# An Efficient Mining Behavioral Pattern using Associated Correlated Bit Vector Matrix for in Wireless Sensor Network

Kaki Hemasai[1], K. Jagdeeshwara Rao [2]

*Final M.Sc. Student[1], Lecturer[2]*
*[1, 2] M. Sc Computer Science, Chaitanya Women's PG College, Old Gajuwaka, Visakhapatnam Andhra Pradesh*

***Abstract:***
*Now a day's wireless sensor network interesting research area for discovering behavioural patterns WSNs can be used for predicting the source of future events. By knowing the source of future event, we can detect the faulty nodes easily from the network. Behavioural patterns also can identify a set of temporally correlated sensors. This knowledge can be helpful to overcome the undesirable effects (e.g., missed reading) of the unreliable wireless communications. It may be also useful in resource management process by deciding which nodes can be switched safely to a sleep mode without affecting the coverage of the network. Association rule mining is the one of the most useful technique for finding behavioural patterns from wireless sensor network. Data mining techniques have recent years received a great deal of attention to extract interesting behavioural patterns from sensors data stream. One of the techniques for data mining is tree structure for mining behavioural patterns from wireless sensor network. By implementing the tree structure will face the problem of time taking for finding frequent patterns. By overcome that problem we are implementing associated correlated bit vector matrix for finding behavioural patterns of nodes in a wireless sensor network. By implementing this concept we can overcome time complexity and also get most correlated patterns of wireless sensor networks.*

**Keywords:** *Data Mining, Association rule Mining, Wireless Sensor Network, Behaviour Pattern, Associated Correlated Frequent Patterns.*

## I. INTRODUCTION

Wireless Sensor Network generates 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 , 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 behavioural 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 we are proposed an efficient associated correlated bit vector matrix for find the frequent item sets and associate correlated frequent pattern sets. The Data Mining in WSN are used to extract useful data from the huge amount of unwanted dataset. The need of mining to get knowledgeable data and discovers the behavioural patterns. As there are many Association techniques in data mining to find out the Frequent Patterns as per . The Association rule can apply on static data and stream data. The frequent patterns are those items, Sequences or substructure which reprise from the available dataset by providing the user specified frequencies. Whenever you want to find out the frequently occurred data apply association rules which will find out the frequent patterns from the dataset.

Mining play main role to mine frequent item set in many data mining tasks. Over data streams, the frequent item set mining is mine the approximation set of frequent item sets 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.

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 behavioural patterns of sensor nodes evolved from meta-data describing sensor behaviours. The application of fine grain monitoring of physical environments can be highly benefitted from discovering behavioural patterns (i.e., associated patterns) in WSNs. These behavioural 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 behavioural pattern mining by predicting the occurrence of an event from a particular node, but no such event reported in subsequent iteration. As behavioural 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, behavioural patterns can also use to identify a set of temporally correlated sensors, thus improving operational aspects in WSNs.

## II. RELATED WORK

Frequent pattern mining has been an important subject matter in data mining from many years. A remarkable progress in this field has been made and lots of efficient algorithms have been designed to search frequent patterns in a transactional database. Agrawal et al. (1993) firstly proposed pattern mining concept in form of market based analysis for finding association between items bought in a market. This concept used transactional databases and other data repositories in order to extract association's casual structures, interesting correlations or frequent patterns among set of. Frequent patterns are those items, sequences or substructures that reprise in database transactions with a user specified frequency. An item set with frequency greater than or equal to minimum threshold will be considered as a frequent pattern. For example in market based analysis if the minimum threshold is 30% and bread appears with eggs and milk more than three times or at least three times then it will be a frequent item set. Frequent pattern mining can be used in a variety of real world applications. It can be used in super markets for selling, product placement on shelves, for promotion rules and in text searching. It can be used in wireless sensor networks especially in smart homes with sensors attached on Human Body or home usage objects and other applications that require monitoring of user environment carefully that are subject to critical conditions or hazards such as gas leak, fire and explosion .These frequent patterns can be used to monitor the activities for dementia patients. It can be seen as an important approach with the ability to monitor activities of daily life in smart environment for tracking functional decline among dementia patients .

In 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 item sets 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. A vision for mining fine-grained urban traffic knowledge from mobile sensing, 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 modelling techniques, thereby satisfying privacy preserving and traffic modelling 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 modelling is an integration of traffic principles and learning/optimization techniques.). Frequent Item set 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 in spite of its technical challenges like balanced data distribution and inter-communication costs. Dist-Eclat and BigFIM focusing on speed and optimized to run on large datasets respectively.

In 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 use-ful or unsuccessful in reflecting true correlation among sensor data.

## III.PROPOSED SYSTEM

Mining play main role to mine frequent item set in many data mining tasks. Over data streams, the frequent item set mining is mine the approximation set of frequent item sets 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.

In this paper we are proposed an efficient correlated association rule mining for mining behavioural patterns from wireless sensor network. For mining association correlated patterns cab be done by performing the two steps. In the first step we are finding frequent patterns of wireless sensor networks and second step is to test whether they are associated correlated patterns or not based on the corr confidence of each pattern in a transaction dataset. By performing those two operations we are implementing an efficient correlated bit vector matrix for finding behavioural patterns from a wireless sensor network. The implementation procedure of associated correlated bit vector matrix is as follows.

### *A. Associated Correlated Bit Vector Matrix:*

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 associated correlated bit vector matrix which is used to generate associated patterns. The generation frequent patterns of sensor stream of data is as follows.

Generation of Bit Vector Matrix:

In this module we can retrieve the transactional data set of sensor items from the data base. Take the each transaction and generate bit vector matrix. The implementation of bit vector matrix is as follows.

