A Comprehensive Overview on Multidimensional Frequent Pattern Mining

Krishna Chaitanya Sanagavarapu*

Network Engineer, Star Tech Networks, Fremont, CA

Abstract – To start with, the great incessant example mining calculations (for example Apriori, FPgrowth) have been centered around mining learning at single idea levels, i.e., either crude or rather high idea level. Notwithstanding, usually attractive to find learning at numerous idea levels. Second, all things considered, applications, numerous measurements, for example, store areas, might be related with exchanges. Joining measurement data into the mining procedure can deliver designs with progressively point by point information. This paper provides a detailed study on multidimensional frequent pattern mining

Index Terms : Data Mining, Multidimensional Frequent Pattern Mining

---------------------------------X---------------------------------

1. INTRODUCTION

For instance, design (store area: BC, IBM Laptop, HP Epson Color Printer) with a help edge of 60% not just educates us of the relationship between two things – IBM Laptop and HP EpsonColor Printer, yet in addition calls attention to that such a mix every now and again happened in the stores situated in British Columbia. Thirdly, recently proposed calculations for multidimensional incessant example mining embraced an Apriori-like technique. It is notable that the Apriori technique depends on iterative example age and various database examines. Subsequently, the proficiency of the Apriori strategy may endure in circumstances of producing long examples. As of late, a novel calculation, FP-development (Han et al 2000, 2004), is proposed to mine regular examples. FP-development is demonstrated to accomplish a superior framework exhibition than customary continuous example mining calculations. Ultimately, the exemplary successive example mining calculations embrace a uniform help limit. However in all actuality, the base help isn't uniform. Outstanding things regularly have either much lower or a lot higher help than general cases. In addition, in most exchange databases, things show up at various reflection levels. Hence, a uniform edge may prompt either produce uninteresting examples at high idea level or miss imperative examples at crude dimension. This part proposes a model for mining with different help requirements and investigates an approach to stretch out FP-development to multidimensional continuous example mining.

2. PROBLEM FORMULATION

A pattern or an item-set, p, is one dimension D_j or one item A_k, or a set of conjunctive items and dimensions $D_i \wedge \ldots \wedge D_j \wedge A_k \wedge \ldots \wedge A_l$, . $A_i \wedge \ldots \wedge A_j \in \tau$. The support of a pattern P is the number of transactions that contain P versus the total number of transactions. Pattern P is frequent if its support satisfies the minimum pattern generation threshold.

The problem of mining multi-dimensional frequent patterns is attacked by implementing the proposed algorithm. The proposed algorithm is able to discover associations between items and dimensions as well as associations among items. The proposed algorithm improves the effectiveness of frequent pattern mining by pushing various support constraints inside the mining process.

3. FREQUENT PATTERN MINING

The process of discovering the complete set of frequent patterns is also called "frequent pattern mining". Its definition is given below.

Definition

Let t = $\{i_1, i_2, \ldots i_m\}$ be a set of items. Let D be a set of transactions, where each transaction T is a set of items such that t T. Patterns are essentially a set of items and are also referred to as item sets, the two terms – "item sets" and "patterns" alternatively. An item set that contains k items is a k -item set. The occurrence of an item set is the number of transactions that contain the item set. This is also known as frequency or support count of the item set.

The task of frequent pattern mining is to generate all patterns (or item sets) whose occurrences (or support) are greater than or equal to the user-specified minimum support. Researchers have been seeking for efficient solutions to the problem of frequent pattern mining since 1993.

4. MULTI-DIMENSIONAL FREQUENT PATTERN MINING

Real transaction databases usually contain both item information and dimension information. Moreover, taxonomies about items likely exist. This section explores the problem of multi-dimensional frequent pattern mining and an example of multidimensional frequent pattern mining as shown in Table 1.

Table 1

An All Electronics Database Illustration

Store Location	Trans-ID	List of Item Ids
BC	001	(TV, Color TV, Sony Color TV); (Computer, Laptop, IBM Laptop); (Printer, Color Printer, HP Epson Color Printer)
ON	001	(Printer, Color Printer, HP Epson Color Printer)
BC	002	(TV, Color TV, Sony Color TV); (Computer, Laptop, IBM Laptop)
ON	002	(Computer, Laptop, IBM Laptop)
BC	003	(TV, Color TV, Sony Color TV); (Computer, Laptop, IBM Laptop)

5. STUDY AND IMPLEMENTATION OF FP-GROWTH ALGORITHM

It is noticed that the bottleneck of the Apriori method rests on the candidate set generation and test. An algorithm called FP-growth given by Han et al (2004) is reported to be faster than the Apriori algorithm. The high efficiency of FP-growth is achieved in the following three aspects. They form the distinct features of FP-growth.

