Neurocomputing 149 (2015) 957–965

Contents lists available at ScienceDirect

Neurocomputing

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

Maximum margin classification based on flexible convex hulls

Ming Zeng, Yu Yang*, Jinde Zheng, Junsheng Cheng

State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, Hunan University, Changsha 410082, PR China

ARTICLE INFO

Article history:
Received 16 January 2014
Received in revised form
14 May 2014
Accepted 14 July 2014
Communicated by: Shiliang Sun
Available online 2 August 2014

Keywords:
Flexible convex hull
Maximum margin classification
Kernel method
Convex hull
Affine hull
Support vector machine

ABSTRACT

Based on defining a flexible convex hull, a maximum margin classification based on flexible convex hulls (MMC-FCH) is presented in this work. The flexible convex hull defined in our work is a class region approximation looser than a convex hull but tighter than an affine hull. MMC-FCH approximates each class region with a flexible convex hull of its training samples, and then finds a linear separating hyperplane that maximizes the margin between flexible convex hulls by solving a closest pair of points problem. The method can be extended to nonlinear case by using the kernel trick, and multi-class classification problems are dealt with by constructing binary pairwise classifiers as in support vector machine (SVM). The experiments on several databases show that the proposed method compares favorably to the maximum margin classification based on convex hulls (MMC-CH) or affine hulls (MMC-AH).

1. Introduction

Over recent years, as a robust methodology for classification, support vector machine (SVM) [1] has been successfully used in a wide variety of applications including computer vision [2,3], text categorization [4,5], bioinformatics [6,7] and fault diagnosis [8,9]. Moreover, some fruitful methods combining SVM with other learning strategies have also been proposed and bring performance improvements to SVM. Ji et al. [10] proposed a new learning paradigm named multitask multiclass privileged information support vector machine that can take full advantages of the multitask learning and privileged information. Sun and Shawe-Taylor [11] presented a general framework for sparse semi-supervised learning and subsequently proposed a sparse multi-view support vector machine. Wang et al. [12] introduced a novel active learning support vector machine algorithm with adaptive model selection which traces the full solution path of the base classifier before each new query, and then performs efficient model selection using the unlabeled samples. The basic idea of SVM is to construct a separating hyperplane that maximizes the geometric margin which is defined as the distance between the separating hyperplane and the closest samples from two sample sets. From geometrical point of view, in linearly separable case the SVM optimization problem of finding the maximum margin between two sample sets is equivalent to finding the closest pair of points on the respective convex hulls of the two sample sets, and the optimal separating hyperplane is chosen to be the one that perpendicularly bisects the line segment connecting the closest pair of points [13]. That is to say, SVM can be regarded as a maximum margin classification based on convex hulls (MMC-CH) which actually approximates each class region with a convex hull, and the two closest points on the convex hulls determine the hyperplane for separating the convex hulls. In practice, the samples we can obtain are always finite, but the real number of samples belonging to the class region of these samples should be actually infinite. SVM approximates the class region of these samples with a convex hull rather than the isolated samples themselves, thus the samples extends to be infinite.

Actually, relevant researchers have also proposed some other analogous geometric models to approximate class regions, such as affine hulls, hyperspheres and hyperellipsoids. In the spirit of the geometric interpretation for SVM, Zhou et al. [14] and Cevikalp et al. [15] have independently presented a maximum margin classification based on affine hulls (MMC-AH). In contrast to SVM, this method approximates class regions with affine hulls of their sample sets rather than convex hulls, and then the separating hyperplane can be determined by solving the closest pair of points problem. Tax and Duin [16,17] introduced a method called support vector data description (SVDD) for novelty or outlier detection. SVDD tries to find a smallest hypersphere containing almost all sample points for approximating the class region and classifies an unknown sample according to the Euclidean distance from this sample to the center of the hypersphere. Moreover, the hypersphere approximation is also extended to multi-class classification

* Corresponding author.
E-mail address: yangyu@hnu.edu.cn (Y. Yang).

http://dx.doi.org/10.1016/j.neucom.2014.07.038
0925-2312/

problems. Lee and Lee [18] employed a hypersphere for approximating the class region of each sample set and then utilize these approximations to classify an unknown sample via Bayesian decision rule. Wei et al. [19] also proposed another one-class classification method, which replaces the hypersphere with a hyperellipsoid for the class region approximation in SVDD and thus utilizes the Mahalanobis distance rather than Euclidean distance as the decision rule. No matter for one-class or multi-class classification problems, these classification methods based on geometric models have one thing in common, namely, a kind of geometric model (convex hulls, affine hulls, hyperspheres or hyperellipsoids etc.) is used to approximate the class region of each sample set and then the classification models are built based on certain decision rule. Therefore, these methods offer us a good idea for classification that one can start with geometric approximation models to class regions.

The convex hull is the smallest convex set containing given finite samples, so it could approximate the class region very tightly. Generally speaking, the natural class region almost always extends beyond the convex hull of its finite samples, especially in high-dimensional spaces. In this sense, the unrealistically tight convex hull is typically a substantial under-approximation to the class region. As opposed to the convex hull, the affine hull gives a rather loose approximation, because it does not constrain the positions of the sample points within the affine subspace. Besides, linear separability of samples does not necessarily guarantee the separability of corresponding affine hulls [15]. It is more restrictive from this point of view. Affine hulls go to infinity in every direction thus they will not overlap only if they are parallel to each other. The hypersphere model is particularly suitable for samples with a spherical distribution and high clustering degree; otherwise, large empty space will appear inside the hypersphere. In this case, the hypersphere also gives a relatively loose approximation to the class region. The large empty space is likely to misclassify samples from other class regions or outliers generated by noise and interference to such approximated class region. Compared with the hypersphere, the hyperellipsoid model takes the distribution information of samples into consideration and thus approximates the class region more tightly. However, in the case of high-dimensional small samples the covariance matrix tends to be singular, which brings some algorithm obstacles to the classification methods based on hyperellipsoids in high-dimensional feature spaces. Although some skills have been developed to improve the calculation of the covariance matrix, but they still have many more parameters owing to the need to represent the covariance matrix.

