SPATIAL INFORMATION BASED SUPPORT VECTOR MACHINE FOR HYPERSPECTRAL IMAGE CLASSIFICATION

Bor-Chen Kuo [1] kbc@mail.ntcu.edu.tw
Chih-Sheng Huang [1] chih.sheng.huang821@gmail.com
Chih-Cheng Hung [2] Chung@spsu.edu
Yu-Lung Liu [3] Liu720402@hotmail.com
I-Ling Chen [1] esther.x10@gmail.com

[1] Graduate Institute of Educational Measurement and Statistics, National Taichung University, Taichung, Taiwan, R.O.C.
[2] School of Computing and Software Engineering, Southern Polytechnic State University, GA, U.S.A.
[3] Department of Computer Science and Information Engineering, Asia University, Taichung, Taiwan, R.O.C.

1. INTRODUCTION

In this study, we propose a novel spatial information based support vector machine for hyperspectral image classification, named spatial-contextual semi-supervised support vector machine (SC^{3}SVM). This approach modifies the SVM algorithm [1] for hyperspectral image classification by using the spectral and spatial-contextual information. The concept of SC^{3}SVM is to utilize the information from the pixels of a neighborhood system in the spatial domain and a novel spatial-contextual term is applied to the constraint of the SC^{3}SVM. We expect this novel SC^{3}SVM to strengthen the capability for classifying the pixels, which come from different land-cover classes but have very similar spectral properties [2].

2. METHODOLOGY

Let $\boldsymbol{X}$ be a hyperspectral d-dimensional image and a set of training dataset $D = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \cdots, (\boldsymbol{x}_n, y_n)\}$ with $\boldsymbol{x}_i \in R^d$ is available, where $\{\boldsymbol{x}_i \,|\, \boldsymbol{x}_i \in R^d\}_{i=1}^n \subset \boldsymbol{X}$ is a subset of $\boldsymbol{X}$ and $\{y_i\}_{i=1}^n$ is the corresponding set of labels. Support vector machine (SVM) is to find a separating hyperplane in the feature (Hilbert) space for a binary classification problem, therefore we assume $y_i \in \{+1, -1\}$ of the pattern $\boldsymbol{x_i}$. Let $\partial \boldsymbol{x_i}$ represent a local neighborhood system of the generic pixel $\boldsymbol{x}_i$ and $\partial \boldsymbol{x_i}$ can be a first-order or second-order neighborhood system. The proposed SC^{3}SVM is defined according to a learning process that is made up of three phases: i) learning supervised SVM to classify the image, ii) learning SC^{3}SVM with both spectral and spatial-contextual information, iii) multiclass strategy of SC^{3}SVM.

A. *Phase 1*: learning Supervised SVM to classify the image.

The semi-labeled image is obtained by training a standard supervised SVM with the training set $\boldsymbol{D}$. The soft-margin SVM algorithm is performed by the following corresponding dual Lagrange function to maximize is defined as:

$$\max_{\alpha} \quad \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j \phi(\boldsymbol{x}_i)^T \phi(\boldsymbol{x}_j)$$

$$\text{subject to} \quad \begin{matrix} \sum_{i=1}^{n}\alpha_i y_i = 0 \\ 0 \le \alpha_i \le C \end{matrix}, \quad \forall i = 1, \cdots n\ .$$

where artificial variable α_i's are Lagrange multipliers corresponding to the training patterns $\boldsymbol{x}_i$. According to the Mercer's theorem, we can substitute $\phi(+)^T \phi(*)$ with a kernel function $\kappa(+,*)$.

B. *Phase 2* **:** Iterative SC^{3}SVM learning

We take into account the semi-labeled image of the neighborhood contextual patterns corresponding to their original pattern. We define the constrained minimization problem associated with the learning of SC^{3}SVM as the following:

$$\min_{\boldsymbol{w},\xi} \quad \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{n}\xi_i$$

$$\text{subject to} \quad \begin{matrix} y_i\{\boldsymbol{w}^T\phi(\boldsymbol{x}_i) + b + \gamma(m^{y_i=+1}(\partial\boldsymbol{x}_i) - m^{y_i=-1}(\partial\boldsymbol{x}_i))\} \ge 1-\xi_i \\ \xi_i \ge 0 \end{matrix}, \quad \forall i = 1, \cdots, n$$

where γ is the parameter term that controls the influence of spectral information and spatial information. $m^{y_i=+1}(\partial\boldsymbol{x}_i)$ and $m^{y_i=-1}(\partial\boldsymbol{x}_i)$ is the spatial-contextual information. $m^{y_i=+1}(\partial\boldsymbol{x}_i)$ represents the number of the neighbor pixels of $\boldsymbol{x}_i$ belongs to class $+1$ and $m^{y_i=-1}(\partial\boldsymbol{x}_i)$ represents the number of the neighbor pixels of $\boldsymbol{x}_i$ belongs to class -1. According to Lagrange theorem and Karush–Kuhn–Tucker conditions, the original constrained minimization problem can derive as a dual maximization problem:

$$\max_{\alpha} \quad J(\alpha) = \sum_{i=1}^{n}(1 - y_i\gamma(m^{y_i=+1}(\partial\boldsymbol{x}_i) - m^{y_i=-1}(\partial x_i)))\alpha_i - \frac{1}{2}\sum_{i-1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)$$

$$\text{subject to} \quad \sum_{i=1}^{n}\alpha_i y_i = 0 \quad 0 \le \alpha_i \le C \quad , \forall i = 1, \cdots, n$$

