Neurocomputing 126 (2014) 36–44

Contents lists available at ScienceDirect

Neurocomputing

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

Diversity measures for one-class classifier ensembles

Bartosz Krawczyk *, Michał Woźniak

Department of Systems and Computer Networks, Wroclaw University of Technology, Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland

ARTICLE INFO

Article history:
Received 1 June 2012
Received in revised form
29 November 2012
Accepted 10 January 2013
Available online 30 July 2013

Keywords:
One-class classification
Multiple classifier system
Machine learning
Diversity measure
Combined classifier

ABSTRACT

One-class classification is one of the most challenging topics in contemporary machine learning and not much attention had been paid to the task of creating efficient one-class ensembles. The paper deals with the problem of designing combined recognition system based on the pools of individual one-class classifiers. We propose the new model dedicated to the one-class classification and introduce novel diversity measures dedicated to it. The proposed model of an one class classifier committee may be used for single-class and multi-class classification tasks. The proposed measures and classification models were evaluated on the basis of computer experiments which were carried out on diverse set of benchmark datasets. Their results confirm that introducing diversity measures dedicated to one-class ensembles is a worthwhile research direction and prove that the proposed models are valuable propositions which can outperform the traditional methods for one-class classification.

1. Introduction

Nowadays methods of automatic classification are the focus of intense research and number of approaches are being used for the practical implementation in the computer systems which support human decision as medical decision support systems [16] or make autonomic decision as IDS/IPC [35], spam filters [10], or fraud detection [3]. There are plethora of propositions [13,9], nevertheless according to "no free lunch theorem" proposed by Wolpert [51] there is not a single classifier that is suitable for all the tasks, since each of them has its own domain of competence. Additionally, for many practical cases the traditional methods cannot achieve the satisfying quality what is usually caused by the insufficient learning material. For such a case usually combined methods of collective decision are used [29]. The key factor for such a system is to assure a satisfactory pool of diverse classifiers, because using similar ones cannot improve classification accuracy of a combined classifier. A strategy for generating the classifier ensemble should guarantee its diversity improvement therefore let us enumerate the main propositions how to get a desirable committee:

- *Dataset splitting*: The individual classifiers could be train on different datasets i.e., we can divide the original dataset into partitions or generate a number of subsets through data splitting, a cross-validated committee, bagging, or boosting [40], in the hope that classifiers trained on different inputs would be complementary.
- *Feature set splitting*: The individual classifiers can use the selected features only. This split can be done manually, driven by some objective function [25] or using a random method as the Random Subspace [21]. One should note that the last mentioned method does not always improve the ensemble quality due to its full randomness—therefore several modifications were proposed in recent years, such as using some supervision over the subspace creation process or choosing only the relevant subspaces [33].
- *Class set splitting*: Usually it could be easy to decompose the problem into simpler ones. Let us note that human being usually learns the complex concept (e.g., character recognition) by split this process into several binary problems, where there are two strategies used—one-versus-all (OVA) and one-versus-one (OVO). Of course the key problem is how to choose a fusion method that can recover the whole set of possible classes. One of the best known approach is Error-Correcting Output Codes proposed by Ditterich et al. [12] which will be described later in this work.
- *Using different classifier models*: The last and intuitive methods are to use individual classifiers trained on different models or different versions of models. Such an approach can exploit the strengths of the classifiers which stay behind so-called individual bias, because according to Watanabe's Ugly Ducking Theorem [48] the high quality classification is impossible without a bias. On the other hand the most of the machine learning methods used for classifier training use an heuristic optimization procedures. E.g., C4.5 decision tree induction uses greedy search [39], which does not guarantee that an optimal

* Corresponding author.
E-mail addresses: bartosz.krawczyk@pwr.wroc.pl (B. Krawczyk), michal.wozniak@pwr.wroc.pl (M. Woźniak).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.01.053

tree structure is found. Therefore finding a pool of would-be classifiers using different starting points and merging them could be an interesting proposition.

The separate problem is how to measure the diversity [37,4]. The measure should evaluate the differences between the individual models in the pool in hope that combining different classifiers will lead to an improvement of the committee quality. We can achieve it if a pool of individual classifiers is mutually complimentary [2] i.e., an incompetence area of the pool (the subset of a feature space where all individual classifiers make the wrong decision) is small [38]. Well known diversity measures focus e.g., on minimizing the possibility of a coincidental failure [31].

In this work we will concentrate on the problem of one-class classification (OCC). OCC seeks to distinguish one specific class from the more broad set of classes (e.g., selecting carrot from other vegetables). The target class is considered as a positive one, while all other are considered as negative ones. As we see the OCC is a quite similar to binary classification but the difference is how the one-class classifier is trained. During the learning only examples target class (known also as positive examples) are being presented to learner. An example of one-class classification with a target concept and existing outliers not included in the training process is presented in Fig. 1.

