Online Anomaly Detection Method Based on BBO Ensemble Pruning in Wireless Sensor Networks

Zhiguo Ding[1,2], Minrui Fei[1] , Dajun Du[1], and Sheng Xu[1]

[1] Shanghai Key Laboratory of Power Station Automation Technology, School of Mechatronics Engineering and Automation, Shanghai University, Shanghai, 200072 China
[2] College of Mathematics, Physics and Information Engineering, Zhejiang Normal University, Jinhua, Zhejiang, 321004, China
dingzhiguo@shu.edu.cn, mrfei@staff.shu.edu.cn,
{ddj559,xsheng1980}@163.com

Abstract. Online anomaly detection in wireless sensor networks (WSNs) has been explored extensively. In this paper, exploiting the spatio-temporal correlation existed in the sensed data collected from WSNs, an online anomaly detector for WSNs are built based on ensemble learning theory. Considering the resources constrained in WSNs, ensemble pruning based on bio-geographical based optimization (BBO) is conducted. Experiments conducted on a real WSN dataset demonstrate that the proposed method is effective.

Keywords: Online Anomaly detection, Ensemble pruning, Biogeography-based Optimization (BBO), Wireless sensor network (WSN).

1 Introduction

A wireless sensor network (WSN) typically consists of a large number of small, low-cost sensor nodes, which are integrated with sensing, processing and wireless communication capabilities [10], and have been received considerable attention for multiple types of applications. However, wireless sensor network are highly susceptible to suffer various kinds of failures, such as hardware malfunctions, energy depletion, and intrusion, etc, which results in the observations anomalous. Under the context of resource constraints in WSNs, identifying anomalies data timely becomes much more important, which can save the network resources and help decision-making swiftly.

Up to now, there are many anomaly detection techniques specifically developed for WSNs emerged [10]. Ensemble learning, as the first concerns method in machine learning community, has attracted many researchers attention. A large body of theoretical and empirical research shows that ensemble learning can improve the generalization performance. However, ensemble learning method requires build and store multiple detectors which incur large amount of computation and storage resource requirement. Consequently, ensemble pruning is a possible strategy to obtain the better (at least same) performance compared to the initial ensemble. From the perspective of resource saving in WSNs, the pruning [11] is a necessary strategy for anomaly detection in WSNs.

S. Ma et al. (Eds.): LSMS/ICSEE 2014, Part I, CCIS 461, pp. 160–169, 2014.

In this paper, considering the spatio-temporal correlation of sensed data in WSNs, a distributed anomaly detection method for WSNs is proposed based on the ensemble learning theory. In order to reduce the high communication requirements caused by broadcast multiple detectors, ensemble pruning based on the BBO [7] is used. The pruned ensemble detector is then used to detect anomalous data.

The remaining of this paper is organized as followed. In section 2, we presented our propose anomaly detection method; in section 3, experiment and results analysis are outlined; Conclusion is presented in section 4.

2 Proposed Online Anomaly Detection Method

2.1 Problem Statement

Generally, the WSNs is composed of a large amount of sensor nodes that can be self-organized into clusters. It can be represented as a graph $G=(V,E)$, where $V=\{v_1,v_2,...v_{|V|}\}$ is a finite set of vertices and $E=\{e_1,e_2,...e_{|E|}\}$ is a finite set of edges, vertex ($v_i, i=1,...,|V|$) and edge ($e_i, i=1,...,|E|$) refers to sensor nodes and the one-hop or multi-hop communication link reachable between sensor v_i and v_j, respectively.

In WSNs, some clusters are formed based on node geographical positions information. In order to concisely describe our proposed anomaly detection method, a relatively small sub-network consisted of some sensor nodes deployed densely is taken into account, which forms a cluster C_i consisting of one cluster head and a number of sensor node represented as CH_i and $N_{i,j}: j=1...|C_i|$, respectively.

