

ASSOCIATION RULE MINING ALGORITHM: A REVIEW

Shikha Dubey (MCA)
Assistant Prof. D. Y. P. I. M. R
Pimpri, Pune-18

shikha.dubey26@gmail.com

Dr. Shivaji D. Mundhe
Director SIMCA
Pune-41

drshivaji.mundhe@gmail.com

ABSTRACT: In this paper, we provide the preliminaries of basic concepts about association rule mining and review 15 scholarly articles related to association rule mining techniques. Of course, a single article cannot be a complete review of all the literature, yet we hope that the references cited will cover the major theoretical issues. This exhaustive literature survey is a preliminary to the proposed research which is to be further carried on. Data mining is used for mining data from databases and finding out meaningful patterns from the database. Many organizations are now using these data mining techniques. This review of literature focuses on how data mining techniques are used for different application areas for finding out meaningful pattern from the database. This paper also includes the steps involved in experimental design. In the conclusion is mentioned the areas which can increase the Efficiency of Association Rules Algorithms.

KEYWORDS: Association Rules, Rule Induction Technique, Apriori Algorithm, Neural Network, Decision Tree.

INTRODUCTION:

Business Intelligence tools enable the organization to make real time decisions at all levels; i.e., strategic, tactical and operational, using advanced analytics and powerful data mining tools. Recent technical achievements in data storage and

database management systems have opened windows of opportunities for developing novel optimizing strategies for all kinds of organizations. The proposed research will introduces a new methodology to enhance organizational performance based on data mining approach. Data mining is a business process for exploring large amounts of data to discover meaningful patterns and rules. The proposed methodology focuses on e business activities and using a multilayer data mining approach, provides high level of business intelligence for enterprises that utilize ecommerce as their main transactional model. To improve the performance of business there is a need to design and implement an effective algorithm based on association rule and therefore the researcher has attempted to work on it.

It is presumed that the result of the proposed research work will help the industry for finding the accurate and useful mining information related to their data and also help in research of similar aspects and for anyone who likely to develop the interest in similar type of work. The proposed work will also help to reduce the present problem related to mining such as repetitive I/O disk scan, huge computation involved during the candidate generation and high memory dependency. So this work is proposed to reduce costly, repeated database scans as well as CPU time without compromising on accuracy.

Steps involved in conducting an experimental study

Identify and define the problem.

Formulate hypotheses and deduce their consequences

Construct an experimental design that represents all the elements, conditions, and relations of the consequences

1. Select sample of subjects.
2. Group or pair subjects.
3. Identify and control non experimental factors.
4. Select or construct, and validate instruments to measure outcomes.
5. Conduct pilot study.
6. Determine place, time, and duration of the experiment.

LITERATURE REVIEW

Literature review of 15 scholarly articles is presented below in tabular form. The various columns of the table reflect the purpose, the themes arising out of the review and the gaps identified.

RESEARCH OBJECTIVES:

- i) To understand and explore the methods employed in various algorithms of data mining.
- ii) To design efficient algorithm for mining Multiple-level association rule from large databases.
- iii) To improve the efficiency of algorithm by reducing the repeated database scan operations.
- iv) To compare the performance of proposed algorithm (AM e-biz) with the existing algorithms.
- v) To implement the algorithm and test it for real time datasets to reduce costly, repeated database scans

