

A Novel Data Preparation Approach for Target Based Association Rule Mining

Prachi Singh, Rajendra Kumar Gupta

Madhav Institute of Technology & Science Gwalior, Gwalior, India

Email: Singh.prachi5@gmail.com, rkgiitm@gmail.com

Received 25 September 2014; revised 20 October 2014; accepted 18 November 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Target based association rules extract the correlation between targets in and around strategically specified regions. These rules can be used to predict and estimate the future results and find the faulty targets, etc. To extract these rules efficiently and effectively data preparation is required. The existing mechanisms of data preparation have the problem of redundancy in the data. As a consequence, extra energy is required for the sensors which used to monitor the targets. In this paper, authors propose a novel data preparation approach for target based association rule mining from the point of coverage wireless sensor networks, which reduces the redundancy in the data and thus enhances the performance of the network.

Keywords

Knowledge Discovery, Point of Coverage Wireless Sensor Networks

1. Introduction

Knowledge discovery process has been proved a very effective technique to find out hidden, previously unknown and useful information (knowledge) from the wireless sensor networks [1]-[4]. It is a complex process which involves various steps such as domain understanding, knowledge definition, data preparation (data cleaning, data reduction, etc.) and finally data mining. Association rule mining has become the one of the popular technique of knowledge discovery which is introduced by Agrawal *et al.* [5] in 1993 to find association among objects from the transactional databases. This technique can also be used to find patterns from wireless sensor networks, which can be used to predict future results and missing values and find faulty nodes. Using this, Samarah *et al.* [6] [7] have proposed a technique that is known as the sensor association rules to find the relation between the nodes in the wireless sensor networks. In this technique the sensor itself acts as the main object, rather than the sensors' readings.

In order to prepare the data required for obtaining sensor association rules, each sensor node should profile its

activity over time and notify to the sink about the timestamps of detected events. This technique uses the meta-data, which describes the sensors' behavior, not the data reported by the sensor about the surrounding environment [6]. The extensions of this technique have been used in target based association rules (TARs) that can be used to enhance the Quality of Service (QoS) of a point of coverage wireless sensor network (POCW) [8].

POCW is a network that consists of a set of targets, a set of sensor nodes and a sink node, as shown in **Figure 1**. In this the sensor nodes deployed in the region, around the targets, which are responsible for monitoring the activities around the targets. These types of networks are generally used in military applications for monitoring the border areas over time. The main objective of TARs is to discover the relationship among a set of targets that are monitored by wireless sensor networks and are used to enhance the Quality of Service of the network, to predict the source of the future events, etc. [8] [9].

To generate TARs efficiently, preparation of Target Database is required. Target Database consists of the collection of activity sets of targets that are sent by the sensor nodes to sink node. This activity data can be gathered using existing routing protocols, or using data collection algorithms. Existing data collection algorithms are not designed for mining purpose and they do not consider the redundancy existing between targets' activity data. That is why special data collection algorithms are required for mining.

Figure 2 shows the main steps of target based association rule mining (TARM). It has six steps. The first three are used to perform data preparation; the fourth and the fifth are used to perform mining, and the sixth one is used to perform analysis on generated patterns.

In this paper authors propose a novel data preparation approach for target based association rule mining from the point of coverage wireless sensor networks, which reduces the redundancy in the data and thus enhances the performance of the network.

This paper is organized as follows: Section 2 explains the preliminaries of target based association rules; Section 3 reviews some of the related works; Section 4 introduces the mechanism for preparing the data to generate the TARs; Section 5 includes the experimental results to evaluate the performance; Section 6 concludes this paper.

2. Preliminaries of Target Based Association Rule Mining

As mentioned in the previous section, target based association rules are based on the common intervals of events occurrences around the targets. To find these associations, time is assumed to be divided into number of time slots, of size λ . Each sensor node is responsible for profile activities of the targets for a given profile time (T_{prof}) in its sensing range. Profiling time is the time period that is determined by the application to profile the activities of the targets. Sensor nodes also have the set of buffers to store the activities of each target. Each buffer will include an entry for each time slot within the given profile time and have size equal to (T_{prof}/λ). Each sensor profiles targets into its sensing range over time, if any event is detected around any target in the current time slot, an entry is set into the buffer in that time slot. A set of slot numbers for which there is a set bit in the buffer, called as the activity set (AS), is then send to the sink node, at the end of the profile time. The sink node, then prepares the target database, T by accumulating the activity sets of each target.

