ISSN: 2456 - 3935
International Journal of Advances in Computer and Electronics Engineering
Volume: 02 Issue: 03, March 2017, pp. 28 – 32

# Mining of Behavioral Patterns Using Prefix Sharing ASP Tree in Wireless Sensor Networks

Anitha Shanmugam, Dr. Abirami Thiayagarajan

Assistant Professor, Department of Information Technology, Kongu Engineering College, India[1]

Assistant Professor (SRG), Department of Information Technology, Kongu Engineering College, India[2]
anithame@kongu.ac.in[1], abiananthmca@gmail.com[2]

**Abstract:** *A new type of associated sensor behavioral patterns for Wireless Sensor Networks (WSNs) is defined to capture the true correlation among sensor data. For this, a highly compact tree structure called associated sensor pattern tree (ASP-tree) and a mining algorithm that can efficiently discover patterns from sensor database (SD) with a single scan is devised. Existing work uses occurrence frequency of patterns to extract knowledge in order to generate sensor association rules. These techniques often generate huge number of rules, most of which are non-informative or fail to reflect true correlation among sensor data. Frequent Pattern Growth algorithm finds frequent item sets without generating any candidate item set and scans database just twice. This algorithm concentrates only on the item in the transaction and not the utility of the item. All the products are treated uniformly and all the rules are mined based on the count of the product. So, in order to mine useful information, concept of Associated Sensor Pattern tree is constructed so that the number of sensor data gets reduced leading to reduction in memory space. Execution time is also decreased so that the sensor can quickly capture the data and send it to sink nodes.*

**Keyword:** *Mining; Wireless Sensor Networks; Behavioral Patterns; Associated Sensor Pattern Tree; Sensor Database; Frequent Pattern Growth Algorithm;*

## 1. INTRODUCTION

Data mining techniques, well established in the traditional database systems, recently became a popular tool in extracting interesting knowledge from sensor data streams (SDSs). Using knowledge discovery in WSNs, one particular interest is to find behavioral patterns of sensor nodes evolved from meta-data describing sensor behaviors. The application of fine-grain monitoring of physical environments can be highly benefitted from discovering behavioral patterns (i.e., associated patterns) in WSNs.

These behavioral patterns can also be used to predict the cause of future events which is used to detect faulty nodes, if any, in the network. For example, possibility of a node failure can be identified using behavioral pattern mining by predicting the occurrence of an event from a particular node, but no such event reported in subsequent iteration. As behavioral patterns reveal a chain of related events, source of the next event can be identified. For e.g. in an industry, fault in a particular process may trigger fault in other processes. In addition, behavioral patterns can also used to identify a set of temporally correlated sensors, thus improving operational aspects in WSNs.

The main challenges in mining associated sensor patterns are:

1. Suitable formulation to discover associated sensor patterns maintaining the downward closure property in order to ensure search space reduction
2. Inventing a compact tree structure which is able to capture the data content in one scan over sensor datasets in order to yield better mining performance
3. Building adaptive tree structure by eliminating unimportant or obsolete information in the data stream and capturing the latest information makes memory usage in an optimal way thereby extracting complete set of recent associated sensor patterns

These challenges can be rectified by defining a new type of behavioral patterns for WSNs, termed as associated sensor patterns to capture true correlation among sensor data. To discover such patterns, the proposed work devise a highly compact tree structure called associated sensor pattern tree and a mining algorithm that can efficiently discover patterns from sensor database (SD) with a single scan.

## 2. RELATED WORK

Mining Frequent itemsets from transactional data streams is a very challenging task as there is uninterrupted, unrestrained and structured sequence of data elements generated at a fast rate in a data stream. In order to improve the stream data analysis, it is necessary to extract frequent itemsets from more current data. For this purpose, a sliding window mechanism is used. Further, usage of memory resources must be taken into account regardless of the amount of data generated in the stream. Proposed RA-FIG (Resource Adaptive Frequent Item Generation) algorithm in [1] accounts for the computational resources like memory availability and dynamically adapts the rate of processing based on the available memory. Extensive experimental analysis shows that the proposed algorithm is efficient in terms of resource utilization and accuracy when finding recent frequent itemsets from a data stream.

