Associated Sensor Patterns Mining of Data Stream from WSN Dataset

Snehal Rewatkar[1], Amit Pimpalkar[2]
G. H. Raisoni Academy of Engineering & Technology
Nagpur, India.

Abstract - Data mining is the process to discover probably beneficial definite information from the large transactional databases. Association rule mining is most common technique of data mining. It aims at discovering associations between attributes in the large databases. It is used for various applications and also applies on Wireless Sensors Network (WSN) dataset. It is a difficult task to extract the interesting knowledge from WSN in mining techniques. In this paper, we extract the useful information from WSN dataset by using Associated Sensor Pattern Mining of data Stream (ASPMS) algorithm. For finding the frequent patterns among sensors, we apply three algorithms i.e. Apriori, FP-growth and ASPMS algorithm. Results shows in comparative manner, that our technique is time and storage efficient with less memory scan in finding frequent patterns than Apriori and FP Growth algorithm.

Keywords: Data mining, Association rule, WSN, frequent pattern, Apriori, FP-growth, ASPMS.

1. Introduction

Mining play main role to mine frequent item set in many data mining tasks. Over data streams, the frequent itemset mining is mine the approximation set of frequent itemsets in transaction with given support and threshold. It should support the flexible determine between mining accuracy and processing time. When the user-specified minimum support threshold is small, it should be time efficient. To propose an efficient algorithm the objective is generates frequent patterns in a very less time. Frequent patterns are very meaningful in data streams such as in network monitoring, frequent patterns relate an indicator for network attack to excessive traffic. In sales transactions, frequent patterns correspond to the top selling products with their relationships in a market. If we consider that the data stream consist of transactions, each items being a set of items, then the problem definition of mining frequent patterns can be written as given a set of transaction and finds all patterns with frequency above a threshold.

The first traditional algorithm for association rule is Apriori. In Apriori, discovery of interesting rules, multiple database scans, large number of candidate item set generation and discovery of are the main issues. Also in FP Growth algorithm, required more storage space for find the frequent pattern mining and expensive to build the tree. In order to decrease the multiple scanning of database, a new method of association rule mining for pattern generation is proposed in this paper i.e. ASPMS algorithm. This method works efficient with reduction in multiple times scanning of database and less memory space. Thus ASPMS is more efficient than the traditional algorithms.

In existing works, Share-Frequent Sensor Pattern (SFSP) found interesting knowledge by considering the non-binary frequency from sensor databases or streams. It generates sensor association rules to extract the knowledge for frequency patterns. SFSP capture the temporal relationships among the sensor nodes during their event detection process. In this Paper to find frequent patterns we have applied ASPMS, Apriori and FP-Growth algorithm in this paper .ASPMS work with associated sensor pattern stream tree (ASPS tree); it's an innovative tree structure. It is especially design to find frequent pattern among sensors in WSN [12] and it is a sliding window based algorithm. It generates all associated pattern with only one scan of dataset. ASPMS utilize the less memory as it compresses the same frequency nodes into single node using Branch Sort Method (BSM). BSM is keep the frequent items in tree after restructure phase. In this method, according to new insertion list order if path is not sorted, first it removed from tree with deleted non-frequent items, sorted into a temporary array according to new insertion list order and then again inserted into tree in order.

The purpose of this paper to mine the frequent item set over sensor data streams and in this regard creates following influences:

- We describe a new type of interactive patterns for WSNs, called as associated sensor patterns to capture the correlation among sensor data.
- To find frequent patterns, we develop a highly compressed tree structure, called ASPS-tree and a mining algorithm that can well discover patterns with a single scan.
- The ASPMS algorithm is based on sliding window technique extracts associated sensor patterns over a sensor data stream in efficient way.

The experiment shows that the proposed algorithm ASPMS manage high accurate mining data, increase the speed as well as the memory uses is less than the existing mining algorithm.

