International Journal of Research Studies in Computer Science and Engineering (IJRSCSE)
Volume 2, Issue 3, March 2015, PP 1-5
ISSN 2349-4840 (Print) & ISSN 2349-4859 (Online)
www.arcjournals.org

Infrequent Pattern Mining from Weighted Transactional Data Set

Sujatha Kamepalli

Research scholar, Computer Science and Engineering Department, Krishna University, Machilipatnam, Andhra Pradesh.
sujatha.cb.srinivas@gmail.com

Raja Sekhara Rao Kurra

Director, Sri Prakash College of Engineering and Technology, Thuni, Andhra Pradesh.
krr_it@yahoo.co.in

Sundara Krishna .Y.K

Professor, CSE Department, Krishna University, Machilipatnam, Andhra Pradesh.
yksk2010@gmail.com

Abstract: *Item set mining is one of the popular data mining techniques in which frequent and infrequent patterns can be mined. Currently the research focuses on infrequent pattern mining. The generating of infrequent item set is valid for the data coming from the distinct real life application background like Statistical disclosure risk evaluation from census data and Fraud detection. Infrequent Weighted Association Mining (IWAM) is one of the main areas in data mining for extracting the rare items in high dimensional datasets. This paper explains about the recent methods proposed for infrequent weighted item set mining. One method is IWI and MIWI to mine the minimal infrequent weighted item sets which uses FP-tree. These methods work efficiently with real weighted data. Another method is based on clustering. This method also scales well and performance has been improved with time and space complexity.*

Keywods: *Infrequent item set, Weighted Association Mining, Clustering, Rare items, and Data mining.*

1. INTRODUCTION

Data mining is known for discovering earlier, suitable, original, useful and reasonable patterns in large databases. Due to the availability of huge amount of data and the need to transform such data into useful information and knowledge, data mining has grown to be the most widely used technique in the society as a whole. Thus data mining can be used for applications ranging from market analysis, fraud detection and customer retention, to production control and science exploration [8]. One of the data mining techniques termed as item set mining is an exploratory data mining strategy is broadly used for generating precious correlation between the data. The first effort for performing item set mining [1] was focused on generating frequent item sets i.e. patterns whose observed frequency of frequency of incidence in the source data is more than the given threshold. In the real life world some applications of frequent item sets are;

1. Medical image processing,
2. Biological data analysis.

In the current years, the concentration of research society has also been focused on the problem of infrequent item set mining i.e. generating the item sets whose frequency of occurrence in the analyzed data is less than or equal to a maximum threshold.[2]. The generating of infrequent item set is valid for the data coming from the distinct real life application background like:

(a) Statistical disclosure risk evaluation from census data and
(b) Fraud detection [3], [4], [5].

Mining infrequent association rules is one of the vital issues in the field of data mining due to its wide range applications. Traditional association rules are derived from frequent item sets, which consider occurrence of items but don't reflect other factors, an example would be profit or price. Weighted Association rule mining has recently been proposed, by which transactions are attached with weighted values according to some measure. However, the exact significance of an item set couldn't be easily recognized by static measures. The problem of weighted association rule mining will be to extract the complete variety of association rules which satisfies a support constraint as well as a weight constraint within the dataset.

Infrequent Weighted Association Mining (IWAM) is one of the main areas in data mining for extracting the rare items in high dimensional datasets. Traditional Association rule mining algorithms produce large

number of candidate sets along with the database scans. Due to large number of transactions and database size, traditional methods consume more time to find the relevant association rules with the specified threshold [6].

