Knowledge-Based Systems 000 (2016) 1–14

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

An one-class classification support vector machine model by interval-valued training data

Lev V. Utkin[a,*], Yulia A. Zhuk[b]

[a] Peter the Great St.Petersburg Polytechnic University, Russia
[b] ITMO University, Russia

ARTICLE INFO

Article history:
Received 24 April 2016
Revised 19 December 2016
Accepted 21 December 2016
Available online xxx

Keywords:
One-class classification
SVM
Uncertainty trick
Interval-valued data
Expected risk

ABSTRACT

A modification of the well-known one-class classification support vector machine (OCC SVM) dealing with interval-valued or set-valued training data is proposed. Its main idea is to represent every interval of training data by a finite set of precise data with imprecise weights. This representation is based on replacement of the interval-valued expected risk produced by interval-valued data with the interval-valued expected risk produced by imprecise weights or sets of weights. In other words, the interval uncertainty is replaced with the imprecise weight or probabilistic uncertainty. It is shown how constraints for the imprecise weights are incorporated into dual quadratic programming problems which can be viewed as extensions of the well-known OCC SVM models. Numerical examples with synthetic and real interval-valued training data illustrate the proposed approach and investigate its properties.

1. Introduction

Most classification problems deal with data which can be divided into two or many classes. However, in many applications, for example, in systems with sensors to sense the surrounding environment, we have unlabeled observations which generally belong to one class, but some observations show large deviations from other observations. They may indicate about abnormal system behavior. Therefore, one of the important classification tasks is to detect these abnormal observations or to solve the novelty detection problem. In order to separate this kind of classification problems from clustering, the novelty detection is viewed as one-class classification (OCC). The OCC aims to detect anomalous or abnormal observations and separate them from the so-called normal examples [9,10,42]. However, as indicated by Khan and Madden [26], OCC problems are different from the conventional binary or multi-class classification problems in the sense that in OCC, the negative class is either not present or not properly sampled. The problem of classifying positive example or normal observation in the absence of appropriately-characterized negative cases (or outliers) is a very interesting and important task which can be applied to various applications. One of the most common ways to define anomalies is by saying that anomalies are not concentrated [39]. The importance of the abnormal behavior detection stimulated to develop a series of interesting OCC models [5,6,38,40,42–44]. Their detailed review and comparison can be found in papers [2,10,24,26,27,34].

A large part of the OCC models are based on using kernel-based methods in the framework of the support vector machine (SVM). The models are called OCC SVMs. It is interesting that Khan and Madden [27] in their review of the OCC models divide them into only two groups: OCC SVM and non-OSVM. The second group, non-OSVM, consists of models which are not based on the SVM framework.

First of all, we have to point out an approach proposed by Tax and Duin [43,44]. According to the approach, the OCC problem is solved by distinguishing the normal observations from all other possible data points by means of finding a hyper-sphere around the normal observations, which contains almost all points in the data set with the minimum radius. This approach is called the Support Vector Data Description (SVDD). In a nutshell, the approach considers the trade-off between the number of errors made on the training set (number of target objects rejected) and the size of the sphere (its radius). By adapting the kernel function, this approach becomes more flexible than just a sphere in the input space. The SVDD model will be studied in detail in Section 5.

Another interesting way to geometrically enclose a fraction of the training data is via a hyperplane and its relationship to the origin. The corresponding approach was proposed by Schölkopf at el. [38,40]. It will be studied in detail in Section 2. Under this approach, a hyperplane is used to separate the training data from the origin with the maximal margin, i.e., the objective is to separate

* Corresponding author.
E-mail addresses: lev.utkin@gmail.com (L.V. Utkin), zhuk_yua@mail.ru (Y.A. Zhuk).

http://dx.doi.org/10.1016/j.knosys.2016.12.022
0950-7051/

off the region containing the data points from the surface region containing no data. It should be noted that both the approaches provide the same results when a symmetric kernel is used.

The third approach is the linear programming approach to the OCC proposed by Campbell and Bennett [5,6]. This model uses linear programming techniques which are much simpler than the standard quadratic programming in the OCC SVM.

Steinwart et al. [42] use the assumption that anomalies are not concentrated in order to consider the problem of finding level sets for the data generating density. The authors of [42] interpret this learning problem as a binary classification problem and compare the corresponding classification risk with the standard performance measure for the density level problem.

Next approaches of the OCC we would like to point out are the so-called Single-Class Mapping Convergence classification algorithm and its fast modification Support Vector Mapping Convergence which were proposed by Yu [52]. The main algorithm computes an accurate boundary of the target class from positive and unlabeled data (without labeled negative data). Its basic idea is to exploit the natural "gap" between positive and negative data by incrementally labeling negative data from the unlabeled data using the margin maximization property of SVM.