2. Literature Survey

Association rule mining was first introduced in [1], where it was initially proposed in term of transactional database. It also applied to generate patterns from sensors nodes in WSN. The problem of discovering association rules of sales transactions between items in a large database. Reflect the numbers of epochs in database that's why binary frequency of pattern not sufficient for finds patterns. Behavioral pattern called share frequent sensor patterns (SFSPs) [2]. It used to discover the patterns. To avoid candidate generation, SFSP worked with SFSP-tree. Also presented parallel and distributed for process of high amount data.

S. Singh, et.al [14], studied the various algorithms for mining the frequent pattern. Extension and modification of basic algorithms like Apriori algorithm, FP growth algorithm, Éclat algorithm, and Maximal frequent itemset (MFI) algorithm. J. Jadhav, et.al [15] suggested a new method for finding frequent pattern in case of incremental database which based on Main Memory database Management System (MMDBMS) with help of HSQLDB. This method takes only one database scan and pass for processing frequent pattern. It works efficiently in single and also multi-processing environment which gives better performance than other existing algorithms.

S. Rewatkar, et.al [18], proposed ASPMS algorithm to find frequent pattern among sensors in WSNs for improving the WSNs quality. It worked with innovative tree structure i.e., ASPS-tree which identifies sets of correlated sensors. The result showed that ASPMS worked efficient than Apriori and FP-Growth algorithm. For finding frequent item sets, many algorithms studied and applied from distributed data for creating and mining efficient association rules [19]. Fat Distributed mining (FDM) find the local support counts and prunes all infrequent local candidate sets. Generation of frequent item set among different sites with globally.

In this article [20], for extraction of frequent item sets, generated of both upward and downward closure property which reduced the total number of scans required for the candidate sets generation. Cache optimization techniques improved the Hit/Miss ratio and increased efficiency of the overall frequent pattern mining process.

L. Yue [22], suggested the analysis of improve Apriori, traditional Apriori and FP Growth for mining frequent patterns form undefined data and undefined data streams and also proposed the different solutions in extending the popular techniques into uncertainty environment i.e. U-Apriori or tree structure UF-growth. The experiment result showed that algorithms had reduced memory and run time for undefined data for good efficiency. Especially used in complete the frequent patterns for mining.

In [21], a method for finding recently frequent item sets over a data stream based on Mining Maximum Frequent Item Sets over Data Streams Using Transaction Sliding Window Techniques (MFITSW) mining frequent item set transaction sensitive sliding window algorithm. When the transaction of large number performed from data warehouse, the space of memory was reduced with the number of transaction and time consumed in candidate itemsets transaction to be scanned. The range of data stream was defined by the size of window. This method encapsulated the knowledge in a data stream. The experimental result showed that acquired accuracy mining, consumed less memory with run significant faster than the existing algorithms for mining frequent item sets.

3. Proposed Method

In this paper, find the frequent pattern by applying Apriori and FP-growth algorithm after preprocessing the dataset. This algorithm finds frequent patterns among sensors in WSN.

