

RESEARCH ISSUES IN ASSOCIATION RULE MINING

Shikha Dubey

Assistant Professor *D.Y.P.I.M.R, PIMPRI

shikha.dubey26@gmail.com

Dr Shivaji Mundhe

DIRECTOR, SIMCA

drshivaji.mundhe@gmail.com

ABSTRACT: Association rule mining is one of the important areas of research receiving increasing attention of researchers all over the world. It is an essential part of knowledge discovery in databases. Over the last ten years it has been developed at a dynamic rate. Although it has been emerged as a new technology but association rule mining is still in a stage of exploration and development. In this paper we present a survey of research work carried by different researchers. Of course, a single article cannot be a complete review of all the research work, yet we hope that it will provide a guideline for the researcher in interesting research directions that have yet to be explored.

KEYWORDS: Association rule mining, Algorithm development, and e-business.

I. INTRODUCTION

Association rules are interesting correlations among attributes in a database. These rules have many applications in areas ranging from e-commerce to sports to census analysis to medical diagnosis. The discovery of association rules is an extremely computationally expensive task and it is therefore imperative to have fast scalable algorithms for mining these rules. In the detailed literature review it was found out that there is a need for efficient techniques for discovering association rules from large databases and for removing redundancy from these rules so as to improve the quality of output.

II. BRIEF INTRODUCTION TO ASSOCIATION RULES:

In a database of transactions D with a set of n binary attributes (items) I ,

A rule is defined as an implication of the form $X \rightarrow Y$ where X, Y is a subset of I and $X \cap Y = \text{Null}$.

The sets of items (for short *item sets*) X and Y are called *antecedent* (left-hand-side or LHS) and *consequent* (right-hand-side or RHS) of the rule respectively. The *support*, $\text{supp}(X)$, of an item set X is defined as the proportion of transactions in the data set which contain the item set.

Their mining requires satisfying a user-specified *minimum support* and a user-specified *minimum confidence* from a

given database at the same time. To achieve this, association rule generation is a two-step process.

- 1) First, minimum support is applied to find all frequent item-sets in a database.
- 2) In a second step, these frequent item-sets and the minimum confidence constraint are used to form rules.

Literature review has been divided in following areas:

- 1) ASSOCIATION RULE MINING
- 2) ALGORITHM DEVELOPMENT
- 3) E- BUSINESS

Following table shows coverage of following issues in AM

- I] Association Rule Mining
- J] Algorithm Design
- K] Enhance e- Business
- L] Only reviewing prior literature
- M] Experimental research in AM
- N] Research in AM using case study
- O] Comparative study of various algorithms and experiments involving AM
- P] Theoretical research in Association Mining involving concepts

Time Line	20 10	20 10	20 11	20 11	20 11	20 11	20 12	20 12	20 12	20 13	20 13	20 13	20 13	20 14	20 14	20 14	20 14	20 14
Paper No.	AM 1	AM 2	AM 3	AM 4	AM 5	AM 6	AM 7	AM 8	AM 9	AM 10	AM 11	AM 12	AM 13	AM 14	AM 15	AM 16	AM 16	AM 17
I			•	•			•	•			•							•
J		•	•				•	•		•	•	•						•
K						•	•		•		•			•		•		
L		•	•	•				•	•	•	•	•	•	•	•		•	•
M						•	•				•					•		
N	•	•	•	•	•			•	•	•					•		•	•
O					•													
P						•	•		•	•	•				•	•	•	

An exhaustive Literature review of various research papers in various international and national journals , scholarly articles, books, doctoral thesis ,project works

was conducted .A brief elaboration of the most relevant ones are shown below.

Serial Number	1
Title	MINING ASSOCIATION RULES BETWEEN SETS OF ITEMS IN LARGE DATABASES
Author	Arun Swamy and Rakesh Agarwal
Published in	IBM Research Lab
Relevance	This algorithm offers novel buffer management and pruning technique.

Review and excerpt:

The work reported in this paper has been done in the context of the Quest project at the IBM AI-maden Research Center. It uses pruning techniques to avoid measuring certain item sets, while guaranteeing completeness. These are the item sets that the algorithm can prove will not turn out to be large. There are two such pruning techniques. The first one, called the remaining tuple optimization", uses the current scan position and some counts to prune item sets as soon as they are generated. This technique also establishes, while a pass is in progress, that some of

the item sets being measured will eventually turn out to be large and prunes them out. The other technique, called the pruning function optimization", synthesizes pruning functions in a pass to use them in the next pass they tested the effectiveness of their algorithm by applying it to sales data obtained from a large retailing company. For this data set, the algorithm exhibited excellent performance. The estimation procedure exhibited high accuracy and the pruning techniques were able to prune out a very large fraction of item sets without measuring them.

Serial Number	2
Title	PRESERVING PRIVACY IN ASSOCIATION RULE MINING
Author	Ahmed HajYasien
Published in	Thesis in Griffith University
Relevance	The researcher analyzed the efficiency of the proposed algorithm theoretically and confirmed the analysis by testing the algorithm on some real world databases.

Review and excerpt:

The researcher suggested future research directions as Association rule mining is of relevance to e-commerce applications. In this thesis, he focused on

the accuracy of the data and blocking the inference channels. However, in practice, some commercial

criteria may also be important and should be considered in the technique implementation.

Serial Number	3
Title	REVIEW OF ASSOCIATION RULE MINING ALGORITHM IN DATA MINING
Author	Xu Chi ; Zhang Wen Fang
Published in	Communication Software and Networks (ICCSN), 2013 IEEE 3rd International Conference on 27 aug 2013
Relevance	This paper presented related formal definitions of association rules and the basic algorithm for association rules mining in data streams

Review and excerpt:

With the database technology, artificial intelligence and mathematical statistics the development of technology, database data mining technology arises at

the historic moment. Based on systematic investigation of association rules mining researches on streams data, analyzed issues and how they were resolved in current literatures.

Serial Number	4
Title	ACCELERATING BUSINESS ANALYTICS APPLICATIONS
Author	Salapura, V, Karkhanis, T
Published in	High Performance Computer Architecture (HPCA), 2012 IEEE 18th International Symposium

Review and excerpt:

Business text analytics applications have seen rapid growth, driven by the mining of data for various decision making processes. Regular expression processing is an important component of these applications, consuming as much as 50% of their total execution time. While prior work on accelerating regular expression processing has focused on Network

Intrusion Detection Systems, business analytics applications impose different requirements on regular expression processing efficiency. They presented an analytical model of accelerators for regular expression processing, which includes memory bus-, I/O bus-, and network-attached accelerators with a focus on business analytics applications

Serial Number	5
Title	APPLYING DATA MINING TO AUTOMATICALLY ACQUIRE USER PREFERENCES IN COMMERCIAL WEB SITES
Author	Shilin Zhang, Mei Gu
Published in	Information Engineering and Computer Science, 2013. IEEE conference
Relevance	An integrated data mining system for NLP in Chinese Commercial Web Sites is presented in this paper.

Review and excerpt:

Data mining are an important field of research. However, there is an important challenge to apply the Data mining technique to NLP applications. It firstly extracted the raw data using NLP technology and then presented the data mining process in detail to acquire

the user preferences. More importantly, it puts forward a new enhanced and integrated method to Acquire User Preferences. Data mining applied to NLP provides a scientific basis for E-Commerce and Decision-making systems

Serial Number	6
Title	MINING & OPTIMIZATION OF ASSOCIATION RULES USING EFFECTIVE ALGORITHM
Author	Sanat Jain, Swati Kabra
Published in	International Journal of Emerging Technology and Advanced Engineering
Relevance	This paper is concerned with discovering positive and negative association rules

Review and excerpt:

They presented an Apriori-based algorithm that is able to find all valid positive and negative association rules in a support confidence framework. The algorithm can find all valid association rules quickly and overcome

some limitations of the previous mining methods. The complexity and large size of rules generated after mining have motivated researchers and practitioners to optimize the rule, for analysis purpose. This optimization can be done using Genetic Algorithm.

