International Journal of Engineering Research and Development
e-ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 9, Issue 9 (January 2014), PP. 67-71

Sensor Association Rules: A survey

Prachi singh
Department of computer science, MITS Gwalior, Madhya Pradesh, India

Abstract:- Recently, knowledge Discovery Process has proven to be a promising tool for extracting behavioral patterns regarding sensor nodes from wireless sensor networks. This paper presents a review of the available literature on the sensor association rules to find behavioral patterns between the sensor data.

Keywords:- Wireless Sensor Networks, Association Rule Mining

I. INTRODUCTION

Wireless Sensor Networks (WSNs) have been successfully applied for detailed observation of a variety of real-world applications, especially in battle fields, smart buildings, the human body, and the other applications that require the fine-grain monitoring of physical environments that are subject to critical conditions, such as fires, toxic gas leaks, and explosions. With such applications come new challenges for information processing in sensor networks. Unlike centralized system a sensor network is subject to a unique set of resource constraints such as finite on board battery power and limited network communication bandwidth. In a typical sensor network, each sensor node operates and has a microprocessor and a small amount of memory for signal processing and task scheduling. So, WSN suffers from lot of problems such as lost messages, delay in data delivery, loss of data, and data redundancy, etc., resulting in poor quality of service (QoS). Various techniques and protocols have been suggested to improve the QoS of WSN.

Recently , the knowledge discovery process, which is a well known process in traditional database systems used to extract patterns from data, has shown to be a promising tool to improve WSN performance and its QoS. Knowledge discovery in WSN (KDW) has been used to extract the following two types of information (knowledge): first is, patterns about the surrounding environment, which are extracted from the data reported by sensor nodes, and the second is, behavioral patterns about sensor nodes, which are extracted from meta-data describing sensor's behaviors.Many techniques have been proposed to solve the mining frequent patterns problem from the databases, these techniques differ mainly in the way they represent the database and in which they generate the frequent patterns. These can be classified into two main approaches: the candidate generation approach and the pattern growth approach. The candidate generation approach enumerates the frequent patterns in a level wise manner, with several scans of database. At each level, the patterns found to be frequent are used to generate the candidates for the next level. Within this approach are the Apriori, Directed Hashing & Pruning, the partitioning algorithm, and Dynamic Itemset Counting etc. A comprehensive survey of recent algorithms for mining association rules is presented in [18]. The common algorithm of candidate generation approach is the Apriori. This algorithm first introduced by Agrawal et al.,[1] in 1993 to discover the correlations between objects in transactional databases. This algorithm was able to predict the items that can be purchased within the same transaction. These predictions have a great impact on making decisions about which item should be put on sale or which items should be placed near to each other.

Association Rule Mining is a costly process. The complexity increases exponentially with the number of the items presented in the database.

In [2], Loo et al., proposes a framework for extracting association rules from sensor networks. The association involves the sensors' values as the main objects in the rule. In their methodology, the time is divided into intervals, and the sensors' values at that interval formulate the transaction to be stored in the database. Each different value of a sensor is regarded as single element and it is assumed that sensors take on a finite number of discrete states. Each transaction is associated with a weight value indicating the validity of the transaction (i.e. the duration of time in which there were no changes in sensors' values). The support of the pattern is then defined to be the total length of all non overlapping intervals in which the pattern occurs. The main differences between this methodology and the sensor association rules proposed in [5], is that in sensor association rules, sensors themselves act as the main objects, not their values, and the support of the pattern is the number of epochs in which the pattern occurs as a subset. The algorithm presented in his work is complex and to handle the streamed data is also difficult and time consuming. Most of the data mining techniques applied to sensor data are based on centralized data extraction, where the data is collected at a single site and then the mining technique is applied. However, there are still some techniques that use the distributed nature of sensors and try to apply distributed mining algorithms.

