Volume No:4, Issue No:7 (December-2018) ISSN No : 2454-423X (Online)

International Journal of Research in Advanced Computer Science Engineering

A Peer Reviewed Open Access International Journal

www.ijracse.com

An Efficient Associated Correlated Bit Vector Matrix for Mining Behavioral Patterns from Wireless Sensor Network

K Ramesh
Department of Computer Science and Engineering,
Sarada Institute of Science, Technology and Management (SISTAM),
Srikakulam, Andhra Pradesh 532404, India.

S Kalyan
Department of Computer Science and Engineering,
Sarada Institute of Science, Technology and Management (SISTAM),
Srikakulam, Andhra Pradesh 532404, India.

ABSTRACT

Now a day's wireless sensor network interesting research area for discovering behavioral patterns wireless sensor network 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. Behavioral 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 behavioral patterns from wireless sensor network. Data mining techniques have recent years received a great deal of attention to extract interesting behavioral patterns from sensors data stream. One of the techniques for data mining is tree structure for mining behavioral 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 behavioral patterns of nodes in a wireless sensor network. By implementing this concept

Key Words: ***wireless sensor network, future event, Data mining techniques, frequent patterns, bit vector matrix.***

INTRODUCTION

Wireless Sensor Network generates [1] a large amount of data in the form of data stream andmining these streams to extract useful knowledge is a highly challenging task. 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 fromsensor stream data is essential for real-time applications, In this proposed work, a new type of sensor behavioral pattern [2] 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 inWSN 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 [3]. 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

Cite this article as: K Ramesh & S Kalyan, "An Efficient Associated Correlated Bit Vector Matrix for Mining Behavioral Patterns from Wireless Sensor Network", International Journal of Research in Advanced Computer Science Engineering, Volume 4 Issue 7, 2018, Page 20-26.

Volume No:4, Issue No:7 (December-2018) ISSN No : 2454-423X (Online)
International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

flexible determine between mining accuracy and rocessingtime. 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) [4]. Using knowledge discovery in WSNs, one particular interest is to find behavioural patterns of sensor nodes evolved meta-data describing sensor behaviours. The application of fine grain monitoring of physical environments can be highly benefitted from discovering behavioural patterns (i.e., associatedpatterns) 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 otherProcesses. In addition, behavioural patterns can also use to identify a set of temporally correlated sensors, thus improving operational aspects in WSNs.

EXISTING SYSTEM

However, association rule mining with real datasets is not so simple. If the minimum support threshold is high, then we can get high value knowledge. On the other hand, when the minimum support threshold is low, an extremely large number of association rules will be generated, most of the mare non-informative. The valid correlation relationships among data objects are buried deep among a large pile of useless rules. Additionally, association rules mining are not able to discover such kind of patterns where the event detected by sensor s1 can increase the likelihood of the event detect by s2. Such kind of patterns are both associated and correlated. To overcome this difficulty, here we combine association and correlation in the mining process to find the associated correlated patterns. To generate associated-correlated patterns that have a certain frequency (support) it is required to generate all the patterns present in the data, i.e., frequent patterns. Once the frequent patterns are determined, the process of generating the associated-correlated patterns is then straight forward. Therefore, devising an efficient algorithm to mine frequent pattern with high-speed in large-scale sensor data has been the real challenge.

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 [5] 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 behavioral patterns from wireless sensor network. For mining association

Volume No:4, Issue No:7 (December-2018) ISSN No : 2454-423X (Online)
International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

correlated patterns can 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 confidence of each pattern in a transaction dataset. By performing those two operations we are implementing an efficient correlated bit vector matrix for finding behavioral patterns from a wireless sensor network. The implementation procedure of associated correlated bit vector matrix is as follows.

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.

