ISSN:0975-9646

U. Sandhya Rani et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (4) , 2012,4743 -4746

Dynamic Behavioral Pattern Mining from Point of Coverage Wireless Sensor Networks

U. Sandhya Rani, P.Subba Rao
Department of CSE, Aditya Engineering College
Surampalem, Kakinada

***Abstract*-- Tremendous and potentially infinite volumes of data streams are often generated by real-time surveillance systems, communication networks, Internet traffic, on-line transactions in the financial market or retail industry, electric power grids, industry production processes, scientific and engineering experiments, remote sensors, and other dynamic environments. Unlike traditional data sets, stream data flow in and out of a computer system continuously and with varying update rates. It may be impossible to store an entire data stream or to scan through it multiple times due to its tremendous volume. To discover knowledge or patterns from data streams, it is necessary to develop single-scan and on-line mining methods. This paper has introduced schedule buffer data preparation mechanism for energy conservation. Secondary memory is used at the sink, A novel algorithm PC-MASW (point of coverage multiple adaptive sliding window) has used for efficient memory management at sink.**

***Keywords*— Multiple Adaptive sliding Window, wireless sensor network, Data Streams.**

I. INTRODUCTION

Wireless sensor network is a collection of sensor nodes that are deployed inside the phenomenon or very closed to it. These Sensor nodes are capable of monitoring various environmental and physical conditions such as temperature, heat, light, pollutants etc. Various applications of sensor networks are habitat monitoring, Industrial sensing and Diagnostics, Infrastructure Protection, Battle field awareness [1].

Challenges in wireless sensor network are

1. Each sensor has limited transmitting range and energy conservation depends upon the transmitting range of the message.
2. Several Sensors are monitoring the same Target so when event is arrived it is monitored by several sensors and every sensor send data about the same event.
3. Sensors lifetime is limited.
4. Sink may not receive all events due to unreliability of wireless sensor network. Some of the events may lose during transmission.

Several techniques have been proposed in the literature to enhance the performance of WSNs, such as clustering, multi-hop transmission, Data aggregation and knowledge discovery techniques[8][10]. Knowledge discovery techniques provide the solution for unreliability of wireless sensor networks.

Pont of coverage sensor network with single user we can directly find association rules but in case of multi user system, user can ask multiple readings. In case of battle field services at the same time one user ask ground based sensors data and other user asks airborne sensors monitored data to acquire information about the movement and deployment of enemy's combat forces in the battle area of responsibility. In those cases clustering is applied on the data first for unsupervised learning of the data.

Clustering of sensor stream is a challenging task for two major difficulties. The reading operation must be carried out linearly and data items could be accessed only a few times. the centroid of clusters in stream often changes over time. Thus, selecting representative points from cluster is meaningful and useful.

For clustering of such evolving data streams, taking some kind of mechanism to evaluate the length of limited history is important because it need not to retain too many historical data. Sliding time window is a classical and natural method which meets such case. However, the length of sliding time window is usually fixed in traditional approaches, and it gives equal treatment to everyone, the life of each item is the same, thus it's unfair to parts of items which are more "important" than others. In order to cope with this problem ADWIN2 is proposed in [11]. ADWIN2 detects the shake in current window, and then adjust the length of the window according to follow rules:

(1) In order to get greater accuracy, the window will grow automatically when the data is stationary.

(2)The window will shrink automatically when change is taking place, and the stale data items are discarded.

The limitation of ADWIN2 is that it takes a broad and passive strategy to adjust the length of sliding window; the adjustment operation may affect many data items, parts of which should not be affected originally.

The rest of this paper is organized as follows. Section 2 provides system architecture. Section 3 presents the data extraction methodologies in the literature. Section 4 contains proposed work. Section 5 presents conclusion.

II. SYSTEM ARCHETECTURE

The architecture in this paper is point of coverage wireless sensor network shown in fig 2.1 consists of a set of sensor nodes $\{S_1,S_2,\ldots,S_n\}$ and set of targets $\{T_1,T_2\ldots,T_m\}$ and sink. Here sensor nodes are deployed randomly with in a field and ensures each Target must be monitored by at least one sensor and several sensors monitor the same target. Each sensor is attached with a flash memory can store monitored events up to the memory capacity.

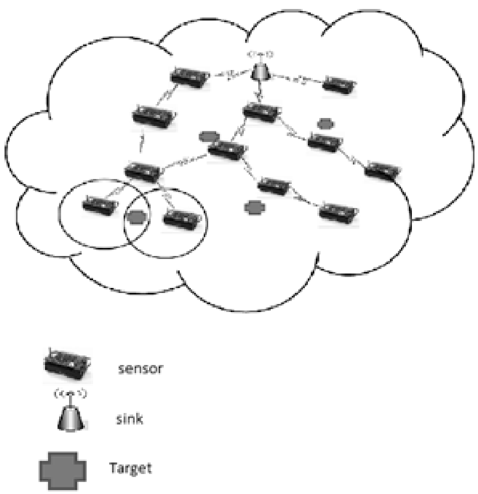

Fig 2.1 POCW architecture

III. DATA EXTRACTION METHODOLOGIES

Two different mechanisms have been proposed for the preparation of data needed for generating Sensor Association Rules [2]. These mechanisms are as follows:

A. *Direct reporting technique*

In this mechanism, sensor behavioral data is transferred to the Sink without any processing by the sensor node. Although the Direct reporting technique sounds simple and no overhead is put on the sensor nodes, it is a costly solution in terms of energy consumption.

