www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 6 Issue 8 August 2017, Page No. 22327-22332
Index Copernicus value (2015): 58.10 DOI: 10.18535/ijecs/v6i8.17

Distributed Data mining In Wireless Sensor Network Using Fuzzy Naïve Byes

Nitika Malik, Pankaj Kumar

Computer Science
Dr. APJ Abdul Kalam Technical University
nitika20_malik@yahoo.co.in

Computer Science
Dr. APJ Abdul Kalam Technical University
unpankaj@gmail.com

Abstract
The wireless sensor nodes are getting smaller, but Wireless Sensor Networks (WSNs) are getting larger with the technological developments, currently containing thousands of nodes and possibly millions of nodes in the future. Therefore, effective and trustworthy event detection methods for the WSN require robust and intelligent methods of mining hidden patterns in the sensor data, while supporting various kinds of dynamicity. Due to the fact that events are often functions of more than one attribute, data fusion and use of more features can help increasing event detection rate and reducing false alarm rate. In addition, sensor fusion can lead to more accurate and robust event detection by eliminating outliers and erroneous readings of individual sensor nodes and combining individual readings. There is a need for intelligent and energy efficient monitoring methods, made possible by novel data mining and classification methods, and the work reported in this paper involves such a novel energy efficient data mining scheme for forest cover type classification based on random forests and random trees & dual event detection decisions. The experimental validation of the proposed data mining scheme on a publicly available UCI machine learning dataset, shows that the proposed random forest and random tree based approach perform significantly better than the conventional statistical classifiers, such as Naïve Bayes, discriminant classifiers, and can lead towards energy efficient, intelligent monitoring and characterization of large physical environments instrumented using Wireless sensor networks.

Keyword : Distributed Data mining, WSN, Fuzzy naïve bayes, classification, prediction etc.

1. Introduction

A wireless sensor network is a wireless network consisting of spatially distributed autonomous devices using sensors to monitor physical or environmental conditions. Sensor devices currently used are computer like devices. They have a CPU, Main memory, Operating system and a suite of sensors. WSNs have been successfully applied for national security and military applications, data collection, monitoring and surveillance and medical care .WSN produces a large dataset. The capabilities for collecting and storing data have far outpaced someone's abilities to analyze, summarize, and extract knowledge from these data. So, the transmission of all sensory data to the sink can be reduced by using data mining techniques. When each sensor node only selects important data, which is usually the fault data, to send to the fusion center, then energy consumption, network traffic can be reduced, and it can extend the lifetime of sensor networks. WSNs have limited computational and energy resource as they are small in size. Raw data collected from the often suffer from inaccuracy and incompleteness. Inaccurate/incomplete data measurements of WSN are often known as WSN anomalies. Anomalies are defined as observations that do not correspond to a well defined notion of normal behaviors. Anomalies in WSNs can be caused by errors, malfunctioning/failure of nodes

and attacks. It is important to effectively detect and respond to anomalies. Besides these anomalies false data can be injected by faulty sensor nodes in various ways by relaying and data aggregation. Data aggregation is essential to reduce data redundancy and/or to improve data accuracy, false data detection is critical to the provision of data integrity and efficient utilization of battery power. In this paper we will focus on rectifying or finding these faults using naïve bayes classification with fuzzy rule based system.

2. Related Work

In [1], Y. Gao et. al. presented the use of data mining methods in understanding building energy performance of geothermal, solar and gas burning energy systems. In their work, classification methodology was used to analyze a combination of internal and external ambient conditions. Developed Classification rules were analyzed for their application to modify control algorithms and to apply results to generalize hybrid system performance. The results of this study can be generalized for an entire building, or a set of buildings, under a single energy network subject to the same constraints.

