SIMILARITY-BASED IMAGE CLASSIFICATION VIA KERNELIZED SPARSE REPRESENTATION

Zhi Zeng, Heping Li, Wei Liang and Shuwu Zhang

Institute of Automation, Chinese Academy of Sciences, Beijing, China

ABSTRACT

We consider the image classification problem based on the similarities between images. The choice of the similarity is related to the particular applications, and it could be based on color, texture, bag-of-features, or even more complex kernels. As long as the pair-wise similarity matrix is transformed into a positive semidefinite one, the similarities of images could be treated as kernels. This transformation makes it possible for kernel methods to solve the similarity-based image classification problem. In this paper, we propose a novel kernelized classification framework based on sparse representation. This new framework casts the classification as finding a sparse linear representation of test image with respect to training images. Unlike the former works, we do this sparse coding procedure through a proposed kernelized orthogonal matching pursuit algorithm, which is performed in inner product space rather than Euclidean space. Through a proper choice of the similarity function, the proposed approach can be applied to diverse image classification problems. Comparative experiments between the proposed method and other existing methods, on two real datasets (*Caltech*-101 and *Face Rec*) show that our method performed better.

Keywords— similarity, similarity-based learning, image classification, sparse representation

1. INTRODUCTION

Several important classification/recognition problems in computer vision, such as object recognition [1], natural scene categorization [2], face recognition [3-6] and etc., can be reduced to a general problem of classifying a whole image (e.g. object, scene, human face, etc.) into one of several predefined semantic categories (e.g. scene categories, object categories, individuals, etc.). These problems that have received considerable attention in the recent past, can be generally called image classification. Though the image classification is usually not a very difficult task for humans, it has been proven to be an extremely challenging problem for machines. In the existing literatures, most of the frameworks for image classification include two main steps: feature extraction and classifier learning.

In the first step, some discriminative features are extracted to represent the image content. These feature extraction methods are diverse and related to concrete problem situations. For instance, many subspace methods [3, 4] are proposed and successfully applied to face recognition. When images are composed of several entities organized in an unpredictable layout, bag-of-features methods [2, 7-9], which represent an image as an orderless collection of local features, have recently demonstrated impressive levels of performance for the image classification tasks. However, these features are usually Euclidean vectors, which mean that the images are explicitly embedded in a Euclidean space. In this paper, we want to eliminate this embedding process, and consider image classification problem based on the similarities between the test image and a set of labeled training images, and the pairwise similarities between the training images.

In the classifier learning step, various multi-class classifiers have been applied to classify images. The most popular two classifiers are *k*-Nearest-Neighbor (*k*NN) and Support Vector Machine (SVM). Moreover, kernel-based classifiers, especially (nonlinear) SVM, which use a kernel function to non-linearly map two images from the input space to the inner product in some space, have been widely applied. Over the years, many new kernels [7-9] have been proposed to improve the classifier's performance. Although the mathematical meaning of a kernel is the inner product in some Hilbert space, a standard interpretation of kernel is the pairwise similarity between different samples [10]. Thus, kernel method can be used to solve the similarity-based classification.

In addition, J. Wright *et al*. [5] proposed a robust face recognition method based on sparse representation (SR). This method uses the SR of each individual test sample directly for classification and adaptively selects the training samples that give the most compact representation. However, this method can not be directly applied to similarity-based classification, as it does sparse coding procedure via l^1-minimization [11], which need explicitly embed the samples in Euclidean space.

Being motivated from the kernelization of SVM, we extend the orthogonal matching pursuit (OMP) algorithm [13] to compute the SR based on the kernel. After that, the obtained SR of test image can be applied to solve the similarity-based image classification problem.

The paper is organized as follows: Section 2 formulates the problem, and the proposed method is represented in Section 3. In Section 4, the experiments are performed to evaluate our method. Finally, we summarize our work in Section 5.

