Scientific Journal Impact Factor (SJIF): 1.711
International Journal of Modern Trends in Engineering and Research
IJMTER
ISSN (ONLINE) : 2349 - 9745
ISSN (PRINT) : 2393 - 8161

Sensor Network Analysis Using Rule Data Mining

Miss Vrushali S. Sakharkar[1], Ashwini Buskat[2], Ketki S. Muzumdar[3], Akhilesh A. Tayade[4]
[1,2,3]*PRMIT & R,BADNERA*
[4]*P.R.POTE COLLEGE OF ENGINEERING*

Abstract-Traditional data mining algorithms proposed to solve the classical association rules mining problem to be applied on sensor based class of applications that generate and use sensor data. In this paper we have analyzed framework of Sensor network in association rule data mining. Wireless Sensor Networks produce large scale of data in the form of streams. Recently, data mining techniques have received a great deal of attention in extracting knowledge from WSNs data. Mining association rules and the sensors data provides useful information for different applications .In this work we have analyzed mining PL T, SP-Tree and FP-growth from WSNs. We analyzed PLT for s100 and s200 with support value versus number of messages. We found that at PLT s200, the number of messages is decreased considerably with increase in support values. We compared PLT and FP- growth the CPU time is decreased considerably with increase in support values. The experimental result shown that in case of SP-Tree and PLT, the CPU time decreases slightly for PLT then there is no change for SP- Tree when support value increases.

Keywords — WSN, Data Mining, PLT, FP-growth, SP-Tree, Association Rule Mining.

I. INTRODUCTION

1.1 Sensor Network

A sensor network is composed of a large number of sensor nodes, which are densely deployed either inside the phenomenon or very close to it. The position of sensor nodes need not be engineered or pre-determined. This allows random deployment in inaccessible terrains or disaster relief operations [8]. On the other hand, this also means that sensor network protocols and algorithms must possess self-organizing capabilities. Another unique feature of sensor networks is the cooperative effort of sensor nodes. Sensor nodes are fitted with an on-board processor. Instead of sending the raw data to the nodes responsible for the fusion, sensor nodes use their processing abilities to locally carry out simple computations and transmit only the required and partially processed data.

WSN Applications: Sensor networks may consist of many different types of sensors such as seismic, low sampling rate magnetic, thermal, visual, infrared, acoustic and radar, which are able to monitor a wide variety of ambient conditions that include the following[9]:

• Temperature,
• Humidity,
• Vehicular movement,
• Lightning condition,
• Pressure,
• Soil makeup,
• Noise levels,

Sensor nodes can be used for continuous sensing, event detection, event ID, location sensing, medical and local control of actuators. The concept of micro-sensing and wireless connection of these nodes promises many new application areas. We categorize the applications into military, environment, health, home and other commercial areas. It is possible to expand this classification with more categories such as space exploration, chemical processing and disaster relief.

1.2 Data Mining

Advances in wireless technologies have led to the development of sensor nodes that is able of sensing, processing, and transmitting. This new trend in sensor technology allow to construct networks, Wireless Sensor Network (WSN), that consist of several sensor nodes with the main function of sensing the surrounding area and send the detected events to a well equipped node called the sink in multi hop fashion. The detected events are transmitted to the sink periodically or if they meet a particular predicate [1]. Frequent item sets play an essential role in many data mining tasks that try to find interesting patterns from databases, such as association rules, correlations, sequences, episodes, classifiers, clusters and many more of which the mining of association rules is one of the most popular problems.

FP-growth, PLT and SP-Tree: In order to count the supports of all generated item sets, FP-growth [6] uses a combination of the vertical and horizontal database layout to store the database in main memory. Instead of storing the cover for every item the database, it stores the actual transactions from the database in a tree structure and every item has a linked list going through all transactions that contain that item. This new data structure is denoted by FP-tree (Frequent- Pattern tree) and is created as follows. The reason to store transactions in the FP-tree in support descending order is that in this way, it is hoped that the FP-tree representation of the database is kept as small as possible since the more frequently occurring items are arranged closer to the root of the FP-tree and thus are more likely to be shared. PLT, the data structure that will be used to store the behavioral data extracted from WASNs and generate the frequent patterns. The PLT's mining process follows the pattern growth approach and generates the frequent patterns in a recursive way. Sensor Pattern Tree (SP-Tree) [3] which is able to capture the information with one scan over the stream of sensor data and store them in a memory-efficient highly compact manner, similar to FP-tree SP-tree is to obtain the frequency of all event-detecting sensors' data and construct a prefix-tree based on that in any canonical order, then reorganize the tree in a frequency descending order. Through the reorganization the SP-tree can maintain the frequently event-detecting sensors' nodes at the upper part of the tree, which, in turn, provides high compactness in the tree structure.

