ABSTRACT

Inspired by the idea of co-training algorithm, in this paper we propose a novel semi-supervised learning algorithm, co-Gaussian Process (co-GP), under a Bayesian framework. Image data are characterized in two distinct views, i.e. two disjoint feature sets. A latent function with a GP prior is employed for each view. In learning process of co-GP, knowledge acquired in each view is transferred by probabilistic labels to the other in turns to enhance learning effect. In this manner, proper parameters are estimated in a bootstrap mode and a satisfying performance can be maintained with only small amount of labeled data. The experiments carried out on multitemporal images validate the proposed algorithm.

Index Terms—Gaussian Process, co-training, change detection, remote sensing

1. INTRODUCTION

Change detection is one of the most important applications in the remote sensing society. Usually change detection aims at discerning areas of changes on two registered remote sensing images acquired in the same geographical area at two different times. To some degree, change detection can be regarded as a particular case of classification, in which the only property to be detected is changes or differences between two multitemporal images. Compared with the existing supervised methods and unsupervised ones, at the operating level, the semi-supervised approaches are based on a more realistic assumption that beginning with a small amount of labeled examples, the learning algorithm can equip itself with enough knowledge by taking advantages of these abundant unlabeled data [1, 2]. G. Camps-Valls etc. presented a kernel-based framework for multitemporal remote sensing image classification and change detection by using of the support vector machine (SVM) [1]. In [2], change detection in multispectral remote sensing images was achieved by a binary semi-supervised support vector machine (S’SVM) classifier which was initialized with a pseudo-training set derived by using a selective Bayesian thresholding. These methods are based on the margin maximization principle and have been proven to be very successful in remote sensing domain.

Gaussian Process (GP) is another interesting kernel-based classification approach which has not yet been investigated in the context of remote sensing. In the paper, we propose a novel change detection approach by combining co-training [3] learning with kernel method. In the approach, data samples are represented by different views, i.e. different sets of features, and a latent function with a GP prior is employed for each view. In learning process, knowledge acquired in one view is transferred by labels to the other in turns to boost the learning performance, and then these functions are jointly optimizes until they come to a consensus. The inference is made by integrating over the latent function values with analytical approximation techniques. Compared with the existing methods, there are several advantages with the proposed approach: 1) the ill-posed classification problems caused by limited quantity and quality of the training samples can be alleviated to some extent with both semi-supervised co-training method and kernel-based GP method; 2) information contained in the wealth of unlabeled samples is exploited under co-training framework; 3) the proposed method permits a Bayesian treatment of the classification problem.

The rest of the paper is organized as follows. Section 2 describes the proposed co-GP framework for change detection. The data sets used in the experiments and the performance of the methods are reported in Section 3. Conclusions are drawn in Section 4.

2. A CO-GP FRAMEWORK FOR CHANGE DETECTION

Considering two coregistered images $X_1$ and $X_2$ size of $M*N$ pixels acquired over the same geographical area but at two different times, our goal is to generate a change detection map $\Omega = \{\omega_1, \omega_2\}$, which represents the changes occurred on the ground between the acquisition dates. Therefore, our aim is to classify the changes from the unchanges in the scene.
2.1. Co-training and Theoretical Principles

Co-training method, originally introduced in [3], working in a bootstrap mode, is a semi-supervised algorithm that uses the initial training set to learn a classifier in each view. In co-training framework, for each \( x_i \in X \) there exist two instances: \( x_i^k \ (k = 1, 2) \). The sets of \( x_i^1 \) and \( x_i^2 \) are called two views of the example. In [3], guarantees under which the co-training algorithm can work well were provided: 1) Given two views \( x_i^1 \) and \( x_i^2 \), there are \( p(x_i^1, x_i^2 | y_i) = p(x_i^1 | y_i)p(x_i^2 | y_i) \); 2) For any pair \( x_i = (x_i^1, x_i^2) \in X \), probability \( p(f_i(x_i^1) \neq f_i(x_i^2)) = 0 \) is satisfied. The essence of co-training is that multiple classifiers work in a bootstrap mode and teach each other in turn to augment the labeled data.

2.2. GP and GP Classification

A GP is a collection of random variables, any finite number of which has a joint Gaussian distribution [4]. GP is fully specified by its mean function \( m(x) \) and covariance function \( k(x, x') \), expressed as:

\[
f \sim GP(m, k).
\]  
(2)

Given the training data \( x \), in the GP model we assume there is a latent function \( f \) underlying the output \( y \), we wish to make prediction on the testing data \( x^* \) by calculating \( p(y^* | x, x^*, y) \) with the GP prior \( p(f) = GP(m, k) \).

