IDEAL REGULARIZED KERNEL FOR HYPERSPECTRAL IMAGE CLASSIFICATION

Jiangtao Peng[*]

Hubei University
Faculty of Mathematics and Statistics

Yicong Zhou[†]

University of Macau
Faculty of Science and Technology

ABSTRACT

This paper proposes an ideal regularized composite kernel (IRCK) framework for hyperspectral images (HSI) classification. In learning a composite kernel, IRCK exploits spectral information, spatial information, and label information simultaneously. It incorporates the labels into standard spectral and spatial kernels by means of ideal kernel according to a regularization kernel learning framework, which captures both the sample similarity and label similarity and makes the resulting kernel more appropriate for HSI classification tasks. With the ideal regularization, the kernel learning problem has a simple analytical solution and is very easy to implement. The ideal regularization can be used to improve and refine state-of-the-art kernels, including spectral kernels, spatial kernels and spectral-spatial composite kernels. The effectiveness of the proposed IRCK is validated on the benchmark hyperspectral data set: Indian Pines. Experimental results show the superiority of our ideal regularized composite kernel method over the classical kernel methods.

Index Terms**—** Hyperspectral image classification, ideal kernel, regularization, composite kernel

1. INTRODUCTION

Various hyperspectral image (HSI) classification methods have been developed in the past decades [1]. Traditional HSI classification methods usually discriminate and classify the pixels by measuring the similarity among different spectral curves. The key to success for these classification methods is to learn an accurate similarity metric between samples.

In order to learn a desirable similarity metric, kernel functions and kernel methods are introduced into HSI classification and have shown good classification performance [2]. Kernel methods can solve the high-dimensional HSI classification problem effectively and are easy to measure the linear/nonlinear relations between hyperspectral samples in Reproducing Kernel Hilbert Space (RKHS) [2]. In the HSI classification, there are mainly three kind of kernels: spectral kernels, spatial kernels and spectral-spatial composite kernels. The commonly used spectral kernels are Gaussian radial basis function (RBF), polynomial, and linear kernels [2]. Spectral kernels are constructed based on the spectral information, while spatial kernels use the spatial information. The representative spatial kernels are spatial local mean or standard deviation feature based kernel [3], mean map kernel [4], and region kernel [5]. Joint consideration of the spectral and spatial textural information, four different spatial-spectral composite kernels are proposed [3], including stacked kernel, direct summation kernel, weighted summation kernel, and cross-information kernel. Similarly, sample-cluster composite kernels [4], spatial and spectral activation-function-based composite kernels [6], generalized composite kernels [7], have been proposed for the spectral-spatial classification of HSIs.

However, almost all of the above-mentioned kernel-based methods learn the standard kernels from the samples alone without considering the labels of a data set. In fact, the label information can be used for kernel learning and to refine the standard kernels. Exploiting the labels explicitly, an ideal kernel is constructed [8]. It assigns the sample pair with a kernel value 1 if they belong to the same class, and a kernel value 0 if they belong to the different classes. The ideal kernel incorporates the label similarities. Based on the ideal kernel, an ideal regularization strategy is recently proposed to learn a data-dependent kernel from the labels and shows good performance [9, 10].

In this paper, we propose an ideal regularized composite kernel (IRCK) framework for spatial-spectral classification of HSIs. In IRCK, we consider the spectral and spatial kernels as the initial kernels, and employ an ideal regularization to refine the initial kernels by incorporating the labels into the standard spectral and spatial kernels. Finally, the regularized spatial and spectral kernels are combined to form a composite kernel for the HSI classification. The proposed IRCK algorithm has the following characteristics:

(1) It is simple and easy to implement. The ideal regularization problem has an analytical solution, and the resulted kernel can be expressed as the summation of the standard kernel and regularized kernels.

(2) It is quite effective. The ideal regularization improves the standard kernels. Because of the spectral variability of

[*]Thanks to the support by the NSFC 41501392 and the NSF of Hubei 2015CFB327.

[†]Thanks to the support by FDCT 106/2013/A3 and by the Research Committee at University of Macau under Grants MYRG113(Y1-L3)-FST12-ZYC, MRG001/ZYC/2013/FST and MYRG2014-00003-FST.

HSI samples, the same material may have different spectral curves and different materials may have similar spectral curves. Without using the label information, the standard kernel is extremely difficult to handle these problems. That is, the standard kernel is inaccurate in measuring the similarity between samples using only the data itself. However, the ideal regularization can solve these problems at a certain extent by incorporating the label information into the standard kernel. It embodies both the sample similarity and label similarity.

(3) It is less sensitive to kernel parameters than general composite kernels.

(4) It is easily adapted to different kinds of kernels (spectral, spatial and composite kernels), and different kernel-based classification algorithms (SVMs and extreme learning machines (ELMs)).

2. IDEAL REGULARIZED KERNEL

2.1. Ideal kernel

Given a set of training samples, $\mathcal{L} = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_\ell, y_\ell)\}$, the ideal kernel [11, 8] is defined as:

$$T(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} 1, & y_i = y_j, \\ 0, & y_i \neq y_j. \end{cases} \tag{1}$$

The ideal kernel leads to a perfect classification inspired from an "oracle": two samples $\mathbf{x}_i$ and $\mathbf{x}_j$ should be considered as "similar" (with kernel value 1) if and only if they belong to the same class ($y_i = y_j$) [11, 8]. In other words, the ideal kernel incorporates the label information and reflects the similarity between labels.

2.2. Ideal regularization

In order to embed the label information into a standard kernel K_0 and to learn a desirable kernel K, an ideal regularization kernel learning framework is proposed [9, 10]:

$$\min_{K \succeq 0} D(K, K_0) + \gamma \Omega(K) \tag{2}$$

where $D(\cdot, \cdot)$ denotes the divergence between the matrices, $\Omega(\cdot)$ is a regularization term, γ is a tradeoff parameter, $K \succeq 0$ means K is a symmetric positive semidefinite matrix. The divergence can be chosen as the von Neumann divergence:

$$D(K, K_0) = \mathrm{tr}(K \log K - K \log K_0 - K + K_0) \tag{3}$$

where $\mathrm{tr}(A)$ denotes the trace of matrix A. The regularization term can be defined as: $\Omega(K) = -\mathrm{tr}(KT)$, which encodes the label information of the given data samples [9, 10]. Then, the solution of (2) is:

$$K^* = \exp(\log K_0 + \gamma T) = K_0 \odot \exp(\gamma T) \tag{4}$$

where $\odot$ denotes the dot product between two matrices.

The Taylor expansion of (4) is:

$$K^* = K_0 + \gamma K_0 \odot T + \frac{\gamma^2}{2!} K_0 \odot T^2 + \cdots \tag{5}$$

The first term on the right hand of the equation is the original kernel, and the rest terms on the right hand consist of the regularized kernels. It demonstrates that the ideal regularized kernel is a linear combination of the original kernel and regularized kernels. When $\gamma = 0$, the ideal regularized kernel is reduced to the original kernel. When γ is very small, only the first order regularization term $K_0 \odot T$ plays a role in the ideal regularization. Because T equals to 1 only for sample pairs belonging to the same class, ideal regularization enhances the kernel similarity values on sample pairs in the same class. In other words, ideal regularization exploits the sample similarity in K_0, and meanwhile uses the label similarity in T to make the samples in the same class more similar.

From the equation (4), we can see that the ideal regularized kernel is a dot product between the original kernel and an exponential ideal kernel. The computation of ideal kernel and exponential ideal kernel are relatively simple, so the computational complexity of ideal regularized kernel is almost the same as that of original kernel.

2.3. Ideal regularization composite kernel

In order to learn an ideal spectral kernel K^w and an ideal spatial kernel K^s and hence an ideal spectral-spatial kernel $K^{ws} = (1-\mu)K^w + \mu K^s$, we propose the following ideal regularization composite kernel (IRCK) optimization framework:

$$\begin{aligned} \min_{K^w, K^s \succeq 0} & D(K^w, K^{w0}) + D(K^s, K^{s0}) + \gamma \Omega(K^{ws}) \\ &= \mathrm{tr}(K^w \log K^w - K^w \log K^{w0} - K^w + K^{w0}) \\ &\quad + \mathrm{tr}(K^s \log K^s - K^s \log K^{s0} - K^s + K^{s0}) \\ &\quad - \gamma\big((1-\mu)\mathrm{tr}(K^w T) + \mu \mathrm{tr}(K^s T)\big) \end{aligned} \tag{6}$$

The optimal solution of (6) is:

$$K^w = K^{w0} \odot \exp(\gamma(1-\mu)T) \tag{7}$$

$$K^s = K^{s0} \odot \exp(\gamma \mu T) \tag{8}$$

And the composite ideal regularized kernel is:

$$\begin{aligned} K^{ws} &= (1-\mu)K^w + \mu K^s \\ &= (1-\mu)K^{w0} \odot \exp(\gamma(1-\mu)T) \\ &\quad + \mu K^{s0} \odot \exp(\gamma \mu T) \end{aligned} \tag{9}$$

If ideal spatial kernel is the mean map kernel K^m learned from the initial mean map kernel K^{m0} [4], the composite kernel is:

$$K^{wm} = (1-\mu)K^w + \mu K^m \tag{10}$$

3. EXPERIMENTAL RESULTS

The Indian Pines data set acquired by the AVIRIS sensor in 1992 is used in the experiment. The image scene contains with 145×145 pixels and 220 spectral bands, where 20 channels were discarded because of atmospheric affection. There are 16 classes in the data. The total number of samples is 10249 ranging from 20 to 2455 in each class.

The proposed ideal regularized kernel classification method is compared with the classical kernel classification methods, including spectral SVM (K^{ω}), spatial SVM (K^{s}), spectral-spatial SVM (SVM with composite kernel, SVM-CK, $K^{\omega s}$) [3], SVM with mean map kernel (K^{m}), and SVM with composite mean map kernel ($K^{\omega m}$) [4]. The classification overall accuracy (OA) on the testing set is recoded. All data are normalized to have a unit ℓ_2 norm. Gaussian kernel is used in all SVM algorithms. For the spatial-based methods, 9×9 neighborhood window is used.

We investigate the performance of the proposed ideal regularized composite kernel methods under different numbers of labeled samples per class. We randomly choose $M = 5, 10, 15, 20, 25, 30, 35, 40$ samples from each class to form the training set, respectively (For the class less than M samples, half of total samples are chosen). The remaining samples consist of testing set. The classification overall accuracies under different numbers of training samples are shown in Table 1. From results in the table, we can conclude:

(1) With the increase of training samples, OAs for all algorithms are greatly improved. The proposed ideal regularized composite kernel methods show a significant improvement over the spectral, spatial, and spectral-spatial SVMs.

(2) The ideal regularized kernels improve the corresponding original kernels. It demonstrates that the ideal regularization can enhance the kernel discriminant ability.

(3) The ideal regularized composite mean map kernel ($K^{\omega m}$-IR) provides the best classification results than other methods. Compared with the original composite mean map kernel (K^{m}-Ori), $K^{\omega m}$-IR increases the OA of different number of labeled samples by 3.7% in average.

(4) The proposed ideal regularized kernel $K^{\omega m}$-IR is quite effective in the case with limited training samples, seeing the results in the case of $M = 5$ or $M = 10$. When the number of labeled samples is limited, the kernel similarity measured by samples is insufficient to reflect the class discrepancy. In this case, the label similarity in ideal kernel can assist the sample similarity to obtain a reliable metric and desirable classification result.

(5) For the spectral SVM (K^{ω}), the ideal regularized kernel has little or no improvements than the original kernel. However, for the spatial SVM (K^{s}), the corresponding ideal regularized kernel largely improves the original kernel. Because the ideal regularization enhances the kernel similarity on samples in the same class, the spectral kernel similarity is relatively inaccurate than spatial kernel similarity, ideal regularization on spectral kernel is less effective than ideal regularization on spatial kernel.

In the following, we investigate the sensitivity of the proposed algorithm on different parameters. We take the SVM-CK-IR algorithm as an example, and choose 15 and 100 labeled samples per class to form the training set and validating set, respectively. We first investigate the effect of regularization parameter γ on SVM-CK-IR. The OAs versus γ are shown in Fig. 1, where the proposed ideal regularized kernel method provides stable results over a wide range of regularization parameters. Then, we analysis the effect of kernel parameters. In SVM-CK, there are two kernel parameters: spatial RBF kernel parameter σ_s and spectral RBF kernel parameter σ_{ω}. The OAs of SVM-CK and SVM-CK-IR versus spatial and spectral kernel parameters are shown in Fig. 2. From the figure, it can be clearly seen that SVM-CK-IR is less sensitive to kernel parameters than the original SVM-CK. The original SVM-CK achieves the best OA at a narrow band while the proposed SVM-CK-IR shows good performance over a wide range of spatial and spectral kernel parameters. The rationality and veracity of kernel function (ideal regularized kernel) can reduce the dependence on the kernel parameters.

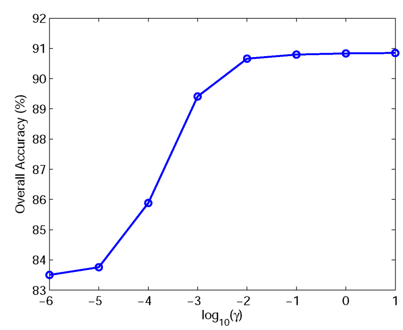
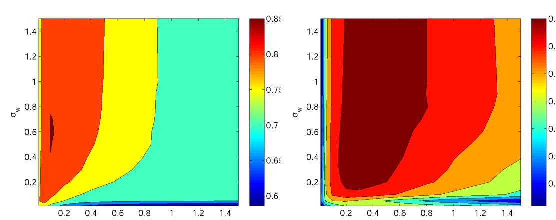


Fig. 1. OA versus regularization parameter γ for SVM-CK-IR



(a) (b)

Fig. 2. OA versus spatial and spectral kernel parameters σ_{ω} and σ_s for SVM-CK (a) and SVM-CK-IR (b).