Sr No	Source & Year	Title	Author	Contents/Purpose	Aspect/Themes Emerging	Relevance/ Gap/Further Research
1	<i>IBM Almaden Research Center 650 Harry Road, San Jose, CA 95120</i> 2012	Mining Association Rules between Sets of Items in Large Databases	Rakesh Agrawal Tomasz Imielinski Arun Swami	We are given a large database of customer transactions. The algorithm incorporates buffer management and novel estimation and pruning techniques. They also present results of applying this algorithm to sales data obtained from a large retailing company, which shows the effectiveness of the algorithm	In this paper they introduced the problem of mining association rules between sets of items in a large database of customer transactions.	According to them database mining is an important new application area for databases, combining commercial interest with intriguing research questions.
2	<i>International Journal of Innovation, Management and Technology, Vol. 3, No. 5, October 2012</i> 2012	Enhancing Organizational Performance through a New Proactive Multilayer Data Mining Methodology: An Ecommerce Case Study	Masoud Pesaran 1 Behbahani , 2 Islam Choudhury , 3 Souheil Khaddaj	This paper introduces a new methodology to proactively enhance organizational performance based on data mining approach. .The proposed methodology focuses on e-business activities and using a multilayer data mining approach, provides high level of business intelligence for enterprises that utilize ecommerce as their main transactional model. These mining structures will	It appears from the preceding discussions and experimentations that the proposed multilayer data mining approach to an e business framework may increase overall amount of business intelligence that an enterprise can gain.	The next step of this research is planned to integrate the methodology into Multidimensional data and cube structures and also deal with fast-generated data streams to support real time decision making.
				act as the platform for applying the multilayer data mining models.	The concept contains a new methodology and its associated mining structures and mining models. The paper used this novel	

					methodology and introduced an optimized framework called EBAF, to provide intelligence for SMEs and help Them to gain competitive advantages.	
3	IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 1, January 2011 ISSN (Online): 1694-0814 2011	An Efficient Approach to Prune Mined Association Rules in Large Databases	¹ D.Narmadha , ² G.NaveenSundar , ³ S.Geetha	Association rule mining finds interesting associations and/or correlation relationships among large set of data items. However, when the number of association rules become large, it becomes less interesting to the user. Thus, this paper presents a novel approach to prune mined association rules in large databases. Further, an analysis of different association rule mining techniques for market basket analysis, highlighting strengths of different association rule mining techniques are also discussed. They wanted to point out potential pitfalls as well as challenging issues need to be addressed by an association rule mining technique.	This paper discusses the problem of selecting interesting association rules through huge volumes of discovered rules. This paper discusses a novel efficient approach to prune mined association rules in large databases. A fairly comparative analysis of different association rule mining techniques for market basket analysis, highlighting strengths of different approaches, potential pitfalls as well as challenging issues need to be addressed by an association rule	It is crucial to help the decision-maker with an efficient post processing step in order to select interesting association rules throughout huge volumes of discovered rules. This motivates the need for association analysis. Their evaluation shows that an efficient approach to prune mined association rules approach should be efficient and produce user interesting rules

					mining technique are also discussed. We believe that the results of this evaluation will	
					help decision maker for making important decisions. They have evaluated the algorithms based on parameters like scalability, quality of filtered rules..	
4	An Improved Algorithm for Mining Association Rules in Large Databases 2011	<i>World of Computer Science and Information Technology Journal (WCSIT)</i> ISSN: 2221-0741 Vol. 1, No. 7, 311-316, 2011	Farah Hanna AL-1 Zawaidah , 2 Yosef Hasan Jbara	Mining association rules in large databases is a core topic of data mining. The proposed approach is derived from the conventional Apriori approach with features added to improve data mining performance. They have performed extensive experiments and compared the performance of their algorithm with existing algorithms found in the literature.	The aim of this paper is to improve the performance of the conventional Apriori algorithm that mines association rules by presenting fast and scalable algorithm for discovering association rules in large databases. The approach to attain the desired improvement is to create a more efficient new algorithm out of the conventional one by adding new features to the Apriori approach. The proposed mining algorithm can efficiently discover the association rules	One of the challenges in developing association rules mining algorithms is the extremely large number of rules generated which makes the algorithms inefficient and makes it difficult for the end users to comprehend the generated rules. Future work includes: 1) extensive empirical evaluation; 2) to real data like retail sales transaction and medical transactions to confirm the experimental

