

# Study of Infrequent itemset mining Techniques

Ms.KalyaniTukaramBhandwalkar,Ms.MansiBhonsle

ME(II),Computer,  
G.H.Raisoni College OfEngg,  
Ahmednagar, Maharashtra,India  
*kalyanibhandwalkar@gmail.com*

**Abstract**—Pattern mining has become an important task in data mining. Frequent itemsets find application in a number of real-life contexts. Most of the past work has been on finding frequent itemsets, but also infrequent itemset mining has demonstrated its utility in web mining, bioinformatics, fraud detection and other fields. The infrequent itemset mining problem is discovering itemsets whose frequency of occurrence in the analyzed data is less than or equal to a maximum threshold. This paper provide a general overview of the different papers related to infrequent patterns and gives the knowledge on different algorithms proposed for mining infrequent patterns which are basis for future research in the field of pattern mining.

**Keywords**— Frequent Itemsets, Infrequent Itemsets, Patterns, DataMining,FpgrowthAlgorithm ,Apriori Algorithm,Pattern Mining

## INTRODUCTION

Data Mining extracts knowledge from large databases to discover existing and newer patterns. Data mining is the technique of automatic finding of hidden valuable patterns and relationships from huge volume of data stored in databases in order to help make better business decisions. Discovering useful patterns hidden in a database plays an essential role in several data mining tasks. Frequent itemsets find application in a number of real- life contexts. Buying a PC first, then a digital camera, and then a memory card, if it occurs frequently in a shopping history database, is a frequent pattern. The market basket analysis is one of the applications of the frequent itemset mining .It analyses customer buying habits by finding associations between the different items that customers place in their “shopping baskets”. For instance, if customers are buying soap, how likely are they going to also buy washing power on the same trip to the supermarket? Such information can lead to increased sales by helping retailers do selective marketing and arrange their shelf space. Most of the past work has been on finding frequent itemsets, but also infrequent itemset mining has demonstrated its utility in different areas.

Patterns that are rarely found in database are considered to be uninteresting and are eliminated using the support measure. Such patterns are known as infrequent patterns. An infrequent pattern is an itemset or a rule whose support is less than the minsup threshold. Infrequent patterns are likely to be of great interest as they relate to rare but crucial cases. Examples of applications where mining rare itemsets include identifying relatively rare diseases, predicting equipment failure, and finding associations between infrequently purchased items. In the market basket domain, indirect associations can be used to find competing items, such as desktop computers and laptops, which states that people whom buys desktop computers won't buy laptops. Infrequent patterns can be used to detect errors. For example, if {Fire = Yes} is frequent, but {Fire = Yes, Alarm = On} is infrequent, then the alarm system probably is faulting. Also ,in the study of finding a better treatment approach for a special disease, researchers would like spend more time on studying an abnormal case rather than reading the millions of records of healthy people To detect such unusual situations, the expected support of a pattern must be determined, so that, if a pattern turns out to have a considerably lower support than expected, it is declared as an interesting infrequent pattern.

## LITERATURE SURVEY

R. Agrawal et al in [28] proposed Apriori algorithm, which is used to obtain frequent itemsets from the database.The itemsets which appear frequently in the transactions are called frequent itemsets. MINIT (MINimal Infrequent Itemsets), which is the first algorithm designed specifically for mining minimal infrequent itemsets (MIIs)[2]. A minimal infrequent item-set is an infrequent item-set that do not have a subset of items which forms an infrequent item-set. MINIT is both minimal and non-minimal (unweighted) infrequent itemset mining from unweighteddata.It is based on SUDA2 algorithm. Also showed that the minimal infrequent itemset problem is NP-complete problem.

Author in [10] has purposed a way to find out weights of items and weights of transactions without pre-assigned weights in the database. HITS model is used to derive the weights of transactions from a database with only binary attributes which is based on Link Ranking Analysis. Based on these weights, w-support is defined to give the significance of item sets which differs from the traditional support in taking the quality of transactions into consideration. An Apriori-like algorithm is proposed to extract association rules whose w-support and w-confidence are above some given thresholds.

Mehdi Addaet. al. described ARANIM algorithm for *Apriori Rare and Non-Present Item-set Mining* to mine rare and non-present itemsets in [7]. Non present items are used to detect what is missing in defective process. The proposed approach is Apriori-like and the mining idea behind it is that if the item-set lattice representing the itemset space in classical Apriori approaches is traversed in a bottom-up manner, equivalent properties to the Apriori exploration of frequent item-sets is provided to discover rare item-sets[7]. Also author proposed an approach based on rare patterns to detect suspicious uses and behaviors in the context of a Web application.

