

One-class support vector machines: analysis of outlier detection for wireless sensor networks in harsh environments

Nauman Shahid · Ijaz Haider Naqvi · Saad Bin Qaisar

Published online: 29 January 2013
© Springer Science+Business Media Dordrecht 2013

Abstract Machine learning, like its various applications, has received a great interest in outlier detection in Wireless Sensor Networks. Support Vector Machines (SVM) are a special type of Machine learning techniques which are computationally inexpensive and provide a sparse solution. This work presents a detailed analysis of various formulations of one-class SVMs, like, hyper-plane, hyper-sphere, quarter-sphere and hyper-ellipsoidal. These formulations are used to separate the normal data from anomalous data. Various techniques based on these formulations have been analyzed in terms of a number of characteristics for harsh environments. These characteristics include input data type, spatio-temporal and attribute correlations, user specified thresholds, outlier types, outlier identification(event/error), outlier degree, susceptibility to dynamic topology, non-stationarity and inhomogeneity. A tabular description of improvement and feasibility of various techniques for deployment in the harsh environments has also been presented.

Keywords Wireless sensor networks · Harsh environments · Outlier detection · Event detection

N. Shahid (✉) · I. H. Naqvi
Department of Electrical Engineering, School of Science and Engineering, Lahore University of Management Sciences, Sector U, DHA, Lahore Cantt 54792, Pakistan
e-mail: naumanshahid6@gmail.com

I. H. Naqvi
e-mail: ijaznaqvi@lums.edu.pk

S. B. Qaisar
Bitsym Research, Bitsym LLC, Bitsym, MI 48823, USA
e-mail: saad@bitsym.com

1 Introduction

1.1 Wireless sensor networks

Wireless sensor networks (WSNs), due to their ability to perform online and distributed processing, have received considerable attention of research community for personal, industrial, business and military domains (Madden et al. 2002; Zhang et al. 2007; García-Hernández et al. 2007; Mainwaring et al. 2002; Cardell-Olivera et al. 2005; Ni et al. 2003; Akyildiz et al. 2002; Dereszynski and Dietterich 2011; Bahrepour et al. 2010, a; Phua et al. 2010; George et al. 2010). Such networks usually consist of a large number of sensor nodes distributed over a large area, with some powerful sink nodes which gather readings of sensor nodes. The sensor nodes can perform sensing, processing and wireless communication among themselves. Some of the most important applications of WSNs include sink initiated query, data gathering and periodic reporting, tracking and event detection. Of all these applications, this survey paper will focus on event detection application of WSNs deployed in harsh environments. A step-by-step formulation of the problem explored in this survey article is given below.

1.2 Harsh environments

Harsh environments are defined as “high stress environments which offer severe monitoring and communication challenges” (Misra et al. 2010). Underground oil, gas, coal, salt mines, forests and volcanic sites (García-Hernández et al. 2007) are typical examples of such environments. A more specific example of such an environment is an underground coal mine which should be monitored regularly to predict any disastrous conditions and to ensure the safety of personnel working inside. Various factors disrupt the communication and monitoring capabilities of sensor nodes in a harsh environment, some of which include high signal attenuation, electrical interference and multiple reflections or echoes, underground topology, instability in mine structures, ionized air, humid and warm conditions, gaseous hazards and noise (Shahid et al. 2012; Tutorial on Wireless Communications and Electronic Tracking 2009; Dario et al. 2005; Akyildiz et al. 2003). These constraints pose severe challenges in event detection applications in harsh environments because a safety assurance solution is highly dependent on monitoring and communication.

The constraints posed by such environments on a WSN can be divided into two major categories.

1. Communication constraints
2. Environmental monitoring constraints

The first set of constraints are caused due to extreme path loss, signal absorption, spreading, rapidly changing time-varying channels, large propagation delay, noise and fading characteristics. All these characteristics belong to a wireless communication channel of a harsh environment; hence, they effect communication among the nodes of a network.

The second set of constraints, also known as monitoring constraints, are directly related to the quality of data measured in a WSN. These constraints are imposed by dynamic changes in the network topology, instability in mine structure and dynamic changes in the data distribution of various attributes being measured in a harsh environment.

1.3 Outlier and event detection in harsh environments

Outlier and Event Detection has been the focus of research community in recent years (Dereszynski and Dietterich 2011; Bahrepour et al. 2010, a; Phua et al. 2010; George et al. 2010; Shahid et al. 2012; Ding et al. 2012; Ozdemir and Xiao 2012; O'Reilly et al. 2012; Zhang et al. 2012, a; Xie et al. 2012, a; Suthaharan 2012; Saada and Meng 2012; Shahid et al. 2012, a; Pham and Pagh 2012; McDonald et al. 2012; Ma et al. 2013; Moshtaghi et al. 2012; Stankovic et al. 2012; Trinidad et al. 2012; Shahid and Naqvi 2011; Giatrakos et al. 2010; Ozdemir and Xiao 2011; Liu 2011; Bezdek et al. 2011; Xue et al. 2006; Sheng et al. 2007; Katdare 2011; Hassan et al. 2011; Moshtaghi et al. 2011, a, b; Bahrepour et al. 2010; Bezdek et al. 2010; Suthaharan et al. 2010, a; Sharma et al. 2010; Rajasegarar et al. 2010, a; Paschalidis and Chen 2010; Keally et al. 2010; Giatrakos et al. 2010). Specifically, it has attracted huge attention for WSNs deployed in harsh environments. Here we briefly define the terms 'outlier' and 'event'.

- An 'event' in a harsh environment can be referred to as a 'disastrous condition' that may be characterized by an unexpected change in environmental conditions, for example, fire, unexpected rise in gaseous concentrations, earthquake etc.
- 'Outlier' is another term that is closely associated with 'event' and can be defined as an observation that differs significantly from the normal set of readings. The earliest definition of outlier was given by Grubbs in Barnett and Lewis (1994) as "*An outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs*". Further definitions can be found in Katdare (2011), Tan et al. (2006). In the context of WSNs we can say: "*Those measurements that significantly deviate from the normal pattern of sensed data*" (Chandola et al. 2009).

Outlying data should be analyzed, as it may indicate a system moving towards a state of natural disaster or event. Various accidents have occurred in underground mines worldwide in the previous years which caused severe financial impact and great loss of human lives <http://connekt.seecs.nust.edu.pk/SAHSE.php>, <http://www.humanite.fr/2006-03-10Societe-Catastrophe-de-Courrieres-une-expression-impropre>, <http://www.genuki.org.uk/big/eng/LAN/Haydock/WoodPitExplosion.html>, <http://www.msha.gov/MSHAINFO/FactSheets/MSHAFACT8.HTM>. Thus, **outlier detection** in harsh environments is essential for **event detection**, data quality assurance and control (QA/QC), fault detection, intrusion detection (Chen et al. 2006; Luo et al. 2006; da Silva et al. 2005; Bhuse and Gupta 2006), focused data collection, (Zouboulakis and Roussos 2007; Krishnamachari and Iyengar 2004; Ding et al. 2005; Ding and Cheng 2009) and adaptive system monitoring (Misra et al. 2010; Shahid et al. 2012).

1.4 Sources of outliers in a WSN deployed in a harsh environment

In a data set collected from WSNs, outliers may occur due to noise and errors, malicious attacks and most importantly **events** (John 1995; Han and Kamber 2006). In fact, an 'event' can actually be described as a sequence of 'outliers or erroneous readings' in a streaming data set (Shahid et al. 2012, a; Shahid and Naqvi 2011) and are considered to be the most important sources of outliers in harsh environments (Dereszynski and Dietterich 2011; Bezdek et al. 2011). For example, a sensor reading indicating a high temperature independent of its surrounding sensors is an outlier, whereas, a continuous stream of high temperature readings on a group of sensors located close together, indicates the presence of an event. Thus, an

event detection scheme can be derived from an **outlier detection** scheme (Shahid et al. 2012; Zhang et al. 2012a; Shahid et al. 2012; Bahrepour et al. 2009).

1.5 Purpose of outlier and event detection techniques for WSNs in harsh environments

The purpose of an outlier and event detection technique for WSNs deployed in harsh environments should be three-fold: outlier detection, event detection and event identification (Shahid et al. 2012; Zhang et al. 2010; Zhang 2010).

- To perform a separation between normal and outlying patterns from the collected data set. This process is known as ‘outlier detection’.
- Determine the potential sources of an outlier, that is to identify if the outlier is due to noise, sensor fault or an event. This process is known as ‘event detection’.
- In case of an event, determine the type of event and attributes that led to that event. This process is known as ‘event identification’

1.6 Impact of harsh environment constraints on the performance of outlier and event detection techniques

Two types of constraints were discussed in Shahid et al. (2012) that are imposed by harsh environments on a WSN: a) communication b) monitoring constraints. In a distributed processing scenario, outliers can be determined via processing on individual nodes of the network. Therefore, communication constraints posed by harsh environments weakly effect the performance of outlier detection algorithms: since such algorithms do not require significant communication between the nodes. However, performance of event detection algorithms may be strongly dependant on the nature of communication between the nodes. State-of-the-art outlier detection techniques have reduced the amount of communication between the nodes by orders of magnitude. Thus, only a few parameters related to the measured data are exchanged between the nodes. Hence, the performance of outlier and event detection algorithms is not significantly affected by the communication constraints.

Data collected by WSNs deployed in harsh environments is subject to temporally and spatially varying distributions. This is the key monitoring constraint of harsh environments that poses significant challenges for outlier and event detection techniques. To ensure a high performance of detection techniques, they should satisfy an additional set of characteristics when being used for WSNs deployed in harsh environments.

1.7 What is this survey about?

Various state-of-the-art techniques focus on outlier and event detection as well as event identification for WSNs. A detailed analysis of state-of-the-art outlier and event detection techniques for WSNs has been presented in Zhang et al. (2010). However, the research community needs to focus more on the application of outlier and event detection techniques for harsh environments. For an outlier and event detection technique to perform well in harsh environments, it should possess various characteristics. These characteristics, their importance and prioritization to the performance of outlier and event detection has been discussed in detail in (Shahid et al. 2012). The review paper (Shahid et al. 2012) also concluded that

clustering and classification based techniques are most suitable for outlier and event detection in harsh environments.

This survey paper is intended to provide a detailed analysis of state-of-the-art classification based outlier and event detection techniques which have been declared to be suitable for WSNs deployed in harsh environments by the authors of (Shahid et al. 2012). It begins with a brief introductory portion on WSNs and their various applications that are currently being explored by the research community, harsh environments, outlier and event detection and their importance from the perspective of WSNs deployed in harsh environments. In forthcoming sections, we discuss a few characteristics that are essential for state-of-the-art outlier and event detection techniques and have been derived mostly from Shahid et al. (2012). A major contribution of this paper, however, lies in the later sections which provide a detailed analysis of classification based outlier and event detection techniques. Further, a number of future research perspectives have been explored to develop optimal outlier and event detection techniques.

1.8 Organization

Rest of this paper is organized as following. Section 2 deals with a brief description of the classification of outlier and event detection techniques and the characteristics that should be satisfied by such techniques when used for harsh environments. This section also concludes that clustering and classification based techniques are most feasible for harsh environment based applications. Section 3 describes the methods for analysis of classification based techniques. Section 4 describes various types of classification based techniques which can be used for outlier and event detection in WSNs and short-lists one-class SVM based techniques as the most suitable for this purpose. Section 5 introduces the reader to a set of definitions and notations used in various one-class SVM formulations. Section 6 presents a detailed analysis of outlier and event detection techniques based on one-class SVM formulations. Section 7 discusses a few problems associated with SVMs and future research directions and Sect. 8 concludes the article.

2 Characteristics and classification of outlier and event detection techniques for WSNs deployed in harsh environments—a brief analysis

A certain set of characteristics should be satisfied by outlier and event detection techniques to be suitable for WSNs deployed in harsh environments. These characteristics were discussed in detail in (Shahid et al. 2012) and various types of techniques were analyzed in terms of these characteristics. This section presents a brief overview of the analysis carried out by the authors of (Shahid et al. 2012) in terms of following:

- Classification of state-of-the-art outlier and event detection techniques for WSNs.
- Characteristics of outlier and event detection techniques for WSNs deployed in harsh environments.
- State-of-the-art techniques suitable for WSNs deployed in harsh environments (based on the discussion in Shahid et al. 2012).

The short discussion carried out in this section will be used to perform a more rigorous analysis of the techniques discussed in the forth-coming sections. We first give a brief overview of the classification of state-of-the-art outlier and event detection techniques for WSNs in the form of Table 1. A detailed explanation of this classification criteria can be

Table 1 Classification of state-of-the-art outlier detection techniques for WSNs in harsh environments [Shahid et al. \(2012\)](#)

	Types and description	Examples
1	Statistical	
a.	Assume a statistical model for data distribution	Dereszynski and Dietterich (2011), Zhang et al. (2012a), Sheng et al. (2007), Wu et al. (2007), Yozo et al. (2004), Jun et al. (2005), Bettencourt et al. (2007), Palpanas et al. (2003), Subramaniam et al. (2006)
b.	Outliers are declared based on their probability to be generated from assumed model	
2.	Nearest neighbor	
a.	Use distance metrics to detect outliers, e.g., euclidian and mahalnobis distance measures	Branch et al. (2006), Zhang et al. (2007), Zhuang and Chen (2006)
b.	All remote data samples are declared as outliers	
3.	Clustering	
a.	Group similar & close data samples into clusters	Zhang et al. (2012), Bezdek et al. (2011), Moshtaghi et al. (2011, 2011a, 2011b), Suthaharan et al. (2010, 2010a), Rajasegarar et al. (2010a, 2012)
b.	Some distance measure from cluster center is used to detect outliers	
c.	Mahalanobis distance measure is commonly used	
4.	Classification	
a.	Learn a classification model in training phase	Bahrepour et al. (2010a), Zhang et al. (2009, 2012), Shahid et al. (2012, 2012a), Shahid and Naqvi (2011), Rajasegarar et al. (2010), Luo et al. (2006), Rajasegarar et al. (2006, 2007, 2008, 2009a), Elnahrawy and Nath (2004), Janakiram et al. (2006), Hill et al. (2007), Gomez-Verdejo et al. (2011)
b.	Classify the unseen data samples in testing phase using learned model	
c.	Support vector machine based techniques are more commonly used	

found in ([Zhang et al. 2007](#); [Shahid et al. 2012](#)). Table 1 gives a brief description of various types of outlier and event detection techniques for WSNs deployed in harsh environments and state-of-the-art examples specific to each of the types.

The authors of [Shahid et al. \(2012\)](#) also presented a set of characteristics that should be satisfied by the state-of-the-art outlier and event detection techniques. A brief description of all these characteristics has been presented in Table 2. All the characteristics presented in Table 2 have been listed down in the order of their decreasing importance to the performance of outlier and event detection techniques. This table also gives examples of various types of techniques which satisfy these characteristics.

