A Lightweight Anomaly Detection Method Based on SVDD for Wireless Sensor Networks

Yunhong Chen[1] · Shuming Li[1]

Abstract
Limited resources and harsh deployment environments may cause raw observations collected by sensor nodes to have poor data quality and reliability, which will influence the accuracy of the analysis and decision making in wireless sensor networks (WSNs). Therefore, anomaly detection must be implemented on the data collected by nodes. Support vector data description based on spatiotemporal and attribute correlations (STASVDD) can efficiently detect outliers. A novel optimization method based on STASVDD (N-STAS-VDD) is put forward in this paper. The proposed method considers that outliers can independently occur in each attribute when the collected data vectors are independent and identically distributed in WSNs. The proposed method applies the concept of core-sets to reduce the computational complexity of the quadratic programming problem in STAS-VDD, consequently reducing the energy consumption of resources-constrained WSNs. In addition, comparing the distributed and centralized detection approach of this method, the results show that the distributed approach has better performance because it relieves the communication burden. Extensive experiments were performed on both synthetic and real WSNs datasets. Results revealed that N-STASVDD achieves low time complexity and high detection accuracy.

Keywords Anomaly detection · Wireless sensor networks · Spatiotemporal and attribute · SVDD · Computational complexity

1 Introduction

Wireless sensor networks (WSNs) have many applications in different fields, such as smart citys, [1] smart grid, [2] environmental monitoring [3] and medical sensing [4]. However, the innate characteristics of WSNs render the sensor node vulnerable to anomalies caused by resource constraints, including energy, memory, computation, bandwidth, and transmission channel. Anomalies are caused by faulty sensor nodes, security threats in the network, or unusual phenomena in the monitoring scope. Therefore, anomaly detection must be

✉ Yunhong Chen
   yunhong1619@163.com

[1]   College of Educational Information and Technology, Hubei Normal University, No. 11 Ci Hu
      Road, Huangshi, Hubei, China

implemented in WSNs [5] so that accurate information can be obtained and effective decisions can be made by information gatherers. Researches have proposed several anomaly detection approaches for WSNs [6–11], such as statistical techniques, nearest-neighbor-based approaches, data mining, and machine learning methods.

In In statistical techniques, a statistical model is established to determine the data distribution, and evaluate the data samples in terms of their suitability for the model. Zhang et al. [12] proposed a statistical outlier detection method based on spatial and temporal correlations of data in WSNs. This method uses the time series to determine the statistical distribution model of the data to achieve the outlier detection, which will lead to a larger amount of computations and affect the energy limited wireless sensor network lifetime. Dereszynski and Dietterich [13] presented a statistical method for identifying valid observations in data streams and distinguishing sensor failures in WSNs; this method exploits the spatial and temporal correlations of the data in real time. Due to the use of real-time approach, it will lead to the increase of the amount of calculations of the outlier detection, which will consume more energy and reduce the lifetime of the network. Li et al. [14] proposed an intrusion detection method based on the statistical distribution in WSNs. These statistical techniques exhibit good detection performance when the underlying data distributions are known. They are not feasible for application in the changing environments of WSNs, wherein data distributions are uncertain.

Nearest-neighbor-based approaches use several well-defined distance notions to calculate the distance between two data samples with similar measured values. A data sample is considered an outlier if it is located far from its neighbors. Branch et al. [15] proposed a distance-based method for outlier detection in WSNs. Zhang et al. [16] presented a distance-based scheme wherein global outliers are identified in snapshots and continuous query processing is performed. The above two methods are outlier detection method based on the reduction of network traffic. Zhuang et al. [17] proposed two in-network outlier cleaning schemes for data acquisition in WSNs. The first scheme uses wavelet analysis to detect outliers of noises or random errors. The second scheme employs distance-based dynamic time warping to detect outliers of random errors for a certain time period. These techniques have great computational complexities because they require the computation of the distances between each pair of data samples.

In recent years, many studies have been conducted on machine learning and data mining approaches for anomaly detection in WSNs [9, 18–23]. Moshtaghi et al. [18] proposed an adaptive method that can create elliptical decision boundaries for anomaly detection in WSNs and maintain the decision boundaries without the need for re-training. Zhang et al. [22] presented two ellipsoidal one-class SVM-based outlier detection techniques for identifying outliers in a distributed and online manner in WSNs. Rajasegarar et al. [24, 25] proposed a distributed approach of one-class quarter-sphere support vector machine (QSSVM) and a centered approach of hyper-ellipsoidal support vector machine (CESVM) for anomaly detection in WSNs; they compared and analyzed the detection accuracy and sensitivity to parameter settings of CESVM and QSSVM. Gol et al. [26] proposed a linear-programming-based fuzzy-constraint SVDD method for anomaly detection in WSNs. In general, data mining and machine learning methods can achieve the desired effect of anomaly detection in WSNs. However, they are hindered by the high computational complexity and large communication overheads for anomaly detection.

Nonparametric approaches for anomaly detection are kernel-based machine-learning methods, which do not require any prior knowledge regarding the data distribution [27, 28]. As such, these approaches are suitable for resource-constrained WSNs, wherein prior knowledge regarding the abnormal behavior of the collected data distribution cannot be

obtained in advance. However, an arising challenge in the implementation of nonparametric anomaly detection is acquiring labeled data for training a classifier. In particular, the training must be implemented frequently in WSNs to adapt to the change in normal behavior over time. Support vector data description (SVDD) [28–30] aims to address this challenge for unsupervised learning problems. Machine learning method can construct the normal area of the data and disregard a few errors or anomalies by the relaxation factor. This method can also deal with nonlinear samples of normal behavior by using a kernel function to map the samples in the input space into high-dimensional feature space. Therefore, this SVDD method is suitable for the problem of outlier detection. However, SVDD based on spatiotemporal and attribute correlations requires the solution for a computationally intensive quadratic programming problem, and is therefore unsuitable for application to WSNs. Moreover, sensor nodes have limited energy in WSNs, and most of the energy is consumed during information transmision rather than calculation [31, 32].

Therefore, the purpose in this paper is to propose a lightweight data mining method based on SVDD as well as perform anomaly detection in a distributed manner in WSNs. The main contributions of this article are as follows:

- We introduce a novel SVDD approach for anomaly detection in WSNs, namely spatiotemporal and attribute SVDD (STASVDD). When the collected data of node is independent and identically distributed in WSNs, the outliers can occur independently in each attribute of the data. STASVDD can solve this problem well, which combines spatiotemporal and attribute correlations of the collected data to implement anomaly detection.
- Given that WSNs have limited energy and that solving the quadratic programming problem by STASVDD will lead to high computational complexity, a novel optimization method based on STASVDD (N-STASVDD) for anomaly detection is proposed by using core-sets, which can reduce the computational complexity of STASVDD from $O(l^3)$ to $O(l)$. In addition, the method is applied in distributed manner to reduce the communication complexity in the anomaly detection of N-STASVDD.

The remainder of this paper is organized as follows. The problem of anomaly detection in WSNs is described in Sect. 2. N-STASVDD for anomaly detection in WSNs is proposed in Sect. 3. The distributed anomaly detection in WSNs is discussed in Sect. 4. In Sect. 5, the proposed algorithms are evaluated using synthetic and real data sets. Finally, the drawn conclusions are enumerated in Sect. 6.

2 Problem Statement

Consider a hierarchical architecture of WSNs deployed in a certain region, where multiple sensor nodes are connected with each other through wireless channel for monitoring m environmental attributes. The network shown in Fig. 1 is a hierarchical topology with seven sensor nodes. Nodes S_2 and S_3 are the direct parents of nodes S_4, S_5, S_6 and S_7, and are also members of the gateway node S_1. Each node S_i is connected to a set of spatially adjacent nodes, represented as $N(S_i)$. It is assumed that each sensor node is configured with $m(m \geq 2)$ different types of sensors, which will sense m-dimensional data at every sampling instant. In one region, the sense data by different adjacent nodes is a high correlation in spatiotemporal and attributes, such as temperature, humidity, pressure etc. At

Gateway node
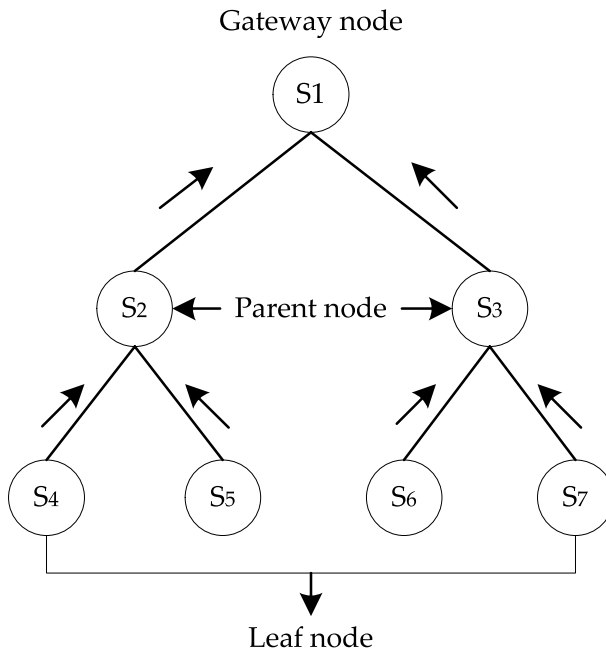
Fig. 1 The architecture on a hierarchical WSNs

each sampling instant k, each node S_i has a data vector x_{km}^i. The b neighboring nodes of S_i in the spatially are represented as S_{ij}, where $j = 1, 2, \ldots, b$. At the kth sampling instant, $\{x_{km}^i, x_{km}^{i1}, x_{km}^{i2}, \ldots, x_{km}^{ij}\}$ denotes the m-dimension data vectors at $\{S_i, S_{i1}, S_{i2}, \ldots, S_{ij}\}$. The problem is to identify normal or abnormal for every new sensed data vector x_{km}^i of the node S_i in real time. An anomaly detection approach based on spatiotemporal and attribute correlations of SVDD will be used to solve this problem.