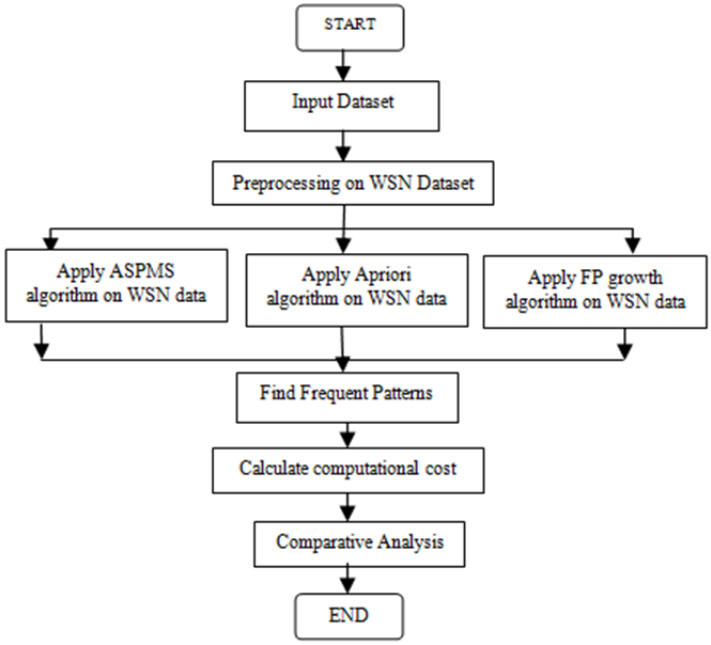

Figure 1.Flow of Proposed System

In proposed system, the first step is preprocessing of dataset. It includes the WSN dataset as an input. It arranged the dataset and store in text file. Preprocessing applied on T10I4D100K dataset, which is synthetic type of dataset with 1, 00,000 transactions. Next step is to applied the association rule mining algorithm i.e. Apriori, FP Growth and ASPMS algorithm. Apriori algorithm is the first traditional algorithm [1]. Apriori is an iterative level wise search Algorithm. It is used for large transactional databases. Apriori is an iterative process where k-itemsets are used to explore (k+1) itemsets.

In each transaction the frequency of an item set is computed by counting its occurrence. Each k-item set must be equal to or greater than minimum support threshold frequency. It generated a large number of candidate itemsets and many times scans the data set as the length of the longest frequent itemsets [14]. There are two steps in Apriori algorithm i.e. the join step and the prune step. Suppose I1, I2, I3……Ik is the itemsets where I1 is the 1-itemset, I2 is the 2-itemset and Ik is the k-itemset. D is the dataset. Ck is the K candidate itemset. In join step, Ck is generated by joining Ik-1 with itself and in prune step, Scan the database to determine the each candidate count in Ck. It should be deleting from the candidate itemsets, when the count is less than the minimums support count. Any (k-1)-itemset that is not frequent cannot be a subset of a frequent k-itemset.

In data mining field, FP Growth algorithm is well-known algorithm for frequent pattern which remove the drawback of Apriori algorithm. It is based on FP tree which is prefix-tree structure used for storing information about frequent patterns [8]. In FP Growth algorithm, first scan the database and find the distinct element with their count, then infrequent patterns are deleted and remaining patterns are sorted in descending order with rearrange the itemsets. Generate FP tree and scan the dataset again. Frequent patterns find in final stage from FP tree. FP Growth algorithm, scan the database only two times than Apriori algorithm and no candidate set of generations. It used divide and conquer approach to decompose the mining problem into a set of smaller problems but one drawback of this algorithm that no new transactions added to database because needs to repeat whole process [14].

Associated Sensor Pattern Mining of data Stream (ASPMS) Algorithm

Frequent Pattern mining techniques find the candidates and frequent patterns generated. In frequent pattern mining techniques for finding frequent patterns contained two problems they are, many times scanned the database and more complex candidate generation process. To find the frequent patterns with single scan of database, we propose a technique ASPMS Algorithm with ASPS-tree which is used to generate associated patterns [12].

ASPMS algorithm can extract useful information for the current window of the sensor from the stream contents in a batch-by-batch manner. ASPS-tree is based on sliding window for WSN that related sensors pattern. The tree is same as FP growth algorithm. The nodes of an ASPS tree in a sensor appearance order and then

restructure the tree in descending order of frequency. After that, compress the tree in a single node by merging the same support sensor node in each branch of the tree. It used BSM method for compresses the same frequency nodes into single node. After restructure phase, BSM is kept the frequent items in tree. In this method if path is not sorted according to new insertion list order, first it removed from the tree, deleted non-frequent items then sorted according to new insertion list order into a temporary array and then again inserted in order into tree. In ASPMS algorithm, new transaction can be added after execution because transaction item divided into window form.