Definition: Target based association rules are defined as $P' \rightarrow P''$ where $P', P'' \subset T$ and $P' \cap P'' = \emptyset$. Here P' and P'' are the frequent patterns of the targets and T is the set of targets in the network. A pattern P is said to be frequent if $\text{Support}(P) \geq \text{Min_Sup}$, where support is given by $\text{Support}(P) = \forall l_i \in P \left| \bigcap AS(T_j) \right|$.

Here $AS(T_j)$ is the activity set of target j . An example of target association rule could be ($T_1 \rightarrow T_2, \lambda, 60\%$). Here T_1 and T_2 are the sets of the targets. This rule says that if an event is detected around targets from T_1 within λ , then there is 60% chance that the event will be detected around the target from T_2 within the same time interval [8].

3. Related Work

In the traditional operational scenario of POCW, all the sensor nodes have to be active all the time. When an event is detected at any particular target then a notification message is sent by the sensors, covering that target, to the sink. Sink maintains a target database, accumulates all the notifications of each target and then store them in the target database for further processing. This mechanism is called as the "All nodes based mechanism" which sounds simple but requires lot of energy to send the notification messages to the sink. Here one more issue is the redundancy, because a target may covered by multiple sensor nodes then, sensors will send multiple redundant notification messages to the sink about the same target. It consumes both extra time and energy [8].

The above issues were addressed by Samarah *et al*, they proposed a mechanism in which a global schedule

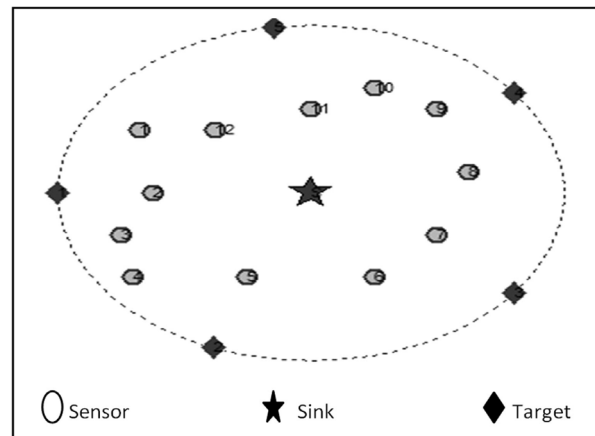


Figure 1. Point of coverage wireless sensor network.

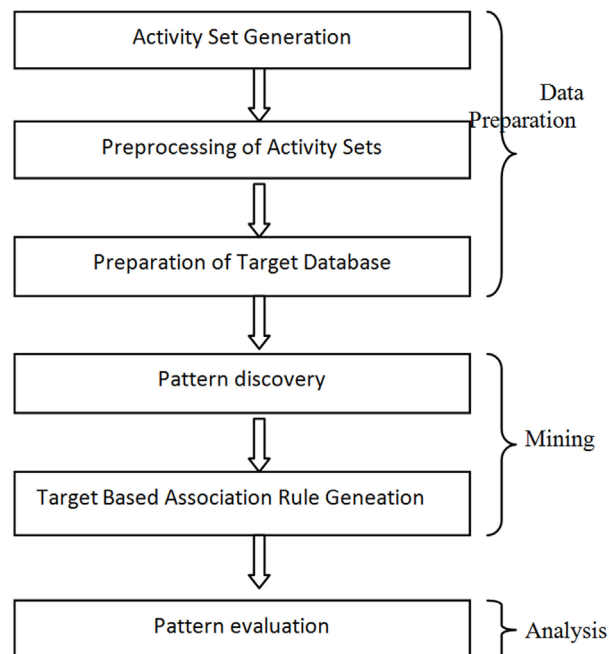


Figure 2. Flow diagram of TARM.

is broadcasted by the sink node which consists of activity time of the sensor node (for which a sensor node should be active for profile the activities) and the set of targets that should be covered by the sensor node. Then the sensor node, during its activity time profiles the targets for activities and prepares a activity set which is a binary bit pattern, “1” for detected event in the particular time slot and “0” for others. This activity set is then sent to the sink. It reduces the required energy to prepare the data for the mining process than the previous one (All-nodes-based mechanism) but there may be the situation where multiple messages have the similar activity sets for different targets in the sensor node. Hence it is the waste of the energy to send the similar or redundant information to the sink for each target by the sensor nodes. In the proposed work this redundancy can be reduces by applying a new parameter refer to as similarity support (sim_sup), on the prepared activity sets which reduces the required energy for data preparation mechanism.