2. BACK GROUND WORK

Kalaiyarasi.P, manikandan.M in "Clustering based Infrequent Weighted Item Set Mining" proposed a clustering based algorithm to find the infrequent weighted item set mining [11]. M.Hamsathvani, D.Rajeswari, R.Kalaiselviin " Survey on Infrequent Weighted Item set Mining Using FP Growth" this survey is focused on the infrequent weighted item sets, from transactional weighted data sets to address IWI support measure is defined as a weighted frequency of occurrence of an item set in the analyzed data. Occurrence weights derived from the weights associated with items in each transaction and applying a given cost function [8]. Sakthi Nathiarasan, Kalaiyarasi, Manikandan in "Literature Review on Infrequent Item set Mining Algorithms" focus on reviewing various Existing Algorithms related to frequent and infrequent item set mining which creates a path for future researches in the field of Association Rule Mining [12]. R. Priyanka, S. P. siddique Ibrahim in" A Survey on Infrequent Weighted Item set Mining Approaches" surveys various method of mining infrequent item set. Finally, comparative way of each method is presented [7]. Sujatha Kamepalli, Raja Sekhara Rao Kurra, Sundara Krishna.Y.K. in " Infrequent Weighted Item Set Mining in Complex Data Analysis weighted association rule mining algorithm" proposed to find infrequent item sets using weighted threshold measures. Proposed approach gives better results on real-time datasets compare to existing weighted models[6].Ms.KalyaniTukaram handwalkar,Ms.MansiBhonsle in "Study of Infrequent item set mining Techniques" 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 [13]. Vaidya Seema Bhagwan, A. B. Bagwan in" A Survey: By Using Frequent Pattern Growth Infrequent Weighted Item set Mining" address the topic of generating rare and weighted item sets, i.e. infrequent weighted item set mining problem. The two new excellence measures are proposed for solving the infrequent weighted item set mining problem. Additionally, the two algorithms are represented which perform IWI and minimal IWI mining professionally. Experimental result represents the competence and usefulness of the proposed approach [2]. Vaidya Seema Bhagwan in "Frequent Pattern Growth to Mine Infrequent Weighted Item-Set" address the subject of producing uncommon and weighted item-sets, i.e. rare weighted item-set mining issue. The two new excellence measures are proposed for answering the infrequent weighted item-set mining issue. Moreover, the two calculations are represented to which perform IWI and negligible IWI mining professionally. Trial result shows to the capability and value of the proposed methodology [14]. Sonia jadhav G., M.bhandari in "A Review on Efficient Mining Approach of Infrequent Weighted Item set" focus on review various Existing Algorithms related to frequent and infrequent item set mining which creates a path for future researches in the field of Association Rule Mining [15].

2.1. Basic Concepts

Frequent item set mining

Item set is frequent if its support satisfies given minimum support threshold. Frequent item set find application in many real life contexts. For example buying a PC first, then a digital camera and then a memory card, if it occurs frequently in shopping history, then it is called frequent pattern. Market basket analysis is one of the frequent item set mining applications.

Infrequent item set mining

Item set is infrequent if its support is less than or equal to predefined support threshold. This method has a great interest as they deal with rare but crucial cases. Applications in infrequent item set mining include identifying rare diseases, predicting equipment failure, and finding association between infrequently purchased items [7].

Weighted items transaction databases

A weighted transaction database (*D*) is defined as follows: *D* comprises a set of transactions T = {t1, t2 …tm}, a set of items I = {i1, i2,…. . in} and a set of positive weights W = {w1, w2,…. . wn} corresponding to each item in *I*. For example, consider the data presented in Table 1 and Table 2. Table 1 presents a data set comprising six transactions $T = \{t1 \ldots t6\}$, and five items $I = \{A, B, C, D, E\}$. The weights of these items are presented in Table 2, $W = \{0.6, 0.1, 0.3, 0.9, \text{and } 0.2\}$.

Table1. *The transaction database*

Transactions	Bought Items
1	A, B, D, E
2	B, C, E
3	A, B, D, E
4	A, B, C, E
5	A, B, C, D, E
6	B, C, D

Table2. *Items weight*

Item	Weight
A	0.6
B	0.1
C	0.3
D	0.9
E	0.2

Infrequent weighted item set mining

In a transactional data base each item may associated with a weight. The infrequent weighted item sets are those whose weighted frequency of occurrence is less than a specified weighted threshold.

2.2. Problem Definition

The significance of a weighted transaction, i.e., a set of weighted items, is commonly evaluated in terms of the corresponding item weights. Furthermore, the main item set quality measures (e.g., the support) have also been tailored to weighted data and used for driving the frequent weighted item set mining process.[9]. This paper deals with the problem of mining infrequent weighted item sets from transactional data base i.e. given a weighted transactional data set T, finds infrequent weighted item sets based on weighted support measures.

3. PROPOSED METHODS

- The method proposed to find infrequent weighted item sets is using fp growth algorithm [9].

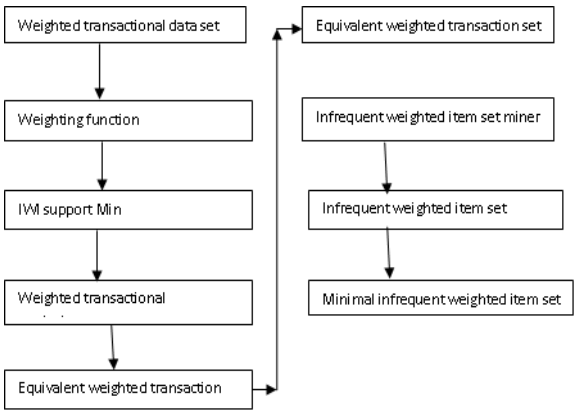
In this method the authors used two support measures to find the infrequent weighted item sets.

