Comparative Study of Anomaly Detection in Wireless Sensor Networks Using Different Kernel Functions

Shashank Gavel, Ajay Singh Raghuvanshi and Sudarshan Tiwari

Abstract Wireless sensor network (WSN) is defined as an autonomous network composed of low power sensor nodes having limited computational, communication, and energy resources. Being short at resources they require efficient use of each resource to prolong network lifetime. Sensor networks are exposed to noise, compromised nodes, low battery levels, and damaged sensors, all these results in anomalous readings or anomaly. Presence of anomaly in system deteriorates the performance of WSN in terms of efficiency, accuracy, and reliability. Hence anomaly detection becomes a major challenge to decide the performance of network. Support vector machine (SVM) is a light weight, learning-based binary classifier that can classify the raw data into normal and anomalous. SVM suffers from computational complexity while handling large datasets, so sequential minimal optimization SVM (SMO-SVM) is used to reduce the complexity. In this paper, a comparative study is made on anomaly detection using SMO-SVM classifier utilizing different kernel functions.

Keywords Wireless sensor network · Anomaly detection · Support vector machine · Sequential minimal optimization · Kernel functions

S. Gavel (✉) · A. S. Raghuvanshi
Department of Electronics and Telecommunication, National Institute of Technology, Raipur, India
e-mail: sgavel.phd2016.etc@nitrr.ac.in

A. S. Raghuvanshi
e-mail: asraghuvanshi.etc@nitrr.ac.in

S. Tiwari
Department of Electronics and Communication, Motilal Nehru National Institute of Technology, Allahabad, India
e-mail: stiwari@mnnit.ac.in

D. Dutta et al. (eds.), *Advances in VLSI, Communication, and Signal Processing*, Lecture Notes in Electrical Engineering 587,
https://doi.org/10.1007/978-981-32-9775-3_8

1 Introduction

Wireless sensor networks (WSNs) is a network made up of small active devices called sensor nodes which are used for monitoring or event detection purpose. Sensor nodes are autonomous low powered devices with sensors that provide different types of sensed data such as humidity, temperature, pressure, and vibrations. Nowadays WSNs have found intensive use in smart cities, smart grid, battlefields, medical sensing, etc. [13]. WSNs are resource constraint networks and the nodes are susceptible to noise, compromised nodes due to intrusion, low battery levels, and damaged sensors giving rise to anomaly in sensed data. These anomalies contribute adversely on the performance and lifetime of network. Hence detection of anomaly in WSN becomes major concern for the efficient use of resources [5].

Anomaly detection techniques that are used for WSN data can be categorized as machine learning, statistical, and signal processing-based approaches. Machine learning approach for detection and classification of anomalies in WSN is gaining a lot of interest by research community [2, 14].

Support vector machine is one of the efficient binary classification technique. The use of SVM for detection and classification has been done in [12]. Although SVM classifies data efficiently sometimes it possesses high computational complexity for the larger datasets. An optimized technique for SVM known as sequential minimal optimization SVM (SMO-SVM) is used in place of simple SVM for improved classification [9].

This paper presents a comparative study of SMO-SVMs classifier by utilizing the different kernel functions. The efficient use of kernel functions is governed by the nature of dataset. The dataset of two standard laboratory has been used to analyze the performance of different kernel functions for SMO-SVMs. The different types of anomaly that exist practically in WSN are manually inserted in each dataset. The performance of anomaly detection is compared using standard performance metrics like accuracy and F1-Measure.

The paper is organized in four sections: Sect. 2 presents literature review and theory of SMO-SVMs, results and discussion are presented in Sect. 3 followed by the conclusion in Sect. 4.