First, an extended prefix tree structure, called frequent pattern tree or FP-tree in short, is used to compress the relevant database information. Only frequent length-1 items will have nodes in the tree, and the tree nodes are arranged in such a way that more frequently occurring nodes will have better chances of sharing than less frequently occurring nodes.

Secondly, an FP-tree-based pattern fragmentation growth mining method – FP-growth, is developed. Starting from a frequent length-1 pattern (as an initial suffix pattern), FPgrowth examines only its conditional pattern base (a “ sub-database” which consists of the set of frequent items co-occurring with the suffix pattern), constructs its conditional FP-tree and performs mining recursively on such a tree. The pattern growth is achieved via concatenation of the suffix pattern with the new ones generated from a conditional FP-tree. Since the frequent pattern in any transaction is always encoded in the corresponding path of the frequent pattern trees, pattern growth ensures the completeness of the result. In this context, FP-growth is not Apriori-like restricted generation-and-test but restricted test only. The major operations of mining are count accumulation and prefix path count adjustment, which are usually much less costly than candidate generation and pattern matching operations performed in most Apriori-like algorithms.

Thirdly, the search technique employed in mining is a partition-based, divide-and conquer method rather than Apriori-like bottom-up generation of frequent pattern combinations. This dramatically reduces the size of conditional pattern base generated at the subsequent level of search as well as the size of its corresponding conditional FP-tree. Inherently, it transforms the problem of finding long frequent patterns to looking for shorter ones and concatenating with the suffix.

The function of FP-growth is to generate all frequent patterns in which database scans are needed only twice. One is used to find out frequent 1-Item-sets and the other is used to construct a FP-tree. The remaining operation is recursively mine the FP-tree using FP-growth. Here, FP-tree resides in main-memory and therefore FP-growth avoids the costly DB scans.

6. CONCLUSION

FP-development calculation is reached out to tackle the issue of multidimensional incessant example mining. The proposed calculation is ensured by the high versatility of FP-development. To build adequacy, the proposed calculation drives different help imperatives into the mining procedure. The proposed calculation is more adaptable at catching wanted information than existing FP-development Algorithm. This paper provided a detailed study on multidimensional frequent pattern mining

REFERENCES

1. Yeshwanth Rao Bhandayker (2016). “Artificial Intelligence and Big Data for Computer Cyber Security Systems” in “Journal of Advances in Science and Technology”, Vol. 12, Issue No. 24, [ISSN : 2230-9659]

2. Sugandhi Maheshwaram (2016). “A Comprehensive Review on the Implementation of Big Data Solutions” in “International Journal of Information Technology and Management”, Vol. XI, Issue No. XVII, [ISSN : 2249-4510]

3. Yeshwanth Rao Bhandayker (2017). “Security Mechanisms for Providing Security to the Network” in “International Journal of

Information Technology and Management", Vol. 12, Issue No. 1, [ISSN : 2249-4510]

4. Sriramoju Ajay Babu, Dr. S. Shoban Babu (2014). "Improving Quality of Content Based Image Retrieval with Graph Based Ranking" in "International Journal of Research and Applications", Volume 1, Issue 1, [ISSN : 2349-0020]

5. Dr. Shoban Babu Sriramoju, Ramesh Gadde (2014). "A Ranking Model Framework for Multiple Vertical Search Domains" in "International Journal of Research and Applications" Vol 1, Issue 1, [ISSN : 2349-0020].

5. Mounika Reddy, Avula Deepak, Ekkati Kalyani Dharavath, Kranthi Gande, Shoban Sriramoju (2014). "Risk-Aware Response Answer for Mitigating Painter Routing Attacks" in "International Journal of Information Technology and Management", Volume VI, Issue I, [ISSN : 2249-4510]

6. Shoban Babu Sriramoju (2017). "Analysis and Comparison of Anonymous Techniques for Privacy Preserving in Big Data" in "International Journal of Advanced Research in Computer and Communication Engineering", Vol 6, Issue 12, DOI 10.17148/IJARCCE.2017.61212 [ISSN(online) : 2278-1021, ISSN(print) : 2319-5940]

7. Shoban Babu Sriramoju (2017). "Review on Big Data and Mining Algorithm" in "International Journal for Research in Applied Science and Engineering Technology", Volume-5, Issue-XI, pp. 1238-1243 [ISSN : 2321-9653], www.ijraset.com

8. Shoban Babu Sriramoju (2017). "OPPORTUNITIES AND SECURITY IMPLICATIONS OF BIG DATA MINING" in "International Journal of Research in Science and Engineering", Vol 3, Issue 6, [ISSN : 2394-8299].

