Effective Sequential Classifier Training for Multitemporal Remote Sensing Image Classification

Yiqing Guo, Student Member, IEEE, Xiuping Jia, Senior Member, IEEE, and David Paul, Fellow, IEEE

Abstract—The explosive availability of remote sensing images has challenged supervised classification algorithms such as Support Vector Machines (SVM), as training samples tend to be highly limited due to the expensive and laborious task of ground truthing. The temporal correlation and spectral similarity between multitemporal images have opened up an opportunity to alleviate this problem. In this study, a SVM-based Sequential Classifier Training (SCT-SVM) approach is proposed for multitemporal remote sensing image classification. The approach leverages the classifiers of previous images to reduce the required number of training samples for the classifier training of an incoming image. For each incoming image, a rough classifier is firstly predicted based on the temporal trend of a set of previous classifiers. The predicted classifier is then fine-tuned into a more accurate position with current training samples. This approach can be applied progressively to sequential image data, with only a small number of training samples being required from each image. Experiments were conducted with Sentinel-2A multitemporal data over an agricultural area in Australia. Results showed that the proposed SCT-SVM achieved better classification accuracies compared with two state-of-the-art model transfer algorithms. Compared with those obtained without the assistance from previous images, the overall classification accuracy was improved from 76.2% to 93.8%, demonstrating that the leverage of a priori information from previous images can provide advantageous assistance for later images in multitemporal image classification.

Index Terms—Multitemporal, classifier training, classification, support vector machines

I. INTRODUCTION

Land cover information extracted from remote sensing images has enabled timely monitoring of land cover dynamics at a large spatial scale [1], [2]. For the generation of land cover maps, supervised classification algorithms have been widely adopted [3]. Ground reference data are required by these algorithms for the training of classifiers [3].

However, traditional supervised classification algorithms such as the Support Vector Machines (SVM) have been challenged by the explosive availability of remote sensing images. While collecting a sufficient number of training samples is critical to achieve satisfactory classification results, it is often labour- and time-consuming and sometimes practically unfeasible [4]–[8]. The problem is more severe for agricultural areas where the accuracy of ground reference data can be quickly outdated, as a result of the frequent land cover changes and spectral characteristic shifts caused by crop phenology and anthropogenic practices such as irrigation and harvesting [9]. As a result, training data sufficiency is often hard to be guaranteed for every image.

Temporal correlation and spectral similarity between multitemporal images have opened up an opportunity to alleviate the small sample size problem. When an incoming image needs to be classified, existing knowledge provided by previous images can be utilized as a priori information. However, it cannot be directly applied to the incoming image, because class data of different images may follow different probabilistic distributions in the feature space. This phenomenon is often referred to as cross-image dataset shifts [10]. The shifts can either originate from changes in the nature of land surface properties, for example, induced by the phenology of vegetation, or from the background noise caused by varied acquisition and atmospheric conditions [11]–[13] and inconsistent sun-target-sensor geometries [14], [15]. Therefore, appropriate strategies need to be applied to tackle the dataset shifts [16].

Domain adaptation is an emerging technique that is able to accommodate the dataset shifts among images [16]–[21]. The technique aims to adapt a priori information extracted from previous images to an incoming image that has shifted spectral characteristics. Compared with a complete retraining of the incoming image, the required number of training samples can be considerably reduced if the assistance from previous images is leveraged [6]. Several domain adaptation algorithms have been proposed to accommodate the dataset shifts in multitemporal image classification. For example, the change-detection-driven algorithms [7], [22] train a classifier for an incoming image with the help of the unchanged samples from a previous image. A change detection step is firstly implemented to identify unchanged pixels, whose labels are then directly transferred to the incoming image. However, as erroneous labels may be transferred as well, the classification error will then propagate from the previous image to the incoming image and accumulate when the algorithm is applied progressively to multiple incoming images. The Bayesian-based algorithms [6] adapt a Bayesian classifier to an incoming image with the statistical parameters calculated from the previous image as an a priori estimate. Then an unsupervised retraining procedure is used to adjust the a priori estimate, making it right for the incoming image. However, the a priori estimate is made with only one previous image instead of a set of previous images, so the temporal trajectory of class spectral characteristics, which is useful information to make a more accurate a priori estimation, is not extracted and utilized in the algorithm.
This study provides a SVM-based sequential classifier training (SCT-SVM) approach for multitemporal remote sensing image classification. With the assistance from a set of previous classifiers, this approach is able to effectively train classifiers for an incoming image at the same location. This approach can be applied progressively to sequential image data, with only a small number of training samples being required from each image. A set of Sentinel-2A multitemporal images over an agricultural area in Australia have been used to demonstrate the effectiveness of the proposed approach.