Serial Number	7
Title	RESEARCH ON ASSOCIATION RULE MINING
Author	Ziauddin, Shahid Kammal, Khaiuz Zaman Khan, Muhammad Ijaz Khan
Published in	Advances in Computational Mathematics and its Applications (ACMA)
Relevance	They presented a survey of research work carried by different researchers since its beginning.

Review and excerpt:

To make frequent pattern mining an essential task in data mining, much research is needed to further develop pattern-based mining methods. Mechanisms

for deep understanding and interpretation of patterns, e.g., semantic annotation for frequent patterns, and contextual analysis of frequent patterns are needed.

Serial Number	8
Title	AN OPTIMIZED DISTRIBUTED ASSOCIATION RULE MINING ALGORITHM IN PARALLEL AND DISTRIBUTED DATA MINING WITH XML DATA FOR IMPROVED RESPONSE TIME.
Author	Dr (Mrs).Sujni Paul
Published in	International Journal of Computer Science and Information Technology, Volume 2, Number 2, April 2010
Relevance	In this paper an Optimized Distributed Association Rule mining algorithm for geographically distributed data is used in parallel and distributed environment so that it reduces communication costs. The response time is calculated in this environment using XML data.

Review and excerpt:

The Future enhancement of this can be cluster the same XML dataset and find out the knowledge extracted out of that. A visual analysis can also be made for the same.

Serial Number	9
Title	GRAPH BASED APPROACHES USED IN ASSOCIATION RULE MINING
Author	Hemant Kumar Sharma
Published in	Thesis
Relevance	He has studied graph based approaches in Association mining

Review and excerpt:

The researcher has not dealt with the time-space tradeoff problem. As the size of frequent itemset increases, computational time for the initial phases increases exponentially with increase in the requirement

in memory space. So, a better way to consider only the relevant transaction or items can be possible field of research. If data cannot fit in the memory than more page faults may occur resulting in the decrease in the performance of the system.

Serial Number	10
Title	RESEARCH ISSUES IN DATA STREAM ASSOCIATION RULE MINING
Author	Nan Jiang and Le Gruenwald

Published in	SIGMOD Record, Vol. 35, No. 1, Mar. 2006
Relevance	There exist emerging applications of data streams that require association rule mining, such as network traffic monitoring and web click streams analysis

Review and excerpt:

The current stream data mining methods require users to define one or more parameters before their execution; however, most of them do not mention how users can adjust these parameters online while they are running. It is not desirable/ feasible for users to wait

until a mining algorithm to stop before they can reset the parameters. This is because it may take a long time for the algorithm to finish due to the continuous arrival and huge amount of data streams

Serial Number	11
Title	AN EFFICIENT INTERESTING WEIGHTED ASSOCIATION RULE MINING
Author	M. Padmavalli, Prof.K.Sreenivasa Rao
Published in	International Journal of Advanced research in Computer Science and Software Engineering Volume 3, Issue 10, October 2013
Relevance	Weighted Association rule mining has recently been proposed, in which transactions are attached with weighted values according to some criteria

Review and excerpt:

Some interesting patterns can be discovered when the hub weights of transactions are taken into account.

Serial Number	12
Title	CHALLENGES AND RESEARCH ISSUES IN ASSOCIATION RULE MINING
Author	Mahesh Kr. Singh 1, Zaved Akhtar 2, Devesh Kr Sharma 3
Published in	International Journal of Electronics and Computer Science Engineering
Relevance	There exist emerging applications of data streams that require association rule mining, such as network traffic monitoring and web click streams analysis. Different from data in traditional static databases, data streams typically arrive continuously in high speed with huge amount and changing data distribution.

