Association Rules for Wireless Sensor Data Based On Fuzzy - Genetic Algorithm

ABIRAMI, T and THANGARAJ, P.

Department of Computer Technology, Kongu Engineering College, Perundurai, Erode, India.
[2]Dept. of CSE, Bannariamman Institute of Technology, Sathy, Tamilnadu, India,
ananthabirami9@gmail.com

Abstract: As wireless sensor networks (WSN) generate a huge amount of data for varied applications, it is important to locate essential knowledge from it. WSN data is generally generated in streams before being forwarded to a sink. WSN performance in adversely affected as raw data leads to higher communication overhead. Frequent patterns are located through association mining. Hence, WSN network data have association mining applied to it only regular raw data patterns are forwarded to the sink which in turn lowers communication overhead. This paper proposes WSN data mining using association rule to extricate patterns. A Fuzzy based genetic algorithm along with the rule is used for efficient extraction.
[ABIRAMI, T,THANGARAJ, P. **Association Rules for Wireless Sensor Data Based On Fuzzy - Genetic Algorithm.** *Life Sci J* 2013;10(4s):554-558] (ISSN: 1097-8135). http://www.lifesciencesite.com. 84

Key words: Wireless Sensor Networks (WSN), Association rules, Genetic Algorithm, Fuzzy Logic.

INTRODUCTION

A wireless sensor network (WSN) is a group of nodes that sense/performs data computations communicating it to others wirelessly [1]. Berkeley motes [2] is a class of inexpensive sensors under this category, where a network is formed by Mica21 Mica2 dots having MTS310 sensor boards, that gather light, temperature, sound and motion readings. But WSN faces challenges in many applications due to limited battery life. Additionally battery replacements are risky when sensors are used in severe environments.

To prolong sensor battery life, network radio transmissions are reduced as they consume energy through sensor nodes [3]. This is offset by shifting majority processing into networked sensor nodes, so that they forward limited/less data to neighbouring nodes/base-stations. Berkeley motes sensors lack both hardware floating point units and memory to handle complicated data mining algorithms as they consume sensor resources quickly. Studies were undertaken for in-network processing algorithms for sensor networks [4]. Related areas are Data Stream Management Systems (DSMS) [5] which filter queries to forward only user relevant data.

A wireless sensor node includes sensing, computing, communication, actuation, and power components integrated on a single/multiple boards, set in a few cubic inches. A WSN has many nodes communicating through wireless channels sharing and processing information being used globally for 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 and in homes to realize smart homes and for patient monitoring for humans [2, 7]. Sensor nodes should organize appropriate network infrastructure on initialization, with multi hop, intra sensor node connections. Continuous/event driven working modes are used to collect acoustic, seismic, infrared or magnetic environment information. Location/positioning information is received through GPS (global positioning system)/local positioning algorithms. This information from a network is processed to ensure a global view of monitoring phenomena/objects. WSN philosophy is that network's aggregate power is enough for special tasks [8] despite limited individual sensor node capability.

WSNs generate much data in streams and forwarded to a sink [9], leading to excessive communication overheads affecting WSN performance. Association mining processes and locates frequent patterns. When applied in-network to WSN, only frequent data patterns need transmission to the sink reducing communication overheads.

This paper proposes mining WSN data using association rule to extract patterns. Genetic algorithm/Fuzzy logic is used with association rule for effective rule extraction from quantitative data. This method uses INTEL dataset with 54 Mica2Dot motes with temperature, humidity and light sensors.

RELATED WORKS

Pirmez et al., [10] proposed a fuzzy-based decision-making mechanism to choose data dissemination protocols which aims to define superior protocol regarding application-specific requirements and network performance for WSNs. The mechanism is simulations dependent and on the definition/execution of a two-tier fuzzy system. The proposed procedure facilitates novel parameters involvement to characterize a WSN. To begin with, popular WSN protocols are studied in various scenarios/application requirements to ensure a knowledge base by simulations. Then, based on simulation results, some fuzzy systems are constructed. A methodology was proposed to guide knowledge base construction and a case study is given to validate the proposed mechanism and also to illustrate its application and efficacy.