Algorithm: ASPMS Algorithm

Input: Sensor database (SD), Initial sensor appearance order (IO), min sup, min all confidence

Output: Complete set of associated sensor patterns

1: **Begin**

2: Divide window into Batches

3: ASPS-tree a prefix-tree with null initialization

4: **while** (Not end of window) do

5: Scan an epoch from the current location in window;

6: Insert the scanned epoch into ASPS-tree according to IO by following tree construction method:

7: **end while**

8: Restructured tree in descending order using merge-sort method;

9: **for** each branch in ASP-tree do

10: **if** same support of two nodes in same branch do

11: Merge the branch in tree using branch sorting method (BSM);

12: **end if**

13: **end for**

14: Find frequent patterns from step 10

4. Experimental Result

In this experiment, we show the result of proposed ASPMS algorithm. We have to implement our sliding window algorithm program on Windows platform. To appraise the performance of our approach, we performed experiments on IBM synthetic dataset T10I4D100K, real life dataset Kosark, other is Mushroom and Retail dataset from frequent itemset mining dataset repository [23].A synthetic dataset T10I4D100K has been used for analysis, which contains 100000 transactions with 870 items, has an average of 5 of items (T5) with average maximum frequent item set size of 4 items (I4) and number of sequences in the dataset (D100K). Kosarak dataset is a real life dataset which contain 41,270 items.

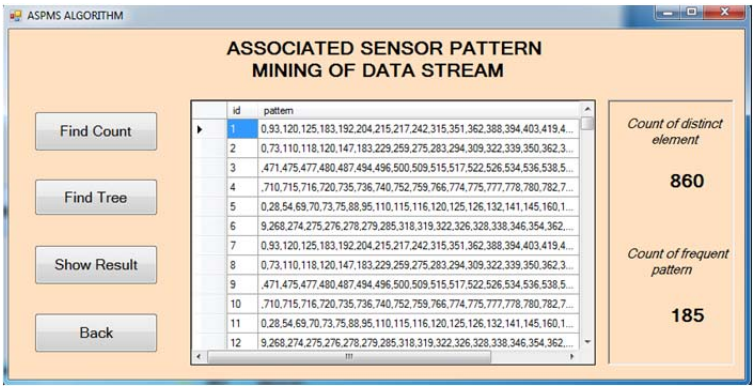

Figure 2.Output screen of frequent patterns by applying ASPMS algorithm.

Figure 2 shows that the frequent pattern by using ASPMS algorithm. Obtain the set of association rules by applying algorithms. By analyzing the data, we can obtain different number of rules and get different frequent patterns of both three algorithms with regard to giving different values of support and confidence. Apriori scan the database for each time a candidate item set is generated, FP growth scans the database only two times and ASPMS scan the database once.

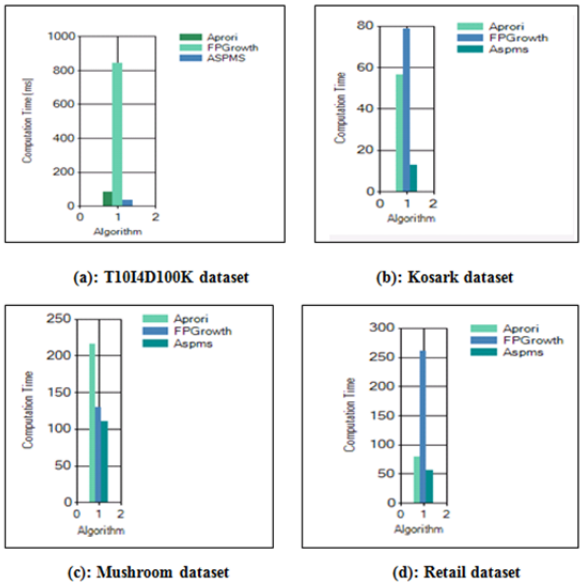

Figure 3- Execution time Comparison Graph

The Results of execution time comparison show in Figure 3. Compare the execution time of all three algorithms which apply on four different dataset.

