfy a user-specified minimum support and a user-specified minimum confidence at the same time. Association rule generation is usually split up into two separate steps:

1. First, minimum support is applied to find all frequent itemsets in a database.
2. Second, these frequent itemsets and the minimum confidence constraint are used to form rules.

While the second step is straightforward, the first step needs more attention.
Many algorithms for generating association rules were presented over time.
Some well known algorithms are Apriori, Eclat and FP-Growth, but they only do half the job, since they are algorithms for mining frequent itemsets. Another step needs to be done after to generate rules from frequent itemsets found in a database.

I. ASSOCIATION RULE MINING APPROACHES

The Table 1. presents the classification of the Association rule mining is a well explored research area, we will discuss about basic and classic approaches for association rule mining

Table 1. Existing Association Rules in Data Mining Techniques and Algorithms

Author	Algorithms	Advantages
Agrawal et al. 1993 [3]	AIS Algorithm. (Agrawal, Imielinski, Swami))	Agrawal et al. (1993) firstly proposed pattern mining concept in form of market based analysis for finding association between items bought in a market.It focus on improving the quality of databases together with necessary functionality to process decision support queries.
Agrawal and Srikant 1994. [4]	Apriori Algorithm.	The AIS is just a straightforward approach that requires many passes over the database, generating many candidate itemsets and storing counters of each candidate while most of them turn out to be not frequent.
Srikant and Agrawal 1996. [5]	Multiple Dimensional ARM.	Multiple dimensional association rule mining is to discovery the correlation between different predicts/attributes. Each attribute/predict is called a dimension, such as: age, occupation and buys in this example. At the same time multiple dimensional association rule mining concerns all types of data such as boolean data, categorical data and numerical dat.
Cheung et al. 1997. [6]	Maintaining of Association Rules.	From the definition of data mining, we can see that the object of data mining is data stored in very large repositories. The giant amount of data poses a challenge of maintaining and updating the discovered rules while the data may change from time to time in different ways. The FUP (Fast UPdate) algorithm was introduced to deal with insertion of new transaction data.
Han and Kamber 2000. [7]	Multiple Concept Level ARM.	Multiple level association rule mining is trying to mine strong association rules among intra and inter different levels of abstraction. For example, besides the association rules between milk and ham, it can generalize those rules to relation between drink and meat, at the same time it can also specify relation between certain brand of milk and ham.
Pei and Han 2000. [8]	Constraints based ARM.	In order to improve the efficiency of existing mining algorithms, constraints were applied during the mining process to generate only those association rules that are interesting to users instead of all the association rules.
Han et al. 2000. [9]	FP-Tree (Frequent Pattern Tree) Algorithm.	The frequent itemsets are generated with only two passes over the database and without any candidate generation process. By avoiding the candidate generation process and less passes over the database, FP-Tree is an order of magnitude faster than the Apriori algorithm.
Das et al. 2001 [10]	Rapid Association Rule Mining (RARM)	RARM is claimed to be much faster than FP-Tree algorithm with the experiments result shown in the original paper. By using the SOTrieIT structure RARM can generate large 1-itemsets and 2-itemsets quickly without scanning the database for the second time and candidates generation.
John D. Holt and Soon M. Chung. 2002. [11]	Hashing and Pruning (IHP) for mining association rules	A new algorithm named Inverted Hashing and Pruning (IHP) for mining association rules between items in transaction databases. The performance of the IHP algorithm was evaluated for various cases and compared with those of two well-known mining algorithms, Apriori algorithm [Proc. 20th VLDB Conf., 1994, pp. 487–499] and Direct Hashing and Pruning algorithm [IEEE Trans. on Knowledge Data Engrg. 9 (5) (1997) 813–825]. It has been shown that the IHP algorithm has better performance for databases with long transactions.
Yi-Chung Hu et al.	Fuzzy Grids	A new algorithm named fuzzy grids based rules mining

