International Journal of Engineering & Technology, 7 (4.10) (2018) 436-439

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



Discovery of Knowledge Using Association Rules in Wireless Sensor Epocs-a Survey

R.M.Rani ,M.Pushpalatha

Department of Computer Science,SRM Institute of Science and Technology,Tamil Nadu,India.
Department of Computer Science,SRM Institute of Science and Technology,Tamil Nadu,India.
**Corresponding author E-mail: rani.r@rmp.srmuniv.ac.in , pushpalatha.m@ktr.srmuniv.ac.in*

Abstract

Data mining and knowledge discovery in huge data streams have recently involved in more applications used for decision making. Currently in wireless sensor networks, various mining techniques are used to discover knowledge on sensor data. Applying mining algorithm in wireless sensor data faces many challenges such as continuous arrival of sensor data, fast and huge data arrival, changes of mining results over time, online mining, data transformation, changing network topology, resource constraints and have emerged into various research problems. In Wireless Sensor Database, this paper presents a review on various approaches of association rule mining algorithms using various techniques forming sensor association rules generating frequent patterns to find upcoming sensor events or sensor fault detection or to estimate the missing sensor readings.

Keywords: *Wireless Sensor Networks,Association Rule Mining, Candidate generation, Frequent Pattern growth, Associated Sensor Patterns.*

1. Introduction

In WSN a huge number of sensor nodes are kept in the various environments to sense the events and detected events are sent to sink node periodically thus generating a stream of data. WSN applications produce large volume of active, distributed and various types of data. This data is efficiently analyzed and converted into meaningful information by applying data mining. The meaningful information is used for decision making in the respective fields. In data mining process, sensor-association rules were used to improve the Quality of service(Qos) in WSN by detecting the faulty sensors, delay in sensor transmission, estimating the missing sensor values. An example of sensor association rules could be (s2, s3→ s4, 80%, "T") which means that if sensor s2 and s3 senses events within time "T" interval, then there is a probability of 80% that s4 detects event within same time interval. If the sensor association rule (s2,s3→s4) is a frequent pattern then if s4 does not occur in the future received patterns then it is predicted that s4 may be a faulty sensor.

Two main approaches classified in Association rule mining are the candidate generation approach and the pattern growth approach which are combined with various techniques and used to generate the frequent patterns. Those frequent patterns from sensor database are used to extract various types of knowledge such as detecting faulty sensors, estimation of missing sensor readings and predicting the source of future events. This paper is reviewed on various Association mining approaches using different techniques on two type of database which are transactional database, wireless sensor database and also focused on objectives, strength, limitations of various algorithms.

2. Candidate Generation Approach

The candidate generation approach is used in Apriori algorithm where it uses several scans of database to create frequent patterns in consequent stages. In each stage 'n' the candidate sets generated in previous stage 'n-1'are used to generate candidate sets for next stage'n' and finally the frequent patterns are identified. Frequent n-itemsets is defined as the itemset which has n items whose frequent count is more than specified threshold value. Initially the frequent 1-itemsets are the candidate set in the first stage and in next stage candidate sets of frequent 2-itemsets are generated by applying join operation of frequent 1-itemsets whose frequent count is greater than or equal to minimum support threshold value. This process is repeated until the generated frequent n-itemsets and frequent n-1 itemsets become equal.

Let an itemset be ITEM={I1,I2,…In}. Let DB be a database consists of transactions each transaction TR is a nonempty itemset such that TR is a subset or equal to ITEM which is identified by the identifier TID. The strength of association rule is measured by two evaluation parameters 'support' and 'confidence'. An association rule is denoted by Q=>R where Q, R are two itemsets which are subset of ITEM and not equal to empty set. The support 's' is a percentage of transactions in DB that contain Q Union R and denoted by the rule Q=>R. The confidence 'c' is the percentage of transactions in DB containing Q that also contain R. The support is considered to be probability of Q Union R and confidence 'c' is conditional probability of (R\Q). The association rule should satisfy both minimum support threshold and minimum confidence threshold value defined by the user. This approach is combined with various techniques in wireless sensor database. The limitations of this approach are it requires multiple database scan, generates more number of association rules and maximum number of them are not informative. This approach is only suitable for offline

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

mining on large transactional data base and not suitable for data stream mining.