Marin-Perianu et al., [11] proposed a distributed fuzzy logic reasoning engine for WSN called D-FLER. To fuse individual and neighborhood observations, the D-FLER implements fuzzy logic to generate reliable and accurate results and it is evaluated through simulation. Simulation was performed in a fire-detection scenario with fire and non-fire input data; by minimizing false alarm, D-FLER accomplished good detection times. D-FLER is employed and memory overhead, execution time and numerical accuracy are analysed on real sensor nodes. Therefore, D- FLER is validated to be efficient and easily employable in resource-constrained sensor nodes.

Usually, the most appropriate fuzzy sets envelop domains of quantitative attributes for fuzzy association rules mining, but as characteristics of quantitative data is unknown it is difficult. And domain experts are unable to provide appropriate unrealistic fuzzy sets. Thus Kaya et al., [12] presented an automated approach for mining fuzzy association rules. A genetic algorithm (GA) based clustering method – to begin with - is introduced to adjust clusters centroids used in the future like triangular membership functions. Then, to generate membership functions along with the use of Clustering Using Representatives (CURE) clustering algorithm a different method belonging to a well-organized clustering algorithm is presented. The proposed GA-based approach is compared to existing approaches described in literature. The experiment was performed on 100K transactions obtained from the US census in 2000, and the results show that with regard to time required for execution and fuzzy association rules the proposed GA-based approach shows good outcome.

Event detection is essential component in many WSN applications. Event description area has not received much consideration till date. To specify event thresholds, major event description/detection schemes depend on precise values. There are often imprecise sensor readings which are unable to be addressed by crisp values. Kapitanova et al., [13] uses fuzzy values as an alternative for crisp values and shows it enhances event detection's accuracy considerably. Compared with other two well-organized classification algorithms, the proposed approach shows improved event detection accuracy. Fuzzy logic rule base's exponentially growing size is its only disadvantage. The main challenge is storing large rule-bases as sensor nodes have limited memory. This issue is handled through introducing many methods that ensure event detection accuracy and lower rule-base size by more than 70%.

Haldulakar et al., [14] proposed a new scheme to store strong rule generation. A general Apriori algorithm is employed for rule generation in optimization methods. A way to optimize rules is through a genetic algorithm. Considering this, a novel fitness function is designed for rule set optimization. This fitness function uses supervised learning concept and then a stronger rule set is efficiently generated using GA. Hence, the proposed scheme provides best results. It is also easily incorporated with other approaches.

MATERIALS AND METHODS

Dataset: Intel lab sensor dataset evaluates the proposed methods. [15]. Dataset has information from 54 sensors including Mica2Dot sensors with weather boards in the Intel Berkeley Research lab. Collected data was time-stamped and though collected in 31 seconds it had information on topology, humidity, temperature, light and voltage value. The dataset comprised of about 2.3 million sensor readings. Thirty minutes of data collected between March 1, 2004, 9:00 AM to March 1, 2004, and 9:30AM includes over 9000 messages in the sink. This was studied to locate intra sensor mote association. Figure 1, reveals data distribution sent by each mote to the sink.

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

Data analysis generated Association rules for regular if/then patterns use criteria support and confidence to identify important relationships. Support reveals frequency of an item's occurrence in a database while Confidence shows the number of times, if/then statements were supposedly 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 areas like telecommunication networks, market and risk management, inventory control etc. Association rule mining finds/locates association rules satisfying predefined minimum support and confidence in a database. Literature is replete with examples of algorithms proposed for discovering association [18, 19].

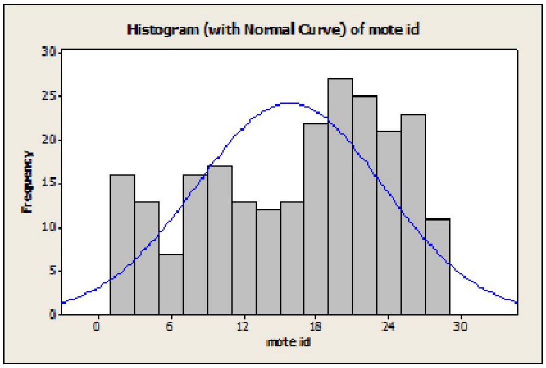

Figure 1: The distribution of data sent by each mote during the 30 minute interval.