2.1 Feasibility of state-of-the-art techniques for harsh environments

State-of-the-art outlier detection techniques for WSNs have been categorized into statistical, clustering, nearest neighbor and classification based techniques in [Shahid et al. \(2012\)](#). Further, it has been concluded in ([Shahid et al. 2012](#)) that clustering and classification based techniques are most feasible for WSNs deployed in harsh environments. We summarize the findings of [Shahid et al. \(2012\)](#) in the form of Table 3. This table assigns a priority order based on the consideration of each of the characteristic in a particular type of technique. The order has been assigned from 1 to 4, where number 1 indicates the highest priority and 4 indicates the lowest priority. For example, clustering based techniques have been assigned

Table 2 Characteristics of state-of-the-art outlier detection techniques for WSNs in harsh environments [Shahid et al. \(2012\)](#)

Characteristics and description	Examples Types	References
1 Multivariate streaming data	Statistical	Zhang et al. (2012a)
a. Attributes measured by multiple sensors should be considered together	Nearest neighbor	–
b. Continually arriving data should be processed	Classification	Shahid et al. (2012, 2012a) , Rajasegarar et al. (2010) , Zhang et al. (2009, 2009a)
	Clustering	Bahrepour et al. (2010a) , Zhang et al. (2012) , Suthaharan (2012) , Moshtaghi et al. (2011, 2011a, 2011b)
2 Temporal correlations	Statistical	Dereszynski and Dietterich (2011) , Zhang et al. (2012a) , Ross et al. (2009)
Consider correlations between newly arrived and historical data samples	Nearest neighbor	–
	Classification	Bahrepour et al. (2010a) , Shahid et al. (2012, 2012a) , Shahid and Naqvi (2011) , Rajasegarar et al. (2008, 2010) , Yozo et al. (2004) , Zhang et al. (2009, 2009a) , Gomez-Verdejo et al. (2011) , Yang et al. (2008) , Wang et al. (2006) ; Tax and Duin (1999)
	Clustering	Zhang et al. (2012) , Bezdek et al. (2011) , Moshtaghi et al. (2011, 2011a, 2011b) , Suthaharan et al. (2010, 2010a) , Rajasegarar et al. (2010a, 2012)
3 Attribute correlations	Statistical	–
Consider correlations between various attributes at a node	Nearest neighbor	–
	Classification	Shahid et al. (2012, 2012a) , Shahid and Naqvi (2011) , Ozdemir and Xiao (2011) , Hassan et al. (2011) , Krishnamachari and Iyengar (2004)
	Clustering	Zhang et al. (2012) , Bezdek et al. (2011) , Moshtaghi et al. (2011, 2011a, 2011b) , Suthaharan et al. (2010, 2010a) , Rajasegarar et al. (2010a, 2012)
4 Spatial correlations	Statistical	Dereszynski and Dietterich (2011) , Zhang et al. (2012a)
Consider correlations between data samples of geographically separated nodes	Nearest neighbor	–
	Classification	Shahid et al. (2012, 2012a) , Shahid and Naqvi (2011) , Rajasegarar et al. (2008, 2010) , Zhang et al. (2009, 2009a) , Yang et al. (2008)
	Clustering	Moshtaghi et al. (2011, 2011b) , Rajasegarar et al. (2012)

Table 2 continued

	Characteristics and description	Examples Types	References
5	In-susceptibility to temporal non-stationarity Should be robust to changes in data distribution with time	Statistical	Bettencourt et al. (2007)
		Nearest neighbor	–
		Classification	Bahrepour et al. (2010a) , Shahid et al. (2012, 2012a) , Rajasegarar et al. (2008) , Zhangh et al. (2009, 2009a)
		Clustering	Zhang et al. (2012) , Bezdek et al. (2011) , Moshtaghi et al. (2011, 2011a, 2011b) , Suthaharan et al. (2010, 2010a, 2012) , Rajasegarar et al. (2010a, 2012) , Subramaniam et al. (2006)
6	In-susceptibility to spatial non-stationarity Should be robust to spatial changes in data distribution	Statistical	–
		Nearest neighbor	–
		Classification	Bahrepour et al. (2010a) , Shahid et al. (2012, 2012a) , Rajasegarar et al. (2008, 2010) , Zhangh et al. (2009, 2009a)
		Clustering	Zhang et al. (2012) , Suthaharan (2012) , Bezdek et al. (2011) , Moshtaghi et al. (2011, 2011a, 2011b)
7	In-susceptibility to topology changes Should be robust to changes in network topology with time Schieferdecker et al. (2011)	Statistical	–
		Nearest neighbor	–
		Classification	Krishnamachari and Iyengar (2004)
		Clustering	–
8	Event / global outlier detection Should detect events or outliers common to the whole network	Statistical	Zhang et al. (2012a) , Wu et al. (2007) , Bettencourt et al. (2007)
		Nearest neighbor	–
		Classification	Bahrepour et al. (2010a) , Zhang et al. (2012) , Shahid et al. (2012, 2012a) , Rajasegarar et al. (2010)
		Clustering	Moshtaghi et al. (2011, 2011b) , Rajasegarar et al. (2012)
9	Event identification Should be able to identify the type of event, if one occurs	Statistical	–
		Nearest neighbor	–
		Classification	Abe (2010) , Liu et al. (2007, 2010) , Hao et al. (2009) , Xu (2009) , Xu et al. (2007) , Aly (2005) , Keerthi et al. (2008)
		Clustering	–

Table 2 continued

	Characteristics and description	Examples Types	References
10	Unsupervised nature	Statistical	–
	Should learn data model during training without requiring labelled input-output data	Nearest neighbor	–
		Classification	Bahrepour et al. (2010a), Shahid et al. (2012, 2012a), Rajasegarar et al. (2010), Zhang et al. (Zhang et al. (2009), Zhang et al. (2009a))
		Clustering	Zhang et al. (2012), Bezdek et al. (2011), Moshtaghi et al. (2011, 2011a, 2011b), Suthaharan et al. (2010, 2010a, 2012), Subramaniam et al. (2006), Rajasegarar et al. (2008, 2010a, 2012)
11	Distributed processing	Statistical	Dereszynski and Dietterich (2011), Zhang et al. (2012a)
	Should be able to distribute processing equally on all nodes of network	Nearest neighbor	Branch et al. (2006), Zhang et al. (2007), Zhuang and Chen (2006)
		Classification	Shahid et al. (2012, 2012a), Shahid and Naqvi (2011), Rajasegarar et al. (2006, 2007, 2008, 2010), Elnahrawy and Nath (2004), Janakiram et al. (2006), Hill et al. (2007), Zhang et al. (2009, 2009a)
		Clustering	Zhang et al. (2012), Bezdek et al. (2011), Moshtaghi et al. (2011, 2011a, 2011b), Suthaharan et al. (2010, 2010a, 2012), Rajasegarar et al. (2012, 2010a), Subramaniam et al. (2006)
12	Online data processing	Statistical	Dereszynski and Dietterich (2011), Zhang et al. (2012a)
	Should process the streaming data as soon as it arrives	Nearest neighbor	–
		Classification	Shahid et al. (2012, 2012a), Shahid and Naqvi (2011), Rajasegarar et al. (2008, 2010) Zhang et al. (2009, 2009a), Yang et al. (2008)
		Clustering	Moshtaghi et al. (2011, 2011b), Rajasegarar et al. (2012)
13	Low complexity	Statistical	Dereszynski and Dietterich (2011), Zhang et al. (2012a), Sheng et al. (2007), Wu et al. (2007), Yozo et al. (2004), Jun et al. (2005), Bettencourt et al. (2007), Palpanas et al. (2003), Subramaniam et al. (2006)
	Should not require excessive computations & communication among nodes of network	Nearest neighbor	–

Table 2 continued

Characteristics and description	Examples Types	References
	Classification	Bahrepour et al. (2010a), Shahid et al. (2012, 2012a), Rajasegarar et al. (2010), Zhang et al. (2009, 2009a)
	Clustering	Zhang et al. (2012), Suthaharan (2012), Giatrakos et al. (2010), Bezdek et al. (2011), Moshtaghi et al. (2011, 2011b), Rajasegarar et al. (2012)

These characteristics have been listed down in the order of their decreasing importance for harsh environments as discussed in Shahid et al. (2012)

Table 3 Feasibility of state-of-the-art techniques for harsh environments

	Characteristics	Clustering	Classification	Nearest neighbor	Statistical
1	Multivariate data	1	1	1	1
2	Streaming data	1	2	3	4
3	Temporal correlations	1	1	1	1
4	Attribute correlations	1	2	3	4
5	Spatial correlations	1	2	3	4
6	In-susceptibility to Temporal non-stationarity	1	2	3	3
7	In-susceptibility to Spatial non-stationarity	1	2	3	3
8	In-susceptibility to Dynamic topology	–	1	–	–
9	Event / global outliers detection	1	1	2	2
10	Event identification	1	–	–	–
11	Unsupervised	1	2	3	4
12	Distributed approach	1	2	3	4
13	Online approach	1	2	3	3
14	Low computational complexity	1	2	4	3
15	Low communication complexity	2	3	4	1
16	Threshold dependence	1	1	1	1

A priority order from 1 (highest) to 4 (lowest) has been assigned to each of the types of techniques depending on their tendency to incorporate the particular characteristic. ‘–’ sign indicates that the technique does not incorporate the characteristic at all. Some of the techniques have been assigned the same priority because they have equal tendency to incorporate the characteristic

a priority of 1 for attribute correlations, which means that clustering based techniques have the highest tendency to consider attribute correlations, whereas the nearest neighbor based techniques have been assigned a priority 4 because they have the lowest tendency to incorporate attribute correlations. ‘–’ sign indicates that the technique does not incorporate the characteristic at all. Some of the techniques have been assigned same priority because they have equal tendencies to incorporate the characteristic.

From Table 3 it is clear that clustering based techniques have the highest priority order (1) for a majority of characteristics. Then come the classification based techniques which have priority orders 1 and 2 for most of the characteristics. Nearest neighbor and statistical based techniques have the lowest priority order (3 and 4) for most of the characteristics. Based on this table and the analysis carried out in [Shahid et al. \(2012\)](#) we arrive at the following increasing order of the techniques for their feasibility in harsh environments.

Nearest Neighbor \approx Statistical < Classification < Clustering

2.2 What is the rest of survey about?

Up till now, we have defined WSNs, harsh environments and explained the concept of outlier and event detection in harsh environments. Further, the characteristics of outlier and event detection techniques for WSNs in harsh environments have been discussed briefly and the feasibility of various techniques has been determined for harsh environments. It has also been concluded that clustering and classification based techniques are most suitable for harsh environments. The above discussion has been derived from [Shahid et al. \(2012\)](#). In the later sections, we give in-depth analysis of state-of-the-art classification based techniques and their feasibility for harsh environments. Rigorous critical analysis has been carried out to determine the feasibility of all techniques in terms of the discussed characteristics. The article ends with future research directions for making state-of-the-art classification based techniques suitable for harsh environments.

Before starting a detailed critical analysis of state-of-the-art classification based techniques, we present a few methods that had been frequently used in literature for measuring the performance of these techniques. We also formulate our methodology for analysis of these techniques which is used in this survey article.

3 Methodology for analysis of classification based techniques

The most commonly used method to analyze the performance of a classification algorithm is to calculate some quantitative measures by testing on various types of data sets and input parameters. This method has commonly been followed by research community for presenting state-of-the-art techniques in various application domains. However, an inherent drawback of this procedure is that various algorithms may outperform other similar algorithms due to difference of experimental setup and conditions. Consider, for instance, the example of an outlier detection algorithm being tested in a harsh environment, such as an underground mine. The same algorithm tested in a underground coal mine may show poor performance as compared to the scenario when it is tested in salt mines due to difference in the nature of two environments. Thus, quantitative measures are not reliable for fair comparison of two techniques meant for the same purpose. To perform a fair comparison of various state-of-the-art outlier and event detection techniques, we introduce *qualitative measures* which use a set of well-defined characteristics to determine the quality of an outlier and event detection technique. Various techniques are compared based on these measures.

Thus, two types of measures can be used to evaluate the performance of outlier and event detection techniques:

- quantitative
- qualitative

3.1 Quantitative performance measures

Three commonly used measures for the performance analysis of outlier and event detection techniques are given below (Abe 2010; Lazarevic et al. 2003; Ganguly 2008):

1. Detection Rate
2. False Alarm Rate / False Positive Rate
3. Receiver Operating Characteristic (ROC) Curve

3.1.1 Detection rate

The ‘detection rate’ of an outlier and event detection technique is defined as the fraction of outlying data that are detected as outliers by the application of outlier detection algorithm. The detection rate may also be described separately as an outlier detection rate or as event detection rate if the technique is able to detect outliers as well as events. Ideally the detection rates should be close to unity. As described earlier, WSNs deployed in harsh environments may suffer from frequent events, thus, it becomes essential for a technique to have a high outlier and event detection rate if it is to be used in harsh environments. A technique with a high outlier and low event detection rate is not suitable for harsh environments. Further, a high outlier and low event detection rate implies that the algorithm is unable to differentiate between various sources of outlier.

3.1.2 False positive rate

Although, a very high detection rate is extremely essential for techniques used in harsh environments, however, it does not imply that a trade-off can be made to allow the normal data to be classified as outliers or events. Thus, an optimal outlier detection technique should not make false detections, that is, normal data should not be classified as outliers or events. The fraction of normal data detected as outliers is known as false positive rate. Ideally, the false positive rate should be as low as possible (close to zero). Similar to the detection rate, false positive rate can also be categorized into outlier and event false positive rates, where the former will provide a measure of normal data samples classified as outliers and later will provide a measure for normal data samples classified as events.

3.1.3 Receiver operating characteristics

The relationship between *detection rate* and *false positive rate* of an outlier and event detection scheme is given by a *receiver operating characteristic curve (ROC)*. An ROC curve is a plot that indicates the variation of the detection rate of a classifier with respect to false positive rate. Ideally, the area under ROC curve should be very close to unity. Thus, for an optimal outlier and event detection technique, the detection rate should be very close to unity even when the false positive rate is very low. If the area under ROC curve for a classifier is less than 0.5, the classifier is just based on a random decision.

3.1.4 Measures for event identification

The detection and false positive rates can only give information about the fraction of outlying data classified as outliers or about the fraction of normal data declared as outliers. Thus, these two measures can be used for detection problems only when there is one target class

(normal for WSN applications). However, in case of data belonging to multiple classes, a more accurate measure is required which takes into account the inter-class miss-classification rates. Such a problem arises in the event identification process in WSNs, where the rate of the occurrence of an event is not the only parameter of key importance. The type of event and the attributes involved in an event should be identified in such applications. For example, in case of occurrence of an event in underground mine, analysis should be carried out to determine if the event is a fire or earthquake. This process requires some additional measures to be associated for event detection and identification algorithms. Following parameters are commonly used for analyzing multi-class problems:

1. Recognition rate
2. Error rate

Suppose that an event identification problem involves determining ‘n’ possible types of an event, then the recognition and error rates can be calculated by formulating an $n \times n$ confusion matrix whose element a_{ij} is the number of class i data classified into class j . Then the recognition rate and error rates can be defined as (Abe 2010):

$$\text{Recognition (R)} = \frac{\sum_{i=1}^n a_{ii}}{\sum_{i,j=1}^n a_{ij}} \times 100(\%)$$

$$\text{Error (E)} = \frac{\sum_{i \neq j, i,j=1}^n a_{ij}}{\sum_{i,j=1}^n a_{ij}} \times 100(\%)$$

where $R + E = 100\%$. Thus, the recognition rate (R) should be close to 100% and the error rate (E) should be as low as possible.

3.2 Qualitative performance measures - THE FEASIBILITY ANALYSIS TABLE

The characteristics for optimal outlier and event detection techniques presented in Sect. 2 and described in detail in Shahid et al. (2012) can be used to formulate a set of qualitative measures for the performance evaluation of these techniques. More specifically, these characteristics can be used to determine the feasibility of a technique for harsh environments. Thus, to determine the feasibility of any state-of-the-art outlier and event detection technique for WSNs in harsh environments we formulate a feasibility analysis table (Table 6) that summarizes the analysis of a technique in terms of the presence or absence of all important characteristic and then concludes with one of the following decisions:

- The technique is feasible for use in harsh environments.
- The technique can be used in harsh environments after some modifications.
- The technique cannot be used in harsh environments.

