Anomaly Detection in Wireless Sensor Networks via Support Vector Data Description with Bayesian Optimization of Hyperparameters

Van Vuong Trinh[1], Kim Phuc Tran[1] and Anh Tuan Mai[2]
[1]Division of Artificial Intelligence, Dong A University, Danang, Vietnam
[2]International Training Institute for Materials Science (ITIMS),
Hanoi University of Science and Technology, Hanoi, Vietnam
email: vanvuong.trinh@gmail.com

I. Introduction

II. Support Vector Data Description and Preliminaries

In this section, we briefly recall the support vector data sescription (SVDD) originally derived in [1] which is an alternative of the known one-class support vector machines (OCSVM) [2] and analogous to support vector machines (SVM) [3].

A. Theory

We have a training data set $\{\mathbf{x}_i\}$, $i = 1, \dots, l$ for which we want to obtain a hypersphere with minimum volume, containing all (or most of) the data points. This is very sensitive to the most outliers in the training set. When one or a few very remote data points are in the training set, a very large hypersphere is obtained which will not represent the data approximately accurate. Therefore, we allow for some data points outside the hypersphere by introducing slack variables $\xi_i \geq 0$. Let a and R being reserved for the center and the radius of the hypersphere, the following primal optimization problem is considered:

$$\underset{R, a, \xi_i}{\text{Minimize}} \; R^2 + C \sum_{i=1}^{l} \xi_i \tag{1a}$$

$$\text{Subject to } \; ||x_i - a||^2 \leq R^2 + \xi_i, \, \xi_i \geq 0 \quad \forall i \tag{1b}$$

where the parameter C control the trade-off between the volume and errors (number of normal data rejected).

In order to extend description capability,

In the feature space ϕ, the primal problem changes into

$$\underset{R, a, \xi_i}{\text{Minimize}} \; R^2 + C \sum_{i=1}^{l} \xi_i \tag{2a}$$

$$\text{Subject to } \; ||\phi(x_i) - \phi(a)||^2 \leq R^2 + \xi_i, \, \xi_i \geq 0 \, \forall i \tag{2b}$$

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j) \tag{3}$$

This leads to the dual optimization problem, which is a standard quadratic program (QP), as follows.

$$\underset{\alpha_i}{\text{Maximize}} \; \sum_{i=1}^{l} \alpha_i K(x_i, x_i) - \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i \alpha_j K(x_i, x_j) \tag{4a}$$

$$\text{Subject to } \; \sum_{i=1}^{l} \alpha_i = 1, \, 0 \leq \alpha_i \leq C \, \forall i \tag{4b}$$

Such an optimisation is feasible if and only if the regularization parameter satisfies the inequality $C \geq 1/l$ with the optimal solution denoted as $\alpha_k^\star$.

The data points corresponding to $\alpha_k^\star > 0$ are consequently referred to as the *support vectors*. These support vectors characterize the scope of the description.

B. Discriminant Function

The image of the sphere center is

$$\phi(a) = \sum_{i=1}^{l} \alpha_i^\star \phi(x_i) \tag{5}$$

By definition, the radius R of the hypersphere is the kernel-based distance from the center a to any of the support vectors on the boundary of the hypersphere, i.e. one correponds to $0 < \alpha_k^\star < C$. Hence, the hypersphere radius R is determined as:

$$\begin{aligned} R^2 &= ||\phi(x_k) - \phi(a)||^2 \\ &= K(x_k, x_k) - 2 \sum_{i=1}^{l} \alpha_i^\star K(x_i, x_k) \\ &\quad + \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i^\star \alpha_j^\star K(x_i, x_j) \end{aligned} \tag{6}$$

for any support vector x_k that has $0 < \alpha_k^\star < C$.