3 The Proposed N-STASVDD for Anomaly Detection in WSNs

This section focuses on the method of N-STASVDD for anomaly detection in WSNs. Firstly, the idea of SVDD based on spatiotemporal correlations (STSVDD) is described for anomaly detection in WSNs. Secondly, the idea of STASVDD is discussed in detail. Finally, the optimization of STASVDD by using the idea of core-set is discussed and its computational complexity is analyzed.

3.1 SVDD Based on Spatiotemporal Correlations (STSVDD)

The basic idea of SVDD classifier [7, 8, 33] is to find the minimum hyper-sphere which contains all possible target data in the feature space. Give a set of training data $X_i = \{x_{1m}^i, x_{2m}^i, x_{3m}^i, \ldots, x_{lm}^i\}$ at the node S_i in the set $N(S_i)$, where $x_{km}^i \in \Re^m (1 < k < l)$ represents m-dimensional data vector corresponding to the number of attributes, and l is the size of the measurements corresponding to l sampling instants. Let X_i at the node S_i be mapped from the input space to feature space via a mapping function $\varphi(\cdot)$. R is the radius

of the minimum hyper-sphere, which is determined by a set of training data X_i, using the idea of SVDD. It is used to identify non-support vectors (NSVs), margin support vectors (MSVs), and non-margin support vectors (NMSVs) on the basis of the Lagrange multiplier α_i values. The sketch map of SVDD is shown in Fig. 2.

Aiming at a new arrived data x_{km}^i of the node S_i at a sampling instant, it is classified as the normal class if the distance between itself and the sphere center is less than or equal to the radius R. On the contrary, it is then classified as outliers. The outliers identified in this way are called local outliers, because it only considers the temporal correlation of the data on a single node S_i. While, the data obtained from a set of $N(S_i)$ has spatiotemporal correlation. Outlier identification is called global outliers in the set $N(S_i)$.

O'Reilly et al. [9] summarizes the outlier detection method of wireless sensor networks. This paper analyzes the method of considering temporal and spatial correlation with better detection performance compared to method of considering temporal correlation. So the method of STSVDD can achieve better detection results to some extent by considering spatiotemporal correlations in WSNs. However, this method does not consider the node data which is independent and identically distributed. When outlier occurs independently in each attribute of the node data, it will cause low anomaly detection rate. Therefore, an effective technology for anomaly detection should combine with attribute correlation on the basis of STSVDD.

3.2 SVDD Based on Spatiotemporal and Attribute Correlations (STASVDD)

Each node S_i of the set $N(S_i)$ in WSNs, consists of multiple sensors for measuring the m attributes of data x_{km}^i. Combined with SVDD formulation, x_{km}^i at each sampling instant k will determine the attribute radius R_A and the corresponding margin support vector based on attribute correlation. Next, the solution of the attribute radius R_A will be described. Given that
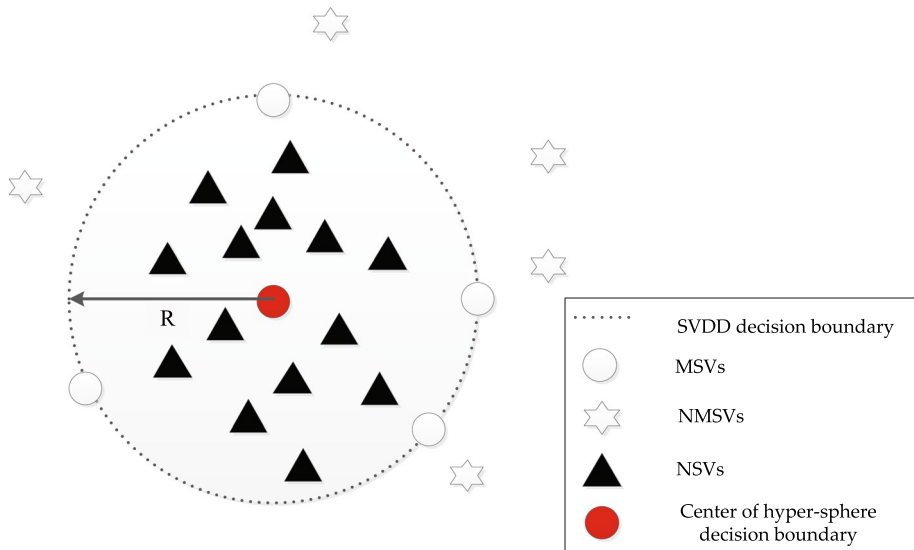

....	SVDD decision boundary
○	MSVs
☆	NMSVs
▲	NSVs
●	Center of hyper-sphere decision boundary

Fig. 2 The sketch map of SVDD

the training data set $X_i = \{x^i_{1m}, x^i_{2m}, x^i_{3m}, \ldots, x^i_{lm}\}$ at each sampling instant k of the node S_i, and let the vector x^i_{km} at each sampling instant k be mapped onto feature space by a mapping function $\varphi(\cdot)$. It will be divided into g portions of $m \times m$ dimensions each. Here g is equal to $\lfloor l/m \rfloor$, where $\lfloor \cdot \rfloor$ is the floor operation. As a result, each X_i corresponding to the node S_i can be expressed as the formula (1).

$$X_i = \{X_{1,s}, X_{2,s}, X_{3,s}, \ldots, X_{\lfloor l/m \rfloor, s}\} \tag{1}$$

Each part of the X_i at node S_i can be showed by $X_{g,s}$, where $g = 1, 2, \ldots, \lfloor l/m \rfloor$; $s = 1, 2, \ldots, m$. Thus, each $X_{g,s}$ is expressed by formula (2).

$$
\begin{aligned}
X_{g,s} &= \begin{bmatrix}
x_{m(g-1)+1,1} & x_{m(g-1)+1,2} & \cdots & x_{m(g-1)+1,m} \\
x_{m(g-1)+2,1} & x_{m(g-1)+2,2} & \cdots & x_{m(g-1)+2,m} \\
\vdots & \vdots & \ddots & \vdots \\
x_{m(g-1)+m,1} & x_{m(g-1)+m,2} & \cdots & x_{m(g-1)+m,m}
\end{bmatrix} \\
&= \begin{bmatrix} x'_{g1} & x'_{g2} & \cdots & x'_{gm} \end{bmatrix}
\end{aligned} \tag{2}
$$

In the matrix $X_{g,s}$, each row of data correspond to a specific sampling instant, and each column of data correspond to a different attribute. A method based on spatiotemporal and attribute correlation is proposed here. Using each column of $X_{g,s}$ as a m-dimension data vector, the attribute radius R_A will be obtained by applying the constrained optimization problem of SVDD. Therefore, m consecutive time measurements for a single attribute are used as a vector for optimization purpose. Compared with the previous method of STS-VDD, the method takes into consideration each row of $X_{g,s}$ as data vector for optimization. Thus, in view of $X_{g,s}$ of the gth row vector of $X_{g,s}$, the primal optimization problem of SVDD can be defined as following:

$$
\begin{aligned}
&Min \; R^2_{Ag} + C \sum_{s=1}^{m} \zeta_{g,s} \\
&s.t. \left\| \varphi'(x'_{gs}) - a_g \right\|^2 \leq R^2_{Ag} + \zeta_{g,s} \\
&\zeta_{g,s} \geq 0, g = 1, 2, \ldots, \left\lfloor \frac{l}{m} \right\rfloor, s = 1, 2, \ldots, m
\end{aligned} \tag{3}
$$

where, $\varphi'(x'_{gs})$ is the image of attribute vector x'_{gs} and acquires via a mapping function $\varphi(\cdot)$. R_{Ag} and a_g denote the radius and center of the hyper-sphere respectively in the feature space, $\zeta_{g,s}$ is the slack variable to allow for a few training data outside the hyper-sphere [9, 13], and the penalty parameter C controls the trade-off between the volume of the hyper-sphere and the number of target data outside the hyper-sphere.