Table 1. Classification accuracies (%) under different number of labeled samples.

M	K^{ω}		K^{s}		$K^{\omega s}$		K^{m}		$K^{\omega m}$	
	Ori	IR	Ori	IR	Ori	IR	Ori	IR	Ori	IR
5	48.16±2.66	48.13±2.61	61.77±2.92	68.79±2.83	62.11±3.56	69.97±3.48	66.96±3.73	69.82±4.01	67.35±3.59	**70.26±3.87**
10	55.66±2.27	56.21±2.44	72.57±3.44	79.09±1.87	73.25±3.25	81.03±1.80	77.22±2.34	81.31±2.48	77.81±2.23	**82.07±2.51**
15	60.35±1.53	61.30±1.23	78.33±2.91	82.59±2.04	78.61±2.47	84.98±1.94	82.70±2.08	86.22±1.94	83.31±2.20	**86.94±2.12**
20	63.13±1.30	64.94±1.38	82.04±0.73	86.55±1.78	82.32±0.87	89.33±1.57	86.94±1.26	91.16±1.71	87.36±1.19	**92.05±1.77**
25	66.48±1.86	68.04±1.61	83.07±1.55	87.73±1.39	83.66±1.31	89.85±0.85	87.88±1.45	91.53±1.31	88.28±1.35	**92.23±1.21**
30	67.84±1.62	68.72±1.67	86.21±1.60	89.57±1.36	86.48±1.59	92.13±1.42	90.27±1.66	93.57±1.20	90.81±1.65	**94.45±1.41**
35	69.67±1.91	71.23±2.06	86.90±1.57	89.88±1.56	87.99±1.27	93.18±0.88	91.87±1.53	94.76±1.04	92.46±1.44	**95.65±0.94**
40	71.08±0.81	72.40±0.99	88.31±1.36	89.08±3.18	89.27±1.36	94.20±0.66	92.60±1.30	95.18±0.67	93.19±1.09	**96.07±0.67**

4. CONCLUSION

In this paper, we have proposed an ideal regularized composite kernel (IRCK) framework for the HSI classification. Different from traditional kernel learning methods, IRCK captures both the sample similarity and label similarity by incorporating the labels into standard spectral and spatial kernels. It exploits spectral information, spatial information, and label information simultaneously. The proposed IRCK has a simple analytical solution and is very easy to implement. Experimental results on Indian Pines data set have shown the superiority of the proposed algorithm.

5. REFERENCES

[1] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE signal proc. mag.*, vol. 31, no. 1, pp. 45–54, Jan. 2014.

[2] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.

[3] G. Camps-Valls, L. Gomez-Chova, J. Muñoz Maré, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.

[4] L. Gómez-Chova, G. Camps-Valls, L. Bruzzone, and G. Calpe-Maravilla, "Mean map kernel methods for semisupervised cloud classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 1, pp. 207–220, Jan. 2010.

[5] J. Peng, Y. Zhou, and C. L. Philip Chen, "Region-kernel-based support vector machines for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 4810–4824, Sep. 2015.

[6] Y. Zhou, J. Peng, and C. L. Philip Chen, "Extreme learning machine with composite kernels for hyperspectral image classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2351–2360, Jun. 2015.

[7] J. Li, P. R. Marpu, A. Plaza, J. M. Bioucas-Dias, and J. A. Benediktsson, "Generalized composite kernel framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4816–4829, Sep. 2013.

[8] J. T. Kwok and I. W. Tsang, "Learning with idealized kernels," in *International Conference on Machine Learning*, 2003, pp. 400–407.

[9] B. Pan, J. Lai, and L. Shen, "Ideal regularization for learning kernels from labels," *Neur. Netw.*, vol. 56, pp. 22–34, 2014.

[10] B. Pan, W. Chen, C. Xu, and B. Chen, "A novel framework for learning geometry-aware kernels," *IEEE Trans. Neur. Netw. Learn. Syst.*, vol. 27, no. 5, pp. 939–951, May 2016.

[11] N. Cristianini, J. Shawe-Taylor, A. Elisseeff, and J. Kandola, "On kernel target alignment," in *Advances in Neural Information Processing Systems*, 2002, vol. 14, pp. 367–373.