Pattern Recognition 61 (2017) 185–196

Contents lists available at ScienceDirect

Pattern Recognition

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

Selective ensemble of SVDDs with Renyi entropy based diversity measure

Hong-Jie Xing [a,*], Xi-Zhao Wang [b]

[a] Key Laboratory of Machine Learning and Computational Intelligence, College of Mathematics and Information Science, Hebei University, Baoding 071002, China
[b] College of Computer Science and Software, Shenzhen University, Shenzhen 518060, China

ARTICLE INFO

Article history:
Received 22 August 2015
Received in revised form
25 May 2016
Accepted 25 July 2016
Available online 27 July 2016

Keywords:
One-class classification
Selective ensemble
Support vector data description
Renyi entropy

ABSTRACT

In this paper, a novel selective ensemble strategy for support vector data description (SVDD) using the Renyi entropy based diversity measure is proposed to deal with the problem of one-class classification. In order to obtain compact classification boundary, the radius of ensemble is defined as the inner product of the vector of combination weights and the vector of the radii of SVDDs. To make the center of ensemble achieve the optimal position, the Renyi entropy of the kernelized distances between the images of samples and the center of ensemble in the high-dimensional feature space is defined as the diversity measure. Moreover, to fulfill the selective ensemble, an ℓ_1-norm based regularization term is introduced into the objective function of the proposed ensemble. The optimal combination weights can be iteratively obtained by the half-quadratic optimization technique. Experimental results on two synthetic data sets and twenty benchmark data sets demonstrate that the proposed selective ensemble method is superior to the single SVDD and the other four related ensemble approaches.

1. Introduction

One-class classification [1–3] is regarded as a machine learning task between supervised learning and unsupervised learning. It can efficiently deal with the problem of extreme class imbalance. In the training phase, only the samples in one-class can be used to train a classifier. Moreover, the testing samples can be classified as normal or novel by the trained classifier. There are many examples of one-class classification in our real world, such as machine fault detection [4], network intrusion detection [5], medical diagnosis [6], credit scoring [7], among others [8,9].

Support vector data description (SVDD) [10] is a generally used method as a one-class classifier. It establishes a hyper-sphere in the form of kernel expansion to distinguish the normal data from the novel data. The kernel function in the decision function maps the samples from the original space into a high-dimensional feature space while the explicit form of the mapping is not needed according to the 'kernel trick' [11]. When certain conditions are satisfied, SVDD is proved to be equivalent to one-class support vector machine (OCSVM) [12,10].

To make one-class classifier achieve more compact classification boundary, Tax and Duin [13] proposed the ensemble of one-class classifiers. They found that the ensemble can obviously improve the classification performance of one-class classifier. Seguí et al. [14] proposed the weighted bagging based ensemble of one-class classifiers. They utilized minimum spanning tree class descriptor as base classifiers. Zhang et al. [15] used locality preserving projection to reduce the dimensionality of the original data, trained several SVDDs upon the reduced data, and combined the outputs of the trained SVDDs. Hamdi and Bennani [16] proposed an ensemble of one-class classifiers by utilizing the orthogonal projection operator and the bootstrap strategy. Wilk and Woźniak [17] constructed the ensemble of one-class classifiers by fuzzy combiner. They utilized fuzzy rule based classifier as base classifier, while used fuzzy error correcting output codes and fuzzy decision templates as ensemble strategies. For tackling malware detection, Liu et al. [19] constructed random subspace method based ensemble of cost-sensitive twin one-class classifiers. Casale et al. [20] proposed the approximate polytope based ensemble of one-class classifiers. The methodology uses the geometrical concept of convex hull to define the boundary of the normal class, while utilizes random projections and ensemble decision process to judge whether a sample belongs to the convex hull in high-dimensional spaces. Furthermore, a tilling strategy was proposed to model non-convex structures. Krawczyk et al. [18] proposed the clustering-based ensemble of one-class classifiers. The clustering algorithm is utilized to split the whole normal class into the disjointed sub-regions. On each sub-region, a single one-class

* Corresponding author.
E-mail address: hjxing@hbu.edu.cn (H.-J. Xing).

http://dx.doi.org/10.1016/j.patcog.2016.07.038
0031-3203/

classifier is trained. Finally, the outputs of all the one-class classifiers are combined together. Aghdam et al. [33] developed a new one-class classification method that can be trained with or without novel data and it can model the observation domain utilizing any binary classification approach. To mine data steams with concept drift, Czarnowski and Jedrzejowicz [34] proposed an instance selection and chunk updating based ensemble of one-class classifiers. Experimental results demonstrate that their method can outperform the well-known approaches for data streams with concept drift.

For the scenarios of two-class classification and multi-class classification, diversity is regarded as a key issue in classifier ensemble. Dietterich [21] compared the effectiveness of three ensemble methods, i.e., randomization, bagging, and boosting for improving the performance of the single decision tree. Through experiments he declared that randomization is competitive with bagging but not as accurate as boosting in the situation with little or no noise in the given training samples. Moreover, Dietterich also observed that the classifiers in the ensemble become less diverse as they become more accurate. Conversely, the classifiers become less accurate as they become more diverse. Kuncheva and Whitaker [22] studied ten measures of diversity between the base classifiers. They concluded that designing diverse classifiers is correct. However, in real-life pattern recognition problems, measuring diversity and utilizing the diversity to efficiently build better classifier ensemble is still an open problem. For the majority vote combiner, Brown and Kuncheva [23] first decomposed the classification error into three parts, i.e., individual accuracy, 'good' diversity, and 'bad' diversity. Moreover, they also declared that a larger value of the good diversity reduces the majority vote error, while a larger value of bad diversity increases the error. Recently, Sidhu et al. [24,25] studied the diversified ensemble approaches for the online stream data.