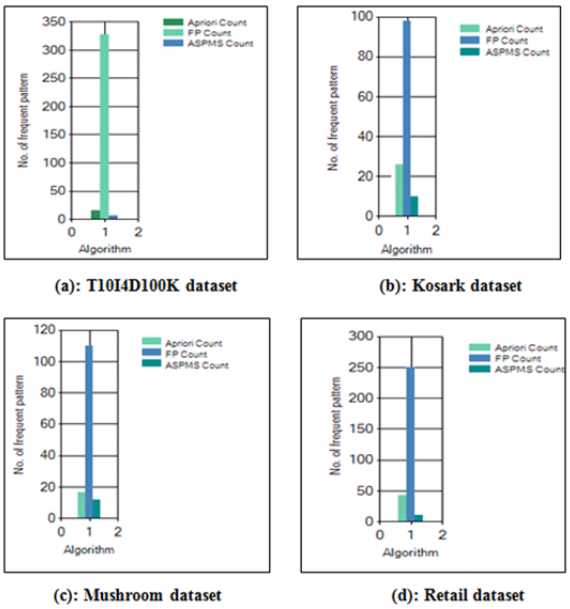

Figure 4- No. of Frequent Pattern Fetches Comparison Graph

Figure4shows, that the graph between number of patterns and algorithms. We compared the performance our algorithm on the basis of the computation time and the number of frequent pattern. The experimental result shows that proposed algorithm ASPMS takes less time for generating frequent pattern and it fetches less frequent pattern i.e., finds the exact frequent patterns than existing algorithm for large transactional dataset.

5. Conclusion

Mining beneficial knowledge from huge amount of transaction data is a difficult task. Major issues in this framework were how to avoid complex candidate generation process, large number of database scans and execution time and memory requirements for large transactional database. In this work, we proposed an efficient one- ASPMS algorithm. It is used to mine the set of all frequent item sets with a sliding window in data streams. We compared the three algorithms with the parameters like computation time, storage structure, efficiency and database scan etc. The experiments indicate that the proposed algorithm manage highly precise mining results with the faster performance for mining frequent item over data-streams. In this paper our analysis i.e. theoretical and experimental shows that the proposed algorithm ASPMS is efficient and scalable and perform better for mining the set of all maximum frequent item sets over the entire existing of the data streams. In future, we will use the ASPMS algorithm for different approaches where mine the data. It may increase the space efficiency.

References

[1] R. Agrawal and R. Srikant, "Fast algorithms for Mining Association Rules", 20th International Conference on Very Large Data Base, pp. 487–499, May1994.

[2] Md. Mamunur Rashid, Iqbal Gondal and Joarder Kamruzzaman, "Share-Frequent Sensor Patterns Mining from Wireless Sensor Network Data", IEEE Transaction Parallel Distribution System, 2014.

[3] K.M. Yu, J. Zhou and W.C. Hsiao, "Load Balancing ApproachParallel Algorithm for Frequent Pattern Mining", Proceedings PaCT, pp.623-631, 2007.

[4] S. Sun and J. Zambreno, "Design and analysis of a reconfigurable platform for frequent pattern mining", IEEE Transaction Parallel Distribution System, Vol.22, no. 9, pp. 1497, 2011.

[5] Anjan Das, "An Enhanced Data Reduction Mechanism to Gather Data for Mining Sensor Association Rules", IEEE Transaction Parallel Distribution System, pp. 120-136, 2011.

[6] C. Nawapornananand V. Boonjing, "An efficient algorithm or mining complete share-frequent item sets using Bit Table and heuristics", Proceeding of ICMLC, Vol.1, pp.96-101, 2012.