Tables 6 and 7 in Sect. 6 show sample feasibility analysis tables that have been used to qualitatively analyze a technique and then determine its feasibility for harsh environments. The left most column of these tables list all the characteristics that have been discussed in Shahid et al. (2012). The characteristics have been further divided into sub-categories which may be used to analyze a technique in more detail. For example, the first characteristic ‘input data’ has been divided into two categories a) data type and b) correlations. The ‘data type’ specifies various data types that the algorithm is able to incorporate (univariate, streaming, sequential etc.) and correlations indicate the types of correlations (temporal, attribute and spatial) that the algorithm is able to incorporate. Similar divisions can be seen throughout the tables. The rightmost columns of these tables indicate the presence of a characteristics by

a ‘Y’ sign, whereas, the absence of a characteristic is indicated by an empty space. In some cases a ‘?’ may also be seen which shows that the technique can incorporate the characteristic after some minor modifications. The communication and computational complexities have been described as ‘high’ or ‘low’ in the tables.

The bottom left of Tables 6 and 7 indicate decisions for the feasibility of a technique for harsh environments. It has been described in Shahid et al. (2012) that a certain set of characteristics has a higher priority for harsh environments than others. The decision on feasibility has been made on the basis of the presence or absence of various characteristics in a technique and their priority for harsh environments. We first divide the entire set of characteristics (Shahid et al. 2012) into two sets, namely *ESSENTIAL*, *NON-ESSENTIAL*:

- **ESSENTIAL:** Multivariate streaming data, spatio-temporal-attribute correlations, insusceptibility to temporal and spatial non-stationarity, insusceptibility to topology changes, event detection and unsupervised nature.
- **NON-ESSENTIAL:** distributed processing, online computations, low complexity.

Now, we formulate the decision for the feasibility of a technique for harsh environments as following:

- **FEASIBLE:** ESSENTIAL & NON-ESSENTIAL are satisfied.
- **INFEASIBLE:** ESSENTIAL characteristics are not satisfied.
- **FEASIBLE AFTER MODIFICATIONS:** ESSENTIAL satisfied, NON-ESSENTIAL not satisfied.

4 Classification based outlier and event detection for WSNs deployed in harsh environments

Classification based techniques have a significant importance in data mining and machine learning community. These techniques tend to learn a classification model during the training phase and classify the unseen data instances during the testing phase. Labeled input-output pairs may be fed to the algorithms during training phase (Stankovic et al. 2012; Ganguly 2008; Bishop 2006). The performance of a classifier depends on its ability to classify the unseen data based on the learned model and is more generally known as its *generalization ability* (Nguyen et al. 2008; Yeung et al. 2007; Steinwart and Christmann 2008; Tax and Duin 2004). Depending on the type of model learned during the training phase, these techniques can be divided into two types:

1. Bayesian based
2. Support vector machine based (SVM)

4.1 Bayesian based approaches

Bayesian based approaches use a probabilistic graphical model to represent a set of variables and their probabilistic dependencies. The learned graphical model can have various nodes connected by edges representing their dependencies on each other. Based on the learned graphical model, these techniques can determine the probability of a newly arrived data sample belonging to a particular class (Zhang et al. 2010). An advantage of these techniques is that they assign a probability measure associated with every newly arrived data sample; hence, these techniques can be used to gain some information about the gradually changing data distributions. Further, these techniques are known as *soft-decision* techniques because

they give information about the probability of data samples belonging to outlying class rather than a hard decision. Despite their numerous advantages, these techniques have not been found to be suitable for WSNs due to following reasons:

- Require significant computational overheads to learn a probabilistic graphical model during the training phase.
- The graphical model may have to be learned time and again due to dynamic nature of data distributions in harsh environments.
- They are not online and adaptive. Thus, a large amount of training data is required initially to learn a model and the training phase is repeated for changing distributions.
- Naive bayes assumption can be used to reduce the complexity of the learning phase, but this assumption also leads to a loss of correlations between various attributes ultimately leading to a reduction in the generalization ability.

4.2 Support vector machine (SVM) based approaches

Support Vector Machine (SVM) based techniques are another type of classification based techniques which learn a classification model during the training phase and then use the learned model to classify unseen data. However, unlike Bayesian based techniques, these techniques have a better generalization ability because they tend to maximize the separation between different classes by making use of mercer kernels. For this reason they are also known as *maximum margin classifiers* (Abe 2010; Scholkopf and Smola 2001; Herbrich 2002). In general, an SVM based technique tends to learn a maximum margin hyperplane between different data classes during training phase. A maximum margin hyperplane is one which has maximum separation from data belonging to each of the target classes. The separation is attained by constructing hyperplanes in input space, if the data is separable, or by mapping the whole data to a higher dimensional feature space by using mercer kernels and then constructing a hyperplane in feature space. The computational complexity of these techniques is less than that of bayesian based techniques because they do not learn the whole graphical model. In fact a hyperplane only requires the determination of two parameters, that is, a weight vector (that determines its slope) and a bias parameter. Depending on the number of target classes, SVM based techniques can be classified as (Tax and Duin 1999; Aly 2005; Bishop 2006; Nguyen et al. 2008; Yeung et al. 2007; Steinwart and Christmann 2008; Tax and Duin 2004; Scholkopf and Smola 2001; Herbrich 2002; Weston and Watkins 1998; Smola and Schölkopf 1998; Weston and Watkins 1999; Mayoraz and Alpaydin 1999; Bredensteiner and Bennett 1999; Schwenker 2000; Schölkopf et al. 2001; Hsu and Lin 2002; Franc and Hlavác 2002; Elisseeff and Weston 2002; Zhu et al. 2003; Ambwani 2003; Cheong et al. 2004; Kecman 2005; Zhang et al. 2006; Navia-Vazquez et al. 2006; Hofmann et al. 2008):

1. Multi-class SVM
2. Binary class SVM
3. One-class SVM

4.3 One-class SVMs for outlier and event detection

Outlier and event detection in WSNs require a model of normal data. All data samples which do not fit the normal data model are declared to be outliers. As the problem of outlier detection requires the model of normal data only so it is a one-class problem. Hence, One-Class SVMs can be used for outlier and event detection in WSNs. In general, one-class SVMs tend to learn a normal data model by enclosing majority of normal data in a geometric figure, where all

the data lying outside the geometric figure are declared to be outliers (Abe 2010). Following are a few advantages of one-class SVM.

- They have a simple geometric representation, that is, they determine a few parameters of the normal data model only. For example, if the normal data is enclosed in a sphere, then one-class SVM tends to determine the center and radius of the sphere only.
- They provide an optimum solution for outlier and event detection by maximizing the separation between normal and abnormal data (Yang et al. 2008).
- Pre-classified abnormal data is difficult to obtain in WSNs, therefore, an unsupervised technique should be used for outlier detection in WSNs. One-class SVMs are unsupervised and can be used for this purpose.
- Provide a sparse solution, that is, only a small number of data samples are only used to formulate the decision boundary between normal and abnormal class.

5 Introduction to one-class SVM formulations

A One-Class SVM formulation defines a boundary (such as a hyper-plane or a hyper-sphere) to separate normal data from anomalous data in a high dimensional space by maximizing the distance between normal and anomalous data. Various formulations for One-Class SVM have been proposed in the literature.

- Hyperplane
- Hypersphere (SOCC)
- Quarter-sphere (QS-SVM)
- Hyperellipsoid (TOCC)
- Centered ellipsoid (CESVM)

5.1 Terminologies and notations used in one-class SVMs

This portion of the paper introduces the reader to a set of terminologies and notations that are commonly encountered in one-class SVM problems. Information presented in this section can help in a better understanding and analysis of SVM formulations discussed in the forthcoming sections.

5.1.1 Classifier

A classifier is a name given to an algorithm or technique that achieves optimal separation of data belonging to different classes. A one-class classifier should achieve a separation between normal and abnormal data samples. A classifier can be described by its model parameters. The quality of a classifier is described in terms of its generalization ability.

5.1.2 Model parameters

Model parameters for one-class SVMs (classifiers) are defined as the parameters used to describe the formulations. A hyperplane formulation is defined by its weight vector w (which determines its slope) and its bias parameter r (which determines its distance from the origin). Hypersphere formulation is defined by its radius R and center c , whereas, a hyperellipsoid formulation, which has more than one clusters is described by the center of clusters c_i and their radii r_i . The quarters-sphere and centered ellipsoid formulations are defined by their radii R . The model parameters are determined by solving some optimization problem.

5.1.3 Training and testing phases

The model parameters of any classifier can be determined by using some example data samples. This process is known as the training phase of a classifier. One-class classifiers are unsupervised because they only require input data samples without their labelling into normal or abnormal samples. Multi-class classifiers are supervised because they require labelled input-output pairs during the training phase. The performance of a classifier can be analyzed by giving unseen data samples at the input during the testing phase.

5.1.4 Generalization ability

The performance of a classifier is tested during the testing phase during which unseen data samples are input to the classifier. The performance during testing phase depends on how well the parameters learned during the training phase are able to perform a separation of unseen data samples. This is known as the generalization ability of a classifier. The generalization ability of all SVM based classifiers is improved by maximizing the margin between normal and abnormal samples.

5.1.5 Maximum margin

Ideally, a one-class classifier should be able to classify the normal and abnormal samples correctly during the testing phase. This is achieved by using the concept of maximum margin between normal and abnormal data classes. The parameters of one-class formulation under consideration are determined in such a way that the separation between normal and abnormal data samples is maximized. This is known as the concept of maximum margin and is guaranteed to improve the generalization ability of SVM based classifiers over other non-SVM based classifiers (Elisseeff and Weston 2002). The parameters corresponding to maximum margin can be determined by solving some optimization problem under some constraints on data samples that ensure maximum separation between the two classes.

5.1.6 Feature space

Let $x \in \chi$ be the data samples and data set defined in the input space, then, each of the data samples $x \in \chi$ can be mapped to a higher dimensional space F by a non-linear mapping function $\phi(x)$. The mapping can be defined as $\phi : \chi \rightarrow F$. If the normal and abnormal data are inseparable in input space, then, this mapping can be used to project the data to a higher dimensional space, where the normal data samples can be easily separated from abnormal samples. The separation in feature space is achieved by formulating the one-class SVM problem in feature space itself (Abe 2010).

5.1.7 Mercer kernel functions

In general, one-class SVMs do not require an explicit treatment of data samples in input or feature space because they involve the computation of norm of vectors in either space and use these norms to detect outliers or events. Thus, mapping from input space to feature space can be carried out without an explicit information about the feature space with the help of Mercer kernel functions. Mercer kernels are symmetric positive semi-definite functions which compute the dot product of vectors in feature space and can be defined as

$K(x, z) = (\phi(x) \cdot \phi(z))$ (Hofmann et al. 2008). Various types of kernel functions can be used for this purpose, some of which are linear, polynomial, gaussian and mahalanobis kernels (Hofmann et al. 2008).

5.1.8 Primal optimization problem

Model parameters for any one-class SVM formulation are usually determined by solving optimization problem. The optimization problem involves minimizing or maximizing some convex objective function under some constraints. The objective function is usually a function of one or more primal variables. A primal variable can be defined as a variable which is directly involved in the determination of the model parameters for one-class SVM. For example, for the hyperplane based formulation, the norm of weight vector (which is a hyperplane parameter) is minimized under some constraints (Boyd and Vandenberghe 2004).

5.1.9 Dual optimization problem and Lagrange function

The primal optimization problem formulated for one-class SVMs can involve more number of variables than the number of constraints, thus, making it difficult to solve. This problem can be alleviated by transforming the primal problem to a dual problem with lesser and low dimensional variables that can be solved using quadratic and linear programming techniques. An unconstrained lagrange function is formed for this purpose which consists of the objective function and constraints weighted by lagrange variables. The lagrange function (also known as lagrangian) is then minimized with respect to the primal variables to obtain some expression in terms of dual lagrange variables only (Boyd and Vandenberghe 2004).

5.1.10 Regularization parameter and slack variables

The model complexity for one-class SVMs, that is, the time taken by optimization problem to return an optimal solution depends on how easily can the normal and abnormal data be separated. Thus, in order to reduce the complexity of learned model and over-fitting caused due to it, some miss-classifications are allowed in the learned model. This is achieved by introducing a regularization term whose weight is controlled by a regularization parameter v . If the cost of miss-classification for each of the data samples is represented by some slack variable ξ_i , then the total miss-classification cost is the sum of all slack variables. The total cost of miss-classification is controlled by weight of $1/(vn)$. If the cost is low then a large number of data samples are allowed to be miss-classified which results in a high detection rate and false positive rate of the classifier. If miss-classification cost is high then the data samples are not allowed to be miss-classified and this results in a low detection rate and low false positive rate of the classifier (Abe 2010).

5.1.11 Support and non-support vectors

The solution of dual optimization problem for one-class SVMs results in the determination of the values of lagrange multipliers for each of the data samples. Depending on these values, the data samples / vectors may be classified as normal or abnormal. The data samples which lie inside the geometric figure are normal and their associated lagrange multipliers have a value of zero. All such vectors are known as non-support vectors. The data samples which lie on or outside the geometric formulation are known as support vectors. Out of the support

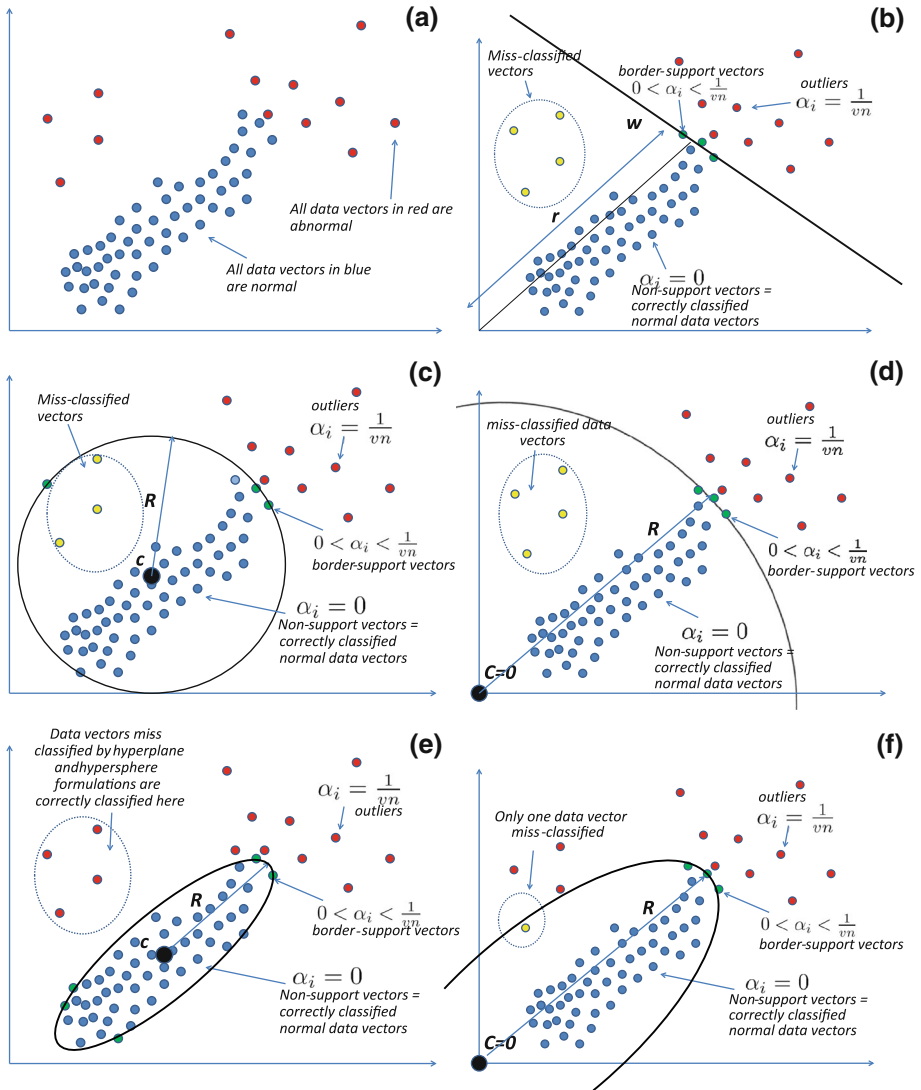


Fig. 1 **a** An example 2D (2-attribute) data set. The *blue* points represent normal data vectors and *red* points indicate abnormal data vectors. **b** Hyperplane formulation of one-class SVM over data set of (a). **c** Hyper-sphere formulation of one-class SVM. **d** Quarter-sphere formulation of one-class SVM. **e** Hyper-ellipsoidal formulation of one-class SVM. **f** Centered ellipsoid formulation of one-class SVM. In all figures, *red*, *blue* and *green* points indicate the data vectors classified as abnormal, normal and support respectively. The *yellow* points indicate the data vectors which have been miss-classified by the formulation. All normal data vectors have a zero value for their associated lagrange multipliers $\alpha_i = 0$. All the data vectors classified as abnormal have their lagrange multiplier values $\alpha_i = 1/vn$. The support vectors have lagrange multiplier values $0 < \alpha_i < 1/vn$

vectors, those that lie exactly on the geometric formulation are known as border support vectors. These vectors define the boundary of one-class SVM formulation (Abe 2010). The support vectors which lie outside the SVM formulation belong to the abnormal class and are also known as non-border support vectors. Classification of data vectors into support and non-support vectors for various one-class SVM formulations is shown in Fig. 1b–f.

Table 4 Notations for SVM formulations

Type	Notations	Explanation
General	n	Size of input data
	l	Dimensions of input data
	$x \in \chi$	Input data set and input space
	$\phi : \chi \rightarrow F$	Mapping to feature space
	$y \in Y$	Output and output space
	$v \in (0, 1)$	Regularization parameter
	α	Lagrange multipliers
	ξ	Slack variables
	Σ_i	Covariance matrix in input space χ
	σ	Covariance matrix in feature space F
	Λ	Positive diagonal matrix
	$P \Lambda P^T$	Eigen decomposition
	$K(x, z)$	Kernel ($\phi(x) \cdot \phi(z)$)
L	Dual Lagrangian	
Hyperplane	r	Distance of hyperplane from origin
	w	weight vector
Hyper-sphere, Hyperellipsoid, Quarter-sphere &	R	radius of hypersphere / quarter-sphere / centered ellipsoid
	r_i	Radius of i th cluster
Centered ellipsoid	c	Center of ball with radius R
	c_i	Center of i th cluster

5.1.12 Euclidean distance

The euclidean distance of any data sample x from a point c is defined as:

$$d(x) = \sqrt{(x - c)^T(x - c)} \tag{1}$$

Suppose that the data collected from a WSN node is enclosed in a ball of radius R and center c . Then, the euclidean distance of any data sample x from center c denotes the distance of x from the center of distribution, assuming a spherical data distribution.

5.1.13 Mahalanobis distance

The mahalanobis distance of a data sample x from a point c is defined as:

$$d(x) = \sqrt{(x - c)^T \Sigma^{-1}(x - c)} \tag{2}$$

The mahalanobis distance measure takes into account the covariance matrix of data. Thus, like euclidean distance, a spherical distribution of data is not assumed. An important property of this measure is that it is unaffected by the absolute values of attributes and takes into account the deviation of data sample from the distribution of each of the attributes.

Table 4 summarizes the notations used in the SVM formulations.

5.2 Description of one-class SVM formulations

5.2.1 Hyper-plane formulation

The earliest known formulation of one-class SVM is the hyper-plane (Schölkopf et al. 2001), as shown in Fig. 1b. The hyper-plane based SVM constructs a maximal margin hyper-plane to separate the normal and anomalous data from the origin. The hyperplane is characterized by a weight vector w and a bias parameter r . Following quadratic optimization program which involves the minimization of the norm of weight vector is solved:

$$\min_{w \in F, \xi \in \mathbb{R}^l, r \in \mathbb{R}} \frac{1}{2} \|w\|^2 + \frac{1}{vn} \sum_{i=1}^n \xi_i - r \tag{3}$$

$$\text{subject to : } (w \cdot \phi(x_i)) \geq r - \xi_i, \xi_i \geq 0 \tag{4}$$

The weight vector w , characterizing the hyperplane, lives in the feature space F , and therefore it is not directly accessible (as the feature space may be extremely high-dimensional). The term on the right hand side of above equation is the regularization term which controls the cost of miss-classification. This term comprises of non-negative slack variables ξ , which allow for some of the anomalies to lie on the wrong side of the hyperplane. The regularization parameter v controls the fraction of data that may be classified as outliers. The overall problem tends to reduce the distance of hyperplane from the origin to determine optimal w and r in such a way that a few miss-classifications are allowed. Instead of the primal problem stated above, the following dual problem, in which all the variables have low dimensions, is solved in practice

$$\min_{\alpha \in \mathbb{R}^l} \sum_{ij} \alpha_i \alpha_j k(x_i, x_j) \tag{5}$$

$$\text{subject to: } \sum_{i=1}^n \alpha_i = 1 \tag{6}$$

$$0 \leq \alpha_i \leq \frac{1}{vn} \tag{7}$$

In the above dual problem, α are the lagrange multipliers and $k(x_i, x_j)$ is the dot product of vectors x_i and x_j in the feature space. This dual problem is obtained by minimizing the lagrangian with respect to primal variables. A detailed derivation of this problem can be found in Schölkopf et al. (2001). The hyperplane formulation for an example synthetic data set (Fig. 1a) is shown in Fig. 1b.

5.2.2 Hyper-sphere formulation (SOCC)

One can also use balls to describe the data in feature space (Tax and Duin 1999), as shown in Fig. 1c. Hyper-sphere based SVM aims at separating the normal and anomalous data by enclosing the normal data in a ball of radius R and centre c in a higher dimensional feature space F . This is attained by solving the following quadratic optimization problem.

$$\min_{R \in \mathbb{R}, \xi \in \mathbb{R}^l, c \in F} R^2 + \frac{1}{vn} \sum_{i=1}^n \xi_i \tag{8}$$

$$\|\phi(x_i) - c\|^2 \leq R^2 + \xi_i, \xi_i \geq 0 \tag{9}$$

The term on the right hand side of above equation is the regularization term that controls the miss-classification cost as described for the hyperplane formulation. The above primal problem tends to reduce the radius R of hypersphere and determine R and center c in such a way that the euclidean distance of a majority of data samples is comparable to the radius R and only a few miss-classifications are allowed. The dual of the above primal problem is:

$$\min_{\alpha} \sum_{i,j} \alpha_i \alpha_j k(x_i, x_j) - \sum_i \alpha_i k(x_i, x_i) \tag{10}$$

$$\text{subject to: } 0 \leq \alpha_i \leq \frac{1}{vn}, \sum_i \alpha_i = 1 \tag{11}$$

where α are the lagrange multipliers and $k(x_i, x_i)$ is the dot product of vectors in feature space. The hypersphere formulation for an example synthetic data set (Fig. 1a) is shown in Fig. 1c.

5.2.3 Quarter-sphere formulation (QS-SVM)

Laskov et al. (2004) proposed to extend the ideas of one-class SVM to one-sided non-negative data. He proposed to fix the center of the fitted sphere in SOCC at the origin, as shown in Fig. 1d. Thus, the problem is reduced to fitting the normal data in a quarter-sphere. This problem, also known as QS-SVM, can be formulated as following:

$$\min_{R \in \mathbb{R}, \xi \in \mathbb{R}^l} R^2 + \frac{1}{vn} \sum_{i=1}^n \xi_i \tag{12}$$

$$\|\phi(x_i)\|^2 \leq R^2 + \xi_i, \xi_i \geq 0 \tag{13}$$

The corresponding dual problem is:

$$\min_{\alpha} - \sum_i \alpha_i k(x_i, x_i) \tag{14}$$

$$\text{subject to: } 0 \leq \alpha_i \leq \frac{1}{vn}, \sum_i \alpha_i = 1 \tag{15}$$

An important fact about quarter-sphere SVM formulation is that its dual problem involves only one set of lagrange multiplier values as compared to hypersphere formulation which requires two sets. This can be attributed to the removal of a primal variable c from the hypersphere problem. Thus, the complexity of quarter-sphere based problem is less as compared to the hyper-sphere problem. The quarter-sphere formulation for an example synthetic data set (Fig. 1a) is shown in Fig. 1d.

5.2.4 Hyper-ellipsoidal formulation (TOCC)

The spherical one-class classifier (SOCC) finds a hypersphere with minimum volume that contains the target data while keeping outlier samples outside. SOCC achieves satisfactory performance only when target samples have same distribution tendency in all orientations. Therefore, performance of SOCC is limited in a way that many superfluous outliers might be mistakenly enclosed. The authors in Wang et al. (2006) have formulated the one-class SVM using hyperellipsoids with minimum effective radii around a majority of the image vectors in the feature space. This hyperellipsoidal formulation involves two phases. First the image

vectors are partitioned into a number of distinct clusters using Ward’s linkage algorithm. Second, the image vectors in each cluster are fixed with a hyperellipsoid that encapsulates a majority of the image vectors in that cluster. The image vectors that do not fall within any of the hyperellipsoids are identified as anomalous. This formulation is also called structured one-class classification (TOCC). The optimization problem in TOCC can be formulated as a series of second-order cone programming problems that can be solved with acceptable efficiency by primal-dual interior-point methods.

$$\min r_i + \frac{1}{vn} \sum_{j=1}^n \xi_j \tag{16}$$

$$\text{s.t.}:\|\Sigma_i^{-\frac{1}{2}}(x_j - c_i)\| \leq r_i + \xi_j, \xi_j \geq 0 \tag{17}$$

The overall problem tends to enclose majority of normal data (mapped to feature space) in one or more clusters in such a way that a few miss-classifications are allowed only. An important fact about hyper-ellipsoidal formulation is that it uses mahalanobis distance of data samples from the center of clusters to detect outliers. Unlike, sphere structured SVMs, which use euclidean distance, the mahalanobis distance ensures a better generalization ability of the classifier. The hyperellipsoid formulation for an example synthetic data set (Fig. 1a) is shown in Fig. 1e.

5.2.5 Centered ellipsoidal formulation (CESVM)

Based on the work presented in Rajasegarar et al. (2007, 2008) presented a centered hyper-ellipsoidal formulation for one-class SVM (CESVM). The aim is to fit a hyperellipsoid in the feature space with minimum effective radius $R(> 0)$, centered at origin, encompassing a majority of the image vectors χ , as shown in Fig. 1f. Like the quarter-sphere problem, this problem can also be formulated as a linear optimisation problem because it requires the determination of a radius only. The primal problem is formulated as.

$$\min_{R \in \mathbb{R}, \xi \in \mathbb{R}^n} R^2 + \frac{1}{vn} \sum_{i=1}^n \xi_i \tag{18}$$

$$\|\phi(x_i)\| \sigma^{-1} \|\phi(x_i)\| \leq R^2 + \xi_i, \xi_i \geq 0 \tag{19}$$

The dual optimization problem is.

$$\min_{\alpha \in \mathbb{R}^n} - \sum_{i=1}^n \alpha_i \|\sqrt{n} \Lambda^{-1} P^T K_c^i\| \tag{20}$$

$$\text{subject to: } \sum_{i=1}^n \alpha_i, 0 \leq \alpha_i \leq \frac{1}{vn} \tag{21}$$

The centered ellipsoid formulation for an example synthetic data set (Fig. 1a) is shown in Fig. 1f.

5.3 Relationship between various one-class SVM formulations

Various one-class SVM formulations discussed in this paper are interlinked with each other. In fact, hyper-ellipsoid based formulation is the most generalized one-class formulation and

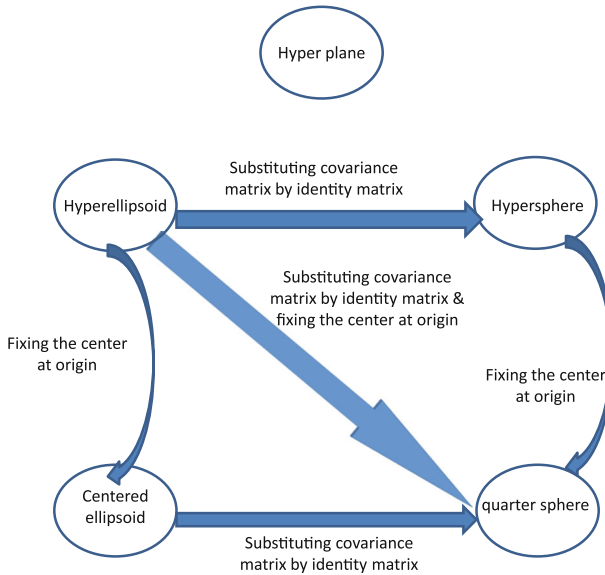


Fig. 2 Relationship between various one-class SVM formulations

various other formulations can be derived from it. The relationship between these formulations is described in the following discussion.

- The hyper-ellipsoid based formulation is generalized form all one-class SVM formulations. A hyper-ellipsoid can be defined by a radius R and centre c . The shape of an ellipsoid can vary from an ellipse to a hypersphere depending on the nature of data distribution along various attributes. The covariance matrix involved in this formulation takes into account the nature of data distribution. Replacing the covariance matrix by identity matrix in Eq. (16) results in the hypersphere formulation of Eq. (8).
- The centered hyperellipsoid formulation of Eq. (18) is also a derivative of the hyperellipsoid formulation and can be obtained by fixing the centre of the hyperellipsoid at the origin, that is, placing $c = 0$ in Eq. (16).
- The quarter-sphere formulation of Eq. (12) is also a derivative of hyperellipsoid and hypersphere based formulations. In fact, QS-SVM formulation can be obtained by fixing the centre of hypersphere at the origin, that is, putting $c = 0$ in Eq. (8). The quarter-sphere formulation can also be obtained from hyperellipsoid formulation by replacing the covariance matrix in Eq. (16) with an identity matrix and putting $c = 0$.
- The hyperplane based formulation is a special SVM formulation which cannot be obtained from other SVM based formulations.

The relationship between one-class SVM formulations is summarized in Fig. 2.

5.4 Comparison of One-class SVM formulations

One-class formulations tend to separate normal and abnormal data by enclosing normal data in a geometric figure such as a hyperplane, hypersphere, hyperellipsoid, quarter-sphere and centered ellipsoid. This section presents a comparison of one-class formulations in terms of model complexity, optimization problem, distance measure used and generalization ability.

5.4.1 Optimization problem and associated complexity

An optimization problem is solved for learning model parameters during training phase, whose complexity depends on the number of parameters to be determined. A hyperplane in feature space is defined by its weight vector w and bias parameter r , therefore, it requires the solution of a quadratic optimization problem. Similarly, hypersphere and hyperellipsoid are also defined by two parameters, that is, their radius and center and require the solution of quadratic optimization problems. Quarter-sphere and centered ellipsoid are only defined by their radii, so they involve a linear optimization problem. Thus, quarter-sphere and centered ellipsoid based techniques are computationally inexpensive and more feasible for implementation on WSNs.

5.4.2 Distance measure

A distance measure is used in all one-class SVM formulations to detect outliers. Hyperplane, hypersphere and quarter-sphere based formulations use euclidean distance of data samples from the center of distribution to detect outliers, whereas, hyperellipsoid and centered ellipsoid use mahalanobis distance to detect outliers. We will shortly prove that classifiers based on mahalanobis distance achieve better generalization as compared to those based on euclidean distance.

5.4.3 Generalization ability and classification performance

Euclidean distance of any data sample from the center of distribution assumes that the data is equally distributed in all directions. Thus, a spherical structure of data is assumed in the classifiers based on euclidean distance. Moreover, this distance measure is highly dependent on the difference among the absolute values of attributes. Mahalanobis distance is another distance measure which is used in certain classifiers such as hyper-ellipsoid and centered ellipsoid. This distance measure is normalized with the inverse of covariance matrix. This normalization removes the effect of absolute values of attributes. Thus, mahalanobis distance denotes the deviation of a data sample from the center of distribution without assuming a spherical nature of data distribution.

