Available online at www.ispacs.com/cacsa
Volume 2017, Issue 1, Year 2017 Article ID cacsa-00074, 7 Pages
doi:10.5899/2017/cacsa-00074
Research Article

Anomaly detection in industrial control systems using evolutionary-based optimization of neural networks

Amin Mansouri[1], Babak Majidi[1*], Abdollah Shamisa[2]

(1) Department of Computer Engineering, Faculty of Engineering, Khatam University, Tehran, Iran.
(2) Department of Electrical Engineering, Faculty of Engineering, Khatam University, Tehran, Iran.

Abstract

Industrial control systems are increasingly used for control and monitoring of important infrastructure. Machine learning algorithms have the ability to discover patterns in large amounts of data and to create diagnosis models based on these patterns. Since modelling a large amount of unlabeled data is costly and time-consuming, the automated machine learning methods have the ability to detect anomalies in industrial control systems effectively. In this paper, first, twenty-four machine learning algorithms are evaluated for anomaly detection in gas distribution control network. Then dimensionality reduction algorithms are used to improve the accuracy of anomaly detection. Finally, by using an evolutionary based optimization for training a neural network, a new algorithm for prediction of anomalies in the SCADA system with high accuracy is proposed. The experimental results show that the proposed algorithm has the ability to detect the anomalies in the gas distribution control network with 97.5% accuracy.

Keywords: Anomaly detection, Industrial automation control systems, Evolutionary algorithms, Neural networks.

International Scientific Publications and Consulting Services

Production and hosting by ISPACS GmbH.

* Corresponding Author. Email address: b.majidi@khatam.ac.ir, Tel: +982189174189

1 Introduction

Anomaly detection systems based on machine learning techniques are used to protect SCADA systems against downtime. Chalamasetty et al. [1] proposed a scalable model for SCADA communications in electricity distribution network that has the ability of effective detection and prevention of anomalies. They also have suggested an anomaly detection technology for avoiding network downtime. The results showed that the proposed method is very effective against various anomalies in the network. Zhang et al. in [2] have compared the detection of various anomalies in SCADA system. They have used modified semi-Markov process model, in SCADA network, to describe the interaction between anomalies in the electrical system. The results shows that with a high level of anomaly, the probability of success diminishes. Almalawi et al. [3] a data-driven clustering approach has been suggested which eliminates the need for experts in the field of SCADA and merely ordinary data for creating of detection models. This method is based on assumption that "normal mode" is a combination of modes and values of multivariate process parameters in a SCADA system which can be clustered in unlimited dense clusters groups and critical states obtained in n-dimensional space as a form of outlier data. Kwon et al. [4] have collected all traffic from digital substations in South Korea for a week based on IEC 61850 protocol. They have suggested a behavior-based anomaly detection method, using multiple media properties and with the aim of obtaining a high detection rate. Singh et al. [5] have proposed a test bed that can simulate SCADA systems. The proposed test bed includes a comparison module that helps to identify potential attacks in the system. Almalawi et al. [6] have suggested an unsupervised learning algorithm that will optimize the SCADA network traffic. They have used numerous pre-processing methods to control the integrity of the outlier data, in the data set of water plants. Kirsch et al. [7] reported their experience in design and implementation of their first survivable SCADA system. Survivability means that the SCADA system, continue to work correctly with minimal degradation performance, even if malicious anomalies have jeopardized a part of the system. To achieve survivability, the proposed system uses the anomaly tolerance method presented in [8] and [9]. In [10] Maglaras and his colleagues have offered an anomaly detection module which is able to detect malicious traffic of the network in SCADA system. They used an unsupervised clustering in their anomaly detection module. Maglaras et al. [11] also have suggested an anomaly detection module based on the combination of their previous method with RBF kernel and K-means recursive clustering. The combination of K-means recursive clustering with previous method will make the anomaly detection module capable of recognizing real warnings of possible anomalies regardless of the parameters of the system. Pan et al. [12] have prposed a framework for anomaly detection for industrial networks.