In [2] and [3], the authors use general SQL primitives to define events in sensor networks. The limitation of this approach is that the events can only be defined by predicates on sensor readings with very simple temporal and spatial constraints connected by AND and OR operators. Madden et al. have extended the SQL primitives by incorporating streaming support where a desired sample rate can be included [5]. Li et al. define events using a sub-event list and confidence functions in SQL [4]. However, SQL is not very appropriate for describing WSN events. Some of its drawbacks include that it: (i) cannot capture data dependencies and interactions among different events or sensor types; (ii) does not explicitly support probability models; (iii) is awkward in describing complex temporal constraints and data dependencies; (iv) lacks the ability to support collaborative decision making and triggers [6]; (v) does not support analysis of the event system. Another approach to formally describe events in WSNs has been the use of extended Petri nets. This was initially proposed by Jiao et al. [7]. The authors design a Sensor Network Event Description Language (SNEDL) which can be used to design Petri nets that specify event logic. Petri nets were also used in

MEDAL [8], an extension of SNEDL that supports the description of additional WSN specific features such as communication and actuation. Both SNEDL and MEDAL, however, use crisp values in the definitions of their Petri nets.

Fuzzy sets and logic were introduced by L. Zadeh in 1965.

Ever since then, numerous fields have taken advantage of their properties. In WSNs, fuzzy logic has been used to improve decision-making, reduce resource consumption, and increase performance. Some of the areas it has been applied to are cluster-head election [10, 11], security [12, 13], data aggregation [14],routing [15, 16], MAC protocols [17], and QoS [18, 19]. However, not much work has been done on using fuzzy logic for event description and detection.

In D-FLER [20] fuzzy logic is used to combine personal and neighbors' observations and determine if an event has occurred. Their results show that fuzzy logic improves the precision of event detection. The use of fuzzy values allows

D-FLER to distinguish between real fire data and nuisance tests. However, the approach used in D-FLER does not incorporate any temporal semantics. In addition, since all of the experiments last only 60 seconds after the fire ignition,

the authors do not analyze the number of false alarms raised by D-FLER.

3. Propesed Work

In the proposed work we are working on a distributed data mining technique with a combination of Naive Bayes and Fuzzy rule based system to detect the fault node. So that we can obtain a network with less data redundancy, energy efficient and accurate data across the network.

3.1 Distributed approach for finding anomaly

In distributed approach, the detection agent is installed in every node. It monitors the behavior of neighboring node within its transmission range locally to detect any abnormal behavior. To perform a real time anomaly detection, some rule based detection techniques are used in a node. Node listens promiscuously to neighboring nodes within its transmission range to collect data

necessary for anomaly detection. The collected data will be analyzed to detect any deviation from normal behavior using neighboring historical data stored in the memory. Once anomalies have been detected, an alarm is sends to alert the base station or neighboring nodes.

3.2 Naïve Bayes Classifier

A naive Bayesian classifier is a simple probabilistic classifier. That based on apply Bayesian theorem with well-built (naive) independence assumption. A naive Bayesian classifier assumes that the absence or presence of the given feature is unrelated to the absence or presence of any other feature, given the particular class variable. Bayesian classifiers use Bayes theorem that may expressed as,

$$p(Cj \mid d) = \frac{p(d|\, Cj\,)p(Cj)}{p(d)}$$

$p(Cj\,|\mathrm{d})$ = probability of instance d being in class Cj,
$p(d|\, Cj)$ = probability of generating instance d given class Cj,
$p(Cj)$ = probability of occurrence of class Cj,
$p(d)$ = probability of instance d occurring.

