

Knowledge Extraction from Wireless Sensor Networks

Priyanka Pandit¹, Swarupa Kamble²

^{1,2}Pune University, Department of Computer Engineering, RMD School of Engineering, Warje, India

Abstract: *As of late, information administration and handling for remote sensor systems are turned into a theme of dynamic exploration in a few fields of software engineering, for example, circulated framework, database frameworks and information mining. WSN creates substantial measure of information streams. Removing data from the information is the most critical undertaking in WSN. Numerous techniques are accessible to extricate the data from WSN. These strategies produce the affiliation rules utilizing recurrence designs. It makes vast number of tenets a large portion of which are non-educational and neglects to reflect genuine connection among the sensor information. This paper shows new sort of behavioral example called as a related sensor design. Related sensor design catches the co-events and worldly connection of the sensor information. To catch such examples a conservative tree structure, called related sensor design tree (ASP-tree) and mining calculation are proposed. At the point when information streams moves through the system there is plausibility of loss of essentialness for the present time of the old data. ASP-tree is further improved to SWASP-tree to catch centrality of late information by receiving sliding perception window and overhauling tree structure in like manner.*

Keywords: Wireless sensor networks, data mining, knowledge discovery, behavioral patterns, sensor data stream.

1. Introduction

A WSN includes of a large number of heterogeneous or homogeneous nodes usually called as sensor nodes, they communicates via wireless media to the concentrator node called as sink node and works cooperatively to monitor the environment. Wireless sensors are integrated as an ad-hoc fashion to initiate a network, that network is able to deliver detected event in multi-hop transmission to sink periodically. In this mode of transmission networks network creates large amount of data which is in the form of stream. There are data mining techniques which have recently received a great attention to extract intersecting knowledge from these stream data. There are three challenges mining techniques need to be address such as the stream nature of the data, limited resources and distributed nature of sensor.

Data mining techniques have shown to be a promising tool to improve WSN performance and quality of services (QOS). Knowledge discovery in WSN has been used to extract information about the surrounding environment, that are deduce from the data reported by sensor nodes and behavioral patterns about sensor nodes, which are evolved from meta-data describing sensor's behaviors.

Discovering behavioral patterns (i.e., associated patterns) from WSNs can be highly useful in applications that require a fine-grain monitoring of physical environments (e.g., buildings, transportation networks, and battlefield) which may face critical situations like fire, toxic gas leaks and explosion. Behavioral patterns can also be used to predict the source of future events. Knowing the source of a future event may lead to detect faulty nodes, if any, in the network. For example, behavioral pattern mining predicts occurrence of an event from a particular node, but no such event is subsequently reported; this indicates possibility of a node failure. It can also identify the source of the next event when behavioral patterns reveal a chain of related events, e.g., fault in a particular process in an industry may trigger fault

in other processes. Behavioral patterns can also reveal a set of temporally correlated sensors. Such patterns can improve operational aspects (e.g., identifying missed reading) and better management of resources (e.g., sleep awake schedule of nodes) in a WSN.

The main challenges in mining associated sensor patterns are: (1) Appropriate formulation to discover associated sensor patterns maintaining the downward closure property to ensure search space reduction; (2) Devising a compact tree structure that is able to capture the data content in one scan over sensor datasets and yield better mining performance; (3) As older information in the data stream may become unimportant or obsolete for the current time period, it is necessary to make the tree structure adaptive so that it effectively captures the latest information, makes optimal use of memory and extracts complete set of recent associated sensor patterns. To facilitate such tree construction, appropriate data observation window size needs to be determined.

2. Literature Survey

Boukerche et al. [1] propose a comprehensive framework for mining Wireless Ad Hoc Sensor Networks (WASNs), which is able to extract patterns regarding the sensors' behaviors. The main goal of determining behavioral patterns is to use them to generate rules that will improve the WASN's Quality of Service by participating in the resource management process or compensating for the undesired side effects of wireless communication. The proposed framework consists of 1) a formal definition of sensor behavioral patterns and sensor association rules, 2) a novel representation structure that we refer to as the Positional Lexicographic Tree (PLT) that is able to compress the data gathered for the mining process and thus allows the fast and efficient mining of sensor behavioral patterns, and 3) a distributed data extraction mechanism to prepare the data required for mining sensor behavioral patterns. Several

experimental studies have been conducted to evaluate our PLT structure and our proposed data extraction algorithms for mining wireless sensor networks.