For one cluster, $C_i=\{CH_i, N_{i1},... N_{im}\}$, which represents a closed neighborhood of node CH_i, $N_{i,j}\in V$. Each sensor node in the sub-network at every time interval Δt measures a data vector. For the cluster head CH_i, the observation is $X^i=(x_1^i, x_2^i, ..., x_d^i)$, where d denotes the dimension. For the j-th neighbor node, $N_{i,j}$, the observation is $X_k^i=(x_{k,1}^i, x_{k,2}^i, ..., x_{k,d}^i)$. Each node in the cluster can do the same work and thus scales well to the whole WSNs.

2.2 Spatial and Temporal (Spatio-Temporal) Correlation of Sensed Dataset

For the sensed dataset, we analyzed the spatio-temporal correlation firstly, which will be used later to build online ensemble detector.

The sensor dataset collected is a time series dataset. A time series is a sequence of value $X=\{x(t), t=1...n\}$ which is a non-random order. For dataset collected form WSNs, Analysis these observations can help to understand the data trend over time and build the appropriate detector. Before training the detector, the foremost requirement is data pre-processing, some data processing method have been emerged and applied to sequence dataset. The commonly methods, such as polynomial fitting, moving averages, differencing, or double exponential smoothing, are used widely [9]. Considering the limited computation resources, one of efficient non-parametric technique, first

differencing, can be used to eliminate the trend and obtain a stationary time series in WSNs, which can be formulated as:

$$X' = \{x'(s,t) = x(s,t) - x(s,t-1) : t = 2,3...n\} \quad (1)$$

Besides, the sensor nodes are always deployed with high density, consequently, the space redundancy existed. A data sequence, $X = \{x(s), s = 1...m\}$, is collected from m sensor nodes with different neighbor locations at a time interval. This dataset can help to understand the spatial correlation structure of data and predict the data value at a location nearby

2.3 Proposed Ensemble Learning Method of Anomaly Detection in WSNs

Considering the spatio-temporal correlations that existed among sensor data and resource constraints in the WSNs, ensemble learning and ensemble pruning were adopted. Our proposed online anomaly detection method includes three parts, ensemble detector training, online anomaly detecting and detector updating, which is depicted in Fig. 1.

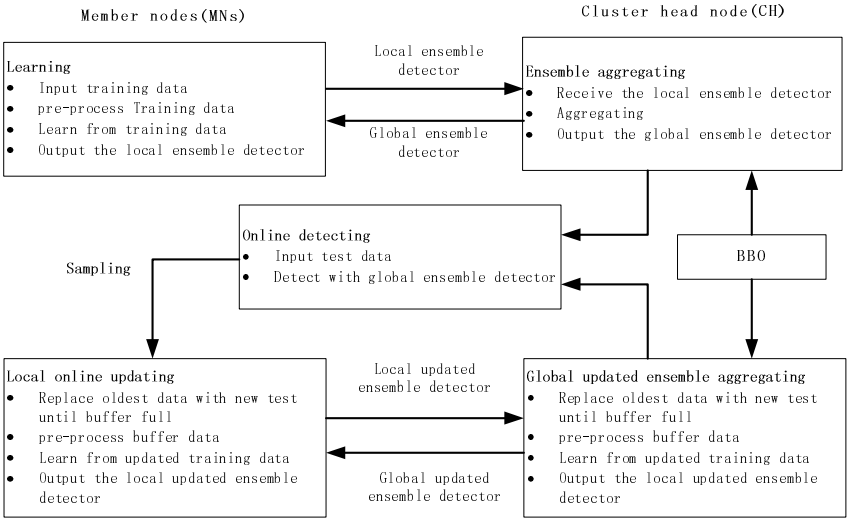

Fig. 1. Ensemble Anomaly Detection Method based on BBO pruning in WSNs

From the Fig. 1, we can see that our proposed method enables each sensor node to globally detect its every new observation normal or anomalous online. Here, distributed detecting is employed to achieve load (communication, computation and storage) evenly in the network and to prolong the lifetime of the whole network. This method can scale well with increase of number of nodes in WSNs due to its distributed processing nature. It has low communication requirements and does not need to transmit any actual observations between cluster head node and its member sensor node which save the communication resource significantly. Next, we described three important procedures mentioned above in detail.

(1) Building the initial Ensemble Detector

Considering the spatio-temporal correlation existed in sensed data in a given cluster, an initial ensemble is constructed by two steps. Firstly, a number of base detectors are firstly trained sequentially for each sensor nodes in a cluster (including the cluster head node itself); In order to build the ensemble detector, the history data are divided into the multiple chunks with the same size and each chunk is used to train a single detector. Because the data distribution maybe changed over time, the previous trained detector maybe useless for the future detection, what's more, the limited memory resource in the sensor node is another constraint to storage too many previous detectors. In practice, according to the space of memory resource, only the latest multiple detectors are kept to build the initial ensemble for one sensor node.

Once the ensemble detector is built, various techniques can be employed to combine the results of each detector. The common used in literature is the majority vote (for classification problem) and weighted average (for regression problem). In our paper, the final ensemble detection result can be calculated by formula (2), w_i denotes weight coefficient, m denotes the number of node and n the number of individual detector of each node). In our paper, for simplicity, the simple average strategy ($w_i = 1$) is employed to combine the finally result.

$$f_{fin}(x) = \frac{1}{n*m}\sum_{i=1}^{n*m} f_i(x) * w_i \tag{2}$$

(2) Ensemble Pruning based on BBO search

Constrained by the limited communication and storage resource in sensor node, the ensemble pruning is necessary.

Given an initial ensemble anomaly detector, $E = \{AD_1, AD_2, ...AD_{n*m}\}$, AD_i is a trained anomaly detector; a combination method C, and a test dataset T. The goal of ensemble pruning is to find an optimal or sub-optimal subset $E' \subseteq E$ which can minimizes the generalization error. Let $f_{i,j}$ $(i=1,2,..m,j=1,2..n)$ be the fitness values measured by the general performance of the detectors, such as true positive rate, false positive rate, accuracy and diversity among different detector so on. Obviously, fitness value F, can be defined as formula (3) based on the results of testing data.

The final fitness function can be defined as:

$$Maximize(\sum_{i=1}^{n*m} f_i * w_i) \text{ , s.t. } \sum_{i=1}^{n*m} w_i = 1 . \tag{3}$$

BBO [7] is a population-based, global optimization method, which has some common characteristics similar to existing evolutionary algorithms (EAs), such as genetic algorithm (GA), particle swarm optimization (PSO), ant colony optimization (ACO) and so on. The details between BBO and these EAs can be seen in [7].

The pseudo-codes of ensemble pruning based on BBO can be described as follows. Here H indicates habit, HIS is fitness, SIV (suitability index variable) is a solution feature.

```
Algorithm: Ensemble_Pruning_BBO(E, T)
Input: E – initial ensemble anomaly detector, T – The number
of maximization iteration
Output: E' – final ensemble anomaly detector
1: BBO parameter initialization
    Create a random set of habitats (populations) {H1, H2,…,
  HN};
    Compute corresponding fitness, i.e., HSI values;
2: Optimization search process
  While ( ! T)
      Compute immigration rate λ and emigration rate u for
    each habitat based on HSI;
    /* Migration */
    Select Hi with probability based on λi;
    If Hi is selected
       Select Hj with probability based on uj;
       If Hj is selected
        Randomly select a SIV form Hj;
        Replace a random SIV in Hi with one from Hj;
       End if
    End if
    /* Mutation */
     Select an SIV in Hi with probability based on the
    mutation rate;
    If Hi(SIV) is selected
       Replace Hi(SIV) with a randomly generated SIV;
    End if
    Re-compute HSI values;
  End while
3: Ensemble pruning
  Get the final ensemble of anomaly detector E* based on the
habitats Hi* with acceptable HSI.
```

(3) Online Update and Relearning

Distribution evolving is occurred possibly and detector updating is necessary. Online detector updating will be accompanied by a relearning procedure. In order to save the computation, communication and memory resources, a comprised strategy, i.e., sampling and delay strategy, was adopted [8]. Simply, a probability p can be specified, which samples the subsequent new observation as a training data for new detector. Here some heuristic rule should be employed to guide its value, for example, if the dynamics is relatively stationary, the small p should be used; otherwise, the big p should be chosen. When the buffer of a sensor node was replaced by the new data completely, online update is triggered and new detector is trained. The pseudo-codes of algorithm can be described as follows.