3.3 Fuzzy Rule Based Logic

Fuzzy logic is defined as the logic of human thought. In Fig 1 shows Fuzzy based Anomaly Detection method fault data will be detected and also classified the data that falls into the one category that is normal data or the fault data. We can get more accurate values than other methods while using fuzzy logic. From the output, data is classified as is normal data or the fault data. Steps involved in the process:

1. Input will be fuzzified.
2. Fuzzification will be done by membership functions.
3. Deriving inference rules.
4.Defuzzification

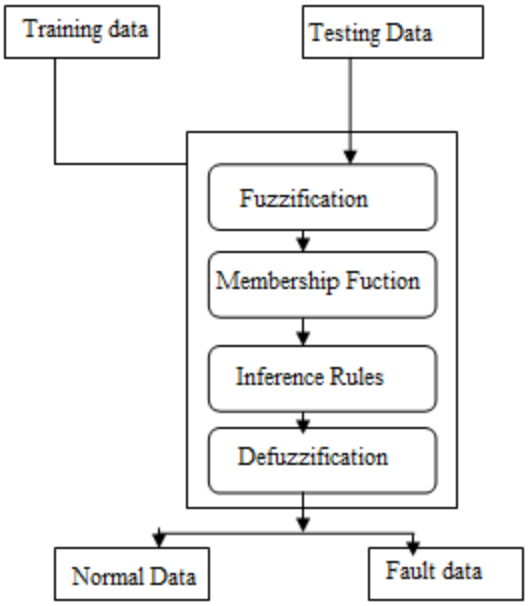

Fig. 1 Fuzzy based detection system

From the fig 1 we can see the overall functionality of the fuzzy logic, proposed a fuzzy logic based intrusion detection system to detect the fault node in WSN. This paper used combination of specification and anomaly based detection methods. This paper used fuzzy inference system for making the fuzzy rules to take the decision. There are some parameters for checking the behavior of node is malicious or not.

3.3.1.Fuzzification process

The Dataset used in the experiment have following attributes which are used as input for the fuzzy logic based intrusion detection system:

1.Network Design
2.Number of nodes
3.Boundary
4.radius
5.Average degree

3.3.2. Membership Functions

Membership function defines the fuzziness in a fuzzy set irrespective of the elements in the set, which are discrete or continuous [18]. Membership function can be thought of as a technique to solve a problems on the basis of experience rather than a knowledge .Available membership functions are trapezoidal, Gaussian , triangular methods.

3.3.3.Inference Rule

Fuzzified output will be compared using inference rules then from the output data will be classified. So that we can say that the data is normal or not.

3.3.4.Defuzzification

This is about output of the Fuzzification process. Output of data classification process:

1.Perfect
2.Imperfect
3.Minimal
4.Failure

3.4 Detection Model

In our study detection model detects and classify the dataset in the various categories such as perfect, imperfect , minimal, failure. These categories are the status of networks used in the data set. This model improve the data accuracy and save the energy in the network by reducing data redundancy. By using two detection models such as naive bayes and fuzzy logic overall accuracy and efficiency will be improved. According to this model prediction for different status under fuzzy rules will be identified easily and a secure and efficient network will be obtained.

In this approach ,in fig 2 we first Load the train Data and apply the Preprocessing and remove the Duplicate Data. Classification is a famous managed learning technique in data mining. It is used to spiteful meaningful information from large datasets and can be efficiently used for predicting unidentified classes. In this investigation, classification is applied to a cluster dataset to predict 'cluster' for different category of the WSN. The cluster dataset used in this research is real in environment; it was collected from socio-economic data from different areas in Delhi. This paper compares the two different classification algorithms namely, Naïve Bayesian (NB) and Enhance Naïve Byes (ENB) for predicting 'Cluster Category' for different states in WSN.

4. Result

In all previous classical approaches , the accuracy plays an important role in the approval of that model for the application. Result shows main implementation of the user data defined in analysis. The accuracy of the clustering comes out to be 94.75% and 5.28% comes under false alarm rate.

In our study, to further understand the behavior of our fuzzy logic approach, we have compared it to well established classification algorithms: a Naive Bayes classifier. Fuzzy logic is more suitable than other algorithms for WSN event description since, unlike Bayes classifiers and decision trees where values are considered to be nominal, it works with continuous values, which is exactly what the sensor readings are. In addition, specifying the membership functions is more intuitive and simpler than building a probability model.

