Information Sciences 264 (2014) 182–195

Contents lists available at ScienceDirect

Information Sciences

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

Clustering-based ensembles for one-class classification

Bartosz Krawczyk [a,*], Michał Woźniak [a], Bogusław Cyganek [b]

[a] Department of Systems and Computer Networks, Wroclaw University of Technology, Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland
[b] AGH University of Science and Technology, Al. Mickiewicza 30, 30-059 Kraków, Poland

ARTICLE INFO

Article history:
Received 8 October 2012
Received in revised form 5 November 2013
Accepted 14 December 2013
Available online 31 December 2013

Keywords:
One-class classification
Multiple classifier system
Classifier ensemble
Clustering
Soft computing

ABSTRACT

This paper presents a novel multi-class classifier based on weighted one-class support vector machines (OCSVM) operating in the clustered feature space. We show that splitting the target class into atomic subsets and using these as input for one-class classifiers leads to an efficient and stable recognition algorithm. The proposed system extends our previous works on combining OCSVM classifiers to solve both one-class and multi-class classification tasks. The main contribution of this work is the novel architecture for class decomposition and combination of classifier outputs. Based on the results of a large number of computational experiments we show that the proposed method outperforms both the OCSVM for a single class, as well as the multi-class SVM for multi-class classification problems. Other advantages are the highly parallel structure of the proposed solution, which facilitates parallel training and execution stages, and the relatively small number of control parameters.

1. Introduction

Well-known and reliable classifiers tend to fail when faced with new problems such as an atypical class distribution, non-stationary environments, or massive data. Therefore, new methods must be developed to deal with the challenges arising and improve the quality of real-life decision support systems.

One of these newly introduced methodologies is known as one-class classification (OCC) [31], which assumes that during the training stage only objects originating from a single class are available. These are called the target concept and are denoted by ω_T. The purpose of OCC is to calculate a decision boundary that encloses all available data samples, thereby describing the concept [53]. During the execution phase, new objects, unseen during training, may appear. These may originate from one or more distributions and represent data outside the target concept. Such objects, denoted by ω_O, are referred to as outliers.

For a single OCC classifier it may be difficult or even impossible to find a good model owing to limited training data, high feature space dimensionality, and/or the properties of the particular classifier. To avoid a too complex model and overfitting of the training target data, a simpler model with a lower number of features or one that has been trained with smaller chunks of data, can be created. Although the complexity of such a model is reduced, the quality thereof also declines significantly. However, it has been shown that a group of individual OCC models can help alleviate the aforementioned problems.

Here one may use an approach known as multiple classifier systems (MCSs), which is considered to be one of the fastest growing fields in machine learning [26]. MCSs are based on the idea of combining several classifiers into a compound recognition system that can exploit the strengths of individual predictors [60]. Each classifier may output a different decision boundary, and so have different competence areas over the analyzed dataset [7]. When combined, the collective decision

* Corresponding author. Tel.: +48 71 320 35 39; fax: +48 71 320 29 02.
E-mail addresses: bartosz.krawczyk@pwr.wroc.pl (B. Krawczyk), michal.wozniak@pwr.wroc.pl (M. Woźniak), cyganek@agh.edu.pl (B. Cyganek).

0020-0255/$ - see front matter
http://dx.doi.org/10.1016/j.ins.2013.12.019

accuracy can outperform any of the individual predictors. However, several important issues, such as selecting the individual classifiers, as well as choosing a fusion method to establish a group decision, must be considered when designing an MCS. Classifiers used to create the ensemble in an ideal situation should be highly accurate and complement each other (i.e., the ensemble should display high diversity). Adding classifiers that are not diverse with respect to those already in the pool will not improve the accuracy of the compound classifier, but will only increase the overall computational cost [5]. It is worth noting that combination rules, for example, majority voting, could even lead to a deterioration in performance of the ensemble of classifiers [36]. On the other hand, building an MCS with highly diverse but poor quality classifiers will result in a weak committee. Therefore, classifier selection is a critical step in the ensemble design process [15].

MCSs are an attractive yet still largely unexplored, alternative for OCC problems. Most of the works concentrate on practical applications of OCC ensembles. Much still needs to be done to gain insight into the theoretical background to this problem, as well as to draw conclusions on how to build efficient OCC ensembles regardless of the intended application [35].

We propose an approach based on the idea of data clustering in the feature space. OCC models are built based on each of the clusters. In this way we ensure that the pool of predictors is highly diverse and mutually complementary (owing to training on different inputs, i.e., clusters of training objects). This can be seen as an extension of the popular family of ensembles derived from the idea of *clustering and selection* proposed by Kuncheva [37]. So far, two other research teams have worked on this topic, proposing very simple hybrid methods for combining clustering and OCC [38,45].

The contributions of this work are as follows:

- We propose building an ensemble of one-class classifiers based on clustering of the target class. This ensures initial diversity among the classifiers in the pool (as they are based on different inputs) and the correct handling of possible issues embedded in the nature of data, such as a rare distribution or chunks of objects.
- We propose an elastic and efficient framework for this task, which requires only the selection of several components, namely, the clustering algorithm, individual classifier model, and fusion method. These can easily be chosen by the user, as there are practically no limitations on their nature. All other parameters for the method are selected automatically.
- We discuss the possibility of extending our one-class ensemble to an efficient tool for multi-class problems.
- We carry out extensive computational tests on a diverse set of benchmarks that highlight the influence of component selection on the overall method quality and show that the proposed approach outperforms the standard OCC methods as well as a single multi-class support vector machine (SVM) in multi-class classification problems.

Our ensemble is easy to use in many practical applications where it is difficult or even impossible to obtain counter-examples (e.g., machine fault diagnosis), or where, owing to a complex data distribution, the class decomposition approach can lead to a significant improvement in recognition quality over the well known multi-class approaches (e.g., imbalanced classification).

This paper is organized as follows. In the next section the idea of OCC is presented. In Section 3 the architecture of the proposed compound recognition system is explained. The components that must be selected as input for the system are also presented. In Section 4 the experimental results are presented and discussed. The paper ends with the presentation of our conclusions in Section 5.

2. One-class classification

OCC aims to distinguish the target concept objects from possible outliers, and hence it is often referred to as learning in the absence of counter-examples. Although OCC is quite similar to binary classification, the primary difference lies in how the one-class classifier is trained. In standard dichotomy problems it is expected that objects from the other classes tend to come from one direction. Here the available class must be separated from all the possible outliers, which leads to a situation in which a decision boundary must be estimated in all directions in the feature space around the target class. An example OCC problem is depicted in Fig. 1.

OCC is a solution to many real-life problems where there are abundant data for a single class, but it is difficult or even impossible to obtain data for other objects. This is often the case in problems such as image analysis [49], intrusion detection [23], machine fault diagnosis [2], and solid-state fermentation [27].

It is worth noting that there are two different views of OCC:

- As a tool for solving single class problems, where during training only data drawn from the target concept are available e.g., for web page classification [63].
- As a method for the decomposition of multi-class decision task into simpler ones. A canonical decomposition is conducted with the use of binary classifiers [20]. In this approach an M class problem is solved by combining M one-class classifiers, each of which is responsible for the recognition of a different class.

Several methods dedicated to solving OCC problems have recently been introduced. Based on the literature, three main approaches can be distinguished: