ISSN 2047-3338

Fuzzy-Genetic Algorithm Based Association Rules for Wireless Sensor Data

T. Abirami and Dr. P. Thangaraj

***Abstract*– Wireless Sensor Networks (WSN) generates a huge amount of data for efficient application of discovering essential knowledge from it is important. Usually, WSN data is generated in streams and forwarded to a sink. Raw data leads to higher communication overhead adversely affecting WSN performance. Association mining processes data to locate frequent patterns. Thus, when association mining is applied to WSN network data only frequent raw data patterns are sent to the sink, thereby reducing communication overhead. This paper proposes mining of WSN data through an association rule to extricate patterns. A Fuzzy based genetic algorithm is used with the rule for efficient extraction of rules from data.**

***Index Terms*– Wireless Sensor Networks (WSN), Association Rules, Genetic Algorithm and Fuzzy Logic**

I. INTRODUCTION

A wireless sensor network (WSN) is a set of nodes which senses and performs data computations and communicates data with others wirelessly [1]. A class of inexpensive sensors which comes under this category is Berkeley motes [2], where a network can be formed by Mica21Mica2dots having MTS310 sensor boards, which help to gather light, temperature, sound and motion readings. But in several applications, WSN generally facechallenges because of its limited battery life. Also large scale use of sensors in severeenvironments ensures that battery replacements arerisky.

To prolong sensor battery life,the radio transmissions in the network is reduced as network radio transmission consumes energy via sensor nodes [3]. This can be overcome by shifting the majority of processing into networked sensor nodes, so that nodes forward limited and less data to neighbouring nodes/base-stations. But sensors like Berkeley motes lack both hardware floating pointunitsand enough memory to handle complicated data mining algorithms as they would use up sensor resources quickly. Studies were conducted recently for in-network processing algorithmsfor sensor networks [4]. Related areas are DataStream Management Systems (DSMS) [5], which help filtering of queries to forward only user relevant data.

A wireless sensor node comprises sensing, computing, communication, actuation, and power components integrated on a single/multiple boards, all packed in a few cubic inches. A WSNhas numerous nodes communicating through wireless channels sharing information and processing. And they are usually usedglobally forwork such as environmental monitoring and habitat study, battle field surveillance and reconnaissance, search and rescue operations, in factories for condition based maintenance, for infrastructure health monitoring in buildings, in homes to realize smart homes, and for patient monitoring for humans [2], [7]. On initialization, sensor nodes should organize appropriate network infrastructure usually with multi hop, intra sensor node connections. On board sensors collect acoustic, seismic, infrared or magnetic environment information through continuous/event driven working modes. The location/positioning information is received through GPS (global positioning system) or local positioning algorithms. This information is got from a network and processed to ensure building up a global view of monitoring phenomena/objects. The philosophy behind WSN is that in spite of limited individual sensor node capability, the network's aggregate power is enough for special tasks [8].

WSNs generate a huge amount of data in streams which are forwarded to a sink [9]. This data generated leads to excessive communication overheads thereby affecting WSN performance. Association mining processes data to locate frequent patterns. When association mining is applied in-network to WSN, only frequent patterns in data need transmission to the sink which in turn reduces communication overheads.

In this paper, it is proposed to mine WSN data using association rule to extract patterns. Genetic algorithm and Fuzzy logic is used with association rule for effective extraction of rules from quantitative data. The proposed method uses INTEL dataset consisting of 54 Mica2Dot motes with temperature, humidity and light sensors.

II. RELATED WORKS

For WSNs, Pirmez et al., [10] proposed a fuzzy-based decision-making mechanism for choosing data dissemination protocols. The main aim of this work is to define superior protocol regarding application-specific requirements and

[1]Assistant Professor (Senior Grade), Department of Computer Technology, Kongu Engineering College Perundurai Erode Tamilnadu, India
[2]Professor & Head, Bannari Amman Institute of Technology, Sathyamangalam Erode Tamilnadu, India

performance of the network. The mechanism mainly depends on the simulations performed and also on the definition and execution of a two-tier fuzzy system. The proposed mechanism facilitates the involvement of novel parameters in order to characterize a WSN. Initially, the popular WSN protocols are studied in various scenarios and application requirements to provide a knowledge base by simulations. Next, based on the simulation results obtained, a collection of fuzzy systems are constructed. To guide the construction of the knowledge base, a methodology was proposed. For the purpose of validation of the proposed mechanism and to illustrate its application and efficacy, a case study is portrayed.

Marin-Perianu et al., [11] proposed a distributed fuzzy logic reasoning engine for WSN called D-FLER. In order to fuse individual and neighborhood observations, the D-FLER implements fuzzy logic for the purpose of generating reliable and most accurate result. The D-FLER is evaluated by simulation. Using both fire and non-fire input data; simulation was performed in a fire-detection scenario. By minimizing the false alarm rate, the D-FLER accomplishes good detection times. Additionally, on real sensor nodes the D-FLER is employed and the memory overhead, the execution time and the numerical accuracy are analysed. Therefore, D- FLER is validated to be efficient and easy to be employed in resource-constrained sensor nodes.

In general, the most appropriate fuzzy sets are able to envelop the domains of quantitative attributes for fuzzy association rules mining, but as the characteristics of quantitative data are not known, which makes it difficult. And the domain experts cannot always give the most appropriate fuzzy sets that are unrealistic also. This motivated Kaya et al., [12] to present an automated approach for mining fuzzy association rules. Initially, a genetic algorithm (GA) based clustering method is introduced to adjust the centroids of the clusters that are used in the future like the midpoints of triangular membership functions. Then, for generating the membership functions along with the use of Clustering Using Representatives (CURE) clustering algorithm a different method is presented that belongs to one of the well-organized clustering algorithms among the literatures available. The proposed GA-based approach is compared with other existing approaches described in the literature. On 100K transactions obtained from the US census in the year 2000, the experiment was performed and the results show that in terms of time required for execution and interesting fuzzy association rules, the proposed GA-based approach exhibits best outcome.

In many WSN applications, the essential component is the event detection. The event description area has not received the required consideration. To specify event thresholds the major current event description and detection schemes depend on using the precise values. Often there are imprecise sensor readings that cannot be effectively addressed by the crisp values. Kapitanova et al., [13] uses fuzzy values as an alternative for crisp values and illustrates that it considerably enhances the event detection's accuracy. Compared to other two well-organized classification algorithms, the proposed fuzzy logic approach shows improved event detection accuracy. The exponentially growing size of the fuzzy logic rule-base is the only disadvantage of employing fuzzy logic scheme. The main challenge faced, is to store large rule-bases because the sensor nodes consist of limited memory only. This issue is dealt by introducing many methods that maintains the event detection accuracy and also minimizes the rule-base size by more than 70%.

Haldulakar et al., [14] proposed a new scheme for the strong rule generation. For the generation of rules which are implemented in the optimization methods, a general Apriori algorithm is employed. The most significant way to optimize the rules is achieved by the genetic algorithm. Considering this, a novel fitness function is designed for the optimization of the rule set. This fitness function utilizes the supervised learning concept and then using GA the stronger rule set is efficiently generated. Therefore, the proposed scheme provides best results and for further efficacy it is easily incorporated with the other approaches.

III. MATERIAL AND METHODS

Dataset

Intel lab sensor dataset evaluates the proposed methods [15]. Dataset contains data from 54 sensors including Mica2Dot sensors with weather boards, installed in the Intel Berkeley Research lab.The data collected was time-stamped and collected 31 seconds includedinformation of topology, humidity, temperature, light and voltage value. The dataset includes a log of about 2.3 million sensor readings.Thirty minutes of data collected between March 1, 2004, 9:00 AM to March 1, 2004, and 9:30AMincludes over 9000 messages received in the sink and which was studied to locate the association between sensor motes. Data distribution sent by each mote to the sink is revealed in Fig. 1.

