# Investigation of Spatio Temporal Associations in Wireless Sensor Networks

T.Abirami Research Scholar,Department of CSE Kongu Engineering College

#### ABSTRACT

Wireless sensor network applications involve discovering patterns from observed events. Generally, prior knowledge of the patterns is not available. The data collected by the sensors are delivered to the sink and offline analyses on the data to extract patterns are conducted. This large volume of data collected affects the performance of the sensor network negatively due to the large communication overhead. The large overhead is a serious obstacle for deploying long lived and large scale sensor networks. In this paper, data mining techniques like Association mining to discover frequent patterns, and their spatial and temporal properties is studied. As the association mining is applied in-network, patterns and not the raw data streams are forwarded to the sink, thus reducing the communication overhead is reduced significantly. In this paper, it is proposed to investigate the association of data received in the sink from various nodes across the network.

#### Keywords

Wireless Sensor networks, data mining, association rules, spatiotemporal patterns.

#### **1. INTRODUCTION**

Wireless Sensor Networks (WSN) are widely used for environmental and security monitoring. Small wireless sensors have the capacity to sense, compute, store and transmit; it integrates with each other to form a network. In a WSN, the sensors monitor the immediate surroundings, and the data is transmitted to a well-equipped node called the Sink. The major constraints faced by the sensor are on resource like energy, memory, computational speed and communication bandwidth. WSN is successfully applied for environmental sensing, industrial monitoring, structural monitoring and so on. Typically, the readings i.e the raw data from every node is delivered to the sink which acts as the gateway for the data to be forwarded from the network to a server. Patterns in the data are analyzed offline, but this results in transmitting a large amount of data through the network leading to communication overhead. This overhead is a serious obstacle for deploying long-lived and large-scale sensor network. To reduce the overhead, data mining techniques is applied in-network to find data patterns in the raw data streams and only the patterns mined at the sensor nodes are transmitted to the sink. The user defines a set of events, and the system discovers the spatio-temporal patterns among the raw data, and it is transmitted to the sink.

Data mining technique such as Association rule mining is applied to discover patterns or associations encoded in the data [1]. Association rule is in the form of  $A \rightarrow B$  where A is the antecedent and B is the consequent and A, B is sets of predicates. The association rule is based on concepts of support and confidence. The support is the probability of an transaction/event in the database containing both the antecedent and the consequent P.Thangaraj, PhD Professor & Head /CSE Bannari Amman Institute of Technology

and the confidence is the probability that a record that contains the antecedent also contains the consequent. If  $I = \{i_1, i_2, ..., i_n\}$  is a set of items, a transaction *T* is a subset of *I*, and dataset *D* is set of transaction. Association rule then means finding rules in the form of

$$R \Longrightarrow i[S,C] \tag{1}$$

where  $R \subseteq I$  and  $i \in I$ , *S* is the support and *C* is the confidence. The support, *supportD*(*X*) of an item *X* in the dataset can be defined as

$$SupportD(X) = \frac{countD(X)}{|D|}$$
(2)

where countD(X) is the number of transactions in *D* containing *X*. The user specifies a minimum support (*min\_sup*) and confidence value (*min\_conf*). An itemset is said to be frequent if its support is greater than the *min\_sup* value specified. Number of algorithms are proposed for discovering association rules from large database [2, 3, 4].

The *apriori algorithm* [2] is on the most popularly used algorithms for discovering association rules. The algorithm first discovers all frequent itemsets  $I_F \subseteq I$  which has a value of support equal to or greater than *min\_sup*. The algorithm merges all the frequent itemsets until no more  $I_F$  are found. On generation of the frequent itemsets, it is split in any possible way into a rule antecedent  $R \subseteq I$  and a rule consequent  $i \in I$  such that  $R \cup i = I_F$  and  $R \cap i = \phi$ . The confidence is calculated for each rule candidate and the rule is output if the confidence is above *min\_conf*.

Spatial data are the data related to objects that occupy space. Spatial data contains topological and/or distance information and is organized by spatial indexing structures. In non-spatial association rule mining, it seeks to find associations among transactions that are encoded explicitly in a database, spatial association rule mining seeks to find patterns in spatial relationships that are typically not encoded in a database but are rather embedded within the spatial framework of the geo-referenced data [5]. These spatial relationships must be extracted from the data prior to the actual association rule mining.

The user defines *maxscope* i.e the upper bounds of distance in which the event occurs and *maxhistory* which is a time frame. The sensors collect event notifications occurring within *maxscope* and keep a history with size *maxhistory* of these events. Association

mining rule is applied on the data collected to discover patterns. Every node mines patterns in the following form:

$$A_{1} \wedge \dots \wedge A_{m} \Longrightarrow E[S, C] \tag{3}$$

where event *E* occurs at node *n* with support *S* and confidence *C* given that antecedents  $A_i$  holds true. Antecedents are in the form of

$$\mathbf{A}_i = \left( E_i, D_i, T_i, N_i \right)$$

Every node sends a subset of discovered patterns to the sink, thus, reducing the communication overhead. The nodes share common time length called epoch. Distances are measured either as Euclidean distances or as number of hops. Spatial association rules describe the implication of one or a set of features by another set of features in spatial databases.