II. WSN IN ASSOCIATION RULE MINING

Time is a critical issue in WSNs and introduces the possibility of temporal relations between sensors. These relations are important in that they can help in predicting the sources of future events. Several techniques can be used to extract these temporal relations, among these techniques data mining has recently received a great deal of attention. However, the stream nature of senor data along with the limited resources of wireless networks, bring new challenges to the data mining techniques that should be addressed Among these challenges are the type of the knowledge to be extracted from the networks and the way to extract the required data to mine the defined knowledge.

III. RELATED WORK

Azzedine Boukerche et al proposed a framework for mining wireless sensor networks. In this framework consists of a new formulation for the sensors' associations' rules problem, distributed extraction

methodology [1], and a compressed representation structure for sensor data. The new formulation captures the temporal relations between sensors, these relations are able to generate the set of correlated sensors which can be used later to estimate the value of another sensor, to predict the future sources of events, or to identify faulty nodes. The distributed extraction tries to maximize the network lifetime through optimizing number of exchanged messages. The measurements used to evaluate the performance of the distributed extraction were the number of messages needed to extract the data from the sensor network and the amount of the data routed to the sink. The comparison is based on the simulator that has been built using MATLAB [5]. In this simulator, they abstracted the underlying communication protocols and it has been assumed that events generation is uniformly distributed over the number of slots within the given historical period along with a certain degree of correlation between sensors that ranges from 0 to 1%.
The measurements used to evaluate the performance of the proposed distributed extraction scheme were the number of messages needed to extract the data from the sensor network and the amount of data that were routed to the sink node. The comparison is based on the simulator that we have built using Matlab version 7.4.0.287 [5]. In order to evaluate the performance of the PLT structure, they compared its performance with the FP-Growth algorithm implemented by Goethals [7]. An Intel 3.00-GHz Pentium 4 with 2 Gbytes of memory has been used in our experiments. Azzedine Boukerche used both (time) and (memory usage) commands in order to measure the CPU time and memory size used.

Syed Khairuzzaman et al proposed a tree structure, called sensor pattern tree (SP-tree) for mining association rules for Wireless Sensor Networks data [3]. The important features of SP-tree are
(i) It can be constructed with one scan over the sensor *epoch*s, which is highly crucial while the streams of sensor data flow;
(ii) It is a frequency-descending tree structure, which enables an efficient FP-growth-based mining technique.
The comparisons were held using the datasets commonly used in frequent pattern mining implementations. They considered both synthetic data generated from and real datasets obtained from. All programs are written in Microsoft Visual C++ 6.0 and run with Windows XP on a 2.66 GHz CPU and 1 GB memory.
Azzedine Boukerche et al introduced the real data is generated from a sensor network deployed at Intel Berkeley Research lab, and wildly used by many researchers. It consists of tuples for environmental sensors generated from a network consist of 54 nodes that report readings every 30 seconds. The ideal reporting will consist of 54 tuples every 30 seconds; however, the radio transmitters and software used for data collection were quite lousy, which led to many missed readings in each epoch [4]. Boukerche found such hard conditional configuration to be the best layout to analyze the performance of their distributed methodology by assuming that a missed reading from a sensor is an undetected event. The measurements were collected using filtering routines applied to the data with different historical periods and minimum support values. Boukerche conducted two experiments using historical periods of 5 and 10 days with minimum support values ranging from 10% to 90% and a time slot size equal to 30 seconds.

IV. RESULTS AND DISCUSSIONS

In this experiment we have analyzed for PLT values s100 and s200. From fig.4.1 it is shown that at PLT s200, the number of messages is decreased considerably with increase in support values. At PLT s100, the number of messages slightly decreased with increase in support values.

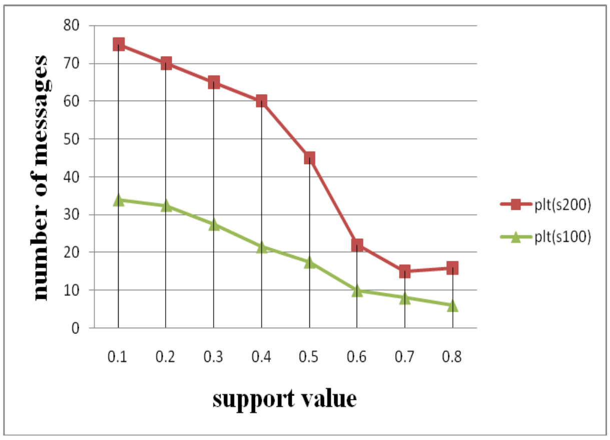

Fig. 4.1 Support value versus number of messages for s200 and s100 using PLT.