User defines *maxscope* as the upper distance limit where an event occurs and *maxhistory* the time frame. Sensors collect event notifications from *maxscope* and record history with size *maxhistory*. Association mining detects patterns from collected data. Every node mines patterns as follows:

$$I_1 \wedge \ldots \wedge I_m \Rightarrow E[S,C]$$

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 transactions are in the form of

$$I_i = (E_i, D_i, T_i, N_i)$$

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

Genetic Algorithm: Genetic Algorithms (GA) optimize association rule mining. A GA chromosome encodes a generalized *k*-rule, *k* being the required length [20], based on quantitative association rules discovery. An association rule of many items in consequent is used, an index stored by first gene, represents end of a previous portion. To encode a rule into a chromosome, antecedent and consequent attributes are sorted in two-segment in a spiral, ascending order. The remaining *k* genes encode items, each representing a pair of values, the first of which is an attribute's index ranging from 1 to a maximum number of database attributes, while the second reveals a gapped interval which in turn is the union of finite base intervals received when uniform discretization process is achieved over database attributes. Compartmentalising categorical attribute domains is not needed as lower/upper bounds coincide. Base interval is represented by an integer number leading to a gapped interval in an integers set. Genetic operators applied to a chromosome are as follows:

Selection: Achieved through computing fitness value with a random number to ensure selection of a chromosome if the product is less than given selection probability (p_s).

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

Mutation: Considering mutation probability (p_m) and fitness value, a chromosome is altered such that the border between antecedent and consequent attributes changes within the same rule. A gene is selected randomly and changes attribute's index with associated gapped interval. The new gapped interval combines base intervals which are now a new attribute sub-domain.

Fuzzy Logic System: A general fuzzy logic system (FLS) structure is seen in Figure 2. Fuzzifier converts crisp input variables x X, where X is a set of possible input variables, to fuzzy linguistic variables through membership functions application. Zadeh defines linguistic variables as ''those whose values are not numbers but words/sentences in natural/artificial language'' [21]. An input variable is usually associated with one/more fuzzy sets based on calculated membership degrees. For example, a temperature value is classified as Low and Medium.

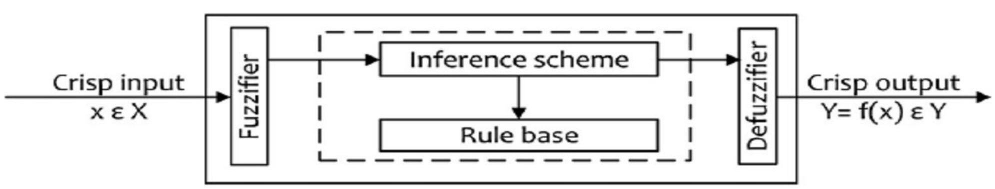

Fig. 2: Generic fuzzy logic system.

If-then statements process fuzzifies values based on predefined rules from domain knowledge. An inference scheme maps input fuzzy sets to output fuzzy sets and a defuzzifier computes crisp results from fuzzy sets output through use of rules. Control actions are based on crisp output value and these steps are called fuzzification, decision making, and defuzzification respectively.

Fuzzification: The fuzzifier converts crisp values into membership degrees through membership function applications which determine association between specific linguistic values. Membership functions have various shapes. Frequently used shapes are triangular, trapezoidal, and Gaussian. Membership functions are defined either by relying on domain knowledge or through various learning techniques application like neural networks [22, 23] and genetic algorithms [24].

Decision making: A rule-base comprises of linguistic statements sets known as rules which are of IF premise, THEN consequent where premise is composed of fuzzy input variables linked by logical functions (e.g. AND, OR, NOT) with consequent being a fuzzy output variable. Rule-base is generated as a large set of all possible value-combination for input linguistic variables making up the premise. Similar to membership function definitions, rule-base derives from domain knowledge or through use of machine learning techniques. Consider a t-input 1-output FLS with rules of the form:

$$R^i: \text{ IF } x_1 \text{ is } S_1^i \text{ and } x_1 \text{ is } S_1^i \ldots\ldots and\ x_1 \text{ is } S_1^i$$

$$\text{THEN } y \text{ is } A^i$$

A triangular norm is a binary operation like AND or OR applied to fuzzy sets provided by membership functions [25].