1. Read each transaction from the data base (D) and get each item of sensor node id $S_i$ .

2. Read all the individual items of sensor nodes until the length of all transactional dataset is completed.

3. After completion of reading process we can sort the all node ids.

4. Find all frequent length of item sets($T_i$) from the data base D

```
If Ti is not null
 For each transaction (Ti) from database
   For each item (Ii) in database D
     If item (Ii) contains Transaction Item sets (Ti)
          BV = 1
              Else
              BV=0
End for.
End for.
End if.
```

### *B. Extracting Maximum Frequent Item Sets from Bit Vector Matrix:*

After completion of bit vector matrix we can find out frequent pattern item sets of wireless sensor network. Each column in the bit vector matrix represents one transaction record. Value 0 in the column means the corresponding transaction record contains the corresponding frequent length-1 item set, vice versa. Therefore, the number of value 1 in each column indicates the corresponding transaction record contains the number of frequent length-1 item sets together. If there is the number of transaction records with the same number of value 1 being larger than the minimum support, the number of value 1 may be the size of maximum frequent item set, vice versa. As a result, a set of values in which each one may be maximum frequent item set's length will be obtained. Then according to each of the values in descending order, a series of candidate item sets will be generated from frequent length-1 item sets and the support of each candidate item set could be calculated according to the Boolean matrix of frequent length-1 item sets. If the support of each candidate item set is larger than the minimum support, the candidate item set is frequent, vice versa.

At last, if the maximum frequent item sets generated from the set of candidate item sets are not empty, the size of candidate item set is required, that is length of maximum frequent item set. Otherwise, it is necessary to continue the previous operation to check the next value until maximum frequent item sets are not empty. If all the maximum frequent item sets are empty, the maximum length of frequent item set is one.

Input: the bit vector matrix, Minimum support value
Output: Maximum Frequent patterns of sensor nodes

Process:
For each column in the bit vector matrix
Calculate number of value one in the current row
End for
Return max[n]
Sort (Max[n])
For each one in the max[n]
Calculate number of columns with the same number of ones

If number> minimum support value
Generate maximum number of candidate item sets from transaction

For each item set in candidate item sets
Calculate support (item set)

If(support(item set)>minimum support count)

Item set is frequent

End if
End for
End if

If maximum item sets is not null
Break;
End if
End for.

All the frequent item sets could be extracted from all the maximum frequent item sets according to the nonempty subsets of frequent item sets being still frequent. And the support of each frequent item set could be calculated, all the strong association rules can be mined from all the frequent item sets.

### *C. Associated Correlated Frequent Patterns:*

A pattern is called associated pattern if its all confidence is greater than equal to given minimum support value. In statistical theory $s_1,s_2 \ldots s_p$ independent nodes in wireless sensor network. The confidence of each sensor node can be calculated by the support of each node in the network. We can calculate the corr confidence more than one node and all those confidence values is greater than equal to minimum support value, take those pattern as associated correlated frequent pattern. By calculating confidence of two sensor, such as $s_1$ , $s_2$ is as follows.

$$\rho(s_1,s_2) = P(s_1,s_2)-P(s_1)P(s_2)/ P(s_1,s_2)+P(s_1)P(s_2)$$

Suppose we are calculating more than two nodes of confidence we are using the following formula.

$$\mathbf{P}=P(s_1,s_2 \ldots s_n)-P(s_1)P(s_2)\ldots P(s_n)/P(s_1,s_2 \ldots s_n)+P(s_1)P(s_2)\ldots P(s_n)$$

After calculating confidence take those values and compare with minimum support values. The confidence of each pattern is greater than equal minimum support value take those patterns as correlated frequent patterns. This process will apply the length of all frequent patterns are completed. By performing this process we can get efficient correlated frequent patterns from the wireless sensor network.

## IV. CONCLUSIONS

In this paper we are proposed an efficient association rule mining finding associated correlated frequent behavioral patterns from wireless sensor network. Our proposed associated correlated bit vector matrix for mining behavioral frequent patterns of wireless sensor network data. By implementing this process we can scan the entire data once and mine many properties is suitable for interactive mining. An extensive analysis of associated correlated bit vector matrix is finding associated frequent patterns mining and out performs the existing algorithm based on execution time and memory usage.

## REFERENCES

[1] 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
[2] Jiinlong, Xu Conglfu, Cben Weidong, Pan Yunhe," Survey of the Study on Frequent Pattern Mining in Data Streams", 2004 IEEE International Conference on Systems, Man and Cybernetics.
[3] R.Agrawal and R. Srikant, "Fast algorithms for Mining Association Rules", 20th International Conference on Very Large Data Base, pp. 487–499, May1994.
[4] Md. Mamunur Rashid, Iqbal Gondal and Joarder Kamruzzaman, "Share-Frequent Sensor Patterns Mining from Wireless Sensor Network Data", IEEE Transaction Parallel Distribution System, 2014.
[5] Imielienskin T. and Swami A. Agrawal R., "Mining Association Rules Between set of items in large databases," in Management of Data, 1993, p. 9.
[6] H.Mannila, R. Srikant, H. Toivonen, and A. Inkeri R. Agrawal, "Fast Discovery of Association Rules," in Advances in Knowledge Discovery and Data Mining, 1996, pp. 307-328.
[7] M.Chen, and P.S. Yu J.S. Park, "An Effective Hash Based Algorithm for Mining Association Rules," in ACM SIGMOD Int'l Conf. Management of Data, May, 1995.
[8] R.Motwani, J.D. Ullman, and S. Tsur S. Brin, "Dynamic Item set Counting And Implication Rules For Market Basket Data," ACM SIGMOD, International Conference on Management of Data, vol. 26, no. 2, pp. 55–264, 1997.