A SURVEY OF POST MINING OF ASSOCIATION RULES AND
HIGH SPEED ASSOCIATION RULE MINING

D.William Albert¹, Dr.K.Fayaz², D.Veerabhadra Babu³

¹Department of Computer Science & IT, Mahatma Gandhi University, Meghalaya, India
²Department of Computer Science & Technology, SK University, Anantapuram, India
³Department of Computer Science & IT, Mahatma Gandhi University, Meghalaya, India

ABSTRACT:

Association rule mining has widely been used to extract trends or patterns that satisfy given support and confidence. The trends are then transformed into business intelligence for expert decision making. However, association rule mining often results in huge number of rules. Users might feel it difficult to identify those rules that are of interest to them. Therefore post mining of association rules is essential to get rid of insignificant rules besides summarizing and visualizing the discovered rules. Recent innovations with respect to GPU (Graphics Processing Unit) technologies help leverage the high speed processing for association rule mining. This paper makes a systematic analysis of various post mining techniques required for discovering actionable knowledge from large number of association rules. It also throws light into the GPU and its usage in high speed association rule mining. We also evaluate the techniques based on the speed, accuracy, time and efficiency.

Keywords – Data mining, association rule mining, post mining techniques, high speed association rule mining

[1] INTRODUCTION

Association rule mining has been around which one of the data mining techniques is widely used in real world applications. The rules obtained bestow the trends that can reveal business intelligence which can be used to make well informed business decisions. According to Boettcher et al. [1] association rule mining discovers the hidden relationships or correlations among the attributes. In this paper we analyzed methods for association rule mining, methods used for post mining of association rules and high speed association rule mining.

Association rule mining which is meant for discovering trends or patterns in given data set has very important utility in the real world enterprises. It is often used to make expert decisions that lead to profit. Various methods or techniques used in association rule mining are explored in [2], [3], [4], [5], [6], [7], [8], [9], [10], [11] as presented in this paper. This paper also focuses on various methods or techniques used for post mining of association rules as explored in [12], [13], [14], [15], [16], [17], [18], [19] and [20]. Finally it throws light into the technological innovations that yield Graphics Processing Unit (GPU) and the usage of it for general purpose computing applications in order to achieve high speed association rule mining. The high speed association rule mining [21], [22], [23], [24] and [25] the architecture of GPU and the results of research pertaining to the high speed association rule mining are analyzed in this paper.

D.William Albert, Dr.K.Fayaz and D.Veerabhadra Babu
Our contributions in this paper include the analysis of association rule mining, its techniques; post mining of association rules and techniques used in the real world; architecture of GPU and its applications for high speed association rule mining. The remainder of this paper is structured as follows. Section II provides information about methods of association rule mining. Section III bestows the analysis of post mining of association rules. Section IV focuses on high speed association rule mining while section V concludes the paper.

[2] METHODS OF ASSOCIATION RULE MINING

Basically association rule mining in the discipline that discovers the hidden correlations among the attributes of a data source. Boukerche and Samarah [2] applied association rule mining on the huge amount of data generated by Wireless Sensor Networks (WSNs). The generated association rules provided the correlations among the sensors that help in making decisions. These researchers also association rule mining mechanisms that are based on coverage and in-network data reduction that leverage correlations among the sensors with respect to locations as well [3]. Association rule mining has been applied to various kinds of data. Shyu et al. [4] applied association rule mining technique on temporal-spatial data obtained from geospatial image databases. Ding et al. [5] applied association rule mining techniques on spatial data. Remote Sensed Imagery (RSI) database was used for their experiments. For compressed and lossless representation of such data, they used Peano Count Tree (PCT) structure. They produced comparatively better results when compared with existing algorithms such as Apriori and FP-growth.

