International Journal of Computer Science and Engineering Communications,
Volume.5, Issue.3 (2017): Page.1593-1599
www.ijcsec.com

Mining Associated pattern from wireless Sensor Networks: A Survey

Surya V Sudhakaran[1] and Vigilkumar V.V[2]
[1]PG Scholar, [2]Assistant Professor
[1,2]Department of Information Technology
[1,2]Govt.Engineering College Idukki, India
suryavsudhakaran@gmail.com[1],vigil.lit@gmail.com[2]

Abstract: *Wireless sensor network (WSN) is a vibrant technology where data management and processing is a topic in several research areas. The main objective of WSN based applications is to make real time decisions. This can remove different challenges like resource-constrained computing, communication capabilities in WSN. These challenges motivate to extract useful data from large set of data base or data streams. There are numerous algorithms are used for mining knowledge from sensor stream. These algorithms are mine frequent pattern from WSN. Here a comparative study of existing data mining techniques and their objectives are presented.*

Keywords *- WSN, data mining, frequent patterns, sensor stream.*

I. Introduction

WSN is an evolving investigation area for different monitoring and surveillance systems such as area monitoring, environment monitoring, industrial and machine health care monitoring and military surveillance. WSNs sense the data as streams, streams are continuous and unbounded. Data mining techniques have been used to extract useful information from WSN data streams. There exist different problems in mining correlated pattern from stream data such as, data stream is continuously arrives with great speed and there is a massive amount of data. There is no sufficient space to store all the stream data and not enough time to scan the whole database. Through determining correlations among the sensor nodes, which is to be considered as the useful knowledge and it is used for future purpose. It find out the distributed nature and limited resources of sensor nodes which cause the delay, predict the source of future events, undesirable side effect of communication and it also finds which nodes can be moved to sleepy mode.

Different algorithms are used to find correlations or patterns from huge databases or data stream. The pattern is called frequent if it occurs many times in the transaction. The main tasks in mining associated patterns are: preserving the downward closure property, provide a compact tree structure that is able to capture the data content in one scan over sensor datasets, and by making the tree structure adaptive it can captures the latest information effectively. We examined how data mining algorithms and techniques will be essential to make the sensor network applications intelligent. The schematic representations of WSN are shown in Fig. 1.This paper focused a review on certain data mining techniques and a comparative study of different algorithms. This is used to identify the relevance of sensor pattern mining from continuous data streams.

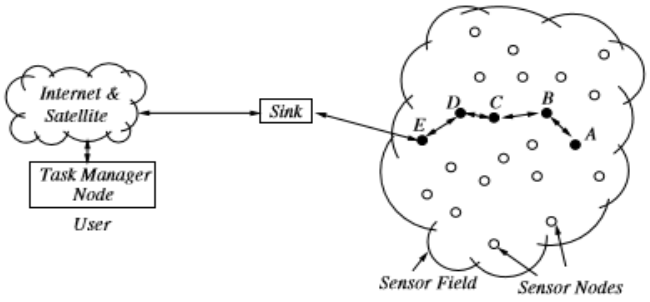

Fig.1. Simple Architecture of WSN [11]

The remaining section of this paper is structured as follows. In Section II, we represent the literature survey. Section III discusses the comparative study. In Section IV Fundamental Challenges of Data Mining in WSNs. Finally, conclusion about the survey is presented in Section V.

II. Literature Survey

Researchers give more importance towards mining frequent and correlated pattern from unbounded, continuous data stream. In this literature survey we focus on algorithms and different techniques related to frequent patterns, sensor database and data streams.

Syed Khairuzzaman.*et.al*. [1] proposed that finding frequent patterns from continuous stream of sensors is very difficult. Numerous frequent pattern mining algorithms are developed for mining process. WSN sense the data in the form of streams, so mining from stream is very important task. So proposes an efficient method for finding frequent items over the data streams. So uses a compact pattern stream tree (CPStree) to capture the latest stream, which removes the old stream data. The whole set of frequent pattern is obtained from CPS tree using FP growth mining technique. FP growth technique requires two database scans to achieve a highly compact tree structure. It uses sliding window method to update the recent data.

Data stream tree (DSTree) structure capture important data from the streams. Leung.*et.al*. [2] proposed that by exploiting properties of DSTree, it can be easily conserved and mined for frequent itemsets and other patterns. Several stream mining algorithms are there, which is categorized into two classes; exact algorithm and approximate algorithm. Exact algorithm find truly frequent itemsets and approximate algorithm find frequent itemsets by using approximate procedures. Which mine the frequent itemsets similar to FP growth.

Joong Hyuk Chang *et.al.* [3] proposed a system in which a sliding window technique to find the newly frequent itemsets over data streams. It uses a prefix tree structure, from that tree capture the frequent pattern and store the mining result in a data structure called D. D consist of set of items, count of itemsets and maximum possible error count. If the local itemsets identified by the new transaction in the buffer, its previous count is incremented. Sliding window frequent (SWF) algorithm uses sliding window to find frequent itemsets. Online mining of frequent itemsets by using stream sliding window is one of the most important difficulties in stream data mining.

