Pattern Analysis on Wireless Sensor Data using Soft Computing Techniques

Dr.T.Abirami
Associate Professor
Department of IT
Kongu Engineering College

Abstract: Wireless sensor networks (WSNs) have amazing potential of being deployed in many places where conventional stressed out or wireless networks are not feasible. The discipline of Wireless Sensor Networks (WSNs) is experiencing a resurgence of interest and a non-stop evolution within the clinical and commercial community. The records gathered by using the sensors used as enter for the gaining knowledge of algorithms is composed of automatically amassed sensor facts and further manually introduced facts. Extracting useful records from uncooked sensor facts requires specific strategies and algorithms. Finally, examine the dataset and evaluate the performance of device learning techniques that are used for pattern classification in this dataset. It conserves the energy and improves a detection performance in sizeable amount. Simulation results indicates that proposed techniques can gain the aim of minimum power, much less sensing time and higher performance. This paper intends to assist new researchers getting into the area of WSNs by using offering a comprehensive survey on latest developments.

Keywords: *Wireless Sensor Networks (WSN), Association Rules, Genetic Algorithm and Fuzzy Logic*

I. INTRODUCTION

A wireless sensor network (WSN) is a hard and fast of nodes which senses and performs information computations and communicates statistics with others wirelessly The task of enhancing sensor electricity consumption is a key difficulty in wi-fi sensor networks (WSNs). This is due to the truth that sensor community lifetime is directly associated with operational lifetime. If sensors can operate for longer durations of time and the frequency of node failures can be reduced, the reliability and adaptableness of the sensor network will therefore improve. Data transmission is very luxurious in phrases of power. Sensor which collect information hand them over to the sink which is observed with the aid of offline records analyses to extract patterns. The lifestyles of a big communique overhead impacts sensor network overall performance negatively. This large overhead becomes a hurdle for the deployment of long term massive scale sensor networks. Data mining technique along with Association rule mining is carried out to find out frequent patterns inside the information.

Sensor nodes that are used to form a sensor network are normally operated via a small battery which has small quantity of power. Therefore, in wireless sensor networks decreasing energy intake of every sensor node is one of the prominent issues to address in the network lifetime, since wireless communications consume full-size quantity of battery power, sensor nodes must be electricity green in transmitting facts. Protocols can reduce transmitted power in two ways. First in which nodes can emit to quick distances together with facts sinks or cluster nodes. The cluster node can then ship the data over a larger distance preserving the energy of the smaller nodes. The 2d is with the aid of decreasing the wide variety of bits (amount of

facts) despatched across the wi-fi community. In this work, the quantity of statistics despatched to sink is decreased the use of affiliation rule mining and to further reduce the energy consumption of the community, most appropriate routes are selected to transmit statistics to the sink primarily based on strength intake. In this paper, it's far proposed to mine WSN facts the use of affiliation rule to extract patterns. Genetic set of rules and Fuzzy logic is used with affiliation rule for effective extraction of rules from quantitative facts.

II. RELATED WORKS

A) Association Rule Mining

Association rule is one of the prime strategies of records mining. Frequent patterns, correlations, institutions or informal structures are detected through affiliation rule among sets of items or gadgets in transactional databases and different storehouses of information. There is a theatrical increase in extent due to the data generated in every day activities. So, via mining affiliation rules, there is an full-size volume of statistics within the database which pastimes many industries which help in many techniques of decision making. Relations are diagnosed between items through the techniques which discover association rules from records thru which some type of human behavior is shown, for instance, reading the buying conduct or the pattern via the gadgets are brought together by using the user. So, a particular neighborhood sample is decided via the association policies that may be without problems interpreted and communicated [1]. Knowledge discovery is the fruit of association rule mining from agricultural databases, which includes information approximately the details concerning soil, cultivation and geographical conditions. Decisions related to selection of crops, resources, proper surroundings and so forth are obtained by using rule mining [2]. Support and self assurance are two primary basic measures of association. Suppose, there are two objects, then the definition of help is the ratio of occurrence of that two gadgets and cumulative transactions. If the guidelines have more help than a user defined guide, then it's far minimum help. Confidence is the opportunity of seeing the rule’s consequence underneath the condition of transactions. Minimum self belief is one where policies offer greater self belief than user-defined consequence.

An Association rule is an implication of the form

$X \Rightarrow Y$, where $X \cap Y = \Phi$ and

X & Y are subsets of all itemset I.

