SEMI-SUPERVISED CONDITIONAL RANDOM FIELD FOR HYPERSPECTRAL REMOTE SENSING IMAGE CLASSIFICATION

Junfeng Wu, Zhiguo Jiang, Haopeng Zhang, Bowen Cai, and Quanmao Wei

Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China
Beijing Key Laboratory of Digital Media, Beijing 100191, China

ABSTRACT

Conditional Random Field(CRF) has been successfully applied to the hyperspectral image classification. However, it suffers from the availability of large amount of labeled pixels, which is labor- and time-consuming to obtain in practice. In this paper, a semi-supervised CRF(ssCRF) is proposed for hyperspectral image classification with limited labeled pixels. Laplacian Support Vector Machine(LapSVM), after extended into the composite kernel type, is defined as the association potential. And the Potts model is utilized as the interaction potential. The ssCRF is evaluated on the two benchmarks and the results show the effectiveness of ssCRF.

Index Terms— CRF, semi-supervised, hyperspectral, remote sensing, classification

1. INTRODUCTION

Over the past decades, hyperspectral image classification has become an important application in the field of remote sensing. Traditionally, supervised methods, such as SVM[1] and CRF[2], dominate this area. Especially, CRF has drew great attention because of its super power in modeling the spatial relationship among adjacent samples. However, it suffers from the availability of large amount of labeled samples, which is labor- and time-consuming to obtain in practice.

It is noteworthy that CRF is a flexible discriminative probabilistic framework, expressed as the sum of two inherent potentials, i.e. association potential and interaction potential. Both of two potentials have diverse forms[2], meeting demands of different tasks. Conceived from this point of view, a semi-supervied CRF(ssCRF) is proposed to solve the problem of hyperspectral image classfication. In ssCRF, the association potential is defined by Laplacian Support Vector Machine (LapSVM)[3], to exploit the unlabeled information. However, instead of using LapSVM directly, it is extended into composite kernel type[4] in our work to consider the effective combination of spectral and texture features. The Potts model[2] is utilized as the interaction potential. The proposed model can achieve three goals when seeking label assignment: 1)The model can achieve higher performance when dealing with inadequate labeled pixels. 2) The model can adaptively solve the feature combination by composite kernels. 3) The model can take into account the spatial coincidence information as a clue for classification.

The ssCRF is evaluated on two well-known benchmarks, including Pavia University scene and Indian Pines scene[5]. The results show that, our model brings significant improvement in labeling accuracy over some state-of-the-art methods, and these gains in accuracy have a significant visual impact on the resulting labeling.

This work was supported in part by the National Natural Science Foundation of China (Grant No. 61501009, 61371134 and 61071137), the Aerospace Science and Technology Innovation Fund of CASC, and the Fundamental Research Funds for the Central Universities.