In a word, it is still worthwhile for us to explore another appropriate geometric model for approximating the class region of samples in the pattern recognition research. Inspired by convex hulls and affine hulls, in this study we define a new geometric model called a flexible convex hull for the class region approximation and propose a novel classification method, i.e., maximum margin classification based on flexible convex hulls (MMC-FCH). The basic goal of MMC-FCH is to find an optimal linear separating hyperplane that yields the maximum margin between flexible convex hulls of sample sets. This can be solved by computing a closest pair of points problem and then the optimal separating hyperplane is chosen to be the one that perpendicularly bisects the line segment connecting the closest pair of points. MMC-FCH can also be extended to nonlinear case by using the kernel trick. To use the proposed method in multi-class classification problems, we can consider most of the common strategies developed for extending binary SVM classifiers to the multi-class cases.

The remaining part of the paper is organized as follows. Section 2 gives the definition of a flexible convex hull. Section 3 introduces the proposed method, MMC-FCH. Section 4 presents our experimental results and Section 5 concludes the paper.

2. Definition of a flexible convex hull

2.1. Motivation

As mentioned above, SVM can be regarded as a maximum margin classification based on convex hulls (MMC-CH), which first approximates each class with a convex hull of its training samples and then finds a hyperplane that maximizes the margin between the two convex hulls. The convex hull of a sample set can be expressed as a linear combination of the sample points from the sample set where all coefficients are non-negative and sum to one. Consider a finite sample set $\boldsymbol{X}=\{\boldsymbol{x}_i\}$ $(i=1,2,\ldots,n)$, $\boldsymbol{x}_i \in \mathbb{R}^p$, then the convex hull of the sample set $\boldsymbol{X}$ can be written as

$$\mathrm{conv}(\boldsymbol{X})=\left\{\sum_{i=1}^{n}\alpha_i\boldsymbol{x}_i \,\middle|\, \sum_{i=1}^{n}\alpha_i=1, 0\le\alpha_i\le 1\right\} \tag{1}$$

where α_i is the combination coefficient of the ith sample point $\boldsymbol{x}_i$. The convex hull model is the tightest possible convex approximation to the class region, and for classes with more general convex forms, it is typically a substantial under-approximation.

Other than convex hull models, maximum margin classification based on affine hulls (MMC-AH) approximates each class with an affine hull [15]. The affine hull of a sample set can be expressed as a linear combination of the sample points where coefficients add up to one without non-negativity constraints. Then, the affine hull of the sample set $\boldsymbol{X}$ can be written as

$$\mathrm{aff}(\boldsymbol{X})=\left\{\sum_{i=1}^{n}\alpha_i\boldsymbol{x}_i \,\middle|\, \sum_{i=1}^{n}\alpha_i=1\right\} \tag{2}$$

where α_i is the combination coefficient of the ith sample point $\boldsymbol{x}_i$. Even though the affine hull is an unbounded and hence typically rather loose model to the class region in contrast to the convex hull approximation, MMC-AH works surprisingly better than SVM (or MMC-CH) especially in high-dimensional spaces with limited number of samples [15]. This is one indication that convex hulls based methods may be too tight to be realistic. Nevertheless, due to the unboundedness of affine hulls, the separability of affine hulls requires they are certainly parallel to each other. If different sample sets have similar or intersecting affine hulls but very different distributions of samples within their affine hulls, MMC-AH would fail to separate the sample sets. Therefore, it seems more reasonable to tighten the affine hull model.

2.2. Flexible convex hulls

Motivated by convex hulls and affine hulls, we define a new geometric model called a flexible convex hull for the class region approximation. Similar to the convex hull and the affine hull, a flexible convex hull can also be expressed as a linear combination of sample points where coefficients add up to one, but it imposes different lower and upper bounds on the coefficients. More formally, the flexible convex hull of the sample set $\boldsymbol{X}$ is define as

$$\mathrm{flex}(\boldsymbol{X})=\left\{\sum_{i=1}^{n}\alpha_i\boldsymbol{x}_i \,\middle|\, \sum_{i=1}^{n}\alpha_i=1, \frac{1-\lambda}{n}\le\alpha_i\le\frac{1-\lambda}{n}+\lambda\right\} \tag{3}$$

where α_i is the combination coefficient of the ith sample point $\boldsymbol{x}_i$ and $\lambda\in(1,+\infty)$ is the flexible factor. The flexible convex hull and the flexible factor both have their own explicit geometric interpretations. For a given λ, an arbitrary sample point $\boldsymbol{x}_i\in\boldsymbol{X}$ extends along the radial direction $\overrightarrow{\bar{\boldsymbol{x}}\boldsymbol{x}_i}$ with respect to λ where $\bar{\boldsymbol{x}}=(1/n)\Sigma_{i=1}^{n}\boldsymbol{x}_i$ is the center of the sample set $\boldsymbol{X}$, and the corresponding extended point can be written as $\boldsymbol{x}_i'=(1-\lambda)\bar{\boldsymbol{x}}+\lambda\boldsymbol{x}_i$, then the convex hull of the new sample set $\boldsymbol{X}'=\{\boldsymbol{x}_i'\}(i=1,2,\ldots,n)$ is exactly the flexible convex hull of the original sample set $\boldsymbol{X}$; moreover, the flexible factor λ is exactly the ratio of the distance between the extended point $\boldsymbol{x}_i'$ and the set