International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-7 Issue-4, September 2018

One class SVMs Outlier Detection for Wireless Sensor Networks in Harsh Environments: Analysis

Bhanu Chander, Kumaravelan

Abstract*: Outlier/Anomaly detection is renewed challenge in data mining, internet of things as well as machine learning communities. In present era Internet of things is emerging with its tremendous applications where wireless sensor nodes are incredibly well structured to accumulate huge amount of raw data from unsystematic sectors and hand over it to authoritative systems such as disaster monitoring, surveillances, green monitoring, and smart city applications etc,. However such authoritative and prediction systems truthfulness subject to reliability of sensor node. Unluckily, sensed data excellence and reliability influenced by circumstances such as sensor faults, intrusion and unusual events within others. As a result it obstructs authoritative decision making as well as prediction, hence there is need of effectual, real time abnormality detection mechanisms for consistent decisions. A key dispute is how to lessen energy consumption and communication overhead in network at the same time identifying anomalies in unsystematic environments. Even though a impressive number of studies, existing anomaly detection algorithms are there still Machine learning numerous appliances has captured massive importance in outlier detection especially for wireless sensor networks (WSNs), notably Support Vector Machine (SVM) based techniques provides effectual outlier detection and classification achievements in harsh environment. This work presents various one class SVM formulations eminently well instructed outlier detection in harsh environments, moreover formulations analyzed in terms of various characteristics include input data, dynamic topology, outlier types, Spatio temporal attribute correlations etc. Brief comparison and characteristics of distinctive one class SVM formulations are described.

Index Terms*: Wireless Sensor Networks, Outlier Detection, Classification, Support vector Machine, Event Detection.

I. INTRODUCTION

Wireless sensor networks (WSNs) is collection small, low cost, limited power with multifunctional devices or sensors which are allied with one sink or more than one sink nodes which have superior computational capacity. Sink is also known as Base station and it is high power-driven device allied with data base devices through satellite links. Each node in WSN network typically equip with a small micro controller, a power source, a wireless radio transceiver, and one or more number of sensors like humidity, temperature and pressure etc. Furthermore each sensor node incorporates with analog to digital converter or digital to analog converter (ADC/DAC). Moreover particular sensor node frame-in with a series of network services in particular area of coverage, data compression, synchronization, localization, data aggregation and security appliance [1-3].

The majority of the WSNs appliances relay on three crucial aspects those are localization, time synchronization and state estimation of sensor node. Node localization means disclose the locality of sensor node. In numerous appliance scenarios locality is exceptionally important to disclose the locality of sensor node based on network transportation. Sensor nodes forward signals to base station, upon that signals method locate sensor nodes locality. Sensor network reliability is influences when its sensor clocks are not synchronized well. Whereas authors programmed populous statistical rules for evaluating clocks off sets, clock skew parameters, capitalize those values to evaluate time synchronization. Probabilistic models have been programmed in favor of proficient, healthy state evaluation in support continuous and discrete time in WSNs similarly Innovation based state estimations in WSN, distributed state estimation, energy efficient state estimation, constrained state estimation for localization, data fusion and collaborative state estimation, micro grid state estimation. WSNs can be used for both continuous time as well as discrete time monitoring. In continuous time WSNs, the time points $t_1, t_2, t_3, t_4, \ldots \ldots t_n$ are integers. Coming to discrete time WSNs, these points are turned to real numbers. For the most part of applications quantity being measured as continuous like humidity, temperature and room pressure. However a sensor node for aggregating, storing continuous measurements exhausts higher energy which is not suitable where sensor energy is constantly limited. Some sensor nodes fabricated in a way that they accumulate continuous estimations then they extract binary values, this done based on some prefixed threshold values. Some other sensor nodes accumulate estimations at all the time although it will store and transmit essential estimations only to the approved base station [1],[2], [16],[17,[19].

At the moment WSNs are set up in an outsized sector to monitor dangerous environmental constraints like rainfall, forest fire, waterfall, and sea water levels, underground oil caves humidity, soil moisture and temperature. The composed sensor data used to various expert systems to support decision making, predict the event occurrences. In general such expert systems are helpful where the events prediction cannot be made in advance furthermore events those under the examinations modify rapidly over time. In order to provide a true overview about observed or surrounding environment, the sensed data should be high quality and reliable.

Revised Manuscript Received on 30 September 2018.
* Correspondence Author
Bhanu Chander*, Department of Computer Science and Engineering, Pondicherry University, Pondicherry (Tamil Nadu), India.
Kumaravelan, Department of Computer Science and Engineering, Pondicherry University, Pondicherry (Tamil Nadu), India.

Unhappily internal as well as external boundaries of sensors in provisions of bandwidth, low power, storage and processing capabilities and deployed exposure environmental conditions compose them to dissimilar kinds of attacks like hardware as well as software flaws and malicious attacks. These results to low quality data which contaminated anomalous assessments. Moreover some unexpected events that may possibly happen in real world will influence regular collected sensed data. Hence there is need to reveal such kind of anomalous definitely and resourcefully in order to sustain the quality of sensed data along with investigates the unusual events [3], [5].

