Volume 5, Issue 3
June 2015

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

FP Growth Tree Implementation In Bank Transactional Databases

Aroosa Hameed

Department of Computer Science, University Of Gujrat, Sub Campus Satellite Town, Rawalpindi.

E-mail:aroosahameed379@yahoo.com

ABSTRACT

As there are significant improvement in IT technologies, these improvements results in a huge amount of data put away in databases and warehouses. Data mining used to mines the data from bulk of data using different views or perspectives and then compresses data into valuable ways. This mined information can be utilized for different business purposes to increase economic development. Now a days, in data mining FP growth algorithm is currently one of the fastest association rules to mine frequent item set. As in the banking sector, numerous transactional databases contain the same set of transactions many times, so in this paper I describe the implementation of the FP Growth tree algorithm in bank transaction databases thus guaranteeing the best performance. The aim of this paper is to ponder and examine the existing association techniques for mining frequent item sets and successful implementation of the FP growth tree algorithm in banking transaction databases to increase the performance of mining frequent item sets.

Keywords: *Association, FP Growth Algorithm.*

1. INTRODUCTION

Analyzes the huge bulk of data from different point of views or perspectives and then summarizes data into useful and meaningful information is called “Data Mining”.

Different data mining procedures can be already defined used to mined data from enormous databases. The one most generally utilized technique is association Rule. Association rule is utilized for mining frequent patterns from hefty databanks. In association rules if-then conditions can be utilized to find associations among unrelated data from the large databases. Different algorithms are used in association rule. The FP growth algorithm is one of the association algorithm that is aimed at finding frequent patterns or item sets from immense databases.

The aim of the research is to implement FP growth algorithm on bank databases, making easier to discover repeated patterns of hefty transactional databases. Generally, a banking database contains a set of customer transactions say T1... Tn. Where each transaction Ti portrays a transaction record of a customer. The transactional database of bank contains over 470 thousand records about customers. To discover frequent patterns, the FP-Growth algorithm has been used.

The algorithm contains two essential strides: compression of the data set in a form of the FP-tree and mining of association rules from the FP-tree. The FP-tree is constructed utilizing two passes over the large database. In the first pass, the algorithm scans the database for all frequent items and after that expels infrequent items from the transaction Ti. As a result, a modified set of transactions T* = T1*…… Tm* is created comprising of only frequent item sets. Then a set of transactions is converted into a compact tree structure called a FP-tree. It is important to note that the FP-tree efficiently compresses the database and stays away from excessive and rehashed data scans as an Apriori type algorithm.

As we have a transactional database of banking systems from which we need to mine the frequent item sets, yet it is hard to mine because of duplication of records and redundancy in the transactional database. So we implement FP growth algorithm to compress all the information at one spot and effectively get frequent item sets and till now there is no implementation of FP growth tree in a transaction database of banking system.

2. RELATED WORK

In the paper [1] [FP-Growth in Discovery of Customer Patterns] the FP growth was used to discover customer frequent shopping patterns. The point of the venture was to outline a stage to incorporate the existing transactional system with new functionalities of rule extraction. In the experimental stage, three practical segments have been distinguished namely, a transactional database with the store commodities tables, software for exchanging information between the stock information system and cash management, and an application to mine association rules. The principle issue with the paper is that the logical databases are used to discover successive patterns of customer which was never remain a good approach in FP growth furthermore they had required extra hardware configurations to execute their investigations. With FP growth, they additionally used some progressions of KDD process which incorporate data cleaning, consolidation, and transformation of data into appropriate form to increase the efficiency of the results.

ITEE Journal
Information Technology & Electrical Engineering
ISSN: - 2306-708X

In the paper [2] [FP Growth Algorithm Implementation] the FP Tree concept was utilized and implemented using Java for a general social survey dataset. They used this approach to determine association rules that occur in the dataset. In this way, they can set up applicable, relevant rules and patterns in any set of records. Along these lines, it was securely presumed that the FP-growth algorithm has a vast future scope in the area of marketing in the organized sector. Henceforth, the issue with this paper is that the inclusion of the FP Tree concept in competitive global markets in the not as good as expected in light of the fact that the numerous other advanced algorithms other than FP growth are used which lower the efficiency of FP tree usage in the competitive global market sector.

In the paper [3] [An Algorithm for Mining Frequent Item sets from Library Big Data] an enhanced FP-Growth algorithm has used called RFP-Growth to abstain from producing intra-property frequent item sets, and to further help its efficiency, implement its Map Reduce version with additional prune strategy. The proposed algorithm was tested using both synthetic and real world library data, and the trial results demonstrated that the proposed algorithm beat existing algorithms.

In the paper [4] [M SUMAN, TANURADHA, K GOWTHAM, A RAMAKRISHNA/ International Journal of Engineering Research and Applications (IJERA)] an algorithm named Apriori-Growth based on Apriori algorithm and FP-Growth algorithm which uses the FP-tree structure to mine frequent pattern. This algorithm combines the advantages of Apriori algorithm and FP-Growth algorithm. The reason to use an Apriori-Growth algorithm is that it doesn't have to produce conditional pattern bases and sub- conditional pattern tree recursively, yet the fundamental issue in this paper is that this strategy only scans the data set twice and builds the FP-tree once while it still needs to generate candidate item sets.