Review and excerpt:

The current data stream mining methods require users to define one or more parameters before their

execution; however, most of them do not mention how users can adjust these parameters online while they are running.

Serial Number	13
Title	THE RESEARCH AND IMPLEMENTATION OF DATA MINING COMPONENT LIBRARY SYSTEM
Author	Peng Peng, Qianli Ma Chaoxiong Li
Published in	Project report of Natural Science Foundation of Guangdong Province, China (07006474) and Science and Technology Planning Project of Guangdong Province, China
Relevance	He dealt with algorithms as components

Review and excerpt:

Compare to the existing data mining software, it can provide data mining service with more humanity

characteristic and higher success rate, which is more suitable for real-time decision analysis application software.

Serial Number	14
Title	Advances in Computer Vision and Information Technology
Author	By K. V. Kale
Published in	I. K. International Pvt Ltd
Relevance	It focused on research issues through automated data mining

Serial Number	15
Title	Smart Big Data, Analytics and Metrics To Make Better Decisions
Author	Bernard B. Marr
Published in	Book
Relevance	How to USE big data to get solid, real-world business results - and putting that in place to improve performance. Big Data gives us a clear understanding, blueprint, and step-by-step approach to build our own big data strategy.

Review and excerpt:

Addresses how the results of big data analytics can be visualized and communicated to ensure key decisions-makers understand them

Serial Number	16
Title	Algorithms and Applications for Spatial Data Mining
Author	Martin Ester, Hans-Peter Kriegel, Jörg Sander
Published in	Geographic Data Mining and Knowledge Discovery, Research Monographs in GIS, Taylor and Francis, 2001.
Relevance	Finding implicit regularities, rules or patterns hidden in spatial databases is an important task, e.g. for geo-marketing, traffic control or environmental studies.

Review and excerpt:

Geographers may be interested in learning spatio-temporal rules describing the process of growth of urban land use.

Serial Number	17
Title	Research on Association Rule Mining
Author	Ziauddin, Shahid Kammal, Khaiuz Zaman Khan, Muhammad Ijaz Khan
Published in	Advances in Computational Mathematics and its Applications (ACMA) 226 Vol. 2, No. 1, 2012,
Relevance	Survey of research work carried by different researchers since its beginning.

Review and excerpt:

Efficient algorithm should be developed for data mining on XML databases. Association rule mining should be applied to social network analysis and mining for business

With the enhanced features in recent computer systems, increasingly larger amounts of data are being accumulated in various fields. The collected data is growing exponentially every year, and it becomes increasingly necessary to use automated tools in order to extract precise and useful information from the collected data. Data mining is a powerful tool that

III. CONCLUSION

enables one to achieve this. Data mining programs have become essential tools in many domains including business (marketing, customer relationship management, scoring and risk management, fraud detection), science (astrophysics, climate modeling, particle physics), biotechnology (understanding diseases, protein identification, drug discovery, personalized medicine), and other fields like internet searches, multimedia, security, etc. 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.

