Constructing of management system on wireless sensor network protocol based on knowledge data association mining technology

QiuFen Wang

College of computer and information engineering
Nanyang Institute of Technology
Henan Nanyang, China
e-mail: wangqiufennylg@163.com

HuiLing Guo

School of Computer Science and Technology
Zhoukou Normal University
Henan Zhoukou, China

Abstract—Data mining is a very important step of knowledge discovery and it is through the establishment of mining model and through the implementation of the corresponding algorithm to complete knowledge formation. This paper analyzes the advantages and disadvantages of the management information system of wireless sensor networks protocols. Data mining tools can find association rules that satisfy the condition. The paper presents constructing of information management system on wireless sensor network protocol based on knowledge data association mining. Experimental results show that the proposed method can improve the efficiency of routing protocol.

Keywords- Data mining; Wireless sensor network; Association rule; Routing protocol; Management system

I.   INTRODUCTION

In the source data (such as a database) stored in the general is the detail data, and sometimes people hope to treatment or observation of these data from a higher level view, conceptual or logical data generalization through different levels to find the data contains, in order to adapt to the requirements of data analysis. One purpose of data mining is based on the microscopic characteristics of these data found that middle higher level concepts are universal, and macroscopic knowledge.

Data mining tools can find association rules that satisfy the condition, but it cannot determine the actual meaning of association rules [1]. The rules related to the understanding of the need to be familiar with business background, rich business experience have enough understanding of the data. In association rule discovery, there may be two subjectively think there is not much between the items, their association rules support and confidence is high, according to the need of business knowledge, experience, from various angles to judge it is an epiphenomenon or has its inherent rationality.

Association rules are involving two or more than two variables. According to whether or not to allow the same dimension repeated, multidimensional association rules can be subdivided into inter dimensional association rules (not permit repeated) and mixed dimensional association rules (rules allow dimension appears in about the same time). Such as "aged 20 to 30, like outing like swimming" is mixed dimensional Association rules. Mining inter dimensional association rules and mixed dimensional association rules to consider different types of data fields.

In most cases, the sensor nodes in the network are powered by batteries, the battery capacity is very limited, and tens of thousands of nodes in wireless sensor network for, replace the battery is very difficult, even impossible. If the nodes in the network because of the ability of exhausted and unable to work, will bring the network topology results change and routing problem is re established, possibly even network into disconnected parts, causing the interrupt of communication.

Wireless sensor network routing protocol is produced mainly because of its normal communication network and ad-hoc network are different [2]. First, because the sensor nodes in the network can not be many, for the establishment of a difference only in the network identity of each node, so the typical IP based protocol could not be applied to wireless sensor network. Association rule mining is the most common method of relational knowledge discovery. Need to be given two thresholds: minimum support and minimum confidence. Mining association rules must satisfy the user specified minimum support degree, which represents a group of projects in the lowest degree of contact with the association, needs to be met. Mining association rules are also must satisfy the minimum confidence a user specified, it reflects an association rules minimum reliability. The paper presents constructing of information management system on wireless sensor network protocol based on knowledge data association mining.

II.   KNOWLEDGE DISCOVERY AND DATA MINING ANALYSIS

Data mining methods are based on mathematical theory, can also be non mathematics; can be interpreted, can also be summed up. From the study of history, they may be scholars and engineering technical personnel database, artificial intelligence, statistics, computer science and other aspects, theory system founded in data mining is discussed in the paper in the process of it.

Data mining is essentially a new business information processing technology. Data mining technology is the application of data, query operation online in the lower

level, raise to decision support, analysis and prediction of more advanced applications. It is through the micro, meso and macro statistics, analysis, synthesis and reasoning of these data, found that the future trend of relevance, between the data and the general summary of knowledge, the knowledge of the information can be used to guide senior business activities.

The record in the database can be divided into a series of meaningful subset, namely clustering. Clustering increases people's understanding of the objective reality, is a concept description and deviation analysis prerequisites. Clustering techniques mainly include pattern recognition method and mathematical taxonomy of traditional.