In [2], vision for mining fine-grained urban traffic knowledge from mobile sensing, especially GPS location traces is introduced. Apart from characterizing human mobility patterns and measuring traffic congestion, this paper expose details such as intersection performance statistics of mobile sensing that are useful for optimizing traffic signal timing. Realizing For such applications realization, co-designing privacy protection algorithms and novel traffic modeling techniques are implemented. By this, privacy preserving and traffic modeling can be simultaneously satisfied. Based on the virtual trip lines (VTL) concept, privacy algorithms are used to regulate where and when the mobile data should be collected. Integration of traffic principles and learning/optimization techniques are also featured. Using two case studies for extracting traffic knowledge, proposed methods are illustrated for urban signalized intersection.

WSNs generate a large amount of data in the form of data stream and mining these streams to extract useful knowledge is a highly challenging task. In literature study [3], existing mechanism use sensor association rules measured in terms of frequency of patterns occurrence. Among the enormous number of rules generated, most of those are not valuable to reproduce true association among data objects. Moreover, mining associated sensor patterns from sensor stream data is essential for real-time applications, but it is not addressed in literature papers. In this proposed work, a new type of sensor behavioral pattern called associated sensor patterns to capture substantial temporal correlations in sensor data simultaneously is introduced to address the above-said problem. In this paper, a novel tree structure called associated sensor pattern stream tree (ASPS-tree) and a new technique using sliding window-based associated sensor pattern mining for WSNs called associated sensor pattern mining of data stream (ASPMS) is proposed. By capturing the useful knowledge of the data stream into an ASPS-tree, ASPMS algorithm mine associated sensor patterns in the current window with frequent pattern (FP)-growth. Extensive experimental analyses show that the proposed technique is very efficient in discovering associated sensor patterns over sensor data stream.

A vision for mining fine-grained urban traffic knowledge from mobile sensing [4], especially GPS location traces. Beyond characterizing human mobility patterns and measuring traffic congestion, this paper is useful for optimizing the timing of a traffic signal by revealing details such as intersection performance statistics. Such applications can be realized by the knowledge of co-designing privacy protection algorithms and novel traffic modeling techniques, thereby satisfying privacy preserving and traffic modeling needs simultaneously. This paper explores privacy algorithms based on the virtual trip lines (VTL) concept to regulate where and when the mobile data should be collected. The traffic modeling is an integration of traffic principles and learning/optimization techniques.).

Frequent Itemset Mining (FIM) is one of the most well known techniques to extract knowledge from data. Combinatorial explosion of FIM methods become more problematic when Big Data is involved in it. Auspiciously, current improvements in the field of parallel programming provide good tool to tackle this problem inspite of its technical challenges like balanced data distribution and inter-communication costs. This paper [5] investigates the applicability of FIM techniques on the MapReduce platform by introducing two new methods for mining large datasets: Dist-Eclat and BigFIM focusing on speed and optimized to run on large datasets respectively.

In [6], FP-Growth Algorithm uses an extended prefix-tree structure for storing compressed and crucial information about frequent patterns named frequent-pattern tree (FP-tree) for mining the complete set of frequent patterns. Even though it is an efficient and scalable method, it faces the issue of discovering frequent sensor data between relevant sensor data only. Moreover, it generates all the associated patterns with twice scan over the synthetic dataset. So, usefulness of the discovered patterns has not been tested on data coming from a real-life context.

Frequent Pattern Growth algorithm finds frequent item sets without generating any candidate item sets and scans database just twice. FP Growth algorithm concentrates only the item in the transaction and not the utility of the item. All the products are treated uniformly and all the rules are mined based on the count of the product. So the concept of weighted items was introduced. Weight association rule mining considers the importance of items such as transaction databases, but items transactions are not taken into consideration. Existing mechanism uses occurrence frequency of patterns to extract the knowledge in order to generate sensor association rules. This technique often produce enormous amount of rules, most of which are not useful or unsuccessful in reflecting true correlation among sensor data.

## 3. MATERIALS AND METHODS

In the proposed system, the sensor database is collected from various sensors at different time slot. The data is used for the construction of ASP-tree which consists of sensor data and its support count value and it is compressed finally. ASP-tree contains certain Characteristics such as the support value of any node in the ASP-tree is greater than or equal to the sum of the total support value of its children.