In order to understand the generalization ability and classification performance of various one-class SVM formulations, let us consider Fig. 1a–f. Figure 1a shows an example synthetic data set which consists of two attributes. All data vectors represented by blue colour in Fig. 1a are normal, whereas, those represented by red colour are abnormal. We now comment on the classification performance of each of the one-class SVM formulations by using them to classify the data of Fig. 1a. Figure 1b shows the classification of this data set into normal and abnormal vectors by using a hyperplane formulation. All the data samples below the hyperplane (characterized by weight vector w and bias r) which are represented by blue colour are classified as normal, whereas, those away from the hyperplane are classified as outliers by this formulation. The green points represent the data vectors which form the support of hyperplane formulation and are known as support vectors (Sect. 5). The yellow points in Fig. 1b show the data vectors which are actually anomalous, but they have been classified as normal by hyperplane formulation.

The classification of same data set using a hypersphere is shown in Fig. 1c. All data samples which lie inside the sphere structure of center c and radius R are classified as normal by this classification and represented by blue colour. Data points in red, which are outside the sphere

are classified as abnormal by this formulation. A comparison of Fig. 1b, c shows that the hypersphere formulation has a better classification performance as compared to hyperplane if the abnormal data points lie far away from the center of distribution. If the data points represented by yellow colour lie more far away from blue points in Fig. 1b, c, then the hypersphere formulation can classify these points as abnormal, whereas, the hyperplane formulation cannot classify these points as abnormal until they lie above the hyperplane. The quarter-sphere formulation (Fig. 1d) has the same performance as that of hypersphere formulation, however its computational complexity is less as compared to that of hypersphere formulation.

The hyperellipsoid formulation for the classification of same data set is shown in Fig. 1e. This formulation classifies all data points correctly because it fits a tight ellipsoid on the data distribution which allows all the data vectors which are away from the data distribution to be classified as abnormal. The centered ellipsoid formulation of Fig. 1f also has a performance similar to that of hyperellipsoid, however, it has an added advantage of reduced computational complexity.

A comparison of Fig. 1b–f shows that the ellipsoid based formulations have a better classification performance as compared to that of hyperplane and sphere based formulations. This better performance can be attributed to the distance measure used in these formulations. The use of mahalanobis distance in ellipsoid based formulations allows the data distribution to be fitted in a tight ellipse whose shape depends on covariance of the data set under consideration. The use of euclidean distance assumes a spherical data distribution and all data samples which are actually anomalous but do not lie far away from the distribution may be miss-classified. The ellipsoid based formulations can also generalize well to changing data distributions because they can update their parameters based on the data covariance, so, they have a better classification performance for unseen data as well. Thus, classification performance and generalization ability of one-class formulations can be arranged in the following increasing order:

hyperplane < hypersphere \approx quarter-sphere < hyperellipsoid \approx centered ellipsoid

5.5 Lessons to take forward

Before proceeding to detailed analysis of outlier & event detection techniques we present some conclusions on one-class SVM formulations that can be derived from the above discussion.

- The hyperplane based formulation has a poor generalization ability and classification performance. Moreover, it involves a quadratic optimization problem. Therefore, it is not possible to use hyperplane based formulation for outlier and event detection on energy constrained WSNs deployed in remote and harsh environments.
- A variant of hyperplane formulation (Gomez-Verdejo et al. 2011), which reduces the computational complexity can be used for WSNs. However, the problem of poor classification performance cannot be solved with this variant. We explore this technique further in next section.
- The hypersphere formulation has a better generalization ability and classification performance as compared to hyperplane formulation, however, a quadratic optimization problem makes it infeasible for implementation on energy constrained WSNs.
- The classification performance of quarter-sphere formulation is comparable to that of hypersphere formulation. However, it involves a linear optimization problem which results in a significant reduction in complexity as compared to that of hypersphere. Thus, it can be used for outlier and event detection in WSNs deployed in harsh environments.

State-of-the-art outlier and event detection techniques based on this formulation have been analyzed in next section.

- Hyperellipsoid based formulation has the best classification performance, however it cannot be used for WSNs because of the complexity associated with its quadratic optimization problem. As compared to the hyper-ellipsoid formulation, the centered ellipsoid formulation involves a linear optimization problem and it can be used for WSNs.
- Thus, quarter-sphere and centered ellipsoid based formulations can be used for outlier and event detection in WSNs deployed in harsh environments. A primary constraint of centered ellipsoid formulation is its communication complexity, which will be discussed in detail in next section.

6 Analysis of state-of-the-art outlier and event detection techniques based on one-class SVM formulations

Now that we have defined various one-class SVM formulations, we proceed to a detailed analysis of each of these formulations and determine their feasibility for WSNs deployed in harsh environments based on the characteristics discussed in Sect. 2.

6.1 General methodology for outlier and event detection by one-class SVMs

One-class SVM formulations follow a specific methodology for outlier and event detection which is outlined below:

6.1.1 Step 1—mapping to feature space

All the data samples collected over a particular time interval are mapped to a higher dimensional feature space. This mapping is required for the separation of data that is inseparable in the input space and can be done by the use of mercer kernels which calculate the dot product of data vectors in feature space.

6.1.2 Step 2—determining appropriate parameters

The data vectors mapped to the feature space are then enclosed in a geometric figure. This requires determining the parameters of geometric figure. For example, a hyperplane can be defined by its weight vector and a bias parameter, whereas a hyper-sphere can be defined by its centre and radius. These parameters are determined by solving a quadratic or linear optimization problem. A primal optimization problem is formulated for this purpose. The model complexity, that is, the complexity of determination of the parameters may be reduced by allowing some errors in the classification. This error can be allowed by introducing some non-negative slack variables and a regularization parameter in the primal optimization problem which allows some of the data samples to be miss-classified. The primal problem is later converted to a dual quadratic or linear problem by the use of lagrange function. We note here that the dual problem will be in terms of lagrange multiplier whose values are obtained for each of the data samples by using optimization algorithms. Hyperplane, hyper-sphere and hyper-ellipsoid are defined by two parameters each, therefore, they require the solution of a quadratic optimization problem, whereas the quarter-sphere and centered hyper-ellipsoid are defined only by their radius so they solve a linear optimization problem.

6.1.3 Step 3—global outlier detection

Once the parameters of geometric figure are determined, each node broadcasts its parameters to other nodes of the network. The central node ultimately receives the parameters of all nodes and merges them to compute global parameters. The merged parameters include weight vector w and bias r for hyperplane, radius R and center c for hypersphere based formulations and an additional covariance matrix for ellipsoid based formulations. The merged parameters are then broadcasted back to all nodes of the network. Each node then classifies the data samples based on their location in the feature space around the geometry. There are two methods to classify the data samples:

1. **Based on the values of lagrange multipliers:** All data samples which lie inside the geometric figure are normal and their associated lagrange multipliers have a value of zero. The data samples which lie on the quarter sphere have a non-zero value for their lagrange multipliers. Such vectors are known as border support vectors. All the data vectors whose lagrange multipliers reach a maximum value equal to $1/(vn)$ are known as outliers because they lie outside the geometric formulation (Fig. 1b–f).
2. **Based on distance measure:** The data samples for all one-class SVM formulations which enclose the data in a closed geometry, such as a hypersphere, hyperellipsoid, quarter-sphere or centered ellipsoid, can be classified as normal or abnormal based on their distance from the center of data distribution. This can be achieved by comparing the euclidean or mahalanobis distance of these samples from the center of geometry with the radius of geometry. All data samples whose distances exceed the radius are known as outliers. Later in this article we will comment on the notion that generalization ability of mahalanobis distance based classifiers is better than those based on euclidean distance.

6.1.4 Event detection and identification

Once a global outlier has been detected at a node of the network, a consensus of other nodes in the network is invoked to get information about the presence or absence of outliers on them. An event is said to have occurred if majority of nodes in a neighborhood show the presence of global outliers. After the occurrence of event, further analysis should be carried out to determine the event cause and attributes involved in the event. This process is known as event type identification.

A summary of outlier and event detection process using one-class SVMs is presented in Fig. 3.

From the discussion carried out in the previous section it was concluded that quarter-sphere (QS-SVM) and centered ellipsoid (CESVM) based formulations are most suitable for outlier and event detection in WSNs because of their low computational complexity. In the discussion that follows, we present a detailed analysis of techniques based on these two formulations. The discussion carried out in this section will highlight that in addition to low computational complexity, the techniques based on these formulations possess a number of additional characteristics which make them suitable specifically for WSNs deployed in harsh environments.

6.2 Techniques based on quarter-sphere formulation (QS-SVM)

Various versions of QS-SVM based outlier and event detection techniques have been proposed in literature. Here we briefly discuss each of these techniques (Table 5).

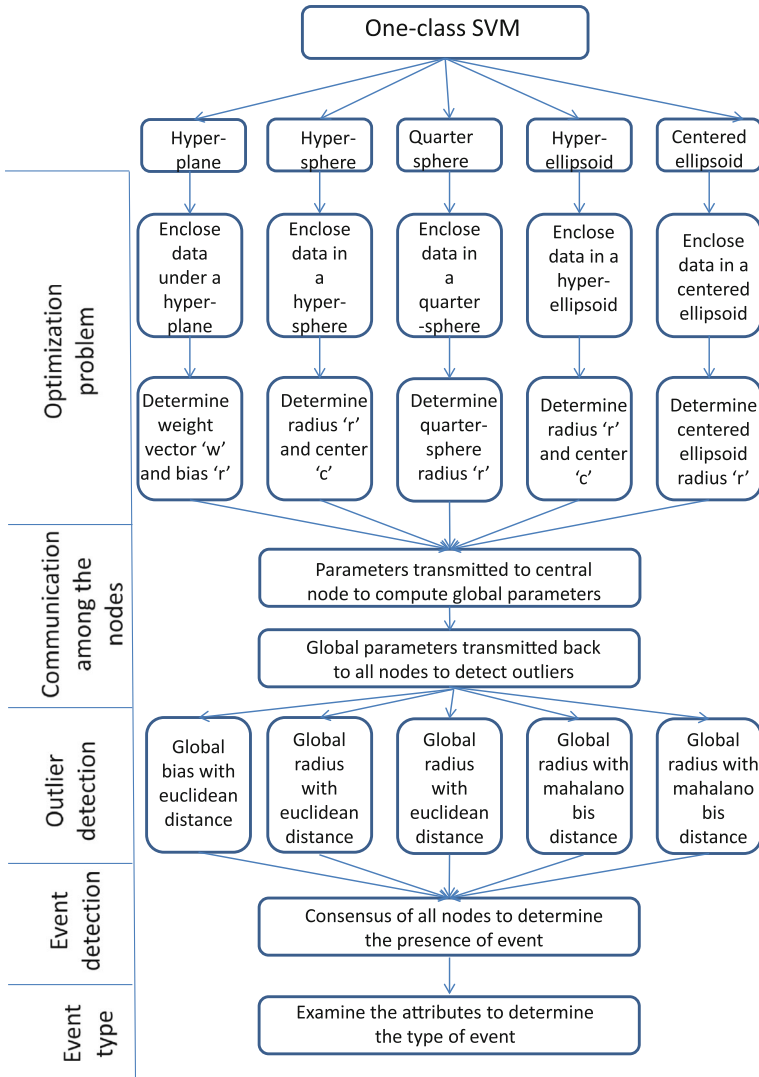


Fig. 3 General strategy for outlier and event detection by various one-class SVM formulations

6.2.1 Offline QS-SVM approach with spatio-temporal correlations

Rajasegarar (2007) presented an SVM based approach for distributed anomaly detection in hierarchical or non-hierarchical network structures. The authors have proposed an offline technique which requires minimal communication between the nodes of network. Assuming a homogeneous environment, that is, where all the measurements produced by sensors have same unknown distribution, each node runs a quarter sphere based anomaly detection algorithm to define a quarter sphere as a boundary for normal data. All the data points lying outside the quarter sphere are outliers, whereas the points lying on the sphere are border support vectors and the points lying inside the sphere are non-support vectors. Each node

identifies the local radius of the quarter sphere and keeps the record of norms of data vector in its memory and then sends its radius information to the parent node. The parent node then collects the radius information from its children and combines them with its own local radius. It then computes the global radius, which is used for global anomaly detection. Four strategies, namely, mean, median, maximum and minimum of the combined radii (from the parent node and children nodes) have been considered for the computation of global radius. Parent node sends the global radius to its children nodes, which then compare the norms of their data vectors with the global radius and classifies them as globally anomalous or normal. A data vector is identified as globally anomalous if its norm is greater than the square of global radius. This technique does not perform event detection. A modification of this technique is presented next which can be used for event detection. (The analysis of this technique is presented in Table 6).

6.2.2 Online QS-SVM approach with spatio-temporal correlations

An online version of the technique (Rajasegarar et al. 2007) (based on QS-SVM) for environmental monitoring applications of WSNs is presented in Yang et al. (2008). This technique shows that by taking advantage of spatial correlations that exist in sensor data of adjacent nodes, the false alarm rate can be reduced and even real-time distinction between events and errors can be made. Initially, each node learns the local radius of the quarter-sphere using its m sequential data measurements, which may include some anomalous data. The one-class quarter-sphere SVM can efficiently find a minimal radius to enclose the majority of these mapped data measurements in the feature space. Each node then locally broadcasts the learned radius information to its spatially neighboring nodes. In fact, a node first collects the radius information from all of its neighbors and then computes a median radius R_{im} of its neighboring nodes as well as a median radius of its closed neighborhood. When a new data measurement x arrives at a node, it computes the Euclidian distance $d(x)_c$ between x and the origin of its centered quarter-sphere in the feature space. Based on the fact that sensor data collected in a densely deployed WSN tends to be correlated in both time and space, a sensor first compares $d(x)_c$ with its own radius R_i . The data x will be classified as normal if $d(x)_c \leq R_i$, which means that x falls on or inside the quarter-sphere at node S_i . Otherwise if $d(x)_c > R_i$, x is a potential (temporal) outlier. In this case, the i th node S_i further compares $d(x)_c$ with the median radius R_{im} of its spatially neighboring nodes. If $d(x)_c > R_{im}$, x will finally be classified as outlier in the neighborhood. This technique also differs between errors and events based on the fact that temporal deviations correspond to events whereas spatio-temporal deviations correspond to errors. (The analysis of this technique is presented in Table 6).

6.2.3 Online & adaptive QS-SVM approach with spatio-temporal correlations

An adaptive and communication efficient version of Yang et al. (2008) is presented in Zhang et al. (2009a). This technique is identical to Yang et al. (2008), in that, it identifies outliers based on spatio-temporal deviations. However the authors have presented three different approaches to online outlier detection namely Instant Outlier Detection (IOD), Fixed-size Time Window-based Outlier Detection Technique (FTWOD) and Adaptive Outlier Detection Technique (AOD). All of these techniques use the quarter sphere based support vector methods to identify outliers, however the difference lies in update of the training set. In IOD the training set is updated at every time instant. This technique is computationally expensive

Table 5 A summary of one-class SVM formulations

Formulation	Description	Model parameters	Optimization	Examples	Distance metric
1 Hyperplane	Normal and abnormal data separated by a hyperplane	Weight vector w bias parameter r	Quadratic	Schölkopf et al. (2001), Gomez-Verdejo et al. (2011)	Euclidian
2 Hypersphere	Normal data enclosed in a hypersphere of radius R . Abnormal data lies outside the hypersphere	Radius R centre c	Quadratic	Tax and Dium (1999)	Euclidian
3 Quarter-sphere	Normal data enclosed in a quarter-sphere of radius R . Abnormal data lies outside the quarter-sphere	Radius R	Linear	Rajasegarar et al. (2007), Yang et al. (2008), Zhang et al. (2009a), Shahid et al. (2012, 2012a)	Euclidian
4 Hyperellipsoid	Normal data enclosed in an ellipsoid of radius R and centre c . Abnormal data lies outside the ellipsoid	Radius R centre c	Quadratic	Wang et al. (2006)	Mahalanobis
5 Centered ellipsoid	Normal data enclosed in an ellipsoid of radius R , centred at origin. Abnormal data located outside	Radius R	Linear	Zhang et al. (2009, 2012)	Mahalanobis

Table 6 Feasibility analysis table for quarter-sphere SVM based outlier & event detection techniques

Characteristics		Rajasegarar et al. (2007)	Yang et al. (2008)	Zhang et al. (2009a)	Shahid et al. (2012)	Shahid et al. (2012a)
1	Input data	Univariate	Y	Y	Y	Y
		Attributes	Y	Y	Y	Y
	Data Type	Multiple	Mixed			
	Complex					
2	Data availability	Temporal	?	Y	Y	Y
		Spatial	Y	Y	Y	Y
	Correlations	Y	Y	Y	Y	Y
	Attribute					
2	User specified thresholds	Supervised				
		Semi-supervised				
	Unsupervised	Y	Y	Y	Y	
	Distance					
3	Outlier types	Neighborhood				
		Significance test				
	Suspected no of outliers	Fixed no.				
	Window width others	Upper bound				
3	Local Outliers	Centralized	Y	Y	Y	Y
		Distributed Approach	Y	Y	Y	Y

Table 6 continued

Characteristics		Rajasegarar et al. (2007)	Yang et al. (2008)	Zhang et al. (2009a)	Shahid et al. (2012)	Shahid et al. (2012a)
	Global					
	Outliers					
	Centralized					
	Distributed					
	Approach					
	Individually					
	Aggregation					
4	Number of outliers	Y	Y	Y	Y	Y
	Single					
	Multiple	Y	Y	Y	Y	Y
5	Outlier Identity	Y	Y	Y	Y	Y
	Errors					
	Noise					
	Events	?	Y	Y	Y	Y
6	Outlier handling	?	?	?	?	?
	Identification					
	Accommodation					
7	Outlier degree	Y	Y	Y	Y	Y
	Scalar					
	Outlier					
	Score					
	Static					
	Dynamic					
	Hierarchical	Y	Y	Y	Y	Y
	Non-hierarchical	Y	Y	Y	Y	Y
9	Network architecture					
	Communication complexity	High	High	High	High	Low
10	Computational complexity	Low	Low	Low	Low	Low
11	Online approach		Y	Y	Y	Y

Table 6 continued

Characteristics	Rajasegarar et al. (2007)	Yang et al. (2008)	Zhang et al. (2009a)	Shahid et al. (2012)	Shahid et al. (2012a)
12 Offline approach	Y				
13 Can be made online	Y				
14 Non-stationarity		Y	Y	Y	Y
15 Type of algorithm used	Temporal	Y	Y	Y	Y
	Spatial				
	Set based				
	Density based				
	Graph based				
	Depth based				
	Model based	Y	Y	Y	Y
16 Feasibility for harsh environments	Distance based				
	Not Possible				
	Possible without improvements	Y	Y	Y	Y
	Possible with improvements				

In this table ‘Y’ correspond to the presence of the characteristic, ‘?’ corresponds to the improvement possibility and space corresponds to the absence of a characteristic

as the radius of quarter-sphere is computed at each time instant and a significant communication cost is incurred for broadcasting this radius information to neighboring nodes. FTWOD requires the update of training set after every n samples. AOD overcomes the computation and communication overhead of IOD and FTWOD by updating the training set only when an outlier is identified. Thus the training set will remain constant until an outlier occurs. This technique reduces the communication overhead significantly. (The analysis of this technique is presented in Table 6).

6.2.4 Online and adaptive QS-SVM approach with Spatio-temporal-attribute correlations (STA-QS-SVM)

The QS-SVM based approaches presented in [Zhang et al. \(2009a\)](#), [Rajasegarar et al. \(2007\)](#), [Yang et al. \(2008\)](#) incorporate temporal correlations of data to detect outliers and spatial correlations to detect global outliers or events. However, none of these approaches incorporate attribute correlations which play an important role in outlier detection ([Shahid et al. 2012](#)). An online and adaptive QS-SVM based approach that incorporates temporal-attribute correlations to detect outliers and spatial correlations to detect events is presented in [Shahid et al. \(2012\)](#). This technique tends to correct the radius of quarter-sphere determined by the approaches of [Zhang et al. \(2009a\)](#), [Rajasegarar et al. \(2007\)](#), [Yang et al. \(2008\)](#) via attribute radius value. The attribute radius value is determined by applying the QS-SVM problem by treating individual attributes as vectors. This results in the exploitation of attribute correlations of the data set. Thus, two radii are calculated, that is temporal and attribute and their mean is taken to form the final quarter-sphere radius. The technique reports significant performance gains over other QS-SVM based techniques. Thus, the radius of quarter-sphere in this technique is determined by temporal-attribute correlations of data set. Those data samples, which fall outside the quarter-sphere are declared to be outliers. Once outliers are detected, a procedure similar to that presented in [Zhang et al. \(2009a\)](#), [Rajasegarar et al. \(2007\)](#), [Yang et al. \(2008\)](#) is invoked to determine the presence of events in the network. (The analysis of this technique is presented in Table 6).

6.2.5 STA-QS-SVM with low complexity

STA-QS-SVM ([Shahid et al. 2012](#)) reports significant performance gains over other QS-SVM based techniques because it incorporates attribute correlations. However, a major drawback of this approach is its high computational complexity. Moreover, the QS-SVM based techniques broadcast the radius of quarter-sphere to central node of the network at each time instant to detect events. An effort to reduce high computational and communication complexity of these techniques has been presented in [Shahid et al. \(2012a\)](#). This paper presents three approaches to outlier and event detection, namely, STA-TSV, STA-TASV & STA-CA. All of these approaches follow the procedure of STA-QS-SVM for outlier and event detection. However, they update and broadcast the radius of quarter-sphere in a different manner as compared to STA-QS-SVM. STA-TSV, broadcasts the radius of quarter-sphere on the arrival of temporal support vectors (TSV) only, whereas, STA-TASV broadcasts the radius of quarter-sphere on the arrival of temporal-attribute support vectors (TASV). The number of temporal or temporal-attribute support vectors in a data set is very small, therefore, these techniques result in a significant reduction in communication complexity of STA-QS-SVM. The performance of these techniques is comparable to that of STA-QS-SVM because only support vectors contribute to the quarter-sphere radius. Hence, broadcast of quarter-sphere

radius at each time instant is not required. STA-CA tends to reduce both the computational and communication complexity of STA-QS-SVM by calculating the attribute radius only on the arrival of temporal support vectors in the data set. Moreover, the temporal-attribute radius is only broadcasted on the arrival of temporal support vectors. (The analysis of this technique is presented in Table 6).

6.2.6 Analysis of QS-SVM based techniques

Following is a brief analysis of QS-SVM based approaches [Zhang et al. \(2009a\)](#), [Shahid et al. \(2012, 2012a\)](#), [Rajasegarar et al. \(2007\)](#), [Yang et al. \(2008\)](#).

- All the QS-SVM based techniques consider multivariate data. Streaming data, however, can only be handled by the techniques which can perform online data processing ([Zhang et al. 2009a](#); [Shahid et al. 2012, a](#); [Yang et al. 2008](#)).
- Spatio-temporal correlations are considered by all the techniques, thus, all the techniques can detect global outliers or events. However, attribute correlations are only considered by [Shahid et al. \(2012, 2012a\)](#). The techniques which consider attribute correlations have a better outlier and event detection performance.
- All the QS-SVM based techniques are unsupervised, that is, they do not require labeled input-output data during the training phase. Moreover, all the techniques support distributed processing.
- All QS-SVM based techniques are online except ([Rajasegarar et al. 2007](#)). As the techniques presented in [Shahid et al. \(2012, 2012a\)](#) are the only techniques that considers attribute correlations, so ([Shahid et al. 2012, a](#)) are robust to dynamically changing data distributions.
- The quarter-sphere formulation, unlike the hyperplane and hyper-sphere formulations, is described by a radius only. Thus, it poses a linear optimization problem as compared to quadratic optimization problem posed by other formulations. This corresponds to a significant decrease in computational complexity making the QS-SVM based techniques suitable for implementation on WSN nodes.
- None of the QS-SVM based techniques can identify the type of event or attributes involved in it.

Based on the analysis of QS-SVM based techniques, it can be concluded that ([Shahid et al. 2012a](#)) is the only technique that is suitable for implementation in harsh environments without modifications.

6.3 Centered ellipsoidal formulation (CESVM)

Only a few techniques based on centered ellipsoidal formulation have been presented in literature. We briefly discuss each of these techniques.

6.3.1 Offline CESVM

The CESVM based algorithm for detection of outliers and events is similar to that of quarter-sphere formulation ([Rajasegarar et al. 2007](#)). Here the data vectors in the input space are first mapped to a higher dimensional space and centered by the mean of kernel functions. Then, a hyperellipsoid, centered at the origin, with minimal effective radius is fitted around the majority of the data vectors in the higher dimensional space. The mahalanobis distance of newly

Table 7 Feasibility analysis table for hyper-ellipsoidal SVM based outlier & event detection techniques

Characteristics		Rajasegarar et al. (2010)	Zhang et al. (2009)	Zhang et al. (2012)
1	Input data	Univariate	Y	Y
		Multiple	Y	Y
	Data Type	Complex		
	Attributes	Sequential Streaming	Y	Y
2	Data availability	Temporal	Y	Y
		Spatial	Y	Y
	Correlations	Attribute	Y	Y
	Supervised			
2	User specified thresholds	Semi-supervised		
		Unsupervised	Y	Y
	Distance			
	Neighborhood			
3	Outlier types	Significance test		
		Suspected no of outliers	Fixed no. Upper bound	
		Window width		
3	Outlier types	Others	Y	Y
		Local outliers		
		Centralized		
3	Outlier types	Distributed	Y	Y
		Individually	Y	Y

Table 7 continued

Characteristics		Rajasegarar et al. (2010)	Zhang et al. (2009)	Zhang et al. (2012)
	Approach	Collaboration		Y
	Centralized			
	Distributed	Individually	Y	Y
	Approach	Aggregation		
4	Number of outliers			
	Global outliers			
	Single	Y	Y	Y
	Multiple	Y	Y	Y
	Errors			
	Noise			
	Events	Y	Y	Y
	Identification	?	?	?
	Accommodation			
7	Outlier degree			
	Scalar			
	Outlier	Y	Y	Y
	Score			
	Static	Y	Y	Y
	Dynamic			
	Hierarchical	Y	Y	Y
	Non-hierarchical	Y	Y	Y
10	Communication complexity	High	High	High
	Computational complexity	Low	Low	Low

arrived data vectors is compared with the radius of hyperellipsoid. All the data vectors that fall on or inside the hyperellipsoid are normal and those falling outside the hyperellipsoid are classified as anomalous. A better detection rate is achieved with hyper-ellipsoids as compared to the quarter-sphere formulation. This can be attributed to the use of mahalanobis distance as a metric for comparison of newly arrived data vectors with the hyperellipsoid radius. However, the hyper-ellipsoidal formulation (Rajasegarar et al. 2008) has more communication overhead as compared to the hyper-sphere (Tax and Duin 1999) and quarter-sphere formulations (Rajasegarar et al. 2007) because it requires broadcasting the covariance matrix to the neighboring nodes in the network. The broadcast of covariance matrix is required to the central node of the network to determine the global covariance matrix. For a d attribute data set, covariance matrix is a $d \times d$ symmetric matrix. Due to symmetric nature of this matrix, a set of $d(d+1)/2$ values have to be broadcasted at each time instant. (The analysis of this technique is given in Table 7).

6.3.2 Online CESVM

The hyper-ellipsoidal based technique presented in Rajasegarar et al. (2010) is offline. An online and distributed approach to outlier detection based on centered hyper-ellipsoids (Rajasegarar et al. 2008) was presented by Zhang et al. (2009). Each sensor node models its hyper-ellipsoid SVM for m observations and then obtains the hyper-ellipsoid radius R_i as well as the corresponding covariance matrix of centered data. Then the local outliers can be determined in real-time by comparing the distances between new observations and the origin with R_i . Each node then communicates its parameters, that is, R_i and the covariance matrix to its spatial neighbors. Due to the fact that the covariance matrix is symmetric, each node only transmits $d(d+1)/2$ elements for covariance matrix, where d is the number of attributes of a sensor observation. These parameters are then merged with similar parameters of other nodes to form global radius and global covariance matrix. Finally each node uses its merged parameters to determine global outliers for its new observations in an online manner. (The analysis of this technique is given in Table 7).

6.3.3 Online CESVM with low complexity

The CESVM based formulation presented in Rajasegarar et al. (2010) has a high computational and communication complexity for the following reasons: a) it requires the calculation of covariance matrix of data b) it requires the computation of mahalanobis distance of newly arrived data samples in feature space c) it requires the broadcast of hyperellipsoid parameters, that is, its radius and covariance matrix to other nodes of network at every time instant. The first two problems have been overcome by the authors of Zhang et al. (2012). This technique presents a method for formulating the hyperellipsoid in the input space, rather than the feature space. It is based on the fact that the data collected from a typical WSN is separable in input space, thus, outliers can easily be separated from normal measurements in the input space. The optimization problem is formulated to determine the radius of hyper-ellipsoid in the input space. Thus, the technique requires the solution of a linear optimization problem and is far less expensive as compared to Rajasegarar et al. (2010). The outliers are detected by comparing the norm of newly arrived data vectors with the radius of hyper-ellipsoid. Once an outlier is detected at a particular node, the decisions of other nodes can be invoked to determine the presence of an event in the network. (The analysis of this technique is given in Table 7).

6.3.4 Analysis of CESVM based techniques

Following is a brief analysis of state-of-the-art centered hyper-ellipsoidal based techniques (Zhang et al. 2009, 2012; Rajasegarar et al. 2010).

- All the CESVM based techniques consider multivariate data. However, only (Zhang et al. 2009, 2012) consider the streaming data.
- Spatio-temporal correlations are considered by all the hyper-ellipsoidal based techniques. An important advantage of CESVM based techniques over QS-SVM techniques is that CESVM based techniques consider attribute correlations of data via covariance matrix. The calculation of covariance matrix, however, requires additional computational complexity and memory. Thus, the CESVM based techniques detect local outliers using temporal-attribute correlations and detect events or global outliers by exploiting spatial correlations between the geographically separated nodes of the network.
- All the CESVM based techniques perform online data processing except (Rajasegarar et al. 2010). Moreover, due to their ability to incorporate attribute correlations, the techniques presented in Zhang et al. (2009, 2012), are robust to dynamically changing data distributions.
- All the CESVM based techniques pose a linear optimization problem and are computationally inexpensive like QS-SVM based techniques. However, an inherent problem of CESVM based techniques is their high communication cost because they require the broadcast of entire covariance matrix among the nodes of the network. Although the technique presented in Zhang et al. (2012) tends to reduce the computational complexity of CESVM based techniques, however, it is still not suitable for outlier and event detection due to high communication cost.
- None of the CESVM based techniques can identify the type of event or the attributes involved in it.

From the analysis carried out in the above discussion it can be concluded that the offline technique presented in Rajasegarar et al. (2010) is unsuitable for WSNs deployed in harsh environments, whereas the techniques presented in Zhang et al. (2009, 2012) can be used for outlier and event detection in harsh environments after some modifications to reduce the communication complexity.

6.4 How to use the techniques based on hypersphere and hyperellipsoid formulations for outlier and event detection in WSNs?

The hypersphere and hyperellipsoid based techniques have a good classification performance but a quadratic optimization problem limits their applicability for WSNs. Thus, these formulations can only be used for outlier and event detection in WSNs if a method can be devised to reduce their complexity. A method based on iterative re-weighted least squares algorithm has been presented to reduce the complexity of hyperplane based formulation. We present this method below and propose a similar solution for hypersphere and hyperellipsoid formulations.

6.4.1 The adaptive hyperplane formulation

The problem of high computational complexity of the hyperplane based formulation (Schölkopf et al. 2001) has been overcome by the technique presented in Gomez-Verdejo et al. (2011). This technique presents an online and adaptive version of a hyperplane

which can update its parameters with the newly arrived data samples. Further, the computational complexity associated with the quadratic optimization problem of hyperplane has been reduced by using iterative re-weighted least squares (IRWLS) method. In IRWLS, the machine structure is updated via growing and pruning mechanisms and the weights are updated using structural risk minimization principles underlying SVMs. The technique is known as adaptive because it uses a forgetting factor that enables the hyperplane parameters to be updated with the most newly arrived data, thus, discarding the effect of old data samples.

7 Bottlenecks of one-class SVMs and future research directions

Although one-class SVM formulations, such as QS-SVM and CESVM satisfy most of the characteristics which are required for outlier and event detection in harsh environments but they have not been extensively explored by the research community for WSN applications due to the following reasons:

7.1 Optimization complexity

The optimization problems required in QS-SVM and CESVM are linear, but for an online technique the solution of this problem is required with the arrival of every new data sample. For example, QS-SVM based technique presented in [Shahid et al. \(2012\)](#) requires the solution of a linear optimization problem at every time instant. Moreover, to detect global outliers and events, the determined parameters are also broadcasted at every time instant. Thus, computation and communication is required at each time instant. Although the techniques presented in [Shahid et al. \(2012a\)](#) reduce communication and computations but more efforts are needed for this purpose. One way to reduce computations can be to use IRWLS algorithm for QS-SVM and CESVM as done in [Gomez-Verdejo et al. \(2011\)](#). Further, CESVM requires even more communication than QS-SVM because the whole covariance matrix has to be broadcasted to central node. The research community should focus on developing communication and computation effective algorithms based on QS-SVM and CESVM without compromising their performance.

7.2 Choice of optimal regularization parameter

An important parameter for one-class SVM problems is the regularization parameter v which decides the fraction of data that is classified as outliers. Small values of this parameter result in high values of $1/vn$, thus, increasing the cost of miss-classification. The increased cost of miss-classification results in a low detection rate and a low false positive rate. Large values of v result in a small value of $1/vn$, thus decreasing the cost of miss-classification. The decreased cost of miss-classification results in a high value of false positive rate and a high detection rate. The choice of optimal regularization parameter can be made by using cross validation. N-fold Cross validation divides the data set into N distinct portions, where some of the portions are used as training set to determine the value of regularization parameter and the remaining data sets are used for testing purposes. An inherent draw back of cross validation is that the procedure has to be repeated time and again for dynamically changing data distributions. The research community should focus on methods to determine the optimal value of regularization parameter in an online manner.

7.3 Event identification

A review of state-of-the-art outlier and event detection techniques based on one-class SVM shows that none of the techniques incorporate an event identification strategy. Event identification strategies play an important role to identify the type of event and attributes involved in event (Shahid et al. 2012). The research community should focus on developing low complexity one-class SVM based event identification algorithms so that the need of manual examination of attributes can be eliminated.

7.4 Soft decision based algorithms

All one-class SVM based techniques analyzed in this paper provide a hard decision for the presence or absence of outliers. Ideally an outlier detection technique should be able to assign an outlier score to each of the newly arrived data samples. This would provide information about gradual change in the state of data samples from normal to outliers which can be used to generate a pre-event warning in harsh environments

8 Conclusion

This paper presents a detailed analysis of various one-class Support Vector machine formulations for outlier and event detection in Wireless sensor networks deployed in harsh environments. These formulations include hyper-plane, hyper-sphere, quarter-sphere and hyper-ellipsoid. The analysis has been carried out in terms of characteristics like input data type, spatio-temporal and attribute correlations, user specified thresholds, outlier types (local and global), type of approach (distributed/centralized), outlier identification (event/error), outlier degree, susceptibility to dynamic topology, non-stationarity and inhomogeneity. A tabular description of improvement and feasibility of various techniques for deployment in the harsh environments has also been presented. We observe from the analysis that the quarter-sphere formulations are most suitable for outlier detection in WSNs because of their low computation and communication complexity and high efficiency. Hyper-plane and hyper-sphere formulations are computationally expensive and do not generalize well to variety of distributions. Hyper-ellipsoid based techniques, although more efficient than quarter-sphere techniques, pose a high communication overhead; hence can be used for energy constrained WSNs only after suitable modifications.

References

- Abe S (2010) Support vector machines for pattern classification. Springer New York Inc
- Akyildiz IF, Su W, Sankarasubramanian Y, Cayirci E (2002) Wireless sensor networks: a survey. *Comput Netw* 38:393–422
- Akyildiz IF, Akan Özgür B (2003) Interplanetary internet: state-of-the-art and research challenges. *Comput Netw* 43:75–112
- Aly M (2005) Survey on multiclass classification methods. *Neural Netw* pp 1–9
- Ambwani T (2003) Multi class support vector machine implementation to intrusion detection. In: *Neural networks, 2003. Proceedings of the international joint conference on*, vol. 3. IEEE, pp 2300–2305
- Bahreppour M, Meratnia N, Havinga PJM (2010) Fast and accurate residential fire detection using wireless sensor networks. *Environ Eng Manag J* 9(2):215–221

- Bahrepour M, Meratnia N, Poel M, Taghikhaki Z, Havinga PJM (November 2010) Distributed event detection in wireless sensor networks for disaster management. In: International conference on intelligent networking and collaborative systems, INCoS (2010) Thessaloniki, Greece. IEEE Computer Society, USA, pp 507–512
- Bahrepour M, van der Zwaag BJ, Meratnia N, Havinga PJM (2010) Fire data analysis and feature reduction using computational intelligence methods. In: Phillips-Wren G, Jain LC, Nakamatsu K, (eds) Advances in intelligent decision technologies. In: Proceedings of the second KES international symposium IDT 2010, Baltimore, Maryland, USA ser. Smart innovation, systems and technologies vol 4. Springer, Berlin Heidelberg, pp 289–298
- Bahrepour M, Zhang Y, Meratnia N, Havinga PJM (December 2009) Use of event detection approaches for outlier detection in wireless sensor networks. In: Proceedings of symposium on theoretical and practical aspects of large-scale wireless sensor networks, The 5th international conference on intelligent sensors, sensor networks and information processing (2009) (ISSNIP 2009), Melbourne, Australia. IEEE Press, Victorian, pp 439–444
- Barnett V, Lewis T (1994) Outliers in statistical data. Wiley, New York
- Bettencourt LMA, Hagberg AA, Larkey LB (Jun 2007) Separating the wheat from the chaff: practical anomaly detection schemes in ecological applications of distributed sensor networks. In: Distributed computing in sensor systems (DCOSS 2007). Santa Fe, NM USA, pp 223–239
- Bezdek J, Rajasegarar S, Moshtaghi M, Leckie C, Palaniswami M, Havens T (2011) Anomaly detection in environmental monitoring networks [application notes]. IEEE Comput Intell Mag 6(2):52–58
- Bezdek J, Havens T, Keller J, Leckie C, Park L, Palaniswami M, Rajasegarar S (July 2010) Clustering elliptical anomalies in sensor networks. In: Fuzzy systems (FUZZ), 2010 IEEE international conference on, pp 1–8
- Bhuse V, Gupta A (2006) Anomaly intrusion detection in wireless sensor networks. J High Speed Netw 15:33–51
- Bishop C (2006) Pattern recognition and machine learning, ser. Information science and statistics. Springer, [Online]. Available: <http://books.google.com.pk/books?id=kTNoQgAACAAJ>
- Boyd S, Vandenberghe L (2004) Convex optimization. Cambridge university press, Cambridge
- Branch J, Szymanski B, Giannella C, Wolff R, Kargupta H (2006) In-network outlier detection in wireless sensor networks. In: Distributed computing systems, 2006. ICDCS 2006. 26th IEEE international conference on, p 51
- Bredensteiner E, Bennett K (1999) Multicategory classification by support vector machines. Comput Optim Appl 12(1):53–79
- Cardell-Olivera R, Kranza M, Smettemb K, Mayerc K (2005) A reactive soil moisture sensor network: design and field evaluation. Int J Distrib Sens Netw 1(2):149–162
- Chandola V, Banerjee A, Kumar V (July 2009) Anomaly detection: a survey. ACM Comput. Surv., vol. 41, pp 15:1–15:58. [Online]. Available: <http://doi.acm.org/10.1145/1541880.1541882>
- Chen J, Kher S, Somani A (2006) Distributed fault detection of wireless sensor networks. In: Proceedings of the 2006 workshop on dependability issues in wireless ad hoc networks and sensor networks, ser. DIWANS '06. New York, NY, USA: ACM, pp 65–72. [Online]. Available: <http://doi.acm.org/10.1145/1160972.1160985>
- Cheong S, Oh S, Lee S (2004) Support vector machines with binary tree architecture for multi-class classification. Neural Inf Process Lett Rev 2(3):47–51
- da Silva APR, Martins MHT, Rocha BPS, Loureiro AAF, Ruiz LB, Wong HC (2005) Decentralized intrusion detection in wireless sensor networks. In: Proceedings of the 1st ACM international workshop on quality of service & security in wireless and mobile networks, ser. Q2SWinet '05. New York, NY, USA: ACM, pp 16–23. [Online]. Available: <http://doi.acm.org/10.1145/1089761.1089765>
- Dario IA, Akyildiz IF, Pompili D, Melodia T (2005) Underwater acoustic sensor networks: research challenges. Ad Hoc Netw Elsevier 3:257–279
- Dereszynski E, Dietterich T (2011) Spatiotemporal models for data-anomaly detection in dynamic environmental monitoring campaigns. ACM Trans Sens Netw (TOSN) 8(1):3
- Ding M, Cheng X (april 2009) Robust event boundary detection in sensor networks—a mixture model based approach. In: INFOCOM 2009, IEEE, pp 2991–2995
- Ding M, Chen D, Xing K, Cheng X (march 2005) Localized fault-tolerant event boundary detection in sensor networks. In: INFOCOM 2005. 24th annual joint conference of the IEEE computer and communications societies. Proceedings IEEE, vol. 2, pp 902–913 vol. 2
- Ding X, Li Y, Belatreche A, Maguire L (2012) Constructing minimum volume surfaces using level set methods for novelty detection. In: Neural networks (IJCNN), The 2012 international joint conference on. IEEE, pp 1–6
- Elisseeff A, Weston J (2002) Kernel methods for multi-labelled classification and categorical regression problems. Adv Neural Inf Process Syst 14:681–687
- Elnahrawy E, Nath B (2004) Context-aware sensors. European workshop on wireless sensor, networks, In, pp 77–93

- Franc V, Hlaváč V (2002) Multi-class support vector machine. In: Pattern Recognition, 2002. Proceedings. 16th International Conference on, vol. 2. IEEE, pp 236–239
- Ganguly AR (2008) Knowledge discovery from sensor data. CRC Press, Boca Raton
- García-Hernández CF, Ibarguiengoytia-González PH, García-Hernández J, Pérez-Díaz JA (2007) Wireless sensor networks and applications: a survey. *IJCSNS Int J Comput Sci Netw Secur* 7(3):264–273
- George S, Zhou W, Chenji H, Won M, Lee Y, Pazarloglou A, Stoleru R, Barooah P (2010) Distressnet: a wireless ad hoc and sensor network architecture for situation management in disaster response. *IEEE Commun Mag* 48(3):128–136
- Gitrakos N, Kotidis Y, Deligiannakis A (2010) Pao: power-efficient attribution of outliers in wireless sensor networks. In: Proceedings of the seventh international workshop on data management for sensor networks. ACM, pp 33–38
- Gitrakos N, Kotidis Y, Deligiannakis A, Vassalos V, Theodoridis Y (2010) Taco: tunable approximate computation of outliers in wireless sensor networks. In: Proceedings of the 2010 international conference on management of data. ACM, pp 279–290
- Gomez-Verdejo V, Arenas-Garcia J, Lazaro-Gredilla M, Navia-Vazquez A (2011) Adaptive one-class support vector machine. *IEEE Trans Signal Process* 59(6):2975–2981
- Han J, Kamber M (2006) Data mining: concepts and techniques. Morgan Kaufmann
- Hao P, Chiang J, Lin Y (2009) A new maximal-margin spherical-structured multi-class support vector machine. *Appl Intell* 30(2):98–111
- Hassan A et al. (2011) A heuristic approach for sensor network outlier detection. *Int J Res Rev Wirel Sens Netw (IJRRWSN)* 1(4):
- Herbrich R (2002) Learning kernel classifiers: theory and algorithms. The MIT press, Cambridge
- Hill DJ, Minsker BS, Amir E (2007) Real-time bayesian anomaly detection for environmental sensor data. In: In Proceedings of the 32nd conference of IAHR
- Hofmann T, Schölkopf B, Smola A (2008) Kernel methods in machine learning. *Ann Stat*, pp 1171–1220
- Hsu C, Lin C (2002) A comparison of methods for multiclass support vector machines. *IEEE Trans Neural Netw* 13(2):415–425
- <http://connekt.seecs.nust.edu.pk/SAHSE.php>
- <http://www.genuki.org.uk/big/eng/LAN/Haydock/WoodPitExplosion.html>
- <http://www.humanite.fr/2006-03-10Societe-Catastrophe-de-Courrieres-une-expression-impropre>
- <http://www.msha.gov/MSHAINFO/FactSheets/MSHAFCT8.HTM>
- Janakiram D, Adi Mallikarjuna Reddy V, Phani Kumar A (2006) Outlier detection in wireless sensor networks using bayesian belief networks. In: Communication system software and middleware, 2006. Comsware 1–6
- John GH (1995) Robust decision trees: removing outliers from databases. In: In Knowledge discovery and data mining. AAAI Press, pp 174–179
- Jun MC, Jeong H, Kuo C-CJ (2005) Distributed spatio-temporal outlier detection in sensor networks
- Katdare M (2011) Pattern based outlier detection in mixed-attribute datasets
- Keally M, Zhou G, Xing G (2010) Watchdog: confident event detection in heterogeneous sensor networks. In: Real-time and embedded technology and applications symposium (RTAS), (2010) 16th IEEE. *IEEE* 2010:279–288
- Kecman V (2005) Support vector machines—an introduction, support vector machines: theory and applications, pp 605–605
- Keerthi S, Sundararajan S, Chang K, Hsieh C, Lin C (2008) A sequential dual method for large scale multi-class linear svms. In: Proceeding of the 14th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 408–416
- Krishnamachari B, Iyengar S (2004) Distributed bayesian algorithms for fault-tolerant event region detection in wireless sensor networks. *IEEE Trans Comput* 53(3):241–250
- Laskov P, Schäfer C, Kottenko I (2004) Intrusion detection in unlabeled data with quarter-sphere support vector machines. In: Proceedings DIMVA, pp 71–82
- Lazarevic A, Ozgur A, Ertöz L, Srivastava J, Kumar V (2003) A comparative study of anomaly detection schemes in network intrusion detection. In: In proceedings of the third SIAM international conference on data mining
- Liu Z (2011) A method of svm with normalization in intrusion detection. *Procedia Environ Sci* 11:256–262
- Liu S, Liu Y, Wang B (2007) An improved hyper-sphere support vector machine. In: Natural computation, 2007. ICNC 2007. Third international conference on, vol. 1. IEEE, pp 497–500
- Liu C, Yang Y, Tang C (2010) An improved method for multi-class support vector machines. In: Measuring technology and mechatronics automation (ICMTMA), 2010 international conference on, vol. 1. IEEE, pp 504–508

- Luo X, Dong M, Huang Y (2006) On distributed fault-tolerant detection in wireless sensor networks. *IEEE Trans Comput* 55(1):58–70
- Ma S, Wang J, Liu Z, Jiang H (2013) Density-based distributed elliptical anomaly detection in wireless sensor networks. *Appl Mech Mater* 249:226–230
- Madden S, Franklin MJ, Hellerstein JM, Hong W (2002) Tag: a tiny aggregation service for ad-hoc sensor networks, in *IN OSDI*
- Mainwaring A, Culler D, Polastre J, Szewczyk R, Anderson J (2002) Wireless sensor networks for habitat monitoring. In: *Proceedings of the 1st ACM international workshop on wireless sensor networks and applications*, ser. *WSNA '02*. New York, NY, USA: ACM, pp 88–97. [Online]. Available: <http://doi.acm.org/10.1145/570738.570751>
- Mayoraz E, Alpaydin E (1999) Support vector machines for multi-class classification. *Engineering Applications of Bio-Inspired Artificial, Neural Networks*, pp 833–842
- McDonald D, Madria S, Ercal F, Birmingham R, Lake T (2012) Ctod: collaborative tree-based outlier detection in wireless sensor networks. In: *Proceedings of the 10th ACM international symposium on Mobility management and wireless access*. ACM, pp 1–10
- Misra P, Kanhere S, Ostry D, Jha S (2010) Safety assurance and rescue communication systems in high-stress environments: a mining case study. *IEEE Commun Mag* 48(4):66–73
- Moshtaghi M, Havens T, Bezdek J, Park L, Leckie C, Rajasegarar S, Keller J, Palaniswami M (2011) Clustering ellipses for anomaly detection. *Pattern Recognit* 44(1):55–69
- Moshtaghi M, Bezdek J, Havens T, Leckie C, Karunasekera S, Rajasegarar S, Palaniswami M (2012) Streaming analysis in wireless sensor networks. *Wirel Commun Mob Comput*
- Moshtaghi M, Leckie C, Karunasekera S, Bezdek J, Rajasegarar S, Palaniswami M (2011) Incremental elliptical boundary estimation for anomaly detection in wireless sensor networks in *Data Mining (ICDM)*, 2011 *IEEE 11th International Conference on*. IEEE, pp 467–476
- Moshtaghi M, Rajasegarar S, Leckie C, Karunasekera S (2011) An efficient hyperellipsoidal clustering algorithm for resource-constrained environments. *Pattern Recognit*
- Navia-Vazquez A, Gutierrez-Gonzalez D, Parrado-Hernandez E, Navarro-Abellan J (2006) Distributed support vector machines. *IEEE Trans Neural Netw* 17(4):1091–1097
- Nguyen X, Huang L, Joseph A (2008) Support vector machines, data reduction, and approximate kernel matrices. *Mach Learn Knowl Discov Databases*, pp 137–153
- Ni L, Liu Y, Lau YC, Patil A, (march 2003) Landmarc: indoor location sensing using active rfid, in *pervasive computing and communications*, (2003) *PerCom 2003*. In: *Proceedings of the first IEEE international conference on*, pp 407–415
- O'Reilly C, Gluhak A, Imran M, Rajasegarar S (2012) Online anomaly rate parameter tracking for anomaly detection in wireless sensor networks. In: *Sensor, Mesh and Ad Hoc communications and networks (SECON)*, 2012 9th annual IEEE communications society conference on. IEEE, pp 191–199
- Ozdemir S, Xiao Y (2011) Outlier detection based fault tolerant data aggregation for wireless sensor networks. In: *Application of information and communication technologies (AICT)*, 2011 5th international conference on. IEEE, pp 1–5
- Ozdemir S, Xiao Y (2012) Ftoda: outlier detection-based fault-tolerant data aggregation for wireless sensor networks. *Secur Commun Netw*
- Palpanas T, Papadopoulos D, Kalogeraki V, Gunopulos D (December 2003) Distributed deviation detection in sensor networks, *SIGMOD Rec.*, vol. 32, pp 77–82. [Online]. Available: <http://doi.acm.org/10.1145/959060.959074>
- Paschalidis IC, Chen Y (September 2010) Statistical anomaly detection with sensor networks. *ACM Trans. Sen. Netw.* vol. 7, pp. 17:1–17:23. [Online]. Available: <http://doi.acm.org/10.1145/1824766.1824773>
- Pham N, Pagh R (2012) A near-linear time approximation algorithm for angle-based outlier detection in high-dimensional data
- Phua C, Lee V, Smith K, Gayler R (2010) A comprehensive survey of data mining-based fraud detection research. *Arxiv*, preprint arXiv:1009.6119
- Rajasegarar S, Leckie C, Palaniswami M (2008) Anomaly detection in wireless sensor networks. *IEEE Wirel Commun* 15(4):34–40
- Rajasegarar S, Leckie C, Bezdek J, Palaniswami M (2010) Centered hyperspherical and hyperellipsoidal one-class support vector machines for anomaly detection in sensor networks. *IEEE Trans Inf Forensics Secur* 5(3):518–533
- Rajasegarar S, Bezdek JC, Leckie C, Palaniswami M (January 2010) Elliptical anomalies in wireless sensor networks. *ACM Trans. Sen. Netw.* vol. 6, pp. 7:1–7:28. [Online]. Available: <http://doi.acm.org/10.1145/1653760.1653767>

- Rajasegarar S, Bezdek J, Leckie C, Palaniswami M (oct 2007) Analysis of anomalies in ibrl data from a wireless sensor network deployment. In: Sensor technologies and applications, 2007. SensorComm 2007. international conference on, pp 158–163
- Rajasegarar S, Bezdek J, Moshtaghi M, Leckie C, Havens T, Palaniswami M (2012) Measures for clustering and anomaly detection in sets of higher dimensional ellipsoids. In: Neural networks (IJCNN), The 2012 international joint conference on. IEEE, pp 1–8
- Rajasegarar S, Leckie C, Palaniswami M (may 2008) Cescvm: centered hyperellipsoidal support vector machine based anomaly detection. In: Communications, 2008. ICC '08. IEEE international conference on, pp 1610–1614
- Rajasegarar S, Leckie C, Palaniswami M, Bezdek J (june 2007) Quarter sphere based distributed anomaly detection in wireless sensor networks. In: IEEE international conference on communications, 2007. ICC '07, pp 3864–3869
- Rajasegarar S, Leckie C, Palaniswami M, Bezdek JC (oct 2006) Distributed anomaly detection in wireless sensor networks. In: Communication systems, 2006. ICCS 2006. 10th IEEE Singapore International Conference on pp 1–5
- Ross G, Tasoulis D, Adams N (2009) Online annotation and prediction for regime switching data streams. In: Proceedings of the 2009 ACM symposium on applied computing. ACM, pp 1501–1505
- Saada M, Meng Q (2012) An efficient algorithm for anomaly detection in a flight system using dynamic bayesian networks. In: Neural information processing. Springer, pp 620–628
- Schieferdecker D, Völker M, Wagner D (2011) Efficient algorithms for distributed detection of holes and boundaries in wireless networks. *Experimental Algorithms*, pp 388–399
- Scholkopf B, Smola AJ (2001) Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT Press, Cambridge
- Schölkopf B, Platt JC, Shawe-Taylor JC, Smola AJ, Williamson RC (July 2001) Estimating the support of a high-dimensional distribution, *Neural Comput.*, vol. 13, pp. 1443–1471. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1119748.1119749>
- Schwenker F (2000) Hierarchical support vector machines for multi-class pattern recognition. In: Knowledge-based intelligent engineering systems and allied technologies, 2000. Proceedings. Fourth international conference on, vol. 2. IEEE, pp 561–565
- Shahid N, Naqvi IH (2011) Energy efficient outlier detection in wsns based on temporal and attribute correlations. In: International conference on emerging technologies
- Shahid N, Naqvi IH, Qaisar SB, (2012) Quarter-Sphere SVM: attribute and Spatio-Temporal correlations based outlier & event detection in wireless sensor networks. In, (2012) IEEE wireless communications and networking conference: Mobile and Wireless Networks (IEEE WCNC 2012 Track 3 Mobile & Wireless). France, Apr, Paris
- Shahid N, Naqvi IH, Qaisar SB (Nov 2012) Real time energy efficient approach to outlier & event detection in wireless sensor networks. In: 13th IEEE international conference on communication systems (2012) IEEE ICCS'12. Singapore, Singapore
- Shahid N, Naqvi I, Qaisar S (2012) Characteristics and classification of outlier detection techniques for wireless sensor networks in harsh environments: a survey. *Artif Intell Rev*, pp 1–36. [Online]. Available: <http://dx.doi.org/10.1007/s10462-012-9370-y>
- Sharma A, Golubchik L, Govindan R (2010) Sensor faults: detection methods and prevalence in real-world datasets. *ACM Trans Sens Netw (TOSN)* 6(3):23
- Sheng B, Li Q, Mao W, jin W (2007) Outlier detection in sensor networks
- Smola A, Schölkopf B (1998) Learning with kernels. Citeseer.
- Stankovic S, Rakocevic G, Kojic N, Milicev D (2012) A classification and comparison of data mining algorithms for wireless sensor networks. In: Industrial Technology (ICIT), 2012 IEEE international conference on. IEEE, pp 265–270
- Steinwart I, Christmann A (2008) Support vector machines (information science & statistics). *Recherche* 67:02
- Subramaniam S, Palpanas T, Papadopoulos D, Kalogeraki V, Gunopulos D (2006) Online outlier detection in sensor data using non-parametric models. In: Proceedings of the 32nd international conference on very large data bases, ser. VLDB '06. VLDB Endowment, pp 187–198. [Online]. Available: <http://portal.acm.org/citation.cfm?id=1182635.1164145>
- Suthaharan S (2012) An iterative ellipsoid-based anomaly detection technique for intrusion detection systems. In: Southeastcon, 2012 Proceedings of IEEE. IEEE, pp 1–6
- Suthaharan S, Alzahrani M, Rajasegarar S, Leckie C, Palaniswami M (dec 2010) Labelled data collection for anomaly detection in wireless sensor networks. In: Intelligent sensors, sensor networks and information processing (ISSNIP), 2010 sixth international conference on. pp 269–274

- Suthaharan S, Leckie C, Moshtaghi M, Karunasekera S, Rajasegarar S (2010) Sensor data boundary estimation for anomaly detection in wireless sensor networks. In: Mobile Adhoc and sensor systems (MASS), 2010 IEEE 7th international conference on. IEEE, pp 546–551
- Tan P, Steinback M, Kumar V (2006) Introduction to data mining. Addison Wesley, Reading
- Tax DMJ, Duijn RPW (1999) Data domain description using support vectors. In: ESANN'99, pp 251–256
- Tax D, Duijn R (2004) Support vector data description. *Mach Learn* 54(1):45–66
- Trinidad M, Valle M, Toledo C (2012) Indoors and outdoors event detection for embedded wireless sensors. *Embedded systems and wireless technology, theory and practical applications*, p 139
- Tutorial on Wireless Communications and Electronic Tracking (2009)
- Wang D, Yeung DS, Tsang ECC (2006) Structured one-class classification. *IEEE Trans Syst Man Cybern Part B Cybern* 36(6):1283–1295
- Weston J, Watkins C (1999) Support vector machines for multi-class pattern recognition. *Proc seventh Eur Symp Artif Neural Netw* 4(6):219–224
- Weston J, Watkins C (1998) Multi-class support vector machines. Citeseer, Tech. Rep
- Wu W, Cheng X, Ding M, Xing K, Liu F, Deng P (2007) Localized outlying and boundary data detection in sensor networks. *IEEE Trans Knowl Data Eng* 19(8):1145–1157
- Xie M, Hu J, Han S, Chen H (2012) Scalable hyper-grid k-*nn*-based online anomaly detection in wireless sensor networks
- Xie M, Hu J, Tian B (2012) Histogram-based online anomaly detection in hierarchical wireless sensor networks. In: Trust, security and privacy in computing and communications (TrustCom), 2012 IEEE 11th international conference on. IEEE, pp 751–759
- Xu T (2009) A new sphere-structure multi-class classifier. In: Circuits, communications and systems, 2009. PACC'S'09. Pacific-Asia Conference on. IEEE, pp 520–525
- Xue W, Luo Q, Chen L, Liu Y (2006) Contour map matching for event detection in sensor networks. In: Proceedings of the 2006 ACM SIGMOD international conference on Management of data, ser. SIGMOD '06. New York, NY, USA: ACM, pp. 145–156. [Online]. Available: <http://doi.acm.org/10.1145/1142473.1142491>
- Xu T, He D, Luo Y (2007) A new orientation for multi-class svm. In: Software engineering, artificial intelligence, networking, and parallel/distributed computing, 2007. SNPD 2007. Eighth ACIS International Conference on, vol. 3. IEEE, pp 899–904
- Yang Z, Meratnia N, Havinga P (dec 2008) An online outlier detection technique for wireless sensor networks using unsupervised quarter-sphere support vector machine. In: Intelligent sensors, sensor networks and information processing, 2008. ISSNIP 2008. International conference on, pp 151–156
- Yeung D, Wang D, Ng W, Tsang E, Wang X (2007) Structured large margin machines: sensitive to data distributions. *Mach learn* 68(2):171–200
- Yozo CP, Hida Y, Huang P, Nishtala R (2004) Aggregation query under uncertainty in sensor networks. Tech, Rep
- Zhang Y (June 2010) Observing the unobservable - distributed online outlier detection in wireless sensor networks. University of Twente, Ph.D. dissertation
- Zhang Y, Meratnia N, Havinga P (2010) Outlier detection techniques for wireless sensor networks: a survey. *IEEE Commun Surv Tutor* 12(2):159–170
- Zhang X, Gu C, Lin J (2006) Support vector machines for anomaly detection. In: Intelligent control and automation, 2006. WCICA 2006. The sixth world congress on, vol. 1. IEEE, pp 2594–2598
- Zhang Y, Hamm N, Meratnia N, Stein A, van de Voort M, Havinga P (2012) Statistics-based outlier detection for wireless sensor networks
- Zhang Y, Meratnia N, Havinga P (2009) Adaptive and online one-class support vector machine-based outlier detection techniques for wireless sensor networks. In: Proceedings international conference advanced information networking and applications workshops WAINA '09, pp 990–995
- Zhang Y, Meratnia N, Havinga P (2012) Distributed online outlier detection in wireless sensor networks using ellipsoidal support vector machine, *Ad Hoc Networks*
- Zhang Y, Meratnia N, Havinga PJM (July 2009) Hyperellipsoidal svm-based outlier detection technique for geosensor networks. In: Third international conference on Geosensor Networks, Oxford, UK, ser. Lecture Notes in Computer Science, vol. 5659. Berlin: Springer, pp 31–41
- Zhang Y, Meratnia N, Havinga PJM (November 2007) A taxonomy framework for unsupervised outlier detection techniques for multi-type data sets, <http://eprints.eemcs.utwente.nl/11366/>, Centre for Telematics and Information Technology University of Twente, Enschede, Technical Report TR-CTIT-07-79
- Zhang K, Shi S, Gao H, Li J (2007) Unsupervised outlier detection in sensor networks using aggregation tree. In: Proceedings of the 3rd international conference on advanced data mining and applications, ser. ADMA '07. Berlin, Heidelberg: Springer, pp 158–169. [Online]. Available: <http://dx.doi.org/10.1007/978-3-540-73871-816>

- Zhuang Y, Chen L (2006) In-network outlier cleaning for data collection in sensor networks. In: In CleanDB, workshop in VLDB. APPENDIX, pp 41–48
- Zhu M, Wang Y, Chen S, Liu X (2003) Sphere-structured support vector machines for multi-class pattern recognition. *Rough Sets, Fuzzy Sets, Data Mining, and Granular, Computing*, pp 589–589
- Zoumboulakis M, Roussos G (2007) Escalation: complex event detection in wireless sensor networks. In: *Smart sensing and, context*, pp 270–285