Similar to the cases of two-class classification and multi-class classification, diversity measure [26,27] acts an important role for the ensemble of one-class classifiers. Krawczyk and Woźniak [28,29] first investigated the diversity of ensemble for one-class classification and formulated five diversity measures. Moreover, Krawczyk and Woźniak [30] studied the relationship between the accuracy and diversity towards the ensemble of one-class classifiers. They proposed a novel ensemble strategy for one-class classification by assuring both high accuracy of individual one-class classifiers and high diversity among these classifiers. Besides the accuracy of individual one-class classifiers and the diversity of ensemble, the combination strategy also affects the performance of the ensemble of one-class classifiers. Menahem et al. summarized the commonly used combination rules and provided a list in literature [31]. However, these combination rules all rely on the estimated probability of sample given the normal class. In the study, the LSE (lease squares estimation)-based weighting [32] is utilized to directly combine the outputs of the decision functions of individual SVDDs.

As aforementioned, the classification boundary of SVDD in the high-dimensional feature space is hyper-sphere. After combined by the LSE-based weighting rule, the boundary of ensemble of SVDDs in the feature space is also a hyper-sphere. Fig. 1 illustrates an ensemble of SVDDs. It can be deduced from Fig. 1 that the performance of the ensemble of SVDDs is determined by its length of radius and location of center. Therefore, the study focuses on finding the optimal radius and center of ensemble rather than the highest diversity of ensemble.

Moreover, although an ensemble of classifiers often achieves better performance than one single classifier, the computational cost for obtaining the ensemble of these classifiers will become expensive when the number of base classifiers is large. To overcome the aforementioned disadvantage, Zhou et al. proved in [35]

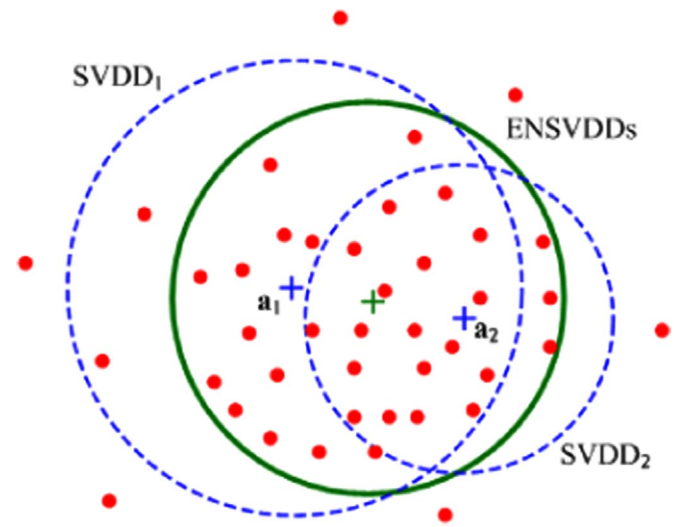

Fig. 1. Schematic diagram of the ensemble of SVDDs. $\mathbf{a}_1$ and $\mathbf{a}_2$ are the centers of the two SVDDs, while ENSVDDs are the ensembles of the two SVDDs.

that it is better to ensemble part of the base classifiers rather than all of them. Li and Zhou [36] proposed a selective ensemble algorithm based on the regularization framework. Through solving a quadratic programming, they get the sparse solution of the vector of combination weights and implement the selective ensemble. Zhang and Zhou [37] proposed a linear programming based sparse ensemble method. Yan et al. [38] proposed a selective neural network ensemble classification algorithm for the incomplete data. It is noted in ensemble learning because the handling of uncertainty plays a key role for classifier performance improvement (e.g. [39,40]) and the selection of base classifier is very sensitive to the overall performance in bio-informatics [41,42]. Nevertheless, the existing selective ensemble approaches mainly concentrate on the supervised learning. Till now, there are too few work upon the selective ensemble of one-class classifiers. Krawczyk and Woźniak [43–46] investigated this issue and proposed four pruning strategies, i.e., multi-objective ensemble pruning, dynamic classifier selection method, firefly algorithm based ensemble pruning, and clustering-based pruning. Experimental results demonstrate that their methods outperform the state-of-the-art algorithms for selecting one-class classifiers from the given classifier committees. Parhizkar and Abadi [47] utilized a modified binary artificial bee colony algorithm to prune the ensemble of one-class classifiers and used the ordered weighted averaging operator to combine the outputs of base classifiers in the pruned ensemble.

In this study, we propose a selective ensemble strategy for SVDD to get the optimal combination weights of base classifiers. The proposed ensemble is mainly based on the Renyi entropy based diversity measure. The main contributions of the present study are as follows:

- The radius of ensemble is defined to be the inner product between the vector of combination weights and the vector of the radii of SVDDs. Therefore, minimizing the radius of ensemble can make the classification boundary of the ensemble of SVDDs as compact as possible.
- The Renyi entropy of the distance variable obtained by the kernelized distances between the images of samples and the center of ensemble in the feature space is defined as the diversity measure. Maximizing the Renyi entropy based diversity can make the center of ensemble attain the optimal position in the feature space.
- An ℓ_1-norm based regularization term of the vector of