The ASP-tree can be constructed in a single scan of the sensor database. For any epoch in a SD, there exists a unique path in the ASP-tree starting from the root. From the use of ASP-tree constructed, the conditional pattern base tree is constructed based on the minimum support count threshold value and conditional pattern base. The basic operations for associated sensor pattern mining from the ASP-tree are: (i) counting length-1 frequent sensors, (ii) construction of conditional pattern-base for each sensor and (iii) conditional tree construction for each conditional pattern-base. Finally, associated patterns are generated from the conditional tree constructed.

### 3.1 ASP -tree construction phase

In the construction of ASP-tree phase, sensor node details are involved in each transaction. In the grid view control, all the records are displayed from which the records can be modified and new values can be updated. In addition, if an item support count is higher than the minimum support count then it will be highlighted. This phase considers only a node with maximum support count otherwise nodes are removed from each transaction. In addition, transaction entries are ordered based on the support count.

These details are stored in 'Ordered' table and viewed by using grid view control. ASP-tree arranges the sensor node according to transaction appearance order in the database and is built by inserting in the transaction one after another for every epoch. At this stage, ASP-tree simply maintains a transaction order-list. It includes each distinct sensor node found in all epochs in the transaction according to their appearance order and contains the support value of each node in the transaction.

### 3.2 ASP -tree compression phase

In the compression of ASP- tree phase, tree is restructured in a descending order of frequency and finally compresses the tree by merging the same support sensor node in a single node in each branch of the tree. The purpose of the restructuring-compression phase is to achieve a highly compact ASP-tree which will utilize less memory and facilitate fast mining process.

### 3.3 ASP mining phase

Construction of a highly compact ASP-tree enables subsequent mining of associated sensor patterns by using pattern growth approach. Similar to the FP-growth mining approach, ASP-tree of decreasing size are mined recursively to generate associated patterns by creating conditional pattern-bases (CPBs) and corresponding conditional trees (CPBTs) without any additional database scan.

The basic operations for associated sensor pattern mining from the ASP-tree are construction of

(i) Conditional pattern-base for each sensor
(ii) Conditional tree for each conditional pattern-base

The three bottom sensor nodes do not satisfy the min sup threshold. Therefore, conditional pattern-base tree of s5 is created by taking all the branches prefixing the sensor node s5. Considering s5 as a suffix, its corresponding three paths are {s1s2s4: 1}, {s2s7: 1} and {s1: 1}. The conditional-tree of s5 is empty which indicates that there is no associated pattern for s5. The sensor node s1, s2, s4 and s7 are not included in the conditional-tree because their support values are less than min sup and all confidence value of {s1s5}, {s2s5}, {s4s5} and {s5s7} are less than minimum of all confidence values. Then, finally, associated patterns are generated from the conditional tree.

Due to the wide diversity of sensors, time consumption of sensors varies greatly. For passive sensors, such as light (L) or temperature sensors, power consumption is negligible in comparison to other devices on a frequent sensor wireless sensor node. On the other hand, for active sensors (such as sonar, soil and gas sensors) time consumption can be significant. Each ASP sensor node can comprise several sensors and each of these sensors typically has its own energy characteristics and in some cases, its own sampling frequency. Sensor, *i*, will have the following sensing energy consumption as indicated in below Equation (1).

$$E_{Tm} = Vdc*Ii*Ti \quad (1)$$

where $T_i$ is the time required for obtaining a single sample from sensor $i$ and $I_i$ is the sensor $i$'s current draw. $T_i$ depends on the start-up ($T_s$), response ($T_r$) and measurement ($T_m$) times of the sensors. For most sensors, as $T_m$ is small in comparison to $T_s$ and $T_r$, only $T_s$ and $T_r$ was considered in calculating $T_i$ considering the value $V_{dc}$= 0.5 as given in below Table II.

The startup time ($T_s$) is the time required for a sensor to reach the ready state after time is engaged, upon which the sensor can give the correct value. It is a well-known factor in the time management of sensors. If a sensing task does not wait for $T_s$, after that the micro controller unit (MCU) requests the sensor to turn on, then the task will receive the wrong value. $T_s$ varies significantly between sensor types.