In this theoretical framework, data mining is considered to be the discovery process knowledge model from the source data set [3]. This is the inheritance and development of machine learning method, is a popular data mining research and system development framework. According to this architecture, we can found on the different knowledge model study the process. At present, the association rules, classification / clustering model, sequential patterns (Sequence Model) and decision tree (Decision Tree) technology and the method of induction and so on discovery mode, as is shown by equation(1).

$$I_{k-1}(X) = I_k(X') = I_k(W(X, P))$$

(1)

Boolean association rule processing value are all types of discrete, it shows the relationship between these variables; and the quantitative association rules can and multidimensional association or multilevel association rules together, carries on processing to the numeric fields, the dynamic segmentation, or directly on the original data processing of course, quantitative association rules can contain type variables.

This algorithm first database logically divided into several disjoint blocks, each individually considered a block and it generates all frequent sets, and then produce frequency collection, used to generate all possible frequency set, finally calculated the support of itemsets. Size selected here block to make each block can be put into main memory, each stage simply by scanning a. And the correctness of the algorithm is composed of every possible frequency set at least ensure frequency set in a block.

Contained in the source data or extreme special case is significantly different from other data knowledge description, it reveals a deviation from the conventional rules of abnormal things [4]. The data in the database often has some abnormal records from a database; detection of specific knowledge contained in these data is of great significance. For example, in the Web site to find the user characteristics of those differences from normal login behavior can prevent illegal intrusion. Specific knowledge can and other data mining technology together, further access to specific knowledge in the mining of common knowledge at the same time, as is shown by equation(2).

$$g(x, y) \in V_j^2 \Leftrightarrow g(2x, 2y) \in V_{j+1}^2, \quad \forall j \in Z$$

(2)

Association rule mining process mainly includes two stages: the first stage must start from the data set out high-frequency project all group (Frequent Itemsets); the second stage from the high frequency items in a group to generate association rules.

Data preparation work to do. Can do a good job in data preparation and depends on the first two points. Data preparation will directly affect the implementation complexity and the target problem. Select the appropriate minimum support and minimum confidence. It relies on the user to estimate the target, if the value is too small, then you will find many useless rules, not only affect the execution efficiency, waste of system resources, and may take the target buried; if the value is too large, it may not find the rules.

All support is greater than minimum support of itemsets called frequent itemsets, referred to as the frequency set. The basic idea of the algorithm is: first find all frequent sets, these itemsets frequent sex appeared at least the same as the predefined minimum support degree. Then generate strong association rules from the frequent set, these rules must meet the minimum support and minimum confidence. And then use the first step to find frequency set to produce the desired rule, produced only contains all rule sets of terms, where each rule of the right of the only one here is used in the definition of the rules. Once these rules are generated, then only those minimum confidence greater than the user specified rules to be left. In order to generate all frequent sets and it is using a recursive method.

Set: I={i1, I2,..., im} is the set of all the project. D is the set of all transactions (i.e. database), each transaction T is a collection of some of the projects, T included in I, and each transaction can be identified by a unique identifier TID. Let X be a set of certain items, if the X contains in T, said the transaction T contains X, association rule is expressed as the following formula [5].

$$u_i = \frac{1}{\sqrt{\lambda_i}} A v_i \quad i = 1, 2, \cdots, r$$

(3)

Data mining technology is regarded as a data compression process. According to this view, association rule, decision tree, clustering algorithm for large data sets are in fact constantly conceptual or abstract compression process. The minimum description length principle can evaluate a compression method, namely the best compression method should be the concept itself is described and regard it as a predictor coding length minimum.

Relational databases are two important problems different from in the traditional mining transaction databases [6]. Logically speaking, relational database is a collection of a series of tables. Therefore, in the mining of relational database, in addition to considering the correlation table attributes, we must also consider the association between attribute tables. Technology and algorithm of transaction database mining traditional research is generally based on single table. Therefore, in mining in relational database, must consider the multi table mining technology.