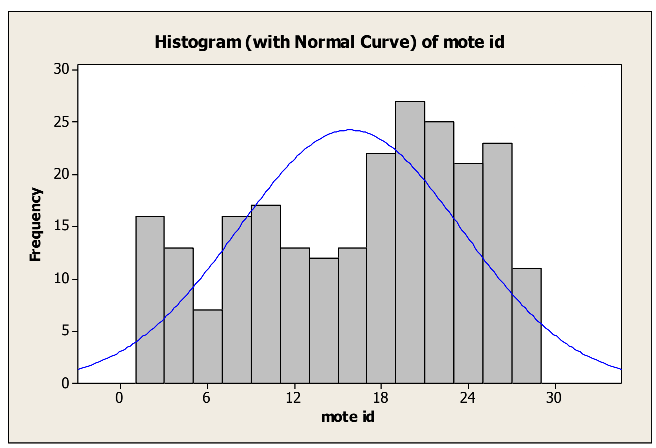

Fig. 1: The distribution of data sent by each mote during the 30 minute interval

IV. ASSOCIATION RULES

If/then statements which are part of Association rules uncover relationships between unrelated data in relational databases/information repositories. Association rules have two parts, an antecedent (if) and the consequent (then). The former is an item within data and consequent is found in combination with an antecedent [16].

Association rules generated by data analysis for regular if/then patterns use criteria *support* and *confidence* to identify important relationships.*Support*reveals the frequency with which an item appears in a database while *Confidence*shows how many times if/then statements were supposed to be true.Association rules extract interesting correlations, frequent patterns, associations or casual structures among item sets in transaction databases/data repositories [17]. Association rules are used in differing areas like telecommunication networks, market and risk management, inventory control etc. Association rule mining finds/locates association rules that satisfy predefined minimum support and confidence in a database. Literature is replete with examples of algorithms proposed for discovering association [18], [19].

User defines *maxscope* the upper distance limit wherean event occurs and *maxhistory*the time frame. Sensors collect event notifications from within *maxscope* and record history with size *maxhistory* of such events. Association mining rule is applied to detect patterns from collected data. Every node mines patterns as follows:

$$I_1 \wedge ... \wedge I_m \Rightarrow E \;\; S, C$$

An event E occurs at node n with support S and confidence C given that antecedents A_i is true. Antecedents for a dataset D, set of transaction, are in the form of

$$I_i = \;\; E_i, D_i, T_i, N_i$$

Every node forwards a subset of discovered patterns to the sink, reducing communication overheads.

V. GENETIC ALGORITHM

Association rule mining is optimized through Genetic Algorithms (GA). Based on quantitative association rules discovery, a GA chromosome encodes a generalized *k*-rule, *k*being the required length [20]. As association rules with many items in consequent is used, an index stored by the first gene, representing end of an earlier portion. To encode a rule into a chromosome, the antecedent and consequent attributes are sorted in two-segment in a spiral, ascending order. The remaining *k* genes encode items; each representsa pair of values. The first value is an attribute's index ranging from 1 to a maximum number of database attributes, while second points to a gapped interval. A gapped interval is the union of finite base intervals received when uniform discretization process is achieved over database attributes. Partitioning the categorical attribute domains is unnecessary as lower and upper bounds coincide. A base interval is represented by an integer number resulting in a gapped interval in a set of integers. Genetic operators applied to a chromosome area follow:

Selection: This is achieved through computing fitness value with a random numberto ensure a chromosome is selected if the product is less than given selection probability (p_s).

Crossover: Selected chromosomes reproduce offspring at a crossover probability (p_c) and the operation includes exchanging a gene segment between first and second chromosomes and vice-versa, all depending on two randomly generated crossover-points.

Mutation: Consideration of both mutation probability (p_m) and fitness value, a chromosome is altered so that the border between antecedent and consequent attributes is changed within the same rule. An operator selects a gene randomly and changes attribute's index with associated gapped interval. The new gapped interval is a combination of base intervals which now form a new attribute sub-domain.