OCC problems are common in the real world where positive examples are widely available but negative ones are hard, expensive or even impossible to gather. Let us consider an engine. It is a quite easy and cheap to collect data about its normal work. Yet gathering observations about its failures is expensive and sometimes even impossible, because one would have to spoil the engine.

Such approach is very useful as well for many practical cases especially when the target class is "stable" and outlier one is "unstable". To explain this motivation let us consider a computer security problem as spam filtering or intrusion detection (IDS/IPS) [36]. The target class which covers the normal, safe messages is unchanged but malicious messages as spam or intrusion is still changing because malicious users are trying to lead security systems on, therefore they are inventing new type of attacks. If we concentrate on normal messages only that we can believe that we are able to train classifier which can distinguish between normal messages and malicious ones without knowledge about outlier class. It could protect our security system against so-called "zero day attack" as well.

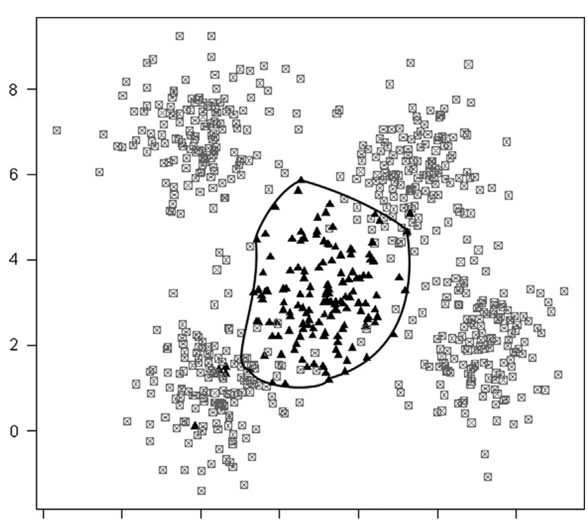

Fig. 1. Example of one-class classification with known target class used for the boundary creation. One-class classification assumes that in the exploratory phase of classification new, unknown objects not belonging to the target class may appear.

Of course designing effective one-class classifier faces the same problems as for the traditional classifiers. The main problem is that for some cases the classifiers cannot achieve a suitable qualities. Therefore this work focuses on the problem how to ensure the appropriate diversity for a pool of one-class classifiers and producing combined one-class classifier and how to design the combined classifiers on the basis of a pool of one-class classifiers for multi-class task. In our previous work we have presented a flexible evolutionary framework for selecting diverse one-class classifiers for the ensemble [27] and discussed the problem of OCC ensemble pruning [26].

In this work we do not deal with the fuser design problem which is of course important and key problem for multiple classifier systems [46], but we will use well-known methods as ECOC [12] or Decision Templates [30] which will be described in the follow-up sections. We realize that the problem of fuzer design for one-class classifier is important and some recommendations and methods could be found in the previous works of authors [49].

The contribution presented in this work can be described by three main ideas:

- Introduction of the concept of diversity for one-class classification and proposal of novel diversity measures for selecting a heterogeneous individual models for one-class multiple classifier system.
- Examining does the increase in the number of one-class classifiers assigned to a single class leads to a performance improvement.
- Exhaustive computational tests that evaluate the quality of proposed methods on the basis of extensive benchmark datasets—for single class tasks and for decomposition of multi-class data with one-class predictors.

This work is organized as follows. The next section gives an overview of one-class classification task and two boundary classifiers used in this paper. Section 3 is devoted to the problem of one-class classifiers fusion. Following section describes our propositions of diversity measures dedicated to one-class classifiers. Section 5 presents extensive experimental investigations, divided into two parts: single-class classification and multi-class decomposition. Final section concludes the paper.

2. One-class classification task

Let us notice the two possible applications of one-class classifiers:

- for single class problems, where during the training only data drawn from target concept is available e.g., for web page classification [52],
- for decomposition of multi-class datasets into simpler decision problems. Canonically decomposition is conducted with the usage of binary classifiers [14]. In this approach a M class problem is solved by the decomposition into M one-class classifiers, each responsible for a different class. Example of such an approach emphasizing the differences between binary and one-class decompositions is presented in Fig. 2.

Several methods have been proposed to solve the OCC problem. In the relevant literature three main approaches could be distinguished: the density estimation, the reconstruction methods and the boundary methods. Let us describe them shortly:

- The estimation of the density of given data by putting a threshold on target concept may be considered as the most