Nestorov and Jukić [6] applied ad hoc association rule mining to data warehouses as they are information rich. Their approach is comprehensive to discover association rules from transactional data as well as historical data which are represented in star schema. Artificial Intelligence (AI) was also used for mining association rules. Huang and Hu [7] applied AI techniques in association with conventional framework to mine association rules for expert decision making. The actual techniques used by them are neural networks and Self-Organization Map (SOM) besides rough set theory. Pang and Kasyanov [8] used SVM classification trees as input for mining association rules. SVM is a famous supervised learning model with underlying algorithms for learning. It is best used to discover patterns and used for regression analysis and classification. Thus they achieved better comprehensibility from the rules that are the result of SVM. They also worked out on the encoding and decoding of rules for knowledge extraction and usage of it in the real world applications.

Couturier et al. [9] improved the presentation of knowledge represented in the form of association rules. They proposed a framework that integrates the data mining process for producing association rules and also visualized the association rules for human computer interaction (HCI). However, they did not integrate the tool with other clustering algorithms. Pan et al. [10] focused on association rule mining with knowledge of a domain expert. As domain experts are aware of the useful support and confidence required for accurate decision making, their technique has assumed
significance. Choi et al. [11] improved the general usage of association rules. They focused on prioritizing the association rules based on the business needs and the associated criteria. They achieved synergic effect of this and the capabilities of data analysis tools. Thus they could address quality, quantity and conflicts problems in association rule mining. Safaei [12] focused on generating fine-grained association rules that can provide comprehensive knowledge for better decision making. The researcher also worked on the post mining process in order to merge association rules that have been discovered in order to make them more meaningful and useful. Au and Chan [13] proposed a novel fuzzy data mining technique that makes use of discovered association rules and find the changes in association rules over time. In fact they mined fuzzy rules in order to discover changes. These researchers could predict well about the changes in the association rules in future based on their fuzzy rules analysis. Lee, Hong and Wang [14] proposed a framework for data mining that produces fuzzy multiple-level association rules from given dataset. Moreover the results are obtained based on the given multiple minimum supports with respect to itemsets and taxonomy. The algorithm they proposed proved to be simple and effective in generating fuzzy multiple-level association rules.

Textual data can also be used for mining association rules. Latiri and Yahia [15] used taxonomy and formal concept analysis (FCA) for generating implicit association rules from textual data. Their approach has two phases. In the first phase they extract formal concepts from textual data explicitly. In the second phase, they use an algorithm in order to extract implicit association rules. Zhong, Li and Wu [16] presented a pattern discovery method that produced association rules as part of text mining. They make some sort of post mining on the generated patterns in order to update then and finding relevant, interesting information. Xie and Liu [17] focused on association rules and attribute implications as two kinds of knowledge. They proposed a framework that makes use of intent reducts and approximate intent reducts in order to discover the two kinds of knowledge. Cherfi, Napoli and Toussaint [18] applied association rule extraction method to textual data. They introduced two classification methods based on classical numerical approach and domain knowledge based approach. They combined the semantic techniques and data mining for post mining of association rules. They built a knowledge model that can be updated incrementally. Singh et al. [19] proposed an optimization algorithm that performs online data mining for generating association rules. They preferred graph to lattice for representing knowledge and generating rules. Martínez-Ballesteros et al. [20] applied evolutionary algorithms for finding gene-gene interrelations. In fact, they extracted these relationships from already discovered quantitative association rules. Their approach also generated some unknown relationships that were considered as hypotheses for further research. Privacy preserving data mining has become very important recently. Giannotti et al. [21] performed experiments on transactional databases in order to mine association rules while preserving privacy of data. They devised an attack model that helped in ensuring the validity of the solution for privacy preserving data mining. Saygin et al. [22] introduced metrics that help in finding how the data mining technique used for mining association rules in actually has privacy preserving element.
Thakkar, Mozafari and Zaniolo [23] proposed an architecture where mining process and post mining are combined into a workflow that provides high quality of services with data sources including the modern data stream management systems. They achieved the high level of mining with SWIM algorithm. Their framework supports many activities such as clustering, trend analysis, and summarization and so on. Continuous post mining process is kept part of the framework that improves quality of service besides reducing the execution time. Jiang and Chen [24] applied post mining of association rules in sensor network. They used post mining on association rules there were meant for sensor data estimation. In fact they focused on informative and non-redundant association rules based on the user interests. Thus the number of rules is reduced which results in improving quality of service. The results of the experiments made by these authors are presented in the following graphs.

