

# Implementing Data Mining Techniques for Smart City Development

Vishakha Mehta <sup>1</sup>, Sarika Khandelwal <sup>2</sup>

*M.Tech. Computer Science Engineering, Geetanjali Institute of  
Technical Studies, Udaipur (India)*

vishakha19mehta@gmail.com,

sarikakhandelwal@gmail.com

**ABSTRACT** - The perception of smart city is increasing rapidly all over the world. There are various technologies that are able to transform a normal city into a smart city. The idea of smart city incorporate the ideology of smart buildings. As there has been considerable research on developing smart environment by blending data mining techniques into smart environments which are assembled with sensors and networks. With the idea of smart environment, our main aim is to reduce the energy consumption by maintaining the comfort level as well. In this study, our aim is to establish and implement techniques like data mining and sensor networks for smart building as to save energy.

## I. INTRODUCTION

A Smart City can be defined as a city assembled with basic infrastructure and latest technologies to provide a premium life style and sustainable environment with the application of smart ideas and technology. A vision of Smart City is to provide more global, protected and adequate services to the citizens and establishing more livable and sustain city. Since last few decades more people are moving to urban areas and because of this, the cities or urban areas have to face more economic and environmental challenges. In order to maintain the condition, efforts are required to accommodate enlarging urban population and their respective challenges. To overcome this situation, government can promote modern technologies to build "Smart City". To develop a city smartly, we need to construct smart buildings first. Now what is a smart building? A smart building consist of features like smart meters and their management, intelligent traffic system, smart sensors to reduce the consumption of energy, fire controlling etc. To develop smart building and working with sensors. Our first step should be research of the techniques that can be useful in the development. For collecting the data and managing this huge amount of data, we should concentrate about the techniques that can work with the data collection from these sensors.

In this study our objective is focusing on the research on developing smart environments and smart building by integrating data mining techniques that are combined with sensors. As for developing smart environment and smart building, we need to focus on energy consumption in building but also managing the satisfaction and comfort of the citizens as well.

## II. LITERATURE REVIEW

Data Mining can be described as a technique of interpret data from data warehouse and summarizing it and make it a useful information. A data warehouse is a system where data is integrated from multiple sources. Data warehouse stored a large number of data for useful analysis and extracting meaningful information from the data.

Data mining techniques implement analyzing of data from different perspective and make data in a categorized manner. On a technical term, data mining is a process of finding patterns and interrelation amidst various number of fields in large database. As the use of internet is increasing rapidly, the size of data is also increasing exponentially. And to extract the useful data from such a large amount of database, data mining technique is very much useful. And For such needs, data mining techniques has recently attracted attention of database users. The main goal of data mining process is analyzing the data and extracting meaningful information from database.

In data mining, there are mainly four problems such as clustering, classification, data analysis and the last one is frequent pattern mining. Among the four major issues in data mining, frequent pattern mining is widely used and studied because of the diverse applications and wide range of data mining problems like clustering and classifications, super market problems, biological data etc. This method can also be implemented by using wireless sensor networks especially in smart building with sensors and some other applications also that needs controlling of environment attached with sensors. It plays important role in association rule mining also.

The main aim of Frequent Pattern Mining problem is searching of frequent pattern, group or structures that appear frequently in database. A frequent pattern can be described as a pattern or set of items that occur periodically.

Agrawal et al. (1993) firstly proposed frequent pattern mining concept and implemented in form of market basket analysis for finding association between items bought in a market. In this algorithm a minimum threshold is considered as a frequent pattern.

Frequent pattern mining is applied to scan the database and extract the similar patterns throughout the scanning of the database. One way of doing this technique is by creating graphical trees of pattern and sub pattern. Firstly, a database is scanned and identify all the structures in first pass and these structures or patterns are called "candidate". And these candidates create a first hierarchical level of a graphical tree. Same thing is done on another passes and candidates are extracted and then the mining of such patterns are done with the help of different techniques used in pattern mining.

The performance of this technique depends on many factors but the major one is the characteristics of databases that was scanned. And item set that is generated after the scanning of database with the frequency greater than or equal to minimum threshold will be taken as a frequent pattern.

Frequent pattern mining technique use tree structure to store the data. A frequent pattern tree (FP Tree) is a structure that is used to store information about the patterns in a database.

The basic frequent pattern mining algorithms are classified in two ways:

1. Candidate Generation
2. Without Candidate Generation

In the first category the algorithms used are Apriori algorithm, AprioriTID, Apriori Hybrid, in which they require candidate generation for mining the frequent pattern and the second one is FP – Growth Algorithm in which candidate keys are not required. In FP Growth algorithm the method is quite different from other earlier used method. In this algorithm it works on generating a structure which is called as FP-Tree with the twice scanning of database.

### III. COMPARATIVE STUDY OF FREQUENT MINING PATTERN TECHNIQUES

Frequent pattern mining is an important techniques used in mining of data. As because of its various applications, it becomes the most widely used techniques among any other techniques in the field of data mining. As there has been a lot of research on this technique in the field of storing a large amount of data and analyzing and extracting some meaningful information from that data, many algorithms are proposed for the doing the same. To make a valid and informative comparison of the techniques let us go through some of the techniques used for data mining and frequent mining pattern.

#### A. Apriori Algorithm

Apriori algorithm is proposed by Agrawal and Srikant in 1994. It is one of the earliest technique used for data mining. As the name suggest Apriori Algorithm is based on the property of Apriori. In this algorithm two processes are executed. One is candidate generation process. In this process, a count of the item set is calculated by scanning the database and then after the scanning is done, large item set are generated by the scanned candidate. Item set which are having a count less than minimum threshold, and this process is repeated till we get large item set. Despite of this algorithm is one of the earliest techniques, it has some disadvantages also. First one is, it generate very complex item set in the candidate generation process as a result it consume large memory and more time. And the second one is the database is scanned many times for candidate generation.

#### B. Rapid Association Rule Mining (RARM)

In this algorithm, it uses a different type of data structure to store the candidate. And so it avoid the complex generation of candidates. It uses a novel tree data structure, known as **Support-Ordered Trie Itemset (SOTrieIT)** and it stores data before any processing or scanning of database. As because of this method, RARM quickly generate large 1-itemset and 2-itemset without scanning of database and also without generating candidate 2-itemset. SOTrieIT is a sorted TrieIT, in this tree nodes are sorted according to their support count. To making SOTrieIT tree structure, the extraction of 1-itemset and 2-itemset are done form each and every transaction. Support count are distinguished by higher and lower support count. Higher ones are stored to the left node of tree and lower are stored to the right node. Performance of RARM is improved as it uses the SOTrieIT by generating large itemset.

#### C. Equivalence Class Transformation (ECLAT)

Apart from the algorithms we discussed above, ECLAT algorithm used vertical data format to store the data. In horizontal data format (TransactionID, Items) are used but in vertical data format (Items, TransactionId) items are stored with the list of transactions. ECLAT algorithm used Depth-First Search Algorithm to traverse in its data structure. Initially in the scanning process of database, list of transaction IDs are maintained for every single item. From the k itemset, k+1 itemset are generated using Apriori method and search techniques used is depth first search. The process of generating these item set continues till no candidate itemset can be found.

#### D. Frequent Pattern (FP) Growth Algorithm

In data mining, Frequent Pattern Growth algorithm is most popular and widely used algorithm. As it overcome the major drawbacks of other data mining techniques. FP Growth algorithm uses a very compressed data structure named as FP-Tree. As is uses this compressed data structure it is different and unique from the other techniques. FP-Growth algorithm is a two-step approach. In the first step database is scanned and frequent pattern tree is generated. In this method, database is scanned two times. In first scanning of database, the support count for every itemset is calculated. The patterns that are useful are stored and that are not are deleted. In the second step, by collecting all the frequent pattern, FP-Tree is prepared and then using FP-Growth algorithm, from the generated FP-Tree, frequent pattern are extracted. There are major three advantages of FP-Growth

algorithm. First one is, in this method database is scanned only 2 times by which the time used in computation is reduced. Secondly, candidates are not generated in this algorithm and the last one is it uses divide and conquer approach and resultant the search space is also reduced.

#### E. Associated sensor pattern mining of data stream (ASPMS)

This technique is new in the field of data mining and unlike other algorithms, this method is used to find frequent pattern from the sensor and other wireless networks. This algorithm also uses a tree structure and it named as associated sensor pattern stream tree (ASPS-tree). In this algorithm database is scanned only once. In this technique, sliding window method is used for pattern mining. The useful information is captured from the stream content for the current window of the sensor in one by one step. The information are stored in form of nodes in an ASPS Tree. Firstly data is stored in order of appearance of sensor and then it re-construct the tree in order of frequency. And then the tree is compressed and merging of same support sensor are done. ASPMS has two steps of mining the data. First the tree is constructed and then sensor pattern are generated and then the mining of these patterns are done. As this method is based on sliding window technique. Initially, the database is divided into equal window size. In which each window have equal number of transactions. Merge Sort is used for sorting the information in descending order and the re-structuring of tree is done. If two nodes in the tree have same support count then these 2 nodes are merged into single node and Branch Sort Method is used for this method and then the last step is to extract the frequent sensor pattern from ASPS Tree.

#### IV. CONCLUSION

To convert a city into a smart city, we have to adapt new technologies and can develop a “Smart City” by providing the basic infrastructure combined with latest technologies. With the idea of smart city, our aim will be to reduce the energy consumption and developing a smart environment. To make this thing happen the use of technology is a must. And by using the technique of sensor network with data mining, we can achieve our goal. In this research paper we found that a data mining technique named **Associated sensor pattern mining of data stream (ASPMS)** can be very useful as it works with sensors and networks, and by implementing such techniques, our goal of Smart City can be achieved successfully.

#### REFERENCES

- [1] R. Agarwal and R. Srikant, —Fast Algorithms for Mining Association Rules.I, International Conference on very large Databases, proc.20th , pp 487-499, june 1994.
- [2] Sumit Aggarwal and Vinay Singal, "A Survey on Frequent pattern mining Algorithms.", International Journal of Engineering Research & Technology (IJERT), ISSN: 2278-0181, Vol. 3 Issue 4, pp 2606-2608, April 2014.
- [3] J. Wang, J. Han, and J. Pei, — Searching for the Best Strategies for Mining Frequent Closed Itemsets.I, International conference on Knowledge Discovery and Data Mining.
- [4] Srinivasan Parthasarathy, and Wei Li Mohammed Javeed Zaki, "A Localized Algorithm for Parallel Association Mining," in *In 9th ACM Symp. Parallel Algorithms & Architectures*.
- [5] R. Agrawal, T. Imieliński, and A. Swami. Mining Association Rules between Sets of Items in Large Databases. *Proc. Conf. on Management of Data*, 207–216. ACM Press, New York, NY, USA
- [6] J. Han, H. Pei, and Y. Yin. Mining Frequent Patterns without Candidate Generation. In: *Proc. Conf. on the Management of Data (SIGMOD '00, Dallas, TX)*. ACM Press, New York, NY, USA
- [7] Borgelt C., "Efficient Implementations of Apriori and Eclat," in 1st IEEE ICDM Workshop on Frequent Item Set.