#### **2.LITERATURE REVIEW**

Romer [6] explored the use of association rule mining to discover frequent event patterns in spatio-temporal data collected by WSN. The proposed method would be useful in exploratory settings where only minimum idea of findings exist before the experiment. Challenges in implementing the proposal in sensor network were also discussed. Memory consumption of discovery algorithm and communication overhead were identified as the major issues during implementation.

Verhein et al., [7] provided a comprehensive definition of spatiotemporal association rules (STARs). The support in the spatialtemporal domain was effectively dealt with. Patterns for mobility data, stationary regions and high traffic regions were introduced. The pattern's temporal characteristics of regions and how it is related to STAR was also discussed. Prune techniques for the search space was demonstrated, which lead to efficient mining.

Tao, et al [8] proposed a new algorithm called weighted association rule mining (WARM) for discovering relationships in weighted settings. The incapacity of the traditional association mining model to treat units differently is solved using weights. The major challenge

faced during using weights is the invalidation of downward closure property. The proposed algorithm overcomes this problem by using weighted downward closure property. Experiments showed that the mining results in the weighted settings conform to the expected hypothesis and that the proposed algorithm is also scalable.

K. K. Loo et al. [9] presented a centralized algorithm for mining inter-stream associations from sensor networks. Methods for representing sensor stream data into transactions using either the simple transformation or the weighted transformation was proposed. An interval-list approach was introduced and experiments showed that it is a viable option in representing a massive sensor network.

Boukerche et al., [10] introduced a new formulation for the association rules that is able to generate the time relations between sensor devices in a particular sensor network. The proposed method generated rules which gives clear correlations between sensors in the network and can be used to make decisions about the network performance. The methodology is useful for predicting the sources of future events too. A distributed methodology is used to prepare the data needed in the mining process and to maximize the network lifetime. In this distributed methodology sensors perform optimization based on local computation to decide whether it will participate in sending data or not. Experimental results have shown that the distributed extraction solution is able to reduce the number of exchanged messages and the data size by 50% compared to a direct transmission of the data.

In this paper, it is proposed to investigate spatio-temporal relationship among the wireless sensor nodes using association rules based on prediction. The proposed method uses INTEL dataset consisting of 54 Mica2Dot motes with temperature, humidity and light sensors.

#### **3.METHODOLOGY**

The experiments were conducted using Intel Lab Sensor Data [11]. The dataset consists of sensor data collected for a month from 54 sensor node with epoch duration of 31 seconds. The sensors collected timestamped topology information containing temperature, light, humidity and voltage readings. The dataset consist of 2.3 million epoch readings collected from all sensors. The data is represented as shown in Figure 1.

Date	Time	Epoch	Moteid	Temp	Humidity	Light	Voltage
Yyyy- mm-dd	hh:mm :ss.	int	int	real	real	real	real

Figure 1: The data captured in the sink

In this paper thirty minutes of data collected between first march 2004, 9:00 AM to first March 2004, 9:30AM consisting of over 9000 data messages received in the sink was studied to find the association between the sensor motes. The distribution of the data sent by each mote to the sink is shown in figure 2.



# Figure 2: The distribution of data sent by each mote during the 30 minute interval.

A predictive apriori approach [12] is used to determine the association among the motes in sending data to the sink. In predictive apriori, balance between confidence and support is found to maximize the accuracy of the predictions on unknown data. The predictive apriori algorithm does not have a fixed confidence and support thresholds, but try to find the *n* best rules. The prior  $\pi$  over all association rules with a specific length is estimated, followed by frequent item set generation with dynamic minsup threshold and all the association rules are generated and finally the redundant rules are removed in the predictive apriori algorithm. The algorithm of the predictive apriori [12] is given.

• Enter *n* the desired rules dataset with items

 $a_1, a_2, ..., a_k$ 

• Draw a number of association rules [x ⇒ y] with *i* items at random. Measure their confidence at support greater

than 0. Let  $\pi_i(c)$  be the distribution of confidence

• For all c measure 
$$\pi(c) = \frac{\sum_{i=1}^{k} \pi_i(c) {\binom{k}{i}} (2^i - 1)}{\sum_{i=1}^{k} {\binom{k}{i}} (2^i - 1)}$$

- Let  $X0 = \{\emptyset\}$ ; Let  $X1 = \{\{a_1\}, \dots, \{a_k\}\}$  be all item sets with one single element
- For i = 1 ... k 1 While (i = 1 or X<sub>i</sub>-1 ≠ Ø), determine the set of candidate item sets of length i as Xi = {x ∪ x' | x, x' ∈ X<sub>i</sub> -1, |x ∪ x'|=i}. Generation of Xi can be optimized by considering only

item sets x and  $x' \in Xi-1$  that differ only in the element with highest item index. Eliminate double occurrences of item sets in Xi.