II. METHOD

A. Proposed Framework

When an incoming image needs to be classified, the small sample size problem often arises if training samples are difficult to collect. In order to mitigate this problem, we propose to leverage a priori knowledge gained from the classification of previous images at the same location. The assistance from previous images is provided at the classifier-level instead of the data-level. Specifically, the parameters of classifiers instead of raw data (i.e., pixel values and/or their corresponding labels) from the previous images are used.

According to the sequence of sensing date, previous images and their classifiers are indexed $t-1$, $t-2$, ···, $t-N$, where $N$ is the total number of previous images, and the incoming image and its classifier are indexed $t$.

Consider that a classifier needs to be determined for Image $t$. We propose a sequential classifier training approach that consists of two steps: prediction and fine-tuning, as shown in Fig. 1. In the prediction step, a rough classifier (Classifier $t^*$) is firstly estimated based on a set of previous classifiers (Classifiers $t-1$, $t-2$, ···, $t-N$). In the fine-tuning step, the predicted classifier (Classifier $t^*$) is adjusted and finalized as Classifier $t$ with current training samples. Because the fine-tuning is based on the predicted classifier, compared with the direct training of a classifier for Image $t$ without any ancillary information, the required number of training samples can be smaller. After Classifier $t$ is determined, it can be included into the set of previous classifiers, while the oldest classifier (Classifier $t-N$) may be removed. The updated set is then used as a priori knowledge to assist the training of the next classifier (Classifier $t+1$).

The proposed approach is named as SCT-SVM as it is based on the well-established classification algorithm of Support Vector Machines (SVM). In a $m$-dimensional feature space, a linear SVM classifier $f(x)$ for a binary classification problem can be expressed as [23]:

$$f(x) = w^T x + b, \quad (1)$$

where $x = (x_1, x_2, \cdots, x_m)^T$ is the feature vector; $w = (w_1, w_2, \cdots, w_m)^T$ is the weight vector and $b$ is the bias factor. For convenience, we stack the classifier parameters $w$ and $b$ into a single vector $p = (w_1, w_2, \cdots, w_m, b)^T$. Because each classifier $f(x)$ can be uniquely determined by its parameters $p$, the terms classifier and classifier parameters will be used synonymously in the following. The binary SVM classifier can handle the multi-class case by applying the one-against-one or one-against-the-rest strategies [3].

Figure 2 illustrates the classifier parameter space defined in two dimensions (e.g., $p_1$ and $p_2$), where the prediction/fine-tuning steps are detailed. Different classifiers often appear at different points in this parameter space, due to the temporal variations among multitemporal images. The temporal trend of previous classifiers ($p^{(t-1)}$, $p^{(t-2)}$, ···, $p^{(t-N)}$) is firstly utilized to predict a rough classifier ($\hat{p}^{(t)}$) for the current image (Image $t$), and it is then fine-tuned into an updated position ($\hat{p}^{(t)}$). The prediction and fine-tuning steps will be detailed below in the following Subsections II-B and II-C respectively.

B. Classifier Prediction

The aim of classifier prediction is to determine a rough classifier for the current image (Image $t$). The prediction is based on the temporal trend of a set of previous classifiers. The trend can be due to the background variations under different acquisition conditions or originated from changes in the nature of land surface properties such as the phenology of crops and forests, or both. The former factor causes random and unpredictable changes. The latter factor induces relatively smooth and predictable temporal transition, due to the continuity of the changes with time. We propose to perform a principal component transform, and use the first transformed component to extract the overall temporal trend.

The trend can be due to the background variations under different acquisition conditions or originated from changes in the nature of land surface properties such as the phenology of crops and forests, or both. The former factor causes random and unpredictable changes. The latter factor induces relatively smooth and predictable temporal transition, due to the continuity of the changes with time. We propose to perform a principal component transform, and use the first transformed component to extract the overall temporal trend. We propose to leverage a priori knowledge gained from the classification of previous images at the same location. The assistance from previous images is provided at the classifier-level instead of the data-level. Specifically, the parameters of classifiers instead of raw data (i.e., pixel values and/or their corresponding labels) from the previous images are used.