In this paper, 24 data mining algorithms are compared for anomaliy detection on the gas control network. The dimentionality reduction algorithms, PCA, ICA, GHA, SVD and SOM were used to improve the accuracy. Following that a series of evolutionary based algorithms including BBO, PBIL and GrayWolf are used for training of a Neural Networ to predict the anomaies in the gas control network. The experimental results shows that the proposed algorithm is capable of prediction of industrial network anomalies with 97.5% accuracy.

The rest of the paper is organized as follows. After reviewing the literature related to industrial SCADA system the security structure is explored in the second section and in the third section biological algorithms for training of neural networks are analyzed. In section 4 the proposed algorithm for optimizid anomaly detection in gas distribution industrial networks is presented. Finally the performance of the proposed algorithm is presented and analyzed.

2 Anomaly detection in industrial control systems

There is a great interest in the scientific community for anomaly detection systems for the industrial control network traffic. Supervisory Control and Data Acquisition (SCADA) systems are increasingly used in varous inportant industrial system and their reliability is very important to avoid unnecessary downtime and finacial loss [3]. SCADA, also are increasingly used in control and monitoring of important infrastructure. These systems, includes gas pipelines, power plants, railways, water and wastewater refinement and even some air conditioning systems. In the past, many of these systems were separated from other networks but in recent years these systems have been merged with the Internet and internal networks of companies. Connecting these systems to other networks has brought more control for operators and also reduced the costs for companies. This connection, also has caused a lot of security concerns for these systems that used to work isolatedly not so long ago. One of these concerns is remote control of the systems. If a vulnerability exists in one of these systems, this vulnerability allows a user to remotely take control of the SCADA system whcih can lead to system failure.

SCADA systems, provide control and visualization of critical infrastructure systems. These systems generally consist of four sections [13]. The first level includes sensors [13]. These sensors collect data about the system and include monitoring and measuring sensors. Actuators include pumps, motors and so on. The second level is Programmable Logic Controllers (PLC). These components control and collect the information that identifies the system state. Controllers are generally named Remote Terminal Units (RTU). RTU associated with the first level of SCADA system, for example, saves the sensor data in predetermined registers. The third level of a SCADA system, is the regulatory controls [13]. Regulatory controls are usually addressed by the Main Terminal Unit (MTU). MTU is a unit which works with the RTU. For example in a gas station MTU can send command to RTU to turn on the pump. MTU will also send a query to read RTU register which contains the current pressure measurement. There are many communication protocols which allow such communication. For example, Profibus, Fieldbus, Modbus and DNP3. Fourth level is Human-Machine Interface (HMI) which is used by operator to display the data collected from sensors by MTU. HMI typically includes a graphical display system and sub-system. HMI is also used for modification of parameters and states in SCADA system through communication with MTU.

3 Evolutionary based Neural Networks for anomaly detection

Grey Wolf Optimizer (GWO) algorithm is a meta-heuristic algorithm inspired by nature which is based on a hierarchical structure and social behavior of wolves in the hunt. The algorithm have been introduced by Mirjalili et al. [14]. GWO is a population-based algorithm which can be generalized to large-scale systems. In implementation of this algorithm, four types of gray wolves, i.e., alpha, beta, delta, and omega are used to simulate the leadership hierarchy in which the three main steps of hunting, i.e., searching for prey, surrounding of prey and attacking to prey are iplemented. The average number of wolves herd is between 5 and 12 wolves. In each herd there are four main rating:

1. Alpha: Leader of the group. These wolves are the dominant wolf in the herd and manage tasks such as place of rest or the way of hunting.
2. Beta: these wolves are helping the Alpha wolves in decision making and are also able to be elected as their successors.
3. Delta: Delta wolves are often in lower position than Beta wolves and are mostly the old wolves, hunters and wolves that are taking care of babies.
4. Omega: they are in the lowest rank in the hierarchy and have the least right compared to the other members of the group. These wolves eat after all and are not involved in the decision-making process.