2. THE FORMULATION OF SSCRF

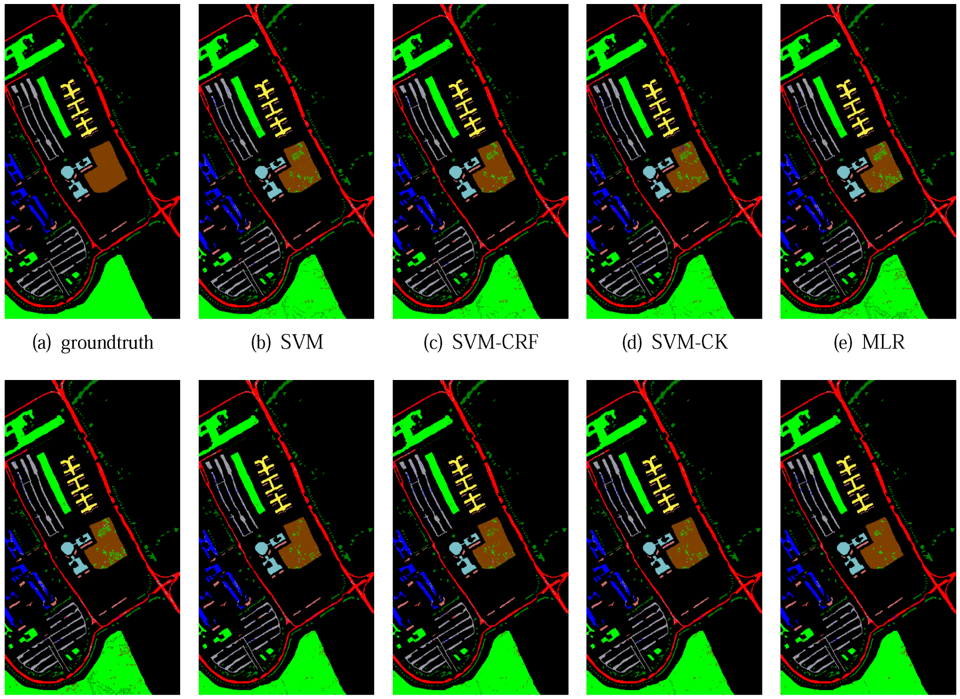
The ssCRF is modeled as the sum of association potential $\varphi(\cdot)$ and interaction potential $\psi(\cdot)$, as expressed in Eq. (1), where λ is a trade-off coefficient. $X_l = \{x_i\}_{i=1}^{n_l}$ is the set of labeled pixel site, corresponding to labels $Y_l = \{y_i | y_i \in \{1,\cdots,L\}\}_{i=1}^{n_l}$, and $X_u = \{x_i\}_{i=1}^{n_u}$ is the set of unlabeled pixel site. L is the number of category. $\mathcal{N}_i$ is the spatial neighbors of site x_i and Z is a normalizing constant known as the partition function. Γ and μ are to-be-leaned parameters.

$$\begin{aligned} P(Y_l|X_l, X_u; \Gamma, \mu) = \frac{1}{Z} \exp\big(\sum_{x_i \in X_l} \varphi(y_i, x_i, X_u; \Gamma) \\ + \lambda \sum_{x_i \in X_l} \sum_{x_j \in \mathcal{N}_i} \psi(y_i, y_j, x_i, x_j; \mu)\big) \end{aligned} \quad (1)$$

In ssCRF, the association potential builds the relationship between one pixel site and its corresponding label in a semi-supervised manner. Interaction potential models the spatial coincidence between a certain pixel and its spatial neighborhoods.

2.1. Association Potential

In the previous CRF model, Multinomial Logistic Regression (MLR) and SVM[2], are often utilized as the association potential. However, both of them are the supervised, losing sight of the ubiquitous and more readily available unlabeled pixels.



(a) groundtruth (b) SVM (c) SVM-CRF (d) SVM-CK (e) MLR

(f) MLR-CRF (g) LapSVM (h) ssLapSVM (i) LapSVM-CK (j) ssCRF

Fig. 1. The classification results of different methods on Pavia University scene

For solving the above problem, LapSVM is designed to be the association potential. Nevertheless, instead of using LapSVM directly, it is developed into the composite kernel type to consider the effective combination of spectral feature $f^s(x_i)$ and texture feature $f^t(x_i)$, rather than concatenating different feature simply to be a long feature vector. The composite kernel LapSVM (LapSVM-CK) is expressed as following,

$$\begin{aligned}\mathcal{F}^* \approx \arg\min \frac{1}{n_l}\sum_{i=1}^{n_l} V(y_i, \mathcal{F}(x_i)) + \gamma_A \|\mathcal{F}(x_i)\|_{K_c}^2 \\ + \gamma_I \sum_{i,j=1}^{n_l+n_u} \big((1-\tau)\omega_{ij}^s + \tau\omega_{ij}^t\big)(\mathcal{F}(x_i) - \mathcal{F}(x_j))^2\end{aligned} \tag{2}$$

where, K_c denotes the composite kernel, and $\mathcal{F}(x) = \sum_{i=1}^{n_l+n_u} \alpha_i\big((1-\beta)K^s(f^s(x_i), f^s(x)) + \beta K^t(f^t(x_i), f^t(x))\big) + b$. ω_{ij}^s and ω_{ij}^t are the weights matrix, defined as the heat kernel in the spectral feature space and texture feature space, respectively. The composite kernel allows us to introduce a priori knowledge in the classifier by designing specific β and τ. And then, the association potential is written as

$$\varphi(y_i, x_i, X_u; \Gamma) = \log\Big(\sum_{l=1}^{L} \mathbf{1}(y_i = l) p_*(l|x_i, X_u)\Big) \tag{3}$$

where, $\Gamma = \{\boldsymbol{\alpha}, b\}$ is the parameters in association potential, $\mathbf{1}(\cdot)$ is the indicator function and p_* is the multiclass probability obtained by LapSVM-CK with one-versus-all scheme[6].