Once α_i ($i = 1, \cdots, n$) are determined, any generic pattern belonging to the investigated image can be classified according the following decision rule:

$$\begin{aligned} y(\boldsymbol{x}^{new}) &= \mathrm{sgn}(f(\boldsymbol{x}^{new})) \\ &= \mathrm{sgn}(\overline{\boldsymbol{w}}^T\phi(\boldsymbol{x}^{new}) + \overline{b} + \gamma(m^{y_i=+1}(\partial\boldsymbol{x}^{new}) - m^{y_i=-1}(\partial\boldsymbol{x}^{new}))) \\ &= \mathrm{sgn}(\sum_{i=1}^{n}\alpha_i y_i \kappa(\boldsymbol{x}_i, \boldsymbol{x}^{new}) + \overline{b} + \gamma(m^{y_i=+1}(\partial\boldsymbol{x}^{new}) - m^{y_i=-1}(\partial\boldsymbol{x}^{new}))) \end{aligned}$$

If $f(\boldsymbol{x}^{new}) > 0$ then $\boldsymbol{x}^{new}$ should be assigned to class $+1$. Conversely, the other will be assigned to class -1.

C. Multiclass strategy of SC^3SVM

This study develops two multiclass strategies, one-against-all strategy (OAA) [4] and one-against-one strategy (OAO) [5] , for SC^3SVM. In the OAA-base SC^3SVM approach is to train the separability hyperplane of k-th class versus others. Hence, we can see k-th class as positive class $(+1)$ and the remaining classes will be see as negative class (-1). In OAO-based SC^3SVM approach is to train the separability hyperplane of the class c_1 (positive class, $+1$) versus c_2 (negative class, -1), and it will ignore the other classes' information. When the neighborhoods of the training pattern not belong to these classes (class c_1 or c_2), the spatial contextual information from these neighborhood pixels is ineffective, even make a misjudgment, in training process. Therefore, we ignore some spatial information of neighborhood pixels which not belong to this step of OAO-based SC^3SVM learning.

3. SOME EXPERIMENTAL RESULTS

In this study, the hyperspectral image, Indian Pine site dataset (IPS), is applied to evaluate the performance of SC^3SVM. There are 16 different land-cover classes available in the original ground-truth, and they are Alfalfa, Corn-notill, Corn-min, Corn, Hay-windowed, Grass/trees, Grass/pasture-mowed, Grass/pasture, Oats, Soybeans-notill, Soybeans-min, Soybeans-clean, Wheat, Woods, Bldg-Grass-Tree-Drives, Stone-steel towers. In our experiment, we have chosen randomly 10% of the samples for each class from the IPS reference data as training samples, which is the same method in [6], and we take the whole image as the testing set to evaluate the performances. For investigating the performances of the spatial-based classifier, we apply a reference algorithm (a spectral–spatial classification scheme, EM+SVM), proposed by [6], into our experiment. Finally, the spatial postregularization (PR) of classification map is performed in EM+SVM and SC^3SVM, which presented and can find some details from [6]. Some results are shown in Table 1. Note that the best performances of each validation measures are highlighted in shadow cell.

Table 1. The overall accuracies, kappa coefficients, and average accuracies in percentage of the experimental classifiers for IPS dataset.

Classifier		Overall Accuracy (%)	Kappa Coefficient (%)	Average Accuracy (%)
SVM_OAO		84.4	82.3	85.5
SVM_OAA		86.5	84.6	83.8
SVM+EM	*before PR*	91.3	90.0	81.6
	after PR	92.8	91.8	82.5
SC^3SVM _OAO	*before PR*	92.9	92.0	94.5
	after PR	94.8	94.1	96.5
SC^3SVM _OAA	*before PR*	93.3	92.3	91.2
	after PR	96.4	95.9	95.8

ACKNOWLEDGMENTS

This paper is partially supported by the National Science Council grant (NSC 98-2221-E-142-005), and the authors would like to thank Prof. D. Landgrebe for providing the Indian pine site image data.

REFERENCES

[1] V. N. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed.New York: Springer-Verlag, 2001.

[2] Q. Jackson, and D.A. Landgrebe, "Adaptive Bayesian Contextual Classification Based on Markov Random Fields," *IEEE Trans. Geosci. Remote Sens.,* vol. 40, no. 11, pp. 2454-2463, 2002.

[3] Bruzzone, L. and Persello, C., " A Novel Context-Sensitive Semisupervised SVM Classifier Robust to Mislabeled Training Samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, issue 7, pp. 2142-2154, 2009.

[4] L. Bottou, C.Cortes, J. Denker, H.Drucker, I. Guyon, L.Jackel, Y. LeCun, U.Muller, E. Sackinger, P. Simard, and& V. Vapnik. ''Comparison of classifier methods: a case study in handwriting digit recognition. '' *In Proc. Int. Conf. on Pattern Recognition*, pp. 77-87, 1994.

[5] S. Knerr, L. Personnaz, and G. Dreyfus, "Single-layer Learning Revisited: a Stepwise Procedure for Building and Training a Neural Network," *In J. Fogelman, editor, Neurocomputing: Algorithms, Architectures and Applications. Springer-Verlag*, 1990.

[6] Yuliya Tarabalka, Jón Atli Benediktsson, and Jocelyn Chanussot, ''Spectral–Spatial Classification of Hyperspectral Imagery Based on Partitional Clustering Techniques.'' *IEEE Trans. Geosci. Remote Sens.*, Vol.47, No.8, pp. 2973-2987, Aug. 2009.