International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

A Survey: By Using Frequent Pattern Growth Infrequent Weighted Itemset Mining

Vaidya Seema Bhagwan[1], A. B. Bagwan[2]

Abstract: ***The connection of frequently holding in data in which items may weight differently represented frequented weighted itemsets. Though, in some situation, for example when there is necessitate to diminish a certain cost function, generating rare data correlation is more motivating than mining frequent one. Here in this paper address the topic of generating rare and weighted itemsets, i.e. infrequent weighted itemset mining problem. The two new excellence measures are proposed for solving the infrequent weighted itemset 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.***

Keywords: Classification, Infrequent itemsets, association rule, FP-growth, frequent itemsets.

1. Introduction

One of the data mining technique term as itemset mining, the technique is broadly used for generating precious correlation between the data. The first effort for performing itemset mining [1] was focused on generating frequent itemsets 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 application of frequent itemsets are; 1. Medical image processing, 2. Biological data analysis. Though, numbers of established approaches disregard the interest of each item in within the investigate data. The idea of weighted itemset has also been established for permit treating items differently on the basis of their relevance in the frequent itemset mining process. According to each data item the weight is associated and distinguishes its local significance within each transaction.

In the current years, the concentration of research society has also been focused on the problem of infrequent itemset mining i.e. generating the itemsets whose frequency of occurrence in the analyzed data is less than or equal to a maximum threshold. In [7], [8] algorithms proposed for generating the minimal frequent itemsets do not subsist any infrequent subset. The generating of infrequent itemset 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 [7], [8], [9]. Although the established infrequent mining algorithms suffer from their incapability to take local item interestingness into account during the mining phase. There are two cases, it itemset quality measures used in [4], [5], [6] for constrain the frequent weighted itemset mining process which are not directly relevant to achieve the IWI mining task effectively, on the other hand, infrequent itemset miners are to the best of our knowledge unable to cope with weighted data.

Here in this paper tackles the innovation of infrequent and weighted itemsets. To tackle this subject, the IWI support measures are defined as a weighted frequency of occurrence of an itemsets in the analyzed data. After applying the given cost function to the weights associated with items in each transaction the occurrence weights are derived. Particularly we take attention on the two distinct IWI support measures: (i) The IWI support min measure, which relies on the minimum cost function. (ii) The IWI support max measures, which relies on the maximum cost function. In this paper we tackle the following basics problems:

a. IWI and minimal IWI mining driven by a maximum IWI support min threshold,
b. IWI and minimal IWI mining driven by a maximum IWI support max threshold.

To achieve above tasks, we represent two algorithms which are, infrequent weighted itemset miner and minimal infrequent weighted itemset miner. The algorithm performs the IWI and MIWI mining driven by the IWI support thresholds. Both these algorithms are FP growth like algorithm which satisfied the following features: (i) FP tree node pruning determined by the maximum IWI support constrained and (ii) cost function independence i.e., they work in the same way regardless of which constraint is applied, (iii) early stopping of the recursive FP tree search in MIWI miner to avoid extracting non minimal IWIs.

Rest of the paper is organized as: in section two we are discussed about the existing work done by the researchers i.e., in this section we discussed about the literatures of the proposed work. In section three we discussed the proposed system. In section four we discussed the conclusion drawn for the proposed system and after that the references are present.

2. Literature Review

[1] Introduced the frequent itemset mining technique which is broadly used in data mining technique. The items which are belonging to transactional data are treated uniformly in the existing approached. In [4], the author focus on generating more revealing association rules i.e.,, the weighted associations rules for allowing the problem of differentiating items based on their interest on intensity within each transaction. Though, weights are initiated only during the rule generation step after performing the traditional frequent itemset mining process. In [5], author first time try to pushing items weights into the itemset mining process. For deriving apriori based itemset mining

process the work produces the anti-monotonicity of the proposed weighted support constraint. In many cases the weights are not Preassigned this would be happen in the [4] and [5] only. To tackle this problem, [6] examine the transactional data set was presented as a bipartite hub authority graph and estimated by means of the well known indexing strategy that is HITS [11], in order to automate item weight assignment. Weighted item support and confidence quality indexes are defined accordingly and used for driving the itemset and rule mining phases. Our approach is distinct from the above mentioned approach since it focuses on mining infrequent itemsets from weighted data instead of frequent ones.