2 Literature Review and Theory of SMO-SVM

2.1 Literature Review

The concept of SVMs has been empirically applied in various fields such as machine learning, pattern recognition, and categorization of text. SVM as learning algorithm work efficiently for low dimension data but sometimes restricts the researcher to train the high dimensional data. So this becomes a problem while using this technique for high dimensional dataset problems. A new optimized algorithm for SVM was

proposed in [9] to overcome the difficulty of simple SVMs and to make the system perform better. This technique was SMO-SVMs which instead provides better results for complex quadratic programming problems. The main motivation behind the use of SMO-SVMs lies in its lesser computational complexity and hence lightweight. SMO-SVM can be used using different learning kernel functions, and to select appropriate kernel function is a major challenge for the classification of different types of anomaly present in WSN [1, 7].

The basic idea behind the SVMs is to fit the data into a hyperplane or hypersphere between different classes. The hyperplane is subspace in which data is fitted accordingly and the dimension is less than its ambient space. This hyperplane creates a large separation between the different classes of data. The method involves mapping of data into a higher dimensional space to make separation easier. After the data is mapped, kernel functions are applied for approximating the dot products between the mapped vectors into the feature space to find the hyperplane. This helps better in identifying the class of normal data with the anomalous data.

2.2 Sequential Minimal Optimization SVM

Here we consider a hypersphere in place of hyperplane, for SVM a set of training data is considered in a feature space, $X = (x_1, x_2, x_3, \ldots, x_n)$ where $x_i \in R^d$ $(1 \leq i \leq n)$ represents the d-dimensional data and n is the size of the training data. The data from the feature space is then trained. Considering this the optimization problem to be solved is given below:

$$\min R^2 + C \sum_{i=1}^{n} \varepsilon_i \tag{1}$$

$$s.t. \quad ||x_i - a||^2 \leq R^2 + \varepsilon_i, i = 1, 2, \ldots, n.$$
$$\varepsilon_i \geq 0, i = 1, 2, \ldots, n.$$

where a and R are the center and radius of the hypersphere, respectively, in the feature space, ε_i is the slack variable which allows few training data outside the hypersphere and the penalty parameter C controls the trade-off between the number of target data outside and volume of the hypersphere [10, 12]. The mapping function $\phi(.)$ is used to map the data of input class to feature space in terms of $\phi(x_i)$.

This mapping variable $\phi(x_i)$ replaces the value of x_i. This allows the function to calculate the inner product of two vectors in feature space. The inner product is given by

$$K(x_i, x_j) = \phi(x_i) * \phi(x_j) \tag{2}$$

Although simple SVM is computed in the form of linear classifier where x is input and y is output, the objective function representing the linear classifier becomes

$$f(x) = w^T x + b \tag{3}$$

where w is a normal vector and b represents the threshold value. Since we are using binary classification to judge the anomalous and non-anomalous data, the value prediction will be in the terms of $y = 1$ if $f(x) \geq 0$ and $y = 0$ if $f(x) < 0$. By considering the inner product as given in (2), the function in (3) is given by

$$f(x) = \sum_{i=1}^{n} \alpha_i y_i K(x_i, x_j) + b \tag{4}$$

where α_i represents the Lagrangian multiplier. Lagrange multiplier helps in finding the local maxima and minima of provided function [4, 6].

To optimize the above function Lagrangian multiplier plays an important role. This Lagrangian multiplier is to be optimized, and the constrained minimization problem is to be solved. The initial value of lagrangian multiplier which is to be updated is shown below:

$$\alpha_j = \alpha_j - \frac{y_j(E_i - E_j)}{\eta} \tag{5}$$

In the above Eq. (5) E_k represents the error on kth value of training example and η is the second order derivative of objective function. The individual representation of the variable are shown below:

$$E_k = f(x_k) - y_k \tag{6}$$

$$\eta = 2K(x_i, x_j) - K(x_i, x_j) - K(x_j, x_k) \tag{7}$$

Thus depending upon the value (5) the new value of final Lagrangian multiplier becomes α_i, below value shows the new optimized value of the multiplier:

$$\alpha_i = \alpha_i + y_i y_j (\alpha_{j-1} - \alpha_j) \tag{8}$$

So the optimization problem is solved by replacing the new value of α_i from (8) to (4).