Following [4], computing the distribution of the latent variable corresponds to a test case

\[
p(f^* | x, y, x^*) = \int p(f | x, x^*, f)p(f | x, y) df.
\]  
(3)

For binary classification, the output of regression is ‘squashed’ through a logistic function to guarantee the probabilistic value within the range of [0, 1]. Laplace approximation method and Expectation propagation (EP) method are two popular analytic approximations.

2.3. Co-GP Classifier

To improve the performance of the classification, we must be able to choose the optimal or sub-optimal prior values in light of the data. To do this, we introduce hyperparameters \( \theta \) and make inference about them via a bootstrap manner. By transferring the knowledge learned by one GP classifier from a view to the other, the hyperparameter \( \theta \) is alternately optimized. This is referred to as training a co-GP model. In our approach, the seed data \( \{(x_i, y_i)\}_{i=1}^{T} \) are characterized into two different views. Considering the two properties of the co-training algorithm, we choose color feature and Gabor feature as the two views. This process can be described as following: First, a set of features for the training seeds are selected, and an initial \( \theta_0 \) is updated as \( \theta_1 \) by a conjugate gradient optimization routine. Using \( \theta_1 \) we can obtain a new label set \( y_1 \) for the training seed set \( x \). Then the label set \( y_1 \) is fed to update the hyperparameter with the other feature set. In this way, new hyperparameter \( \theta_2 \) and label set \( y_2 \) are obtained. This process is repeated several times and the GP priors \( p(f_j) \), \( j = 1, 2 \) for each view are estimated. It is noticeable that in this process only the seed data \( \{(x_i, y_i)\}_{i=1}^{T} \) are used and no other unlabeled data is added to this process. The main advantages of this scheme are that: 1) By alternately optimizing the hyperparameter \( \theta \), the strength of different views of the seed data are full employed and more robust prior information is obtained, which in turn improve the performance of the GP classification; 2) A small amount of training data are needed for co-GP but maintains a satisfying performance.

Based on the GP prior \( p(f_j) \), co-GP prediction is accomplished by learning a GP classifier \( f_j \) in each view and retrain each other in turn to argument the labeled data. With the prior \( p(f_j) \) obtained in co-GP training, a GP classifier \( f_j \) is trained only using the attribute set \( x_i^j \), i.e. a color feature set. A label set \( y_j^* \) which are the most confident predictive labels made by the GP classifier \( f_j \) are added to the labeled data set \( y \). GP classifier \( f_2 \) is trained on this new training \( y \) only using the Gabor feature set \( x_i^2 \), then predictive labels with the highest confidence are put into the pool of labeled set \( y \). The whole process is repeated for several iterations until the two classifiers reach a consensus. Based on the predicted labels from the two views, a final change map is created by a voting scheme.

In traditional co-training algorithm, labeled data with little information are eliminated and the elimination is random and equal for every sample in the set \( y \). However, it is no longer a smart way in our approach. The reasons are those: as a semi-supervised method, the seeds, selected by the human beings, usually provide quite abundant and valuable information to finish the detection task. These seeds to some degree represent the most confident labels. Therefore, it is wise to put uneven weights on the seeds and the other labeled samples. In the paper, we keep all the seed labels from being discarded. By this way, the algorithm is remained more robust and the instability of original co-training algorithm is revised to a certain extent.
Fig. 1. Original images (a-b); Change maps obtained by the proposed approach (c) and main error distribution of our proposed approach (d); Single-view GP method with (e) both Gabor and Color feature (S_GP_G&C); (f) Gabor feature (S_GP_G); and (g) Color feature (S_GP_C); (h) Reference map.

Fig. 2. (a) Change map obtained by SVM; (b) Difference of the result obtained by SVM and Co-GP.

3. EXPERIMENTS

Two experiments were designed to validate the proposed method. In the first experiment, we compared the proposed method with conventional single-view GP methods [5]. In the second one, we chose traditional SVM [6] as a comparison to the proposed method.

