Research Inventy: International Journal of Engineering And Science
Issn (e): 2278-4721, Issn (p):2319-6483, PP: 17-20
www.researchinventy.com

Automated Techniques for Association Rules: A Two Decades Review

Subrata Datta[1*], Atreyee Datta[2], Soumita Sen[3]
[1]*Neotia Institute of Technology, Management & Science, D.H. Road, 24 PGS (S), WB, Email: subrataju2008@gmail.com*
[2]*Modern Institute of Engineering and Technology, Bandel, Dist. Hooghly, WB, Email: atreyeedatta12@gmail.com*
[3]*Techno International Batanagar, Maheshtala , WB, Email: sou.2a8.mita@gmail.com*

Abstract
Association rule mining is a popular data mining technique used for knowledge discovery from the real world datasets. Due to the progress in science and technology, data is generating with more complexity. Finding significant association rules from the complex data using domain dependent traditional techniques often creates different types of hazards to the users. In this connection, researchers have introduced automated techniques for significant association rules. This paper presents a literature review on the automated techniques for significant association rules introduced during the last two decades. This review process focuses on the necessity and role of automated techniques for significant association rules along with their technological backgrounds. Furthermore, we have compared the efficiency of the automated techniques over the traditional techniques. In addition, this review work identifies the limitations of the automated techniques and recommends future research directions to mitigate the limitations.
Keywords: *Data Mining, Knowledge Discovery, Association Rule, Automated Technique*

I. Introduction

Association rule mining [1, 2, 3, 4] is considered as one of the important data mining techniques [5] for knowledge discovery [6]. Technically, association rules are generated from the interesting patterns [7, 8]. However, selection of interesting patterns depends on the specified minimum support (minsup) threshold. The same itemset may be frequent or rare depending on the specified minsup. Henceforth, proper specification of minsup is necessary for mining quality itemsets from the database. Most of the traditional approaches require domain knowledge for specification of the threshold parameters. In reality, databases are available without any domain knowledge. The scenario creates barriers in discovering significant association rules. In this connection, researchers have introduced a number of automated techniques for significant association rules to overcome the barriers discussed above. We have reviewed the automated techniques for significant association rules introduced in the literature during the last two decades since 2001 to 2021. This review work first identifies the automated techniques for significant association rules and thereafter, categorizes them. We have also discussed the roles, technological backgrounds and limitations of the automated techniques in this paper. This review work also presents a comparative study among the automated techniques. In addition, efficiency of the recent automated techniques are compared with the traditional techniques based on real-life dataset.

The rest of the paper is organized as follows. Section II presents the background for quality association rules. Section III presents the scope of automated techniques. Section IV describes the major automated techniques. Section V presents the taxonomy of the automated techniques. Section VI presents a comparative study among the automated techniques. Finally, section VII presents conclusion and future research direction.

II. Background

Association rule mining (ARM) consists of two consecutive sub-processes such as interesting pattern mining [7, 8] and rule generation [1]. Support-confidence [1] is a popular framework used for ARM. Let, TDB is a transactional database consists of a finite set of items called I. In addition, it also contains a finite set of transactions called T where each transaction is a proper subset of I. A k-itemset is a union of k number of distinct items. Support is an objective interestingness measure that states the probability of occurrence of the itemsets. For example, if X is an itemset, then support (sup) of X is $P(X)$ where P stands for the probability function. For generating interesting patterns (itemsets), a support threshold (minsup) is fixed up. The itemsets with support greater than or equal to the minsup is called frequent itemsets. From the set of frequent itemsets, association rules are generated. Another objective interestingness measure called confidence is used for the

assessment of the interestingness of the rules. Confidence (conf) expresses the conditional probability between the two itemsets such as $P(Y|X)$. An association rule is a mathematical implication and is expressed in the form of 'if-then' relationship. For example, if X and Y are two frequent itemsets, then $X \rightarrow Y$ is an association rule provided that $X \cap Y = \Phi$, $sup(X, Y) \geq minsup$ and $conf(X \rightarrow Y) \geq minconf$.