In [3], Key Romer et al, proposes an in-network data mining technique to discover frequent patterns of events with certain spatial and temporal properties. In their framework, each sensor should be aware of the events that are within a certain distance from itself (this distance may be a Euclidian distance or number of hops). The sensor then collects these events and applies a mining algorithm to discover the pattern that satisfies a given parameters. The main difference between Romer's methodology and the sensor association rule proposed in [5], is in type of events in the discovered pattern, in which he targeted all the possible values of a sensor as objects, whereas in [5] targets sensors in an abstract manner regardless of the sensor value, which will reduce the amount of the data exchanged between the nodes if we assume in-network solution.

In [4] Miahail and Le, propose a data model to store and update the data generated from sensor networks. In their work, they adopt a centralized extraction methodology where all the data is collected at sink for further analysis. They focus on generating association rules between pairs of sensors instead of generating all associations. The direct application of their extracted rules is to determine the set of the sensors to be used to predict a missed value of other sensors.

II. SENSOR ASSOCIATION RULES

The concept of sensor association rules (SAR) first introduced by Azzedine Boukerche and Samer Samarah [5]. In their work they introduce a new formulation for the association rules, which is able to generate the time relations between sensor devices in a particular sensor network. In this formulation they allow traditional data mining algorithms proposed to solve the classical association rule mining problem to be applied on sensor based class of applications that generate and use sensor data. The generated rules will give a clear picture about the correlations between sensors in the network and can be used to predict the sources of the future events. In their work they also suggested two possible methodologies for extracting the data from wireless sensor networks. The first is a direct transmission, in which the data is transmitted to the central site without any optimization from the nodes. The second method considers the overall limited resources in the network and each node tries to optimize the number of messages it will send.

Sensor association rules is among the first knowledge discovery techniques that have been proposed to generate behavioral patterns. It discovers the temporal relationships between sensor nodes in detecting events. An example of sensor association rules could be ($A \rightarrow B$, λ, 60%). Here A and B are the sets of the sensors. The means of this rule is that if sensors from A detect events within time λ, then there is 60% chance that sensors from B detects events within that same time interval. The main benefit of Sensor Association Rules (SAR) in many applications is the ability to predict sources of future events, and the ability to identify the sets of temporally correlated sensors.

Definition: Let $A= \{s_1,s_2,s_3\ldots\ldots,s_m\}$ be a set of sensors in a particular sensor network. We assume that the time is divided into equal sized slots $\{t_1,t_2,\ldots..,t_n\}$ such that $t_{i+1} - t_i = \lambda$, for all $1 < i < n$. λ is the size of each time slot. $T_{his} = t_n - t_1$ is the historical period of the data defined during the data extraction process. A set $P = \{s_1,s_2,\ldots..,s_k\} \subseteq A$ is called a pattern of sensors.

A sensor database, DS, is defined to be a set of epochs in which each epoch is a couple $E(E_{ts},P)$ such that P is a pattern of sensors that report events within the same time slot. E_{ts} is the epoch's time slot. We say an epoch $E(E_{ts}, P)$ supports a pattern P_1 if $P_1 \subseteq P$. frequency of the pattern P_1 in DS is defined to be the number of epochs in DS that support it.

$Freq(P_1, DS) = | \{E(E_{ts}, P) | P_1 \subseteq P\} |$.

Sensor association rules are implications of the form $P' \rightarrow P''$ where $P' \subset A$, $P'' \subset A$ and $P' \cap P'' = \Phi$. The frequency of the rule ($P' \rightarrow P''$) is the frequency of the pattern ($P' \cup P''$). The confidence of the rule is defined as follows:

$Conf(P' \rightarrow P'') = Freq(P' \cup P'', DS) / Freq(P', DS)$

The major impacts of Sensor Association Rules that could bring benefit to many applications are the ability to predict sources of future events, and the ability to identify sets of temporally correlated sensors.

