Composite kernels conditional random fields for remote-sensing image classification

Junfeng Wu, Zhiguo Jiang, Jianwei Luo and Haopeng Zhang

The problem of classifying a remote-sensing image by specifically labelling each pixel in the image is addressed. A novel method, named composite kernels conditional random field (CKCRFs), which embeds multiple kernels into a classical CRFs model is proposed. Rather than manually selecting kernel-like CRF, CKCRFs chooses the appropriate kernel by training. Moreover, a genetic programming-based decision-level fusion framework is proposed to tackle the problem of feature selection. It can select the appropriate features suitable to each category. Evaluations show that CKCRFs outperform CRFs and KCRFs, and CKCRFs with the fusion scheme is better than that without the fusion step.

Introduction: With the development of imaging technology, remote-sensing image (RSI) understanding has attracted increasing attention, and it can be applied to many situations like resource identification and natural disaster observation. This Letter addresses the problem of classifying a RSI, by specifically giving a label (e.g. grass) for each pixel in the image.

There are many existing works on RSI classification; among them the conditional random field (CRF) is popular because of its ability to model the context of data [1, 2]. Kernel conditional random fields (KCRFs) [3] improve CRFs by embedding Mercer kernels into CRFs and can better measure the similarity of data. However, the issue with KCRFs is that a person has to manually choose the appropriate kernel and its parameters.

In this Letter, we propose a novel method, named composite KCRFs (CKCRFs), which embeds multiple kernels into the classical CRFs. In this way, an appropriate kernel is composited by a family of base kernels, and the weight coefficients involved can be chosen automatically by learning on samples.

Moreover, feature selection is a key problem in RSI classification, since for different categories, the suitable features to represent them are different. For example, a colour feature is better to represent the grass category than the length width extraction algorithm (LWEA) [4], which is better to describe a road otherwise. To automatically choose features for each category, we propose a genetic programming (GP) [5] based fusion system that is established on decision level with CKCRFs as the base classifier.

In this Letter, we evaluate the two proposed methods: CKCRFs and CKCRFs-fusion. In CKCRFs, a series of radial basis function (RBF) kernels with different parameters are used. For CKCRFs-fusion, we first train an individual CRFs for each feature, and the decision values of each CRFs are fed to the GP-based fusion framework to learn the final classifier.

CKCRFs: In this Letter, a new discriminative model named CKCRFs is presented, and Fig. 1 shows an illustration of our model.

As shown in Fig. 1, given an undirected graph model $G = (N, E)$, for each node $N_i$ and the observation $y_i$, the conditional distribution over the label $x_i$ is defined as

$$p(x_i | y_i) = \frac{1}{z_i} \exp(A(x_i, y_i)) + \sum_{j \in E} (I_j(x_i, x_j, y_i, y_j))$$  (1)

where $z_i$ is a normalising constant known as the partition function. Note that $y_i$ is the feature constructed on the super-pixel level rather than the lattice-like level. In (1), the association potential $A_i$ is an $L$-class soft max regression classifier given by (2)

$$A_i(x_i, y_i) = \sum_{l=1}^L I(x_i = l)$$

$$\times \log \frac{\exp(\sum_{k=1}^p \alpha_{yk} (\sum_{xj=1}^{x} K_k(x_j, y_j)))}{\sum_{l=1}^L \exp(\sum_{k=1}^p \alpha_{yl} (\sum_{xj=1}^x K_k(x_j, y_j)))}$$  (2)

subject to $\sum_{k=1}^p \mu_k = 1$ and $0 \leq \mu_k \leq 1$

where $K_k$ is the $k$'th kernel of $k$ composite kernels, with coefficient $\mu_k$. $\{y_{i1}, r = 1, ..., n\}$ denotes the set of vectors forming the implicit space in the association potential for the $k$'th kernel. $I(x_i = l)$ is the indicator function and $\alpha_{yk}, r = 1, ..., n$ are the model parameters in the association potential.