In all experiments, we choose color feature and Gabor feature as the two views (i.e., two features,) for one image. View 1 is simply RGB values of the original multifractal images and their corresponding differences, i.e., a 9-dimensional vector \( \{r_1, g_1, b_1, r_2, g_2, b_2, r_1-r_2, g_1-g_2, b_1-b_2\} \);

View 2, was the Gabor features [7] extracted from the original multifractal images and the differences of those Gabor features, i.e., a 15-dimensional vector \( \{g_{a1}, g_{a2}, g_{a3}, g_{a1} - g_{a2}, g_{a2} - g_{a3}\} \). The total number of seeds \( d \) was set to less than 2500 pixels in all our tests. This was a compromise between time complexity and performance. The seeds used for training in our experiments are marked with blue frame in the original images, showing in Fig. 1(a-b).

All of the results are evaluated in terms of: 1) overall error rate (PE); 2) false alarm rate (PF); 3) missed alarm rate (PM); 4) Kappa coefficient. A reference map was manually generated according to a detailed visual analysis.

We carried out our experiments on several different sets of multifractal remote sensing images. For space limited, here, we just show some results on two typical data sets. The first set contains a pair of Ikonos 2m resolution optical images which were acquired before and after the tsunami on January 10, 2003 and December 29, 2004 respectively, over Aceh, Sumatra, Indonesia. After co-registration, each image had the size of 3000×2880 pixels. In this paper, experimental results on four typical partial images with the size of 256×256 pixels are shown. Fig. 1(a) and (b) show a sample image pair. This area in the images was mainly covered by green vegetation, and the remarkable changes to be detected were that original covered vegetation submerged and seacoast eroded by flood.

For the first experiment, all the change maps are presented in Fig. 1 for visual comparison. According to these change maps, we can find that the proposed approach gave a better performance than all the other ones produced by single-view based ones (showing in Fig. 1(e)-(g)). Many detailed but significant changes were not detected by the single-view approaches, while some unchanged parts were falsely alarmed. Contrarily, the proposed method was very effective to locate all the changes, and even most of the tiny changes were successfully detected. To give a more intuitive comparison, we offered the difference maps between the proposed approach and the reference map in Fig. 1(c). These comparison, we offered the difference maps between the
proposed approach and the reference map in Fig. 1(c). These results reveal that the main errors are located on the edges between the changed and unchanged area. The reason may be that the location information was not fully exploited in experiment. The reference map is shown in Fig. 1(h). Table I gives a quantitative evaluation of all the change results. Based on these figures, one can find that our approach has a much lower error rate (about 63% less than the lowest error of single-view method with Gabor feature). The co-GP classifier has the highest kappa coefficient of agreement. By analyzing these results, one can deduce that the proposed co-GP framework carries the detection precision to a new and high level.

The results obtained in the second experiment confirm the validity of the GP classification. The change maps in Fig. 2 show that GP classifier can compete seriously with the state-of-the-art SVM classifier. With the same quantity of training data and feature sets, both the two kernel-based methods produced relative good change results, while our proposed approach provided a better accuracy. The difference map of the two kernel-based methods is illustrated in Fig. 2 (b). A quantitative comparison of the two methods is reported in Table I. Concerning the error typology, the SVM method had a higher false alarm rate. The reason may be that our co-GP strategy improved the GP detection precision in a bootstrap mode, which in turn proved that the co-GP framework used in our approach was quite effective. From these analyses, one can deduce that the GP classification is a potential but promising method for change detection.

Other experimental results on Indonesia data set were presented in Fig. 3 without detailed analyses. The Kappa coefficient values of these three samples are 0.8276, 0.9160, and 0.8891 corresponding to change results in Fig. 3(a)-(c).

Fig. 4 illustrates the change map obtained by the proposed approach on Lebanon dataset (a pair of Quickbird 0.4m resolution color images with the size of 285×285 pixels, taken over downtown area in Beirut, on November, 2003 and February, 2005 respectively). Kappa coefficient value is 0.8805.

4. CONCLUSION

In the paper, we proposed an effective method for remote sensing change detection which is treated as a binary classification problem. Satisfactory results are achieved on a small training set under a co-GP framework. The capabilities of GP classifiers for remote sensing imagery change detection have been thoroughly investigated. The experimental results confirmed the effectiveness of the proposed algorithm and also have positively indicated GP classifier can compete seriously with the state-of-the-art SVM classifier. Future works will involve 1) including the location information into the feature sets to boost the change detection precision along the edges; and 2) matching the GP kernel adaptively with different data.

REFERENCES