2.1. Apriori

In [1]R.Agarwal at el applied Apriori Algorithm on transactional database to generate frequent patterns using candidate generation approach. The major disadvantages were the algorithm uses multiple numbers of database scans which leads to have more execution time. The objective is to reduce the memory usage, the candidate set which is not required for further stage are moved from memory to secondary storage. The limitation of this technique is only suitable for offline mining on large transactional data base and not suitable for wireless sensor data since the new incoming sensor data were received continuously and so the result may be changed which was based on previously arrived data. In [2] Agrawal applied the technique Apriori-id along with Apriori to avoid the database scan after the initial pass for counting the support value of candidate sets and thus the identification of the candidate itemsets are generated with the help of previous pass . The advantage is that the size of the previous stage candidate itemsets is smaller than the database and also database scan is done only for one time.

2.2. Incremental

In [3]Gurmeet Motwani Singh et al proposed incremental algorithm applied only for transactional database stored in dataware house. Whenever there is an updation on datawarehouse, the incremental algorithm is executed on only updated data. So this algorithm does not required repeated rescans of the whole database and mining applied only on streams of incomming transactional data into dataware house. The major drawback is not suitable for sensor data where the data arrives very fast and huge.

2.3. Window Association Rule Mining (Warm)

In [5] M.Haltchev et al adapted a centralized methodology and modified Apriori algorithm with sliding window concept called Window Association Rule Mining(WARM) for estimat-ing the missing sensor reading values. The sliding window concept technique used data cube data structure for storing the incoming stream of sensor data. The disadvantages of this methodology is it occupied more memory space and take more time to compute the estimated value, but it achieved better accuracy of the estimated value when compared with other methodologies.

2.4. In-Network Mining

In [6] Romer used in-network data mining technique to generate event patterns by applying distributed mining. In this methodology the events were collected in each sensor node and mining algorithm was applied in the node itself to find frequent patterns. The drawback of this methodology increased the buffer cost and processing power in each sensor node and also communication overhead during event collection. Also it is applied only for limited maximum distance sensor nodes and maximum time bound. The local frequent patterns from each sensor node is collected at sink node and mining is applied to generate global frequent patterns.

2.5. Genetic Algorithm and Fuzzy Logic Association Rule Mining.

In [9] T.Abirami et at proposed an genetic algorithm and fuzzy logic association rule mining on WSN. In this algorithm the sensor data were collected from neighbour sensor nodes and mining algorithm is executed in the same node to find the frequent data patterns. It greatly decreased the communication overhead. The major drawbacks are that each sensor node required more computational power and buffer.

2.6. Data Extraction Distributed (Dde)

In [10] Azhar Mahmood proposed a Distributed Data Extraction(DDE) which extracted the sensor data from network. The pre-processing methods are applied on sensor network such as extraction , reduction and each cluster head receives the data with reduced data size from its sensor nodes. The association rule mining is applied on cluster head and then frequent patterns are transmitted to sink. This technique maximized the data accuracy and data are extracted efficiently. This technique is used to calculate the missing sensor values. The disadvantages are increased buffer cost, sensor computational power due to the application of pre-processing technique on the sensor nodes.

2.7. Novel Association Rule Mining.

In[16] Anjan Das proposed an algorithm where frequent patterns are mined on sensor nodes connected in a network and finally send to sink. Whenever the sensor triggers in the particular slot time, then the corresponding time slot-id is added to the list which is maintained by the sensor. At the end of the time period, the frequent count of each sensor occurrence can be calculated using the list created in each sensor. For example if sensor 's1' creates the list [1,2,3,5], then it indicates that s1 has triggered at timeslots 1,2,3,5 and thus frequent count sensor 's1' is 4. Similarly all the sensors connected in the network, its frequent count is calculated for each time period. Once the frequent sensors are identified, the frequent patterns are generated through interaction between the frequent sensors. The strategy depth first search is applied to minimize the computing power, storage and energy of sensors connected in a network.

