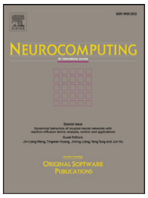
Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom



Ramp loss one-class support vector machine; A robust and effective approach to anomaly detection problems

Yingjie Tian[a,b], Mahboubeh Mirzabagheri[a,b,c], Seyed Mojtaba Hosseini Bamakan[b,c,d,*], Huadong Wang[a,b], Qiang Qu[d]

[a] Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing 100190, China
[b] Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences, Beijing 100190, China
[c] School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China
[d] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

ARTICLE INFO

Article history:
Received 28 October 2017
Revised 12 February 2018
Accepted 7 May 2018
Available online xxx

Communicated by Dr Zhao Zhang

Keywords:
Anomaly detection
One-class SVM
Ramp loss function
Non-convex problem

ABSTRACT

Anomaly detection defines as a problem of finding those data samples, which do not follow the patterns of the majority of data points. Among the variety of methods and algorithms proposed to deal with this problem, boundary based methods include One-class support vector machine (OC-SVM) is considered as an effective and outstanding one. Nevertheless, extremely sensitivity to the presence of outliers and noises in the training set is considered as an important drawback of this group of classifiers. In this paper, we address this problem by developing a robust and sparse methodology for anomaly detection by introducing Ramp loss function to the original One-class SVM, called "Ramp-OCSVM". The main objective of this research is to taking the advantages of non-convexity properties of the Ramp loss function to make robust and sparse semi-supervised algorithm. Furthermore, the Concave–Convex Procedure (CCCP) is utilized to solve the obtained model that is a non-differentiable non-convex optimization problem. We do comprehensive experiments and parameters sensitivity analysis on two artificial data sets and some chosen data sets from UCI repository, to show the superiority of our model in terms of detection power and sparsity. Moreover, some evaluations are done with NSL-KDD and UNSW-NB15 data sets as well-known and recently published intrusion detection data sets, respectively. The obtained results reveal the outperforming of our model in terms of robustness to outliers and superiority in the detection of anomalies.

1. Introduction

The history of anomaly detection or outlier detection can be traced back to the studies have been done by the statistic community at the beginning of nineteen century [13]. Because of the importance of anomaly detection during the time, many researchers from various domains have noted this problem and a broad range of techniques from generic to specific methods have been proposed [6].

In [6], the authors defined the anomaly detection as a problem of finding those data samples which do not follow the patterns of the majority of data points. In fact, anomaly detection is the problem of distinguishing between normal data points with the well-defined patterns or signatures and those that do not conform to the expected profiles. Anomaly detection has a wide range of applications such as, fraud detection [36,48], healthcare monitoring [3], fault detection [11,55], event detection [36] and intrusion detection [25,27,32,35,39,46]. Although in some cases, the anomalies and outliers are considered as the same concept and sometimes interchangeably. In our research, we distinguish them in such a way that we want to reduce the impact of outliers and noises in the training set on the proposed method to have a better detection of abnormal classes.

The focus of our research is on those anomalies that occur in computer networks. The main challenges in this field include the massive volumes of network traffic, the streaming nature of traffic data, a high number of false alarm rate and lack of labeled data for the attacks. Among the aforementioned challenges, availability of labeled data is considered as a significant factor that affects the chosen technique to be a supervised, semi-supervised or unsupervised. Since preparing labels for attack classes are costly, semi-supervised and unsupervised techniques become more favor-

[*] Corresponding author at: Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

E-mail addresses: tyj@ucas.ac.cn (Y. Tian), mirzabagheri888@mails.ucas.ac.cn (M. Mirzabagheri), smhosseini@outlook.com, hosseini@siat.ac.cn (S.M.H. Bamakan), huadong.wang@samsung.com (H. Wang), qiang@siat.ac.cn (Q. Qu).

https://doi.org/10.1016/j.neucom.2018.05.027
0925-2312/© 2018 Elsevier B.V. All rights reserved.

able in anomaly detection problem [6,53]. In fact, the normal class is widely available and this problem can be formulated as a one-class classification problem.

As defined by Giacinto et al. [17], One-class classification problem refers to those two-class problems which the main class is the well-sampled, or in other words, it has a well-defined signature, whereas the other one is severely undersampled because of its extremely diverse nature and also the difficulty in obtaining a significant number of clear and well-defined training patterns. The main objective of one-class classification technique is to distinguish between a set of target objects and the rest of existing objects, which are defined as anomalies or outliers [17]. It means in this case the machine learning techniques are trained just based on one class and the task is to determine whether a new test data is the member of this specific class or not. This situation can be found in many cases such as fault detection in an industrial process, or anomaly detection in network traffic analysis.

Many researchers address this problem and vast ranges of techniques have been proposed as they can be categorized into three groups; density methods such as Parzen window estimator [12], reconstruction methods such as k-means [24] and boundary methods such as One-class SVM [41] and Support Vector Data Description (SVDD) [44].

By considering the boundary based classification methods, in a two-class classification problem, the basic idea of SVM is to establish a hyperplane to separate these two classes of objects with a maximum margin [19,51,52,54]. However, in a case of a one-class problem, a separating hyperplane is constructed in such a way that has the maximum margin between the normal data points and the origin [41]. A new data sample will be classified, as a normal one if it is located within the boundary and conversely, it would detect as an abnormality when it lies outside of the boundary [54]. The focus of our research is on one-class SVM technique as a brief review is presented in Section 2.

Although One-class SVM (OC-SVM) is considered as an effective and outstanding classification technique for anomaly detection problems, this classifier suffers from the problem of sensitivity to the presence of outliers and noises in the training sets. Here, outliers refer to those data points that deviate from the majority of the others. In the real world problem, data sets mostly contain outliers because of some reasons such as instrument failure, formatting errors and non-representative sampling. The sensitivity of OC-SVM to the outliers comes from the convex property of the Hinge loss function. It causes as far as outliers are from the decision boundary, they get larger losses. Therefore, the outliers will shift the decision boundary toward themselves and as a result, it decreases the generalization power of OC-SVM.

The main contributions of this paper are proposed as follows:

(a) Ramp loss One-class SVM is developed as a robust and sparse anomaly detection methodology.
(b) Since Ram-OCSVM is a non-differentiable non-convex optimization problem, the Concave–Convex Procedure (CCCP) is introduced to solve it.
(c) The efficiency and robustness of the proposed method have been examined by different data sets including artificial data, some UCI benchmark data sets, and two network anomaly detection data sets.

2. Related works

Since the time of introducing the first model of One-class SVM by Scholkopf et al. [41] in 2001, many researchers have attempted to introduce some improvements to the basic model [14,26,31]. Guillermo et al. [18] proposed a modified version of OC-SVM to deal with the abrupt change detection problem. The proposed method is based on finding the area of the input space somewhere most of the data points are located. Since this area is changing with the time windows, the model is named “One-class Time-Adaptive SVM”. A weighted version of OC-SVM is proposed by Bicego and Figueiredo [3] named “WOC-SVM”. In their research, a weight factor which represents the importance of corresponding data point has been used to train the model. The authors utilized the WOC-SVM to introduce a soft clustering algorithm. In [57], the authors tried to further improve the WOC-SVM by introducing a novel instance-weighted strategy. In their method k-nearest neighbor is used to denote a weight to those data samples which are near the boundary of the training set. Higher weights are assigned to the data samples which are close to the boundary of the data distribution and conversely lower weights are assigned to the data samples which located in the interior of the training set.

In specific, some researchers forced on the sensitivity of OC-SVM to the outliers and the modifications are proposed to make this classifier more robust. Amer et al. [1] in 2013 proposed two enhanced versions of OC-SVM called them “robust one-class SVM” and “eta one-class SVM” to make it more effective in unsupervised anomaly detection problem. In the former model, the authors did some modification respect to the slack variables, in such a way that they are proportional to the distance to the centroid of the kernel space. On the other hand, in eta OC-SVM the number of non-zero slack variables which contribute to minimize the objective function, is controlled by introducing 0–1 variable η_i. Here, if the η is equal to 1, it refers to normal data samples, and for the outliers η will be equal to 0 [1]. Since the objective function of this problem consists of a convex quadratic problem and a linear problem, so the objective is not jointly convex. Thus, it needs to be relaxed by a semi-definite programming problem. According to the Amer et al. [1], eta OC-SVM shows promising results compared to the other methods in terms of sparsity and area under the curve (AUC).

In 2014, Yin et al. [54] also addressed the sensitivity of the OC-SVM to the noises and outliers in the training sets in the context of the fault detection problem. Therefore, to depress the pressure of these points on the decision function of OC-SVM, an adaptive penalty factor is introduced to develop a robust one class SVM. In the OC-SVM, a slack variable ξ_i is introduced to allow some data points to locate outside of the boundary. And the number of these points are controlled by the penalty factor $1/\nu l$. It means that the possibility of locating data points outside of the decision boundary will be increased if the value of the penalty factor is small. However, in the proposed robust OC-SVM, the penalty factor is adjusted by taking into account the distances between the data samples and the center of the data set [54].

Motivated by our previous research that we develop a precise, sparse and robust methodology for multi-class intrusion detection problem based on the Ramp Loss K-Support Vector Classification-Regression, named “Ramp-KSVCR” [20] and the aforementioned works, we addressed the sensitivity of One-class SVM to the outliers by introducing a non-convex loss in order to reduce the impact of unexpected points in the data sets. However, during the reviewing process of this paper, we found that the similar idea is used by Xiao et al. [50]. Although, both of ideas are developing the Ramp loss based OC-SVM, the approaches to formulating and considering this problem is different. In addition to develop the proposed model based on sequential minimal optimization (SMO) [7], in order to make our model more applicable in the large-scale setting, Alternating Direction Method of Multipliers (ADMM) is used to solve sub-quadratic programming problems in each iteration of CCCP [4,20,47]. Moreover, the main focus of this paper is to address the problem of computer network anomaly detection. Hence, besides to do the comprehensive experiments and perform parameters sensitivity analysis on two artificial data sets and some general anomaly detection data sets, to show the superiority of our

Please cite this article as: Y. Tian et al., Ramp loss one-class support vector machine; A robust and effective approach to anomaly detection problems, Neurocomputing (2018), https://doi.org/10.1016/j.neucom.2018.05.027