According to the sequence of sensing date, previous images and their classifiers are indexed $t-1$, $t-2$, ···, $t-N$, where $N$ is the total number of previous images, and the incoming image and its classifier are indexed $t$.

Consider that a classifier needs to be determined for Image $t$. We propose a sequential classifier training approach that consists of two steps: prediction and fine-tuning, as shown in Fig. 1. In the prediction step, a rough classifier (Classifier $t^*$) is firstly estimated based on a set of previous classifiers (Classifiers $t-1$, $t-2$, ···, $t-N$). In the fine-tuning step, the predicted classifier (Classifier $t^*$) is adjusted and finalized as Classifier $t$ with current training samples. Because the fine-tuning is based on the predicted classifier, compared with the direct training of a classifier for Image $t$ without any ancillary information, the required number of training samples can be smaller. After Classifier $t$ is determined, it can be included into the set of previous classifiers, while the oldest classifier (Classifier $t-N$) may be removed. The updated set is then used as a priori knowledge to assist the training of the next classifier (Classifier $t+1$).

The proposed approach is named as SCT-SVM as it is based on the well-established classification algorithm of Support Vector Machines (SVM). In a $m$-dimensional feature space, a linear SVM classifier $f(x)$ for a binary classification problem can be expressed as [23]:

$$f(x) = w^T x + b, \quad (1)$$

where $x = (x_1, x_2, \cdots, x_m)^T$ is the feature vector; $w = (w_1, w_2, \cdots, w_m)^T$ is the weight vector and $b$ is the bias factor. For convenience, we stack the classifier parameters $w$ and $b$ into a single vector $p = (w_1, w_2, \cdots, w_m, b)^T$. Because each classifier $f(x)$ can be uniquely determined by its parameters $p$, the terms classifier and classifier parameters will be used synonymously in the following. The binary SVM classifier can handle the multi-class case by applying the one-against-one or one-against-the-rest strategies [3].

Figure 2 illustrates the classifier parameter space defined in two dimensions (e.g., $p_1$ and $p_2$), where the prediction/fine-tuning steps are detailed. Different classifiers often appear at different points in this parameter space, due to the temporal variations among multitemporal images. The temporal trend of previous classifiers ($p^{(t-1)}$, $p^{(t-2)}$, ···, $p^{(t-N)}$) is firstly utilized to predict a rough classifier ($\hat{p}^{(t)}$) for the current image (Image $t$), and it is then fine-tuned into an updated position ($\hat{p}^{(t)}$). The prediction and fine-tuning steps will be detailed below in the following Subsections II-B and II-C respectively.

The aim of classifier prediction is to determine a rough classifier for the current image (Image $t$). The prediction is based on the temporal trend of a set of previous classifiers. The trend can be due to the background variations under different acquisition conditions or originated from changes in the nature of land surface properties such as the phenology of crops and forests, or both. The former factor causes random and unpredictable changes. The latter factor induces relatively smooth and predictable temporal transition, due to the continuity of the changes with time. We propose to perform a principal component transform, and use the first transformed component to extract the overall temporal trend.

The trend can be due to the background variations under different acquisition conditions or originated from changes in the nature of land surface properties such as the phenology of crops and forests, or both. The former factor causes random and unpredictable changes. The latter factor induces relatively smooth and predictable temporal transition, due to the continuity of the changes with time. We propose to perform a principal component transform, and use the first transformed component to extract the overall temporal trend.
Fig. 1. The framework of the proposed sequential classifier training approach for multitemporal remote sensing image classification.

Fig. 2. Illustration of the two step classifier training in the classifier parameter space.

Fig. 3. Principal component analysis in the classifier parameter space.

Fig. 4. Regression between first principal components and sensing dates.

Fig. 4. The first principal components after the transformation, $q_1^{(t-1)}, q_1^{(t-2)}, \ldots, q_1^{(t-N)}$, are now associated with the image sensing dates, $d^{(t-1)}, d^{(t-2)}, \ldots, d^{(t-N)}$, to show the intervals between each image collection, which can be irregular. A regression is then performed in order to predict a classifier for Image $t$. The polynomial fitting function is selected in this study:

$$q_1^{(i)} = a_0 + a_1 d^{(i)} + \cdots + a_r (d^{(i)})^r,$$

where $r$ is the order of the function, and $a_0, a_1, \ldots, a_r$ are the fitting coefficients. The sensing dates used here are integer counts of days starting from a consistent reference date.