There are two measures of rule interestingness i.e. Support (σ) and Confidence (C). They reflect the usefulness and certainty of the rules. The rule $X \Rightarrow Y$ (support $\varphi = 10\%$, confidence $C = 80\%$) indicates 10% of all the transactions under analysis involves simultaneous purchase of items X and Y by customers and 80% of confidence shows that customers who purchased item X also brought item Y[3]. Association rule is used to relate objects to each other and group them together. Association rule is classified in numerous ways, based on type of values (Boolean or Quantitative), dimensions of data (Single dimension or Multidimensional) and level of abstractions involved (Single level or Multilevel).

For mining, various algorithms have been proposed for the association rules and can be decomposed in two phases.

1. Frequent item sets whose support and confidence values should be more than user specified minimum support (φ) and minimum confidence (C) values respectively.
2. Desired association rules are found through frequent items and so the parameters should satisfy minimum support (φ) and minimum confidence (C).

With regard to Association rule mining, Apriori set of rules is a massive accomplishment. Until now, it's far one of the most famous rule mining set of rules. Moreover, the belief is that itemset is of lexicographic order. The candidate itemsets are generated by Apriori by means of connecting massive itemsets of previous bypass and the subsets that are small inside the previous bypass are deleted from the database. By considering huge itemsets of the previous pass, reduction in the variety of candidate big itemsets is achieved. Most of the health-related cellular utility based on sports and fitness monitoring. Currently, the survey shows that fast development in healthcare utility and ambient –assisted living solutions [4].This kind of increase in demand and device ends in the notion of context focus to aware social process [5].Moreover, sizeable data's has amassed from a heterogeneous gadget that results in new collective intelligence. Hence customers flip as part of massive urban fantastic organisms [6].Finally, it moves toward the following generation, developments and solution implemented in context-aware technology. For illustrating those, we used sensing infrastructure for smart towns and collective action for crowd sourcing of information.

B) Genetic Algorithms for Mining Association Rules

A Genetic Algorithm (GA) is implemented to generate balanced and strength efficient facts aggregation spanning bushes for wireless sensor networks. In a statistics amassing round, a single pleasant tree consumes lowest strength from all nodes but assigns extra load to a few sensors. Association rule mining is optimized via Genetic Algorithms (GA). Based on quantitative association policies discovery, a GA chromosome encodes a generalized okay-rule, okay being the required length [7]. As association policies with many objects in consequent is used, an index stored by using the primary gene, representing end of an in advance portion. To encode a rule into a chromosome, the antecedent and consequent attributes are taken care of in two-segment in a spiral, ascending order. The ultimate okay genes encode items; every represents a pair of values. The first fee is an attribute's index starting from 1 to a maximum quantity of database attributes, whilst second factors to a gapped c language. A gapped c language is the union of finite base intervals received whilst uniform discretization technique is done over database attributes.[14] Partitioning the categorical attribute domains is senseless as decrease and top bounds coincide.

A base interval is represented via an integer quantity resulting in a gapped interval period in a fixed of integers. Genetic operators implemented to a chromosome area follow: *Selection:* This is achieved through computing fitness price with a random quantity to make sure a chromosome is selected if the product is less than given choice probability (ps). *Crossover:* Selected chromosomes reproduce offspring at a crossover probability (pc) and the operation includes exchanging a gene phase among first and 2nd chromosomes and vice-versa, all relying on randomly generated crossover-points.

Mutation: Consideration of both mutation probability (pm) and fitness cost, a chromosome is altered so that the border between antecedent and consequent attributes is modified inside the equal rule. An operator selects a gene randomly and changes attribute's index with associated gapped interval. The new gapped interval is a mixture of base intervals which now form a brand new characteristic sub-domain.