Density based clustering method is by measuring the number of objects contained by a zone to form the final goal. If an area density exceeds the specified value, it needs to be further decomposed into finer group, until the user acceptable results. Compared to the clustering method based on partitioning, can discover arbitrary shape clusters of spherical outside, but also can be a very good filtering outlier (Outlier) data, for large data sets and spatial database has good adaptability.

Association rule mining problem can be divided into the following two sub problems: find the transaction database in D all is greater than or equal to the user specified minimum support degree item sets (itemset). With minimum support of itemsets called maximal itemsets. Item set support refers to the number of itemsets include using maximal itemsets; association rules to generate the required, for each maximal itemsets A, find all nonempty set a A, if the ratio of support (A) /support (a) >=minconfidence.

Data mining system in constraint guidance users, can improve the mining efficiency and accuracy. The research on it is a very broad topic. In the visualization and interactive data mining, the use of user input constraint is a prerequisite for visualization and interactive mining. The relational database, because of the complexity of its properties (such as a large number of properties exist), attribute Association contains storage and multi table or a multi-level concept and other issues, the problem of mining constrained data becomes more important

III. INFORMATION MANAGEMENT SYSTEM ON WIRELESS SENSOR NETWORK PROTOCOL

In wireless sensor networks, node cans only activities for data forwarding and topology control can determine which nodes as the forwarding nodes, while determining the neighbor relationship between nodes. Refers to the sensor nodes send data acquisition to the central node in wireless sensor network data fusion, the central node for data fusion, and send the data to sink node after fusion. While the center node selection is an important content of topology control, it is to help it.

The convergence point on receipt of these low rate data, to the data reach the neighbor node sends the fastest enhanced message; enhanced message said the sink node sends the data requirements of high rate. The corresponding neighbor nodes according to the same way, followed by transfer of enhanced information to its neighbor nodes, until the arrival of the data source, the main path to construct the data transmission, data will send to sink through the path.

When a new node is added to the sensor network S, it will be according to the following steps: adding routing backbone network a) node S to the one hop neighbor nodes broadcast within a Req message; b) receiving Req message if a node has been added to the routing backbone network, it returns a Res message, including the distance parameter of its value in the message; c) the node S waiting for T time, it check the distance parameter Res message to return values in the choice of parameter values, the minimum distance of nodes (more than 1) as the parent node, the parent node at the same time will be saved to the parent node in the collection and recording the distance parameter self arrives at the gateway node values (i.e. distance

parameters of the parent node value plus 1), then the message to a neighbor node broadcast Res, as is shown by equation(4) [7].

$$\tilde{r}(n) = \sum_{i=1}^{N} \gamma^{n-i} d^*(i) x^H(i)$$
(4)

Planar structure refers to all nodes in the network in the routing function on the same position, without introducing hierarchical management mechanism. Its advantage is that there are no special nodes in the network, network traffic is uniformly dispersed in the network, and the routing algorithm is easy to implement [8]. The disadvantage is the scalability small, to a certain extent, limiting the size of the network. Hierarchical Routing: should be relative and planar routing protocol routing protocol is a hierarchical structure. It uses the concept of clusters of the level division of sensor nodes.

SPIN negotiation process with the 3 handshake mode. Step1: run SPIN protocol source nodes in the transmission of DATA information before the first to broadcast the adjacent node contains DATA description mechanism of ADV information. Step2: the neighbor nodes need the DATA information, request information to the source to send REQ. Step3: a source node based on the received REQ information, selectively DATA messages to the neighbor nodes corresponding.

Flooding (Flooding) protocol is a kind of traditional wireless routing protocol. The terms of the agreement, each node from receiving information from other nodes, and it is send a broadcast form to other neighbor nodes. And so it went on, the information data is sent to the destination node. But this protocol is easy to cause information "Implosion" (Implosion) and "overlap" (Overlap), resulting in a waste of resources. So based on the flooding protocol, and it is proposed gossip (Gossiping) protocol.