[1] is used to find the infrequent patterns and non-present patterns but it does not considers any pruning strategy. It is better to implement any pruning strategy to improve the complexity of the proposed method. Paper [22] proposed Talky-G and Walky-G algorithms. It uses a depth-first strategy for traversal. IFP\_min algorithm described in [3] uses concept of residual tree to reduce computation time. The IFP\_min algorithm recursively mines the minimally infrequent itemsets (MIIs) by dividing the IFP-tree into two sub-trees: projected tree and residual tree[3]. Pattern-growth based algorithms are computationally faster on dense datasets.

Apriori-rare is a modification of the Apriori algorithm used to mine frequent itemsets. To retrieve all rare itemsets from minimal rare itemset (mRIs), a prototype algorithm called "A Rare Itemset Miner Algorithm (Arima)" was proposed in [8]. *Arima* generates the set of all rare itemsets, splits into two sets: the set of rare itemsets having a zero support and the set of rare itemsets with non-zero support. If an itemset is rare then any extension of that itemset will result a rare itemset [8].

Paper [24] presented WSFI (Weighted Support Frequent Itemsets) algorithm where user can specify the weight for each item. A WSFP-Tree store compressed important information about frequent patterns which is an extended FP-tree. It mines the frequent itemsets in only one scan from the data stream .

In [4] proposed FPGrowth- like algorithms i.e. IWI and MIWI for discovering infrequent itemsets by using weights for differentiating between relevant items and not within each transaction. The IWI-support measure is defined as a weighted frequency of occurrence of an itemset in the analyzed data. Occurrence weights are derived from the weights associated with items in each transaction by applying a given cost function. They mainly focuses on IWI- support-min measure and IWI support-max measure which are described as : (i) The IWI-support-min measure, which relies on a minimum cost function, i.e., the occurrence of an itemset in a given transaction is weighted by the weight of its least interesting item, (ii) The IWI support-max measure, which relies on a maximum cost function, i.e., the occurrence of an itemset in a given transaction is weighted by the weight of the most interesting item. As per the analysis of [5] MIWI is the most effective algorithm, which computes in very less computing time, improves the efficiency of performance when the database is large, computes the weighted transaction among the existing algorithms.

## CONCLUSION

Frequent Itemset Mining has attracted plenty of attention but much less attention has been given to mining Infrequent Itemsets. This paper surveys different research papers that proposed various algorithms which are basis for future research in the field of pattern mining. This paper explains different application areas where the infrequent patterns are used. Identifying infrequent patterns efficiently from large datasets and the interesting patterns from the discovered patterns are the challenging tasks in the field of infrequent itemset mining.

## REFERENCES:

- [1] Sujatha Kamepalli, Rajasekhara Rao Kurra and Sundara Krishna Y. K., "Apriori Based: Mining Infrequent and Non-Present Item Sets from Transactional Data Bases," International Journal of Electrical & Computer Science IJECS-IJENS Vol:14 No:03