Samarah et al. [2] Recently, Knowledge Discovery Process has proven to be a promising tool for extracting the behavioral patterns of sensor nodes, from wireless sensor networks. In this paper, we propose a new kind of behavioral pattern, named Target-based Association Rules (TARs). TARs aim to discover the correlation among a set of targets monitored by a wireless sensor network at a border area. The major application of the Target-based Rules is to predict the location (target) of a missed reported event. Different data preparation mechanisms for accumulating the data needed for extracting TARs have been proposed. We refer to these mechanisms as AI-Node, Schedule-Buffer, and Fused-Schedule-Buffer. Several experiment studies have been conducted to evaluate the performance of the three proposed data preparation mechanisms. Results show that the Fused-Schedule-Buffer scheme outperforms the selected schemes in terms of energy consumption.

Tanbeer et al. [3] The FP-growth algorithm using the FP-tree has been widely studied for frequent pattern mining because it can dramatically improve performance compared to the candidate generation-and-test paradigm of Apriori. However, it still requires two database scans, which are not consistent with efficient data stream processing. In this paper, we present a novel tree structure, called CP-tree (compact pattern tree), that captures database information with one scan (insertion phase) and provides the same mining performance as the FP-growth method (restructuring phase). The CP-tree introduces the concept of dynamic tree restructuring to produce a highly compact frequency-descending tree structure at runtime. An efficient tree restructuring method, called the branch sorting method, that restructures a prefix-tree branch-by-branch, is also proposed in this paper. Moreover, the CP-tree provides full functionality for interactive and incremental mining. Extensive experimental results show that the CP-tree is efficient for frequent pattern mining, interactive, and incremental mining with a single database scan.

Gruenwald et al. [4] In wireless sensor networks, a significant amount of sensor readings sent from the sensors to the data processing point(s) may be lost or corrupted. In this research we propose a power-aware technique, called WARM (Window Association Rule Mining), to deal with such a problem. In WARM, to save battery power on sensors, instead of requesting the sensor nodes (MS), the readings of which are missing, to resend their last readings, an estimation of the missing value(s) is performed by using the values available at the sensors relating to the MS through association rule mining. The paper then presents the performance studies comparing WARM with existing techniques using the real traffic data collected by the Department of Transportation in Austin, Texas.

Tong et al. [5] Study the problem of mining frequent value sets from a large sensor network. We discuss how sensor stream data could be represented that facilitates efficient online mining and propose the interval-list representation.

Based on Lossy Counting, we propose ILB, an interval-based online mining algorithm for discovering frequent sensor value sets. Through extensive experiments, we compare the performance of ILB against an application of Lossy Counting (LC) using a weighted transformation method. Results show that ILB outperforms LC significantly for large sensor networks.

3. System Architecture

Because of late progression in hardware industry, Wireless Sensor Networks (WSNs) is utilized for different applications, for example, security, farming and natural observing. WSN might contain many small, minimal effort, battery-controlled gadgets that screen physical qualities (dampness, temperature, and light) and that self-sort out into systems that can settle on self-sufficient choices (turn on/off actuators), and are a piece of a bigger appropriated administration and control framework (e.g. watering system framework). As every hub is an information source, a sensor system can create extensive arrangements of information, which are perfect possibility for information mining procedures. Notwithstanding, sensor systems are compelled in their capacity to convey their information to a unified preparing server where information mining would regularly happen. Sensors are constrained as far as accessible vitality for transmission, computational force, memory, and interchanges data transfer capacity. Dispersed information mining (DDM) strategies give answers for these limitations by putting parts of the information mining process, for example, information examining, collection, and demonstrating on individual sensors, and also groups of sensors. These exercises and arrangement in the sensor system shift by the kind of information mining being embraced, for example, order, forecast, time arrangement investigation, grouping, and irregularity location.