Gray wolves hunting process includes three main phases:

1. Finding, tracking and pursuing of the prey.
2. Approaching and surrounding and misleading of the prey untill it stops moving.
3. Attacking the prey.

The behaviur of the wolf pack in the hunting is modelled in five phases:

- Modeling of the hierarchical structure:
  Optimization is done by alpha, beta and delta wolves. The alpha is assumed as the main director of the algorithm and one beta and delta wolf take part in main part of the hunt and the rest are their followers.
- Surrounding the prey process modeling
- Estimation process modeling

Gray wolves have the ability to estimate the position of the prey. For modeling of this process we have the following steps: We do not have any idea about the position of hunting in the initial search. It is assumed that wolves alpha, beta, delta have a better basic knowledge about initial position of prey. The answer to these three candidate position is calculated as follows: the wolf alpha, beta, delta estimate the position of the prey. The rest of the wolves position randomly around hunting wolves and the positions are updated. We always keep our top 3 positions as answer.

- Attack process modeling

When the prey is surrounded by wolves and stops, the invasion with the leadership of alpha wolf begins. The gray wolves algorithm requires that all wolves update their positions according to the position of alpha, beta and delta wolves.

- Searching for the prey

Searching process is exactly the reverse of the attack. When searching for the prey, wolves are away from each in order to track it, while after tracking the prey, wolves are close to each other in the attack phase.

In general, the algorithm can be summarized as follows:

- The efficiency of all answers are calculated and the three top answers as alpha, beta and delta are chosen until the end of the algorithm.
- In each iteration three top answers, (wolves alpha, beta and delta) have the ability to estimate the position of the prey.
- In each iteration after determining the position of wolves alpha, beta and delta, the rest of answers will be updated.
- At the end of iterations position of alpha wolf is introduced as the optimal point.

The most important thing in training of MLP using metaheuristic algorithms is that the problem should be formulated in a way that is appropriate for use of metaheuristic algorithms. Weights and biases in training of MLP are important variables. A trainer should find a set of weight and bias values with the highest accuracy. So our variables in this problem are the weights and biases. After declaring variables, we need to define our objective functions for gray wolves algorithm. As mentioned above, the aim in training of a MLP, is achieving the highest classification accuracy rate, estimation or prediction for the experimental samples. A general criteria for the assessment of MLP, is Mean Square Error (MSE). In this benchmark, a set of test samples have been calculated in MLP in accordance with the following equation and the difference between the output and the amount of MLP is achieved by the following equation:

$$MSE = \sum_{i=1}^{m} \left(o_k^i - d_k^i \right)^2 \tag{3.1}$$

where m is the number of outputs, d_k^i is the intended output of the i^{th} input unit when the k^{th} prototype is used and o_k^i is the actual output for the i^{th} input. The performance of an MLP is evaluated based on the average MSE on all samples tested in accordance with the following formula:

$$\overline{MSE} = \sum_{k=1}^{s} \frac{\sum_{i=1}^{m} \left(o_i^k - d_i^k \right)^2}{s} \tag{3.2}$$

The s is the number of tested samples, m is number of outputs, $d^k{}_i$ is output of the i^{th} input unit when the k^{th} prototype is used, and $o^k{}_i$ is the actual output of i^{th} input unit when the k^{th} prototype used as input. Finally, the problem of training of a MLP with the variables and the average MSE for gray wolves algorithm is calculated. Figure 1 shows the overall process of training of MLP with GWO. Gray wolves algorithm continuously changes the weight and bias in order to minimize the average MSE of the tested samples.