The consistency of sensor node is also influence by one or more anomalous node attendance. Anomalous node which behave completely different from other normal sensor nodes in that network region. Infrequently, an external user can formulate a normal node to anomalous; collect sensed information from the network. Those anomalous nodes may transfer possessed data or incapable to communicate other nodes in that network structure. For many reasons preliminary detection of corresponding nodes is essential. Naturally irregular nodes figure out as a result of its proceedings or activities in comparison with its bordering nodes or with other nodes.

II. HARSH ENVIRONMENTS

A harsh environment has high stress and that presents rigorous monitoring and communication challenges. Examples for such environments are volcanic sites, underground oil, Gas, salt; gold, coal mines, deep forest regions etc. In underground gas mines we have to continually observe, visualize for any harmful, unfavorable conditions to secure human resources working inside. There are different issues which disturb the connections and monitoring capabilities of sensor nodes in harsh environments such as underground topology, noise, ionized air, high signal attenuation, multiple reflections, and instability in mine structures. Those constrains mentioned above partition into two major categories based on WSN prospective those are environmental monitoring constraints and communication constraints. Environmental monitoring constraints are straightforwardly associated with quality of data in sensor node. Monitoring constraints institute by topology changes of the network, changes in various attributes, unsteadiness or fluctuation in mine structures. Subsequent to communication constraints which are mostly generate by large propagation delay, path loss, spreading, signal loss, variations in time channels etc. these all constraints are interrelated to communication channel of WSN which impact the connection between sensor nodes [1],[2],[3],

III. OUTLIER AND EVENT

The foremost explanation regarding outlier is expressed by Grubbs in 1994 “an outlier or outlying observation is one that appears to deviate markedly from other member of samples from which it occurs”. Here we briefly characterize outlier as well as event terms, an event is disastrous condition which characterize by unanticipated changes in environmental conditions. Compare to outlier, event probability of occurring is very low. Examples for events are earthquakes, forest fire, and unexpected gaseous rise in underground gas mines. Outlier has some nearest relation with events and defined as an observation which is completely diverging from normal set of readings compared to its surrounded ones. Outlier must be inspecting it may perhaps point out a situation touching close to event or natural failure. Most likely resources regarding outliers in records gathered from WSNs are noise or errors, malicious attacks plus events. Noise or errors are result of fault sensor nodes as much as possible we can eliminate noise by applying some filtering techniques. Malicious attacks steal the information with and without disturbing the network. An event can defined as sequences of outlier or erroneous readings in streaming data. Event detection scheme indicated by outlier detection schemes for getting additional idea about this, a sensor reading showing high air pressure independently than its neighboring nodes is an outlier, whereas group of sensor nodes positioned close jointly point out continuous high air pressure indicates the occurrence of an event. [1], [5], [6],[11],[13],[14].

The intention in relation to outlier, event recognition system in favor of WSN installed in any harsh environment is for outlier finding, event revealing moreover event identification. The process of outlier detection is separating standard and abnormal sample from composed sensed data. The process of event revealing determines the sources of outlier whether it is due to sensor fault or event indication. The process of event identification determines nature of event along with attributes which are governing toward create that event.

A.Outlier and Event Detection Technique

Here we providing various outlier detection techniques especially designed for WSN

Statistical based techniques: Consider a statistical model for data distribution, based on their probability generated from assumed model outliers are declared.

Cluster based techniques: It groups similar or close data samples as clusters, cluster center detect outlier based on some measurements from it. Mostly mahalanobis distance measure is used to cluster data points.

Nearest neighbor based techniques: Distance based metrics are accomplished to detect outliers. Generally Euclidian and mahalanobis metrics are used.

Classification based techniques: Classification procedures are ultimate suitable in outlier determination for WSNs rather than all other techniques. It learns a classifier in training mode, classify unseen data samples in testing phase using learned model. Support vector machine based methods are mostly used in classification.

B. Characteristics of Outlier Detection Techniques

Multivariate data: Attributes considered by various sensor nodes must be taking for consideration. Operations performed on continuous arriving data.

Attribute correlation: Consider all the correspondences among discrete attributes found in one node.

Temporal correlation: Consider both historical data sample with newly arrived data samples.

Spatial correlation: Recognize interrelationship over data fragments of topographically scattered sensor nodes.

Distributed processing: In network all nodes have equal distributed processing.
Online data processing: It should progression on continuous streaming data as quickly as it came.
Unsupervised Nature: Techniques should learn data models from training with unsupervised or unlabelled data. If model learn from labeled data it will not suitable for WSN, because of its dynamic nature.
Event identification: If any event occurs, it should able to identify the type of event.
Low complexity: Techniques should have low communication and computation cost among sensor nodes in network.
Local/global outlier: Method should clearly notice local and global outliers.

C. Methodologies for Classification Techniques

The most frequently used schemes to study the success of any classification techniques are evaluate quantitative operations by experimenting on different data sets. Some algorithms perform well in particular or specified regions and those algorithms will not work properly in other region because difference of experimental setups and atmosphere. For fair comparison between event and outlier detection techniques there is need of some qualitative measure which will produce well defined characteristics for differentiation between event and outlier detection techniques [6], [7], [8].