III. Scope of Automated Techniques

This section lists the scope of automated techniques.

(i) Automated specification of support threshold
(ii) Automated specification of MIS threshold for rare itemset mining
(iii) Automated specification of weight for weighted association rule mining
(iv) Automated rule discovery

Major Automated Technique

The list of existing approaches with the automated techniques is shown in Table 1.

Taxonomy of the Automated Techniques

Figure 1 presents the detailed taxonomical structure of the automated techniques.

Comparative Study among the Automated Approaches

Table 2 and Table 3 show the comparative results of the automated approaches with respect to weight automation and threshold automation respectively.

IV. Conclusion and Future Research Direction

This research provides a review study of automated techniques for significant association rules. A taxonomy of the automated techniques is presented in this paper. We have reviewed the mining techniques and policies of the automated approaches and thereafter, present a comparative study among the automated techniques. In addition, experimental studies are also presented to show the efficiency of the automated approaches. During the last two decades, a number of automated techniques have been introduced in the literature. These techniques are suitable for specific purposes. Unification of the automated techniques is the future research direction.

Year	Approach	Technique is based on	Role
2006	CLS_Apriori [9]	CLS	Automated specification ofsupport threshold
2006	Alatas et al. [10]	GA	Automated discovery ofpositive and negative association rules
2008	Sun et al. [11]	Hub and Authority	Automated weight assignmentto the items
2009	Automated_Apriori [12]	CLS and Cumulativesupport	Automated specification ofsupport threshold
2010	Koh et al. [13]	Item valency and PCA	Automated weight specification to the items andrule pruning
2011	Automated_Apriori_Rare[14]	Average support and Mediansupport	Level wise automated specification of supportthreshold
2012	WeightTransmitter [15]	Propagation of landmarkweights	Automated weight assignmentto the items
2013	Pears et al. [16]	Graph based connectivity	Automated weight specification to the items andrule pruning
2015	XWeightTransmitter [17]	Propagation of landmarkweights	Automated weight assignmentto the items
2016	EARAC [18]	L-Matrix and Cogency	Automated specification ofthresholds.
2017	Can et al. [19]	GSA	Automated discovery ofassociation rules
2020	AutoMMS-FPM [20]	Item influence and statisticalmeasures	Automated specification ofMIS to the items
2021	LWARM [21]	Inter-item link	Automated weight assignmentto the items