IV. REFERENCES

- 1] Rakesh Agrawal¹, Tomasz Imielinski², Arun Swami³: *Mining Association Rules between Sets of Items in Large Databases IBM Almaden Research Center 650 Harry Road, San Jose, CA 95120 2012*.
- 2] Bay Vo¹, Bac Le²: *fast algorithm for mining generalized association: International Journal of Database Theory and Application through IEEE proceedings. Vol. 3, No. 5, October 2012*
- 3] Bikramjit Saikia¹: *study of association rule mining and different hiding techniques: Thesis Ph.D. NIT, Rourkela*
- 4] Vikram Pudi¹: *Efficient discovery of concise Association rules from large Databases World of Ph.D Thesis IIT, Bangalore.*
- 5] M. Jayakameswari¹: *Design and Development of data mining system to estimate cars promotion using improved id3 algorithm, International Journal of Advanced Research in Computer and Communication Engineering, 2010*
- 6] Pappa, Gisele L: *Automating The Design Of Data Mining Algorithms, 2010, XIII.*
- 7] Ramakrishna Shrikant¹, Rakesh Agarwal²: *Comparison Of Various Association Rule Mining Algorithm On Frequent Itemsets in International Journal of Advanced Research in Computer Science and Software Engineering.*
- 8] Jayna Donga¹, Kanu Patel²: *Algorithms for Association Rule Mining – A General Survey and Comparison SIGKDD Explorations Volume 2 issue 1, 2000.*
- 9] Prof. Lizhe Wang Yan Bai²: *An Overview Of Association Rule Mining Algorithms in IEEE Proceedings of international conference on BIG DATA 2013*
- 10] Xu Chi¹, Zhang Wen Fang²: *Review of Association Rule Mining Algorithm in Data Mining, A Business and Information Management appeared in KDD-98, New York, Aug 27-31, 1998*
- 11] Tian-ru Li¹, Yang Xu²: *Parallel and Distributed Computing, Applications and technologies, 2003. PDCAT'2003. Proceedings of the Fourth International Conference of IEEE.*
- 12] Sotiris Kotsiantis¹, Dimitris Kanellopoulos²: *Association Rules Mining: A Recent Overview Venkateswari S et al. / GESTS International Transactions on Computer Science and Engineering*
- 13] Dr. S.P. Shantharajah¹, S. Suriya², R. Deepalakshmi³: *A Complete Survey On Association Rule Mining With Relevance To Different Domain International Journal Of Advanced Scientific And Technical Research.*
- 14] Thabet Slimani¹, Amor Lazzez²: *Efficient Analysis Of Pattern And Association Rule Mining Approaches: IEEE Proceedings*
- 15] K. Vanitha¹, R. Santhi²: *Evaluating The Performance Of Association Rule Mining Algorithms Journal of Global Research in Computer Science, vol. 11, no. 5, September/ October 2010*
- 16] Ahmed HajYasien¹: *Preserving Privacy In Association Rule Mining Ph D thesis in Griffith University.*
- 17] Xu Chi¹, Zhang Wen Fang²: *Review of association rule mining algorithm in data mining Communication Software and Networks (ICCSN), 2013 IEEE 3rd International Conference on 27 aug 2013*
- 18] Salapura¹, Karkhanis, T²: *Accelerating business analytics applications High Performance Computer Architecture (HPCA), 2012 IEEE 18th International Symposium .*
- 19] Shilin Zhang¹, Mei Gu²: *Applying Data Mining to Automatically Acquire User Preferences in Commercial Web Sites in Information Engineering and Computer Science, 2013. IEEE conference*
- 20] Sanat Jain¹, Swati Kabra²: *Mining & Optimization of Association Rules Using Effective Algorithms in International Journal of Emerging Technology and Advanced Engineering .*
- 21] Ziauddin¹, Shahid Kammal², Khaiuz Zaman Khan, Muhammad Ijaz Khan: *Research on Association Rule Mining ACMA, Vol. 2, No. 1, pp. 226-236, 2012*
- 22] Dr Sujni¹: *An Optimized Distributed Association Rule Mining Algorithm In Parallel And Distributed Data Mining With Xml Data For Improved Response Time International Journal of Computer Science and Information Technology, Volume 2, Number 2, April 2010*
- 23] Hemant Kumar Sharma¹: *Graph Based Approaches Used In Association Rule Mining Ph.D Thesis: 2010*
- 24] Nan Jiang¹, Le Gruenwald²: *Research Issues In Data Stream Association Rule Mining, SIGMOD Record, Vol. 35, No. 1, Mar. 2006.*
- 25] M. Padmavalli¹, Prof. K. Sreenivasa Rao²: *An Efficient Interesting Weighted Association Rule Mining, International Journal of Advanced research in Computer Science and Software Engineering Volume 3, Issue 10, October 2013.*