Fig. 2 Flowchart of fusion framework

Similarly, the formulation of interaction potential $I_p$ is written as in (3), defined by the difference between the current node and the nodes surrounding it. $q^k$ is the coefficient of the $k$'th kernel. $\{y_{ij}, r = 1, ..., n\}$ are the set of vectors forming the implicit space in interaction potentials. $\beta_r, r = 1, ..., n$ are the model parameters in this potential

$$I_p(x_i, x_j, y_i, y_j) = \sum_{r=1}^p \beta_r \sum_{k=1}^p \eta_k (K_k(x_i, y_i) - K_k(x_j, y_j))$$

$$I_p = +1, \quad \text{if } x_i = x_j$$

$$-1, \quad \text{other}$$  (3)

subject to $\sum_{k=1}^p \eta_k = 1$ and $0 \leq \eta_k \leq 1$

As for all training RSI $T$, we solve the classification problem by finding the model parameters $\tau = (\alpha, \beta, \mu, \eta)$ that can maximise the objective function displayed in (4), where $A_i$ and $I_p$ are expressed as (2) and (3)

$$\hat{\tau} = \arg \max \sum_{r=1}^p \log p(\xi_i | y_i, r)$$

$$p(\xi_i | y_i, r) = \frac{1}{z_i} \exp(A(x_i, y_i)) + \sum_{j \in E} (I_j(x_i, x_j, y_i, y_j))$$

$$z_i = \sum_{x_i \in \{0, 1, ..., L-1\}} \exp(A(x_i, y_i)) + \sum_{j \in E} (I_j(x_i, x_j, y_i, y_j))$$  (4)

subject to $\sum_{k=1}^p \mu_k = 1$ and $0 \leq \mu_k \leq 1$

$$\sum_{k=1}^p \eta_k = 1 \quad \text{and} \quad 0 \leq \eta_k \leq 1$$

The genetic algorithm (GA), that aims at finding the global optimum, is special for our model. It has been also theoretically and empirically proven to be a robust search technique. The best individual (model parameters) will be picked out as the final result once the optimisation criterion is met.

GP-based fusion framework: The fusion framework is shown in Fig. 2, which is a two-tier GP-based system, including bottom and top tiers.

At the bottom tier, there are $L$ GP trees with respect to each category individually, and for each GP tree, the input is the output of $n$ CKCRFs.
trained on each feature. For example, \( p_{1\alpha} \) is the probability of category \( L \) from CKCRFs trained only on the \(\alpha\)th feature. Usually, the number and the behaviour of leaves differ from each other among these GP trees. However, at the top tier, the function is fixed, which returns the index of the maximal value as the final label of the input.

The GP-based fusion framework is established on the GP system provided by [3]. The functions of both two tiers are displayed in Table 1. The fitness of the whole fusion system is measured by the classification accuracy. Mathematically, \( \text{acc} = \frac{N_{\text{correct}}}{N} \), in which \( N \) is the number of total samples and \( N_{\text{correct}} \) is the number of samples classified correctly.

| Table 1: Functions of top tier and bottom tier |
|-----------------|-----------------|-----------------|
| Bottom tier     | Top tier         |                  |
| Function        | Input            | Output          |
| add, min, std2, mean | double, double   | double          |
| sin, cos, tan   | double           | double          |
| if              | double, double, double | double |
| IndexMax4       | double, double, double | integer |

**Experimental result:** The dataset is gathered from Google Earth, including 35 for training and 10 for testing. The spatial resolution of each image is \(<0.5\ m\). All the training images are manually labelled into four categories, including road, house, tree and grass (see Fig. 3).

**Table 2: Key parameters of GA**

| Individuals | 60 | Crossover | 30 | Generation gap | 0.8 |
|-------------|----|-----------|----|----------------|-----|
| Generations | 200 | Mutation | 0.4 | Precision of variable | 30 |

In the experiments, to better respect the boundaries of local entity, the super-pixel fragments are built by perform watershed segmentation algorithm with a relatively high threshold \( th=0.8 \). Then four kinds of features, including bag of words style of Gabor with 130 bins, LWEA [4], normalised colour histogram with 120 bins and occurrence texture measure [4], are extracted for each super-pixel fragment. They are concatenated to be the association feature. The average intensity in \((0.01, 0.05, 0.1, 0.5, 1, 10, 20, 30, 40)\). \( \{p_{1\alpha}\} \) and \( \{p_{2\alpha}\} \) are the set of all association features and interaction features, respectively. The key parameters of the GA are set as in Table 2, obtained by the orthogonal experimental method [6]. In CKCRF-fusion, four base CKCRFs are trained on the four features mentioned above. Then, the fusion framework works to obtain the final labels. The parameters of the GP are set as those of the GA, because we find that the parameter settings have little influence on the final result.