Life cycle is defined from the WSN start to be as the observer needed information available so far for the duration of the. For working in special environment of WSN, the physical maintenance or replacement of sensor failure is often difficult or impossible, so the routing protocol to prolong the sensor nodes as working time may be, make the first sensor node moment of death off as long as possible, thus making monitoring the entire network to extend the time, as is shown by equation (5).

$$D_k^{k-n+1}(x) = \sum_{i=k-n+2}^{k} D_i(x)$$
(5)

According to whether the data in the transmission process of polymerization process can be divided into routing protocols, it is data aggregation and non data aggregation routing protocol. Data aggregation can reduce the amount of communication, but the need for time synchronization technology support, and the transmission delay increase. F) Routing according to whether the specified by the source node, can be divided into routing protocol and routing protocol for the source non source station station. Source station routing protocol without establishing, and it is maintaining routing information, so

as to save storage space, reduce the communication overhead. But if the network scale is larger, the routing information overhead data in Baotou are large, and if the network topology changes frequently, will cause the routing failure.

ADV announced the data sent by the REQ request, hope to receive the data, encapsulating data with DATA. The protocol has the advantages of: small ADV news relieved implosion problem; through data naming solves the overlap problem; node according to its resources and the application of information to decide whether ADV circular, avoid blind problem of resource utilization, effectively saving energy [9]. But the drawback is: when all of the neighbor nodes to generate or receive data nodes do not need the data, will cause data cannot continue forwarding, so far unable to get data node. When most of the nodes in the network are potential sink point, the problem is not serious, but when sink is less, it is a very serious problem; and when a sink points are required for any data, the peripheral node energy easily exhausted; while reducing the data implosion.

In the hierarchical structure of the network, the network node has some associated component cluster. In the clusters, it is usually with an elected according to a certain rule known as the cluster head node. In addition to the cluster head node, the node of the members of the general function of relatively simple does not need to maintain routing information complex. Hierarchical routing shortcomings is that the cluster head node to become the bottleneck of the network, so the routing algorithm with fault tolerance requirements; at the same time, a great challenge to cluster load balancing is also distributed clusters.

IV. DEVELOPMENT OF INFORMATION MANAGEMENT SYSTEM ON WIRELESS SENSOR NETWORK PROTOCOL BASED ON KNOWLEDGE DATA ASSOCIATION MINING TECHNOLOGY

The data in data warehouse is organized according to topic. The stored data can provide information from a historical point of view. In the face of multiple data source, and it is through the data warehouse cleaning and after conversion can provide ideal environment for data mining and knowledge discovery [10]. If a data warehouse model has a support multidimensional data model or the multidimensional data cube model, then the operator based on multidimensional data cube can achieve high efficiency of calculation and fast access.

Wireless sensor network also can be measured as its main structural basis. Single layer network in the sensor density is big; the gateway will appear overload phenomenon. This phenomenon will occur in the passive communication, so the single layer network is not suitable for emergency monitoring. In addition, the monolayer is unable to realize the transmission of long sensor time. Transaction, using the cluster of data fusion, and then achieve the effective utilization of energy and saving energy purpose [11]. Choose the way to construct cluster from the sensor node is energy saved method based on typical, as is shown by equation (6).

$$H_e = - \sum_{l=0}^{L-1} P(l) \log_2 p(l)$$

(6)

The design of a knowledge discovery system should have the key steps. Generally speaking, from the large source data are found in the process of the value of knowledge can be summarized as follows: from the data source to extract data of interest, and put it into the organization of data for mining; then, algorithm for generating calls the corresponding knowledge required; finally carries on the appraisal to the generation of knowledge mode, and the valuable knowledge is integrated into the intelligent system in the enterprise. Reduce energy waste of idle listening, but still maintain a reasonable throughput. Through T-MAC simulation, no duty cycle CSMA and the duty ratio of fixed S-MAC compared with the typical, found a constant load when T-MAC and S-MAC energy-saving similar, as is shown in Fig .1.
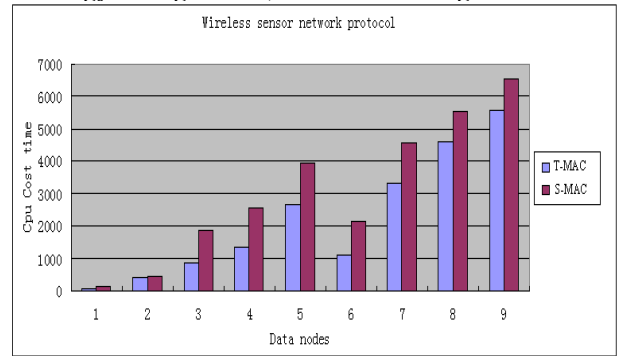