2.2. Interaction Potential

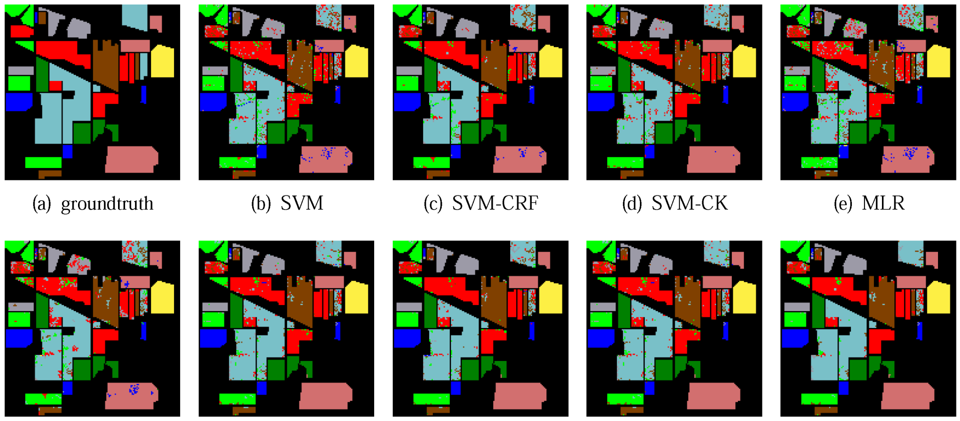
By modeling the spatial coincidence, interaction potential smooths out possible wrongly classified pixels to promote a locally coincident classification result. Mathematically, the spatial term in our work is formulated as

$$\begin{aligned}&\psi(y_i, y_j, x_i, x_j; \mu) = t_{ij}\exp(-\mu\|f(x_i) - f(x_j)\|^2) \\ &x_j \in \mathcal{N}_i \quad \& \quad x_j \in X_l \\ &t_{ij} = \begin{cases} +1 & y_i = y_j \\ -1 & y_i \neq y_j \end{cases}\end{aligned} \tag{4}$$

In Eq.(4), $f(x) = [f^s(x), f^t(x)]$ and the spatial neighbors $\mathcal{N}_i$ of a pixel x_i in X_l may be one of the following typical instances, such as Von Neumann neighborhood(V), Moore neighborhood(M) or extended Moore neighborhood(E)[7].

2.3. Model Implementation

After the modeling procedure, the complete implementation of the ssCRF includes the following steps: 1) model learning, and 2) model inference. In model learning phase, the optimal parameters Γ^* in association potential are first learned



(a) groundtruth (b) SVM (c) SVM-CRF (d) SVM-CK (e) MLR

(f) MLR-CRF (g) LapSVM (h) ssLapSVM (i) LapSVM-CK (j) ssCRF

Fig. 2. The classification results of different methods on Indian Pines scene

Table 1. The accuracy (%) on the two hyperspectral dataset

	Pavia University			Indian Pines		
	AA	OA	kappa	AA	OA	kappa
SVM	92.53	90.64	87.72	86.11	80.45	77.44
SVM-CRF	93.74	91.95	89.43	90.53	85.93	83.70
SVM-CK	92.34	91.26	88.46	89.04	84.41	81.95
MLR	90.33	87.09	83.20	80.51	75.23	71.43
MLR-CRF	91.43	88.12	84.50	85.03	80.62	77.56
LapSVM	92.34	91.82	89.20	87.35	83.29	80.61
ssLapSVM	93.57	93.47	91.37	**94.11**	90.73	89.23
LapSVM-CK	93.50	93.43	91.32	91.64	88.24	86.33
ssCRF	**94.63**	**94.70**	**92.99**	93.99	**91.56**	**90.15**

alone[3]. Then the parameter μ in interaction potential is learned[8] by maximize the Eq. (5). In the model inference step, mean-field[8] is adopted to find the optimal label configuration Y_t over the set X_t.