[7] A. F. Chowdhury, T. K. Syed and J. Byeong-Soo, "An Efficient Method for Incremental Mining of Share-Frequent Patterns" in International Asia-Pacific Web Conference, 147-153, 2010.

[8] Jiawei Han,Jia Pei, Yiwen Yin and RunyingMao,"Mining Frequent Patterns without CandidateGeneration: A Frequent-Pattern Tree Approach", Data Mining and Knowledge Discovery, 8, 53–87, 2004.

[9] Ramesh C. Agarwal, Charu C. Aggarwal and V.Prasad, "A Tree Projection Algorithm for Generation of Frequent Itemsets", Vol. 61, Issue 3, pp 350-371, March 2001.

[10] M.M. Rashid, I.Gondal and J.Kamruzzaman, "Mining Associated Patterns from Wireless Sensor Networks", IEEE Transaction on Computers, August 2014.

[11] Akshita Bhandari, Ashutosh Gupta and Debasis Das, "Improvised Apriori algorithm using frequent pattern tree for real time applications in data mining", Vol. 46, Pages 644–651, 2015.

[12] Shamila Nasreen, Muhammad Awais Azam, Khurram Shehzad, Usman Naeem and Mustansar Ali Ghazanfar, "Frequent Pattern Mining Algorithms for Finding Associated Frequent Patterns for Data Streams: A Survey",EUSPN, pp.109 – 116 -2014.

[13] Jiawei Han, Jian Pei, and Yiwen Yin, "Mining Frequent Patterns without Candidate Generation", 2000.

[14] Sukhbir Singh and Dharmender Kumar, "Frequent Pattern Mining Algorithms: A Review", International Journal of Advanced Engineering and Nano Technology (IJAENT), Volume-1, September 2014.

[15] Jyoti Jadhav, Lata Ragha and Vijay Katkar, "Incremental Frequent Pattern Mining", International Journal of Engineering and Advanced Technology (IJEAT), Volume-1, August 2012.

[16] Ms. Deepika Fole and Asst Prof. Omprakash Dewangan, "Cross Level Frequent Pattern Mining Using Dynamic Programming Approach", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Volume 4, May 2015.

[17] Pratima O. Fegade and Prof. Dinesh D. Patil, "Mining Frequent Itemsets for improving theEffectiveness of Marketing and Sales" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5, 3736-3738, 2014.

[18] Snehal Rewatkar and Amit Pimpalkar, "ASPMS Based Approach to Mine Frequent Pattern from WSN Dataset", International Journal of Advanced research in Computer Science and Software Engineering (IJARCSSE), Vol. 5, pp 151-154, November 2015.

[19] Bhagyashri Waghamare and Bharat Tidke, "Review:Association Rule for Distributed Data", International Journal of Computer Science & Communication Networks,Vol 5(2),58-61.

[20] Vaibhav Kant Singh, Vijay Shah, Yogendra Kumarrin, Anupam Shukla, A.S. Thoke, Vinay Kumar Singh, Chhaya Dule and Vivek Parganiha, "Proposing an Efficient Method for Frequent Pattern Mining", International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:2, No:12, 2008.

[21] Neeraj And Anuradha, "Mining Maximum Frequent Item Sets Over Data Streams Using Transaction Sliding Window Techniques", International Journal of Information Technology Convergence and Services (IJITCS) Vol.3, No.2, April, 2013.

[22] Liwen Yue, "Review of Algorithm for Mining Frequent Patterns from Uncertain Data", IJCSNS International Journal of Computer Science and Network Security, VOL.15 No.6, June 2015.

[23] Frequent Itemset Mining Dataset Repository, http://fimi.ua.ac.be/data/T10I4D100K.dat., https://archive.ics.uci.edu/ml/datasets/Mushroom, https://archive.ics.uci.edu/ml/datasets/Online+Retail, http://www.philippefournierviger.com/spmf/index.php?link=datasets.php,