It is worth noting that the order of the fitting function is selected according to the pattern of the temporal trend of classifiers. In order to achieve better fitting accuracies, higher-order polynomials may be adopted, but the overfitting risk will increase as well. So the balance between fitting accuracy and overfitting problem should be considered when choosing the fitting function.

After determining the fitting coefficients $a_0$ and $a_1$ with the data of previous classifiers, the first principal component of
the predicted classifier for Image $t$, $\hat{q}_t^{(t)}$, can be calculated as:

$$\hat{q}_1^{(t)} = a_0.$$  \hspace{1cm} (6)

The rest of the principal components $\hat{q}_2^{(t)}, \hat{q}_3^{(t)}, \cdots, \hat{q}_{m+1}^{(t)}$ can be set to zero.

The inverse of the principal component transform is then performed, followed by the mean-decentralization, to obtain the predicted classifier in the original classifier parameter space:

$$p^{(t)*} = G^{-1}\bar{q}^{(t)*} + \frac{1}{N} \sum_{i=t-1}^{t-N} p^{(i)}.$$  \hspace{1cm} (7)

C. Classifier Fine-Tuning

A domain adaptation algorithm is developed as the second stage of the proposed SCT-SVM. It fine-tunes the predicted classifier $p^{(t)*}$ to a more accurate position $p^{(t)}$ with training samples of the current image (Image $t$). The algorithm is designed with anticipation that the true classifier should be located not far from the predicted position. This is a reasonable expectation, given that the predicted classifier is based on the historical information extracted from previous images and the changes often have continuity with time. The aim of the fine-tuning is set as to achieve a higher classification accuracy on the training samples of Image $t$. In this way, contributions from the predicted classifier and the training samples are incorporated and balanced in the algorithm, as detailed in the following.

The a priori information provided by the predicted classifier is utilized by restricting the fine-tuned classifier from departing too far away from the predicted position. To achieve this goal, the distance between the first $m$ elements of $p^{(t)}$ (i.e., $w^{(t)}$) and $p^{(t)*}$ (i.e., $w^{(t)*}$) is minimized. More specifically, a set of non-negative slack variables $m = \{\mu_j\}_{j=1}^m$ are introduced, where the $j$th element restricts the upper and lower bounds of the deviation of $p_j^{(t)}$ from $p_j^{(t)*}$, resulting in the following constraints:

$$-\mu_j \leq p_j^{(t)} - p_j^{(t)*} \leq \mu_j \quad \text{and} \quad \mu_j \geq 0, \quad j = 1, 2, \cdots, m.$$  \hspace{1cm} (8)

The information provided by the training samples is utilized by taking into account the classification error of the fine-tuned classifier on the training samples $\{x_i, y_i\}_{i=1}^n$, where $x_i$ is the feature vector of the $i$th training sample and $y_i \in \{+1, -1\}$ is its corresponding label. Similar to those in the standard support vector machines, a set of non-negative slack variables $\xi = \{\xi_i\}_{i=1}^n$ are used to allow soft-margin classification in order to tackle overlapping classes. Consequently, the following constraints are designed:

$$y_i \left( \sum_{j=1}^m p_j^{(t)} x_{i,j} + p_{m+1}^{(t)} \right) \geq 1 - \xi_i \quad \text{and} \quad \xi_i \geq 0, \quad i = 1, 2, \cdots, n,$$  \hspace{1cm} (9)

where $x_{i,j}$ is the $j$th entry in $x_i$.

With the constraints in Eqs. (8) and (9), the following optimization problem can be constructed:

$$\min \frac{1}{2} \sum_{j=1}^m (p_j^{(t)})^2 + C \sum_{i=1}^n \xi_i + F \sum_{j=1}^m \mu_j$$  \hspace{1cm} (10)

The first term in the objective function accounts for the margin space of the fine-tuned classifier. Like that in the standard support vector machines, the term aims to adjust the separating hyperplane to a position that generates the maximum margin space between classes. The second and third terms aim to minimize the degree of violation of the constraints. The positive constants $C$ and $F$ are regularisation parameters, which control the weights of the second and third terms relative to the first term. The values of $C$ and $F$ control the amounts of contributions from training samples and previous images, respectively.