In the paper [5] [An Implementation of the FP growth Algorithm] FP growth algorithm is actualized by utilizing C language which contains two variants of the basic operation of computing a projection of an FP-tree (the fundamental data structure of the FP-growth algorithm). Moreover, anticipated FP-trees are pruned by removing items that have become infrequent because of the projection. But the second projection strategy might have the capacity to yield preferably better results over the first.

In the paper [6] [nonordfp: An FP-Growth Variation without Rebuilding the FP-Tree] they portrayed an implementation of a FP pattern growth based frequent item set mining algorithm. They demonstrated a reduced, memory effective representation of an FP-tree that supports the most important requirements of the core algorithm, with a memory format that permits fast traversal. The implementation based on data structure and several further optimizations. The data structure presented here can oblige the top-down recursion approach, accordingly promote lessening memory need and calculation time.

3. PROPOSED MODEL

As we have a transactional database of banking systems from which we want to mine the frequent item sets, yet it is hard to mine the because of duplication of records and redundancy in the transactional database. So we implement FP growth algorithm to compress all the information at one place and easily get frequent item sets and till now there is no implementation of FP growth tree in a transaction database of banking system.

Volume 5, Issue 3
June 2015

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

IMPLEMENTATION.

The flow chart of complete FP Growth tree implantation is given in figure 1.1:

To implement FP Growth tree algorithm on transactional database, we take some dummy transactions given below:

Transactional ID	**Items**
1	(MoneyTransfer, CashWithDrawal, CashWithDrawal)
2	(CashWithDrawal, CashWithDrawal, Deposits)
3	(Deposits, MoneyTransfer, CashWithDrawal)
4	(MoneyTransfer, CashWithDrawal, Deposits)
5	(Deposit, Deposit,MoneyTransfer)

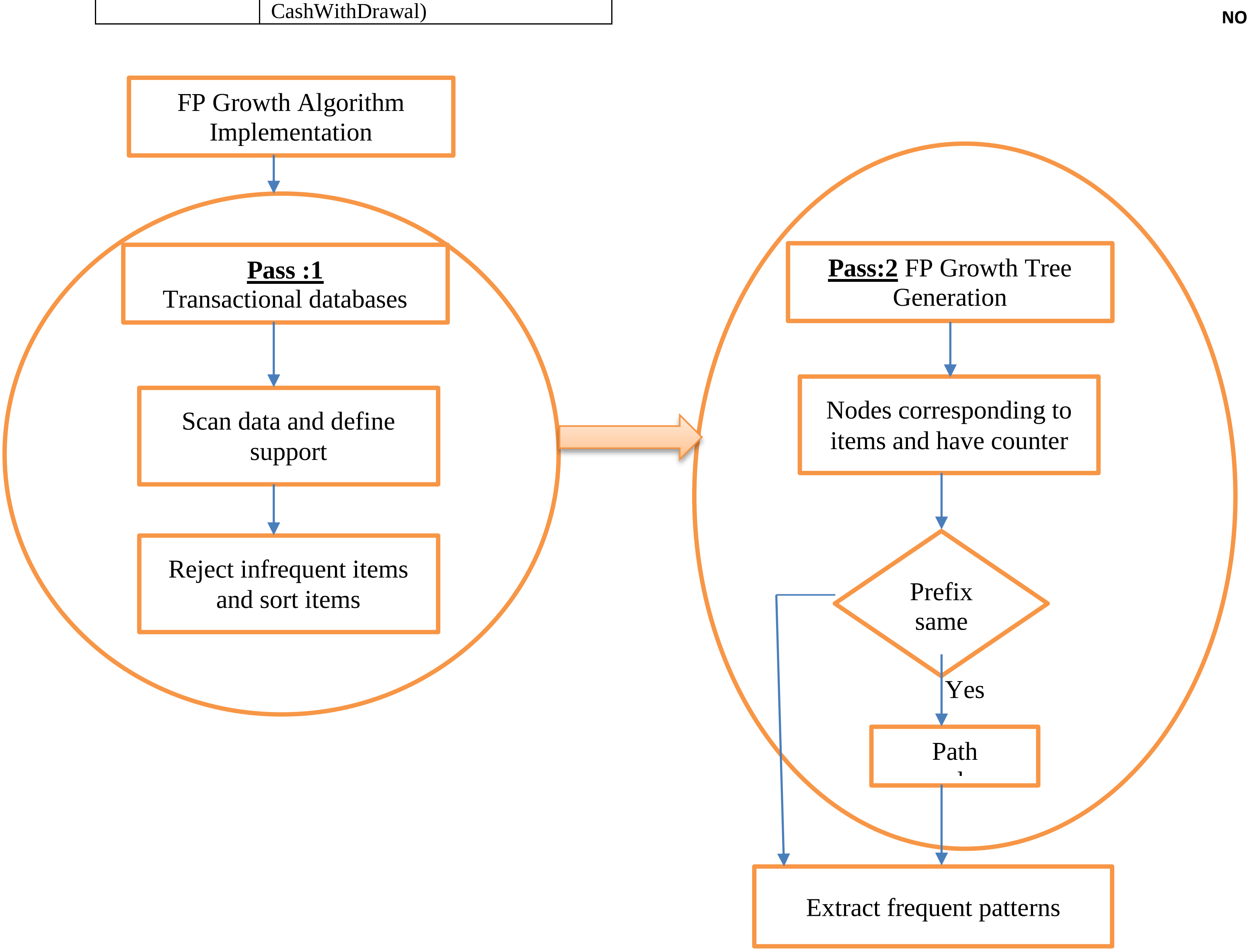
Pass 1:

✓ Support of giving items is:

Support (Money Transfer) =4
Support (CashWithDrawal) =6
Support (Deposit) =5

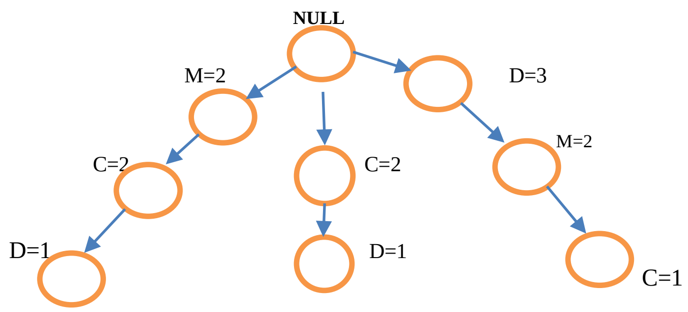
✓ Sorted in descending order:

Items order:
CashWithDrawal, Deposit, Money Transfer



Pass 2:

FP Growth Tree:



4. CONCLUSION

The aim of this research paper was to implement FP Growth tree algorithm in bank transactional databases using the core operation of the FP Growth algorithm This implementation targets the need of using a FP Growth tree algorithm to obtain association rules between related data. Furthermore, different algorithms are used to mine frequent item sets, but these may have different drawbacks thus for large and complex data, time and complexity factors also increases. I believe that this implementation will help the global market to improve their economic development. However, I acknowledge, much work remains to further validate and extend this implementation.

5. ACKNOWLEDGMENT

I would like to thank all those who dedicated their energies, resources and time to the success of this research. The foremost, thank you goes to my ever encouraging teachers. Last but not the least my thanks goes to my friends and family for their help, support and efforts.

REFERENCES

[1]. M. Salam, Dr. Siffat Ullah Khan "Software Development Multi-Sourcing Relationship Management Model (Sdmrmm) Protocol (A Systematic Literature Review)" ISSN 2278-0661 Volume 3, Issue 2 (July-Aug. 2012), PP 24-31

[2]. Thomas Ph. Herz, Florian Hamel, Falk Uebernickel, Walter Brenner "Towards a Multi sourcing Maturity Model as an Instrument of IT Governance at a Multinational Enterprise" Proceedings of the 44th Hawaii International Conference on System Sciences – 2011

[3]. "Outsourcing comes of age: The rise of collaborative partnering" PricewaterhouseCoopers

[4]. B.Huber, Agile Multisourcing Critical Business Trend - Concept and Background, Technology Partners International, Inc., Housten, TX, 2008

[5]. http://www.investopedia.com/terms/o/outsourcing.asp

[6]. http://searchcio.techtarget.com/definition/Multisourcing

[7]. Multisourcing Conflicts: What Does a Manager Focus On?

[8]. Barbara ŁobodaTransaction Costs Lowering as a Motive for Multi sourcing in the Information Technology *2012*

[9]. N. Levina and N. Su, Global Multisourcing Strategy

[10]. Competing across borders "How cultural and communication barriers affect business"

[11]. Thomas Ph. Herz, Florian Hamel, Falk Uebernickel and Walter Brenner "Deriving a Research Agenda for the Management of Multisourcing Relationships Based on a Literature Review" 8-1-2010

[12]. The Keys to Successful Multisourcing

[13]. Bobby Varanasi, COP Chairman & Chief Executive Officer
MATRYZEL CONSULTING "Multi sourcing- A Comparative to Outsourcing." September 2006 Volume I, Issue 2

[14]. Shaun Daly and Michael Brito "Multisourcing Service Integration New Roles, New Rules, New Results" May 2012

[15]. White paper Assembling the Jigsaw Service Integration and Management in a Multi sourced Operating Model

[16]. Case Study "Multi sourcing Services Integration Helps Australian National Audit Office Control Costs and Improve IT Service Quality"

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

[17]. The Rise of Multi sourcing Service Integrators Latest Thinking

[18]. Affan Yasin and Adeel Yasin "How to Overcome Communication Barriers in Global Software Development?"

[19]. 10 Strategies for Overcoming Language Barriers

AUTHOR PROFILES

Aroosa Hameed received her BS degree in Information Technology from University of Gujrat, in 2015. She is a research student of University Of Gujrat. She has done his research by implementing best practices of data warehouse in bank database.