B. *Distributed Extraction*

This mechanism is designed to put more computational load on the sensor nodes by equipping each sensor with additional storage space to store behavioral data for the given historical period.

IV. PROPOSED WORK

Data extraction methods in the literature are not suitable for point coverage sensor network because in point of coverage sensor networks sensors are deployed randomly and more than the requirement. If all sensors are working continuously then network life time is decreased. To increase the network life time this paper uses Schedule buffer data preparation mechanism is used for data extraction from sensor nodes.

In point coverage application sensor association rules are not suitable.

Multiple adaptive sliding window technique is applied to data streams at the sink for identify the dynamic behavior of the events.

A. *Schedule Buffer data preparation mechanism*

It is modified version of distributed extraction technique. Here assumption is network has been deployed and that all its elements initialized without a problem and the sink knows how many sensor nodes are working as expected, their position and the position of the targets by using global positioning system.

First, the sink divides the sensor nodes into several disjoint sets. Each set can completely cover all the Targets in network. Using disjoint sets to take turns monitoring prolongs the lifetime of the network instead of having all the sensor nodes monitoring the targets at the same time [3]. In this technique first sink broad cast some parameters to the sink. Those are profile time and slot size. At any time only one disjoint set is active and sensors in that disjoint set quantize the time into regular intervals based on slot size and historical period. Then sensors in the active set take the snap shots of these events along with Target id's and store these values up to the schedule time. if the schedule time is completed then values are send to the sink.

Routing procedure:

Tree based routing protocol [4][5] is shown in fig 4.1. The entire nodes in the network form a tree with different level. And the node of the tree has some degree constraint which is depend on the level of the node. The distance between the two levels is equal to the Radio range approximately equal to the maximum distance cover by the sensor node.

After Tree is formed for routing, when profile time (how long target should be profiled) is completed child nodes send data to its parent node. Here parent node is chosen based on their state (i.e. Active state and sleeping state of the sensor). At any time only one parent node is selected and sends data to that parent node. Then parent node aggregates its data with children data and sends it to its parent node. Finally the node near to the base station sends the collected data to the base station.

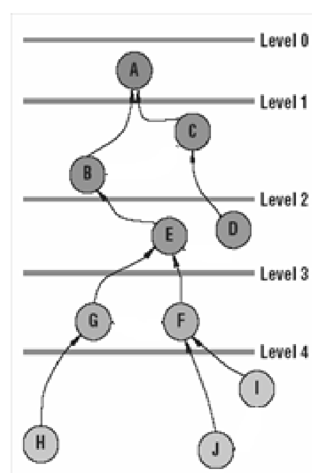

Fig 4.1 Tree based Routing protocol

B. *Target Association Rules*

To generate associations among a set of targets, we have to collect behavioral data that describe the activity of the targets over the time. In this section, we present a formal definition for all the main concepts needed to generate Target-based Association Rules. Let T = { T1, T2, . . T_n } be a set of targets in a particular POCW field. Let S = { S1, S2…S_n } and be the set of sensor nodes deployed within the network. Target Association Rules are based on the common intervals of events' occurrences at targets.

Definition 4.1 Let Ti be a target in a particular POCW. AS(Ti) = { t1, t2 . . . tn } such that n <= (profile time/slotsize) and B_{TI} (Tk)=1 for all 1<=k<=m , is then defined as the Activity Set of Target Ti.

Definition 4.2. A Target Database (TD) is defined as the set of targets, covered by a particular sensor network, along with Their Activity Sets. Table 2 shows an example of a target database.

Definition 4.3. P = {T1, T2 . . . Tk } such that P ⊆ T is defined as a pattern of targets.

Definition 4.4. The support of pattern P = { T_1, T_2, . . . T_k } in the Target Database is defined by the cardinality of the set that is produced by intersecting all the activity sets of targets in this pattern

$$Support(p) = \left| \bigcap_{\forall l \in p} AS(T_j) \right| \quad (4.1)$$

Definition 4.5. A pattern P is said to be frequent if its support is greater than or equal to a given minimum support.

Definition 4.6. Target-based Rule is defined as the implication $P' \Rightarrow P''$ where $P' \subset T$, $P'' \subset T$, and $P' \cap P'' = \Phi$.

Definition 4.7. The support of the rule ($P' \Rightarrow P''$) is defined as the support of the pattern ($P' \cup P''$) in the Target Database

Definition 4.8.C confidence of the rule is defined by

$$Conf(P' \Rightarrow P'') = Support(p' \cup p'') / Support(p') \quad (4.2)$$

Knowledge Discovery for Target-based Rules is the process of generating all the rules in the Target Database that meet predefined minimum support and confidence percentage. One important application for TAR is the prediction of the association among targets. An example of this application is the rule that states T1, T2 ⇒ T3. If events e1 and e2 happen at T1 and T2 at the same time interval, then there is an event happening at T3 in the same time interval with probability equal to the rule's confidence. This knowledge will help in the case of message lost from the sensor node covering T3. If the cardinality of activity set of any target is less than minimum support than that Target will not participate in generating association rule.