Defuzzification: Executing rules in a rule-base creates multiple shapes representing modified membership functions. Common defuzzifiers are centres of gravity, centre of singleton and maximum methods [25].

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

The center of gravity develops a centroid shape by superimposing shapes due to rule application. The defuzzifier output is the centroid's x-coordinate. Defuzzification process is made simple when center of singleton procedure is used, as all rule membership functions are defuzzified separately. Every membership function is becomes a singleton representing a center of gravity function. The simplification is through singletons being determined during system design.The center of singleton method is almost similar to center of gravity method. Maximum methods class determines output by selecting a membership function with maximum value. If maximum is a range, then lower, upper, or middle values are considered for output value based on method. With methods, rule with maximum activity always determines output value. As maximum methods class shows discontinuous output on continuous input, they are unsuitable for use in controllers.

RESULTS AND DISCUSSION

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

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

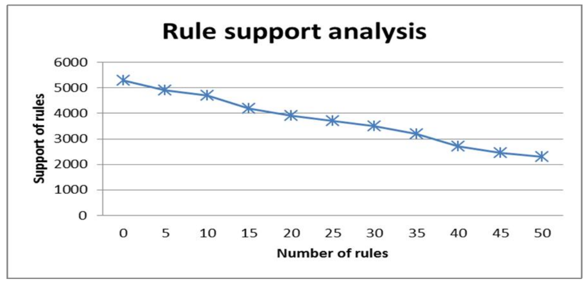

Figure 3: Rule Support Analysis

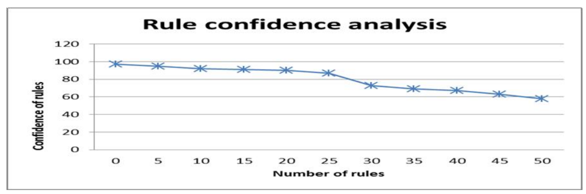

Figure 4: Rule Confidence Analysis

Figure 3 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. Similarly, Figure 4 shows rule support analysis performing better than a genetic algorithm. The proposed fuzzy genetic algorithm shows improved performance.

Figure 5 explains relation between rules number obtained with regard to LHS size. Maximum rules number is obtained when LHS size is 2 and then decreases with LHS increase.

CONCLUSION

This paper proposes to mine WSN data to extract patterns. Genetic algorithm and Fuzzy logic are used with association rule for rule extraction from quantitative data. This method is evaluated through INTEL dataset. Fuzzy logic and Genetic algorithm are used with association rule for rule extraction. Experiments show that the proposed fuzzy genetic algorithm is effective along with Association rule mining.

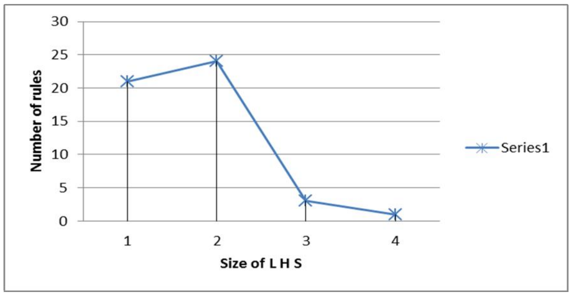

Figure 5: Number of Rules vs. Size of LHS

REFERENCE

[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 Pmcessing Magarine, Vol. 19, pp. &SO, 2002.

[4] Rakesh Agrawal, and Ramakrishnan Srikant, Fast Algorithms for Mining Association Rules, Proceedings of the 20th VLDB, pp. 487499, 1994.

[5] S. Madden, and MI. Franklin, Fjording the Stream: An 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 Eumpean Workrhop 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] Luci Pirmez, 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 Systems Conference '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] Rupali Haldulakar, Prof. Jitendra Agrawal, “Optimization of Association Rule Mining through Genetic Algorithm”, International Journal on Computer Science and Engineering, Vol3, No. 3 Mar 2011, pp:1252-1259.

[14] Krasimira Kapitanova, 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, Dimitris Kanellopoulos, 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.

2/20/2013