Neurocomputing 124 (2014) 105–110

Contents lists available at ScienceDirect

Neurocomputing

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

E2LSH based multiple kernel approach for object detection

Ruijie Zhang *, Fushan Wei, Bicheng Li

Zhengzhou Information Science and Technology Institute, China

ARTICLE INFO

Article history:
Received 24 December 2012
Received in revised form
24 July 2013
Accepted 29 July 2013
Communicated by T. Heskes
Available online 19 August 2013

Keywords:
Object detection
E2LSH
Multiple kernel learning
Hadamard product

ABSTRACT

Multiple kernel learning (MKL) methods is widely used in object detection. The conventional MKL methods employ a linear and stationary kernel combination format which cannot accurately describe the distributions of complex data. This paper proposes an E2LSH based clustering algorithm which combines the advantages of nonlinear multiple kernel combination methods—E2LSH-MKL. E2LSH-MKL is a nonlinear and nonstationary multiple kernel learning method. This method utilizes the Hadamard product to realize nonlinear combination of multiple different kernels in order to make full use of information generated from the nonlinear interaction of different kernels. Besides, the method employs E2LSH-based clustering algorithm to group images into subsets, then assigns cluster-related kernel weights according to relative contributions of different kernels on each image subset to realize nonstationary weighting of multiple kernels to improve learning performance. Finally, E2LSH-MKL is applied to object detection. Experiment results on datasets of TRECVID 2005 and Caltech-256 show that our method is superior to the state-of-the-art multiple kernel learning methods.

1. Introduction

Object detection has a wide range applications in image and video indexing and retrieval, such as pattern recognition [1,2], multi-object detection, tracking [3,4], etc. Usually, a object detector is learned by training on labeled datasets using machine learning methods. For an unknown image or video key frame, put it directly into the trained detectors to obtain the concept type. However, object detection is a much challenging task. The essential reason lies in that the images within a category would still exhibit diversity [5]. We may refer to such phenomena as “intra-class diversity”.

Currently, multiple kernel learning (MKL) methods [6] are widely used in object learning [7,8] and have shown an excellent detection performance. Compared to a single kernel in the traditional support vector machine (SVM) [9], MKL learns an optimal kernel combination and the associated classifier simultaneously, providing an effective way of fusing informative features and kernels. Therefore, MKL is a more flexible kind of kernel-based learning model. The theories and applications in recent years have shown that MKL achieves a more superior performance than single kernel based classifiers [10].

However, existing MKL classifiers mostly adopt a stationary combination of kernel functions, that is these methods basically adopt a uniform similarity measure over the whole input space. When a category exhibits high variation as well as correlation with other categories in appearance, they are difficult to cope with the complexity of data distribution. Therefore, several sample-based methods [11,12] have been proposed to capture the characteristics of individual samples. For example, a sample-specific ensemble kernel learning method is proposed in [11] to explore the relative contributions of distinct kernels for each sample. In practice, such methods have yielded promising discriminative power. But expensive computation is incurred to learn sample-based similarity measures. More importantly, heavily respecting individual samples may overwhelm the intrinsic properties of a category so as to make the classifier less reliable.

Considering the intra-class diversity of images, Yang [13] proposes a nonstationary combination of kernel functions Group Sensitive Multiple Kernel Learning (GS-MKL) to perform object learning [5]. In GS-MKL, K-Means is firstly used to partition image features into several clusters and then multiple kernel learning classifiers are trained based on these feature clusters. In this way, different kernel weights are assigned to different feature clusters. However, K-Means clustering algorithm has some drawbacks. Firstly, K-Means is variance-based [1], it will award more clusters to high frequency areas of the feature space, leaving less clusters for the remaining areas. Since frequently occurring features are not necessarily informative, this over-sampling of dense regions is inappropriate [14]. Secondly, K-Means does not support dynamic expansion of clustering data [15], therefore once new images are added, it needs to cluster the new data set again. Thirdly, it is computationally very expensive as it involves several distance calculations of each data point from all the centers in each iteration. Besides, the initial cluster centers

* Correspondence to: Henan Province, Zhengzhou, Jinshui District, Jianxue Street 7#, 1001 Mailbox 835#, 450002, China. Tel.: +86 18637119588.
E-mail addresses: rjz_wonder@163.com (R. Zhang), weifs831020@163.com (F. Wei), lbc_lm@163.com (B. Li).