$$\begin{aligned}
&\mu^* \approx \underset{\mu}{\operatorname{argmax}} \sum_{x_i \in X_l} \log p(y_i|x_i, \mathbf{\Gamma}^*; \mu) \\
&=\underset{\mu}{\operatorname{argmax}} \sum_{x_i \in X_l} \log \left(\frac{1}{z_i} \exp\big(\varphi(\cdot\,; \Gamma^*) + \lambda \sum_{x_j \in \mathcal{N}_i} \psi(\cdot\,; \mu)\big) \right) \\
&z_i = \sum_{y_i \in \{1,\cdots,L\}} \exp\left(\varphi(\cdot\,; \Gamma^*) + \lambda \sum_{x_j \in \mathcal{N}_i} \psi(\cdot\,; \mu) \right)
\end{aligned} \tag{5}$$

3. EXPERIMENTS

The performance of ssCRF is evaluated on two well-known benchmarks, i.e. Pavia University scene and Indian Pines scene. Comparisons are made among ssCRF, its interaction potential (LapSVM-CK) and 7 other related methods, including SVM, SVM-CK[4], SVM-CRF[2], MLR, MLR-CRF[2], LapSVM[3], ssLapSVM[3], in terms of average accuracy (AA), overall accuracy(OA) and kappa coefficient. The average result of 20 independent experiments are reported.

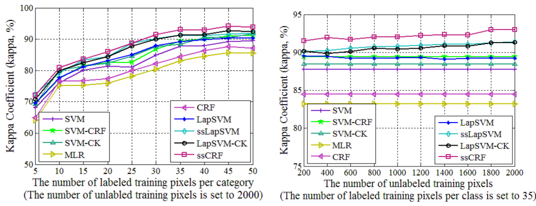
3.1. Performance analysis on Pavia University scene

In this dataset, 50% pixels of each category are randomly selected for testing. From the remaining pixels, 35 pixels of each category are randomly chosen to form labeled training set and 2000 pixels randomly selected as the unlabeled training set, leaving others to be validation set. The spectral response, scaled to the range [0, 1], is used as the spectral feature and the mean of spectral response of neighboring pixels in a window (3×3) is used as the texture feature[4]. In addition, Moore neighborhood is used to define the spatial neighborhood in association potential. λ_A and λ_I are tuned in $[10^{-6}, 10^6]$ via five-fold cross-validation. RBF kernel is used and its parameter is tuned by five-fold cross-validation in the range of $[10^{-3}, 10^3]$. The trade-off coefficient λ is set to 5. And $\beta = 0.95$, $\tau = 0.9$.

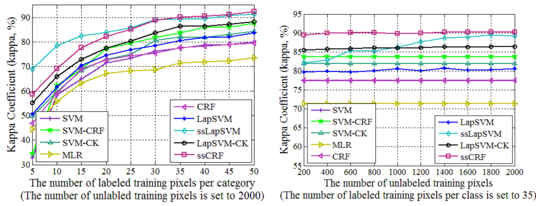
The classification accuracy of different algorithm is summarized in **Tab.**1. It can be seen that the overall performance of ssCRF exceeds all other methods, in terms of all the three criteria. The prominent performance of our method can be further confirmed by the accurate and pleasing classification results displayed in **Fig.**1.

3.2. Performance analysis on Indian Pines dataset

The experiment setup on this scene is the same as that on Pavia University scene. The trade-off coefficient λ is set to 20, and $\beta = 0.85$, $\tau = 0.9$. Evaluation criteria shown in **Tab.**1 provides the performance overview of each method. It shows that LapSVM-CK significantly improves the accuracy of its prototype, i.e. LapSVM, by considering the feature combination. ssCRF exhibits the best performance among all algorithms, which can be confirmed by visual inspection shown in **Fig.**2. ssLapSVM yields the most challenging result to ssCRF. Although ssLapSVM takes a lead in AA criteria by a tiny difference, ssCRF can improve OA and more reliable kappa



(a) Pavia University



(b) Indian Pines

Fig. 3. The performance of different methods versus different number of labeled and unlabeled pixels

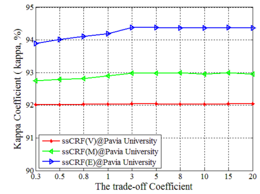
with the non-ignorable 0.83%, 0.92% difference.