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

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

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

Find all frequent length of item sets (Ti) 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.
```

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 off requent 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

Volume No:4, Issue No:7 (December-2018) ISSN No : 2454-423X (Online)
International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

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 [6].

Requirement Analysis

A Software Requirements Specification (SRS) is a complete description of the behavior of the system to be developed. It includes a set of use cases that describe all the interactions the users will have with the software. Use cases are also known as functional requirements. In addition to use cases, the SRS also contains non-functional (or supplementary) requirements. Non-functional requirements are requirements which impose constraints on the design or implementation (such as performance engineering requirements, quality standards, or design constraints).

Hardware Requirements

1. VDU: Monitor/ LCD TFT / Projector
2. Input Devices: Keyboard and Mouse
3. RAM: 512 MB
4. Processor: P4 or above
5. Storage: 10 to 100 MB of HDD space.

Software Requirements

1. Operating System: Any Operating System
2. Run-Time: OS Compatible JVM

System Design

The Unified Modeling Language (UML) is a general-purpose, developmental, modeling language in the field of that is intended to provide a standard way to visualize the design of a system.The Unified Modeling Language (UML) [7] offers a way to visualize a system'sarchitectural blueprints in a diagram (see image), including elements such as:

- Any activity
- Individual component of the system
- And how they can interact with the other components
- How the system will run
- How entities interact with others (components and interfaces)
- External user interface

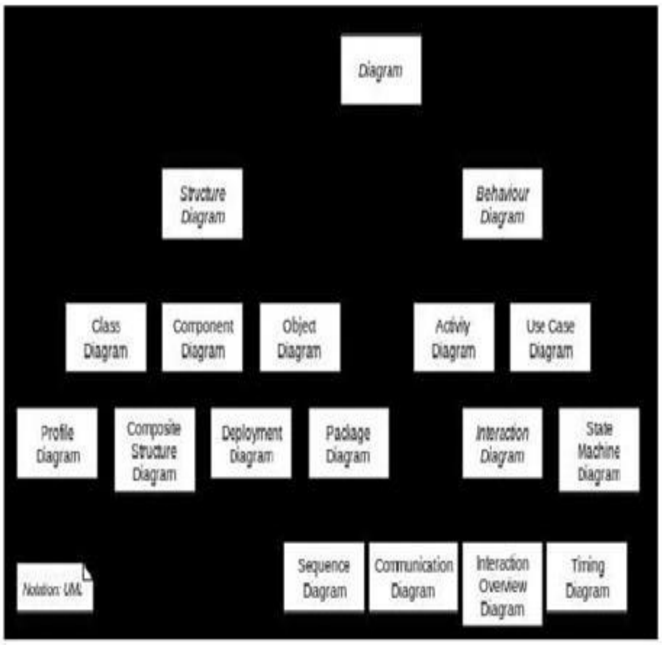


Fig 1 Block diagram of various UML diagrams