In order to solve the optimization problem of Eq. (3) with these constraints, Lagrange function is constructed as follows:

$$
\begin{aligned}
L_g\left(R_{Ag}, a_g, \zeta_{g,s}, \alpha_{g,s}, \gamma_{g,s}\right) &= R^2_{Ag} + C \sum_{s=1}^{m} \zeta_{g,s} \\
&- \sum_{s=1}^{m} \alpha_{g,s} \left(R^2_{Ag} + \zeta_{g,s} - \left\| \varphi'(x'_{gs}) - a_g \right\|^2 \right) - \sum_{s=1}^{m} \gamma_{g,s} \zeta_{g,s}
\end{aligned} \tag{4}
$$

In the above equation, Lagrange function for the parameter g expansion is obtained as follows:

$$
\begin{bmatrix}
L_1(R_{A1}, a_1, \zeta_{1,s}, \alpha_{1,s}, \gamma_{1,s}) \\
L_2(R_{A2}, a_2, \zeta_{2,s}, \alpha_{2,s}, \gamma_{2,s}) \\
\vdots \\
L_{\left\lfloor \frac{l}{m} \right\rfloor}\left(R_{A\left\lfloor \frac{l}{m} \right\rfloor}, a_{\left\lfloor \frac{l}{m} \right\rfloor}, \zeta_{\left\lfloor \frac{l}{m} \right\rfloor,s}, \alpha_{\left\lfloor \frac{l}{m} \right\rfloor,s}, \gamma_{\left\lfloor \frac{l}{m} \right\rfloor,s}\right)
\end{bmatrix}
$$

$$
=
\begin{bmatrix}
R_{A1}^2 & 0 & \cdots & 0 \\
0 & R_{A2}^2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & R_{A\left\lfloor \frac{l}{m} \right\rfloor}^2
\end{bmatrix}
$$

$$
\times \left(
\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}
- \sum_{s=1}^{m}
\begin{bmatrix} \alpha_{1,s} \\ \alpha_{2,s} \\ \vdots \\ \alpha_{\left\lfloor \frac{l}{m} \right\rfloor,s} \end{bmatrix}
\right)
$$

$$
+ \sum_{s=1}^{m}
\begin{bmatrix}
\alpha_{1,s}\left(\left\| \varphi'(x'_{1s}) - a_1 \right\|^2 - \zeta_{1,s} \right) \\
\alpha_{2,s}\left(\left\| \varphi'(x'_{2s}) - a_2 \right\|^2 - \zeta_{2,s} \right) \\
\vdots \\
\alpha_{\left\lfloor \frac{l}{m} \right\rfloor,s}\left(\left\| \varphi'\left(x'_{\left\lfloor \frac{l}{m} \right\rfloor s}\right) - a_{\left\lfloor \frac{l}{m} \right\rfloor} \right\|^2 - \zeta_{\left\lfloor \frac{l}{m} \right\rfloor,s} \right)
\end{bmatrix}
\tag{5}
$$

$$
+ \sum_{s=1}^{m}
\begin{bmatrix}
(C - \gamma_{1,s})\zeta_{1,s} \\
(C - \gamma_{2,s})\zeta_{2,s} \\
\vdots \\
\left(C - \gamma_{\left\lfloor \frac{l}{m} \right\rfloor,s}\right)\zeta_{\left\lfloor \frac{l}{m} \right\rfloor,s}
\end{bmatrix}
$$

$$
+ \sum_{s=1}^{m}
\begin{bmatrix}
\alpha_{1,s}\left(\left\| \varphi'(x'_{1s}) - a_1 \right\|^2 - \zeta_{1,s} \right) \\
\alpha_{2,s}\left(\left\| \varphi'(x'_{2s}) - a_2 \right\|^2 - \zeta_{2,s} \right) \\
\vdots \\
\alpha_{\left\lfloor \frac{l}{m} \right\rfloor,s}\left(\left\| \varphi'\left(x'_{\left\lfloor \frac{l}{m} \right\rfloor s}\right) - a_{\left\lfloor \frac{l}{m} \right\rfloor} \right\|^2 - \zeta_{\left\lfloor \frac{l}{m} \right\rfloor,s} \right)
\end{bmatrix}
$$

$$
+ \sum_{s=1}^{m}
\begin{bmatrix}
(C - \gamma_{1,s})\zeta_{1,s} \\
(C - \gamma_{2,s})\zeta_{2,s} \\
\vdots \\
\left(C - \gamma_{\left\lfloor \frac{l}{m} \right\rfloor,s}\right)\zeta_{\left\lfloor \frac{l}{m} \right\rfloor,s}
\end{bmatrix}
$$

where each $\alpha_{g,s} \geq 0$, $\gamma_{g,s} \geq 0$, $\forall g = 1, 2, \ldots, \left\lfloor \frac{l}{m} \right\rfloor$, $s = 1, 2, \ldots, m$ is the Lagrange multipliers and x'_{gs} is the sth column vector corresponding to $X_{g,s}$.

Define $L = \begin{bmatrix} L_1 & L_2 & \cdots & L_{\lfloor \frac{l}{m} \rfloor} \end{bmatrix}$ and $R_A = \begin{bmatrix} R_{A1} & R_{A2} & \cdots & R_{A\lfloor \frac{l}{m} \rfloor} \end{bmatrix}$. Using KKT conditions [34], L should be minimized with respect to $R_{Ag}, a_g, \zeta_{g,s}$ and maximized with respect to $\alpha_{g,s}$ and $\gamma_{g,s}$. In order to find the stationary point of the Lagrange function, it will set partial derivatives of L equal to zero. That is, $\frac{\partial L_f}{\partial R_{Ah}} = 0$ and $\frac{\partial L_f}{\partial \zeta_{h,s}} = 0$ for $f \neq h$, where $f = 1, 2, \ldots, \lfloor \frac{l}{m} \rfloor$ and $h = 1, 2, \ldots, \lfloor \frac{l}{m} \rfloor$, the Jacobi matrices are expressed as follows:

$$
J\left(R_{A1} \ R_{A2} \ \cdots \ R_{A\lfloor \frac{l}{m} \rfloor} \right) = \begin{bmatrix} \frac{\partial L_1}{\partial R_{A1}} & 0 & \cdots & 0 \\ 0 & \frac{\partial L_2}{\partial R_{A2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\partial L_{\lfloor \frac{l}{m} \rfloor}}{\partial R_{A\lfloor \frac{l}{m} \rfloor}} \end{bmatrix} \tag{6}
$$

In the same way, the Jacobi matrix for a and $\zeta_{g,s}$ can be obtained similar to (6). Now putting the Jacob equation equal to zero, the following equations are obtained.

$$
\sum_{s=1}^{m} \begin{bmatrix} \alpha_{1,s} \\ \alpha_{2,s} \\ \vdots \\ \alpha_{\lfloor \frac{l}{m} \rfloor, s} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \tag{7}
$$

$$
\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{\lfloor \frac{l}{m} \rfloor} \end{bmatrix} = \sum_{s=1}^{m} \begin{bmatrix} \alpha_{1,s} \varphi'(x'_{1s}) \\ \alpha_{2,s} \varphi'(x'_{2s}) \\ \vdots \\ \alpha_{\lfloor \frac{l}{m} \rfloor, s} \varphi'\left(x'_{\lfloor \frac{l}{m} \rfloor s}\right) \end{bmatrix} \tag{8}
$$

$$
\begin{bmatrix} \alpha_{1,s} \\ \alpha_{2,s} \\ \vdots \\ \alpha_{\lfloor \frac{l}{m} \rfloor, s} \end{bmatrix} = \begin{bmatrix} C - \gamma_{1,s} \\ C - \gamma_{2,s} \\ \vdots \\ C - \gamma_{\lfloor \frac{l}{m} \rfloor, s} \end{bmatrix} \tag{9}
$$

where $\sum_{s=1}^{m} \alpha_{g,s} = \alpha_{g,1} + \alpha_{g,2} + \cdots + \alpha_{g,m}$, $\sum_{s=1}^{m} \alpha_{g,s} \varphi'(x'_{gs}) = \alpha_{g,1} \varphi'(x'_{g1}) + \alpha_{g,2} \varphi'(x'_{g2}) + \cdots + \alpha_{g,m} \varphi'(x'_{gm})$.