Algorithm: Online_Updating (E', p)
Input: *E'* – Current pruned ensemble anomaly detector, *p* – Sampling probability
Output: *E** – Updated pruned ensemble anomaly detector
For each sensor node
Retain the new observation with probability *p*;
If buffer is replaced completely
Train new detector and transmit its summary to cluster head;
E=Ensemble_Pruning_BBO(E', T)*
Broadcast E to its member sensor node for subsequent anomaly detection.*

3 Experimental and Analysis

3.1 Dataset and Data Pre-processing

IBRL datasets [1] was used in our paper, which was collected using a WSN deployed in Intel Research laboratory at University of Berkeley and commonly used to evaluate the performance of some existing models [2][5][6][8]. This network consists of 54 Mica2Dot sensor nodes which were deployed in a period of 30 days from 29/02/2004 until 05/04/2004. Fig. 2 shows the location of each node in the deployment [5]. Four types of measures data, that is light, temperature and humidity as well as voltage, were collected by this network. Those measurements were recorded in 31s interval. Because the measurement variables have little changes over the time, this dataset is considered a type of static datasets for many researchers. In our experiments, to evaluate the anomaly detection algorithm, some artificial anomalies are created by randomly modify some observations [6].

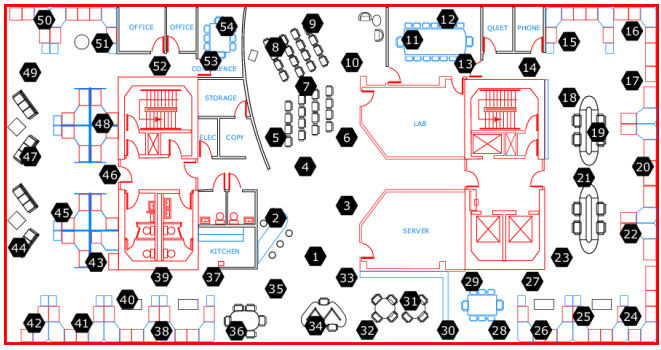

Fig. 2. Sensor nodes location in the IBRL deployment

In our experiment, a cluster (consisted of 4 sensor nodes, i.e., N7, N8, N9 and N10) and dataset (collected from these four nodes on 29th/02/2004) is chosen. The data

distribution can be seen in Fig. 3. From Fig. 3 an obvious fact is that data distribution in a cluster is almost same which are well proved that spatial correlation exists. There are some trivial differences largely due to packet loss Following accepted practices, we replaced missing data points with the average values of the data points within a certain amount of time [2].

Because the IBRL dataset has no label attributes and regarded as all observation are normal, to evaluate the performance of our proposed method, some anomaly data points should be inserted. In our paper, a number of 30 data points of artificial anomalies for each sensor were injected consecutively in each dataset to calculate the true positive rate (TPR) and false negative rate (FPR) as well as detection accuracy (ACC). The anomalies were generated using a normal randomizer with slightly deviate statistical characteristics from the normal data characteristics [6].

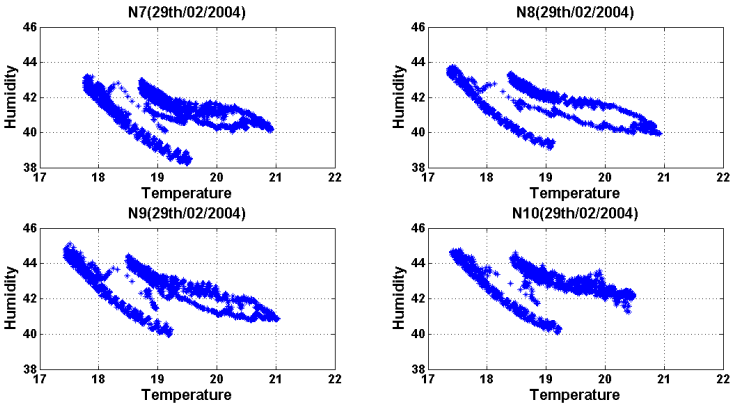

Fig. 3. The data distribution of Node 7, Node 8, Node 9 and Node 10 on Feb. 29. 2004