Here with fuzzy logic approach ,experiment worked under four fixed states – perfect, imperfect , minimal and failure. Selecting any of these state meant that a rule is defined and random attributes were given as input to the machine , these inputs are selected from the dataset used to train the machine. On applying the fuzzy naïve bayes algorithm we get more efficient and accurate prediction values for different states.

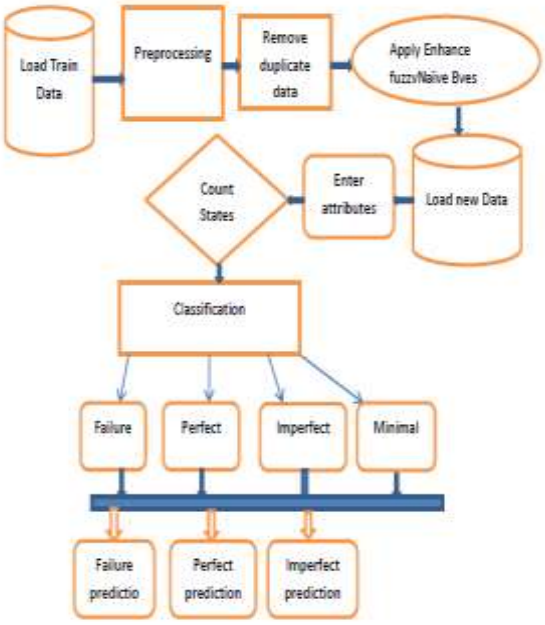

Fig 2:Detection model for data mining using fuzzy naive bayes

5. Conclusion

The main purpose of sensors networks for any detection like fire, earthquake etc. it is to collect

the monitored original data, and provide basic information and decision support for monitoring center. Also, data mining algorithm has to be sufficiently fast to process high-speed arriving data. The sensor scenario may often require in-network processing, wherein the data is processed to higher level representations before further processing. On this way, individual nodes access and process local information and in order to achieve a collective decision, they must communicate to neighbor nodes, to send local and partial models and negotiate a common decision. In this case, whole data cannot be stored and must be processed immediately by their compressing and filtering for more effective mining and analysis in order to generate actionable insights from massive, disparate and dynamic data, in real time or near real time. This reduces the transmission costs, and the data overload from a storage perspective.

The aim of this paper was to make a comparative analysis between different classification algorithms, applied on nominal and real time recorded network data, and to see which of applied methods has the best prediction performances in order to reduce sensor node activity and bandwidth. For evaluation of classification methods next measures can be used: accuracy, speed, time to construct the model (training time), time to use the model (classification/prediction time), robustness (handling noise and missing values), scalability, interpretability, understanding and insight provided by the model and other measures (e.g., goodness of rules). For simulation results the standard measures for evaluation of the accuracy of the predictive model (the number of perfect and imperfect classified on the basis of, sensor id, Network, Reading, No of nodes, Average degree, boundary, event radius and result) were applied while prediction value was used for a more detailed analysis of the class attribute distribution. According to chosen evaluation measures, fuzzy Naïve byes which was used as a base classification model, has shown the best prediction power in performed experiments. Even applied data mining methods are efficient, none of them can be considered as unique or general solution. On the contrary the selection of a correct data mining algorithm depends of an application and the compatibility of the observed data set. Thus, each situation should be considered as a special case and choice of adequate predictor or classifier should be performed very carefully based on empirical arguments. Real time data set used in those experiments is just an example, and for getting better and more accurate results the larger data sets should be used.

6. Future Work

Our future work includes investigating applicability of other machine learning methods for distributed and online event detection in WSN and defining a generic mechanism to detect more events. Many different types of methods are combined to overcome individual limitations and benefit from each other's merit and measure the performance of data reduction method in wireless sensor networks. it can be remarked that designing a new anomaly based IDS is a true challenge. Since, it must satisfy the performance aspects as well as the security aspects. Also, a newly feature selection methods can be adopted. In addition, relying on a newly data mining method rather than traditional classifiers based on neural networks to properly select the clustering parameters can enhance the refining process. it can be remarked that designing a new anomaly based IDS is a true challenge. Since, it must satisfy the performance aspects as well as the security aspects. Also, a newly feature selection methods can be adopted. In addition, relying on a newly data mining method rather than traditional classifiers based on neural networks to properly select the clustering parameters can enhance the refining process.