Volume No:4, Issue No:7 (December-2018) ISSN No : 2454-423X (Online)

International Journal of Research in Advanced Computer Science Engineering

A Peer Reviewed Open Access International Journal

www.ijracse.com

USE CASE DIAGRAM

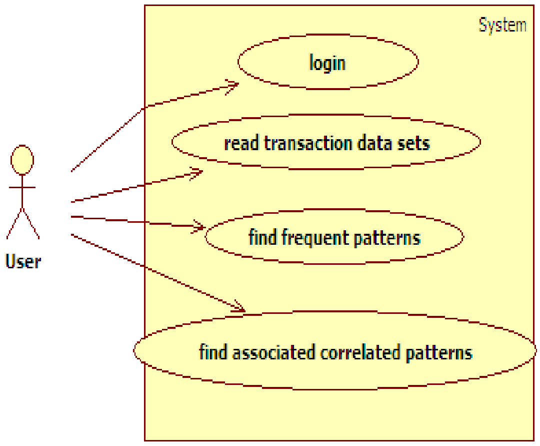


Fig 2 Use Case Diagram

Sequence Diagram

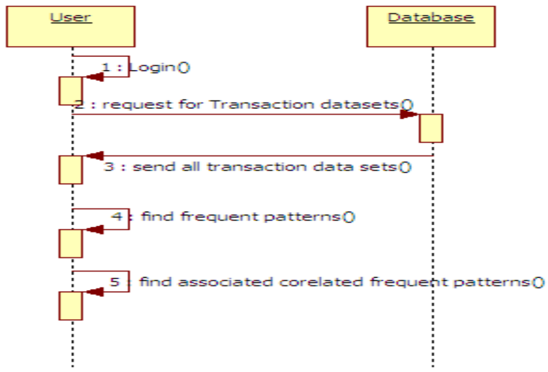


Fig 3 Sequence Diagram

State Chart Diagram

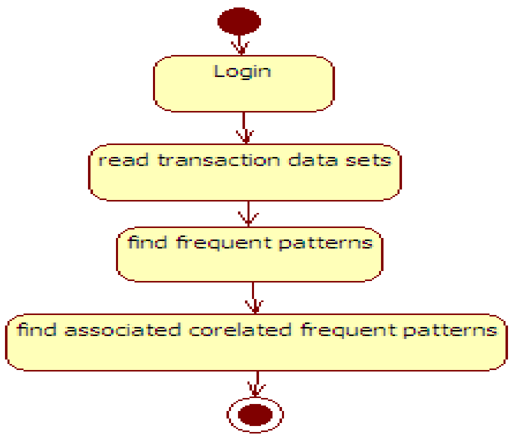


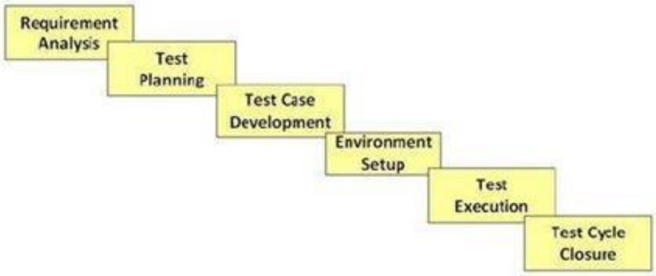
Fig 4 State Chart Diagram

TESTING

SOFTWARE TEST LIFE CYCLE:

Every line of code that is written has to be tested thoroughly to check that the requirement is met with correct logic and its bug free. Testing phase has to be followed in a proper life cycle so that entire code is tested and no test cases are missed out.

The different stages in Software Test Life Cycle (STLC): -



Requirement Analysis

During this phase, test team studies the requirements from a testing point of view toidentify the testable requirements. The QA team interact with various stakeholdersto understand the requirements in detail. Requirements could be either functional or nonfunctional automation feasibility for the given testing project is also done in this stage

Activities:

Identify types of tests to be performed.

Gather details about testing priorities and focus.

Prepare Requirement Traceability Matrix (RTM) [8].

Identify test environment details where testing is supposed to be Carried out

TEST PLANNING

This phase is also called Test Strategy phase. Typically, in this stage, a Senior QAmanager will determine effort and cost estimates for the project and would prepare andfinalize the Test Plan

Activities

Preparation of test plan/strategy document for various types of Testing

Test tool selection

Volume No:4, Issue No:7 (December-2018) ISSN No : 2454-423X (Online)

International Journal of Research in Advanced Computer Science Engineering
A Peer Reviewed Open Access International Journal
www.ijracse.com

Test effort estimation

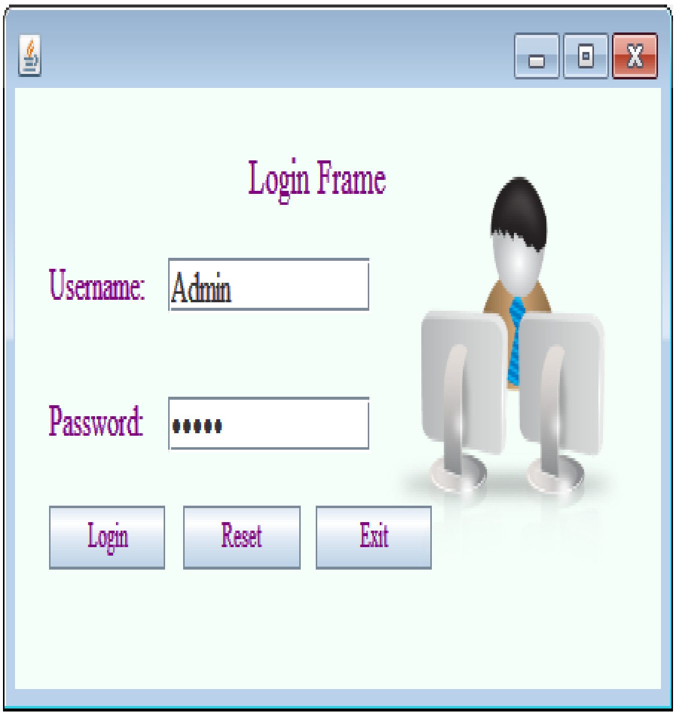
Resource planning and determining roles and responsibilities.