3. Frequent Pattern Growth Approach

The frequent pattern approach used a frequent pattern tree(FP-tree) data structure to mine frequent item sets without the help of candidate item sets. This approach uses divide and conquer strategy by compressing the database into frequent pattern tree(FP tree) and this tree holds the item set related details. Then the compressed database is divided into conditional databases each associated with pattern segments. The associated datasets with each pattern segments are mined separately. This approach when compared with candidate generation approach, may greatly decreased the size of item sets in the search space while examining the pattern growth and scans the whole database for only one time. This type of approach is followed in many algorithms with various techniques which uses Closed item-sets based association rules mining (CARM), Positional Lexicographic Tree (PLT), Fd-tree, Sensor Pattern(SP-tree), Associated Sensor Pattern(ASP),Total From Partial(TFP), Sliding Window ASP(SWASP), Adaptive SWASP(ASWASP), Share Frequent Sensor Pattern(ShrFSP),Parallel shrFSP(PshrFSP) tree on WSN.

3.1. Closed Item-Sets based Association Rules Mining (CARM)

In[7][8] N. Jiang proposed a data estimation technique called CARM(Closed item-sets based Association Rules Mining) which derives frequent subsets from the initially generated frequent item sets that are stored in memory to avoid the re-occurence of frequent item sets. This algorithm used tree data structure to store the current and updated closed item sets and derived the association rules which are related to the current data between the sensors in the present sliding window. This technique calculated the missing values by discovering the relationship between the sensors.

In[12] J.Han et al proposed the FP-growth algorithm used a frequent pattern tree(FP-tree) data structure to mine frequent itemsets without candidate generation in transactional database. It is used for mining both long and short frequent patterns.

3.2. Positional Lexicographic Tree (PLT)

In [13] Bourkerche at al proposed Positional lexicographic tree structure (PLT)where it used sensor id's to generate the patterns instead of their sensor values. The drawback of this tech-nique required two database scans for constructing tree structure and the sensor id's were converted to a vector which required extra mapping mechanism. This technique consumed more processing time when compared to SP(Sensor Pattern) and processing time was de-creased when support value increases. This technique is used to estimate the value of missing sensor values and used to identify faulty nodes.

3.3. Sensor Pattern (Sp)

In[15]S.K. Tanbeer adapted a tree-based data structures called Sensor Pattern(SP-tree) which generates association rules with one database scan. The advantages of this technique were less memory consumption and run time which outperformed PLT. The disadvantage was SP occupied more memory and runtime when compared to ASP and SWASP. In [17] Manisha proposed an algorithm Total From Partial which generates entire association rules from sensor data and give better performance in sparse data set.

3.4. Associated Correlated Sensor Pattern (ACSP)

In[18] Md.mamunar Rashid at el proposed sensor behavioral pattern called associated-correlated sensor patterns which capture not only association like co-occurrences but also the substantial temporal correlations implied by such co-occurrences in the sensor data. This technique used a prefix tree-based structure called associated-correlated sensor pattern-tree (ACSP-tree), which facilitates frequent pattern (FP) growth-based mining technique to generate all associated-correlated patterns from WSN data with only one scan over the sensor database.

3.5. Associated Sensor Pattern (Asp)

In [19]Mamunur Rashid et at proposed an algorithm ASPT (ASSOCIATED SENSOR PATTERN TREE) where it has used associated sensor pattern (ASP) mining algorithm along with on highly compressed tree structure which was efficient for discovering associated sensor pattern. This associated sensor patterns captures association like co-occurrences and the strong temporal correlations implied by such co-occurrences in the sensor data. It required only single data base scan and so the run time of ASP was high when compared with cominer++ for different sets of data. It occupied less memory because it utilized highly compressed tree structure. It was also noticed that there were linear increase in execution time with increase in database size. The major drawback of this algorithm is that data stream mining cannot be achieved because it does not use sliding window concept. The memory requirement of ASP was less compared with FPT,PLT.