Quantitative measurements: Detection rate (DR), false positive rate (FPR) and Receiver operating characteristic curves (ROC) are three main operations to evaluate success of any abnormality/outlier techniques. DR is nothing but how much data is correctly scaled as abnormal/outliers as a result of outlier detection technique. In any application technique detection rate must be high, a technique with low detection rate will not more useful in harsh environments. False positive rate can be defined as ordinary data must not classify like outliers; the amount of ordinary data revealed like outliers is called FPR. FPR have to exist as lowered as possible. Even though tremendous detection rate necessary for practices but it must not allocate normal data as outliers. Receiver operating characteristic curve shows the relation flanked by detection rate with false positive rate. An ROC curve stands a plot that point out deviations of detection rate through false positive rate, area under the ROC curve must exceptionally close up to unity.

Quantitative measurements: Qualitative performance measures used to analyze the feasibility of any event as well as outlier detection technique for WSNs. It analyze summary of techniques in presence or absence significant distinctiveness moreover conclude the technique used for harsh atmospheres or not. The technique applies for harsh atmosphere following a few alterations, or technique not applicable for harsh atmospheres, or finally the technique used only for harsh environments. Feasibility decision resting upon the foundation regarding presence or absence of separate distinctiveness within applied technique. We divide quantitative measurements into two categories: First Category is Essential Feasibility--spatio-temporal attributes correlation, Distributed approach, Topology changes, Multivariate streaming data, interference to temporal and spatial non stationary. Second category is Non-Essential Feasibility—low complexity, online computations, global outliers.

IV. CLASSIFICATION BASED APPROACH FOR OUTLIER AND EVENT DETECTION FOR WSNS

Classification base techniques are most preferable on behalf of anomaly recognition in machine learning and data mining communities. Classification techniques are likely to develop classification learner or replica for the duration of preparation phase with labeled data headed for classify any unnoticed data during testing phase. Remaining outlier or event detection techniques uses distance metrics, pre labeled input output data samples, data distribution probabilities, fuzzy logic systems which consume more energy, storage space, computational complexities. The successes of classifier build upon generalization means capability to categorize unnoticed data through the learned model. Subject to category of model/replica learning in training/preparation phase, classification procedure separated into two categories. Bayesian model classification and Support Vector Mechanism based classification [2], [4].

D.*Bayesian model classification:* Compare to further machine learning models Bayesian model classification need less number of training samples. Bayesian model utilizes probability distribution models to represent variables and their probabilistic dependencies devoid of over fitting. The learned model has a mixture of nodes attached through edges represent their beliefs on rest. Learned methods find out the possibility of recently arrived data sample fit in to appropriate class. Bayesian techniques also acknowledged as soft-decision procedures for the reason that rather than hard decision they will only contribute intelligence regarding data samples. There are some disadvantages also there; Bayesian models are not either online or adaptive so to learn a model outsized quantity data samples are needed and if there is any altering take place training phase must repeat/change. On the road to find out a probabilistic replica in preparation/training phase Bayesian model suffer with significant amount of computational overhead. This computational overhead in learning phase may reduce by utilizing Naïve Bayesian model but this model again suffering with loss of correlation among various attributes. Moreover due to dynamic changes in WSNs, learned graphical model repeat time to time [1],[3], [4]

E. *Support Vector Machine: Support* vector machine (SVM) techniques learn a classification model during training phase moreover used that model to classify any unseen or new arrived data. Briefly SVM makes a maximum margin hyper plane among different classes. This hyper plane separate maximum distances data fit in to different destination classes. Compared to Bayesian models, these techniques show some great generalization ability, reason is by using kernels SVM maximize separation between different classes efficiently. Computational complexity of SVM techniques is low compared to Bayesian models since they do not discover entire figurative model.SVM based classification techniques can be categorized into one-class SVM, Binary class SVM and Multi-class SVM [9],[11],[14],[20].

V.INTRODUCTION TO NOTATIONS USED IN SVMS

Generalization ability: Generalization ability of classifier is nothing but how well classifier parameter trained all through the preparation phase is proficient to do partition of newly arrived or unobserved data. Performance analysis of any classifier is depends on testing phase which depends on parameters learned in training phase.

Maximum margin: one-class SVM must separate ordinary data form unobserved/unusual data correctly,

this is done by highest border among ordinary and unusual data. The parameter formulation standardized in such a way to make easy separation of ordinary and unusual sample is maximized. As a result of resolve a few optimization problems under a few limitations will define parameters which corresponding to maximum margin.

Kernel functions: kernels are balanced clear semi-definite specific objectives which evaluate dot product of vectors in feature space furthermore define as K(x,y) = (Q(x), Q(y)). Linear, polynomial Gaussian, mahalanobis are various kernels presently working.

*Model parameters***:** model parameters in favor of one-class SVM explain in parameters which express formulations. Hyper plane formulations express through its weight vector w and bias parameter r. hyper sphere formulation express via its radius r plus center c. while hyper ellipsoid formulation express through center of cluster c_i and their Radii r_i . Quarter-sphere as well as centered ellipsoids define through their Radii r only. All of these parameters are need to solve optimization problem.

*Training and testing phase***:** Model parameters of cluster can regulate via few model data samples. One class SVMs are unsupervised reason is they use unlabelled data in training means they need input data sample not including labeling into ordinary and unusual data. Multi class SVM is supervised because they need labeled input out data sample in training phase.

*Regularization and slack variable***:** the complexity of one class SVM is depends on the quadratic expansion dilemma on the way to make most favorable result. Favorable solution calculates on how effortlessly ordinary, unusual data are distributed. On the way to decrease complication we introduce some miss classification to learner which is restricted via regularization parameter. The charge of miss classification of complete data samples is sum of slack variable. If slack variable cost low it results into high detection rate, false positive rate of classifier. If slack variable cost low it results into low detection rate, false positive rate of classifier.

Support vector, Non-support vector and Boarder support vector: the data sample or vectors which are inside the geometric region along with their correlated Lagrange multiplies enclose zero significance called non support vectors. The data examples or vectors which are outer the geometric region recognized as support vectors. The data examples which rest on the geometric region call as boarder region vectors.

Euclidian distance: Euclidian space for every data illustrations form a point c is denoted like

$$\text{ED (z)} = \sqrt{((x-c)^T(x-c)}$$

If we are using Euclidian distances to measure distances between data sample for outlier or event detection in support vector machine it will results in form of spheres data distribution.

Mahalanobis distance: Mahalanobis space for every data illustrations from a point c is denoted like

$$\text{MD (z)} = \sqrt{(x-c)^T\,\Sigma^{-1}(x-c)}$$

If we are using mahalanobis metric to measure the distances of data samples for outlier or event detection in support vector machine it will form a geometric region of ellipsoid data distribution.

ONE-CLASS SUPPORT VECTOR MACHINES

Hyper plane formulation: Primary classification regarding one-class SVM is Hyper plane SVM; it formulates the highest boundary among ordinary and unusual data from origin. Hyper plane SVM described through its weight vector w along with bias vector r. hyper plane mainly suffer from quadratic optimization problem which involves reduce of weight vector [6],[19].Overall problem tending toward shrink the space of hyper plane from the origin on the way to conclude the optimal w as well as r via allowing some miss-classification data samples.

$$\text{Min } \frac{1}{2}\|w\|^2 + \frac{1}{vn}\sum_{i=1}^{n}\xi_i - r$$

$$\text{Subject to: } (w, \phi(x_i)) \geq r - \xi_i ,$$

$$\xi_i \geq 0$$

The terms on right hand side indicates the regularization which controls miss classification cost, slack variable allows some of anomaly variables to fall on wrong side of the hyper plane. Dual form of problem is as follows

$$\text{Min } \sum_{ij} \alpha_i \alpha_j \, k(x_i, x_j)$$

$$\text{Subject to } \sum_{i=1}^{n} \alpha_i = 1$$

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

Here α are Lagrange multipliers and k (x_i, x_j) is inner dot product of vectors in feature space

Yashwant singh et (2013)in [13],—modeled a distributed fixed partitioning SVM (DFP-SVM) for handle constant data proficiently in favor of notice events especially in distributed manner to reduce energy efficiency in WSN nodes. Here training sample are divided into fixed size partitions and rather than sending entire information, each cluster has incremental step which will send fractional information means in each step it will send current estimation rather than whole data. Consider there are N cluster heads in WSN, evaluation at one cluster depend on evaluation of preceding CH and all the data samples at that cluster head.

Hyper sphere formulation: Hyper sphere SVM formulates a sphere of radius r and center c in high dimensional space to separate normal and abnormal data sample [9],[10]. In this method we use Euclidian distance measurements where greater part of data sample is identical to radius R, some amount of miss classification is allowed.

$$\text{Min } R^2 + \frac{1}{vn}\sum_{i=1}^{n}\xi_i$$

$$\left\|\Phi(x_i) - a_\Phi\right\|^2 \leq R^2 + \xi_i,\ \xi_i \geq 0$$

Right hand side term indication for regularization which controls miss classifications in hyper sphere and overall equation reduce radius R regarding hyper sphere and determine R and c in such a way allowing only some miss-classification. The dual formation is written as

$$\text{Min } \sum_{ij} \alpha_i \alpha_j \text{ k } (x_i, x_j) - \sum_i \alpha_i \text{ k } (x_i, x_i)$$

$$\text{Subject to: } 0 \le \alpha_i \le \frac{1}{vn}, \quad \sum_i \alpha_i = 1$$

From the above equations α indicates Lagrange multipliers, k (x_i, x_j)indicats dot product of vectors in feature space.

Support vector data description (SVDD): SVDD is a classifier which aims to construct hyper-sphere with minimum value that contains all the data samples [15].

SVDD has more flexible boundary value than hyper sphere boundary value to adapt irregularly shape target data sets that will help to extract normal data from anomalous data. Given X= $\{ x_1 x_2 x_3 \ldots . x_n \}$ a set of training data where $x_i \in R^d$ $(1 \le i \le l)$ correspond to d –dimensional records with l is size of training records. Then SVDD shows like follow.

$$\text{Min } R^2 + C \sum_{i=1}^{l} \xi_i$$

$$\text{Subject to } \|x_i - a\|^2 \le R^2 + \xi_i, \quad \text{where i = 1, 2, 3.....} l$$

$$\xi_i \ge 0, \text{ where i = 1,2,3….} l$$

Here (r and a) indicates of radius as well as center of hyper sphere in feature space, here symbol ξ_i is the slack variable which assign some miss classification in training data. In SVDD standard class summarized from input space to feature space via mapping function ɸ (.) The intention of this is in the direction of formulate samples extra solid in feature space compare to input space. In feature space inner/dot products of two vectors can be determine via implementing mercer kernel tasks.

$$\text{k } (x_i, x_j) = \phi(x_i).\phi(x_j)$$

Lagrange's function constructed as

$$\text{L } (\text{R}, a_\phi, \xi_i, \alpha_i, \gamma_i) = R^2 + C \sum_{i=1}^{l} \xi_i$$

$$- \sum_{i=1}^{l} \alpha_i (R^2 + \xi_i - \|\phi(x_i) - a_\phi\|^2) - \sum_{i=1}^{l} \gamma_i \xi_i$$

Here $\alpha_i = (\alpha_i, \alpha_i \ldots \ldots \alpha_i)^T \ge 0$ and $\gamma_i = (\gamma_i, \gamma_i \ldots \ldots \gamma_i)^T \ge 0$ are Lagrange multipliers.

Zhen Feng et al (2017) in [15] proposed a method SVDD based anomaly detection technique. On the way to decrease the computational complication of SVDD train plus test phases authors used SMO based second order approximation to reduce computation in training phase moreover quick decision based rules for unobserved data sample in train phase is implemented to accelerate the training speed. SMO algorithm is used for two reasons first thing improve Lagrange multiplier to violate KKR (Karush Kuhn Tucker) plus to meet up KKR circumstances another one is difficulty of functioning set collection. Initialize slack variable, kernel of SVDD; solve the QP problem by applying SMO method on second order approximation. Compute radius, weight vector and estimate pre-image of hyper sphere center. Then classify the unseen data based on the calculated radius, weight, image vectors. Van vuongntrinh et al (2017) in [23] modeled a SVDD based outlier detection applying Mahalanobis kernels. Mahalanobis kernels provide better measurements than RBF based kernels. In this method SVDD established a discriminative function for progressively detect outliers. Those data sample which lies inside the effective region named as normal, outside the region data sample named as abnormal or outliers.

Quarter sphere Support vector Machine (QS-SVM): QS-SVM is prepared in the direction of lower computational complication from hyper spherical SVM with removing primal variable c, moreover QS-SVM require single set of Lagranges multiplier standards for its dual problems where hyper plane need two sets. Consider a data vector $X_j = \{X_i$, n= 1, 2, 3,…n} mapped to higher dimensional feature space. The mapped vector $\phi(xi)$ in feature space is called as image vectors. Optimization problem of QS-SVM could be formulated as

$$\text{Min } R^2 + \frac{1}{vn} \sum_{i=1}^{n} \xi_i$$

$$\text{S. to : } \|\phi(x_i)\|^2 \le R^2 + \xi_i,$$

$$\xi_i \le 0$$

Here ξ_i i = 1,2,3 … n slack variable, it allows some misclassification to the classifier means some image vector lie outer the sphere.. Regularization parameter use to control image vectors that lie outer sphere. The Lagranges task for this optimization written as

$$\text{L } (\text{R}, \xi_i, \alpha_i, \gamma_i) = R^2 + \frac{1}{vn} \sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} \gamma_i \xi_i - \sum_{i=1}^{n} \alpha_i (R^2 - \|\phi(x_i)\|^2 + \xi_i$$

Here $\alpha_i \ge 0, \gamma_i \ge 0, \forall i$ are the Lagrange multipliers Corresponding dual form written as

$$\text{Min } - \sum_{i=1}^{n} \alpha_i \text{ k } (x_i, x_j)$$

$$\text{Subject to } \sum_{i=1}^{n} \alpha_i = 1$$

$$0 \le \alpha_i \le \frac{1}{vn}, \qquad i = 1 \ldots n$$

Here dual problem is linear optimization so $\{\alpha_i\}$ can easily obtain from available linear optimization. Compared with some existing SVM which suffer from quadratic optimization problem increase computational complexity, coming to QS-SVM has advantage of linear optimization to solve computational complexity.

Rajasegarar et al (2007) in [6] stated a SVM based distributed anomaly detection method in offline network which involve low communication among nodes in the set of connections. Every node individually implements quarter sphere abnormality uncovering innovation to characterize quarter sphere border line those record points lie outside quarter sphere boundary are named as anomalies. Each node calculate local radius moreover maintain testimony of norms in its memory and send it to parent node. Parent node calculates its own radius record and combines it with collected records to form a global radius then send back to children nodes to determine global radius values. Yang et al (2008) in [23] presents online QS-SVM based anomaly detection through spatial correlations of adjacent sensor nodes. Method also provided notification regarding real time difference among events and errors. Zhang et al (2009) in [21] promoted three methods namely fixed size time window based outlier detection technique (FTWOD). Instant outlier detection (IOD), Adaptive outlier detection (AOD). Every method has its own training time periods and detection rates. Shahid et al (2012) in [19] described adaptive-online QS-SVM procedure where temporal connections identify outliers in addition to spatial connections identify events. Normally Radius of quarter sphere was resolve through temporal correlation attribute.

Hugo martin's et al (2015) in [11] established method online outlier detection with supportive least square SVM under consideration of Kernel Hilbert Space with radial basis function all of this performed through sliding window based learning.

Hyper ellipsoid formulation: Spherical one class SVM provides satisfactory results when sample have same distribution tendency in all the directions. So there is some chance that normal data can mistakenly identify as outliers. Hyper ellipsoid formulation can be effectively identifies outlier by using minimum radii throughout the greater part of image vectors inside feature space. Image vectors forms discrete clusters and then fixed with hyper ellipsoid that can encapsulate a greater part of image vectors within that cluster. Those image vectors which are not fall in any of constructed hyper ellipsoids are called as anomalous values.

More importantly hyper ellipsoid formulation utilizes Mahalanobis gap measurement concerning data sample from center of clusters en route for identify anomalous values [9],[10],[12].

Presume m data vectors regarding d variables in input space using some mapping function variables mapped into feature space. Hyper ellipsoid SVM functions the majority of mapped data vectors into feature space by making ellipsoid center at origin with minimum effective radius.

$$\text{Min } r_i + \frac{1}{vn}\sum_{i=1}^{n}\xi_i$$

$$\text{Subject to: } \left\|\Sigma_i^{-\frac{1}{2}}(x_j - c_i)\right\| \le r_i + \xi_i,\ \xi_i \ge 0$$

D.Wang, Yeung et al (2006) in [24] designed a method hyper ellipsoid one class SVM for anomaly detection. Multiple hyper ellipsoids are fitting through smallest efficient Radii, these data points fitted outer of calculated radii named outliers. Although it shows good classification results but suffering with quadratic optimization. Yang Zhang et al (2012) in [12] Distributed online outlier detection in WSN using ellipsoidal SVM, Ad Hoc Networks, Elsevier B.V- described a online and distributed method of hyper ellipsoidal one class SVM technique for anomaly detection by considering correlation among sensor data attributes. To resolve changed sensor data samples behavior they consider theory of spatio-temporal correlation.

Centered ellipsoidal formulation: Centered ellipsoid formulation aims toward fitting a hyper ellipsoid in feature space through least efficient Radius (R>0). Unlike other SVMs except QS-SVM, centered ellipsoidal SVM originate like linear optimization difficulty since it requires commitment of radius only.

$$\text{Min } R^2 + \frac{1}{vn}\sum_{i=1}^{n}\xi_i$$

Subject to: $\Phi(x_i)\,\Sigma^{-1}\,(\Phi(x_i))^T \le R^2 + \xi_i,\ \xi_i \ge 0$, i =1, 2, …m

Here manage the mapped data vectors to become as outliers. ξ_i Slack variable which allow some misclassification to classifier means it maps some data vectors to be outside hyper ellipsoid region. Σ^{-1} opposite of covariance matrix Σ of mapped data vectors.

$$\Sigma = \frac{1}{m}\sum_{i=1}^{n}(\Phi(x_i) - \mu)\,(\Phi(x_i) - \mu)^T,$$

$$\mu = \frac{1}{m}\sum_{i=1}^{m}\Phi(x_i)$$

Lastly dual form hyper ellipsoidal SVM turns to continuous optimization issue characterize as

$$\text{Min } - \sum_{i=1}^{n}\alpha_i\ \left\|\sqrt{m}\,\Omega^{-1}A^T K_c^i\right\|^2$$

$$\text{Subject to: } \sum_{i=1}^{n}\alpha_i = 1,$$

$$0 \le \alpha_i \le \frac{1}{vn},\ \text{i= 1, 2, 3, …. m}$$

Data vectors with $\alpha = \frac{1}{vm}$ those data vectors space to the origin are superior to effective radius R from hyper ellipsoid named as Outliers. The data directions along $\alpha_i = 0$ sinking within the ellipsoid region is named as Normal data vectors. The data directions along $0 \le \alpha_i \le \frac{1}{vn}$ those lies on the surface of ellipsoid are called as margin support vectors.

Rajasegarar et al (2012a) implemented a method same as QS-SVM, radius of children nodes send to parent node and parent node calculates global radius, sent back to children nodes for outlier classification. However hyper ellipsoidal has more communication than QS-SVM reason is it needs to broadcast covariance matrix to adjacent nodes. Centered node with the help of this broadcasted covariance matrix generate a global covariance matrix. Rajasegarar et al (2012) in [10] presented CE-SVM based method which is unsuitable for WSNs because it require to send radius plus covariance matrix to other nodes regularly. Need to compute mahalanobis distances for each new arrived data sample.

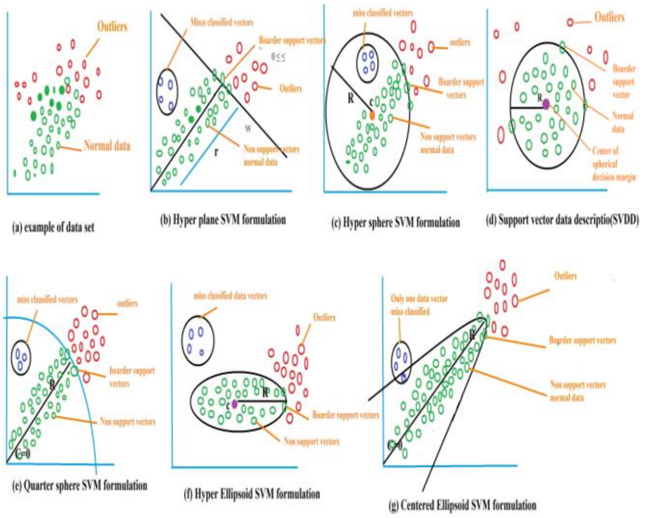

Figure 1. One Class SVM Formulations

VI. COMPARISON OF ONE CLASS SUPPORT VECTOR MACHINES

Every Support vector machine namely hyper-plane, hyper-sphere, and Quarter-sphere, hyper-ellipsoid, centered ellipsoid SVMs are modeled to differentiate normal data and abnormal data samples by making unlabelled data into geometric region [6],[9],[11],[12],[13],18],[20],[21],[22].

1) Hyper plane, hyper sphere and hyper ellipsoidal SVM be described via two parameters weight vector w as well as bias parameter r, therefore they require the solution of quadratic optimization problem. On the other side remaining Quarter sphere SVM, Centered ellipsoidal SVM are defined only by their respective Radii, which makes them to involve linear optimization problem. So QS-SVM and CE-SVM has less computational complexity and feasible for implementation in WSNs.

2) Ellipsoidal SVM formulation shows better generalization abilities compare to hyper plane and hyper sphere formulations because of their distance metrics used in these formulations. Euclidian distance makes spherical shape data distribution and all data samples which are anomalous but not as far from centered may lead to miss-classification. Coming to Mahalanobis distance makes ellipsoidal form of data distribution in tight ellipse whose shape depends on covariance of data set. Many researchers prove that mahalanobis distance obtain better results than Euclidian distances.
3) Hyper plane, Hyper sphere and Quarter sphere based SVM uses Euclidian distance metrics for outlier detection and hyper ellipsoidal, centered ellipsoidal SVM uses Mahalanobis distance based metrics to detect outlier detection.
4) Euclidian space measure extremely build upon dissimilarity between fixed values, coming to mahalanobis distance measure depend upon inverse of covariance matrix, which makes it to remove affective absolute value attributes. Thus mahalanobis distance denotes the deviation of a data sample from center of distribution.
5) Hyper plane SVM has poor classification and generalization abilities furthermore it associate quadratic optimization difficulty. Consequently hyper plane based formulations not suitable for power restricted WSNs those positioned in dangerous, harsh environments.
6) The hyper sphere SVM formulation has somehow perform good generalization ability as well as classification performances compared with hyper plane SVM, however it is not so feasible to implement on energy constrained WSNs, and reason for this is like hyper plane SVM it also suffer with quadratic optimization problem.
7) The Quarter sphere SVM has better classification performance than hyper plane as well as hyper sphere moreover QS-SVM involve with linear optimization problem which reduces the computational complexity. So considering all these reasons QS-SVM is feasible to implement on WSNs for outlier and event detection.
8) Hyper ellipsoid formulation also suffer with quadratic optimization problem so it will not used in WSNs. but it is better abilities than hyper plane and hyper sphere formulation. Centered ellipsoidal formulation has better generalization ability as well as classification performance than all other one class SVM, moreover it involve with linear optimization problem, Centered ellipsoidal SVM are best suited for WSNs deployed in inconsistent, unstructured environments for outlier end event detection.
9) All the existed QS-SVM based outlier detection techniques multivariate as well as online streaming data; moreover all QS-SVM techniques consider spatio temporal correlations of sensor nodes so they can handle both local and global outliers. But out of them some techniques only consider attribute correlations for better outlier and event detection performance. QS-SVM techniques mostly performed in distributed processing and perform with unsupervised mode means they need only unseen data in training phase. Compared to sphere, plane SVM, QS-SVM detection performance is good but none of QS-SVM techniques mention clearly about type of event and attributes involved in it. Most of the QS-SVM techniques are online mode but some of in offline mode, online mode streaming data QS-SVM based techniques are best suited of WSN installed in harsh environments because of it handle dynamic changes of nature.
10) Centered ellipsoid SVM based approaches consider both multivariate and streaming data likewise QS-SVM, CE-SVM based techniques also consider spatio temporal correlations an important advantage over all one class SVM techniques is that CE-SVM considers the attribute correlations of data via covariance matrix and it requires additional computational and storage memory for calculating covariance matrix. Thus CE-SVM techniques based on sensor temporal characteristic correspondence notice local outliers in addition to based on sensors spatial correspondences notice events and global outliers. All of CE-SVM based techniques involve in linear optimization problem in order to reduce computational cost but problem is CE-SVM techniques has high communication cost reason is they need to send covariance matrix to all the sensor nodes in that network there are some principle are added to reduce the communication cost but they are not perfectly suited for online outlier, event detection functionalities. Likewise QS-SVM, CE-SVM approaches also did not mention clearly about kind of event along with characteristics involved in it.
11) Hyper plane, hyper sphere has good classification accuracy but computational complexities to solve quadratic optimization not allow them to apply on WSNs. So they can efficiently apply on outlier detection in WSNs when they have a method to trim down the complexity. Iterative re-weights least squares, SMO algorithm, error correction are some methods to trim down the complexity in hyper plane.

VII. RESEARCH PROBLEMS

From all the above mentioned notes that we discussed till now shows Support Vector Machines satisfy all the conditions and requirements for effective event as well as outlier detection but there are some limitations which are not explored by research communities to apply them on energy constrained WSNs. Here we providing some research problems research community may focus on it while working on SVM.

1) SVM based event and outlier detection techniques in WSNs separate normal and abnormal data from collected sensor data surrounded environmental region but they did not mention event strategies. Event strategy provides useful identifications to mention nature of event along with type of characteristics related toward particular event. Research community focus on this to trim down computational complexity of SVM and build event identification algorithms.

2) Among all SVM techniques QS-SVM and CE-SVM techniques require linear optimization problem which will reduce computational complexity than other SVMs. But for online type detection techniques this linear optimization problem performed at each time on each new arrived data sample.
3) Moreover in QS-SVM and CE-SVM high communication cost to identify global outliers as well events. The calculated parameters broadcast to each and every node in the network at each instant time particularly in CE-SVM calculated covariance matrix is broadcasted to essential node every time. Here research community made some techniques but more work is needed, research communities can attention to reduce communication and computational complexities of QS-SVM,
CE-SVM based techniques without compromising their classification performances.
4) SVM techniques endow with hard decision for identifications of outliers in the given samples. Other than these research communities focus on outlier rank or score card which can provide about gradual changes in the state.
5) Important thing in one-class SVM is choice of regularization specification that allows some portion record to lies outside of the calculated regions those data vectors are called as outliers. Enlarged rate of miss-classification outcome into small recognition rate as well as small false positive rate, diminished rate of miss-classification outcome into great false positive rate and great detection rate. Research communities spotlight on schemes to establish optimal significance of regularization constraint.

VIII. CONCLUSION

Outlier revealing over WSNs records is demanding as well as serious issue, due to its enlarge appliances like noise data, fault discovery, intrusion and incident finding. Machine learning based Support vector machine classification techniques are best suited for outlier anomaly detection in WSNs in provisions of communication along with computational complexity than other existed techniques. This article provides an analysis of various SVM Formulations hyper plane, hyper sphere, quarter sphere, hyper ellipsoid, and centered ellipsoid are briefly described in terms of various characteristics. A comprehensive comparison and characteristics SVM formulations discussed, advised as well as research problems are noted.

REFERENCES

1. Xie M, Han S, "Anomaly detection in WSN: A survey," Journal of computer Applications, 34(4),1302-1325, 2011.
2. O'Reilly C, Gulhak, "Anomaly detection in WSN in a non stationary environment," IEEE Communication survey tutor, 16(3), 1413-1432, 2014.
3. Bhanu chander, "A Analysis of machine learning in WSN," International journal of engineering and technology, 7(4.6), pp 185-192, 2018.
4. P.N.Tan, "knowledge discovery from sensor data, 2006.
5. Banarjee A, Burlina P, " A Support vector machine for anomaly detection in hyper spectral imagenay" IEEE Remote, 44(8),2282-2291, 2006.
6. Suthuran rajasegarar, Christopher Leckie, "Quarter sphere based distributed anomaly detection in wireless sensor networks" IEEE Trans. ICC 2007.
7. C.D.Brown, and H.T.Davis, "Receiver operating characteristics curves and related decision measures: A tutorial," chemetrics and intelligent laboratory systems, Vol 1, No 1, pp 24-38, 2006.
8. Hao P, "A new maximal-marginal special structured mutli class classification," intelligent application,30(2) pp-98-111, 2009
9. Abe S, "Support vector machine for pattern classification," Springer, 2010.
10. Sutharan rajasegarar, "An iterative ellipsoid based anomaly detection in WSN," IEEE proceedings, 2012.
11. Hugo martins, Luis Palma, "A support vector machine based technique for detection of outlier in transient times," IEEE trans. 2015
12. Yang Zhang, Nirvana Meratnia, "Distributed online outlier detection in wireless sensor networks using ellipsoidal SVM," Ad-hoc networks, Elsevier B.V all rights reserved, 2012.
13. Yaswanth singh, Suma Saha, "Distributed event detection in wireless sensor networks," IEEE 15th international conference on computer modeling and simulation, 2013
14. Nauman shahid, Ijaz Haider Naqvi, "One class support vector machine: analysis of outlier detection for WSN in harsh environments," Springer, Artificial intelligence reviews, 2013.
15. Zhen Feng, Jingqi Fu, "A new approach of anomaly detection in wireless sensor networks using SVDD," International journal of distributed sensor networks, Vol.3(1), 2017
16. Raihan, Hossain, "A Novel anomaly detection algorithm for sensor data under uncertainty," Springer, Soft computing, 2016
17. Aditi chatterjee, Das, " State estimation and anomaly detection in wireless sensor networs," Springer, Emerging WSN technolohies, 2018
18. Rajasegarar S, Leckie, "Hypersphereical cluster based distributed anomaly detection in wireless sensor networks" journal of parll distrb communications,74(1),1833-1847,2014
19. Shaid, Naqvi, "Quarter sphere SVM: Spatio temporal outlier detection in WSN," IEEE Communications, 2012.
20. Zhu F and Wei JF, " A New SVM reduction strategy of large scale training sample sets, ICMT-2013, pp 815-819, 2013
21. Zhang, Mariena " Adaptive online one class SVM based outlier detection for WSN," WAINA-2009, PP 990-995, 2009.
22. Van vuongntrinh. Anomaly detection in WSN via SVM description, 2017 4th NAFOSTED Conference on Information and Computer Science. 2017.
23. Yang et al, "An online outlier detection in wsn Using unsupervised quarter sphere Support vector machine," ISSNIP 2008, PP151-156,2008
24. D.Wang, D.S Yung, "Structured one class classification," IEEE Trans system man, cyber, 2006.