The system proposed by Yin Lee *et.al.* [4] used an effective bit-sequence based, one-pass algorithm, called MFI- TransSW (Mining Frequent Itemsets within a Transaction sensitive Sliding Window), to mine the set of frequent itemsets from data stream. MFI-TransSW algorithm consists

of three stages: window initialization, sliding and pattern generation. First, all transaction is encoded to bit sequence representation in the window initialization phase. Second, MFI-TransSW uses the left bit-shift technique to slide the windows efficiently in the window sliding. Finally, the complete set of frequent itemsets within the current sliding window is generated by a level-wise method in the pattern generation phase.

Iqbal Gondal *et.al*. [5] Proposes a new behavioral pattern called associated correlated sensor pattern.Here uses a prefix tree based structure called ACSP tree.Here all confidence and corr-confidence disclose true temporal relationship among data objects.FP growth technique is used to generate the pattern. The frequent patterns are discovered in two steps: in first step mine all frequent pattern using ACSP tree.In second step, test whether they are associated or not. Young Kim *et al.* [6] proposed two algorithms, CoMine alpha and CoMine gamma. Here the database consists of set of transactions and each transaction consists of set of items. Here mining of correlated pattern done in two aspects: First, several popularly used measures related to correlation mining are reexamined. Second, develop two algorithms. Compute the size of the universe using the FP-tree.

Mamunur Rashid *et al.* [7] proposed that binary frequency of a pattern only reflects the number of epochs which contain that pattern in the dataset. Data mining techniques have been used to extract useful knowledge through discovering relationship among sensor nodes ie, behavioral pattern. Here consider a shared frequent sensor pattern (SFSP), which is useful in fine grain monitoring of physical environment. A parallel and distributed approach is used to execute data mining tasks concurrently. Mamunur Rashid *et al.* [8] proposed an associated sensor pattern stream tree (ASPS tree).Which uses sliding window method with FP growth mining techniques. ASPS tree has two phases: insertion phase and reconstruction compression phase. After getting pattern from ASPS tree, build conditional pattern base and conditional tree. ASPS tree can capture the recent data from datastream.

Kay Romer [9] presents a distributed approach to find both spatial and temporal correlated pattern. The raw data is collected at the sink node for later offline analysis. Here after the mining compact pattern is sent to the sink. To limit the potentially huge search space for the discovery of such interesting patterns, we apply a constrained form of data mining. Iqbal Gondal *et al.* [10] proposed that WSN generate large amount of data in the form of streams. Here mining a new type of behavioral pattern called associated sensor pattern, which capture co-occurrences and temporal correlations. Pattern extraction will improve WSN performance. To mine these patterns ASP mining algorithm is used with pattern growth approach.

III. Comparative study

Different algorithms and tree structures for mining frequent pattern from sensor data streams are refers in Table I. The major problem of these approaches is, they only consider the temporal correlation to find the frequent itemset. Table II represents the comparison between different parameters, data types, attributes etc.

Table I. Comparative Study

Ref No	Objective	Tree and algorithm	Advantages	Disadvantages
[1]	Mining of frequent pattern	Sliding Window algorithm	1) Maximum possible error should be decrea-sed as the window slid.	Transaction in the current window should be scanned many times.
[2]	Mining of regular frequent pattern	Data Stream Tree	1) Keep a list of frequ-ency count. 2) No false positives or false negatives.	Leave some garbage nodes, when mining delayed. Massive storage overhead.
[3]	Mining of frequent pattern	Compact pattern stream tree(CPS tree)	1) Constantly updates itself by extracting the expired transactions. 2) The tree will not contain garbage nodes.	Keep the window size unchanged even though the dataflow rate changes.
[4]	Mining the correlated pattern	MFI-TransSW algorithm	1) The errors of outputs should be minimum. 2) Dynamically mainta-ining all transactions	Candidate generation and test problem.
[5]	Mining the correlated pattern	Associated correlated sensor pattern (ACSP Tree)	1) No candidate generat- ion and test problem.	Inefficient for handling high speed data.
[6]	Mining of correlated pattern	CoMine Algorithm	1) Reduce the search space	Not suitable for data stream. Cannot keep batch by batch information
[7]	Mining of associated correlated pattern	Share frequent sensor pattern tree (SFSP tree)	1) Efficient load balan-cing and memory usage.	Centralized storage.
[8]	Mining of associated correlated pattern	Associated Sensor Pattern Stream tree (ASPS tree)	1) Satisfy both null-invariance and downw-ard closure properties	High tree construction cost
[9]	Mining of frequent pattern	Association-rule mining	1) Find both spatio-temporal pattern	Most of them are fail to reflect the true relation among data
[10]	Mining of associated correlated pattern	Associated Sensor Patterns tree(ASP tree)	1) Provide fast mining. Window size is constantly changing	Time Consuming