Le Gruenwald et al, [7] proposed a new approach to find associations between the sensors in any sensor network. They incorporate the time factor to estimate the missing sensor values. They propose a data estimation technique, FARM, which uses association rule mining to discover intrinsic relationships among sensors and incorporate them into the data estimation while taking data freshness into consideration. FARM uses association rule mining to find related sensors. The contribution of FARM is:

- Incorporate the temporal aspect into association rules and estimation
- Compact data streams and allow a large history to appropriately influence sensor rules
- Guarantee retrievability of original data from its compact form

Azzedine Boukerche et al, proposed in [5], a new representation structure for mining association rules from wireless sensor networks that is a tree structure which is known as the PLT (Positional

Lexicographic Tree), a better solution than the FP-tree proposed in [], to find the frequent sensor patterns from the set of epochs (DS). The mechanism proposed in his work uses a partitioning mechanism in PLT, that is used to locate the conditional vectors of a particular pattern easily, instead of following the node's link as in the FP-tree. These partitions are independent hence, we don't need the entire structure to be in the main memory, as opposed to the FP-tree that requires the whole structure to be in the main memory. It also uses a numerical value that is known as the comparison value of the position vectors, is used to locate the insert vectors accelerate the process of accessing the PLT structure, whereas the FP-tree doesn't include such values. PLT requires smaller variable sizes for storing the positional values, since it uses the lexicographic distance to represent sensors, as opposed to the FP-tree that uses the actual sensor's identifier.

In [9], Samer Samarah et al, proposed coverage based association rules for wireless vehicular ad hoc networks, to discover the correlation among the set of locations monitored by the network. Coverage based rules have been designed specifically for sensor networks that guarantee a k-coverage property for the area under monitoring. Here k-coverage property states that in a k-coverage sensor network, all k nodes within a specific area are expected to detect the same event, as a result all k sensor nodes will have the same activity sets during the process of profiling their behaviors when preparing the data needed for generating sensor association patterns.. Author proposed a relaxation version of sensor association rules that discovers the correlation between a set of locations (areas) rather than individual sensor nodes.

Sayed Khairuzzaman Tanbeer et al in [10] proposes a new tree-based data structure called sensor pattern tree (SP-tree) to generate association rules from WSNs data with one database scan. This SP-tree is able to capture the information with one scan over the stream of sensor data and store them in memory-efficient highly compact manner, similar to FP-tree. The main idea of 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. Once the SP-tree is constructed, we apply the efficient FP-growth mining technique on it. The basic operations in the FP-growth based pattern growth mining approach are counting the set of frequent event-detecting sensors, constructing conditional pattern base for each of such sensor, and constructing new conditional tree from each conditional pattern base. [19]

In [11], Azzedine et al, proposed an in network data reduction mechanism to reduce the amount of data (about sensor's behaviors) by removing some of the data's redundancies. The proposed in network data reduction is implemented on top of a data gathering tree, which they refer to as a minimum nodes data gathering tree (MNDGT) and in cooperation with the distributed extraction mechanism. The MNDGT takes in consideration that not all sensor nodes are going to participate in formulating the sensor association rules, and it is constructed in such a way that those nodes that will participate in formulating sensor association rules are included, plus the nodes that are needed to maintain the minimum distance to the sink. The MNDGT is designed in such a way that the activity sets of different nodes can meet at intermediate nodes where a reduction technique can take place.

In [16], Anjan Das, proposes an enhanced data reduction mechanism to gather data for mining sensor association rules. In this the he tried to enhance the algorithm proposed by Azzedine et al in [11], by removing more redundancy between sensor activities.

In [12], Samer et al, proposed target based association rules for point of coverage wireless sensor networks. They propose a new type of behavioral patterns, which they refer to as target based association rules, to discover the correlation among a set of targets monitored by a network at border region. Such association can be the main driver for predicting the source of future events. The main difference between TAR and SAR is that in TAR the interest will be in capturing the association between the targets, instead of the sensor nodes. He also proposes two mechanisms to find the frequent target patterns. First is the all-nodes based data preparation mechanism which is similar to the direct reporting mechanism, requires each sensor node to be active all the time, without a predefined schedule. Although this solution sounds simple and does not put extra load in the network, it is costly solution in terms of energy consumption. Second is the schedule-buffer based data preparation mechanism, which solves the problems in the previous mechanism. A global schedule has been defined by the sink. The global schedule defines the activity time for each sensor and the set of targets that should be monitored by each sensor. This solves the energy consumption problem because not all the sensors have to be active all the time. But to follow the global time schedule each sensor node in same schedule should have to synchronize with each other to generate the activity set.