From the last equation $\alpha_{g,s} = C - \gamma_{g,s}$ and using $\alpha_{g,s} \geq 0, \gamma_{g,s} \geq 0$, the following inequality can be obtained.

$$
0 \leq \alpha_{g,s} \leq C \tag{10}
$$

Resubstituting (7)-(9) into (5) results in:

$$
\begin{bmatrix}
L_1(R_{A1}, a_1, \zeta_{1,s}, \alpha_{1,s}, \gamma_{1,s}) \\
L_2(R_{A2}, a_2, \zeta_{2,s}, \alpha_{2,s}, \gamma_{2,s}) \\
\vdots \\
L_{\lfloor \frac{l}{m} \rfloor}(R_{A\lfloor \frac{l}{m} \rfloor}, a_{\lfloor \frac{l}{m} \rfloor}, \zeta_{\lfloor \frac{l}{m} \rfloor,s}, \alpha_{\lfloor \frac{l}{m} \rfloor,s}, \gamma_{\lfloor \frac{l}{m} \rfloor,s})
\end{bmatrix}
$$

$$
= \sum_{s=1}^{m}
\begin{bmatrix}
\alpha_{1,s} \left\| \varphi'(x'_{1s}) - a_1 \right\|^2 \\
\alpha_{2,s} \left\| \varphi'(x'_{2s}) - a_2 \right\|^2 \\
\vdots \\
\alpha_{\lfloor \frac{l}{m} \rfloor,s} \left\| \varphi'\left(x'_{\lfloor \frac{l}{m} \rfloor s}\right) - a_{\lfloor \frac{l}{m} \rfloor} \right\|^2
\end{bmatrix}
\tag{11}
$$

$$
= \sum_{s=1}^{m}
\begin{bmatrix}
\alpha_{1,s} \varphi'(x'_{1s}) \bullet \varphi'(x'_{1s}) \\
\alpha_{2,s} \varphi'(x'_{2s}) \bullet \varphi'(x'_{2s}) \\
\vdots \\
\alpha_{\lfloor \frac{l}{m} \rfloor,s} \varphi'\left(x'_{\lfloor \frac{l}{m} \rfloor s}\right) \bullet \varphi'\left(x'_{\lfloor \frac{l}{m} \rfloor s}\right)
\end{bmatrix}
$$

$$
- \sum_{s=1}^{m} \sum_{r=1}^{m}
\begin{bmatrix}
\alpha_{1,s} \alpha_{1,r} \varphi'(x'_{1s}) \bullet \varphi'(x'_{1r}) \\
\alpha_{2,s} \alpha_{2,r} \varphi'(x'_{2s}) \bullet \varphi'(x'_{2r}) \\
\vdots \\
\alpha_{\lfloor \frac{l}{m} \rfloor,s} \alpha_{\lfloor \frac{l}{m} \rfloor,r} \varphi'\left(x'_{\lfloor \frac{l}{m} \rfloor s}\right) \bullet \varphi'\left(x'_{\lfloor \frac{l}{m} \rfloor r}\right)
\end{bmatrix}
$$

Now, using the kernel trick, [25] in the feature space the dot product of two vectors in (11) can be calculated by a kernel function (12). Hence, the dual formation of the problem (3) will become (13).

$$
K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)
\tag{12}
$$

$$
\max \sum_{s=1}^{m}
\begin{bmatrix}
\alpha_{1,s} K(x'_{1s}, x'_{1s}) \\
\alpha_{2,s} K(x'_{2s}, x'_{2s}) \\
\vdots \\
\alpha_{\lfloor \frac{l}{m} \rfloor,s} K\left(x'_{\lfloor \frac{l}{m} \rfloor s}, x'_{\lfloor \frac{l}{m} \rfloor s}\right)
\end{bmatrix}
$$

$$
- \sum_{s=1}^{m} \sum_{r=1}^{m}
\begin{bmatrix}
\alpha_{1,s} \alpha_{1,r} K(x'_{1s}, x'_{1r}) \\
\alpha_{2,s} \alpha_{2,r} K(x'_{2s}, x'_{2r}) \\
\vdots \\
\alpha_{\lfloor \frac{l}{m} \rfloor,s} \alpha_{\lfloor \frac{l}{m} \rfloor,r} K\left(x'_{\lfloor \frac{l}{m} \rfloor s}, x'_{\lfloor \frac{l}{m} \rfloor r}\right)
\end{bmatrix}
\tag{13}
$$

$$
\text{s.t. } \sum_{s=1}^{m}
\begin{bmatrix}
\alpha_{1,s} \\
\alpha_{2,s} \\
\vdots \\
\alpha_{\lfloor \frac{l}{m} \rfloor,s}
\end{bmatrix}
=
\begin{bmatrix}
1 \\
1 \\
\vdots \\
1
\end{bmatrix}
$$

$$
0 \leq \alpha_{g,s} \leq C, g = 1, 2, \ldots, \left\lfloor \frac{l}{m} \right\rfloor, s = 1, 2, \ldots, m
$$

From Eq. (13) the value of $\alpha_{g,s}$ can be obtained using quadratic optimization technique. X_i is composed of $\left\lfloor \frac{l}{m} \right\rfloor$ sets of $\alpha_{g,s}$, which respectively corresponds to $\left\lfloor \frac{l}{m} \right\rfloor$ parts of X_i. Each set includes m number of $\alpha_{g,s}$, which can be further classified according to the results of $\alpha_{g,s}$. The data vectors corresponding to $\alpha_{g,s} = 0$, which are called non-support vectors and fall inside the hyper-sphere. The data vectors corresponding to $0 < \alpha_{g,s} < C$, which are called margin support vectors. Their distances to the center of hyper-sphere indicate the radius of hyper-sphere. The data vectors corresponding to $\alpha_{g,s} = C$, which are called non-margin support vectors and fall outside the hyper-sphere. Their distances to the center of hyper-sphere is larger than the radius of the hyper-sphere. Thus the corresponding sample points of these data vectors are considered to be outliers.

For a given set of training data X_i of node S_i in WSNs, the attribute radius R_{Ag}^2 corresponding to $X_{g,s}$ can be calculated by the following formula.

$$R_{Ag}^2 = \left\| x'_{gs} - a_g \right\|^2 \tag{14}$$

where x'_{gs} is margin support vector and a_g is the center of hyper-sphere in each part of X_i. The final attribute radius of X_i is then acquired by taking the mean of all $\left\lfloor \frac{l}{m} \right\rfloor$ radii. The center of the corresponding sphere can be determined by calculating the mean of all $\left\lfloor \frac{l}{m} \right\rfloor$ spheres.

$$R_A = mean\left\{ R_{A1}, R_{A2}, R_{A3}, \ldots R_{A\left\lfloor \frac{l}{m} \right\rfloor} \right\} \tag{15}$$

$$a = mean\left\{ a_1, a_2, a_3, \ldots a_{\left\lfloor \frac{l}{m} \right\rfloor} \right\} \tag{16}$$

The algorithm is given as follows, which is the determination of spatiotemporal and attribute radius at each node.

Step 1   Let X_i be the $l \times m$ data at sensor node S_i. The rows of X_i represent l sampling instants and the columns of X_i represent m attributes.

Step 2   Get $X_{g,s}$ by dividing X_i into $\left\lfloor \frac{l}{m} \right\rfloor$ parts.

Step 3   Construct the spatiotemporal and attribute optimization problem L_g for each $X_{g,s}$.

Step 4   Determine the Lagrange Multipliers $\alpha_{g,s}$ for each L_g.

Step 5   Obtain the center of sphere R_{Ag} and the radius a_g for each L_g.

Step 6   Calculate $R_A = mean\left\{ R_{A1}, R_{A2}, R_{A3}, \ldots R_{A\left\lfloor \frac{l}{m} \right\rfloor} \right\}$ and $a = mean\{a_1, a_2, a_3, \ldots a_{\left\lfloor \frac{l}{m} \right\rfloor}\}$ for X_i.

As you can see from the above procedure of algorithm, the obtained radius R_A and the center point a of hyper-sphere are used to detect the abnormal state of the node data in WSNs. However, STASVDD requires the solution of a computationally-intensive quadratic programming problem in the process of obtaining the decision boundary. The runtime complexity is of $O(l^3)$, where l is the number of training samples. Since energy is very vital to resource constrained WSNs, it is necessary to reduce computation complexity of quadratic programming problem in STASVDD.

3.3 A Novel Optimization STASVDD by Using Core-Sets (N-STASVDD)

On the broader perspective, the problem of sphere-finding in STASVDD is similar to the minimum enclosing ball problem (MEB) in computational geometry [35–37]. MEB problem is to computer a ball of minimum radius enclosing a given set of data vectors. The MEB algorithm combined with the idea of core-sets has the computational time that is only linear in the number of samples in the literature [36]. Therefore, inspired by this idea, a novel optimization STASVDD is proposed, which reduces the computation complexity of STASVDD from $O(l^3)$ to $O(l)$.

The main procedure is as follows. Firstly, how to determine an initial core-set that only contains m normal samples, and how to obtain the initial radius $R_{A,1}$ of the sphere. Secondly, the execution process of the proposed method is introduced in STASVDD, how to implement iterative procedure with the core-set of samples instead of all l training samples. Finally, the computational complexity of QP (quadratic programming) will be of $O(n^3) \ll O(l^3)$ because the size of the core-set is $n \ll l$. It is proved that the number of iterations is independent of l, and the proposed method has a linear computational complexity.

There are two key issues to deal with the initialization process. On the one hand, it is necessary to select m normal samples of time continuous as much as possible. The ideal choice is to obtain the m samples in X_i that are the most adjacent to the sample mean. However, in the kernel-induced feature space, it will cost $O(l^2)$ time in order to obtain the m samples nearest to the sample mean. It is self-contradictory that the goal of runtime is only linear in l. Since the data obtained by sensor node are usually normal at the beginning in WSNs, the initial m samples are called l_0 that are fixed to select at the beginning of the m sampling instants from X_i. The STASVDD is run on these l_0 samples to get a sphere with the center a_0. The sample y is able to choose from these l_0 samples which is the most adjacent to a_0. On the other hand, the initial radius $R_{A,1}$ of the sphere is set that is yet key issue. Theoretically, the smaller $R_{A,1}$ will be more appropriate so that the initial sphere does not contain any outlier. Hence, a sample x_0 is first selected from l_0 sample above. Meanwhile, the the sample $z \in X_i$ that is the farthest distance from the sample x_0 is searched out. Define $B = \|x_0 - z\|$. R_M is radius, which can be determined by using the MEB algorithm in the sample set of X_i. It is obvious that $B \geq R_M$. $R_{A,1} = B/p$ is initialized, where $p > 1$, suerdefined constant and determine the number of iterations, such that $R_{A,1}$ is a much smaller number. Therefore, the following expression is established.

$$R_{A,1} \geq R_M/p \tag{17}$$

3.3.1 Execution Process

After initialization, a set of samples is added to core-set incrementally, which is a multiple of m. The center, radius and the core-set are expressed as a_t, $R_{A,t}$, X_{it} at the tth iteration, respectively. Moreover, the value of C is assumed to have been given in STASVDD, which is an upper bound on the fraction of outliers.

The process of STASVDD using the formulation of the core-set is given as follows.

Step 1   Initialize $R_{A,1}$ and y as mentioned above. Set $X_{i1} = \{l_0\}$, $a_1 = y$ and $t = 1$.

Step 2   Find the set Q_t of samples in X_i that fall outside the $(1 + \sigma)$-sphere $G_{a_t(1+\sigma)R_{A,t}}$. In other words,

$$Q_t = \left\{ x \in X_i \big| \|x - a_t\| > (1 + \sigma)R_{A,t} \right\} \tag{18}$$

Step 3   If the size of Q_t is smaller than $1/C$, the expected number of outliers, then terminate.

Step 4   Otherwise, enlarge the sample size of core-set X_{it} by including the sample in Q_t that is closest to a_t and do not belong to the sphere with the radius of $R_{A,t}$. Denote the enlarged core-set by $X_{i(t+1)}$.

Step 5   Run STASVDD on $X_{i(t+1)}$, also acquire the new center a_{t+1} and the radius $R_{A,(t+1)}$ of the sphere.

Step 6   Perform the constraint that

$$R_{A,(t+1)} \geq (1 + \delta\sigma)R_{A,t} \tag{19}$$

where, δ is a small constant defined by the user. That is, the radius at each iteration must be increased by at least $\delta\sigma R_{A,t}$.

Step 7   Increment t by one and then return to step 2.

3.3.2 Analysis of Computational Complexity

In the MEB problem, it can be indicated that the number of iterations is of $O(1/\sigma^2)$ in similar steps as above [35]. Even the number of iteration is of $O(1/\sigma)$ when the farthest sample is used in each iteration [37]. Because of the presence of a slack variable in the STASVDD formulation, it is not directly applied here. However, the computational complexity of the above algorithm is analyzed that is only linear in the number of training samples l.

Consider first step 1. As l_0 is fixed, both running the initial STASVDD and searching of y only spend $O(1)$ time. In identifying the initial radius $R_{A,1}$ and searching of z spend $O(l)$ time. Thus the total cost of the time is $O(l)$ in the process of initialization. At the tth iteration, combining formula (17) with (19), at least increase in $R_{A,t}$ is shown in the following formula (20). Obviously, R_M is an upper bound on the radius of the acquired sphere. Therefore, the total number of iterations is no more than $p/\delta\sigma = O(1/\sigma)$.

$$\delta\sigma R_{A,t} \geq \delta\sigma R_{A,(t-1)} \geq \cdots \geq \delta\sigma R_{A,1} \geq \frac{\delta\sigma}{p}R_M \tag{20}$$

At each iteration, a sample set consisting of m continuous samples will be added to the core-set in step 4. Consequently, the size of X_{it} is mt and a_t is a linear combination of mt ϕ'-mapped samples. Thus, step 4 needs to spend time which is $O(mtl)$ at the tth iteration, and running STASVDD needs to spend time which is $O\big((m(t+1))^3\big) = O(t^3)$. The other steps spend only constant time. Thus, the total cost of the time is $O(mtl + t^3)$ for the tth iteration.

The total cost of the time is shown as the formula (21) for the whole process in N-STASVDD including initialization time and $W = O(1/\sigma)$ of iterations time.

$$
\begin{aligned}
T &= O(l) + \sum_{t=1}^{W} O\big(mtl + t^3\big) \\
&= O(n) + \left(\sum_{t=1}^{W} mt\right)O(l) + \sum_{t=1}^{W} t^3 \\
&= O\big(mW^2 l + W^4\big) = O\left(\frac{m}{\sigma^2}l + \frac{1}{\sigma^4}\right) = O(l)
\end{aligned}
\tag{21}
$$

Remark 1 Based on the above analysis, the computational complexity of N-STASVDD can be seen from the formula (21), which is $O(l)$ for a fixed σ. However, the computational

complexity of the STASVDD is $O(l^3)$ because it needs to solve the quadratic optimization problem of the formula (13). Therefore, the performance of computational complexity in N-STASVDD is significantly improved.

4 Distributed Anomaly Detection in WSNs

According to network architecture in the second section, an approach of distributed anomaly detection is used in wireless sensor networks deployed in hostile environment. Each sensor node with multiple sensors collects a set of measurements monitored environment at every sampling instant. The purpose of this article is mainly to discuss the local and global anomaly detection for the collected data by nodes in WSNs. Local anomalies are identified using similarities among data within a single sensor node. Global anomalies are identified considering similarities on the union set of measurements representing multiple sensor nodes on the network.

Local anomalies can be detected by using the data of a single node, so it needs no communication overhead. Global anomalies can be detected by using the data of multiple sensor nodes, so it will generate some communication overhead and consume the energy of sensor node. The scheme of centralized anomaly detection needs to gather all the sensor measurements to the gateway. However, these data communication will consume the energy in the network and are bound to reduce the lifetime of the network [32, 38]. Thus, the distributed approach of energy efficient is suitable for anomaly detection in WSNs. These urge us to propose a distributed anomaly detection scheme based on STASVDD that can be used to detect local and global anomalies for the data collected by sensor nodes in WSNs. This scheme is described as follows.

Each sensor node S_i runs the N-STASVDD algorithm on its local measurements and acquires the local radius. The local radius is used to determine whether the new measurement is abnormal. Each sensor node S_i transmits its radius information to its parent node S_p. The parent node computes the global radius based on the mean strategy, which combines radius information from its own and its children nodes. The parent node sends back the global radius to all of its children nodes. For a new received data, the child node uses the global radius to determine whether it is a global anomaly.

Research on the problem of anomaly detection, a typical topology of WSNs is applied in this paper, as shown in Fig. 1. Taking the network topology structure as an example, our scheme for distributed anomaly detection is analyzed in WSNs. The local radius of any single node is used to detect local anomaly in the network, which can be obtained by running the N-STASVDD algorithm on the local measurements. The global radius can be obtained at any parent node of the hierarchy in the network and is used to detect global anomaly. For example, the global radius of the parent node S_2 can be obtained by computing the mean value of the radii from its own and its children S_4 and S_5. The node S_4 or S_5 will implement global anomaly detection that needs to apply the global radius from the node S_2. Similarly, if the node S_2 or S_3 uses the global radius from the node S_1 to detect global anomaly, then the global radius is considered by the local radius information of all nodes in the network in Fig. 1. Therefore, this illustrates that the distributed approach can be flexibly to realize anomaly detection for the local region or global region of the network on the basis of actual requirement.

Remark 2 In the process of distribution anomaly detection, the approach only requires to exchange the radius information and not to exchange the other information between the parent node and the children node, so the communication complexity is $O(1)$ on each link. Compared with the centralized approach, all data is centralized to the central node for anomaly detection,so the communication complexity is $O(lm)$ on each link. This distributed approach greatly reduces communication overhand and thus effectively prolongs the network lifetime. Moreover, with the expansion of the scale of the network, it can significantly increase the amount of data communicated in the centralized approach. Meanwhile, it is meaningless that the anomaly detection will consume a large amount of limited energy in WSNs. Rajasegarar et al. [24, 33] and other related literature have been compared with distributed and centralized outlier detection methods. The results show that the distributed method is superior to the centralized method. Thus,the distributed approach is more suitable for data anomaly detection in WSNs because the approach only needs to transmit a small amount of data and can implement the local anomaly detection as required.

The amount of data transmission and data calculation in wireless sensor networks are the main energy consumption of network nodes. The greater the amount of data to be calculated and the amount of data to be transmitted, the greater the energy consumption will be, and vice versa.Furthermore, the distributed approach is not restricted by the hierarchical topology of the network and can be applied to any topology of the network. The parent node and the child node can be flexible to determine and effectively detect global anomaly. Therefore, this distributed approach has certain robustness to the fault nodes in the network, so as to improve the accuracy of the anomaly detection in resources-constrained WSNs.