![Figure 1 – Average and Maximum Estimation of Accuracy for Traffic Monitoring Application [24]](image)

![Figure 2 – Running Time for Traffic Monitoring Application](image)
As can be seen in Figure 1, 2 and 3 it is evident that the proposed approach for post mining in [24] consumes less memory besides giving high accuracy. However, it takes relatively more time to complete the processing. Marinica and Guillet [25] explored ontologies for postmining of association rules in an interactive mode. Their framework for interactive post mining is as presented in Figure 1.

As can be seen in Figure 4 it is evident that association rules are extracted from database. Afterwards, the user knowledge which is in the form of ontology, rule schemas and operators is applied for post mining which results in filtered rules. The filtered rules are more accurate and can be used directly for decision making. Sulthana and Murugeswari [26] also focused on interactive post mining using ontologies. Their approach is named as “ARIPSO” which is meant for filtering or pruning the discovered rules using rule schema and ontologies.
[4] HIGH SPEED ASSOCIATION RULE MINING

Owens et al. [27] described GPU (Graphics Processing Unit) as powerful, programmable and highly parallel processing unit for general purpose computing application for high speed processing. The very important characteristic of GPU is parallel computing which is required by the real world applications. The power of GPU outperforms that of the conventional CPU (Central Processing Unit). The rapid growth in GPU computing has attracted many researchers in the field to focus on leveraging the power of GPU for general purpose applications. The architecture of modern GPU is as shown in Figure 5.

![Figure 5 – Illustrates the architecture of Modern GPU][27]

As can be seen in Figure 5, it is evident that the GPU architecture is flexible for new hardware innovations besides providing parallel programming a boost. These are the two important characteristics that make GPU useful in real world applications. When GPU is used for data mining, for instance association rule mining, there exists high speed processing with respect to association rule mining. The CPU and GPU performance comparison is shown in Figure 6.

![Figure 6 – Performance of CPU vs. GPU][27]
As can be seen in Figure 6, the scan performance of graphics based GPU (OpenGL); direct-compute GPU (CUDA) is presented. The results reveal that OpenGL and CUDA which are forms of GPU computing exhibit high performance.

Daga et al. [28] analyzed the efficiency of using a fused CPU and GPU in order to leverage parallel computing. They used kernel benchmarks and micro benchmarks and actual applications in order to demonstrate the efficiency of fused approach. These authors made experiments with Nacate APU which is the fused form of both CPU and GPU, Radeon HD 5870, and Radeon HD 5450 in terms of size and bandwidth with respect to host to device and device to host. The results are as shown in Figure 7 and Figure 8.

![Figure 7 – Bandwidth Test (Host to Device) [28]](image)

As shown in Figure 7, it is evident that the size and bandwidth are represented by horizontal and vertical axes respectively. The results revealed that the fusion approach exhibits better performance.

![Figure 8 – Bandwidth Test (Device to Host) [28]](image)

As shown in Figure 8, it is evident that the size and bandwidth are represented by horizontal and vertical axes respectively. The results revealed that the fusion approach exhibits better performance.
[5] CONCLUSION

In this paper we studied the association rule mining which is one of the techniques in data mining. We focused on post mining of association rules and high speed association rule mining as well. The post mining helps in reducing the number of discovered rules while the high speed association rule mining leverages the power of Graphics Processing Unit (GPU). Our analysis has resulted in certain important facts which are described here. Post mining when incorporated into the mining framework provides better performance when compared to the linear usage. Post mining could efficiently reduce the scope of rules and improved the decision making process. GPU helped to make use of recent innovations in computing to speed up the process of data mining. For instance, it can be used for high speed association rule mining. In fact GPU leveraged the speed of general purpose computing besides making the parallel computing possible. Another interesting fact is that the fusion of CPU and GPU in some architecture also proved to be effective in high speed processing required for mining association rules.
REFERENCES


A Survey of Post Mining of Association Rules and High Speed Association Rule Mining