4. Proposed Work

In this section an efficient data preparation mechanism is proposed for target based association rule mining, called as Reduce Redundancy (RR) schedule buffer based data preparation mechanism.

To define this mechanism it is assumed that a global schedule has been defined by the sink. The global schedule specifies the set of targets that should be monitored by each sensor and the activity time of each sensor. Each sensor contains number of buffers to store targets activities, which is divided into number of time slots λ for capture the activity at targets T in given profiling period T_{prof} . During profiling period when an event is detected at a particular target T_i , an entry is set into current time slot of that target's buffer $B(T_i)$. At the end of this profiling period, sensor scans its buffers for each target and prepare the activity set AS_{T_i} . The Activity Set is a set of time slot identifiers for which there is a set bit in the targets' buffer. If the cardinality of activity set is greater or equal to the minimum support (min_sup) then it finds the similarity between buffer entries of targets by performing the bit and operation between the buffers of each target. If the result is greater or equal to the similarity support (sim_sup) then it sends a combine message of targets to the sink which satisfies the similarity support. Otherwise it sends individual message of each target to the sink.

Algorithm 1 gives a formal description for this mechanism.

Algorithm 1. RR-Schedule-Buffer-based-Data-Preparation.

Sink Node:

1. Broadcast global schedule;
2. Broadcast mining parameters (T_{prof} , λ , min_sup, sim_sup);
3. For each received activity set AS_{T_i} ;
4. Insert (AS_{T_i} , TD);
5. End for.

Sensor Node:

1. Upon receiving mining parameters and node schedule;
2. Slot number = 1;
3. Time = current time;
4. While (current time < time + T_{prof});
5. If (current time < time + (slot number * λ));
6. If there is a detected event
 - a. Let;
 - b. Set;
 - c. Else;
 - d. Slot number = slot number + 1;
 - e. End if;
7. End while;
8. For each target T_i within the sensor's coverage;
9. $AS(T_i)$;
10. If cardinality;
11. If
 - a. Send (sensor id, T_i , T_j , $AS(T_i)$) to the sink;
 - b. Else;
 - c. Send (sensor id, T_i , $AS(T_i)$);
 - d. End if;
12. End if;
13. End for.

5. Experimental Results

In this section, we discuss a comparison analysis of the data preparation process for point of coverage wireless sensor network, using schedule buffer based and RR-schedule buffer based mechanisms. Two metrics are used to evaluate performance of the network. First is the number of messages, required to report the activity sets of targets to the sink node, and second is the average energy consumption for each sensor node. These metrics are generated based on a simulator using Matlab version 9. In simulation we have assumed that targets are located at fixed points in a grid 100×100 cm. We have used 10 targets and 25 sensors deployed randomly around the targets. For calculating the energy consumption we have used the first order radio model introduced in [10]. Equations (1) and (2) show the energy consumption for sending k bit message across distance d [11] [12].

$$E_{TX}(k,d) = E_{elec} * k + E_{amp} * k * d^2 \quad (1)$$

$$E_{RX}(k) = E_{elec} * k \quad (2)$$

Here $E_{TX}(k,d)$, represents the energy consumption for sending k bit message across d distance. E_{elec} represents the energy consumption to run the transmitter circuit, which is 50 nJ/bit. E_{amp} represents the energy consumption for the transmitter's amplifier. $E_{RX}(k)$ represents energy consumption to run receiver circuit. We have assumed $k = 10$ bits and $d = 30$ cm. For storage device we assumed a Toshiba 16MB NAND flash memory which costs $0.136 \mu\text{J}$ to read, write and erase a bit [11] [12]. Equation (3) shows the average energy consumption of data preparation for target based association rule mining, in μJ .

$$E_{avg} = E_{TX}(k,d) + E_{RX}(k) + 136 \quad (3)$$

Figure 3 shows the number of messages to report the activity sets to the sink node at 50% minimum support and 50% similarity support over different profiling times. As we can see the number of messages generated by RR-schedule buffer based mechanism is 40% to 60% less than the schedule buffer based mechanism. **Figure 4** shows the number of messages required to report the activity sets to the sink node at 50% minimum support and 70% similarity support over different profiling times. Here we can see clearly that the number of messages generated by RR schedule buffer based mechanism is 60% to 85% less than the schedule buffer based mechanism. **Figure 5** shows the average energy consumption at 50% minimum support and 50% similarity support, over different profiling time. We can see clearly the average energy consumption of RR schedule buffer based is 50% to 60% less than the schedule buffer based mechanism. **Figure 6** shows the average energy consumption at 50% minimum support and 70% similarity support, over different profiling times. We can see the average energy consumption by RR schedule buffer based mechanism is 60% to 90% less than the schedule buffer based