7. References

[1] Y. Gao, E. Tumwesigye, L. Allan, B. Cahill, K. Menze 2010 Using Data Mining in Optimisation of Building Energy Consumption and Thermal Comfort Management. In Proc.2nd International Conference on Software Engineering and Data Mining, IEEE.

[2] Cornell Database Group-Cougar, \http://www.cs.cornell.edu/bigreddata/cougar/ ."

[3] R.Govindan,J.Hellerstein,W.hong,S.Madden, M.Franklin,and S.Shenker,\The sensor Network as a database,Computer Science Department,University of Southern California,Technical repoet 02-771,2001.

[4] S.Li,S.H.Son,and J.Stankovic ,\ Event detection services using data service using data service middleware in distributed sensor network," IPSN,pp.502-517,2003.

[5] S.Madden,M.Franklin,J.Hellerstein,and W.hong,\The design of an acquisitional query processor for sensor networks," in SIGMOD,2003,pp.491-502.

[6] Michael Franklin, "Declarative interfaces to sensor networks," Presentation at NSF Sensor Workshop,2004.

[7] B.Jiao, S.Son, and J. Stankovic, \ GEM : Genetic event service middleware for wireless sensor networks," INSS , 2005.

[8] K.Kapitanova and S.Hon.Son, \ MEDAL: A Compact event description and analysis language for wireless sensor networks," INSS, 2009.

[9] O.Amft,M.Kusserow and G. Troster, "Probabilistic parsing of dietary activity events," BSN, pp.242-247,2007.

[10] I.Gupta, D.Riordan, and S. Sampalli, " Cluster-head election using fuzzy logic for wireless sensor networks," in CNSR,2005,pp.255-260.

[11] J.Kim, S.Park, Y.Han, and T. chung, "CHEF: Cluster head election mechanism using fuzzy logic in wireless sensor networks," ICACT, pp.654-659, 2008.

[12] H.Lee and T.Cho, "Fuzzy logic based key disseminating in ubiquitous sensor networks." ICACT, pp.958-962,2008

[13] B.Kim, H.Lee, and T.Cho, " Fuzzy key dissemination limiting method for the dynamic filtering -based sensor network," in ICIC,2007 ,pp.263-272.

[14] B.Lazzerini,F. Marcelloni, M.Vechio, S.Croce, and E. Monaldi, "A fuzzy approach to data aggregation to reduce power consumption in wireless sensor networks," NAFIPS,pp 436-441, 2006.

[15] J.Kim, T.Cho, "Routing Path generation for reliable transmission in sensor networks using GA with fuzzy logic based fitness function,"ICCSA,pp.637-648,2007.

[16] S.-Y.Chiang,and J.-L. Wang, "Routing analysis using fuzzy logic system in wireless sensor networks," KES,pp,966-973,2008.

[17] Q.Ren and Q.Liang, " Fuzzy logic-optimized secure media access control(fsmac) protocol wireless sensor networks," CIHSPS,pp.37-43,2005.

[18] S.A.Munir,Y.W.Bin,R.Biao, and M.Jian, " Fuzzy logic based congestion estimation for QoS in wireless sensor networks," WCNC, pp.4336-4341,2007.

[19] F.Xia, W.Zhao,Y.Sun, and Y.-C.Tian, "Fuzzy Logic control based QoS management in wireless sensor network,"Sensors,pp.3179-3191,2007.

[20] M.Martin-Perianu and P.Havinga, " D-FLER: A distributed fuzzy logic engine for rule- based wireless sensor networks", in UCS,2007,pp.86-101.