9. Guguloth Vijaya, A. Devaki, Dr. Shoban Babu Sriramoju (2016). "A Framework for Solving Identity Disclosure Problem in Collaborative Data Publishing" in "International Journal of Research and Applications", Volume 2, Issue 6, 292-295, [ISSN : 2349-0020]

10. Shoban Babu Sriramoju (2017). "Heat Diffusion Based Search for Experts on World Wide Web" in "International Journal of Science and Research", https://www.ijsr.net/archive/v6i11/v6i11.php, Volume 6, Issue 11, pp. 632 - 635, #ijsrnet

11. Dr. Shoban Babu Sriramoju, Prof. Mangesh Ingle, Prof. Ashish Mahalle (2017). "Trust and Iterative Filtering Approaches for Secure Data Collection in Wireless Sensor Networks" in "International Journal of Research in Science and Engineering", Volum 3, Issue 4, [ISSN : 2394-8299]

12. Monelli Ayyavaraiah, Shoban Babu Sriramoju (2016). "A Survey on the Approaches in Targeting Frequent Sub Graphs Mining" in "Indian Journal of Computer Science and Engineering (IJCSE)", Volume 9, Issue 2, [e-ISSN : 0976-5166 p-ISSN : 2231-3850], DOI : 10.21817/indjcse/2016/v9i2/180902017

13. Dr. Shoban Babu, Prof. Mangesh Ingle, Prof. Ashish Mahalle (2017). "HLA Based solution for Packet Loss Detection in Mobile Ad Hoc Networks" in "International Journal of Research in Science and Engineering", Volume 3, Issue 4, [ISSN : 2394-8299]

14. Ramesh Gadde, Namavaram Vijay (2017). "A SURVEY ON EVOLUTION OF BIG DATA WITH HADOOP" in "International Journal of Research in Science and Engineering", Vol-3, Issue-6, pp. 92-99 [ISSN : 2394-8299].

15. Shoban Babu Sriramoju (2015). "A Framework for Keyword Based Query and Response System for Web Based Expert Search" in "International Journal of Science and Research" Index Copernicus Value: 78.96 [ISSN : 2319-7064].

16. Sriramoju Ajay Babu, Dr. S. Shoban Babu (2014). "Improving Quality of Content Based Image Retrieval with Graph Based Ranking" in "International Journal of Research and Applications", Volume 1, Issue 1, [ISSN : 2349-0020]

17. Dr. Shoban Babu Sriramoju, Ramesh Gadde (2014). "A Ranking Model Framework for Multiple Vertical Search Domains" in "International Journal of Research and Applications" Vol 1, Issue 1, [ISSN : 2349-0020].

18. Mounika Reddy, Avula Deepak, Ekkati Kalyani Dharavath, Kranthi Gande, Shoban Sriramoju (2014). "Risk-Aware Response Answer for Mitigating Painter Routing Attacks" in "International Journal of Information Technology and Management", Volume VI, Issue I, [ISSN : 2249-4510]

19. Shoban Babu Sriramoju (2014). "An Application for Annotating Web Search Results" in "International Journal of Innovative Research in Computer and Communication Engineering" Vol. 2,Issue 3,

[ISSN(online) : 2320-9801, ISSN(print) : 2320-9798]

20. Namavaram Vijay, S Ajay Babu (2017). "Heat Exposure of Big Data Analytics in a Workflow Framework" in "International Journal of Science and Research", Volume 6, Issue 11, pp. 1578 - 1585, #ijsrnet

21. Shoban Babu Sriramoju (2014). "Multi View Point Measure for Achieving Highest Intra-Cluster Similarity" in "International Journal of Innovative Research in Computer and Communication Engineering" Vol 2,Issue 3, [ISSN(online) : 2320-9801, ISSN(print) : 2320-9798]

22. Ajay Babu Sriramoju, Namavaram Vijay, Ramesh Gadde (2017). "SKETCHING-BASED HIGH-PERFORMANCE BIG DATA PROCESSING ACCELERATOR" in "International Journal of Research in Science and Engineering", Vol-3, Issue-6, pp. 92-99 [ISSN : 2394-8299].