3.3. Parameters analysis

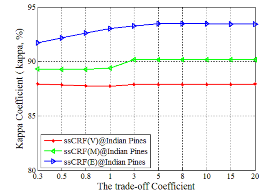
The performance of different methods versus the number of labeled and unlabeled pixels are shown in **Fig.**3. The kappa curve of ssCRF gradually rises as sufficient labeled training pixels are in place. On Pavia University scene, it can clearly identified that ssCRF gets higher accuracy than its competitors at all instances. And on Indian Pines scene, when the number of labeled training pixels reaches 35 for each category, the performance of ssCRF overwhelms all competing methods. The fluctuation of kappa coefficient against the variation of trade-off coefficient λ and spatial neighborhood is shown in **Fig.**4. It can be observed that the proposed ssCRF is not sensitive to the variation of λ. But the definition of spatial neighborhood has a great influence to ssCRF. And ssCRF with the Extended Moore Neighborhood, i.e. ssCRF(E), performs better than ssCRF with the other two spatial neighborhood, i.e. ssCRF(V) and ssCRF(M). In general, ssCRF showes promising performance when extended Moore neighborhood is adopted and a relatively large value (> 5) is set to λ.

4. CONCLUSION

In this work, a semi-supervised CRF model is developed to address the problem of hyperspectral image classification. Our model can effectively improve the classification accuracy when only limited labeled-training set is available. Experiments conducted on two well-known hyperspectral datasets demonstrates the outstanding performance of our method.



(a) Pavia University



(b) Indian Pines

Fig. 4. The performance of ssCRF with different trade-off coefficient λ and spatial neighborhood

5. REFERENCES

[1] M. Fauvel, J. A. Fauvel, J. Chanussot, J. R. Sveinsson, 'Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles,' *IEEE Tran. Geosci. Remote Sens.*, vol. **46**(11), pp.3804-3814, Nov. 2008.

[2] Y.F. Zhong, X.M. Lin and L.P. Zhang, 'A support vector conditional random fields classifier with a mahalanobis distance boundary constraint for high spatial resolution remote sensing imagery,' *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. **7**(4), pp.1314-1330, Apr. 2014.

[3] L.X. Yang, S.Y. Yang, P.L. Jin and R. Zhang, 'Semi-supervised hyperspectral image classification using spatio-spectral laplacian support vector machine,' *IEEE Geosci. Remote Sens. Lett.*, vol. **11**(3), pp.651-655, Mar. 2014.

[4] Gustavo C-V, Luis G-C, Jordi M-M, Joan V-F and Javier C-M, 'Composite kernels for hyperspectral image classification,' *IEEE Geosci. Remote Sens. Lett.*, vol. **3**(1), pp.93-97, Jan. 2006.

[5] R.R. Ji, Y. Gao, R.C. Hong and Q. Liu, 'Spectral-Spatial Constraint Hyperspectral Image Classification,' *IEEE Trans. Geosci. Remote Sens.*, vol. **52**(3), pp.1811-1824, Mar. 2014.

[6] Y.F. Gu, S.Z. Wang and X.P. Jia, 'Spectral unmixing in multiple-kernel hilbert space for hyperspectral imagery,' *IEEE Geosci. Remote Sens. Lett.*, vol. **51**(7), pp.3968-3981, Jul. 2013.

[7] M. Espłnola, J.A. Piedra-Fernndez and R. Ayala, 'Contextual and hierarchical classification of satellite images based on cellular automata,' *IEEE Trans. Geosci. Remote Sens.*, vol. **53**(2), pp.795-809, Feb. 2015.

[8] A. Kae, K. Sohn, H. Lee and E. Learned-Miller, 'Augmenting CRFs with Boltzmann Machine Shape Priors for Image Labeling,' *IEEE Conf. Comput. Vis. Pattern Recog.(CVPR)*, pp.2019-2026, Nov. 2013.