3.2 Performance Evaluation Metrics and BBO Parameters

In order to evaluate our proposed method, some commonly used performance evaluation metrics for anomaly detection are used in our paper, such as detection accuracy (ACC), true positive rate (TPR) and false positive/alarm rate (FPR). It was described as follows:

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN) \tag{4}$$

$$TPR = TP/(TP+FN) \tag{5}$$

$$FPR = FP/(FP+TN) \tag{6}$$

where TP means number of samples correctly predicted as anomaly class, FP means number of samples incorrectly predicted as anomaly class, TN means number of samples correctly predicted as normal class and FN means number of samples incorrectly predicted as normal class; Further, the TPR and FPR can be calculated by following formulas:

BBO is employed to prune the initial ensemble. For BBO parameter setting, the migration model is same as that present in [7], other parameters are set as: Habitat (population) size S=30, the number of *SIVs*(suitability index variables) in each island n={20, 40, 60, 80}, maximum migration rates E=1 and I=1, and mutation rate m=0.01, λ, μ are the immigration rate and emigration rate, respectively. The elitism parameter ρ=2.

HIS (habitat suitability index) is a fitness function. Here, which is evaluated by *F-measure (F-score)*, which is a measure of a test's accuracy. It considers both the precision probability and recall probability of binary classification problem.

$$F-measure=\frac{\left(1+\beta^2\right)precision*recall}{\beta^2*\text{precision}+\text{recall}}=\frac{\left(1+\beta^2\right)*TP}{(1+\beta^2)*TP+\beta^2*FN+FP} \tag{7}$$

F-measure can be interpreted as a weighted average of the precision and recall, the big *F-measure* means that the precision and recall are both big. Consequently, a good detector is analogous to a habitat with a high *HSI* which is included in the final ensemble detector, and a poor detector is analogous to a habitat with a low *HSI*, which may be discard from the finally ensemble detector. In our paper, the $\beta = 1$.

3.3 Results Presentation and Discussions

One class SVM, due to its attracting advantages, is especially favored by anomaly detection methods [3]. In our paper, it is used to train the base learner of anomaly detection. The dataset for each sensor node was divided into two parts: about 66% are used for training the local detector, the remainder is test set for evaluating our proposed method. Online Bagging [4] is used to build our initial ensemble detector. In our paper, three experiments are done, i.e., local ensemble anomaly detection only considering the temporal correlation of each sensor node, global ensemble anomaly detection considering the spatio-temporal correlation and the global pruned ensemble anomaly detection based on BBO. For the page limitation, only later two experimental results are presented in **Table 1** and **Table 2**, respectively.

Table 1 shows the global detection performance of each sensor node, Here, each member node sent its local ensemble detector to cluster head, combing with the cluster head itself built local ensemble detector, a global ensemble detector is built. Then cluster head broadcast this global ensemble detector, each member node use this global detector to online test the local observation.

Table 1. Global Detection Performance Based on Global-Ensemble Detector

Combined Ensemble Size	N7			N8			N9			N10		
	ACC	TPR	FPR	ACC	TPR	FPR	ACC	TPR	FPR	ACC	TPR	FPR
20	0.9467	0.8333	0.0486	0.9300	0.7778	0.0603	0.9467	0.7500	0.0423	0.9500	0.7857	0.0420
40	0.9700	0.7500	0.0208	0.9433	0.8333	0.0496	0.9710	0.8938	0.0246	0.9650	0.8929	0.0315
60	0.9700	0.8333	0.0243	0.9733	0.8889	0.0213	0.9800	0.9375	0.0176	0.9783	0.9357	0.0196
80	0.9817	0.9583	0.0174	0.9800	0.9444	0.0177	0.9767	0.9375	0.0211	0.9780	0.9714	0.0217

However, sending the local ensemble detector and global ensemble detector between member node and cluster head will arouse massive communication requirement. Consequently, despite the global ensemble detector has a good detection performance, it is impracticable in the real application due to communication cost. Here, ensemble pruning is activated based on BBO for global ensemble detector in cluster head, **Table 2** show the result of detection performance of pruned global ensemble detector. An obvious facts is that the size of global ensemble decreases sharply (at least 60%) and thus save communication resources to some extent, while the detector performance keep the same compared the initial global ensemble detector.