Genetic Algorithms are chaotic and non-determinist seek methods, which use real lifestyles models to solve complicated and at instances intractable troubles [8]. It functionality is

primarily based at the Darwin's principle of biological evolution via natural and sexual choice. In fact, these feasible answers are chromosomes. The reasonable solutions are selected after the populace is studied in every iteration under genetic operators. But there are some problems of the genetic algorithm, including the slow convergence rate, premature convergence and sometimes local optimum solution. The value of comparing the path is taken as the fitness value in GA, which contains all constraint conditions in looking optimum path. The element including path length, electricity consumption and the network energy equilibrium should be considered for possible route fitness value. The feasible path fitness function of n nodes is defined as follows:

$$k(X) = w_d g(X) + w_e h(X) + w_l m(X)$$

w_d,w_e and w_l are respectively length, energy consumption and the network energy equilibrium. g(X) is the path length, h(X) is energy consumption of path(20), m(X) is consumption of network energy equilibrium, the total length of path is as follows:

$$g(X) = \sum_{i=1}^{n-1} d^2(p_i, p_{i+1})$$

is the distance between the node p_i and p_{i+1}. Network energy equilibrium consumption is as follows: A chromosome shows a path of transmitting data, which is divided into feasible path and unfeasible path. Different chromosome consists of different sensor node sequence, so the length of chromosome is variable and it can be showed by two-dimensional array or chain structure. As a result, the energy resources of heavily loaded nodes will be depleted earlier than others. The proposed GA extends network lifetime.

C) Fuzzy Logic System

A general fuzzy logic system (FLS) structure is shown in Fig.1. The fuzzifier converts crisp input variables X, where X is a set of possible input variables, to fuzzy linguistic variables through application of corresponding membership functions. Zadeh defines linguistic variables as "those whose values are not numbers but words/sentences in natural/artificial language" [9]. An input variable is usually associated with one or more fuzzy sets based on calculated membership degrees.[13] For example, a temperature value can possibly be classified as Low and Medium.

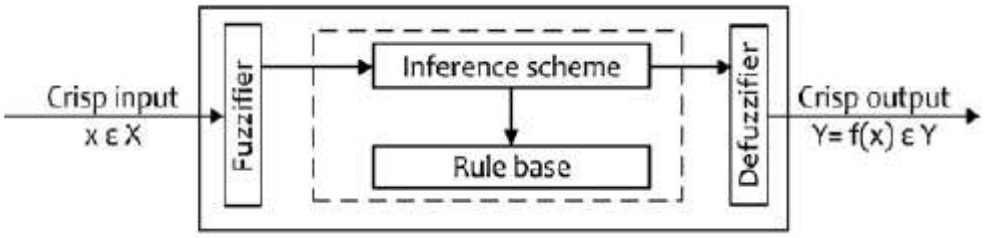

Fig.1.Generic Fuzzy Logic System

Fuzzified values are processed through if–then statements in accordance to a set of predefined guidelines were given from domain expertise. An inference scheme maps enter fuzzy units to output fuzzy units and a defuzzifier computes a crisp end result from the fuzzy units output the usage of the rules. Control moves are based at the crisp output cost. The

classified steps are referred to be Fuzzification, Choice making and Defuzzification.

Fuzzification: The fuzzifier converts crisp values into club degrees via membership feature applications. Membership characteristic determines the association between unique linguistic values. Membership capabilities have numerous shapes and some regularly used shapes are triangular, trapezoidal, and Gaussian. Membership features are defined either via relying on domain expertise or via software of numerous learning techniques like neural networks [10], [11] and genetic algorithms [12].

Decision making: A rule-base contains of a hard and fast of linguistic statements referred to as guidelines that are of the shape IF premise, THEN consequent in which premise consists of fuzzy enter variables linked with the aid of logical functions (e.G., AND, OR, NOT) with the consequent being a fuzzy output variable. The rule-base is generated as a massive set of each feasible value-aggregate for input linguistic variables constituting the premise. Similar to definitions of club capabilities, rule-base is derived from both domain understanding or through use of system learning strategies.

Defuzzification: Executing policies in rule-bases generates multipleshapes that represent modified membership features. Common defuzzifiers are middle of gravity, center of singleton, and most strategies [13]: The middle of gravity approach receives a centroid shape by way of superimposing shapes as a result of rule software. The defuzzifier output is the x-coordinate of this centroid.

The defuzzification method is simplified when middle of singleton manner is used, as each rule membership characteristic is defuzzified separately. Each club function is reduced to a singleton representing the center of gravity characteristic. The simplification is done via singletons being determined during machine design.[16] The centre of singleton approach approximates center of gravity approach. The class of most strategies determines the output through selecting a membership characteristic with a maximum fee. If the most is a range, then lower, upper, or the middle price is taken into consideration for output price based totally at the technique. With these methods, rule with maximum hobby determines output cost always. As the class of most methods indicates discontinuous output on non-stop input; these strategies are unsuitable to be used in controllers.