2. PROBLEM FORMULATION

In this section, we formulate the similarity-based image classification problem. We are given a labeled training set of n images $\{(x_1, l_1), (x_2, l_2), \ldots, (x_n, l_n)\}$ in this problem. Here, each x_i is an input image and $l_i \in \{1, 2, \ldots, k\}$ is the corresponding category label. Instead of direct access to the feature of the images, we have a similarity function $S(\cdot, \cdot)$ for any pair of samples. Then, a $n \times n$ similarity matrix S with (i, j)-entry $S(x_i, x_j)$ can be obtained. In the test stage, we need estimate the category label for an unknown test image t based on its semantic content, or its similarities to the training images $S(t, x_i)$, $i = 1, 2, \ldots, n$. A $n \times 1$ vector $\boldsymbol{v}$ with ith element $S(t, x_i)$ is used to represent these similarities.

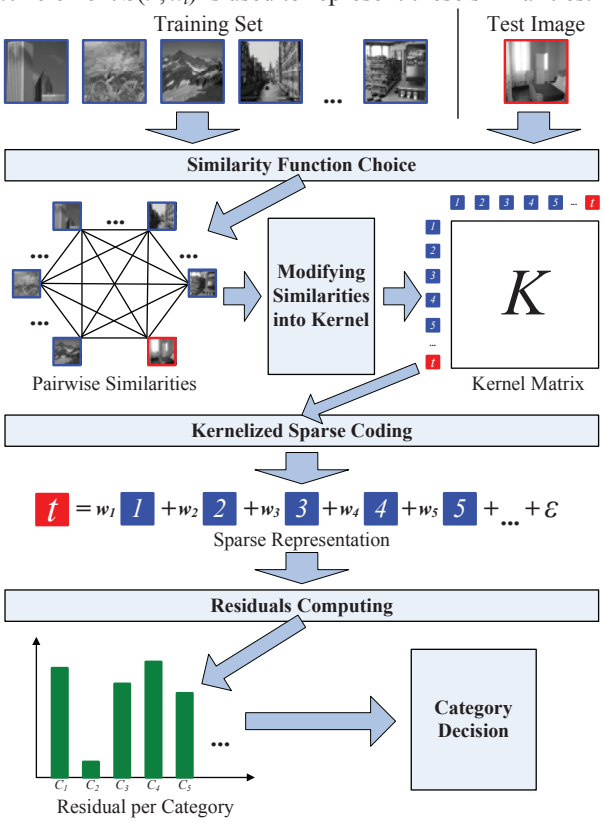

Fig. 1. Overview of the proposed image classification framework.

It is important to note that the function $S(\cdot, \cdot)$ can be any well-defined similarity function or kernel, and its choice depends on the application. In this paper, we suppose $S(\cdot, \cdot)$ is symmetric, which means $S(a, b) = S(b, a)$. This assumption does not affect the usage of our method because most similarity functions satisfy the symmetry.

Now, we are ready to show the details of our approach.

3. PROPOSED METHOD

In this section, we introduce the proposed similarity-based kernelized image classification framework. Fig. 1 is a summary of our approach. After the choosing of the similarity function, we need first modify the similarity matrix into kernel. This step is equal to make the similarity matrix positive semidefinite (PSD), which is necessary to utilize kernel-based classification methods. Then, based on the obtained kernel matrix, the SR of the test image in the training set is computed by a kernelized sparse coding algorithm, which is developed from the OMP algorithm and called kernelized OMP in this paper. The obtained SR can be used to classify the test image, which is similar to the method in [5].

3.1. Modify Similarities into Kernels

A popular approach to similarity-based classification is to treat the given similarities as inner products in some Hilbert space [10]. This approach treats similarities as kernels and applies a certain kernel-based classification algorithm to do the classification task.

However, the theory of kernel method requires the kernel matrix K must be PSD, and many similarity functions do not satisfy this property. Thus, it is necessary to modify the similarities into kernels.

In [10], the authors have discussed four methods to modify similarities into kernels, which include spectrum clip, spectrum flip, spectrum shift and spectrum square. The common key idea of these four different modification methods is to alter the negative eigenvalues of similarity matrix to be non-negative. Compared these four methods, we choose spectrum clip to modify similarity matrix in this paper. This is due to its good performance and easy to modify the training and test similarities in a consistent fashion.

Since similarity matrix S is assumed to be symmetric, it has an eigenvalue decomposition $S = U^T \Lambda U$, where U is an orthogonal matrix and Λ is a diagonal matrix of real eigenvalues, that $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. Spectrum clip is to modify the similarities by linear transformation, that is, $K = PS$ and $\boldsymbol{s} = P\boldsymbol{v}$. P is the corresponding transformation matrix and $P = U^T \Lambda' U$, where $\Lambda' = \mathrm{diag}(I_{\{\lambda_1 \geq 0\}}, \ldots, I_{\{\lambda_n \geq 0\}})$ and $I_{\{\cdot\}}$ is the indicator function.

It is important to note that this modification step is not necessary when the similarity is already PSD.

3.2. Kernelized Sparse Coding

The goal of sparse coding is to represent input vectors approximately as a weighted linear combination of a small number of "basis vectors". It is solving the minimization problems:

$$\min_w \|w\|_0 \quad \text{subject to} \quad t = Xw \tag{1}$$

or

$$\min_w \|w\|_0 \quad \text{subject to} \quad \|t - Xw\|_2 \le \varepsilon \tag{2}$$

where $\|\cdot\|_0$ is the l^0 norm, counting the nonzero entries of a vector.

Exact determination of sparest representation proves to be an NP-hard problem [5]. Thus, approximate solutions are considered instead. The approaches come in two flavors. The one is convex relaxation, which replaces the combinatorial sparse approximation problem with a related convex program. A well-known approach is the basis pursuit (BP) [11]. It suggests a convexification of the problem posed in (1) and (2) by replacing the l^0 norm with an l^1 norm and solving by linear programming. J. Wright *et al*. [5] used this pursuit algorithm in their SR-based classification. The other is greedy algorithms that select the "basis vectors" sequentially. These methods include matching pursuit (MP) [12] and OMP [13].

These methods are all designed to solve the SR problem in Euclid space; they can not be directly applied to our similarity-based or kernel-based problem. However, we found that the greedy algorithms are only involving the computation of inner products between the signal and "basis vectors". Thus, they can be kernelized. In this paper, we kernelize the OMP algorithm due to its good performance [13], and present the kernelized orthogonal matching pursuit (KOMP) algorithm. It is given as follows:

Algorithm 1. Kernelized Orthogonal Matching Pursuit

1. **Input**: a predefined parameter γ, a $n \times n$ kernel matrix K with (i, j)-entry $K(x_i, x_j)$ and a $n \times 1$ vector $\boldsymbol{s}$ with ith element $K(x_i, t)$, where $x_1, x_2, \ldots x_n$ are the training images, t is the test image and $K(\cdot, \cdot)$ is the kernel function.

2. Initialize $\boldsymbol{w} = \boldsymbol{0}$, active set $\theta = \{\}$, $l = 1$ and $R_0 = K(t, t)$.

3. Select index λ_l by solving an easy optimization problem:

$$\begin{aligned} \lambda_l &= \arg\max_{\omega \in [1,n]-\theta} \left| \langle x_\omega, r \rangle \right| \\ &= \arg\max_{\omega \in [1,n]-\theta} \left| \langle x_\omega, t \rangle - \sum_{i \in \theta} w_i \langle x_\omega, x_i \rangle \right| \\ &= \arg\max_{\omega \in [1,n]-\theta} \left| s_\omega - K_\omega w \right| \end{aligned} \tag{3}$$

where $r = t - \sum_{i \in \theta} w_i x_i$ is the residual, s_ω is the ω-th element of vector $\boldsymbol{s}$ and K_ω is the ω-th row of matrix K.

4. Set active set $\theta = \{\lambda_l\} \cup \theta$. Let K' be a submatrix of K that contains only the elements corresponding to the active set θ, and let $\boldsymbol{s}'$ and $\boldsymbol{w}'$ be subvector of $\boldsymbol{s}$ and $\boldsymbol{w}$ corresponding to the active set θ. Renew $\boldsymbol{w}'$ by solving a least-squares minimization problem:

$$\begin{aligned} w' &= \arg\min_{w'} \left\| t - \sum_{i \in \theta} w_i x_i \right\|_2 \\ &= \arg\min_{w'} (w'^T K' w' - 2 w'^T s') = K'^{-1} s' \end{aligned} \tag{4}$$

5. Compute R_l as follow:

$$R_l = \left\| t - \sum_{i \in \theta} w_i x_i \right\|_2 = K(t,t) - 2 w'^T s' + w'^T K' w' \tag{5}$$

6. If $R_l > 0$ and $(R_{l-1} - R_l)/R_{l-1} > \gamma$, set $l = l+1$ and goto step 3; otherwise return $\boldsymbol{w}$ as the solution.

7. **Output**: The $n \times 1$ representation vector $\boldsymbol{w}$ with ith element w_i.

By using the Algorithm 1, which is only based on the kernel matrix, we can obtain the SR of the test image with respect to training images.

3.3. Classification Based on Sparse Representation

Given a new test image t from one of the classes in the training set, we first compute its SR vector $\boldsymbol{w}$ via the KOMP algorithm. Based on the global SR, one can design many possible classifiers to resolve it. In this paper, we adopt the one used in [5], which classify test image t based on how well the representation coefficients associated with all training images of each object reproduce t. Algorithm 2 below summarizes the complete classification procedure:

Algorithm 2. Similarity-based Classification via Sparse Representation

1. **Input**: a $n \times n$ similarity matrix S with (i, j)-entry $S(x_i, x_j)$ and a $n \times 1$ vector $\boldsymbol{v}$ with ith element $S(x_i, t)$.

2. Modify similarity into kernel by spectrum clip, and obtain a $n \times n$ kernel matrix K and a $n \times 1$ vector $\boldsymbol{s}$.

3. Get the SR vector $\boldsymbol{w}$ by KOMP.

4. Compute the residuals $r_i(t)$ for $i = 1, \ldots, k$:

$$r_i(t) = \left\| t - \sum_{x_j \in C_i} w_j x_j \right\|_2 = K(t,t) - 2 w^{iT} s^i + w^{iT} K^i w^i \tag{6}$$

where $\boldsymbol{w}^i$, $\boldsymbol{s}^i$ and K^i are subvectors (submatrix) of $\boldsymbol{w}$, $\boldsymbol{s}$ and K corresponding to category i.

5. **Output**: identity $(t) = \arg\min_i r_i(t)$.

4. EXPERIMENTAL EVALUATION

In order to evaluate the performance of our approach, we test it on two different datasets: *Caltech*-101 [1] and *Face Rec* [6]. For easy comparison, we directly use the data presented by [10][1], which include the similarities between images and the randomized partitions. The parameter γ is chosen by 10-fold cross-validation on the training set. For each dataset, all experiments are repeated 20 times with different randomly selected training and test images, and finally the mean and standard deviation of the classification errors are reported. We compare our approach with four different similarity-based classification methods presented

[1] The datasets along with the randomized partitions are available at http://idl.ee.washington.edu/SimilarityLearning/

in [10]: *k*NN, kernel ridge interpolation (KRI) weights for *k*NN, SVM-*k*NN and SVM.

4.1. Caltech-101 Dataset

The *Caltech*-101 dataset is an object recognition benchmark dataset and consists of 8677 images from 101 object categories. The significant variation in color, pose and lighting makes this dataset quite challenging. In our experiment, we use the pyramid match kernel [7] on SIFT feature [14] to compute the similarities between images. Due to this similarity is PSD, spectrum clip is not performed in this experiment. Table 1 gives the experimental results:

Table 1. Test misclassified rate (%) on *Caltech*-101 dataset

*k*NN	KRI-*k*NN	SVM-*k*NN	SVM	Our method
41.55 (0.95)	30.13 (0.42)	36.82 (0.60)	33.49 (0.78)	**25.74** (0.40)

As shown in Table 1 above, our method significantly outperforms the baseline methods and achieves best results on *Caltech*-101 dataset.