2003. [12]	Based Rules Mining Algorithm (FGBRMA)	algorithm (FGBRMA) is proposed to generate fuzzy association rules from a relational database. The proposed algorithm consists of two phases: one to generate the large fuzzy grids, and the other to generate the fuzzy association rules. A numerical example is presented to illustrate a detailed process for finding the fuzzy association rules from a specified database, demonstrating the effectiveness of the proposed algorithm.
Feng-Hsu Wang and Hsiu-Mei Shao. 2004. [13]	Hierarchical Bisecting Medoids Algorithm (HBM)	The author proposed a new clustering method, called HBM (Hierarchical Bisecting Medoids Algorithm) to cluster users based on the time-framed navigation sessions. Those navigation sessions of the same group are analyzed using the association-mining method to establish a recommendation model for similar students in the future. Finally, an application of this recommendation method to an e-learning web site is presented, including plans of recommendation policies and proposal of new efficiency measures. The effectiveness of the recommendation methods, with and without time-framed user clustering, are investigated and compared.
Yuh-Jiuan Tsay and Jiunn-Yann Chiang. 2005. [14]	Cluster-Based Association Rule (CBAR)	The author presented an efficient algorithm named cluster-based association rule (CBAR). The CBAR method is to create cluster tables by scanning the database once, and then clustering the transaction records to the *k*-th cluster table, where the length of a record is *k*. Moreover, the large itemsets are generated by contrasts with the partial cluster tables. This not only prunes considerable amounts of data reducing the time needed to perform data scans and requiring less contrast, but also ensures the correctness of the mined results.
Guoqing Chen et al. 2006. [15]	Gain based Association Rule Classification (GARC)	The author presented a new approach for constructing a classifier, based on an extended association rule mining technique in the context of classification. The characteristic of this approach is threefold: first, applying the information gain measure to the generation of candidate itemsets; second, integrating the process of frequent itemsets generation with the process of rule generation; third, incorporating strategies for avoiding rule redundancy and conflicts into the mining process.
Frans Coenen and Paul Leng. 2007. [16]	Classification Association Rule Mining (CARM)	Classification Association Rule Mining (CARM) systems operate by applying an Association Rule Mining (ARM) method to obtain classification rules from a training set of previously classified data. The rules thus generated will be influenced by the choice of ARM parameters employed by the algorithm (typically support and confidence threshold values). In this paper we examine the effect that this choice has on the predictive accuracy of CARM methods.
He Jiang et al. 2008. [17]	Weighted Association Rules (WARs)	The weighted association rules (WARs) mining are made because importance of the items is different. Negative association rules (NARs) play important roles in decision-making. But the misleading rules occur and some rules are uninteresting when discovering positive and negative weighted association rules (PNWARs) simultaneously.
Yuanyuan Zhao et al. 2009. [18]	Weighted Negative Association Rules	The Negative association rules become a focus in the field of data mining. Negative association rules are useful in market-basket analysis to identify products that conflict with each other

	(WNARs)	or products that complement each other. The negative association rules often consist in the infrequent items. The experiment proves that the number of the negative association rules from the infrequent items is larger than those from the frequent.
Tongyan Li and Xingming Li. 2010 [19]	Association Rules Mining based Alarm Correlation Analysis System (ARM-ACAS)	A novel Association Rules Mining based Alarm Correlation Analysis System (ARM-ACAS) to find interesting association rules between alarm events. In order to mine some infrequent but important items, ARM-ACAS first uses neural network to classify the alarms with different levels. In addition, ARM-ACAS also exploits an optimization technique with the weighted frequent pattern tree structure to improve the mining efficiency.
WeiminOuyang et al. 2011. [20]	Mining Fuzzy Association Rules	They propose mining fuzzy association rules to address the first limitation. In this they put forward a discovery algorithm for mining both direct and indirect fuzzy association rules with multiple minimum supports to resolve these three limitations.
IdhebaMohamad Ali O. Swesi et al. 2012. [21]	Interesting Multiple Level Minimum Supports (IMLMS) Algorithm	A new approach (PNAR_IMLMS) for mining both negative and positive association rules from the interesting frequent and infrequent item sets mined by the IMLMS model. The experimental results show that the PNAR_IMLMS model provides significantly better results than the previous model.
AnjanaGosainet al. 2013. [22]	Traditional Algorithm for Association Mining Rules	The traditional algorithms for mining association rules are built on binary attributes databases, which has two imitations. Firstly, it cannot concern quantitative attributes; secondly, it treats each item with the same significance although different item may have different significance.
Yiyong Xiao et al. 2014. [23]	Variable Neighbourhood Search (VNS) Algorithm	A variable neighbourhood search (VNS) algorithm is developed to solve the problem with near-optimal solutions. Computational experiments are performed to test the VNS algorithm against a benchmark problem set. The results show that the VNS algorithm is an effective approach for solving the MTFWS problem, capable of discovering many large-one frequent itemset with time-windows (FITW) with a larger time-coverage rate than the lower bounds, thus laying a good foundation for mining ARTW.
Vahid Beiranvand et al. 2014. [24]	Multi-Objective Particle Swarm Optimization Algorithm (MOPAR)	This paper deals with the numerical ARM problem using a multi-objective perspective by proposing a multi-objective particle swarm optimization algorithm (i.e., MOPAR) for numerical ARM that discovers numerical association rules (ARs) in only one single step. To identify more efficient ARs, several objectives are defined in the proposed multi-objective optimization approach, including confidence, comprehensibility, and interestingness. Finally, by using the Pareto optimality the best ARs are extracted.

II. CONCLUSION

In this paper we briefly reviewed the existing association rules mining in data mining applications. This review would be helpful to researchers to focus on the various issues of data mining. In future course, we will review the various classification algorithms and significance of evolutionary computing (genetic programming) approach in designing of efficient classification algorithms for data mining.

This study is prepared to our new project titled secure multi party algorithm for mining of association rules.

III. REFERENCES

[1] Piatetsky-Shapiro, Gregory (1991), Discovery, analysis, and presentation of strong rules, in Piatetsky-Shapiro, Gregory; and Frawley, William J.; eds., Knowledge Discovery in Databases, AAAI/MIT Press, Cambridge, MA.

[2] Agrawal, R.; Imieliński, T.; Swami, A. (1993). "Proceedings of the 1993 ACM SIGMOD international conference on Management of data - SIGMOD '93". p. 207. doi:10.1145/170035.170072.