Table 1: Approaches with Automated Techniques

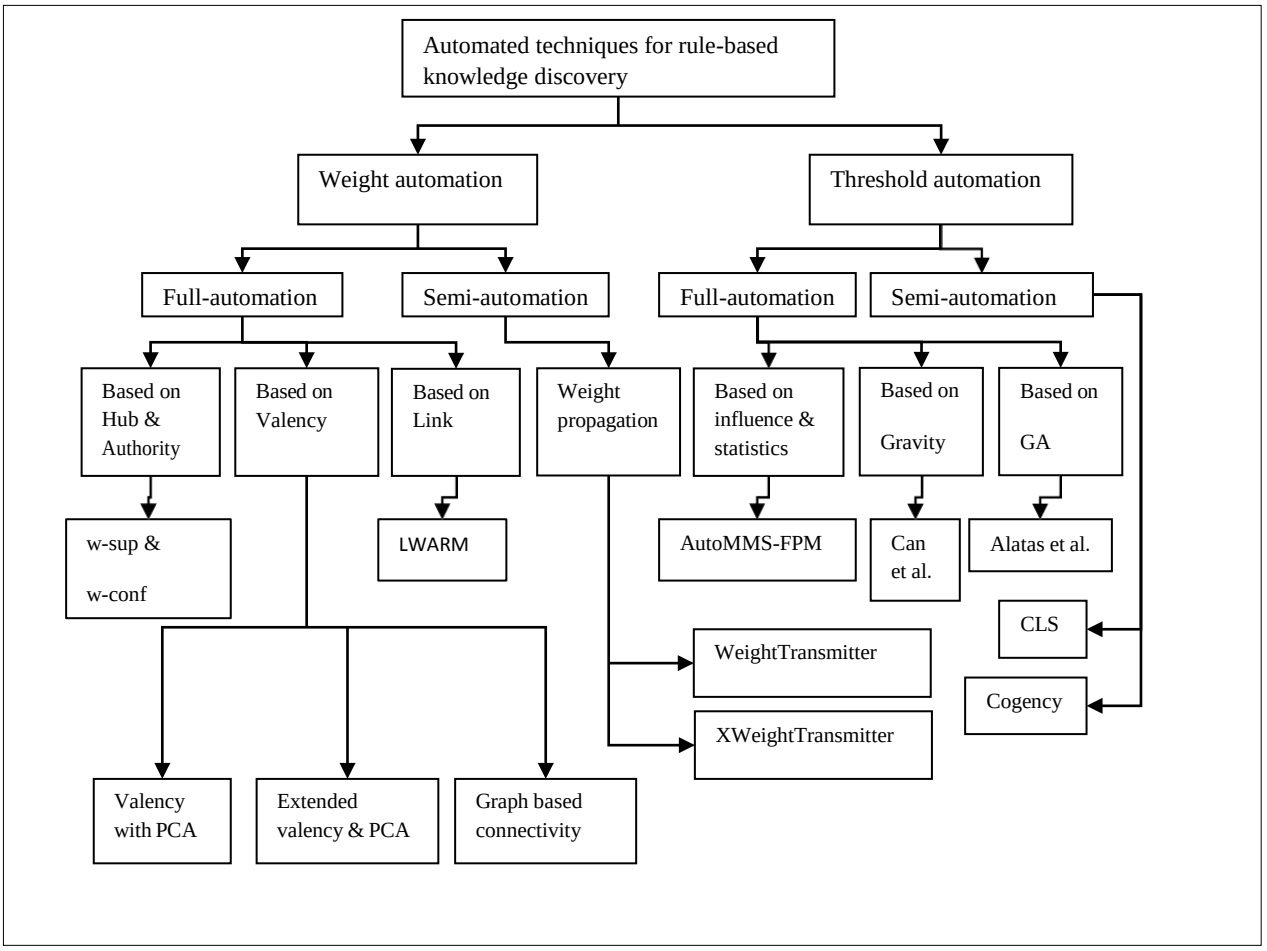

Fig. 1. Taxonomy of the automated techniques for association rules

Approach	**Fully automated**	**Mining without any threshold**	**Accounts indirect connectivity**	**Weight inference**
LWARM [21]	√	×	√	×
Sun et al. [11]	√	×	×	×
Koh et al. [13]	√	√	×	×
Pears et al. [16]	√	√	√	×
WeightTransmitter [15]	×	×	×	√
XWeightTransmitter [17]	×	×	×	√

Table 2: Comparative study among the automated approaches with respect to weightautomation

Approach	Fully automated	minsup/ minconf automation	MIS automation	Avoids thresholds
CLS_Apriori [9]	×	√	×	×
Alatas et al. [10]	√	×	×	√
Automated_Apriori [12]	×	√	×	×
Automated_Apriori_Rare [14]	×	√	×	×
EARAC [18]	×	√	×	×
GSA [19]	√	×	×	√
AutoMMS-FPM [20]	√	×	√	×

Table 3: Comparative study among the automated approaches under threshold automation

References

[1] Mining association rules between sets of items in large databases, R. Agrawal, T. Imielinski, and A. Swami, in: ACM SIGMOD'93, Washington DC, USA, pp. 207-216, 1993.

[2] An efficient algorithm for unique class association rule mining, M. Nasr, M. Hamdy, D. Hegazy, and K. Bahnasy, Expert Systems with Applications, vol. 164, art. 113978, 2021.