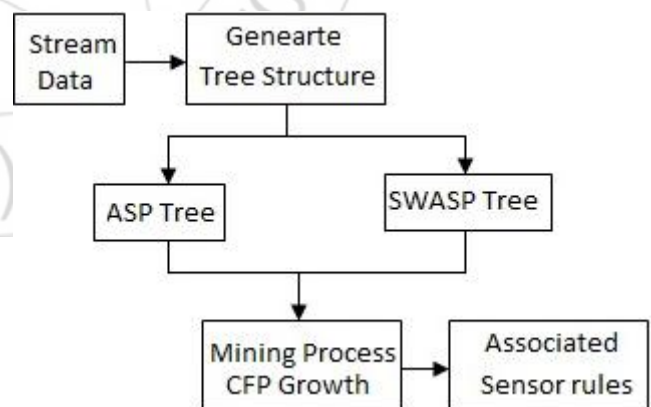


Figure 1: System Architecture

4. Proposed Methods

4.1 ASP Algorithm

Different techniques applied in frequent pattern mining to find candidates and frequent patterns generated. In frequent pattern mining techniques, there are two problems. First is that many times scanned the database and second is more complex candidate generation process for finding frequent patterns. We propose a technique ASPMS Algorithm with

ASPStree. ASPStree algorithm is used to generate associated patterns. ASPMS algorithm can extract associated sensor patterns in the current window with FP growth like pattern-growth method after getting useful information from the ASPStree. ASPMS perform single scan of database and more flexible for transactions. This algorithm can extract useful information from the stream contents for the current window of the sensor in a batch-by-batch manner. ASPStree is based on sliding window that associated sensors pattern for WSN. The nodes of an ASPStree in a sensor appearance order and then restructure the tree in a frequencydescending order. Then finally compress the tree by merging the same support sensor node in a single node in each branch of the tree. By using BSM, it compresses the same frequency nodes into single node. BSM is keep the frequent items in tree after restructure phase. In this method if path is not sorted according to new insertion list order, it is removed from the tree, deleted non-frequent items, sorted according to new insertion list order into a temporary array and then again inserted into tree in order.

4.2 SWASP Algorithm

In sliding window based sensor information stream mining, a window comprises of different nonoverlapping clusters and every bunch contains a non-unfilled arrangement of ages. A case of sensor information stream is appeared, where the information stream of eight ages are isolated into four clusters and three clumps constitute one window. The SWASP tree, as depicted beneath, catches imperative information from the stream content for the present window where each hub in the tree keeps up sensor-id and batchby-clump bolster data. The SWASP tree development continues window by window. Like the ASP-tree, the SWASP-tree development has two stages: insertion stage and rebuilding pressure stage. Moreover, we utilize an upgrade stage to remake the tree developed in the past window as the window propels. SWASP-tree contrasts from DSTree altogether as the last does not have rebuild pressure stage and does not perform recurrence list redesign operation for the hubs which are not went to amid the new approaching bunch, abandoning some excess hubs in the tree structure. This might bring about a DSTree to end up over-burden with hubs that are no more important for the present window. SWASP-tree structure is additionally entirely distinctive than that of CPS-tree. CPS tree development is in view of a predefined thing request (e.g., lexicographic thing arrange) and characterizes two sorts of hubs (customary and tail) while in SWASPtree all hubs are dealt with just as. Also, it doesn't perform pressure. The rebuilding furthermore, pressure in SWASP-tree ensure that SWASP-tree will have less memory prerequisite than DSTree and CPS-tree.

4.3 CFP Growth Algorithm

CFP-growth algorithm has been proposed to overcome the rare item problem. It is defined as, at high minsup value, the frequent patterns with low minsup are missed, and at low minsup value, the number of frequent patterns explodes. It is based on multiple minimum support frameworks The frequent patterns mined with this framework do not satisfy downward closure property and algorithm follows different

types of pruning techniques to reduce the search space. In the CFP Growth, each item in the transaction database is denoted with the support constraint known as minimum item support (MIS) and minsup of an itemset is represented with the minimal MIS value among all its items.

5. Experimental Evaluation and Graphs

The investigations were led with the datasets manufactured information (Intel Lab data) prior too used to think about the execution of example mining in remote sensor system . In this investigations were performed in both thick and inadequate information sets. Table demonstrates the qualities of intel lab datasets. Analyzed the execution of proposed complete from fractional tree approach with our calculation. The calculations were actualized in Java program dialect and keep running with Windows 2007 on a 3.20 GHz CPU and 4 GB memory. The execution of calculations with various information sets is appeared through beneath Fig. . The execution time incorporated the tree creation, era and affiliation guideline era times.