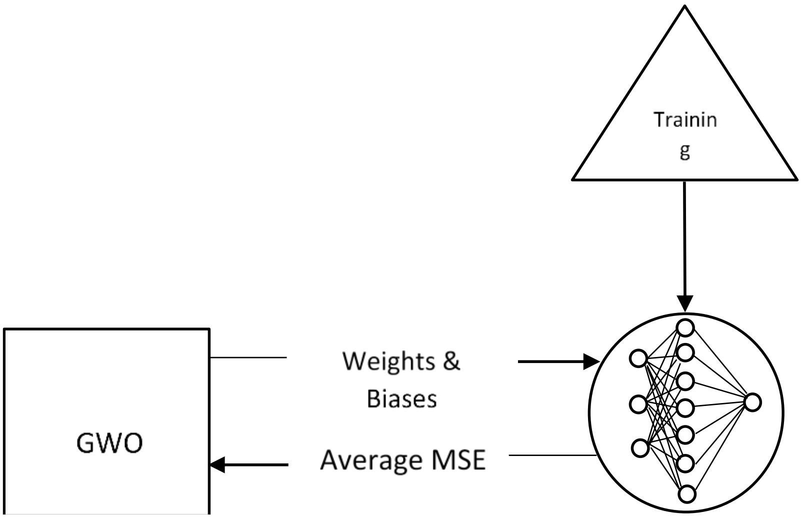

Figure 1: Process of training of MLP with GWO.

4 Experimental results

The data from transmission network between remote terminal units and a central control unit in SCADA system of gas pipeline in Mississippi State University is used for experimental results [16]. The dataset is collected with a new framework for simulating of real anomalies in a gas pipeline. The dataset contains three separate categories of features: network information, payload information and labels. For simulation 24 algorithms have been run on the dataset, and then by checking the accuracy of the algorithms, neural network algorithm has been selected for optimization. These algorithms have been run by Rapidminer 6.5.002 software on the data set. First the algorithms were performed without normalization and then the results were obtained once again with normalization. All algorithms were evaluated in the form of 10 folds cross validation. The simulations are implemented on a DELL laptop with CPU Intel Core i5 and 2.40 GHz frequency. Then the dimentionality reduction algorithms, PCA, ICA, GHA, SVD and SOM were conducted on AutoMLP algorithm to improve the accuracy. Then three optimization algorithms are used for tranining of a neural network. These algorithms include: BBO, PBIL and Gray Wolf. Basic parameters that were used for the implementation of these algorithms are as demonstrated in Table 1.

Table 1: Evolutionaary algorithms parameters.

Population size	**Generation limitation**	**Number of genomes**	**Mutation chance**
50	249	1484	**0**

The gas data set included 97019 members that 96019 members of it were used to training of the neural network and also the rest of the data was used for testing. The the evolutionary based optimization manages to improve the anomaly detection results for the gas dataset as demonstrated in Table 2:

Table 2: Evolutionary based optimization results for the gas dataset.

Optimization algorithms with ANN	Accuracy (%)
BBO	90.90
PSO	90.75
ACO	91.74
ES	90.75
PBIL	90.75
Gray Wolf	97.98

5 Conclusion

In this paper, twenty-four machine learning algorithms are evaluated for anomaly detection in gas distribution control network. Then dimensionality reduction algorithms, (e.g. PCA, ICA, GHA, SVD and SOM algorithm) combined with neural networks are used to improve the accuracy of anomaly detection. Finally, by using an evolutionary based optimization for training of a Neural Networks, a new algorithm for high accuracy prediction of anomalies in the SCADA network system is proposed. The experimental results show that the proposed algorithm has the ability to detect the anomalies in the gas distribution network with 97.5% accuracy.

References

[1] G. K. Chalamasetty, P. Mandal, T. Tzu-Liang, Secure SCADA communication network for detecting and preventing cyber-attacks on power systems, Clemson University Power Systems Conference (PSC), (2016) 1-7.
https://doi.org/10.1109/PSC.2016.7462865

[2] Y. Zhang, L. Wang, Y .Xiang, Power System Reliability Analysis With Intrusion Tolerance in SCADA Systems, IEEE Transactions on Smart Grid, 7 (2016) 669-683.
https://doi.org/10.1109/TSG.2015.2439693