					between the data items in large databases.	results in the real life domain;
5	Intech ISBN: 978-953-307-067-4, 2010	Discovering Web Server Logs Patterns Using Generalized Association Rules Algorithm	Mohd Helmy Abd Wahab ¹ , Mohd Norzali Haji Mohd ² and Mohamad Farhan Mohamad Mohsin ³	Web Usage Mining is an aspect of data mining that has received a lot of attention in recent year .In this paper they have focused on this aspect of data mining.	Commercial companies as well as academic researchers have developed an extensive array of tools that perform several data mining algorithms on log Files coming from web servers in order to identify user behavior on a particular web site.	Thus, there is a need for robust methods that integrates different intelligent Techniques that is free of any assumptions about the noise contamination rate
					Performing this kind of investigation on the web site can provide information that can be used to better accommodate the user's needs	
6	Rakesh Agrawal ¹ Ramakrishnan Srikant ² 2009	Fast Algorithms for Mining Association Rules	IBM Almaden Research Center 650 Harry Road, San Jose, CA 95120	They considered the problem of discovering association rules between items in a large database of sales transactions. They presented two new algorithms for solving this problem that are fundamentally different from the known algorithms.	The algorithms presented in this paper have been implemented on several data repositories, including the AIX _le system, DB2/MVS, and DB2/6000. They have also tested these algorithms against real customer data,	They did not consider the quantities of the items bought in a transaction, which are useful for some Applications. Finding such rules needs further Work.
7	ARM: A HYBRID	Zahir Tari ¹ Wensheng Wu	Springer journal	Most of the approaches for association rule mining focus	Based upon their study of	Future work mayl consist of

	ASSOCIATION RULE MINING ALGORITHM	2,		<p>on the performance of the discovery of the frequent item sets. They are based on the algorithms that require the transformation of data from one representation to Another, and therefore excessively use resources and incur heavy CPU overhead. They Proposes a hybrid algorithm that is resource efficient and provides better performance. . In addition, they propose a comparison algorithm (CmpApr) that compares candidate item sets with a transaction, a filtering algorithm (FilterApr) that reduces the number of comparison operations required to find frequent item sets. ARMhas better performance and scales linearly.</p>	<p>association rule mining, they have proposed a sequential algorithm, ARM, heterogeneous which achieves better performance with the available resources and displays near-linear scale-up behavior. . In the early iterations, it requires fewer resources than <i>Partition</i>, and in the late iterations, it performs faster than <i>FilterApr</i>.</p>	<p>extending the proposed algorithms in a context of environments. Because usually there are multiple jobs running at the same time, the local resources, i.e., CPU, memory, disk, and communication resources, change over time. The challenge is to maximize performance.</p>
8	Algorithms for Association Rule Mining – A General Survey and Comparison	Jochen Hipp ¹ , Ulrich Guntzer ² , Gholamreza Nakhaeizadeh	SIGKDD Explorations Volume 2 issue 1	Today there are several efficient algorithms that cope with the popular and computationally expensive task of association rule mining. In this paper they have explained the	In this paper the authors have dealt with the algorithm aspects of association rule mining. The advantages and	
		3,		<p>fundamentals of association rule mining and a general framework is given.</p>	<p>disadvantages of all of them are listed based on the experiment carried out. There is not a single algorithm which is beating others.</p>	
9	Association Rule Mining: A Survey	Qiankun Zhao ¹ , Sourav S ² , Bhowmick ³	Technical Report, CAIS, Nanyang Technological University, Singapore, No. 2003116 ,	In this paper, we surveyed the list of existing association rule mining techniques	This investigation is prepared to our new project titled <i>mining historical changes to web delta</i> .	Thru this we come to know of existing algorithms of mining.
10	Integrating	Bing Liu ¹ ,	Appeared in <i>KDD-</i>	Classification rule mining	This paper	The research