3.6. Sliding Window Association Sensor Pattern (Swasp)

In [20]Mamunur Rashid et at enhanced SWASP(Sliding Window association sensor pattern) with an algorithm using overlapped sliding window concept where the sensor patterns are generated on currently receiving sensor data stream. The author compared the run time of ASP with comine+ , SWISPP where it required only the run time for different minimum all-confidence value. The window size can be changed dynamically using the utilization factor called Adaption SWASP (ASWASP) where the window size varies on incoming flow of sensor data. SWASP occupies less memory compared to PLP, FP tree and SP tree**.**

3.6 .Share-Frequent Sensor Patterns (Sfsps)

In [21] Md.Mamunur Rashid at el proposed a new behavioral pattern called share-frequent sensor patterns(SFSPs) using trigger values of sensors epochs. The algorithm used sensor trigger values from each epochs to calculate measure values of sensor set to find the share frequent sensor patterns. This method used a share-frequent sensor pattern tree (ShrFSP-tree) to improve pattern growth mining technique. This share-frequent sensor pattern tree (ShrFSP-tree) is also enhanced as PShrFSP-tree which can be used in Parallel and distributed heterogeneous processing environment.

4. Other Approaches

The approaches discussed in the previous sections are belonged to horizontal partitioning. The other approaches also uses vertical formats which uses DFS with less complex data structures and takes less computing time when compared with frequent pattern approaches. This vertical format approaches are applied for transactional databases and also used in data streams with sliding window technique. In vertical format the scalability shall be achieved by including more number of sensors in transposed database.

5. Conclusion and Future Work

In this paper the various existing association rule mining based on the approaches are reviewed. The Association Rule Mining in wireless sensor data not only discover all the interesting relationships in large amount of data, and also used for estimating missing sensor values , identifying faulty sensors and predicting future events. The existing algorithms can also be classified based on the application of sensor association rules, various approaches, and techniques applied which are represented in the Table 1 and Table 2. The pros and cons of various existing association rule mining algorithms for wireless sensor data and transactional database are discussed in this paper. The two main approaches in the Association Rule mining algorithms are the candidate generation approach and the pattern growth approach with various techniques are discussed and focussed on objective, strength and limitations. The researchers can also apply mining technique on the sensor behavioural data to extend the lifetime of sensors and thus improving Quality of service(Qos). The researchers can improve its efficiency by contributing some new techniques in mobile wireless sensor network to overcome the limitations of the existing approaches.

Table 1: Candidate Generation Approaches

Ref.No	Year	Candidate Generation Approaches	Objective
[1][2]	1994	APriori	Discovery of Association rules in transactional database
[3]	2002	Incremental	Generate Sensor Association rules
[5]	2005	Window Association Rule Mining(WARM)	Estimate the missing sensor values
[6]	2006	Distributed mining of spatial Temporal event patterns	Generate frequent event patterns.
[9]	2012	Fuzzy based genetic algorithm	Generate Sensor Association rules
[10]	2013	Distributed Data Extraction(DDE)	Estimate the missing sensor values
[16]	2012	Novel Association Rule Mining (NARM)	Frequent Association patterns

Table 2: Frequent Pattern Growth Approaches

Ref.No	Year	Frequent pattern Growth Approaches	Objective
[7][8]	2007	Closed item-sets based association rules min-ing(CARM)	Estimate the missing sensor values
[13]	2008	Positional Lexicographic Tree(PLT)	Generate Sensor association rules
[14]	2010	Fd-tree(Fd)	Used to mine association rule in datastream
[15]	2009	Sensor Pattern Tree(SP)	Generate Sensor association rules
[17]	2014	Total From Partial Tree(TFP)	Generate Sensor association rules
[18][19]	2014	Associated Correlated Sensor Pattern Tree(ACSP)	Generate frequent Sensor association rules
[20]	2015	Sliding Window Associated sensor Pat-tern(SWASP)	Generate frequent sensor association rules and predicts faulty sensors
[21]	2015	Share Frequent Sensor Pattern(ShrFSP)	Generate frequent Sensor Association rules
[21]	2015	Parallel Share Frequent Sensor Pattern(PShrFSP)	Generate frequent sensor association rules in parallel and distributed pro-cessing environment.

References

[1] R. Agrawal, T. Imielinski and A. N. Swami, "Mining Association Rules between Sets of Items in large Databases," Proc. ACM SIG-MOD Conference on Management of Data, pp. 207-16, 1993.

[2] R. Agrawal and R. Srikant, "Fast Algorithms for Mining AssociationRules," Proc. of the 20th VLDB Conf, pp. 487-99, 1994.

[3] Approximate Frequency Counts over Data Streams,2002 Gurmeet Motwani Singh, Manku ,Rajeev

[4] G.S Manku and R. Motwani, "Approximate frequency counts over-data streams," Proc. on VLDB, pp. 346-357, 2002.