The [12] and [13] are the related research issues in probabilistic frequent itemset mining. It required mining the frequent itemsets from the uncertain data in which item occurrence in each transaction is uncertain. To tackle this problem, probabilistic model and have been created and integrated in the Apriori based or projection based algorithms. Though, chances of item occurrence might be remapped to weights, the semantic behind the probabilistic and weighted itemset mining is fundamentally different. The chances of occurrence of an item within a transaction might be totally uncorrelated with its relative importance. For instance, an item that is very likely to occur in a given transaction may be deemed the least relevant one by a domain expert. The proposed approach is distinct from the above mentioned approaches as it specifically tackles the infrequent itemset mining task.

Similar attempt has been dedicated for generating rare correlation between data, i.e. the infrequent itemset mining problems [7], [8], [9], [15], [16], [17]. In [7], [8] for generating minimal unique itemsets for structured data set that is the shortest itemsets with absolute support value equal to 1, a recursive algorithm is proposed. The algorithm proposed in [18], this algorithm was extend by [7], [8] by specially addressing the algorithm scalability issues. In [9] the author first tackle the issue of generating minimal infrequent itemsets that is the itemset which satisfy a maximum support threshold and does not contain any infrequent subset from transactional datasets. In [17], author proposed FP growth like algorithm for mining minimal infrequent itemsets. Author initiates the concept of residual tree for reducing the computational time that is FP tree associated with a generic item i which represent data set transaction obtained by removing i. As similar like the [17], in this paper we proposed an FP tree based approached to mining the infrequent itemsets. Although different from the above mentioned approaches, we face the problems of treating items distinctly, which are based on their relative importance in each transaction, in the discovery of infrequent itemsets from weighted data. Additionally, distinct form [17], we assume a different item pruning strategy tailored to the traditional FP tree structure to perform IWI mining efficiently. In [15], [16] an attempt to develop infrequent itemsets mining positive and negative association rules has been made. Because infrequent itemset mining is considered an intermediate step, their focus is fundamentally different from that of this paper.

3. Proposed Approach

i. Problem Statement

In a given weighted transactional data set T, an IWI support measure based on a weighting function W, and a maximum IWI support threshold t, here we tackles the following tasks:

A. Generating all IWIs which satisfy t in T
B. Generating all MIWIs which satisfy t in T.

Task A demand for discovering all IWIs, and task B choose only minimal IWIs, which represent the smallest infrequent item combination satisfying the constraint.

If W is the minimum weighting function then the IWI support min measure is painstaking and both the task choose all IWI or MIWI which include at least one lowly weighted item in the each transaction. Otherwise, if W is the maximum weighted function, the IWI support max measure is considered and both the task choose all IWI or MIWI that include only lowly weighted items within each transactions.

ii. Algorithms

In this section we represent the two algorithms which are infrequent weighted itemset miner and minimal infrequent weighted itemset miner, which tackles the task A and B which we discussed in the previous section. The proposed algorithms are FP growth like miners whose main features are as follows: (i) the use of equivalence property, for adopting weighted transactional data to transactional FP tree based itemsets mining, and (ii) the exploitation of a novel FP-tree pruning strategy to prune part of the search space early. The proposed two algorithms are as follows:

a. The infrequent weighted itemset miner algorithm

IWI miner is a FP growth like mining algorithm which carry out projection based itemset mining. The algorithm carrying out the following FP growth mining steps: (i) creation of FP tree and (ii) recursive itemset mining from the FP tree index. Distinct from the FP tree IWI miner generates infrequent weighted itemsets instead of frequent ones. For achieving this task, there are some changes in the FP tree growth which are as follows: (i) A novel pruning approach for pruning part of the search space early and (ii) a slightly modified tree structure, which allows storing the IWI support value associated with each node.

b. The minimal infrequent weighted itemset miner algorithm

The MIWI mining process is similar to IWI mining process. Since the MIWI miner focus on discovering only minimal infrequent patterns, the recursive extraction in the MIWI mining procedure is stopped as soon as the infrequent items are occurred.