We have to point out some modifications of the OCC SVM. In particular, Kwok et al. [32] propose the so-called single-class minimax probability machines which offer robust novelty detection with distribution-free worst case bounds on the probability that a pattern will fall inside the normal region. Bicego and Figueiredo [4] study a weighted OCC model. According to their approach, every data point has a weight indicating the importance assigned to each point of the training set. A question of getting the corresponding weights for every point is a main drawback of the approach. Utkin [45], Utkin and Zhuk [48] propose robust modifications of the OCC SVM by using imprecise statistical models [49] in order to assign the weights and to robustify the OCC SVM.

Zhu and Zhong [54] propose an OCC method which uses an additional information hidden in the training data. This is a group information related to training errors or slack variables in the standard OCC SVM. In accordance with the group information, additional constraints on the slack variables are introduced. In fact, the authors use one of the features, which divides the training data into a finite number of subsets. Chen et al. [14] present an OCC algorithm which uses tensor as input data and aims to separate almost all samples of target class from the origin with maximal margin. The first advantage of the algorithm is that the use of direct tensor representation is helpful to retain the data topology more efficiently. The second advantage is that the tensor representation can greatly reduce the number of parameters and overcome the overfitting problem.

In order to improve the OCC performance and to increase the OCC accuracy, Cyganek [15] presents an interesting method which allows extensions of a single OCC SVM into an ensemble of OCC SVMs. The main idea underlying the method consists in prior clustering of the input data into smaller and more compact partitions which are then used to train the OCC SVM classifiers. Following this ensemble-based approach, Krawczyk et al. [31] propose to use a clustering algorithm to partition the feature space in order to train one-class classification classifiers for each cluster. The main advantage of the proposed method is that individual classifier strengths can be taken into account in the combined classifiers trained on the basis of clusters. A very interesting modified weighted OCC SVM algorithm for constructing efficient one-class ensembles augmented with the incremental training of the weights additionally supported by forgetting mechanism is introduced by Krawczyk and Wozniak [29]. The authors propose to assign weights of a special form to training examples such that, in case of data streams, samples from the incoming chunk are used together with the samples from previous chunks. The ensemble-based approach to OCC problems has shown outperforming results [15,29,31]. Therefore, Krawczyk [28] presents a novel technique for forming efficient OCC ensembles. It is important to point out that the proposed technique covers a series of interesting ideas and the corresponding algorithms, including the swarm intelligence approach implemented as a firefly algorithm, a weighting scheme for modifying the influence of the selected classifiers. These original ideas lead to a powerful ensemble-based OCC algorithm. Another approach for constructing ensembles for OCC problems based on dynamic classifier selection is presented by Krawczyk and Wozniak [30]. The authors propose the following three efficient measures for calculating the competence of given one-class classifiers for each object of a training set: the minimal difference measure computed according to correctness and degree of uncertainty of the analyzed classifier at a given validation object; the full competence measure reflecting how a given classifier opts strongly for one of the considered classes; the entropy measure based on the popular entropy criterion. Liu et al. [33] also present an ensemble-based OCC algorithm which aims to deal with multi-modality, multi-density, arbitrarily shaped distributions and the noise of target samples. It performs density analysis on the target class to build a tree-shaped structure to represent the density distribution of the target samples. By using the tree-shaped structure, several local dense subsets of the target class are constructed to describe groups of target samples which are close to each other and have similar density. The algorithm aggregates several simple OCC models trained on every local dense subset by modular ensemble.

It should be noted that the above OCC SVM models assume that training sets consist of precise or point-valued data. However, training examples in many real applications can be obtained only in the interval form. Interval-valued data stem from imperfection of measurement tools or imprecision of expert information, from missing data. Interval training sets may arise in situations of specific processing and representation of point-valued data such as recording monthly interval temperatures in meteorological stations, daily interval stock prices, etc. Another source of interval data is the aggregation of huge data-bases into a reduced number of groups [35]. For example, patient records with unusual symptoms, containing daily interval temperatures, the interval-valued blood pressure, may indicate potential health problems for a particular patient. The identification of such unusual records by interval-valued data can be regarded as an OCC problem. Another example is the identification of a certain atypical behavior in the environmental monitoring when many factors (temperature, rainfall) are recorded for some time intervals, for example, for a day or a week. We also have to mention swarm robotic systems where a huge number of robots equipped with several sensors which may include GPS units, temperature sensors, altimeters, imaging systems, etc. Similar sensors of different robots provide with the information about the same object approximately at the same time. Therefore, the set of the corresponding measurements in this case can be regarded as a single set-valued or interval-valued training example. The efficient joint functioning of all robot subsystems significantly depends on the quick detection of an anomalous behavior of the total swarm robotic system. Interesting real-life applications of the anomaly detection where interval-valued data may be available are also provided by Zhang et al. [53].

There are many approaches to handle interval-valued data in classification problems. The largest part of the approaches is based on replacement of intervals with their precise or point-valued representation, for example, by replacing intervals with their middle points [35]. These approaches can be successfully used when intervals are not large and the area produced by the interval intersections is rather small. At the same time, they may lead to a low

Please cite this article as: L.V. Utkin, Y.A. Zhuk, An one-class classification support vector machine model by interval-valued training data, Knowledge-Based Systems (2016), http://dx.doi.org/10.1016/j.knosys.2016.12.022