0925-2312/$ - see front matter Crown Copyright © 2013 Published by Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2013.07.027

in K-means algorithm are randomly generated, thus the clustering results are unstable and susceptible to noise data points. And the final cluster results heavily depend on the selection of initial centers which cause it to converge at local optimum.

To overcome these problems, we propose a E2LSH-MKL based object detection method. Exact Euclidean LSH (E2LSH) [16] is a scheme of Locality Sensitive Hashing (LSH) [17] realized in Euclidean space and has a wide range application in fast approximate nearest neighbor search and indexing for large scale high-dimensional data. The basic idea of LSH is to use several locality sensitive hashing functions to map high-dimensional data points into low-dimensional space ensuring that the closer points in high-dimensional space are still close to each other in low-dimensional space. According to this idea, we employ E2LSH instead of K-Means to perform image features' clustering. In E2LSH-MKL, firstly, E2LSH is used to partition image features into different clusters. Secondly, an E2LSH clustering based multiple kernel learning classifier is trained based on these feature clusters, which assign different kernel weights to different feature clusters, thus to capture intra-class diversity of images effectively. Finally, an unknown image or video key frame is input to the trained E2LSH-MKL classifier to obtain its object type.

The remainder of this paper is organized as follows. In Section 2 we briefly introduce the related theories, including GS-MKL and E2LSH clustering algorithm. In Section 3, we introduce E2LSH-MKL and give out its coefficients optimization procedure. Section 4 depicts E2LSH-MKL based object detection algorithm in detail. We present the experimental results in Section 5. Finally we conclude this paper in Section 6.

2. Preliminaries

2.1. GS-MKL

Multiple kernel learning was first proposed by Pavlidis [18] to fuse heterogeneous data for gene functional classification and then extended to multi-class classification in [19]. Canonical MKL learns a convex kernel combination and the associated classifier jointly, thereby selecting informative features and discriminative kernels. The combination of multi-kernels is defined as follows:

$$K(x_i, x) = \sum_{m=1}^{M} \beta_m K_m(x_i, x) \tag{1}$$

where $\{\beta_m\}_{m=1}^{M}$ are kernel weights, $\sum_{m=1}^{M}\beta_m = 1$ and $\beta_m \ge 0$. M is the total number of kernels. K_m is a positive definite kernel associated with a reproducing kernel Hilbert space (RKHS). Each K_m can employ different kernel functions and use different feature subsets or data representations.

Canonical MKL employs a uniform kernel combination over the whole input space. Instead of learning a global kernel combination, Yang [5] employs a nonstationary kernel weighting method, Group Sensitive Multiple Kernel Learning (GS-MKL). GS-MKL learns a set of group-sensitive kernel combination to adapt with the complexity of data distribution. The kernel weights in GS-MKL not only depend on the kernel functions, but also on the groups that the two images belong to. Let $c(x_i)$ and $c(x_j)$ represent the group ids of image x_i and x_j, respectively. The combined kernel form in Eq. (1) can be rewritten as

$$K(x_i, x_j) = \sum_{m=1}^{M} \beta_m^{c(x_i)} \beta_m^{c(x_j)} K_m(x_i, x_j) \tag{2}$$

where $\beta_m^{c(x_i)}$ and $\beta_m^{c(x_j)}$ are the group-sensitive kernel weights of x_i and x_j. Let G denote the total group number, then $\beta_m^{c(x)} \in \{\beta_m^1, ..., \beta_m^g, ..., \beta_m^G\}$ for $m \in (1, 2, ..., M)$.