4. Conclusion

In this paper the problem of generating the infrequent itemset by using the weights for distinguishing among applicable items and not within each transaction. We proposed the two FP growths like algorithms which achieve IWI and MIWI mining competently. The value of the exposed patterns has been authenticated on the data coming from a real life context with the help of domain experts.

In future, we have to incorporate the proposed system in an advanced decision making system which sustain domain expert targeted actions based on the characteristics of the discovered IWIs.

References

[1] A. Manning and D. Haglin, "A New Algorithm for Finding Minimal Sample Uniques for Use in Statistical Disclosure Assessment," Proc. IEEE Fifth Int'l Conf. Data Mining (ICDM '05), pp. 290-297, 2005.
[2] A.M. Manning, D.J. Haglin, and J.A. Keane, "A Recursive Search Algorithm for Statistical Disclosure Assessment," Data Mining and Knowledge Discovery, vol. 16, no. 2, pp. 165-196, http:// mavdisk.mnsu.edu/haglin, 2008.
[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] A. Frank and A. Asuncion, "UCI Machine Learning Repository," http://archive.ics.uci.edu/ml, 2010.
[5] C.-K. Chui, B. Kao, and E. Hung, "Mining Frequent Itemsets from Uncertain Data," Proc. 11th Pacific-Asia Conf. Advances in Knowledge Discovery and Data Mining (PAKDD '07), pp. 47-58, 2007
[6] C.K.-S. Leung, C.L. Carmichael, and B. Hao, "Efficient Mining of Frequent Patterns from Uncertain Data," Proc. Seventh IEEE Int'l Conf. Data Mining Workshops (ICDMW '07), pp. 489-494, 2007.
[7] D.J. Haglin and A.M. Manning, "On Minimal Infrequent Itemset Mining," Proc. Int'l Conf. Data Mining (DMIN '07), pp. 141-147, 2007.
[8] F. Tao, F. Murtagh, and M. Farid, "Weighted Association Rule Mining Using Weighted Support and Significance Framework," Proc. nineth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '03), pp. 661-666, 2003.
[9] 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.
[10] IBM, "IBM Quest Synthetic Data Generation Code," http:// www.almaden.ibm.com/, 2009.
[11] J. Han, J. Pei, and Y. Yin, "Mining Frequent Patterns without Candidate Generation," Proc. ACM SIGMOD Int'l Conf. Management of Data, pp. 1-12, 2000.
[12] J.M. Kleinberg, "Authoritative Sources in a Hyperlinked Environment," J. ACM, vol. 46, no. 5, pp. 604-632, 1999.
[13] K. Sun and F. Bai, "Mining Weighted Association Rules without Preassigned Weights," IEEE Trans. Knowledge and Data Eng., vol. 20, no. 4, pp. 489-495, Apr. 2008.
[14] M.L. Antonie, O.R. Zaiane, and A. Coman, "Application of Data Mining Techniques for Medical Image Classification," Proc. Second Intl. Workshop Multimedia Data Mining in Conjunction with seventh ACM SIGKDD (MDM/KDD '01), 2001.
[15] M.J. Elliot, A.M. Manning, and R.W. Ford, "A Computational Algorithm for Handling the Special Unique Problem," Int'l J. Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 10, no. 5, pp. 493-509, 2002.
[16] R. Agrawal, T. Imielinski, and Swami, "Mining Association Rules between Sets of Items in Large Databases," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '93), pp. 207-216, 1993.
[17] T. Bernecker, H.-P. Kriegel, M. Renz, F. Verhein, and A. Zuefle, "Probabilistic Frequent Itemset Mining in Uncertain Databases," Proc. 15th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '09), pp. 119-128, 2009.
[18] X. Wu, C. Zhang, and S. Zhang, "Efficient Mining of both Positive and Negative Association Rules," ACM Trans. Information Systems, vol. 22, no. 3, pp. 381-405, 2004.
[19] X. Dong, Z. Zheng, Z. Niu, and Q. Jia, "Mining Infrequent Itemsets Based on Multiple Level Minimum Supports," Proc. Second Int'l Conf. Innovative Computing, Information and Control (ICICIC '07), pp. 528-531, 2007.
[20] 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