In this paper we have basically used the kernel functions value as mentioned in (2) for different kernels which are shown below [3]:

1. Gaussian Kernel

$$K(x_i, x_j) = e^{\gamma ||x_i - x_j||^2}$$

2. Linear Kernel

$$K(x_i, x_j) = x_i^T x_j$$

3. Sigmoid Kernel

$$K(x_i, x_j) = (\gamma(x_i^T x_j + 1))^d$$

4. Polynomial (poly3) Kernel

$$K(x_i, x_j) = \tan(\gamma x_i^T x_j + 1)$$

These four kernel functions are most commonly used as they perform better for working set selections.

3 Results and Discussion

For the analysis of suitability of kernel function with anomalous data, we have used SMO-SVM as a classifier to our paper. Data conditioning of data set from Labeled Dataset collection [11] and IBRL Dataset [8] is done, and the different types of anomalies are inserted to analyze the compatibility of different kernel function. Following are the datasets taken for analysis:

1. Multihop and Singlehop Datasets [11]:
   This dataset is taken from the network containing both multihop and singlehop scenario for anomaly detection. Both the multihop and singlehop network is setup and depending upon that the readings from the sensor node are taken.
2. IBRL Datasets [8]:
   This dataset was taken from the well-known Intel Berkeley Lab experiment where 54 sensors were deployed. We have managed to take and conditioned the data from *node*8 and *node*9 as they are close to each other and shows similar behavior.

Following types of noise are inserted to the dataset:

1. Random Noise:
   This type of anomaly occurs where the sensor node supplies data and at the same time transient disturbance happens depending upon the random time distribution. This anomaly is used in multihop datasets from [11] example is shown in Fig. 1.
2. Regenerative Feedback Noise:
   This type of anomaly occurs in the network when the data from the sensor node