Table 3 shows comparisons of different methods, including classical CRFs and KCRFs (best with RBF kernel, \( \delta=18.5\)). We can see from Table 3 that CKCRFs show more promising performance than CRFs and KCRFs significantly, with an error reduction of about 20%, and CKCRF-fusion achieves a better result than CKCRFs. Fig. 3 shows some examples of labelling results for all the methods. In Fig. 3a, we can see that CKCRF-fusion correctly labels the long white house while others don’t, and CKCRFs are little better than CRFs and KCRFs. In Fig. 3b, CKCRFs correctly categorise the grass area adjacent with the tree compared to other methods. Fig. 4 is one example of our GP-based fusion framework. Not surprisingly, for the road category, the LWEA feature is chosen. However, for the house category, two other features, including bins style on Gabor and optical coherence microscopy, are selected. This validates that for each category the representational capacity of different features is different.

**Table 3: Super-pixel-wise classification accuracy and error reduction**

| Method  | Accuracy | Error reduction (%) |
|---------|----------|---------------------|
| CRFs    | 0.8036   | –                   |
| KCRFs   | 0.8218   | 9.267               |
| CKCRFs  | 0.8584   | 27.90               |
| CKCRFs (fusion) | 0.8753 | 36.51               |

**Fig. 3** Result of RSI classification with CRF, KCRF, CKCRF and CKCRF with fusion scheme

**Fig. 4** One example of GP-based fusion framework

**Conclusion:** In this Letter, we propose a novel method CKCRF that combines composite kernels and CRFs together, and apply it to solve the RSI classification problem. Evaluation shows that CKCRFs outperform CRFs and KCRFs significantly. Besides, a novel fusion scheme based on GP is exploited to perform feature selection for different categories, and the experimental results of CKCRF-fusion validate its effectiveness.

**Acknowledgments:** This work was supported by the National Natural Science Foundation of China (grants 61071137, 61201315 and 61371134) and the 973 Program of China (grant 2010CB327900).

© The Institution of Engineering and Technology 2014 27 May 2014 doi: 10.1049/el.2014.1964

One or more of the Figures in this Letter are available in colour online.

Junfeng Wu, Zhiguo Jiang, Jianwei Luo and Haopeng Zhang (Beijing Key Laboratory of Digital Media, Image Processing Centre, School of Astronautics, Beihang University, Beijing 100191, People’s Republic of China)

E-mail: patrickwu0609@hotmail.com

**References**

1. Hoberg, T., Rottensteiner, F., and Heipke, C.: ‘Classification of multi-temporal remote sensing data using conditional random fields’. IAPR Workshop on Pattern Recognition in Remote Sensing (PRRS), Istanbul, Turkey, August 2010, pp. 1–4.
2. Zhong, P., and Wang, R.: ‘Modelling and classifying hyperspectral imagery by CRFs with sparse higher order potentials’, IEEE Trans. Geosci. Remote Sens., 2011, 49, (2), pp. 688–705.
3. Lafferty, J., Zhu, X., and Liu, Y.: ‘Kernel conditional random fields: representation and clique selection’. Proc. 21st Int. Conf. Machine Learning, Banff, Alberta, Canada, July 2004, p. 64.
4. Shackelford, A.K., and Davis, C.H.: ‘A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas’, IEEE Trans. Geosci. Remote Sens., 2003, 41, (9), pp. 1920–1932.
5. Kozar, J.R.: ‘A source of information about the field of genetic programming and the field of genetic and evolutionary computation’, 8 July 2007. Available at: www.genetic-programming.org
6. Leung, Y.W., and Wang, Y.: ‘An orthogonal genetic algorithm with quantization for global numerical optimisation’, IEEE Trans. Evol. Comput., 2001, 5, (1), pp. 41–53.