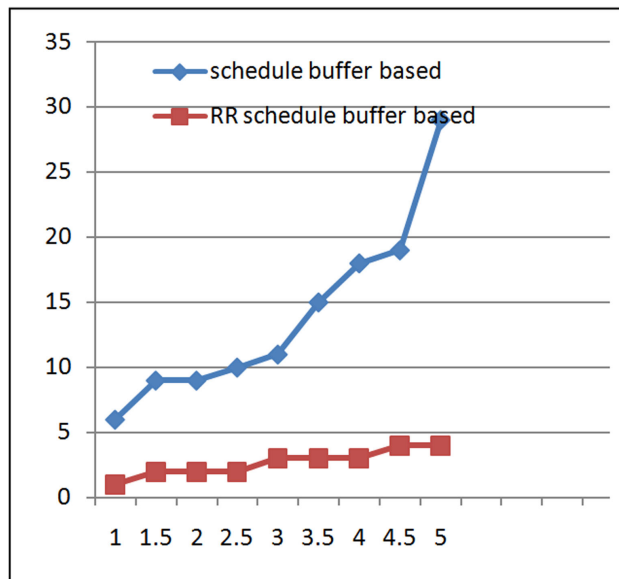


Figure 3. Number of messages (50% min_sup, 50% sim_sup).

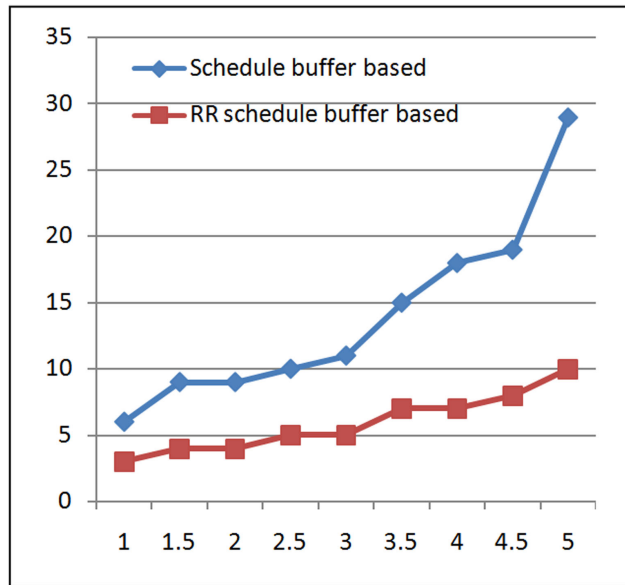


Figure 4. Number of messages (50% min_sup, 70% sim_sup).

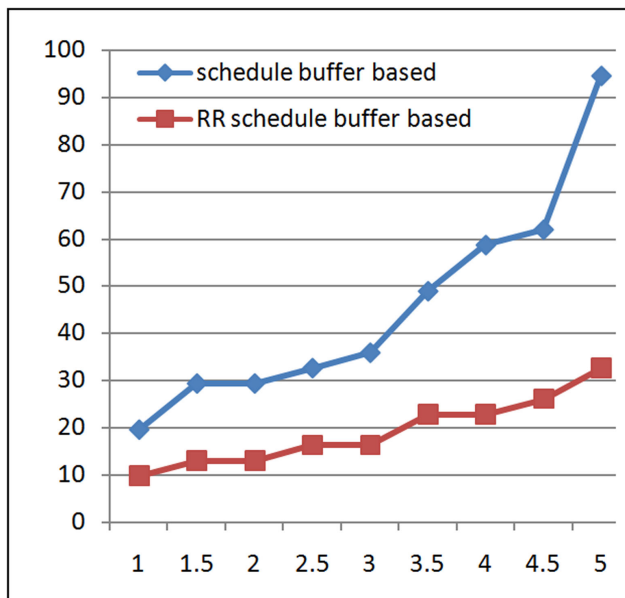


Figure 5. Average energy consumption (50% min_sup, 50% sim_sup).

mechanism.