4.2. Face Recognition

The *Face Rec* dataset consist of 945 sample faces of 139 people from the NIST Face Recognition Grand Challenge dataset. So there are 139 categories and one for each person. Similarities for pairs of the original three-dimensional face data were computed as the cosine similarity between integral invariant signatures based on surface curves of the face [6]. Table 2 gives the experimental results:

Table 2. Test misclassified rate (%) on *Face Rec* dataset

*k*NN	KRI-*k*NN	SVM-*k*NN	SVM	Our method
4.23 (1.43)	4.15 (1.32)	4.23 (1.25)	4.18 (1.25)	**4.02** (1.31)

As shown in Table 2 above, our method outperforms the baseline methods on *Face Rec* dataset.

It is important to note that the chosen similarity is not PSD in this experiment. Thus, spectrum clip is carried on to modify it into a kernel. To demonstrate this necessity, we give the comparison between with and without spectrum clip. The comparison results are shown in Table 3.

Table 3. Comparison between with and without spectrum clip on *Face Rec* dataset

with spectrum clip	**4.02** (1.31)
without spectrum clip	4.81 (1.50)

Without spectrum clip, our kernel-based method is illegal, and its test misclassified rate on the dataset falls to 4.81%, which is much worse than the baseline methods.

5. CONCLUSION

In this paper, we propose a general similarity-based kernelized classification framework to solve a variety of image classification problems. At first, spectrum clip is used to modify similarities into kernels. Then, a proposed KOMP algorithm is performed to do the sparse coding in inner product space. Finally, the SR of the test image in the training set is utilized to find the right category label of the test image. By properly choice of similarity, the proposed approach can be applied to diverse image classification problems. Though experimental results on two real datasets show that our method outperforms the baseline similarity-based learning methods, more experiments are need in the future work.

6. ACKNOWLEDGEMENT

This work has been supported by the National Science and Technology Supporting Program of China under Grant No. 2008BAH26B02-3, 2008BAH21B03-04 and 2008BAH26B03.

7. REFERENCES

[1] L. Fei-Fei, R. Fergus and P. Perona, "Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories," in *Proc. CVPR*, 2004.

[2] A. Bosch, A. Zisserman and X. Munoz, "Scene classification using a hybrid generative/discriminative approach," *IEEE Trans. PAMI*, vol. 30, no. 4, pp. 712-727, April 2008.

[3] P. Belhumeur, J. Hespanda and D. Kriegman, "Eigenfaces versus Fisherfaces: recognition using class specific linear projection," *IEEE Trans. PAMI*, vol. 19, no. 7, pp. 711-720, July 1997.

[4] X. He, S. Yan, Y. Hu, P. Niyogi and H. Zhang, "Face recognition using Laplacianfaces," *IEEE Trans. PAMI*, vol. 27, no. 3, pp. 328-340, March 2005.

[5] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. PAMI*, vol. 31, no. 2, pp. 210-227, February 2009.

[6] S. Feng, H. Krim and I. A. Kogan, "3D face recognition using Euclidean integral invariants signature," in *Proc. IEEE Workshop Statistical Signal Processing*, 2007.

[7] K. Grauman and T. Darrell, "The pyramid match kernel: efficient learning with sets of features," *JMLR*, vol. 8, pp 725-760, April 2007.

[8] S. Lazebnik, C. Schmid and J. Ponce, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories," in *Proc. CVPR*, 2006.

[9] Z. Lu and Horace H.S. Ip, "Image categorization with spatial mismatch kernels," in *Proc. CVPR*, 2009.

[10] Y. Chen, E. K. Garcia, M. R. Gupta, A. Rahimi and L. Cazzanti, "Similarity-based classification: concepts and algorithms," *JMLR*, vol. 10, pp 747-776, March 2009.

[11] S. Chen, D. Donoho and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129-159, 2001.

[12] S. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3397-3415, 1993.

[13] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Trans. Information Theory*, vol. 50, no. 10, pp. 2231-2242, Oct. 2004.

[14] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, no. 2, pp. 91-110, 2004.