Pattern Recognition 47 (2014) 854–864

Contents lists available at ScienceDirect

Pattern Recognition

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

Approximate polytope ensemble for one-class classification

Pierluigi Casale [a,*], Oriol Pujol [b], Petia Radeva [b]

[a] ACTLab, Signal Processing Group, TU Eindhoven, Den Dolech 2, Eindhoven, 5612AZ, The Netherlands
[b] Department de Matemàtica Aplicada i Anàlisi, Universitat de Barcelona, Gran Via 585, Barcelona 08007, Spain

ARTICLE INFO

Article history:
Received 5 November 2012
Received in revised form
22 July 2013
Accepted 4 August 2013
Available online 22 August 2013

Keywords:
One-class classification
Convex hull
High-dimensionality
Random projections
Ensemble learning

ABSTRACT

In this work, a new one-class classification ensemble strategy called approximate polytope ensemble is presented. The main contribution of the paper is threefold. First, the geometrical concept of convex hull is used to define the boundary of the target class defining the problem. Expansions and contractions of this geometrical structure are introduced in order to avoid over-fitting. Second, the decision whether a point belongs to the convex hull model in high dimensional spaces is approximated by means of random projections and an ensemble decision process. Finally, a tiling strategy is proposed in order to model non-convex structures. Experimental results show that the proposed strategy is significantly better than state of the art one-class classification methods on over 200 datasets.

1. Introduction

In pattern recognition, a particular typology of problems is defined when only data related to one target class are available. These problems are known as one-class classification problems. These classification tasks naturally arise when target data can be effortlessly collected while counterexamples are scarce or difficult to obtain [1]. Typical one-class problems are the prediction of mean time before a machinery failure [1,2] or the problem of banknotes verification [3]. In the former case, examples of non-regular operations can only be found in the presence of cracks and malfunctions. In the latter case, while it is possible to easily provide all the examples necessary to model the examples of valid banknotes, it is hard to define a proper sampling of examples belonging to the counterexample set. Effective one-class classification strategies use density estimation methods and boundary methods to model the target class. Gaussian Model, Mixture of Gaussian Model and Parzen Density Estimation are density estimation methods widely used. Density estimation methods work well when there exists *a-priori* knowledge of the problem at hand and a big amount of data is available. On the other hand, boundary methods only intend to model the boundary of the problem without focusing on the complete description of the underlying distribution. Well known approaches to boundary methods are k-centers [4], Nearest Neighbors [2]) and extensions of Support Vector Machines (SVM) to the one-class setting [5–7]. Support Vectors Data Description (SVDD) [2] represents the state of the art in one-class classification. SVDD computes the minimum hypersphere containing all the data in a multi-dimensional space, providing an elegant and intuitive understanding about the solution of the classification problem. Indeed, many classification problems can be efficiently solved when addressed from the geometrical point of view. In particular, when the geometrical framework is taken into account, the *convex hull*, i.e., the smallest polytope containing the full set of points, may represent an even more general structure than the hypersphere.

The convex hull has always been considered a powerful tool in geometrical pattern recognition [8–11]. Bennet et al. [12,13] showed that there exists a geometrical interpretation of the SVM related to the convex hull, i.e., finding the maximum margin between two classes is equivalent to finding the nearest neighbors in the convex hull of each class when classes do not overlap.

Nevertheless, using the convex hull in real applications is limited by the fact that its computation in high dimensional spaces has an extremely large computational cost. Although advanced solutions have been proposed to overcome the limitation of building the convex hull in high dimensional spaces [14,15], there exists a wide range of pattern recognition problems where the use of the convex hull is still unsuitable specially when limited computational resources are considered. Hence, theoretical methodologies for building a convex hull in constrained scenarios are worth to be investigated. In addition, detecting whether a point lies inside or outside of the convex hull in high dimensional spaces still remains an open problem. Random projections and high dimensional geometry lie at the heart of several important approximation algorithms [16]. Random projections have been widely used in pattern recognition applications as a tool for dimensionality

* Corresponding author. Tel.: +31 402473547.
E-mail address: p.casale@tue.nl (P. Casale).

0031-3203/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.patcog.2013.08.007

reduction [17–19]. The random projections technique is based on the idea that high dimensional data can be projected into a lower dimensional space without significantly losing the structure of the data. This data structure preservation has been proved by Johnson and Linderstrauss (JL) [20] and it is ensured with high probability if data are projected into a destination space having dimensionality proportional to the logarithm of the cardinality of the dataset. The capability to reduce the dimensionality of the problem without a significant computational effort and loss of structure allows to create very simple and powerful learning techniques. The simplest learning algorithm based on random projections [21] consists in two steps: project the data in a random subspace and run the learning algorithm in that space. Some works suggest that using random projections is equivalent to using the '*kernel trick*' [22]. Other works [23] suggest that random projections can help in the feature selection process and can also provide specific insights in the construction of large margin classifiers. Moreover, even picking a random separator on data projected down to a line, there is a reasonable chance to get a valid weak hypothesis.