6. Conclusion & Future Work

Target based association rule mining is a knowledge discovery process which has been proven to be very important and useful in point of coverage wireless sensor networks (POCW). It finds the correlation between the set of targets around a network. In this paper an efficient mechanism is proposed to collect data for mining the target based association rules. Results indicate clearly that the proposed mechanism outperforms existing mechanisms. The target based association rules are very useful for estimation and prediction for future situations and results in a sensor network. Although the proposed mechanism is more effective in terms of reduction in redundancy than the existing mechanisms, it puts an extra load on sensor nodes. Effective distribution of loads on

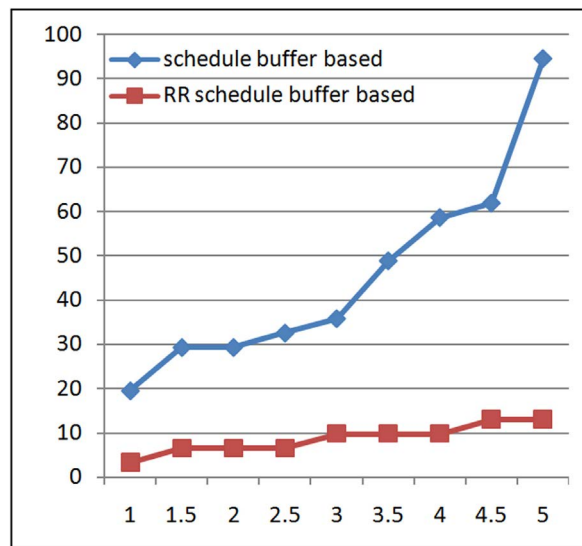


Figure 6. Average energy consumption (50% min_sup, 70% sim_sup).

sensor nodes can be the scope for future work.

References

- [1] Stankovic, S.V., Rakocevic, G., Kojic, N. and Mlicev, D. (2012) A Classification and Comparison of Data Mining Algorithms for Wireless Sensor Networks. 2012 *IEEE International Conference on Industrial Technology*, Athens, 19-21 March 2012, 265-270.
- [2] Halatchev, M. and Gruenwald, L. (2005) Estimating Missing Values in Related Sensor Data Streams. *Advances in Data Management*.
- [3] Romer, K. (2006) Distributed Mining of Spatio-Temporal Event Patterns in Sensor Networks, NCCR-MICS.
- [4] Loo, K.K., Tong, I., Kao, B. and Cheung, D. (2005) Online Algorithms for Mining Inter-Stream Association from Large Sensor Networks. HKU CS Tech Report.
- [5] Agrawal, R., Imielinski, T. and Swami, A.N. (1993) Mining Association Rules between Sets of Items in Large databases. *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, **22**, 207-216.
- [6] Boukerchem, A. and Samarah, S. (2007) A New Representation Structure for Mining Association Rules from Wireless Sensor Networks. University of Ottawa, Ottawa.
- [7] Boukerche, A. and Samara, S. (2008) A Novel Algorithm for Mining Association Rules in Wireless ad Hoc Sensor Networks. *IEEE Transactions on Parallel and Distributed Systems*, **19**, 865-877. <http://dx.doi.org/10.1109/TPDS.2007.70789>
- [8] Samarah, S., Habyalimana, A.S. and Boukerche, A. (2009) Target-Based Association Rules for Point-of-Coverage Wireless Sensor Networks. *IEEE Symposium on Computers and Communications*, Sousse, 5-8 July 2009, 938-943.
- [9] Samarah, S. and Boukerche, A. (2011) Target Association Rules: A New Behavioral Patterns for Point of Coverage Wireless Sensor Networks. *IEEE Transactions on Computers*, **60**, 879-889. <http://dx.doi.org/10.1109/TC.2010.227>
- [10] Heinzelman, W.R., Chandrakasan, A. and Balakrishnan, H. (2000) Energy-Efficient Communication Protocol for Wireless Microsensor Networks. *Proceedings of the 33rd Hawaii International Conference on System Sciences*, 4-7 January 2000, 8020-8030.
- [11] Mathur, G., Desnoyers, P., Ganesan, D. and Shenoy, P. (2006) Ultra-Low Power Data Storage for Sensor Networks. *Proceeding of the fifth IEEE/ACM Conference on Information Processing in Sensor Networks*, Nashville, 19-21 April 2006, 374-381.
- [12] Toshiba 128-MBIT (16M8BITS/8Mx16BITS) CMOS NAND E2PROM.

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or [Online Submission Portal](#).