Table II. Comparison of Data Mining Technique for WSN

Research Work	Correlation		Parameters				Attributes		Data Type	
	Attribute	Temporal	Corr - confidence	Error	All-confidence	Support	Homogeneous	Heterogeneous	Static	Data Stream
A Sliding Window Method for Finding Recently Frequent Itemsets over Online Data Streams [1]	√			√		√		√		√
DSTree: A Tree Structure for the Mining of Frequent Sets from Data Streams [2]	√					√	√			√
Sliding window-based frequent pattern mining over data streams [3]	√					√	√			√
Mining frequent itemsets over data streams using efficient window sliding techniques [4]	√					√		√		√
ACSP-Tree: A Tree Structure for Mining Behavioral Patterns From Wireless Sensor Networks [5]		√	√		√	√		√		√
CoMine: Efficient Mining of Correlated Patterns [6]	√				√	√	√		√	
Share-Frequent Sensor Patterns Mining from Wireless Sensor Network Data [7]	√				√			√		√
Mining Associated Sensor Patterns for Data Stream of Wireless Sensor Networks [8]		√			√	√		√		√
Distributed Mining Of Spatio-temporal Event Patterns In Sensor Networks [9]		√				√	√			√
Mining Associated Patterns from Wireless Sensor Networks [10]		√			√	√		√		√

IV. Basic Challenges of Data Mining in WSNs

Based on some reasons data mining techniques in WSN has some challenges.

i) Data mining in WSN:- Data mining in wireless sensor networks is the method of extracting useful information from continuous and unbounded data streams. Streams are arrives continuously in high speed with huge amount. Normally the data mining techniques are deals with static data with and use multiple scanning to the data base. So the conventional data mining techniques are not suitable for the large amount of stream data.

ii) Resource Constraint:- The sensor nodes are resource constraints in terms of power, memory, bandwidth, and computational power. The main challenge for WSNs is to satisfy the mining accuracy requirements.

iii) Huge data arrival at high speed:- In many domains sensor data is arrived at very high speed and they are unbounded. The spatial and temporally correlated pattern plays an important role in data mining.

iv) Online Mining:- Most of data mining techniques are deal with offline data. Thus a challenge for data mining techniques is how to process distributed streaming data online.

v) Data Transformation:- Sensor nodes are limited in terms of bandwidth, transforming original data over the network is not feasible. The challenge for data mining technique is how to efficiently represent data and discovered patterns over network for transmission.

vi) Dynamic Network Topology:- Sensor nodes are dispersed in many areas. Moreover, sensor nodes may move between different locations at any time. Such dynamicity increases the difficulty of designing an appropriate data mining technique for WSNs.

V. Conclusion

The above survey aimed at studying different papers about pattern mining. Various papers were studied, most of them deal with mining of associated pattern from WSN. From the database, frequent pattern is extracted from the raw data stream. Time and space consuming is the major problem in pattern mining approach. Most of the systems only consider temporal correlation to discover frequent pattern. To find both spatial and temporal correlated pattern use a parallel distributed approach with FP growth technique.

Acknowledgment

I would like to take this opportunity to express my gratitude to my guide Prof. Vigilkumar V.V, for his exemplary guidance, valuable feedback and constant encouragement throughout the duration of the project. His valuable suggestions were of immense help throughout my project work. His perceptive criticism kept me working to make this project in a much better way.

REFERENCES

[1] Joong Hyck Chang, A Sliding Window Method for Finding Recently Frequent Itemsets over Online Data Streams, *Journal 20, 753-762* (2004).

[2] Byeong-Soo Jeong, Sliding window based frequent pattern mining over data streams, *Information Sciences 179* (2009) 38433865.

[3] Hua-Fu Li,Suh-Yin Lee, Mining frequent itemsets over data streams using efficient window sliding techniques, *Expert Systems with Applications 36* (2009).

[4] Md. Mamunur Rashid, Iqbal Gondal, ACSP-Tree: A Tree Structure for Mining Behavioral Patterns From Wireless Sensor Networks, *IEEE Conference on Local Computer Networks*.

[5] Md. Mamunur Rashid, Iqbal Gondal, Share-Frequent Sensor Patterns Mining from Wireless Sensor Network Data, *IEEE Transcactions* DEC 2015.

[6] Md. Mamunur Rashid, Iqbal Gondal, Mining Associated Sensor Patterns for Data Stream of Wireless Sensor Networks, *Barcelona, Spain*, November 2013.

[7] Md. Mamunur Rashid, Iqbal Gondal, Mining Associated Patterns from Wireless Sensor Networks, *IEEE Transactions on Computers, Vol. 64*, July 2015

[8] Carson Kai-Sang Leung (2006) DSTree: A Tree Structure for the Mining of Frequent Sets from Data Streams, *Proceedings of the Sixth International Conference on Data Mining.*

[9] Young-Koo Lee,Won-Young Kim, Y. Dora Cai, CoMine: Efficient Mining of Correlated Patterns.

[10] Kay Romer, Distributed Mining of Spatio-temporal Event Patterns In Sensor Networks. *Institute for Pervasive Computing*.

[11] Wireless Sensor Networks, National Instruments, http://www.ni.com/white-paper/7142/en/