In [15], author proposes one more mechanism that is the fused-schedule-buffer based data preparation mechanism, which is an improved version of the Schedule-Buffer based mechanism. The improvement consists of allowing the sensor nodes to collaborate with each other during the data preparation process. Although it reduces the redundancy in the messages to be sent to the sink but it requires much load on individual sensor.

In [14], Nan Jiang et al, proposed a post mining of non redundant association rules for sensor estimation. They focus on post mining of non-redundant and informative association rules that match the user interests. The generated association rules are then applied to sensor network databases of a traffic monitoring site for missing data estimation purpose, in which data missing by a sensor a estimated using the data generated by its related sensors. Author proposed an algorithm that is developed based on non-redundant informative association rules which means all rules cannot be derived from other rules which means all rules cannot be derived from other rules and the left hand side and right hand side of the selected rules contain the input itemset.

In [13], JunTan et al, proposed a Fd-tree method which requires no scanning of the whole data stream and to only scan the updated transactions once without involving candidate sets generation. A data stream is na ordered sequence of items that arrives in timely order. Traditional association rules mining algorithms are developed to work on static data, therefore, cannot be applied directly in data streams. For example, both the Apriori and the FP-tree mining [19] approaches belong to batch mining. That is, they must process all the transactions in a batch way. In data stream, new transaction insertion and old transaction deletions may lead to previously discovered association rules no longer interesting, and new interesting association rules may also appear. To solve this problem author proposes a new method that uses a Fd-tree structure to store all the information and requires (1) no scanning of the original database, (2) to only scan the updated transactions once without involving candidate sets generation.

TABLE I

S. no	Author	Year	Key Approach	Association between	Applications
[4]	M. Halatchev et al	2005	Centralized extraction	Pairs of sensors	Estimate the missing sensor values
[3]	Key Romer et al	2006	Spatial & temporal features of events	Sensor's values	Estimate the missing sensor values
[5]	Boukerche et al	2007	Positional Lexicographic Tree (PLT)	Sensors	Predict the source of the future events
[7]	Gruenwald et al	2007	Freshness Association Rule Mining (FARM)	Sensor's values	Estimation of missing sensor values
[8]	Boukerche et al	2007	Distributed extraction	Sensors	Estimation of missing sensor values, source of the future events, fault detection
[9]	Samarah et al	2008	Coverage based association rules	Locations	Predict the location of the future events in vehicular adhoc & sensor network based applications
[10]	Tanbeer et al	2009	SP-tree	sensors	Sensor association rules
[11]	Boukerche et al	2009	Minimum node data gathering tree (MNDGT)	sensors	Sensor association rules
[12]	Samarah et al	2009	Target Association Rules	Targets	Predict the source of the future events
[13]	Jun Tan et al	2010	Fd-tree	Sensor's values	Estimation of missing sensor values
[16]	Anjan Das	2012	In-network mechanism	sensors	Estimation of missing sensor values

In [17], Anjan Das proposes a novel association rule mining mechanism in wireless sensor networks. In this he proposes an in network mechanism to find frequent sensor patterns in the sensors themselves. So, the sensors send only the frequent sensor patterns to the sink, not the sensor's complete activity sets. He proposes a mechanism to calculate the frequent sensor patterns in the sensor patterns along with support to the sink. Although this work reduces redundancies in the messages to be sent to the sink, it put extra load on the individual sensor node and it requires extra memory in each sensor node to store the information to calculate the frequent sensor

patterns in the sensors themselves. Table I highlights the key features proposed for WSN in different works discussed in this paper.