5 Experiment

5.1 Simulation Scenario

In this section, the performance of our proposed method is evaluated by applying it to synthetic and real data sets, and compared with the method of FCSVDD and LP-FCSVDD in the literature [26]. All experimental evaluations are performed on MATLAB with the data sets and are run on Intel Core i5 CPU, 3.30 GHZ, and the OS is windows 7. All of the experimental the parameters for this N-STASVDD method are set as follows: the initial value of l_0 is set to m, p in Eq. (17) is set to 5,and δ in Eq. (19) is set to 0.01. The synthetic dataset consists of three features with a mixture of Gaussian distributions. For each attribute of the Gaussian distribution, its mean is randomly selected from $(0.3 - 0.6)$ and the variance 0.03, and uniformly distributed outlier ranging between [0.60, 1] is added to each feature of the dataset with the ratio of 5%. The data set for 20 sensor nodes is created and combined. The combined data comprise 3000 data samples of three features, including 5% abnormal data. The entire data set is normalized to the range [0, 1]. Among, the training set consist of 2200 data samples and the testing set consists of 800 data samples.

In the first experiment, the synthetic dataset with three attributes is applied in the proposed methods. The RBF kernel is used as the distance based kernel for this evaluation, which can be represented as $k_{rbf} = \exp\left(\left\|y_i - y_j\right\|^2 / \eta^2\right)$ for data vector y_i and y_j, where η is the width parameter of the kernel function.

In each simulation, the measured values of the false positive, the true positive, the false positives rate (FPR) and the true positive rate (TPR) were recorded. The false positive means that a anomalous measured value is detected as normal by the detector. The true positive means which an actual normal measured value is correctly identified by the detector. The FPR is calculated as the percentage ratio between the false positives and the actual anomalous measurements. The TPR is calculated as the percentage ratio between the true positives and the actual normal measurements. In order to compare several methods of STSVDD, STASVDD, N-STASVDD, FCSVDD and LP-FCSVDD, receiver operating characteristic (ROC) curves were acquired for each anomaly detection scheme. The ROC curve plots the TPR versus the FPR by varying one of the parameters of the detection scheme while the others are fixed. The value of AUC can be obtained by calculated the area under the ROC curve, which has better performance if the value is more close to 1.

Figure 3 illustrates the AUC curves for STSVDD, STASVDD, N-STASVDD, FCSVDD and LP-FCSVDD in the synthetic dataset by using the RBF kernel. Results are reflected in an exponential interval with different η parameter in the range $2^{-10} \sim 2^{40}$ and set the value of σ to 0.2 in N-STASVDD. From the experimental results, we can see that the introduced STASVDD has better performance than STSVDD and FCSVDD. Meanwhile, the proposed optimal method of N-STASVDD has a comparable performance contrast to STASVDD and is slightly better than LP-FCSVDD.

Figure 4 compares the time complexity of STSVDD, STASVDD, N-STASVDD, FCS-VDD and LP-FCSVDD with different number of training dataset. Here, RBF kernel is used and set the parameter η to 1. Seen from the figure, the time complexity of STSVDD and STASVDD is almost the same and slight superior to FCSVDD, which shows that the proposed STASVDD method is effective. When the training dataset is small, the time complexity of STSVDD, STASVDD and FCSVDD are faster than N-STASVDD and LP-FCS-VDD. This is because N-STASVDD and LP-FCSVDD has to run the *QP* multiple times
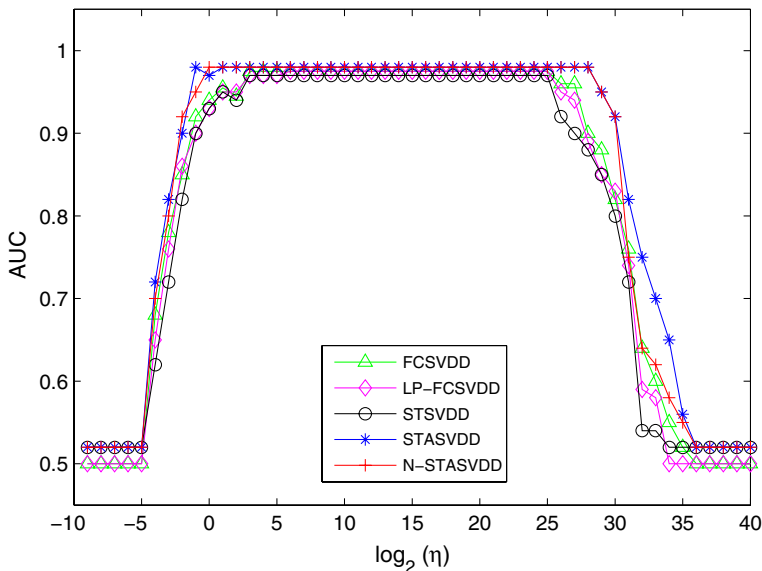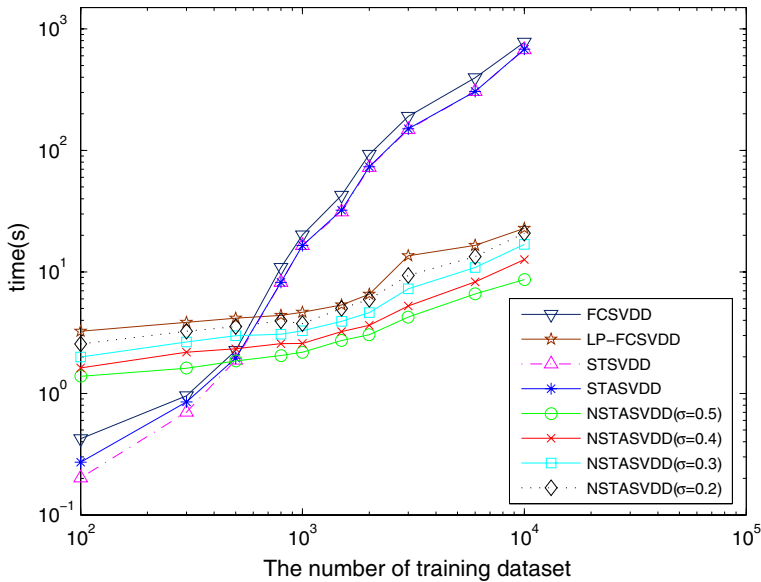

Fig. 3 The graph of AUC

Fig. 4 The graph of the time complexity

under this situation. However, with the continuous expansion of the training dataset, this can be clearly seen from Fig. 4 that N-STASVDD gets significantly faster than other three approaches. Meanwhile, the time complexity of our proposed N-STASVDD is superior to LP-FCSVDD. This is because the N-STASVDD method uses the idea of core-set to reduce the computational complexity from $O(l^3)$ to $O(l)$. Therefore, when the number of samples is greater than 800, the computational complexity of this method is greatly reduced.

5.2 Real Scenario

In the second experiment, the real dataset is obtained from a cluster of neighboring sensor nodes which is derived from a wireless sensor networks deployed in Grand-St-Bernard. Figure 5 illustrates the deployment. This sub-network consists of seven sensor nodes, namely nodes 2, 3, 6, 7, 11, 13 and 14. The sensor node record ambient temperature, surface temperature, solar radiation, relative humidity, soil moisture, watermark, rain meter, wind speed and wind direction measurement at 2 min interval. A continuous time period of 3000 data recorded is used in our experiment in September 2007. To validate our proposed method, we select five attributes to carry out the experiment, namely ambient temperature, solar radiation, relative humidity, soil moisture and wind speed. The obtained sensor data was standardized to zero mean and unit variance, using a data conditioning approach as in the literature [33]. Besides some of outliers, which account for 5% of the normal data, are generated randomly and introduced to the normal data. A three-level hierarchical structure of wireless sensor node as shown in Fig. 1, was formed with node 7 as gateway node, nodes 11 and 13 as the intermediate parent nodes, and the others as leaf nodes.

The purpose of experiment is to compare the performance of the proposed distributed anomaly detection approaches. Several methods were performed in MATLAB and some of
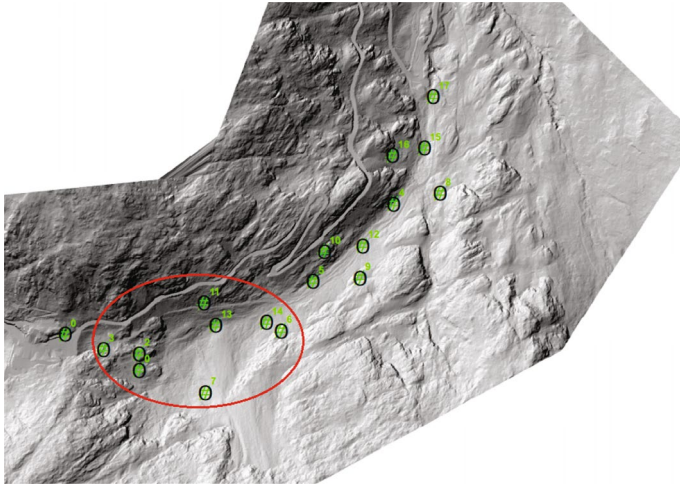
Fig. 5 Grand-St-Bernard deployment in SensorScope System

the function derived from PRtools and DDtools are utilized. Here, we mainly evaluate the STSVDD, STASVDD, N-STASVDD, FCSVDD and LP-FCSVDD strategy for anomaly detection. The radius of the sphere R is computed using any border support vector. RBF kernel function was considered in the evaluation. The training set is composed of 80% data samples and the testing set is composed of 20% data samples. Results are reported for the global radius calculation at the most top parent node (gateway node) in the network topology. The strategy of the global radius computation adopts the median value of all radii for the distributed detection scenario.