VI. FUZZY LOGIC SYSTEM

Figure 2 depicts a general fuzzy logic system (FLS) structure. The fuzzifier converts crisp input variablesx $\in$ X, where X is a set of possible input variables, to fuzzylinguistic variables through application of corresponding membershipfunctions. Zadeh defines linguistic variables as "those whose values are not numbers but words/sentences in natural/artificial language" [21]. An inputvariable is usually associated with one or more fuzzy sets based on calculated membership degrees. Forexample, a temperature value can possibly be classified as Low and Medium.

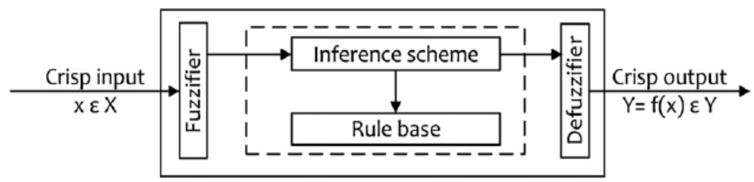

Fig. 2: Generic fuzzy logic system

Fuzzified values are processed through if–then statementsaccording to a set of predefined rules got from domainknowledge.An inferencescheme maps input fuzzy sets to output fuzzy sets and a defuzzifier computes a crisp result from the fuzzy setsoutput using the rules. Control actions are based on the crisp output value. The above three stepsare called fuzzification, decision making and defuzzification respectively.

Fuzzification: The fuzzifier converts crisp values into membership degrees through membership function applications.Membership function determines the association between specificlinguistic values. Membership functions have various shapes and some frequently usedshapes are triangular, trapezoidal, and Gaussian. Membership functions are defined either by relyingon domain knowledge or through application of various learning techniques like neural networks [22], [23] and genetic algorithms [24].

Decision making: A rule-base comprises of a set of linguistic statements known as rules which are of the form IF premise, THENconsequent where premise is composed of fuzzy inputvariables linked by logical functions (e.g., AND, OR, NOT) with the consequent being a fuzzy output variable. The rule-baseis generated as alarge set of every possible value-combination for input linguistic variables constituting the premise. Similarto definitions of membership functions, rule-base is derived from either domain knowledgeor by use of machine learningtechniques. Consider a *t*-input 1-output FLS with rules ofthe form:

$$R^i : \text{IF } x_1 \text{ is } S_1^i \text{ and } x_1 \text{ is } S_1^i \text{.....} \textit{and } x_1 \text{ is } S_1^i \text{ THEN } y \text{ is } A^i$$

A triangular norm isa binary operation like AND or OR applied to fuzzysets provided by membership functions [25].

Defuzzification: Executing rules in rule-bases generates multipleshapes that represent modified membership functions. Commondefuzzifiers are center of gravity, center of singleton, and maximum methods [25]:

The center of gravity approach gets a centroid shape by superimposing shapes as a resultof rule application. The defuzzifier output is the x-coordinate of this centroid.

The defuzzification process is simplified when center of singleton procedure is used,as every rule membership function is defuzzified separately. Each membership function isreduced to a singleton representing the center of gravity function. The simplification is done through singletons being determined during system design. The centre of singleton method approximates center of gravity method.The class of maximum methods determines the outputby selecting a membership function with a maximumvalue. If the maximum is a range, then lower, upper, or the middle value is considered for output value based on the method. With these methods, rule with maximum activity determines output value always.As theclass of maximum methods shows discontinuous outputon continuous input; these methods are unsuitable for use in controllers.

VII. RESULTS AND DISCUSSION

Table 1 tabulates the results of Rule support analysis, and Rule confidence analysis. Fig. 3 and Fig. 4 show the graphs of the same.

Fig. 3 shows the rule confidence analysis for the rules obtained. The number of rules obtained for confidence of 90% is around 20 which are much higher than when only genetic algorithm is used. The number of rules obtained increases proportionally with the decrease in confidence. Similarly, Fig. 4 shows the rule support analysis which also performs better than the genetic algorithm. The proposed fuzzy genetic algorithm achieves better performance.

Fig. 5 shows the relation between the numbers of rules obtained with respect to the size of LHS. Maximum number of rules is obtained when the size of LHS is 2 and the thereafter decreasing with an increase of LHS.

Table 1: Support and Confidence analysis results

Number of rules	Support of rules	Confidence of rules %
0	5300	97
5	4900	95
10	4700	92
15	4200	91
20	3900	90
25	3700	87
30	3500	73
35	3200	69
40	2700	67
45	2450	63
50	2300	58

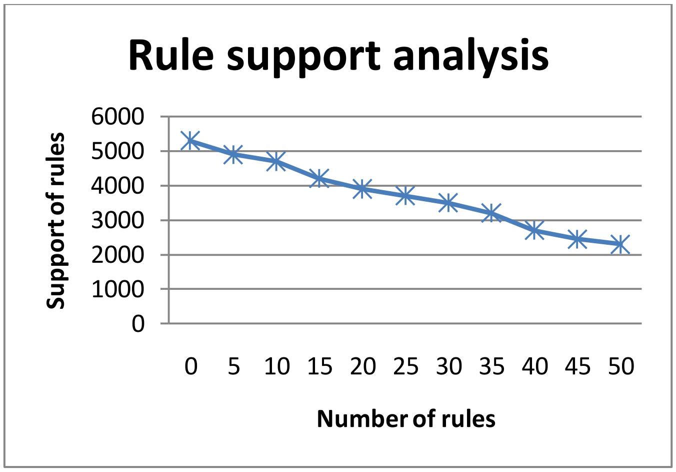

Fig. 3: Rule Support Analysis

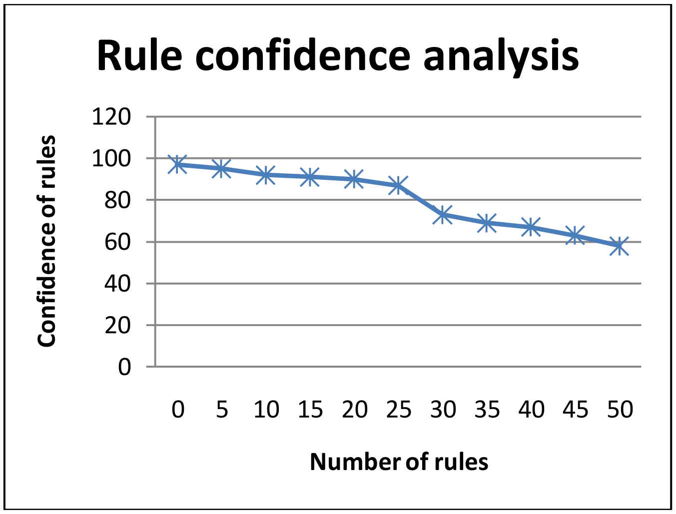

Fig. 4: Rule Confidence Analysis

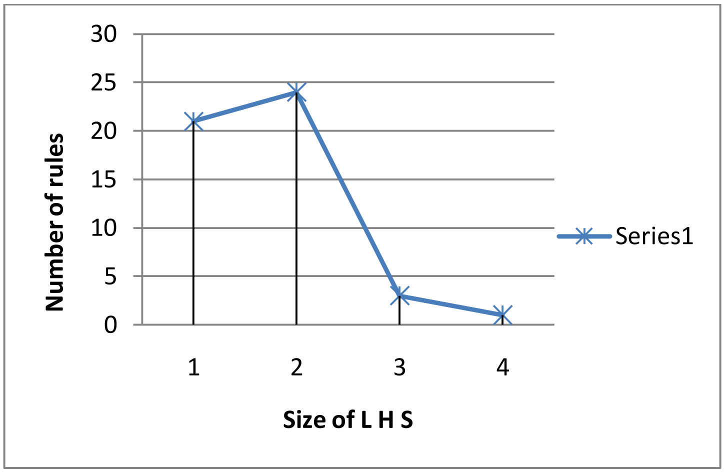

Fig. 5: Number of Rules vs. Size of LHS

VIII. CONCLUSION

In this paper, it is proposed to mine WSN data using association rule to extract patterns. Genetic algorithm and Fuzzy logic is used 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.