• **Output** *best*[1]... *best*[*n*], the list of the *n* best association rules

## **4.RESULT**

The rules generated with proposed predictive apriori accuracy greater than 70% is shown in Table I. Table II shows the Apriori algorithm with confidence level of 1.

 TABLE I : ASSOCIATION IDENTIFIED IN THE

 PROPOSED METHOD

m3	тб	==>	m2	
m4		==>	m3	
m14		==>	m9	
m16		==>	m9	
m2	m4	==>	m3	m6
m2	m11	==>	m9	
m4	m6	==>	m2	m3
m6	m9	==>	m16	
m6	m10	==>	m7	
m8	m9	==>	m1	m11
m8	m11	==>	m1	m9
m10	m11	==>	m7	
m2	m6	==>	m3	
m3	m7	==>	m2	

Table II shows the rules generated using Apriori with confidence level of 1.

TABLE II: RULES GENERATED USING APRIORI ALGORITHM

m3	m6		==>	m2	
m4			==>	m3	
m14			==>	m9	
m16			==>	m9	
m8	m9		==>	m1	
m8	m11		==>	m1	
m2	m4		==>	m3	
m4	m6		==>	m2	
m2	m4		==>	mб	
m2	m11		==>	m9	
m4	m6		==>	m3	
m4	m7		==>	m3	
m3	m14		==>	m9	
m3	m16		==>	m9	
m6	m10		==>	m7	
mб	m16		==>	m9	
m6	m9		==>	m16	
m10	m11		==>	m7	
m8	m11		==>	m9	
m8	m9		==>	m11	
m14	m16		==>	m9	
m8	m9	m11	==>	m1	
m1	m8	m11	==>	m9	
m1	m8	m9	==>	m11	
m8	m11		==>	m1	m9
m8	m9		==>	m1	m11
m3	m4	m6	==>	m2	
m2	m4	m6	==>	m3	
m2	m3	m4	==>	m6	
m4	m6		==>	m2	m3
m2	m4		==>	m3	m6
m3	m6	m7	==>	m2	

Comparing Table I and II it can be seen that the proposed method is able to discover the associations to make predictive analysis such as node failure, asymmetric links. All associations with accuracy greater than 75% are identified in the Apriori algorithm also.

### **5.CONCLUSION**

In this paper it was proposed to find associations among motes in a wireless sensor network based on the packets received in the sink. Associations based on the received traffic can be effectively used to identify mote failures and link failures. A novel feature extraction method from the datastream was proposed and association among the motes were identified based on the type of data traffic sent was analyzed using predictive apriori algorithm. The proposed method at accuracy levels greater than 75% was able to identify all associations among the motes. Further work needs to be done for larger duration of the streamed data as only 30 minutes of data were considered in this study.

### REFERENCES

- AGRAWAL, R., IMIELINSKI, T., and SWAMI, A., 1993, Mining association rules between sets of items in large databases. In Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 207-216.
- [2] R. Agrawal and R. Srikant. Fast Algorithms for Mining Association Rules. In *VLDB 1994*, Santiago de Chile, Chile, September 1994.
- [3] J. Han, J. Pei, and Y. Yin. Mining Frequent Patterns without Candidate Generation. In *SIGMOD 2000*, Dallas, USA, May 2000.
- [4] Fernando Berzal, Juan C. Cubero, Nicolas Marín, José-María Serrano, "TBAR: An efficient method for association rule mining in relational databases," *Data & Knowledge Engineering*, Vol. 37, No. 1, 2001, pp. 47-64.
- [5] K. Koperski and J. Han. Discovery of Spatial Association Rules in Geographic Information Databases. In Proc, 4<sup>th</sup> International Symposium on Large spatial databases, pg:47-66, Portland, Maine, Aug 195.
- [6] Roemer, K.: Distributed mining of spatio-temporal event patterns in sensor networks. In: Proc. of the 1<sup>st</sup> Euro-American Wkshp. on Middleware for Sensor Networks (EAWMS). (2006)

- [7] F. Verhein and S. Chawla, "Mining spatio-temporal association rules, sources, sinks, stationary regions and thoroughfares in object mobility databases," in DASFAA, 2006.
- [8] F. Tao, F. Murtagh, and M. Farid. Weighted association rule mining using weighted support and significance framework. In Proc. 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 661–666, Washington DC, 2003.
- [9] K. K. Loo, I. Tong, B. Kao, and D. Cheung, "Online Algorithms for Mining Inter-Stream Associations From Large Sensor Networks," In: PAKDD, 2005.
- [10] Boukerche, A.; Samarah, S.; , "An Efficient Data Extraction Mechanism for Mining Association Rules from Wireless Sensor Networks," *Communications, 2007. ICC '07. IEEE International Conference on*, vol., no., pp.3936-3941, 24-28 June 2007
- [11] Intel Lab Sensor Data. http://berkeley.intelresearch.net/labdata/.
- [12] T. Scheffer. Finding association rules that trade support optimally against confidence. Intelligent Data Analysis, 9(4):381 – 395, 1995.