ISSN: 2456 - 3935
International Journal of Advances in Computer and Electronics Engineering
Volume: 02 Issue: 03, March 2017, pp. 28 – 32

TABLE II EXECUTION TIME ANALYSIS-
FP GROWTH – ASP TREE

| Light Sensor | $I_i$ (sec) | $T_m$ (ms) | | $T_i$ (ms) | $E_{Tm}$ (ms) |
|---|---|---|---|---|---|
| | | $T_s$ (ms) | $T_r$ (ms) | | |
| L1 | 0.10 | 0.014 | 0.018 | 0.04 | 0.002 |
| L2 | 0.20 | 0.023 | 0.028 | 0.05 | 0.005 |
| L3 | 0.30 | 0.036 | 0.042 | 0.06 | 0.009 |
| L4 | 0.40 | 0.041 | 0.048 | 0.07 | 0.014 |
| L5 | 0.50 | 0.052 | 0.061 | 0.09 | 0.022 |
| L6 | 0.60 | 0.063 | 0.074 | 0.11 | 0.033 |
| L7 | 0.70 | 0.075 | 0.087 | 0.12 | 0.042 |
| L8 | 0.80 | 0.082 | 0.096 | 0.14 | 0.056 |

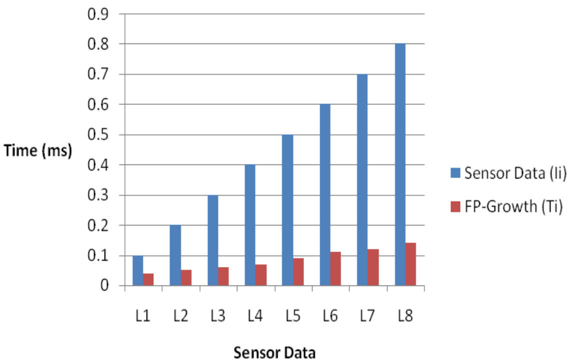


*Figure 1 Execution time analysis-FP growth analysis*

From the above graph indicated in Figure 1, it is analyzed that the execution time of FP-Growth using light sensor is reduced by 36% on the average, compared to individual sensor data in mining frequent itemsets.

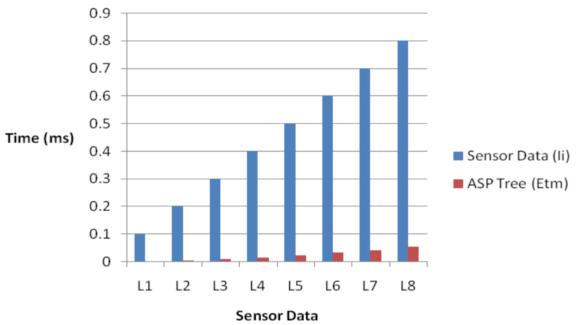


*Figure 2 Execution time analysis-ASP tree analysis*

From the above graph depicted in Figure 2, it is inferred that the execution time of ASP-tree, on the average, is reduced by 42.5% compared to individual sensor data from light sensor without tree formation in mining frequent itemsets.

## 4. CONCLUSION AND FUTURE WORK

The proposed system introduced a new type of behavioral patterns called associated sensor patterns that capture co-occurrences. The proposed idea generates temporal correlations among sensors. To extract such patterns, a prefix tree structure called ASP-tree that stores sensor data in a compact manner is devised. Based upon this tree, a mining algorithm called ASP is proposed which effectively mines associated sensor patterns over sensor database in only one scan. This algorithm is used to predict frequently occurring item details in the transaction dataset and time complexity rate for tree construction was calculated. The tree is dynamically updated as per user needs. It also helps to reduce the execution time and transfer the data quickly to destination node.

Future work will be based on giving weights to items in mining sensor data. This can be used to mine secured sensor items in addition with minimum support count items. In addition, tree will be constructed for both frequent and infrequent moved items in dataset.

## REFERENCES

[1] J. Chandrika and K. R. A. Kumar,(2012) 'Resource adaptive technique for frequent itemset mining in transactional data streams,'Int. J. Comput. Sci. Net. Security, vol. 12, pp. 87–92,

[2] Chandrika.J and K.R.Ananda Kumar (2012) 'Mining Frequent itemsets from transactional data streams'.