[3] Agrawal, R., Imielinski, T., and Swami, A. N. 1993. Mining association rules between sets of items in large databases. In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, P. Buneman and S. Jajodia, Eds. Washington, D.C., 207-216.

[4] Agrawal, R. and Srikant, R. 1994. Fast algorithms for mining association rules. In Proc. 20th Int. Conf. Very Large Data Bases, VLDB, J. B. Bocca, M. Jarke, and C. Zaniolo, Eds. Morgan Kaufmann, 487{499.

[5] Srikant, R. and Agrawal, R. 1996. Mining quantitative association rules in large relational tables. In Proceedings of the 1996 ACM SIGMOD international conference on Management of data. ACM Press, 1-12.

[6] Cheung, D. W.-L., Lee, S. D., and Kao, B. 1997. A general incremental technique for maintaining discovered association rules. In Database Systems for Advanced Applications. 185-194.

[7] Han, J. and Kamber, M. 2000. Data Mining Concepts and Techniques. Morgan Kanufmann.

[8] Pei, J. and Han, J. 2000. Can we push more constraints into frequent pattern mining? In Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM Press, 350-354.

[9] Han, J. and Pei, J. 2000. Mining frequent patterns by pattern-growth: methodology and implications. ACM SIGKDD Explorations Newsletter 2, 2, 14-20.

[10] Das, A., Ng, W.-K., and Woon, Y.-K. 2001. Rapid association rule mining. In Proceedings of the tenth international conference on Information and knowledge management. ACM Press, 474 - 481.

[11] John D. Holt, Soon M. Chung, "M ining association rules using inverted hashing and pruning", Information Processing Letters, Volume 83, Issue 4, 31 August 2002, Pages 211-220, Elsevier.

[12] Yi-Chung Hu, Ruey-Shun Chen, Gwo-Hshiung Tzeng, "Discovering fuzzy association rules using fuzzy partition methods", Knowledge-Based Systems, Volume 16, Issue 3, April 2003, Pages 137-147, Elsevier.

[13] Feng-Hsu Wang, Hsiu-Mei Shao, "E ffective personalized recommendation based on time-framed navigation clustering and association mining", Expert Systems with Applications, Volume 27, Issue 3, October 2004, Pages 365-37, Elsevier.

[14] Yuh-Jiuan Tsay, Jiunn-Yann Chiang, "BAR: an efficient method for mining association rules", Knowledge-Based Systems, Volume 18, Issues 2–3, April 2005, Pages 99-10, Elsevier.

[15] Guoqing Chen, Hongyan Liu, Lan Yu, Qiang Wei, Xing Zhang," A new approach to classification based on association rule mining ",Decision Support Systems, Volume 42, Issue 2, November 2006, Pages 674-68, Elsevier.

[16] Frans Coenen, Paul Leng, "T he effect of threshold values on association rule based classification accuracy", Data & Knowledge Engineering, Volume 60, Issue 2, February 2007, Pages 345-360, Elsevier.

[17] He Jiang; Yuanyuan Zhao; Xiangjun Dong, "Mining Positive and Negative Weighted Association Rules from Frequent Itemsets Based on Interest," Computational Intelligence and Design,ISCID '08. International Symposium on , vol.2, no., pp.242,245, 17-18 Oct. 2008.

[18] YuanyuanZhao,He Jiang; RunianGeng; Xiangjun Dong, "Mining Weighted Negative Association Rules Based on Correlation from Infrequent Items," Advanced Computer Control,ICACC '09. International Conference on, vol., no., pp.270,273, 22-24 Jan. 2009.

[19] Tongyan Li, Xingming Li, "Novel alarm correlation analysis system based on association rules mining in telecommunication networks", Information Sciences, Volume 180, Issue 16, 15 August 2010, Pages 2960-2978, Elsevier.

[20] WeiminOuyang; Qinhua Huang, "Mining direct and indirect fuzzy association rules with multiple minimum supports in large transaction databases," Fuzzy Systems and Knowledge Discovery (FSKD), Eighth International Conference on , vol.2, no., pp.947,951, 26-28 July 2011.

[21] WeiminOuyang," Mining Positive and Negative Fuzzy Association Rules with Multiple Minimum Supports", International Conference on Systems and Informatics, 2012.

[22] AnjanaGosain and ManeelaBhugra," A Comprehensive Survey of Association Rules On Quantitative Data In Data Mining", IEEE Conference on Information and Communication Technologies, 2013.

[23] Yiyong Xiao, Yun Tian, Qiuhong Zhao, "Optimizing frequent time-window selection for association rules mining in a temporal database using

a variable neighbourhood search", Computers & Operations Research, Volume 52, Part B, December 2014, Pages 241-250, Elsevier.
[24] Vahid Beiranvand, Mohamad Mobasher-Kashani, Azuraliza Abu Bakar, "Multi-objective PSO algorithm for mining numerical association rules without a priori discretization ", Expert Systems with Applications, Volume 41, Issue 9, July 2014, Pages 4259-4273, Elsevier.