Whether a data point z is normal or not is in accordance to the kernel-based distance to the center of the sphere.

$$R^2 - ||\phi(z) - \phi(a)||^2 = 2 \sum_{k} \alpha_k^\star K(z, x_k) + v \tag{7}$$

where

$$v = 2 \sum_{i=1}^{l} \sum_{j=1}^{l} \alpha_i^\star \alpha_j^\star K(x_i, x_j) + 1 - R^2 \tag{8}$$

C. Evaluation Measures via Geometric Mean Analysis

While training a SVDD does not require the availability of abnormal data, its evaluation does.

$$g = \sqrt{Acc^{+} \cdot Acc^{-}} \qquad (9)$$

III. Anomaly Detection Method for Wireless Sensor Networks using LOF and SVDD

A. Description of Aproach

B. Local Outlier Factor

C. Bayesian Optimization for Hyperparameter Selection

Similar to other kernel methods, the performance of SVDD models is strongly affected by the kernel parameters.

IV. Simulation Results

[4]

A. Intel Berkeley Research Laboratory Data

We consider a data set gathered from a wireless sensor network deployment at the Intel Berkeley Research Laboratory (IBRL) [5]. A wireless sensor network consisting of 54 *Mica2Dot* sensor nodes was deployed in the IBRL for a 30 day (720 hour) period between 28th Feb 2004 and 5th April 2004 [10]. Figure 1 shows the deployed node locations in the laboratory. The sensors collect five measurements: light in Lux, temperature in degrees celsius, humidity (temperature corrected relative humidity) ranging from 0% to 100%, voltage in volts and network topology information in each 30 second interval. Node 0 is the gateway node. Other nodes transmit their data in multiple hops to the gateway node. The furthest node in the network is about 10 hops away from the gateway node. During the 30 day period, the 54 nodes collected about 2.3 million readings.

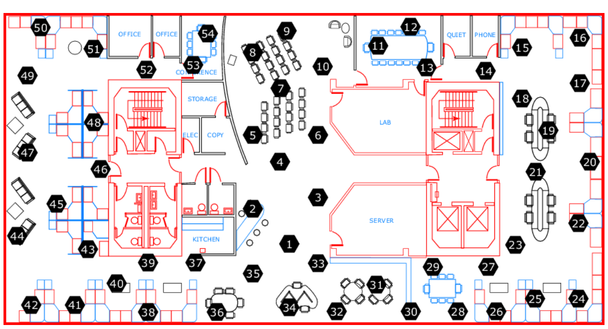

Fig. 1: A map of IBRL, with mote locations labeled in black.

In this paper we consider the IBRL data set obtained from 54 nodes, namely node IDs from 1 to 54, during the first 10 days period collected on March 2004. While the lab in Figure 1 has a total of 55 sensors (including the gateway node), only 54 of them provided data during the 10 days time window examined in this paper. Also, only two features, namely temperature and humidity, are taken into account. However, since node M_5 did not contain any humidity data during this time window, we only

It is notable that the notes contain a number of outliers.

Nodes 15 and 18 provided unrealistic data during some time intervals, i.e. too high or too low.

B. Numerical Analysis of Parameters Optimization Algorithm

Platform: 2.6 GHz Intel(R) Core(TM) i7 and 16GB of RAM.

Primal solution of SVDD is obtained using ILOG CPLEX 12.7.0.

NLopt nonlinear optimization package [6]

V. Conclusion and Future Work

References

[1] D. M. J. Tax and R. P. W. Duin, “Support Vector Data Description,” *Machine Learning*, vol. 54, no. 1, pp. 45–66, 2004.

[2] B. Schölkopf, R. Williamson, A. Smola, J. Shawe-Taylor, and J. Platt, “Support Vector Method for Novelty Detection,” *Advances in Neural Information Processing Systems*, vol. 12, pp. 582–588, 2000.

[3] V. N. Vapnik, *Statistical Learning Theory*, 1998, vol. pp.

[4] S. Suthaharan, M. Alzahrani, S. Rajasegarar, C. Leckie, and M. Palaniswami, “Labelled data collection for anomaly detection in wireless sensor networks,” *2010 Sixth International Conference on Intelligent Sensors, Sensor Networks and Information Processing*, pp. 269–274, 2010.

[5] P. Buonadonna, D. Gay, J. M. Hellerstein, W. Hong, and S. Madden, “TASK: Sensor network in a box,” *Proceedings of the Second European Workshop on Wireless Sensor Networks, EWSN 2005*, vol. 2005, pp. 133–144, 2005.

[6] S. G. Johnson, “The nlopt nonlinear-optimization package,” http://ab-initio.mit.edu/nlopt.