Figure 1. Information management system on wireless sensor network protocol based on knowledge data association mining.

In a sensor network, sensor nodes may have different functions. Although in the initial state can assume that all nodes have the same function. However, once the application requirements change, you need to have some special function nodes. In order to solve this contradiction, some hierarchical routing protocol used in the ordinary nodes in the cluster head selection (CH) method. LEACH was first proposed in wireless sensor networks in the hierarchical routing protocol. A lot of hierarchical routing protocols are proposed subsequently on the basis of it.

V. SUMMARY

Data mining is a very important step of knowledge discovery; it is through the establishment of a mining model and through the implementation of the corresponding algorithm to complete knowledge formation. Of course, there are many tools can help complete the data mining. The paper presents constructing of information management system on wireless sensor network protocol based on knowledge data association mining. These tools are provided including many association rules, classification, clustering, and decision tree and many kinds of model and algorithm.

REFERENCES

[1] Xiaoyan Wan, "Research on Data Mining Technology of Association Rule", JCIT, Vol. 8, No. 6, pp. 628 ~ 635, 2013.

[2] Zhi Chuan, "A New MAC Protocol based on Adaptive Energy-efficient for Wireless Sensor Networks", JDCTA, Vol. 7, No. 3, pp. 119 ~ 126, 2013.

[3] Hassan Ansari , "Combining Association Rules and Clustering in Data Mining ", RNIS, Volumn 9, pp. 30 ~ 36, 2012.

[4] Yan Hai, Xiu-li Li, "A General Temporal Association Rule Frequent Itemsets Mining Algorithm", IJACT, Vol. 3, No. 11, pp. 63 ~ 71, 2011.

[5] JIANG Fei, "Research on Association Rule Mining of Adaptive Genetic Simulated Annealing algorithm", JCIT, Vol. 8, No. 5, pp. 876 ~ 883, 2013.

[6] Somboon Anekritmongkol, Kulthon Kasemsan, "Boolean Normalized By SQL for Association Rule Mining", IJACT, Vol. 5, No. 14, pp. 138 ~ 149, 2013.

[7] Chifaa TABET HELLAL, Mohamed LEHSAINI, Hervé GUYENNET, "An Enhanced Fault-tolerant Version of LEACH for Wireless Sensor Networks", IJACT, Vol. 6, No. 6, pp. 50 ~ 57, 2014.

[8] Shaoguo Xie, Yanjun Hu, Yi Wang, Lele Qiu, Yingguan Wang, "Probability-based Localization in Wireless Sensor Networks", IJACT, Vol. 5, No. 9, pp. 670 ~ 678, 2013.

[9] Lasheng Yu, Nsoita Ivan, Li Jie, Linong Li, Renjie Liu, "Research and Implementation of an Effective Data Gathering Algorithm for Wireless Sensor Network", JCIT, Vol. 8, No. 4, pp. 98 ~ 107, 2013

[10] Sheng-Chu Su, Chien-Hung Lin, Tzu-Chin Chao, "Data Mining Technique on Bank Service Satisfaction Research", JCIT, Vol. 8, No. 6, pp. 845 ~ 854, 2013

[11] Yuan Wang, Yihua Zhang, Lei Cai, "Research on China Accounting Firms Competitiveness Based on Data Mining Methods", AISS, Vol. 5, No. 4, pp. 524 ~ 535, 2013