23. Mounica Doosetty, Keerthi Kodakandla, Ashok R., Shoban Babu Sriramoju (2012). "Extensive Secure Cloud Storage System Supporting Privacy-Preserving Public Auditing" in "International Journal of Information Technology and Management", Volume VI, Issue I, [ISSN : 2249-4510]

24. Shoban Babu Sriramoju, Madan Kumar Chandran (2014). "UP-Growth Algorithms for Knowledge Discovery from Transactional Databases" in "International Journal of Advanced Research in Computer Science and Software Engineering", Vol 4, Issue 2, [ISSN : 2277 128X]

25. Namavaram Vijay, Ajay Babu Sriramoju, Ramesh Gadde (2017). "Two Layered Privacy Architecture for Big Data Framework" in "International Journal of Innovative Research in Computer and Communication Engineering" Vol 5, Issue 10, October 2017 [ISSN(online) : 2320-9801, ISSN(print) : 2320-9798]

26. Shoban Babu Sriramoju, Azmera Chandu Naik, N.Samba Siva Rao (2014). "Predicting The Measurability Of Data From Malicious Insiders" in "International Journal of Computer Engineering and Applications" Vol V, Issue II, [ISSN : 2321-3469]

27. Ajay Babu Sriramoju, Dr. S. Shoban Babu (2014). "Analysis on Image Compression Using Bit-Plane Separation Method" in "International Journal of Information Technology and Management", Vol VII, Issue X, [ISSN : 2249-4510]

28. Shoban Babu Sriramoju (2014). "Mining Big Sources Using Efficient Data Mining Algorithms" in "International Journal of Innovative Research in Computer and Communication Engineering" Vol 2, Issue 1,January 2014 [ISSN(online) : 2320-9801, ISSN(print) : 2320-9798]

29. Ajay Babu Sriramoju, Dr. S. Shoban Babu (2013). "Study of Multiplexing Space and Focal Surfaces and Automultiscopic Displays for Image Processing" in "International Journal of Information Technology and Management" Vol V, Issue I, [ISSN : 2249-4510]

30. Dr. Shoban Babu Sriramoju (2014). "A Review on Processing Big Data" in "International Journal of Innovative Research in Computer and Communication Engineering" Vol-2, Issue-1, [ISSN(online) : 2320-9801, ISSN(print) : 2320-9798]

31. Amitha Supriya (2017) "Implementation of Image Processing System using Big Data in the Cloud Environment." International Journal for Scientific Research and Development 5.10: pp. 211-217.

32. Shoban Babu Sriramoju, Dr. Atul Kumar (2012). "An Analysis around the study of Distributed Data Mining Method in the Grid Environment : Technique, Algorithms and Services" in "Journal of Advances in Science and Technology" Vol-IV, Issue No-VII, [ISSN : 2230-9659]

33. S.A. Supriya (2017). "A Survey Model of Big Data by Focusing on the Atmospheric Data Analysis." International Journal for Scientific Research and Development 5.10: pp. 463-466.

34. Shoban Babu Sriramoju, Dr. Atul Kumar (2012). "An Analysis on Effective, Precise and Privacy Preserving Data Mining Association Rules with Partitioning on Distributed Databases" in "International Journal of Information Technology and management" Vol-III, Issue-I, [ISSN : 2249-4510]

35. Shoban Babu Sriramoju, Dr. Atul Kumar (2011). "A Competent Strategy Regarding Relationship of Rule Mining on Distributed Database Algorithm" in "Journal of Advances in Science and Technology" Vol-II, Issue No-II, [ISSN : 2230-9659]

36. Shoban Babu Sriramoju, Dr. Atul Kumar (2011). "Allocated Greater Order Organization of Rule Mining utilizing Information Produced Through Textual facts" in "International Journal of Information

Technology and management" Vol-I, Issue-I, [ISSN : 2249-4510]

37. Monelli Ayyavaraiah (2016). "Review of Machine Learning based Sentiment Analysis on Social Web Data" in "International Journal of Innovative Research in Computer and Communication Engineering" Vol 4, Issue 6, [ISSN(online) : 2320-9801, ISSN(print) : 2320-9798]

38. Monelli Ayyavaraiah (2016). "A Study on Large-Scale Cross-Media Retrieval of Wikipedia Images towards Visual Query and Textual Expansion" in "International Journal for Research in Applied Science and Engineering Technology", Volume-6, Issue-II, pp. 1238-1243 [ISSN : 2321-9653], www.ijraset.com

39. Monelli Ayyavaraiah (2016). "Nomenclature of Opinion Miningand Related Benchmarking Tools" in "International Journal of Scientific & Engineering Research" Vol 7,Issue 8, [ISSN 2229-5518]

40. Siripuri Kiran (2016). 'Decision Tree Analysis Tool with the Design Approach of Probability Density Function towards Uncertain Data Classification', International Journal of Scientific Research in Science and Technology(IJSRST), Print ISSN : 2395-6011, Online ISSN : 2395-602X,Volume 4 Issue 2, pp. 829-831, URL : http://ijsrst.com/IJSRST1841198

Corresponding Author

Krishna Chaitanya Sanagavarapu*

Network Engineer, Star Tech Networks, Fremont, CA