[3] M. M. Rashid, I. Gondal and J. Kamruzzaman, (2013 )'Mining associated sensor patterns for data stream of wireless sensor networks,' in Proc. 8th ACM Workshop Perform. Monitoring Meas. Heterogeneous Wireless Wired Netw., , pp. 91–98

[4] X. J. Ban and M. Gruteser (2012) 'Towards fine-grained urban traffic knowledge extraction using mobile sensing,' in Proc. ACM SIGKDD Int. Workshop Urban C Comput., pp. 111–117.

[5] Sandy Moens ,Bart Goethals (2013) 'Frequent Itemset Mining for Big Data',2013.

[6] A. Boukerche and S. Samarah (2008), 'Novel algorithm for mining association rules in wireless ad-hoc sensor networks,' IEEE Trans. Parallel Distrib. Syst., vol. 19, no. 7, pp. 865–877.

[7] S. K. Tanbeer, C. F. Ahmed and B. S. Jeong (2009), 'An efficient singlepass algorithm for mining association rules from WSN,' IETE Tech. vol. 26, p. 280.

[8] S. K. Tanbeer, C. F. Ahmed, B. S. Jeong, and Y. K. Lee, (2009. ) 'Sliding window- based frequent pattern mining over data streams,'Inf. Sci., vol. 179, pp. 3843–3865.

[9] H. F. Li and S.-Y. Lee, (2009 ) 'Mining frequent itemsets over datastreams using eefficient window sliding techniques,"'Expert Syst. Appl., vol. 36, pp. 1466–1477.

[10] S. Samarah, B. Azzedine, and S. H. Alexander, (2011.)'Target association rules:A new behavioral patterns for point of coverage wireless sensor networks,"'IEEE Trans. Comput., vol. 60, no. 6, pp. 879–889, Jun.

### Authors Biography

**Dr. T. Abirami** received B.Sc., degree in Computer Technology from Kongu Engineering College, Bharathiyar University, Coimbatore, in 2000; Master of Computer Applications degree from Kongu Engineering College in Anna University, Chennai in 2003 and M.E., degree in Computer Science from Kongu Engineering College in

ISSN: 2456 - 3935
International Journal of Advances in Computer and Electronics Engineering
Volume: 02 Issue: 03, March 2017, pp. 28 – 32

Anna University, Chennai in 2009. She completed her Ph.D in Wireless Sensor Networks from Anna University, Chennai. Presently, she is working as an Assistant Professor (Senior Grade), in the department of Information Technology at Kongu Engineering College, Perundurai, Erode, Tamilnadu, India. She has been in the teaching profession for the past 12 years. She is awarded with a UGC Major Research Project entitled "A Weighted Association Rule Mining Algorithm for Energy Conservation in Wireless Sensor Networks" for the period from the year 2010 to 2014 and also TNSCST student project. She is a life member of CSI. She has received "Best Faculty" award for the year 2009-10. She has Published 11 papers in International Journals and 20 Papers in International and National Conferences. Her academic interests include Wireless Sensor Networks, Data Mining and Mobile Application Development. She has organized 15 national level seminar, workshop and conferences for the benefit of faculty members and students. She has attended 30 Seminars, FDP's, and Workshops organized by various Engineering colleges.

**S. Anitha** received B.E Degree in Electronics and Communication Engineering from Coimbatore Institute of Engineering and Information Technology, Coimbatore in 2006 and M.E Degree in Computer Science and Engineering from Kongu Engineering College in 2009. From 2009 to 2010 she worked as a Lecturer in the department of IT, Velalar College of Engineering and Technology. Currently she is working as an Assistant Professor in the Department of IT, Kongu Engineering College, Perundurai. Currently, she was sanctioned with a UGC Minor Research Project entitled, "Pattern improvement in Rank-based Association Rule Mining for Energy Conservation in Wireless Sensor Networks" She is a life member of ISTE and CSI. She has published 8 papers in National and International Journals and 5 papers in National and International Conferences. Her academic interests include Wireless Sensor Networks and its security, Computer Graphics and Multimedia. She has organized 5 national level seminar, workshop and conferences for the benefit of faculty members and students. She has attended 12 Seminars, FDP's, and Workshops organized in and around the college.