[4] Mining and ranking association rules in support, confidence, correlation and dissociation framework, S. Datta, and S. Bose, in: S. Das et al. (eds.) proc. of 4th FICTA, Springer, AISC, vol. 404, pp. 141-152, 2015.

[5] A survey on data mining techniques in recommender systems, M.K. Najafabadi, A.H. Mohamed, and M.N. Mahrin, Soft Computing, vol.23, pp.627-654, 2019.

[6] Knowledge discovery from remote sensing images: a review, L. Wang, J. Yan, L. Mu, and L. Huang, WIREs Data Mining and Knowledge Discovery, vol. 10, issue 5, pp. e1371, 2020.

[7] Frequent itemset mining: a 25 years review, J.M. Luna, P. Fournier-Viger, and S. Ventura, WIREs Data Mining and Knowledge Discovery, vol. 9, issue 6, pp. e1329, 2019.

[8] Interesting pattern mining using item influence, S. Datta, K. Mali, S. Ghosh, R. Singh, and S. Das, in: S.C. Satapathy et al (eds.) Advances in Decision Sciences, Image Processing, Security and Computer Vision, LAIS, Springer, vol. 3, pp. 426-434, 2020.

[9] Automated support specification for efficient mining of interesting association rules, W.Y. Lin, M.C. Tseng, Journal of Information Science, vol. 32, no. 3, pp. 238-250, 2006.

[10] An efficient genetic algorithm for automated mining of both positive and negative quantitative association rules, B. Alatas, and E. Akin, Soft Computimg, vol. 10, pp. 230-237, 2006.

[11] Mining weighted association rules without preassigned weights, K. Sun, F. Bai, IEEE TKDE, vol. 20, no. 4, pp. 489-495, 2008.

[12] An automated association rule mining technique with cumulative support thresholds, C.S.K. Selvi, A. Tamilarasi, Int. J. Open Problems in Compt. Math., vol. 2, no. 3, pp. 427- 438, 2009.

[13] Valency based weighted association rule mining, Y.S. Koh, R. Pears, W. Yeap, in: M.J. Zaki et al. (eds.) proc. of PAKDD'10, Part I, LNAI 6118, Springer-Verlag, Berlin, Heidelberg, pp. 274-285, 2010.

[14] Mining rare itemset with automated support thresholds, C.S.K. Selvi, A. Tamilarasi, Journal of Computer Science, vol. 7, no. 3, pp. 394-399, 2011.

[15] WeightTransmitter: Weighted association rule mining using landmark weights, Y.S. Koh, R. Pears, G. Dobbie, in: P. Tan et al. (eds.): PAKDD'12, Part II, LNAI 7302, pp. 37-48, 2012.

[16] Weighted association rule mining via a graph based connectivity model, R. Pears, Y.S. Koh, G. Dobbie, W. Yeap, Information Sciences, vol. 218, pp. 61-84, 2013.

[17] A graph based approach to inferring item weights for pattern mining, R. Pears, S. Pisalpanus, Y.S. Koh, Expert Systems with Applications, vol. 42, pp. 451-461, 2015.

[18] Explorations of association rules by applying automated cogency, B. Kiruthika, and M.S. Hema, IJSER, vol. 7, no. 3, pp. 30-34, 2016.

[19] Automatic mining of quantitative association rules with gravitational search algorithm, U. Can, and B. Alatas, International Journal of Software Engineering and Knowledge Engineering, vol. 27, no.3, pp. 343-372, 2017.

[20] Mining frequent patterns partially devoid of dissociation with automated MMS specification strategy, S. Datta, K. Mali, S. Ghosh, IETE Journal of Research, 2020, https://doi.org/10.1080/ 02672063.2020.1127243

[21] Weighted association rule mining over unweighted databases using inter-item link based automated weighting scheme, S. Datta, K. Mali, and S. Ghosh, Arabian Journal for Science and Engineering, vol. 46, pp.3169-3188, 2021.