REFERENCES

[1] Akyildiz, W Su, Y. Sankarasubramaniam, and E. Cayirci, Wireless Sensor Networks: A Survey, Computer Networks. Vol. 38, pp. 393422, 2002.

[2] MICA2 Mote Datasheet, http : //www.zbow.m/Procl2~cts/P~ocl?~~t~pdf files/Wirelesspdf/6020 - 0042 – 05 4-MICA2.pdf (Accessed 08/06/2006)

[3] V. Raghunathan, C. Schurgers, S. Park, and M.B. Srivastava, Energy Aware Wireless Microsensor Networks, IEEE Signal PmcessingMagarine, Vol. 19, pp. &SO, 2002.

[4] RakeshAgrawal, and RamakrishnanSrikant, Fast Algorithms for Mining Association Rules, Proceedings of the 20th VLDB, pp. 487499, 1994.

[5] S. Madden, and MI. Franklin, Fjording the Stream: Architecture for Queries over Streaming Sensor Data, Pmceedings of the 18th International Conference on Data Engineering. pp. 55, California, 2002.

[6] E. Elnahrawy, and B. Nath, Context-Aware Sensors, Pmceeedings of the 1st EumpeanWorkrhop on Wireless Sensor Networks, pp.77-93, Springer-Verlag, Berlin, Heidelberg, 2004.

[7] F. Zhao and L. J. Guibas. Wireless Sensor Networks: an Information Processing Approach. Morgan Kaufmann publisher, 2002.

[8] C. S. R. Murthy, and B. S. Manoj. Ad Hoc Wireless Networks Architecture and Protocols. Prentice Hall PTR, New Jersey, 2004.

[9] C. Intanagonwiwat, R. Govindan, D. Estrin, J. Heidemann, and F. Silva. "Directed Diffusion for Wireless Sensor Networking", IEEE/ACM Transactions on Networking, Vol. 11, no. 1, pp: 2-16, 2003.

[10] LuciPirmez, Flávia C. Delicato, Paulo F. Pires, Ana L. Mostardinha and Nelson S. de Rezende, "Applying fuzzy logic for decision-making on Wireless Sensor Networks", Fuzzy SystemsConference '07, Proc., IEEE (2007) 1{6

[11] M. Marin-Perianu and P. J. M.Havinga. D-FLER: A distributed fuzzy logic engine for rule-based wireless sensor networks. In International Symposium on Ubiquitous Computing Systems (UCS), pages 86–101, 2007.

[12] Kaya M, Alhajj R (2005) Genetic algorithm based framework for mining fuzzy association rules. Fuzzy Sets Syst 152(3): 587–601

[13] RupaliHaldulakar, Prof. JitendraAgrawal, "Optimization of Association Rule Mining through Genetic Algorithm", International Journal on Computer Science and Engineering, Vol3, No. 3 Mar 2011, pp:1252-1259.

[14] KrasimiraKapitanova, Sang H. Son, Kyoung-Don Kang, "Using fuzzy logic for robust event detection in wireless sensor networks," Ad Hoc Networks 10 (2012) 709–722

[15] Intel Lab Sensor Data. http://berkeley.intel-research.net/labdata/.

[16] Sotiris Kotsiantis, DimitrisKanellopoulos, 2006, "Association Rules Mining: A Recent Overview", GESTS International Transactions on Computer Science and Engineering, Vol.32 (1), pp. 71-82

[17] 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.

[18] R. Agrawal and R. Srikant.Fast Algorithms for Mining Association Rules. In VLDB 1994, Santiago de Chile, Chile, September 1994.

[19] J. Han, J. Pei, and Y. Yin.Mining Frequent Patterns without Candidate Generation. In SIGMOD 2000, Dallas, USA, May 2000

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

[21] 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.

[22] 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.

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

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

[25] G.J. Klir, B. Yuan, Fuzzy Sets and Fuzzy Logic: Theory and Applications, Prentice-Hall Inc., Upper Saddle River, NJ, USA, 1995.