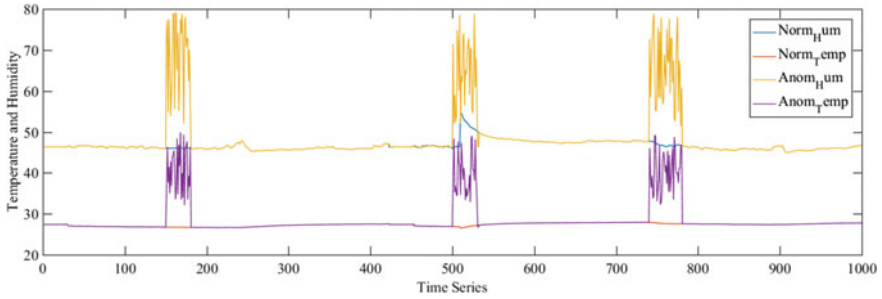


Fig. 1 Normal and anomalous dataset with respect to time series

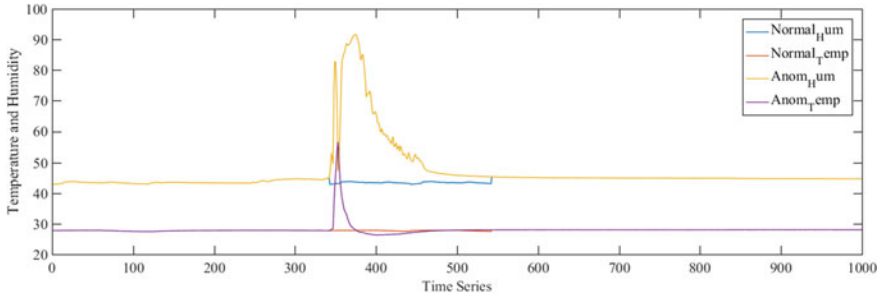


Fig. 2 Normal and anomalous dataset with respect to time series

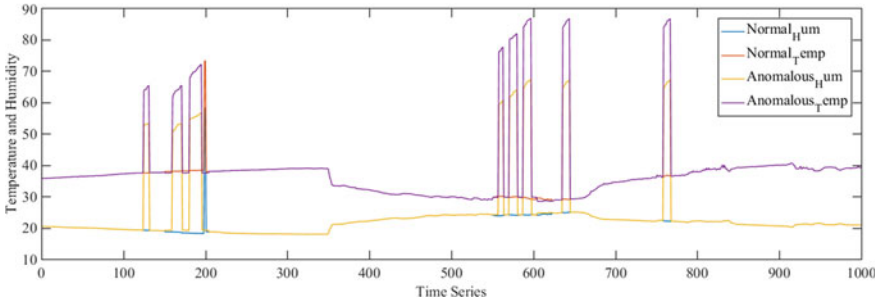


Fig. 3 Normal and anomalous dataset with respect to time series

keep on increasing as compared to the normal data. This anomaly is used in singlehop datasets from [11] example is shown in Fig. 2.

3. Shot Noise(Spikes):
   This anomaly occurs when the sensor data shows small spikes between the data. This anomaly is inserted to Node8 and Node9 of dataset from [8], example is shown in Fig. 3.

The comparison of the kernel functions are performed on the basis of performance metrics shown in (9) and (12). In which g is the g-means accuracy and $F1 - Measure$ shows the balance between the recall and precision value obtained from the dataset.

$$g = \sqrt{Acc_+ * Acc_-} \tag{9}$$

where

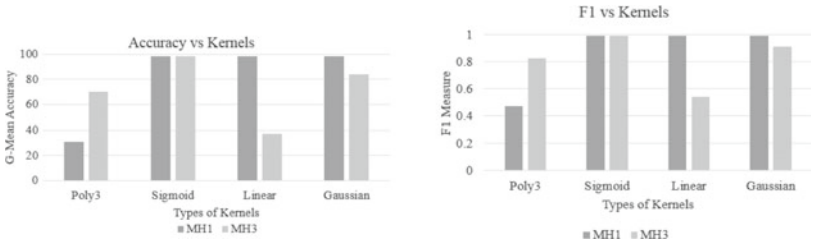
$$Acc_+ = \frac{No.\ \ of\ \ target\ \ samples\ \ correctly\ \ classified}{Total\ \ number\ \ of\ \ target\ \ samples} \tag{10}$$

$$Acc_- = \frac{No.\ \ of\ \ nontarget\ \ samples\ \ correctly\ \ classified}{Total\ \ number\ \ of\ \ nontarget\ \ samples} \tag{11}$$

and the F1-Measure is given by

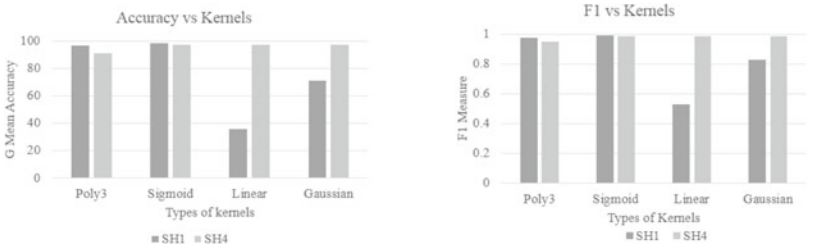
$$F1 - Measure = \frac{2 * Precison * Recall}{Precison + Recall} \tag{12}$$

The graph is plotted in terms of accuracy and F1-Measure and this result shows the compatibility of different kernel functions with SMO-SVM classifier in terms of different kinds of anomalies. The comparison of results in terms of performance metrics can be shown in the figures. Figure 4 (a, b) shows the accuracy and F1-Measure for SMO-SVM method of multihop dataset, Fig. 5 (a, b) shows the accuracy