III. METHODOLOGY

The data sets collected from various sensors at different time slot have included in the proposed system. The sensor data derived from the data sets were collected from 54 sensor node with epoch duration of 31 seconds. The sensors collected time stamped 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 Table 1.

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

Health Routes To Users

Most of the people take a part of character lifestyle due to medical recommendations based totally at the incidence of fitness issues. Hence, this activity normally closer to extra emotions and few facts. Mostly this approach applies to the human beings for selecting activities like sports, exercise but they do no longer have right guidance to comply with and perform it. For example, walking can be extra useful for a person to decide the area to play it. Most of the humans used to decide on it based on distance and weather. Some avoided elements while walking like air pollution, crown density; it is able to result in chronic disorder for human beings. For fixing this kind of issues, our method affords realtime constraints and data from diverse sources particularly real-time facts from clever city infrastructure, usage of routes for humans based on fitness circumstance.

The device has some capabilities like
a. It used to select routes primarily based on best-suit of user's requirement and preferences.
b. Users can follow with smartphones.
c. The System will offer both dynamic and collaborative in following the real-time modification. It consequence relies upon on variable monitoring by way of town sensor
d. Users perform in sensor that offers records and experience to the systems.
e. The user can also inform the system concerning the unexpected circumstance that can also have an effect on other users.
f. The user can show new routes, to enhance the machine.

The schedule of data aggregation tree is created at the base station using E-Span and LPT algorithm. Two sets of experiments are conducted. The simulation parameters for the sensor network are as follows: a) network deployment areas are 50×50 and 100m×100m, b) initial energy of each sensor node is 1 Joule, c) sensor nodes are randomly deployed in the given area, d) each tree is used for 10 rounds, e) each experiment is conducted for 3 simulation scenarios and the average is used for documentation, f)base station is located at the center of the sensor network, g) MAC layer, 802.11, is used in simulation. Table. 2 gives the simulation parameters.

SIMULATION PARAMETERS

Parameter	**Remarks**
Network Area	50×50 m2 and 100×100m2
Number of nodes	25, 50 and 75
Base station location	a) at the center of network field, and b) outside network field at the distance of 100 m.
data packet length	It is assumed that each sensor generates fixed length data packet of size 1000 bits.
Initial Energy	Each sensor is initialized with 1 J

In this study, the amount of data sent to sink is reduced using association rule mining and to further reduce the energy consumption of the network, optimal routes are chosen to transmit data to the sink based on energy consumption. The proposed method is able to discover the associations to make predictive analysis such as node failure, asymmetric links. The rules found forms the basis for coding solutions in the proposed Genetic Algorithm. Intel lab sensor dataset is used for evaluating the proposed work.

IV. RESULTS AND DISCUSSION

Table 3 tabulates the results of Rule support analysis, and Rule confidence analysis.

Number of rules	Support of rules	Confidence of rules %
0	5500	95
5	5300	90
10	5000	86
15	4800	82
20	4400	78
25	4000	74
30	3600	70
35	3200	68
40	3000	65
45	2400	60
50	2000	55

Figure 2 reveals rule confidence analysis for obtained rules. Rules number obtained for confidence of 90% is around 20 higher than when only a genetic algorithm is used. The obtained rules number increases proportionally with decrease in confidence.

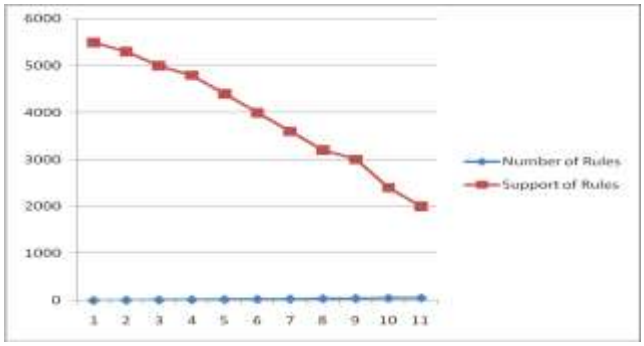

Fig 2. Rule Support Analysis

Similarly, Figure 3 shows rule support analysis performing better than a genetic algorithm. The proposed fuzzy genetic algorithm shows improved performance.