Table 2. Global Detection Performance Based on Global-Ensemble Detector of BBO Pruned

Ensemble	N7			N8			N9			N10		
Size(BBO pruned)	ACC	TPR	FPR	ACC	TPR	FPR	ACC	TPR	FPR	ACC	TPR	FPR
14	0.9480	0.8000	0.0458	0.9327	0.7667	0.0567	0.9500	0.8125	0.0423	0.9533	0.8571	0.0420
23	0.9710	0.7750	0.0208	0.9447	0.8000	0.0461	0.9733	0.9250	0.0239	0.9697	0.9143	0.0276
27	0.9713	0.8500	0.0236	0.9683	0.8333	0.0230	0.9810	0.9563	0.0176	0.9797	0.9357	0.0182
32	0.9820	0.9750	0.0177	0.9750	0.8333	0.0160	0.9820	0.9500	0.0162	0.9830	0.9786	0.0168

4 Conclusion and Future Work

Since the ensemble detector is verified to have better performance than single detector, in this paper, exploiting the spatio-temporal correlation existed in the sensed data collected from WSNs, an online ensemble anomaly detector method is proposed. Due to the computation and communication resource constrained in the WSNs, ensemble pruning is employed to identify the optimal subset of detectors based on BBO method. The experiment results on real dataset demonstrated our proposed method effective and efficient.

Acknowledgments. This work is supported by the National High Technology Research and Development Program of China (2011AA040103-7), the National Key Scientific Instrument and Equipment Development Project (2012YQ15008703), National Science Foundation of China (61104089), Science and Technology Commission of Shanghai Municipality (11jc1404000), Shanghai Rising-Star Program (13QA1401600).

References

1. Intel Berkely Reseach Lab (IBRL) dataset (2004), http://db.csail.mit.edu/labdata/labdata.html
2. Branch, J.W., Glannella, C., Szymanski, B., Wollf, R., Kargupta, H.: In-network Outlier Detection in Wireless Sensor Networks. Knowledge and Information Systems 34, 23–54 (2013)

3. Hejazi, M., Singh, Y.P.: One-class Support Vector Machines Approach to Anomay Detection. Applied Artificial Intelligence 27, 351–366 (2013)
4. Oza, N.C.: Online Bagging and Boosting, Systems, man and cybernetics. In: 2005 IEEE International Conference on, pp. 2340–2345 (2005)
5. Rassam, M.A., Zainal, A., Maarof, M.: An Adaptive and Efficient Dimension Reduction Model for Multivariate Wireless Sensor Networks Applications. Applied Soft Computing 13, 1978–1996 (2013)
6. Rassam, M.A., Zainal, A., Maarof, M.: One-Class Principal Component Classifier for Anomaly Detection in Wireless Sensor Network. In: 2012 Fourth International Conference on Computational Aspects of Social Networks, New York, pp. 271–276 (2012)
7. Simon, D.: Biogeography-based Optimization. IEEE Transactions on Evolutionary Computation 12, 702–713 (2008)
8. Xie, M., Hu, J., Han, S., Chen, H.: Scalable Hyper-Grid KNN-based Online Anomaly Detection in Wireless Sensor Networks. IEEE Transactions on Parallel and Distribution Systems 24, 1661–1670 (2012)
9. Zhang, Y.: Observing the unobservable: distributed online outlier detection in wireless sensor networks, p. 174. Universtiy of Twente, The Netherlands (2010)
10. Zhang, Y., Meratnia, N., Havinga, P.: Outlier Detection Techniques for Wireless Sensor Networks: A survey. IEEE Communications Surveys & Tutorials 12, 159–170 (2010)
11. Zhou, Z.H., Tang, W.: Selective Ensemble of Decision Trees, Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, pp. 476–483. Springer (2003)