- The IWI-support-min measure, which relies on a minimum cost function, i.e., the occurrence of an item set in a given transaction is weighted by the weight of its least interesting item,
- The IWI-support-max measure, which relies on a maximum cost function, i.e., the occurrence of an item set in a given transaction is weighted by the weight of the most interesting item.

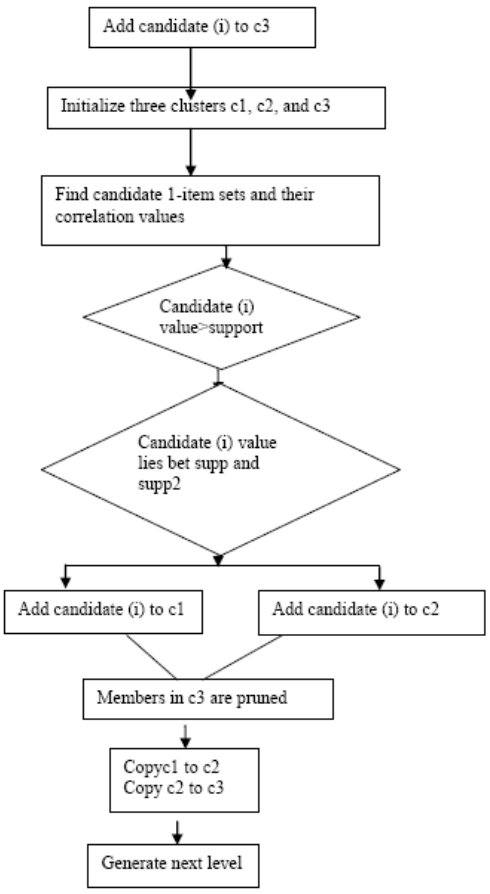
In this paper specifically, the following problems have been addressed:

A) IWI and Minimal IWI mining driven by a maximum IWI-support-min threshold, and
B) IWI and Minimal IWI mining driven by a maximum IWI-support-max threshold.

Task (A) entails discovering IWIs and Minimal IWIs (MIWIs) which include the item(s) with the least local interest within each transaction. Task (B) entails discovering IWIs and MIWIs which include item(s) having maximal local interest within each transaction by exploiting the IWI-support-max measure. To accomplish tasks (A) and (B), we present two novel algorithms, namely Infrequent Weighted Item set Miner (IWI Miner) and Minimal Infrequent Weighted Item set Miner (MIWI Miner), which perform IWI and MIWI mining driven by IWI-support thresholds. IWI Miner and MIWI Miner are FP-Growth-like mining algorithms [10].the flow diagram for the above method is given below.



Another method proposed to find infrequent weighted item sets is based on clustering [11]. The proposed Algorithm is based on the concept of clusters to find infrequent item sets. A cluster is a logical grouping of similar or closely resembling item sets into a single group. The proposed method is explained by the following diagram.



4. CONCLUSION

This paper explains about the recent methods proposed for infrequent weighted item set mining. One method is IWI and MIWI to mine the minimal infrequent weighted item sets which uses FP-tree. These methods work efficiently with real weighted data. Another method is based on clustering. This method also scales well and performance has been improved with time and space complexity. Even though these methods work well there are some limitations. In the first method as the size of the database is complex then it is difficult to prune items within the FP Tree. Max or Min, Equivalent weighted function produces best results on small datasets with limited items. Equivalent weighted function grows as the size of the item set increases. In the second method it has to find the candidate data sets by scanning the data set for number of times.

REFERENCES

[1]. *A. Manning and D. Haglin, "A New Algorithm for Finding Minimal Sample Unique for Use in Statistical Disclosure Assessment", Proc. IEEE Fifth Int'l Conf. Data Mining (ICDM '05), Pp. 290-297, 2005.*

[2]. *Vaidya Seema Bhagwan, A. B. Bagwan"A Survey: By Using Frequent Pattern Growth Infrequent Weighted Item set Mining", International Journal of Science and Research (IJSR) Volume 3 Issue 10, October 2014, ISSN (Online): 2319-7064pp 621-623.*

[3]. *D.J. Haglin and A.M. Manning, "On Minimal Infrequent Item setMining," Proc. Int'l Conf. Data Mining (DMIN '07), Pp. 141-147, 2007.*

[4]. *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.*

[5]. *G. Cong, A.K.H. Tung, X. Xu, F. Pan, and J. Yang, "Farmer: Finding Interesting Rule Groups in Microarray Data sets", Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '04), 2004.*

[6]. *Sujatha Kamepalli Raja Sekhara Rao Kurra Sundara Krishna.Y.K. "Infrequent Weighted Item Set Mining in Complex Data Analysis", International Journal of Computer Applications (0975 – 8887), Volum 103 – No.5, October 2014, Pp12-17.*