To estimate the values of the missing readings, first uses association rule data mining to identify the Targets that are related to the Targets with the missing readings, then uses the current readings at related Targets to calculate the missing values in the current round [9].

C. *PC-MASW Technique*

PC-MASW stands for point of coverage multiple adaptive sliding window technique. In each data item associates with an independent time window, length of which is adjusted only by importance of itself so that the adjustment operation will not affect other data items. Once current time could not be covered with one point's time window, it's treated as inactive items which mean there will be no more items evolved from it. PC-MASW always drops such inactive items as much as possible. This feature is very important and useful because:

- It reduces the time and space requirements strikingly. So benefited from the initiative to discard obsolete data items, PC_MASW has more powerful handling capacity.
- It gives different treatment to different item, such that more influential points would have longer time window.
- There are always new items join in existing clusters, such that each cluster has a sliding time window.
- Different clusters have independent sliding time windows
- The window length of a data item is computed according to its influence.

In evolving data streams, the number of potential cluster is usually in-definability, besides the concept drifts feature, trying to find global cluster in such case is difficult, PC_MASW provides a mechanism to select some items of cluster at different time point as the representation, they could be used to query similar cluster. Given a certain cluster, users could retrieve similar clusters in specified time range quickly via many existing or other suitable evaluation index. Even more valuable is that these representative points could describe how one cluster evolves on the timeline clearly.

For each batch of transactions, discovering clusters of users (grouped by behavior) and then analyzing their navigations by means of a sequences alignment process. This allows us to obtain clusters of behaviors representing the current usage of the sensor target environment. For each cluster having size greater than min Size (specified by the user) store only the summary of the cluster. This cluster is called as micro cluster. This summary is given by the aligned sequence obtained on the sequences of that cluster. Based the on the nature of the cluster association rules will be generated. Later to handle the large volume of data, assign time stamp and life time to each cluster. If any cluster influences the coming data default life time will be increased. If not it will be mark as inactive and store it in the memory. Later association rules will be generated based on that cluster nature.

V. CONCLUSION

In many applications, it is difficult to find strong associations among data items at low or primitive levels of abstraction due to the scarcity of data at those levels and difficult to fit in memory if that data is dynamic. On the other hand, strong associations discovered at high levels of abstraction may represent commonsense knowledge. Therefore, multilevel association rules provide sufficient flexibility for mining and traversal at multiple levels of abstraction. Multilevel association rules can be mined efficiently using concept hierarchies under a support confidence framework. Therefore of this paper investigate the possibility of finding target association rules using multiple adaptive sliding windows for large volume of sensor data stream in Point of Coverage network.

REFERENCES

[1]. F. Zhao and L.J. Guibas, Wireless Sensor Networks: An Information Processing Approach. Morgan Kaufmann Publishers, 2002.

[2]. A. Boukerche and S. Samarah, “A Novel Algorithm for Mining Association Rules in Wireless Ad Hoc Sensor Networks,” IEEE Trans. Parallel Distributed Systems, vol. 19, no. 7, pp. 865-877, July 2008.

[3] M. Cardei and D.-Z. Du, “Improving Wireless Sensor Network Lifetime through Power Aware Organization,” Wireless Networks, vol. 11, no. 3, pp. 333-340, May 2005.

[4].Mohammad Zeynali, Leili Mohammad Khanli, and Amir mollenezad.. TBRP: Novel Tree Based Routing Protocol in Wireless Sensor. NetworkVol. 2, No. 4, December, 2009

[5] K. Akkaya and M. Younis, “A Survey on Routing Protocols for Wireless Sensor Networks,” Ad Hoc Networks, vol. 3, no. 3, pp. 325- 349, 2005.

[6] M. Ilyas, I. Mahgoub, "Handbook of Sensor Networks: Compact Wireless and Wired Sensing Systems", CRC Press, London, Washington, D..C., 2005.

[7]. R. Agrawal, T. Imielinski, and A. Swami, “Mining Association Rules between Sets of Items in Large Databases,” Proc. ACMSIGMOD Conf. Management of Data, pp 207-216, May 1993

[8] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “From DataMining to Knowledge Discovery: An Overview,” Advances inKnowledge Discovery and Data Mining, pp. 1-34, The MIT Press,1996.

[9] M. Halatchev and L. Gruenwald, “Estimating Missing Values in Related Sensor Data Streams,” Proc. 11th Int’l Conf. Management of Data, 2005.

[10] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, “From Data Mining to Knowledge Discovery in Databases,” AI Magazine, vol. 17, no. 3, pp. 37-54, 1996.

[11] A. Bifet and R. Gavalda. Learning from time-changing data with daptive windowing. In SDM, 2007.