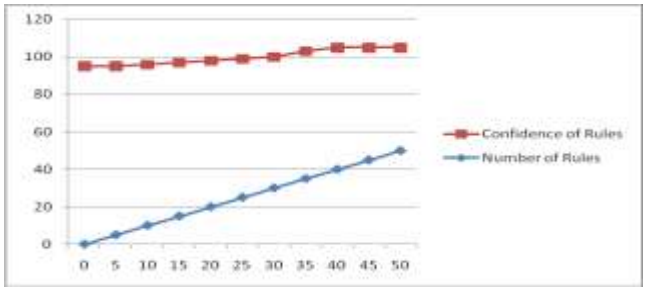

Fig 3. Rule Confidence Analysis

V. CONCLUSION

In this paper, the proposed work to extract pattern from WSN data using association rule mining. Genetic algorithm and Fuzzy logic is used along with association rule for effective extraction of rules from quantitative data. The proposed method uses INTEL dataset for evaluation. Fuzzy logic and Genetic algorithm is used with association rule for effective extraction of rules. Experimental results show that the proposed fuzzy genetic algorithm method is effective in concurrence with Association rule mining. The results that have obtained are encouraging for further extension of the system, by creating a network of sensors so that more information can be obtained by considering and complementing the recent system with semantic technologies for enhancing the data for more diversified and highly accurate predictions.

REFERENCES

[1] CERÓN-RIOS, G., LÓPEZ, D. M., & Blobel, B. (2017, May). Architecture and User-Context Models of CoCare: A Context-Aware Mobile Recommender System for Health Promotion. In PHealth 2017: Proceedings of the 14th International Conference on Wearable Micro and Nano Technologies for Personalized Health 14–16 May 2017 Eindhoven, The Netherlands (Vol. 237, p. 140). IOS Press.

[2] Kaur, E., & Haghighi, P. D. (2016, November). A Context-Aware Usability Model for Mobile Health Applications. In Proceedings of the 14th International Conference on Advances in Mobile Computing and Multi-Media (pp. 181-189). ACM.

[3] Bogdanović, D., & Lazarević, K. (2014). Early Warning System and Adaptation Advice to Reduce Human Health Consequences of Extreme Weather Conditions and Air Pollution. In Handbook of Research on Democratic Strategies and Citizen-Centered E-Government Services. IGI Global.

[4] Rashidi, P., & Mihailidis, A. (2013). A survey on ambient-assisted living tools for older adults. IEEE Journal of biomedical and health informatics, 17(3), 579-590.

[5] Lukowicz, P., Pentland, S., & Ferscha, A. (2012). From context awareness to socially aware computing. IEEE pervasive computing, 11(1), 32-41.

[6] Zambonelli, F. (2012). Toward sociotechnical urban superorganisms. Computer, 45(8), 76-78.

[7] Mata, J., Alvarez, J., Riquelme, J.: An Evolutionary Algorithm to Discover Numeric Association Rules, ACM Symposium on Applied Computing, Madrid, Spain, March 2002.

[8] Maraiya, K., Kant, K. and Gupta, N. "Wireless Sensor Network: A Review on Data Aggregation", International Journal of Scientific and Engineering Research, Vol. 2, No. 4, pp. 1-6, 2011.

[9] L.A. Zadeh, Outline of a new approach to the analysis of complex systems and decision processes, IEEE Trans. Syst Man Cybernet. SMC 3 (1) (1973) 28–44.

[10] S.-i. Horikawa, T. Furuhashi, Y. Uchikawa, On fuzzy modeling using fuzzy neural networks with the back-propagation algorithm, IEEE Trans. Neural Networks 3 (5) (1992) 801–806.

[11] J.-S.R. Jang, Self-learning fuzzy controllers based on temporal backpropagation, IEEE Trans. Neural Networks 3 (5) (1992) 714–723.

[12] Arslan, M. Kaya, Determination of fuzzy logic membership functions using genetic algorithms, Fuzzy Sets Syst. 118 (2) (2001) 297–306.

[13] G.J. Klir, B. Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice-Hall Inc., Upper Saddle River, NJ, USA, 1995.
[14] T Abirami, P Priakanth, "Energy Efficient Wireless Sensor Network using Genetic Algorithm based Association Rules", International Journal of Computer Applications, (2014), pp.33–36
[15] T Abirami, P Thangaraj, "Association Rules for Wireless Sensor Data Based On Fuzzy - Genetic Algorithm", Life Science Journal 2013;10(4s)
[16] T Abirami, 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]