[7]. *R. Priyanka, S. P. Siddique Ibrahim "A Survey on Infrequent Weighted Item set Mining Approaches", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 4 Issue 1, January 2015, ISSN: 2278 – 1323,Pp 199-203.*

[8]. *M.Hamsathvani, D.Rajeswari, R.Kalaiselvi "Survey on InfrequentWeighted Item set Mining Using FP Growth", International Journal of Advanced Research in Electrical, Electronics and InstrumentationEngineering, Vol. 3, Issue 11, November 2014, ISSN (Print): 2320 – 3765, ISSN (Online): 2278 – 8875, Pp: 13128-13131.*

[9]. *Luca Cagliero and Paolo Garza "Infrequent Weighted Item set Mining using Frequent Pattern Growth", IEEE Transactions on Knowledge andData Engineering, Vol. 26, No. 4, 1041-4347 2014, Pp: 1-14.*

[10]. *J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data, 2000, Pp. 1–12.*

[11]. *Kalaiyarasi. P, Manikandan. M "Clustering based Infrequent Weighted Item set Mining" International Journal of Advanced Technology in Engineering and Science, ISSN (online): 2348 – 7550 Volume No.03, Special Issue No. 02, February 2015.*

[12]. *Sakthi Nathiarasan, Kalaiyarasi, Manikandan "Literature Review on Infrequent Item set Mining Algorithms", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3, Issue 8, August 2014, ISSN (Online) : 2278-1021 ISSN (Print) :2319-5940,Pp:7670-7673.*

[13]. *Ms.KalyaniTukaramBhandwalkar,Ms.MansiBhonsle " Study of Infrequent item set mining Techniques International Journal of Engineering Research and General Science, Volume 2, Issue 6, October-November, 2014, ISSN 2091-2730, Pp: 676-679.*

[14]. *Vaidya Seema Bhagwan "Frequent Pattern Growth to Mine Infrequent Weighted Item-Set ", International Journal of Science and Research (IJSR), Volume 3 Issue 12, December 2014, ISSN (Online): 2319- 7064pp.2015-2017.*

[15]. *Sonia jadhav G. M.bhandari "A Review on Efficient Mining Approach of Infrequent Weighted Item set "International Journal of Advance Research in Computer Science and Management Studies, Volume 2, Issue 11, November 2014 ISSN: 2327782 (Online), Pp.470-475.*

AUTHORS' BIOGRAPHY

K. Sujatha is pursuing her Ph. D. in Krishna University, Machilipatnam, A. P. She is interested doing research in data mining. Present she is carrying her research in Infrequent Pattern Mining. She has about 10 international journal publications with high impact factors and indexing. She has two national journal publications. She attended for Two AICTE

sponsored 2-week workshops and attended for a number of FDPs. She is member in Indian Association of Engineers (IAE). She has a total of 11 years experience in teaching. She is working as an Associate Professor in CSE Department, Malineni Lakshmaiah Engineering College, Singaraya konda, Prakasam District. A. P.

Prof. K. Rajasekhara Rao is director of Sri Prakash College of Engineering, Thuni, East Godavari District, and Andhra Pradesh. He holds several key positions in K.L.University, as Dean (Administration) & Principal, K L College of Engineering (Autonomous). Having more than 26 years of teaching and research experience, Prof. Rao is actively engaged in the research related to Embedded Systems, Software Engineering and Knowledge Management. He had obtained Ph.D in Computer Science & Engineering from Acharya Nagarjuna University (ANU), Guntur, and Andhra Pradesh and produced 58 publications in various International/National Journals and Conferences. Prof.KRR was awarded with "Patron Award" for his outstanding contribution, by India's prestigious professional society Computer Society of India (CSI) for the year 2011 in Ahemadabad. He has been adjudged as best teacher and has been honored with "Best Teacher Award", seven times. **Dr. Rajasekhar Rao** is a Fellow of IETE, Life Member's of IE, ISTE, ISCA & CSI (Computer Society of India). Dr.Rajasekhar is nominated as sectional committee member for Engineering Sciences of 100th Annual Convention of Indian Science Congress Association. He has been the past Chairman of the Koneru Chapter of CSI.

Dr. Y. K. Sundara Krishna qualified in Ph. D. in Computer Science and Engineering from Osmania University, Hyderabad. Now, he is working as Professor in the Department of Computer Science, Krishna University, Machilipatnam and presently holding several key positions in Krishna University. His research interests are Mobile Computing, Service Oriented Architecture and Geographical Information Systems and having practical work experience in the areas of Computing Systems including Developing Simulators for Distributed Dynamic Cellular Computing Systems, Applications of Embedded and Win32 clients, Maintenance of Multi-user System Software. He has about 24 international publications and about 2 national publications. Has attended about 5 international conferences and 25 national conferences.