Table 1: Dataset Characteristics

No. of Trans.	No. of Items	MaxTL(MTL)	AvgTL(ATL)
25000	54	130	10.10

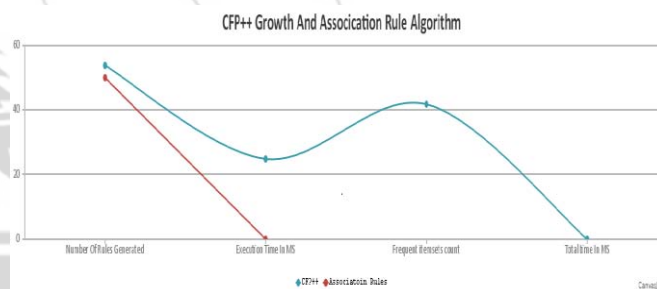


Figure 2: CFP Growth and association rule algorithm

6. Conclusion

In this venture, have presented another sort of behavioral examples called related sensor designs that catch co-events and worldly connections among sensors. To concentrate such examples, we contrived a prefix tree structure called ASP-tree that stores sensor information in a reduced way, and based upon this tree a mining calculation called ASP is proposed which adequately mines related sensor designs over sensor database in one and only output. Since it is not down to earth to keep up all components of sensor information in a tree when information are caught and broke down for a stream, we displayed an upgraded tree structure called SWASP-tree which can proficiently find related sensor designs from stream information in a solitary pass utilizing sliding perception window. Both trees have the assemble once and mine numerous property, making it very suitable for intuitive mining. To streamline asset use, a strategy to powerfully adjust window size is likewise exhibited. Relative execution examinations demonstrate that our strategies are exceptionally powerful and proficient for mining related sensor designs over sensor information and

beat existing calculations in both runtime and memory use, furthermore scale well to handle extensive number of particular sensors and ages.

References

- [1] Boukerche D and S. Samarah, "Novel algorithm for mining association rules in wireless ad-hoc sensor networks," IEEE Trans. Parallel Distributed. Syst., vol. 19, no. 7, pp. 865–877, Jul. 2008.
- [2] S. Samarah, B. Azzedine, and S. H. Alexander, "Target association rules: A new behavioral patterns for point of coverage wireless sensor networks," IEEE Trans. Computer., vol. 60, no. 6, pp. 879–889, Jun. 2011
- [3] S. K. Tanbeer, C. F. Ahmed, and B. S. Jeong, "Efficient single-pass frequent pattern mining using a prefix-tree," Inf. Sci., vol. 179, pp. 559–583, 2009.
- [4] L. Gruenwald and M. Halatchev, "Estimating missing values in related sensor data streams," in Proc. 11th Int. Conf. Manage. Data, 2005, pp. 981–987.
- [5] K. K. Loo, I. Tong and B. Kao, "Online algorithms for mining interstream associations from large sensor networks," in Proc. 9th Pacific- Asia Conf. Adv. Knowledge. Discovery data Mining, 2005, pp. 143–149.
- [6] G. S. Manku and R. Motwani, "Approximate frequency counts over data streams," Proc. 28th Int. Conf. Very Large Data Bases, 2002, pp. 346–357.
- [7] K. Romer, "Distributed mining of spatial-temporal event patterns in sensor networks," in Proc. Eur.-Amer. Workshop Middleware Sensor Networking. Conjunction with Int'l Conf. Distributed. Computer. Sensor Syst., 2006, pp. 103–116 .
- [8] Y. Jennifer, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," Computer. Networking. vol. 52, pp. 2292–2330, 2008.

Author Profile



Priyanka Pandit received B.E. degree in Computer Science in 2012 from Savitribai Phule Pune University and pursuing M.E. from RMDSSOE, Warje, Pune.



Swarupa Kamble is working with RMDSSOE, Warje, Pune as an Assistant Professor. She has experience of 5 yrs in the field of teaching and research and her research interests are Image Processing and Data Mining.