In this work,[1] three main contributions are proposed in the context of one-class classification:

1. The geometric structure of the convex hull is proposed for modeling the boundary of the one-class classification problem. Shrunken or enlarged versions of the baseline convex hull of the training data are used to avoid over-fitting and to find the optimal operating point of the classifier. These versions are called *extended convex polytopes* and their growth is governed by a parameter α. Using this model, a new data point is said to belong to the target class if it lies inside the extended convex polytope. The creation of the extended convex polytope is limited by the fact that its computation is unfeasible in high dimensional spaces.
2. This limitation is circumvented by approximating the D-dimensional expanded convex polytope decision by an ensemble of decisions in very low-dimensional spaces $d \ll D$. This new geometric structure is called *approximate convex polytope decision ensemble*. In these low-dimensional spaces, computing the convex hull and establishing whether points lie inside the geometric structure are both well known problems having very computationally efficient solutions [25,26]. In this work, the effect of projecting and constructing the ensemble using two-dimensional and one-dimensional random spaces is analyzed. As a result, a very efficient and powerful one-class classifier is obtained.
3. However, many real world problems are not well modeled using a convex polytope. Thus, an ensemble of convex polytopes is finally proposed in order to approximate the non-convex boundary of the one-class classification problem. The algorithm is based on a tiling strategy and each convex polytope is approximated by the *approximate convex polytope decision ensemble.*

All the proposed one-class methodologies are validated on a set of 5 toy problems with different cardinalities, 82 one-class problems derived from the UCI machine learning repository, 15 problems related to mobile user verification from walking patterns and 100 text categorization datasets. The paper is organized as follows. In Section 2, the proposed one-class classification method based on the convex hull is described in detail. Its extension for modeling non-convex boundaries is described in Section 3. In Section 4, the validation protocol is described and in Section 5, experimental results are presented. In Section 6 some methodological topics of interest are discussed. In particular, discussions on the number of random projections needed by the proposed methodology, the effect of the expansion parameter and the computational complexity in comparison to state of the art one-class classification methods are addressed. Finally Section 7 concludes the paper.

2. Approximate polytopes decision ensemble for one-class classification

One-class classification can be performed by modeling the boundary of the set of points defining the problem. If the boundary encloses a convex area, then the convex hull, defined as the minimal convex set containing all the training points, provides a good general tool for modeling the target class. The *convex hull* of a set $C \subseteq \mathbf{R}^n$, denoted **conv** C, is the smallest convex set that contains C and is defined as the set of all convex combinations of points in C:

$$\textbf{conv}\ C = \left\{ \theta_1 x_1 + \dots + \theta_m x_m | x_i \in C, \theta_i \geq 0, \forall i; \sum_i \theta_i = 1 \right\}.$$

In this scenario, the one-class classification task is reduced to the problem of knowing if test data lie inside or outside the hull. Although the convex hull provides a compact representation of the data, outliers may lead to shapes of the convex polytope not corresponding to the desired model and a decision using this structure is prone to over-fitting. In this sense, it is useful to define a parameterized set of convex polytopes associated with the original convex hull of the training data. This set of polytopes are shrunken/enlarged versions of the original convex hull governed by a parameter α. This expansion parameter allows to calculate the Receiver Operating Characteristic (ROC) curve and set the optimal operating point for the final one-class classifier. Given the set $C \subseteq \mathbf{R}^n$, the *extended convex polytope* is defined with respect to the center point c and expansion parameter α. Vertices of the convex polytope are defined as in

$$v_\alpha : \left\{ v + \alpha \frac{(v-c)}{\| v-c \|} | v \in \textbf{conv}\ C \right\} \tag{1}$$

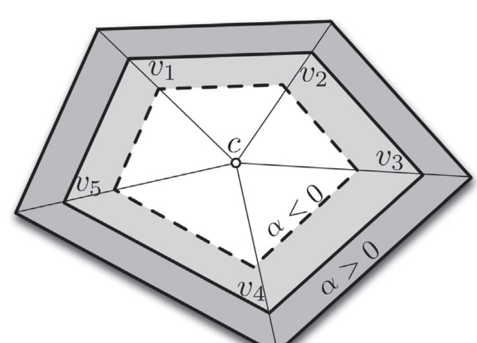

Fig. 1. Illustration of the expanded convex polytope in the 2D space. The central light gray convex polygon represents the original convex hull with vertices $\{v_i, i = 1...5\}$. The outer dark gray polygon corresponds to the enlargement of the original convex hull using $\alpha > 0$. The inner white polygon corresponds to a shrunken version of the convex hull using $\alpha < 0$.

[1] The current paper improves and extends our seminal work [24]. The previous paper shares with the present work the basic idea of using random projections for reducing the complexity of building the convex hull in a high dimensional space. In the current paper the methodology is revised and new contents are introduced, i.e. the use of the expansion factor and a methodology for handling non-convex shapes as well as a large validation section.