The comparison result for PLT and FP-growth from fig. 4.2 is that in case of PLT, the CPU time is decreased considerably with increase in support values. For SP-Tree the CPU time initially is very high and then decreases dramatically. There is no change in CPU time in both the cases after support value .4.

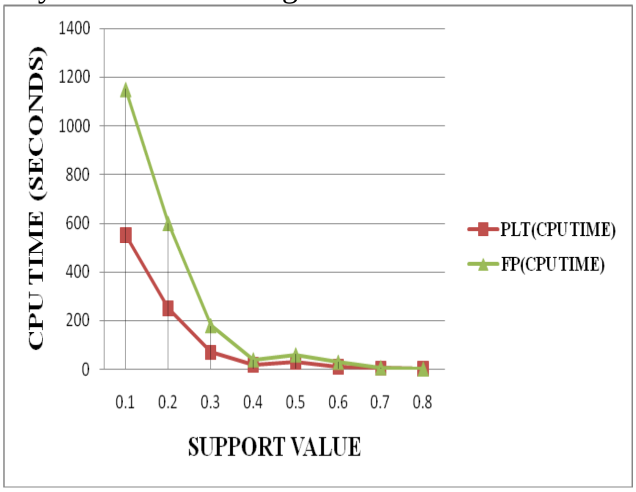

Fig. 4.2 Support value versus CPU Time for S100 using PLT and FP-Growth**.**

From fig. 4.3, it is seen that both PLT and SP-Tree are compared. In case of PLT, the CPU time is decreased with increase in support values. For SP-Tree the CPU time decreases slightly then there is no change after support value 4. But PLT consumes more CPU time than SP-Tree.

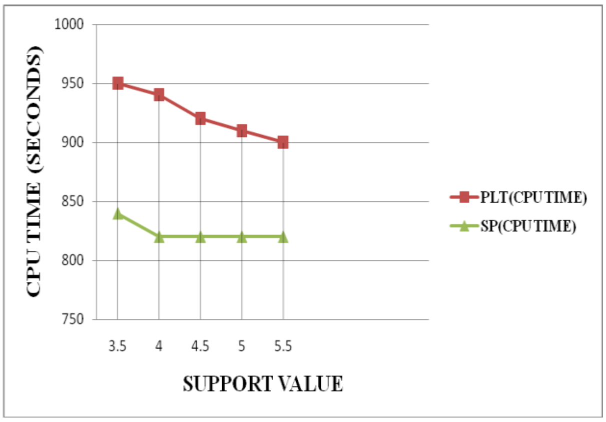

Fig. 4.3 Support values versus CPU Time for S100 using PLT and SP-Tree.

V. CONCLUSIONS AND FUTURE WORK

In this work we have analyzed mining PLT, SP-Tree and FP-growth from WSNs. We analyzed PLT for s100 and s200 with support value versus number of messages. At PLT s200, the number of messages is decreased considerably with increase in support values. The CPU time consumed by PLT is more that of SP-Tree and less than FP-growth. Over all FP-growth consumes initially high CPU time in low support values and SP-Tree consumes considerably less CPU time than PLT.

This comparison is limited with analysis of FP-Tree, PLT-Tree and SP-Tree with less number of nodes but in future it has to be compared with more number of nodes and more messages in different fields.

REFERENCES

[1] Azzedine Boukerche and Samer Samarah, “A Performance Evaluation of Distributed Framework for Mining Wireless Sensor Networks”, ANSS'07 2007

[2] Azzedine Boukerche and Samer Samarah, “A Novel Algorithm for Mining Association Rules in Wireless Ad Hoc Sensor Networks”, IEEE Transactions On Parallel And Distributed Systems, Vol. 19, (7), Page 865, July 2008.

[3] Syed Khairuzzaman Tanbeer, Chowdhury Farhan Ahmed, Byeong-Soo Jeong, Young-Koo Lee “Efficient Mining of Association Rules from Wireless Sensor Networks”, ICACT, Feb. 15-18, 2009.

[4] Azzedine Boukerche and Samer Samarah “An Efficient Data Extraction Mechanism for Mining Association Rules from Wireless Sensor Networks” ICC 2007 proceedings.

[5] http://www.mathworks.com/products/matlab.

[6] “Frequent Item set Mining Implementations,” Proc. IEEE ICDM Workshop on Frequent Item set Mining Implementations (FIMI’04)http://fimi.cs.helsinki.fi, 2004.

[7] Goethals http://www.adrem.ua.ac.be/~goethals.

[8] D.Culler, D. Estrin, and M. B. Srivastava, "Overview of Sensor Networks," Computer, vol. 37 (8), pp. 41-49, August 2004.

[9] S. S. Iyengar and R. R. Brooks, “Distributed Sensor Networks” CRC press ‘Inc., pp 1188 Dec 28, 2004.