[3] A. Almalawi, A. Fahad, Z. Tari, A. Alamri, R. AlGhamdi, A. Y. Zomaya, An Efficient Data-Driven Clustering Technique to Detect Attacks in SCADA Systems, IEEE Transactions on Information Forensics and Security, 11 (2016) 893-906.
https://doi.org/10.1109/TIFS.2015.2512522

[4] Y. Kwon, H. K. Kim, Y. H. Lim, J. I. Lim, A behavior-based intrusion detection technique for smart grid infrastructure, in PowerTech, IEEE Eindhoven, (2015) 1-6.
https://doi.org/10.1109/PTC.2015.7232339

[5] P. Singh, S. Garg, V. Kumar, Z. Saquib, A testbed for SCADA cyber security and intrusion detection, in Cyber Security of Smart Cities, Industrial Control System and Communications (SSIC), International Conference on, (2015) 1-6.
https://doi.org/10.1109/SSIC.2015.7245683

[6] A. Almalawi, X. Yu, Z. Tari, A. Fahad, I. Khalil, An unsupervised anomaly-based detection approach for integrity attacks on SCADA systems, Computers & Security, 46 (2014) 94-110.
https://doi.org/10.1016/j.cose.2014.07.005

[7] J. Kirsch, S. Goose, Y. Amir, D. Wei, P. Skare, Survivable SCADA Via Intrusion-Tolerant Replication, IEEE Transactions on Smart Grid, 5 (2014) 60-70.
https://doi.org/10.1109/TSG.2013.2269541

[8] J. Kirsch, Intrusion-tolerant replication under attack, Ph.D. dissertation, Johns Hopkins University, Baltimore, MD, USA, (2010).

[9] Y. Amir, B. Coan, J. Kirsch, J. Lane, Prime: Byzantine Replication under Attack, IEEE Transactions on Dependable and Secure Computing, 8 (2011) 564-577.
https://doi.org/10.1109/TDSC.2010.70

[10] L. A. Maglaras, J. Jiang, Intrusion detection in SCADA systems using machine learning techniques, in Science and Information Conference (SAI), 2014, (2014) 626-631.
https://doi.org/10.1109/SAI.2014.6918252

[11] L. A. Maglaras, J. Jiang, OCSVM model combined with K-means recursive clustering for intrusion detection in SCADA systems, in Heterogeneous Networking for Quality, Reliability, Security and Robustness (QShine), 2014 10th International Conference on, (2014) 133-134.
https://doi.org/10.1109/QSHINE.2014.6928673

[12] Z. Pan, S. Hariri, Y. Al-Nashif, Anomaly based intrusion detection for Building Automation and Control networks, in 2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA), (2014) 72-77.
https://doi.org/10.1109/AICCSA.2014.7073181

[13] P. S. M. Pires, L. A. H. g. Oliveira, Security Aspects of SCADA and Corporate Network Interconnection :An Overview, in 2006 International Conference on Dependability of Computer Systems, (2006) 127-134.
https://doi.org/10.1109/DEPCOS-RELCOMEX.2006.46

[14] S. Mirjalili, S. M. Mirjalili, A. Lewis, Grey Wolf Optimizer, Adv. Eng. Softw, 69 (2014) 46-61.
https://doi.org/10.1016/j.advengsoft.2013.12.007

[15] S. Mirjalili, How effective is the Grey Wolf optimizer in training multi-layer perceptrons, Applied Intelligence, 43 (2015) 150-161.
https://doi.org/10.1007/s10489-014-0645-7

[16] T. Morris, Z. Thornton, I. Turnipseed, Industrial Control System Simulation and Data Logging for Intrusion Detection System Research, 7th Annual Southeastern Cyber Security Summit. Huntsvile, AL. June, 2015 (3 - 4) (2015).

[17] G. Wei, T. Morris, B. Reaves, D. Richey, On SCADA control system command and response injection and intrusion detection, in eCrime Researchers Summit (eCrime), 2010 (2010) 1-9.
https://doi.org/10.1109/ecrime.2010.5706699