Figure 6 shows graphs of the ROC curves obtained for STSVDD, STASVDD, N-STAS-VDD, FCSVDD and LP-FCSVDD using RBF kernel. η is fixed at 1, C is varied from 0.01 to 1 in intervals of 0.01 and the value of σ is set to 0.2 in N-STASVDD. The graph indicates that the proposed N-STASVDD scheme shows better detection performance than other schemes in Grand-St-Bernard dataset. Among them, the AUC value of STSVDD is 0.9599, the AUC value of STASVDD is 0.9814, the AUC value of R-STASVDD is 0.9825, the AUC value of N-STASVDD is 0.9883, the AUC value of RN-STASVDD is 0.9891, the AUC value of FCSVDD is 0.9620, the AUC value of LP-FCSVDD is 0.9834. Meanwhile, Table 1 shows the time complexity of the above methods in the case of obtaining the

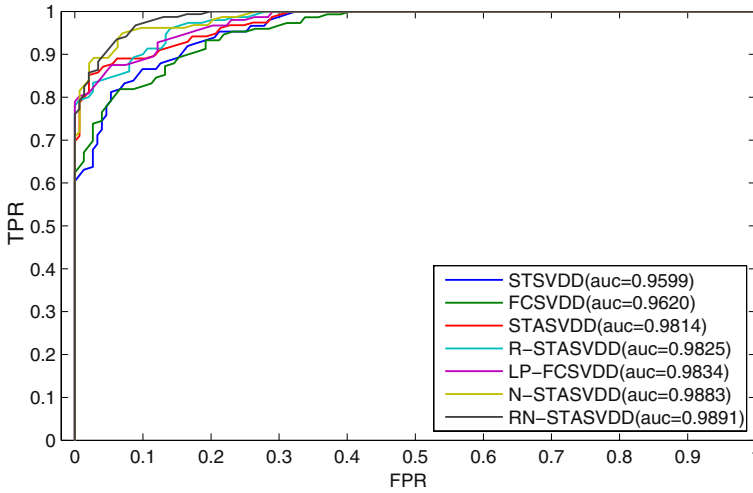| **Table 1** The time complexity of five methods	Method	Time(s)	AUC
	RN−STASVDD	1.465	0.9891
	N−STASVDD	1.536	0.9883
	LP−FCSVDD	2.654	0.9834
	R−STASVDD	24.317	0.9825
	STASVDD	26.425	0.9814
	FCSVDD	32.215	0.9620
	STSVDD	23.643	0.9599

Fig. 6 The graph of ROC curve

ROC curve of Fig. 5. Obviously, as can be seen from the value, STASVDD was superior to FCSVDD with a significant difference. N-STASVDD is slightly better than LP-FCSVDD. Meanwhile, R-STASVDD and RN-STASVDD are the results of the first part of the simulation data set. The result is better than the corresponding STASVDD and N-STASVDD, respectively, because the simulation data set is more normal distribution than the actual data set.

Figures 7 and 8 show that graphs of the FPR and TPR exploit distributed detection scenarios of the above five schemes with varying η and C values. Here, it is required to fix one of the parameters at a time and the value of σ is set to 0.2 in N-STASVDD. In Fig. 7a, FPR gradually increased with the increase of C. In the range of C, the value of the FPR of five schemes is the best when C is equal to 0.1. Among them, the FPR value of N-STASVDD is 3%, slightly lower than the other schemes. In Fig. 7b, the sensitivity of the detection scheme with C can be revealed. Better performance is revealed for values beyond 0.25 of
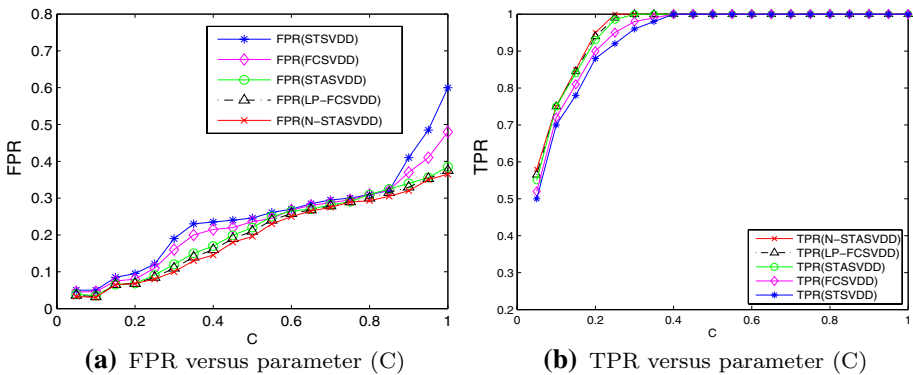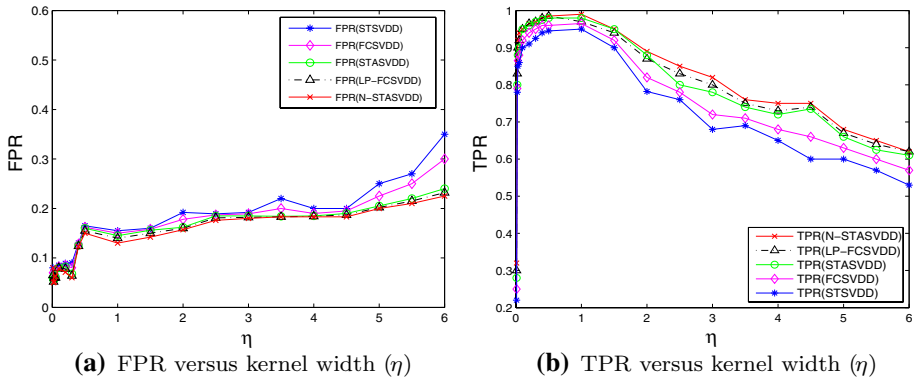

(a) FPR versus parameter (C)

(b) TPR versus parameter (C)

Fig. 7 The graphs of FPR and TPR

(a) FPR versus kernel width (η)   **(b)** TPR versus kernel width (η)

Fig. 8 The graphs of FPR and TPR

five schemes. In Fig. 8a, the FPR value of N-STASVDD is better than the other schemes with varying η values, and the FPR value of N-STASVDD gets the minimum value of 5% when η is equal to 0.02. In Fig. 8b, the best detection performance is reflected for η between 0.02 and 1.5 in five schemes, and then all of TPR value reach more than 85% and the performance of N-STASVDD was best. Seen from these pictures, STASVDD reveals significantly better performance than STSVDD and FCSVDD, and N-STASVDD reveals slightly better performance than LP-FCSVDD. Therefore, these results demonstrate that the distributed N-STASVDD scheme achieves comparable accuracy compared with the other schemes. In general, the proposed N-STASVDD has achieved good performance by using the distributed anomaly detection in WSNs.

6 Conclusions

Several of the existing anomaly detection methods in WSNs are analyzed based on the spatial and temporal correlations of the collected data. However, the collected data are independent and identically distributed, causing outliers to independently occur in each attribute. Thus, spatiotemporal and attribute correlations of the collected data must be considered to improve the detection performance. Therefore, a light-weight N-STASVDD approach is presented in this paper to address the problem of anomaly detection in WSNs.

The proposed approach is based on SVDD combined with spatiotemporal and attribute correlations of collected data in WSNs. Since SVDD is unsuitable for energy-constrained WSNs because it requires the solution for a computationally intensive quadratic programming problem. As for the computation complexity, a novel optimized method that uses core-sets in STASVDD (N-STASVDD) is presented to reduce the computation complexity from $O(l^3)$ to $O(l)$. Given that data transmission is the main energy consumption in WSNs, N-STASVDD performs anomaly detection in a distributed manner. To evaluate and validate the proposed method, both synthetic and the real WSNs dataset deployed in Grand-St-Bernard were used. We compared several methods of STSVDD, STASVDD, N-STASVDD, FCSVDD, and LP-FCSVDD. The results demonstrate that the distributed N-STASVDD achieves better detection accuracy and satisfactory performance in WSNs. This article uses a shorter sampling time of data in WSNs. The issue of a long sampling time needs to be further discussed in the follow-up work. Meanwhile, for the distributed

and centralized detection approaches, the work to be studied in the next stage is the mathematical theoretical analysis and experimental verification of the energy consumption problem.