III. CONCLUSION & FUTURE WORK

In this paper we discussed sensor association rule and its different variations. There are still many challenges that need to be considering in sensor association rules like many of its applications based on estimation of missing values and predict the source of the future events, the energy efficiency still a very crucial issue to apply these techniques on tiny sensor nodes which have the small memory and processing power. Future research scope in sensor association rules is to enhance the efficiency of wireless sensor network with respect to energy constraint.

REFERENCES

[1]. R. Agrawal, T. Imielinski, and A. N. Swami, "Mining Association Rules Between Sets of Items in Large Databases", Proc. ACM SIGMOD conference on management of Data, 1993, pp. 207-216.

[2]. K.K. Loo, L Tong, B. Kao, and D. Chenung, "Online Algorithms for Mining Inter-Stream Association from Large Sensor Networks," Proc. Ninth Pacific-Asia Conf. Knowledge Discovery and data Mining (PAKDD '05), may 2005.

[3]. K. Romer, "Distributed Mining of Spatio-temporal Event Patterns in Sensor Networks," proc. Euro-American Workshop Middleware for Sensor Networks (EAWMS '06) June 2006.

[4]. M. Halatchev and L. Gruenwald, "Estimating Missing Values in Related Sensor Data Streams," Proc. 11th Int'l Conf. Management of Data (COMAD '05), Jan. 2005.

[5]. Azzedine Boukerche and Samer Samarah, "A New Representation Structure for Mining Association Rules from Wireless Sensor Networks", PARADISE University of Ottawa, Canada, 2007

[6]. A Boukerche, Samer Samarah, "A novel algorithm for mining association rules in wireless ad hoc sensor networks", IEEE Transactions on Parallel and Distributed Systems, Vol 19, No 7, July 2008, pp 865-877.

[7]. Le Gruenwald, Hamed Chok, Mazen Aboukhamis, "Using Data Mining to Estimate Missing Sensor Data", Seventh IEEE International Conference in Data Mining- Workshops, 2007

[8]. Azzedine Boukerche and Samer Samarah, "An Efficient data Extraction Mechanism for Mining Association Rules from Wireless Sensor Networks", 2007.

[9]. Samer samara, Azzedine Boukerche, and Yonglin ren, "Coverage-based Sensor Association Rules for Wireless Vehicular Ad hoc and Sensor Networks", PARADISE Research Laboratory – University of Ottawa – Canada, 2008.

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

[11]. Azzedine Boukerche and Samer Samarah. " In-Network Data Reduction and Coverage-Based Mechanisms for Generating Association Rules in Wireless Sensor Networks", IEEE Transactions on Vehicular Technology, Vol. 58, No. 8, Oct. 2009.

[12]. Samer Samarah, Alexandre Shema Habyalimana, Azzedine Boukerche, " target-based Association Rules for Point-of-Coverage Wireless Sensor Networks", PARADISE research Laboratory- SITE, 2009.

[13]. Jun Tan, Yingyong Bu and haiming Zhao, "Incremental Maintenance of Association Rules Over Data Streams ", 2010 International Conference on Networking and Digital Society.

[14]. Nan Jiang, Zhiqiang Chen, "Post Mining of Non-redundant Association Rules for Sensor Data Estimation", 2010 2nd International Conference on Education Technology and Computer (ICETC).

[15]. Samer Samarah, Azzedine Boukerche, "target Association Rules: A new Behavioral Patterns for Point of Coverage Wireless Sensor Networks", IEEE Transactions on Computers, Vol. 60, No. 6, June 2011.

[16]. Anjan Das, "An Enhanced Data Reduction Mechanism to Gather Data for Mining Sensor Association Rules", IEEE, 2011.

[17]. Anjan Das, "A Novel Association Rule Mining Mechanism in Wireless Sensor Networks", IEEE, 2012.

[18]. B. Goethals, "Survey on Frequent Pattern Mining", 2003.

[19]. J. Han, J. Pei, and Y. Yin, "Mining Frequent Patterns Without Candidate Generation", Proc. The 2000 ACM SIGMOD INT. Conference on Management of Data, 2000, pp. 1-12.