- [2] Haglin D. and Manning A. (2007).” On Minimal Infrequent Itemset Mining”. In 2007 International Conference on Data Mining (DMIN’07), Las Vegas, June 25-28, 2007, pp. 141- 147.
- [3] A. Gupta, A. Mittal, and A. Bhattacharya, “Minimally Infrequent Itemset Mining Using Pattern-Growth Paradigm and Residual Trees,” Proc. Int’l Conf. Management of Data (COMAD), pp. 57-68, 2011.
- [4] Luca Cagliero and Paolo Garza “Infrequent Weighted Itemset Mining using Frequent Pattern Growth”, IEEE Transactions on Knowledge and Data Engineering, pp. 1- 14, 2013.
- [5] SakthiNathiarasan, Kalaiyarasi, Manikandan , “Literature Review on Infrequent Itemset Mining Algorithms” International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 8, August 2014
- [6] INFREQUENT PATTERNS, CHAPTER- 5 Introduction to data mining, Pang-Ning Tan, Michael Steinbach, Vipin Kumar Pearson Education, book.
- [7] Mehdi Adda, Lei Wu, Sharon White(2012), Yi Feng ” Pattern detection with rare item-set mining” International Journal on Soft Computing, Artificial Intelligence and Applications (IJSCAI), Vol.1, No.1, August 2012.
- [8] L.Szathmary,A.Napoli, P.Valtchev, Towards rare itemset mining ,in: Proceedings of the 19th IEEE Interational Conference on Tools with Artificial Intelligence , 2007, Volume-1, pp.305-312
- [9] JyothiPillai, O.P.Vyas, “Overview of Itemset Utility Mining and its Applications” International Journal of Computer Applications (0975 – 8887) Volume 5– No.11, August 2010
- [10] K.Sun and F.Bai,"Mining Weighted Association Rules Without Preassigned Weights", IEEE Transactions on Knowledge and Data Engineering, Vol. 20, No. 4, April 2008, pages 489-495
- [11] B.Barber, H.J Hamilton , Extracting share frequency itemsets with infrequent subsets, Data Mining and Knowledge Discovery 7(2) (2003)153-185.
- [12] V. Podpecan , N. Lavrac, I. Kononenkom, A fast algorithm for mining utility-frequent itemsets,in workshop on Constraint-Based Mining and Learning at ECML/PKDD ,2007, pp. 9-20.
- [13] G. Cong, A.K.H. Tung, X. Xu, F. Pan, and J. Yang, “Farmer: Finding Interesting Rule Groups in Microarray Datasets,” Proc. ACM SIGMOD Int’l Conf. Management of Data (SIGMOD ’04), 2004.
- [14] W. Wang, J. Yang, and P.S. Yu, “Efficient Mining of Weighted Association Rules (WAR),” Proc. Sixth ACM SIGKDD Int’l Conf. Knowledge Discovery and data Mining (KDD ’00), pp. 270-274, 2000.
- [15] F. Tao, F. Murtagh, and M. Farid, “Weighted Association Rule Mining Using Weighted Support and Significance Framework,” Proc. ninth ACM SIGKDD Int’l Conf. Knowledge Discovery and Data Mining (KDD ’03), pp. 661-666, 2003.
- [16] JyothiPillai and O.P.Vyas ,”TRANSACTION PROFITABILITY USING HURI ALGORITHM [TPHURI]”, International Journal of Business Information Systems Strategies (IJBISS) Volume 2, Number 1,November 2013
- [17] Barber B., Hamilton, H. J. “Extracting share frequent itemsets with infrequent subsets”, Data Mining and Knowledge Discovery, 7(2) (2003), pp 153-185.
- [18] J. Yang and J. Logan. A data mining and survey study on diseases associated with paraesophageal hernia. In AMIA Annual Symposium Proceedings, pages 829–833, 2006.
- [19] Pang-Ning Tan, Michael Steinbach, Vipin Kumar Introduction to data mining, Pearson Education, book.
- [20] Jiawei Han · Hong Cheng · Dong Xin · Xifeng Yan. Frequent pattern mining: current status and future directions.
- [21] Ling Zhou, Stephen Yau, “Efficient association rule mining among both frequent and infrequent items” Computers and Mathematics with Applications 54 (2007) 737–749.
- [22] Laszlo Szathmary<sup>1</sup>, Petko Valtchev<sup>2</sup>, Amedeo Napoli<sup>3</sup>, and Robert Godin<sup>2</sup> 2012.” Efficient Vertical Mining of Minimal Rare Itemsets” ISBN 978,84{695}{5252}{0}, pp. 269{280}, 2012.
- [23] Diti Gupta and Abhishek Singh Chauhan “Mining Association Rules from Infrequent Itemsets: A Survey”, International Journal of Innovative Research in Science, Engineering and Technology, ISSN: 2319-8753, Vol. 2, Issue 10, pp. 5801 – 5808, 2013
- [24] Younghee Kim, Wonyoung Kim and Ungmo Kim “Mining Frequent Itemsets with Normalized Weight in Continuous Data Streams”, Journal of Information Processing Systems, Vol.6, No.1, March 2010
- [25] Laszlo Szathmary, PetkoValtchev, Amedeo Napoli, “Finding Minimal Rare Itemsets and Rare Association Rules”, Proceedings of the 4th International Conference on Knowledge Science, Engineering and Management (KSEM 2010) 6291 (2010) 16—27.
- [26] JyothiPillai, O.P.Vyas, “CSHURI – Modified HURI algorithm for Customer Segmentation and Transaction Profitability”, International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.2, April 2012, pp 79-89.
- [27] J. Han, J. Pei, and Y. Yin, ”Mining frequent patterns without candidate generation,” Proceedings of ACM SIGMOD International Conference on Management of Data, ACM Press, Dallas, Texas, pp. 1-12, May 2000.

- [28] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," in Proceedings of the 1993 International Conference on Management of Data (SIGMOD 93), May 1993, pp. 207–216.
- [29] *en.wikipedia.org/wiki/Data\_mining*
- [30] *en.wikipedia.org/wiki/Association\_rule\_learning*

IJERGS