Given a labeled image dataset as $D_{train} = \{x_i, y_i\}_{i=1}^{N}$, where x_i is the ith sample and $\{y_i = \pm 1\}$ is the corresponding binary label by human for a given visual concept, our goal is to train a multi-kernels based classifier with a decision function $f(x)$ to predict the object category of an unlabeled image x in dataset D_{test}. The decision function can be given as follows:

$$f(x) = \sum_{i=1}^{N} \alpha_i y_i \sum_{m=1}^{M} \beta_m^{c(x_i)} \beta_m^{c(x)} K_m(x_i, x) + b \tag{3}$$

where $\{\alpha_i\}_{i=1}^{N}$ and b are the coefficients of the classifier, corresponding to the Lagrange multipliers and the bias. $\beta_m^{c(x)}$ are the number of group-sensitive kernel weights, total $G \times M$, $m \in \{1, 2, ..., M\}$, $c(x) \in \{1, 2, ..., G\}$.

2.2. E2LSH clustering

E2LSH is a scheme of LSH realized in Euclidean space and is first proposed by MIT Professor Indyk [16]. E2LSH is widely used in large scale video/image similarity search and rapid retrieval applications. The key idea of LSH is to hash the points using several hash functions so as to ensure that for each function, the probability of collision is much higher for objects which are close to each other than for those which are far apart.

For a point domain S, the LSH family is defined as

Definition 1:. A family $\mathscr{H} = \{h : S \to U\}$ is called locality-sensitive, if for any point $\boldsymbol{q}$, the function $p(t) = \Pr_{\mathscr{H}}[h(\mathbf{q}) = h(\mathbf{v}) : ||\mathbf{q} - \mathbf{v}|| = t]$ is strictly decreasing with t. That is the collision probability of points $\boldsymbol{q}$ and $\boldsymbol{v}$ is decreasing with the distance between them.

According to the choice of hash function h, LSH has many different formats, among which E2LSH is a representative scheme. The LSH function family that E2LSH employed is based on p-stable distributions, $p \in (0, 2]$, which is defined as

Definition 2:. A distribution $\mathcal{D}$ over $\mathfrak{R}$ is called p-stable, if there exists $p \ge 0$ such that for any n real numbers $v_1, ..., v_n$ and variables $X_1, ..., X_n$ with distribution $\mathcal{D}$, the random variable $\sum v_i X_i$ has the same distribution as the variable $(\sum_i |v_i|^p)^{1/p} X$, where X is a random variable with distribution $\mathcal{D}$.

In this paper, all the LSH functions [17] we employed are 2-stable and defined as

$$h(\mathbf{v}) = \left\lfloor \frac{\boldsymbol{\alpha} \cdot \mathbf{v} + \beta}{\varpi} \right\rfloor \tag{4}$$

where $\lfloor \cdot \rfloor$ is the floor operation, $\boldsymbol{\alpha}$ is a d-dimensional vector with components that are selected at random from a p-stable distribution. β is a random variable uniformly distributed in $[0, \varpi]$, and ϖ is a constant. The hash function $h(\mathbf{v})$ maps a d-dimensional vector $\mathbf{v}$ into the integer set.

E2LSH usually combines k Locality Sensitive Hashing functions, a function set is defined as

$$\mathcal{G} = \{g : S \to U^k\} \tag{5}$$

where $g(\mathbf{v}) = (h_1(\mathbf{v}), ..., h_k(\mathbf{v}))$, for each $\mathbf{v} \in \mathbb{R}^d$, k dimensional vector $\mathbf{a} = (a_1, a_2, ... a_k)$ is obtained after the mapping of $g(\boldsymbol{v}) \in \mathcal{G}$. Then a primary hash function $hash_1$ and secondary hash function $hash_2$ are utilized to hash the vector $\mathbf{a} = (a_1, a_2, ... a_k)$, constructing the hash tables and saving the data points. $hash_1$ and $hash_2$ are defined [20] as follows:

$$hash_1(\mathbf{a}) = ((\textstyle\sum_{i=1}^{k} r_i' a_i) \bmod m) \bmod s \tag{6}$$

$$hash_2(\mathbf{a}) = (\textstyle\sum_{i=1}^{k} r_i'' a_i) \bmod m \tag{7}$$

where r_i' and r_i'' are random integers, s is the size of hash tables, and m is a prime number. E2LSH puts data points having the same