	<p>Classification and Association Rule Mining</p> <p>1998</p>	<p>Wynne Hsu², Yiming Ma</p>	<p>98, New York, Aug 27-31, 1998</p>	<p>aims to discover a small set of rules in the database that forms an accurate classifier. Association rule mining finds all the rules existing in the database that satisfy some minimum support and minimum confidence constraints. For association rule mining, the target of discovery is not pre-determined, while for classification rule mining there is one and only one predetermined target. In this paper, we propose to integrate these two mining techniques. The integration is done by focusing on mining a special subset of association rules, called <i>class association rules</i> (CARs).</p>	<p>proposes a framework to integrate classification and association rule mining. An algorithm is presented to generate all class association rules (CARs) and to build an accurate classifier. The new framework not only gives a new way to construct classifiers, but also helps to solve a number of problems that exist in current classification systems.</p>	<p>can be extended to building Of more accurate classifiers by using more sophisticated Techniques and to mine CARs without pre-discretization.</p>
<p>11</p>	<p>A Comparative Study of Association Rules Mining Algorithms</p> <p>2000</p>	<p>Cornelia¹ Robert Györödi² . Stefan Holban³</p>	<p>Appeared in IEEE journal</p>	<p>This paper presents a comparison between classical frequent pattern mining algorithms that use candidate set generation and test and the algorithms without candidate set generation. The compared algorithms are presented together with some experimental data that lead to the final conclusions.</p>	<p>From the experimental data presented it was concluded that the Dyn FP-growth algorithm behaves better than the FP-growth algorithm. First of all, the FP-growth algorithm needs at most two scans of the database, while the number of database scans for the candidate generation algorithm</p>	
					<p>(Apriori) increases with the dimension</p>	

					of the candidate item sets.	
12	Association Rules Mining Algorithm 1999	Zhihua Xiao ¹		Many algorithms to mine the association rules are divided into two stages, the first is to find the frequent set; the second is use the frequent set to generate association rules. This proposal discuss the respective characteristics and .shortcoming of the current algorithms to mine association rules and propose another method to mine faster; unlike the other algorithms,	This algorithm emphasis on the usage of domain knowledge and the preparation of the systematically Using of the association rules.	
13	ASSOCIATION RULE MINING IN ECOMMERCE: A SURVEY 2000	Venkateswari S ¹	Venkateswari S et al. / International Journal of Engineering Science and Technology (IJEST)	Association Rule mining is one of the most popular data mining techniques which can be defined as Extracting the interesting correlation and relation among large volume of transactions. E-commerce applications generate huge amount of operational and behavioral data.	In this paper a survey of association rule mining and frequent item set mining algorithms are presented. How association rule mining can be applied to e-commerce to improve the services of e-commerce enterprises are Discussed in detail. The types of e-commerce data and the data collection methods are also discussed.	Applying association rule mining in e-commerce application can unearth the hidden knowledge from these data.
14	Mining Multiple-Level Association Rules in Large Databases	Jiawei Han ¹ , Yongjian Fu ² ,	IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 11, NO. 5, SEPTEMBER/OCTOBER 1999	A group of variant algorithms is proposed based on the ways of sharing intermediate results, with the relative performance tested and analyzed.	They have extended the scope of the study of mining association rules from single level to multiple concept levels and studied	They also suggested that efficient algorithms can be developed from large databases for the discovery of interesting

	1999				methods for mining multiple-level association rules from large transaction databases.	and strong multiple-level association rules. mining multiple-level correlations in databases is another
						interesting issue to be Studied in the future.
15	<p>Interestingness of association rules in data mining: Issues relevant to e-commerce</p> <p>2005</p>	<p>RAJESH NATARAJA^{N1}</p> <p>B SHEKAR²</p>	<p><i>S-adhan-a</i> Vol. 30, Parts 2 & 3, April/June 2005, pp. 291–309. © Printed in India</p>	<p>They first examine different approaches to address this problem citing their applicability to e-commerce whenever appropriate. They then provide a detailed survey of one important approach, namely interestingness measure, and discuss its relevance in e-commerce applications Such as personalization in recommender systems. We conclude by reiterating the importance of post-processing methods in data mining for effective and efficient deployment of e-commerce solutions</p>	<p>In this paper, they have examined the understandability problem in data mining, and its relevance to e-commerce applications. The genesis of this problem can be traced to the glut of patterns generated by data mining applications. Large numbers of mined patterns render manual inspection impractical and infeasible. In addition, the commonplace and obvious nature of knowledge revealed by most of these patterns necessitates usage of automated methods for selecting the most Interesting, novel, relevant and significant</p>	<p>Interestingness is a perceptual concept whose different facets are difficult to capture and operationalise. Some of these facets are yet to be identified. The ensuing discussions identified some research issues relevant to e-commerce applications. Future research pertaining to interestingness and associated methods is expected to Yield results with more complete interestingness characterizations. With large volumes of data, data mining is expected to play a significant role</p>

					patterns for further action. They have surveyed some of the available post-processing methods in data mining.	in increasing the effectiveness of ecommerce applications.. It is not uncommon to find organizations struggling to make sense of data captured through automated processes. Frameworks and methodologies that aid not only in the selection of
--	--	--	--	--	---	--

relevant and

						significant patterns but also in their effective and efficient deployment need to be researched as they may help firms in leveraging their information advantage
--	--	--	--	--	--	--

Conclusion:

Increasing the Efficiency of Association Rules Algorithms

The computational cost of association rules mining can be reduced in four ways:

- By reducing the number of passes over the database
- By sampling the database
- By adding extra constraints on the structure of patterns
- Through parallelization.

In recent years much progress has been made in all these directions

REFERENCES:

- 1) Rakesh ¹ Agrawal, Tomasz Imielinski ², Arun Swami ³: Mining Association Rules between Sets of Items in Large Databases *IBM Almaden Research Center* 650 Harry Road, San Jose, CA 95120 2012 .
- 2) Masoud Pesaran Behbahani ¹, Islam Choudhury ², Souheil Khaddaj ³: Enhancing Organizational Performance through a New Proactive Multilayer Data Mining Methodology: An Ecommerce Case Study *International Journal of Innovation, Management and Technology*, Vol. 3, No. 5, October 2012
- 3) D.Narmadha ¹, G.NaveenSundar ², S.Geetha ³: An Efficient Approach to Prune Mined Association Rules in Large Databases: *IJCSI International Journal of Computer Science Issues*, Vol. 8, Issue 1, January 2011 ISSN (Online): 1694-0814, 2011
- 4) Farah Hanna AL-Zawaidah ¹, Yosef Hasan Jbara ²: An Improved Algorithm for Mining Association Rules in Large Databases *World of Computer Science and Information Technology Journal (WCSIT)* ISSN: 2221-0741 Vol. 1, No. 7, 311-316, 2011.
- 5) Mohd Helmy Abd Wahab¹, Mohd Norzali Haji Mohd², Mohamad Farhan Mohamad Mohsin³: Discovering Web Server Logs Patterns Using Generalized Association Rules Algorithm, Intech ISBN: 978-953-307-067-4, 2010
- 6) Rakesh Agrawal ¹, Ramakrishnan Srikant²: Fast Algorithms for Mining Association Rules, IBM Almaden Research Center 650 Harry Road, San Jose, CA 95120, 2009
- 7) Zahir Tari ¹, Wensheng Wu ²: ARM: a hybrid association rule mining algorithm Springer journal, 2006
- 8) Jochen Hipp¹, Ulrich Guntzer ², Gholamreza Nakhaeizadeh ³: Algorithms for Association Rule Mining – A General Survey and Comparison *SIGKDD Explorations* Volume 2 issue 1, 2000 .
- 9) Qiankun Zhao¹, Sourav S ², Bhowmick³: Association Rule Mining: A Survey , Technical Report, CAIS, Nanyang Technological University, Singapore, No. 2003116 ,2003
- 10) Bing Liu¹, Wynne Hsu ², Yiming Ma Integrating Classification and Association Rule Mining, Appeared in *KDD-98*, New York, Aug 27-31, 1998
- 11) Cornelia ¹, Robert Györödi², Stefan Holban³: A Comparative Study of Association Rules Mining Algorithms Appeared in *IEEE journal* 2000.
- 12) Zhihua Xiao¹: Association Rules Mining Algorithm Intech ISBN: 978-953-307-067-5, 2010
- 13) Venkateswari S ¹: association rule mining in ecommerce :a survey: Venkateswari S et al. / *International Journal of Engineering Science and Technology (IJEST)* 2000.
- 14) Jiawei Han ¹, Yongjian Fu ²: Mining Multiple-Level Association Rules in Large Databases: *IEEE transactions on knowledge and data engineering*, vol. 11, no. 5, September/ October 1999
- 15) Rajesh Natarajan¹, b Shekar²: Interestingness of association rules in data mining: Issues relevant to e-commerce , *S-adhan`a* Vol. 30, Parts 2 & 3, April/June 2005, pp. 291–309. © Printed in India 2005