[5] M. Halatchev and L. Gruenwald, "Estimating missing values in related sensor data streams" in Proceedings of the 11th International Conference on Management of Data (CO-MAD '05), 2005.

[6] Romer,"Distributed Mining of spatial temporal event patterns in sensor networks",2006.

[7] N. Jiang, "Discovering association rules in data streams based on closed pattern mining," in Proceedings of the SIGMOD Workshop on Innovative Database Research, 2007.

[8] N. Jiang and L. Gruenwald, "Estimating missing data in data streams," Advances in Databases: Concepts, Systems and Applications, pp. 981–987, 2007.

[9] T. Abirami and Dr. P. Thangaraj," Fuzzy-Genetic Algorithm Based Association Rules for Wireless Sensor Data", International Journal of Computer Science and Telecommunications [Volume 3, Issue 11, November 2012

[10] Azhar Mahmood, Ke Shi and Shaheen KhatoonAn Efficient Distributed Data Extraction Method for Mining Sensor Network's Data,2013

[11] Samer Samarah, Member, IEEE, Azzedine Boukerche, Fellow, IEEE, and Alexander Shema Habyalimana "Target Association Rules: A New Behavioral Patterns for Point of Coverage Wireless Sensor Networks" IEEE TRANSACTIONS ON COMPUTERS, VOL. 60, NO. 6, JUNE 2011,

[12] J Jan, J Pei and Y.Yin,"Mining frequent patterns without candidate generation".In Proc.2000 ACMSIGMOD Int Conference Management of Data,pp.1-12, May 2000

[13] Boukerche and S.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, pp. 865–877, 2008.

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

[15] S. K. Tanbeer, C. F. Ahmed, B.-S. Jeong, and Y.-K. Lee, "Efficient mining of associa-tion rules from wireless sensor networks," in Proceedings of the 11th International Confer-ence on Advanced Communication Technology (ICACT '09), pp. 719–724, February 2009.

[16] Anjan Das, "A novel Association Rule Mining Mechanism in Wireless Sensor Net-work" 2012 IEEE

[17] Manisha Rajpoot ,"Efficient Pattern Mining for Wireless Sensor Networks Data" , ,Journal of Information Engineering and Applications, www.iiste.org ISSN2224-5782 SSN 2225-0506 (online) Vol.4, No.5, 2014

[18] Mamunur Rashid, Iqbal Gondal, loarder Kamruzzaman Faculty of Information Tech-nology "ACSP-Tree: A Tree Structure for Mining Behavioral Patterns From Wireless Sensor Networks",Md. Monash University. 38th Annual IEEE Conference on Local Computer Networks

[19] Md. Mamunur Rashid, Iqbal Gondal, Member, IEEE, and Joarder Kamruzzaman, "A Novel Algorithm for mining Behavioral patterns from WSN ",Member, IEEE (2014 Interna-tional Joint Conference on Neural Networks (IJCNN)July 6-11, 2014, Beijing, China-©2014 IEEE)

[20] Md. Mamunur Rashid, Iqbal Gondal, Member, IEEE, and Joarder Kamruzzaman, Mem-ber, IEEE "Mining Associated Patterns from Wireless Sensor Networks ",IEEE TRANSAC-TIONS ON COMPUTERS, VOL. 64, NO. 7, JULY 2015

[21] Md. Mamunur Rashid, Iqbal Gondal, Member, IEEE, and Joarder Kamruzzaman, Mem-ber, IEEE,"Share-Frequent Sensor Patterns Mining from Wireless Sensor Network Da-ta",IEEE TRANSAC-TIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 26, NO. 12, DECEMBER 2015

[22] G. Vijay Kumar, M. Sreedevi & NVS Pavan Kumar," Mining Regular Patterns in Data Streams Using Vertical Format", International Journal of Computer Science and Security (IJCSS), Volume (6) : Issue (2) : 2012

[23] C.Ganesh, B.Sathiyabama, T.Geetha, "Fast Frequent Pattern Mining Using Vertical Data Format for Knowledge Discovery" International Journal of Emerging Research in Management &Technology ISSN: 2278-9359 (Volume-5, Issue-5)May 2016.