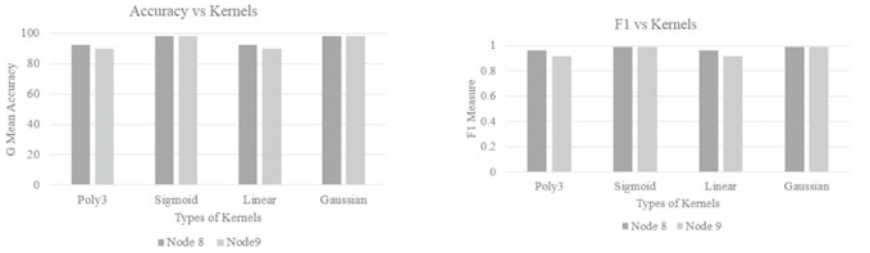
(a) Accuracy for Multihop dataset (b) F1-Measure for Multihop dataset

Fig. 4 Accuracy and F1-Measure for multihop dataset



(a) Accuracy for SingleHop dataset (b) F1-Measurefor SingleHop dataset

Fig. 5 Accuracy and F1-Measure for singlehop dataset



(a) Accuracy for IBRL dataset (b) F1-Measure for IBRL dataset

Fig. 6 Accuracy and F1-Measure for IBRL dataset

and F1-Measure for SMO-SVM method of singlehop dataset, and Fig. 6 (a, b) shows the accuracy and F1-Measure for SMO-SVM method of IBRL dataset. However the system performance for *Random Noise* the Sigmoid and Gaussian kernel show better compatibility with classifier, for *Regerative feedback Noise* the Sigmoid and poly3 kernel show better compatibility with classifier and for *Shot Noise* the all the four kernels show better compatibility with classifier. So it can be concluded that the accuracy of detection depends upon the selection of kernel function.

4 Conclusion

In this article we have used SMO-SVM as a detector and classifier for the WSN datasets. Three different types of noise are inserted in the datasets as explained in the above section, and to analyze this anomalous dataset various kernel functions are used with the SMO-SVM classifier. By analyzing the results obtained we may conclude that for network prone to random noise Sigmoid and Gaussian kernel function are better choice. For network prone to regenerative feedback noise sigmoid and poly3 kernel functions works better. For network prone to shot noise any of the kernel function among the four can be used. We have successfully compared the importance of kernel functions with the classifier and its compatibility for different anomalous conditions. In future we will be using this classifier to construct an Anomaly Detection System.

References

1. Branch, J.W., Giannella, C., Szymanski, B., Wolff, R., Kargupta, H.: In-network outlier detection in wireless sensor networks (2013)
2. Du, D., Qi, B., Fei, M., Wang, Z.: Quantized control of distributed event-triggered networked control systems with hybrid wired–wireless networks communication constraints (2017)

3. Fan, R.E., Chen, P.H., Lin, C.J.: Working set selection using second order information for training support vector machines (2005)
4. Kang, W.S., Choi, J.Y.: Domain density description for multiclass pattern classification with reduced computational load (2008)
5. Leccese, F., Cagnetti, M., Trinca, D.: A smart city application: a fully controlled street lighting isle based on raspberry-pi card, a zigbee sensor network and wimax (2014)
6. Lee, S.W., Park, J., Lee, S.W.: Low resolution face recognition based on support vector data description (2006)
7. Moshtaghi, M., Leckie, C., Karunasekera, S., Rajasegarar, S.: An adaptive elliptical anomaly detection model for wireless sensor networks (2014)
8. Peter, B., Wei, H., C.G.S.M.M.P., Thibaux, R.: Ibrl dataset. http://db.csail.mit.edu/labdata/labdata.html
9. Platt, J.: Sequential minimal optimization: a fast algorithm for training support vector machines (1998)
10. Rajasegarar, S., Leckie, C., Palaniswami, M.: Hyperspherical cluster based distributed anomaly detection in wireless sensor networks (2014)
11. Suthaharan, S., Alzahrani, M., Rajasegarar, S., Leckie, C., Palaniswami, M.: Labelled data collection for anomaly detection in wireless sensor networks (2010)
12. Tax, D.M., Duin, R.P.: Support vector data description (2004)
13. Xie, M., Han, S., Tian, B., Parvin, S.: Anomaly detection in wireless sensor networks: a survey (2011)
14. Zamani, M., Movahedi, M.: Machine learning techniques for intrusion detection (2013)