References

1. Lanza, J., Sánchez, L., Muñoz, L., Galache, J. A., Sotres, P., Santana, J. R., et al. (2015). Large-scale mobile sensing enabled internet-of-things testbed for smart city services. *International Journal of Distributed Sensor Networks*, *11*, 785061.
2. Kbar, G., Al-Daraiseh, A., Mian, S. H., & Abidi, M. H. (2016). Utilizing sensors networks to develop a smart and context-aware solution for people with disabilities at the workplace (design and implementation). *International Journal of Distributed Sensor Networks*, *12*(9), 1550147716658606.
3. Souza, C. P., Carvalho, F. B., Silva, F. A., Andrade, H. A., Silva, N. D. V., Baiocchi, O., et al. (2016). On harvesting energy from tree trunks for environmental monitoring. *International Journal of Distributed Sensor Networks*, *12*, 9383765.
4. Yang, Y., Liu, Q., Gao, Z., Qiu, X., & Meng, L. (2015). Data fault detection in medical sensor networks. *Sensors*, *15*(3), 6066.
5. Xie, M., Han, S., Tian, B., & Parvin, S. (2011). Anomaly detection in wireless sensor networks: A survey. *Journal of Network and Computer Applications*, *34*(4), 1302.
6. Shahid, N., Naqvi, I. H., & Qaisar, S. B. (2015). One-class support vector machines: Analysis of outlier detection for wireless sensor networks in harsh environments. *Artificial Intelligence Review*, *43*(4), 515.
7. Shahid, N., Naqvi, I. H., & Qaisar, S. B. (2015). Characteristics and classification of outlier detection techniques for wireless sensor networks in harsh environments: A survey. *Artificial Intelligence Review*, *43*(2), 193.
8. McDonald, D., Sanchez, S., Madria, S., & Ercal, F. (2015). A survey of methods for finding outliers in wireless sensor networks. *Journal of network and systems management*, *23*(1), 163.
9. O'Reilly, C., Gluhak, A., Imran, M. A., & Rajasegarar, S. (2014). Anomaly detection in wireless sensor networks in a non-stationary environment. *IEEE Communications Surveys and Tutorials*, *16*(3), 1413.
10. Haque, S. A., Rahman, M., & Aziz, S. M. (2015). Sensor anomaly detection in wireless sensor networks for healthcare. *Sensors*, *15*(4), 8764.
11. Feng, Z., Fu, J., Du, D., Li, F., & Sun, S. (2017). A new approach of anomaly detection in wireless sensor networks using support vector data description. *International Journal of Distributed Sensor Networks*, *13*(1), 1550147716686161.
12. Zhang, Y., Hamm, N. A., Meratnia, N., Stein, A., Van de Voort, M., & Havinga, P. J. (2012). Statistics-based outlier detection for wireless sensor networks. *International Journal of Geographical Information Science*, *26*(8), 1373.
13. Dereszynski, E. W., & Dietterich, T. G. (2011). Spatiotemporal models for data-anomaly detection in dynamic environmental monitoring campaigns. *ACM Transactions on Sensor Networks (TOSN)*, *8*(1), 3.
14. Li, G., He, J., & Fu, Y. (2008). Group-based intrusion detection system in wireless sensor networks. *Computer Communications*, *31*(18), 4324.
15. Branch, J. W., Giannella, C., Szymanski, B., Wolff, R., & Kargupta, H. (2013). In-network outlier detection in wireless sensor networks. *Knowledge and information systems*, *34*(1), 23.
16. Zhang, K., Shi, S., Gao, H., & Li, J. (2007). Unsupervised outlier detection in sensor networks using aggregation tree, In *International conference on advanced data mining and applications* (pp. 158–169). Springer.

17. Zhuang, Y., & Chen, L. (2006). In-network outlier cleaning for data collection in sensor networks. In *CleanDB*
18. Moshtaghi, M., Leckie, C., Karunasekera, S., & Rajasegarar, S. (2014). An adaptive elliptical anomaly detection model for wireless sensor networks. *Computer Networks*, *64*, 195.
19. Salem, O., Guerassimov, A., Mehaoua, A., Marcus, A., & Furht, B. (2013). Anomaly detection scheme for medical wireless sensor networks. In *Handbook of medical and healthcare technologies* (pp. 207–222). Springer.
20. Rajasegarar, S., Leckie, C., & Palaniswami, M. (2014). Hyperspherical cluster based distributed anomaly detection in wireless sensor networks. *Journal of Parallel and Distributed Computing*, *74*(1), 1833.
21. Salmon, H. M., de Farias, C. M., Loureiro, P., Pirmez, L., Rossetto, S., Rodrigues, P. H. A., et al. (2013). Intrusion detection system for wireless sensor networks using danger theory immune-inspired techniques. *International Journal of Wireless Information Networks*, *20*(1), 39.
22. Zhang, Y., Meratnia, N., & Havinga, P. J. (2013). Distributed online outlier detection in wireless sensor networks using ellipsoidal support vector machine. *Ad Hoc Networks*, *11*(3), 1062.
23. Kumarage, H., Khalil, I., Tari, Z., & Zomaya, A. (2013). Distributed anomaly detection for industrial wireless sensor networks based on fuzzy data modelling. *Journal of Parallel and Distributed Computing*, *73*(6), 790.
24. Rajasegarar, S., Leckie, C., Palaniswami, M., & Bezdek, J. C. (2007). Quarter sphere based distributed anomaly detection in wireless sensor networks. *ICC*, *7*, 3864–3869.
25. Rajasegarar, S., Leckie, C., Bezdek, J. C., & Palaniswami, M. (2010). Centered hyperspherical and hyperellipsoidal one-class support vector machines for anomaly detection in sensor networks. *IEEE Transactions on Information Forensics and Security*, *5*(3), 518.
26. GhasemiGol, M., Ghaemi-Bafghi, A., Yaghmaee-Moghaddam, M. H., & Sadoghi-Yazdi, H. (2015). Anomaly detection and foresight response strategy for wireless sensor networks. *Wireless Networks*, *21*(5), 1425.
27. Shawe-Taylor, J., & Cristianini, N. (2004). *Kernel methods for pattern analysis, Kernel methods for pattern analysis*. Cambridge: Cambridge University Press.
28. Schölkopf, B., & Smola, A. J. (2002). Learning with kernels: Support vector machines, regularization, optimization, and beyond. In: *Learning with kernels: Support vector machines, regularization, optimization, and beyond* (p. 632). MIT Press.
29. Tax, D. M., & Duin, R. P. (2004). Support vector data description. *Machine Learning*, *54*(1), 45.
30. Wang, D., Yeung, D. S., & Tsang, E. C. (2006). Structured one-class classification. *IEEE Transactions on Systems, Man, and Cybernetics Part B (Cybernetics)*, *36*(6), 1283.
31. Pottie, G. J., & Kaiser, W. J. (2000). Wireless integrated network sensors. *Communications of the ACM*, *43*(5), 51.
32. Raghunathan, V., Schurgers, C., Park, S., & Srivastava, M. B. (2002). Energy-aware wireless microsensor networks. *IEEE Signal Processing Magazine*, *19*(2), 40.
33. Rajasegarar, S., Leckie, C., Palaniswami, M., & Bezdek, J. C. (2006). Distributed anomaly detection in wireless sensor networks. In *2006 10th IEEE Singapore international conference on communication systems* (pp. 1–5). IEEE.
34. Bishop, C. (2007). *Pattern recognition and machine learning (information science and statistics), 1st edn. 2006. corr. 2nd printing edn*. New York: Springer.
35. Bădoiu, M., Har-Peled, S., & Indyk, P. (2002). Approximate clustering via core-sets. In *Proceedings of the thiry-fourth annual ACM symposium on theory of computing* (pp. 250–257). ACM
36. Kumar, P., Mitchell, J. S., & Yildirim, E. A. (2003). Computing core–sets and approximate smallest enclosing hyperspheres in high dimensions. In *ALENEX* (pp. 45–55).
37. Badoiu, M., & Clarkson, K. L. (2003). Smaller core-sets for balls. In *Proceedings of the fourteenth annual ACM-SIAM symposium on discrete algorithms* (Society for Industrial and Applied Mathematics) (pp. 801–802).
38. Zhao, F., Liu, J., Liu, J., Guibas, L., & Reich, J. (2003). Collaborative signal and information processing: An information-directed approach. *Proceedings of the IEEE*, *91*(8), 1199.

Yunhong Chen is currently an associate professor in College of Educational Information and Technology, Hubei Normal University. She received her M.S. degree in CAT (Computer Application Technology) from Huazhong University of Science and Technology. Her research interest lies in Digital Media Technology.

Shuming Li is currently professor in College of Educational Information and Technology, Hubei Normal University. He received his Ph.D.degree from NERCEL (National Engineering Research Center for E-Learning) in Central China Normal University. His research interest lies at Education Information Science and Technology.