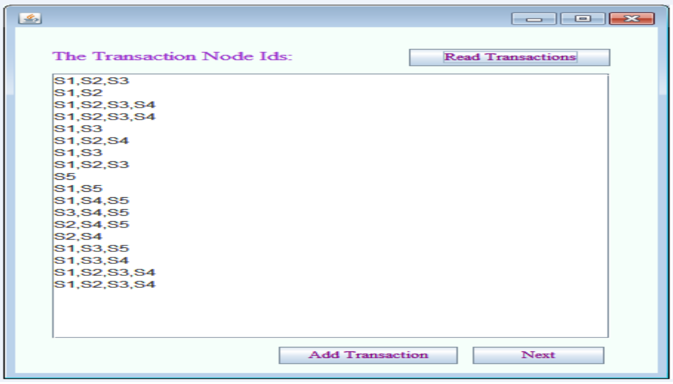
Training requirement

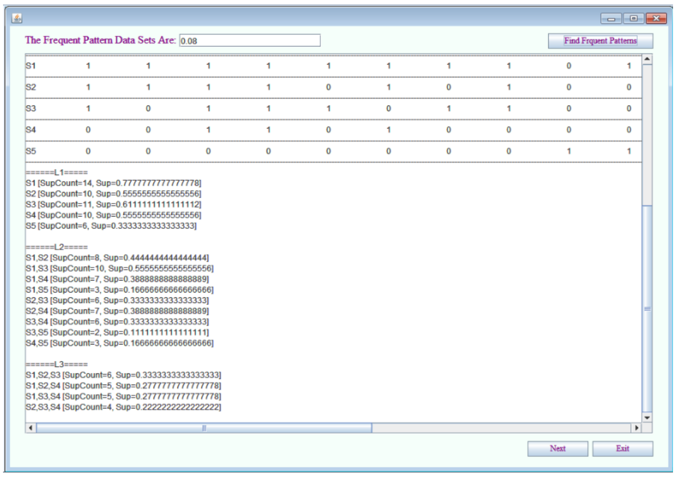
Grey Box Testing

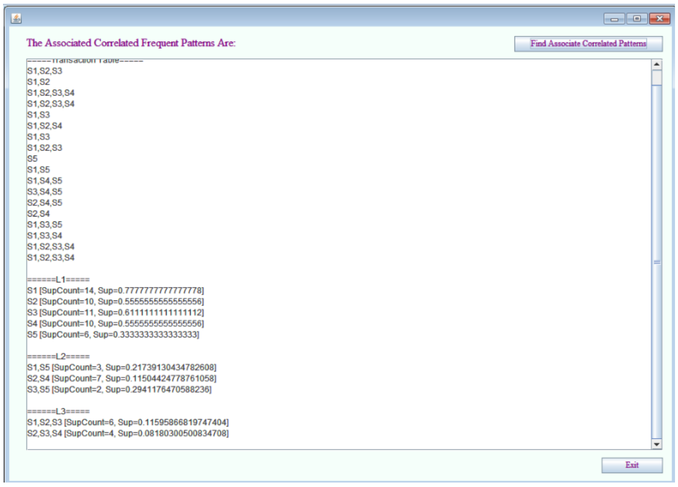
Grey box testing (American spelling: gray box testing) involves havingknowledge of internal data structures and algorithms for purposes of designing the testcases, but testing at the user, or black-box level. Manipulating input data and formattingoutput do not qualify as greybox, because the input and output are clearly outside of the"black-box" that we are calling the system under test. This distinction is particularlyimportant when conducting integration testing between two modules of code written bytwo different developers, where only the interfaces are exposed for test. However,modifying a data repository does qualify as grey box, as the user would not normally beable to change the data outside of the system under test. Grey box testing may also include reverse engineering to determine, for instance, boundary values or error.

RESULTS









CONCLUSION

In this paper we are proposed an efficient association rule mining findingassociated correlated frequent behavioral patterns from wireless sensor network. Ourproposed associated correlated bit vector matrix for mining behavioral frequent patternsof wireless sensor network data. By implementing this process we can scan the entiredata once and mine many properties is suitable for interactive mining. An extensiveanalysis of associated correlated bit vector matrix is finding associated frequent patternsmining and out performs the existing algorithm based on execution time and memoryusage.

REFERENCES

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

[2]. Jiinlong, XuConglfu, CbenWeidong, 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, IqbalGondal 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 59Association 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 Itemset Counting And Implication Rules For Market Basket Data," ACM SIGMOD, International Conference on Management of Data, vol. 26, no. 2, pp. 55–264, 1997.