The dual form of the optimization problem in Eq. (10) can be obtained by constructing the following Lagrangian function $L$:

$$L = \frac{1}{2} \sum_{j=1}^m (p_j^{(t)})^2 + C \sum_{i=1}^n \xi_i + F \sum_{j=1}^m \mu_j$$
$$- \sum_{i=1}^n \alpha_i \left[ y_i \left( \sum_{j=1}^m p_j^{(t)} x_{i,j} + p_{m+1}^{(t)} \right) - (1 - \xi_i) \right]$$
$$- \sum_{i=1}^n \beta_i \xi_i - \sum_{j=1}^m \gamma_j \left[ \mu_j - (p_j^{(t)} - p_j^{(t)*}) \right]$$
$$- \sum_{j=1}^m \delta_j \left[ (p_j^{(t)} - p_j^{(t)*}) + \mu_j \right] - \sum_{j=1}^m \varepsilon_j \mu_j,$$  \hspace{1cm} (11)

where $\{\alpha_i\}_{i=1}^n$, $\{\beta_i\}_{i=1}^n$, $\{\gamma_j\}_{j=1}^m$, $\{\delta_j\}_{j=1}^m$ and $\{\varepsilon_j\}_{j=1}^m$ are non-negative Lagrange multipliers. Equating the partial derivatives of $L$ to zero with respect to $p_j^{(t)}$, $\xi_i$, and $\mu_j$ results in the following equations:

$$p_j^{(t)} - \sum_{i=1}^n \alpha_i y_i x_{i,j} + \gamma_j - \delta_j = 0, \quad j = 1, 2, \cdots, m,$$  \hspace{1cm} (12)

$$\sum_{i=1}^n \alpha_i y_i = 0,$$  \hspace{1cm} (13)

$$C - \alpha_i - \beta_i = 0, \quad i = 1, 2, \cdots, n,$$  \hspace{1cm} (14)

$$F - \gamma_j - \delta_j - \varepsilon_j = 0, \quad j = 1, 2, \cdots, m.$$  \hspace{1cm} (15)
Eliminating $p_j^{(t)}$, $C$, and $F$ by substituting Eqs. (12)-(15) into Eq. (11) results in the following Lagrangian dual function $L_d$:

$$
L_d = -\frac{1}{2} \sum_{i=1}^{n} \sum_{k=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_k y_i y_k x_{i,j}^2 - \frac{1}{2} \sum_{j=1}^{m} \gamma_j^2 - \frac{1}{2} \sum_{j=1}^{m} \delta_j^2 + \sum_{i=1}^{n} \sum_{j=1}^{m} \gamma_j \alpha_i y_i x_{i,j} + \sum_{j=1}^{m} \gamma_j \delta_j + \sum_{i=1}^{n} \alpha_i - m \gamma_j - \delta_j.
$$

Then the following dual optimization problem can be constructed, which is equivalent to the primal problem in Eq. (10) but has a simpler form:

$$
\begin{align*}
\min_{\alpha_i, \gamma_j, \delta_j} & - L_d \\
s.t. \quad & 0 \leq \alpha_i \leq C, \quad i = 1, 2, \ldots, n, \\
& \gamma_j + \delta_j \leq F, \quad \gamma_j \geq 0, \quad \delta_j \geq 0, \quad j = 1, 2, \ldots, m, \\
& \sum_{i=1}^{n} \alpha_i y_i = 0.
\end{align*}
$$

This is a quadratic programming problem and can be solved with standard quadratic programming algorithms [25]. After determining $\alpha_i$, $\gamma_j$, and $\delta_j$, the parameters for the fine-tuned classifier $p_j^{(t)}$ can be calculated as:

$$
p_j^{(t)} = \sum_{i=1}^{n} \alpha_i y_i x_{i,j} - \gamma_j + \delta_j, \quad j = 1, 2, \ldots, m,
$$

$$
p_m^{(t)} = \frac{1}{N_S} \sum_{i \in S} \left[ \frac{1}{y_i} - \sum_{j=1}^{m} p_j^{(t)} x_{i,j} \right],
$$

where $S$ and $N_S$ are the index set and the total number of support vectors, respectively.

D. Discussions

1) Advantages of Classifier-Level Approach:

The proposed SCT-SVM is a classifier-level approach. It transfers the classifier-level knowledge, i.e. the classifier parameters of a set of previous classifiers, to the incoming image. Compared with the data-level approaches that require the availability of pixels and/or their corresponding labels (raw data) from previous images, such as the Domain Adaptation SVM (DA-SVM) [26] and the Domain Transfer SVM (DT-SVM) [27], the proposed approach has the following two merits. Firstly, the required computational load can be reduced considering that the size of raw data is usually large. Secondly, the proposed algorithm is applicable even if the raw data of previous images are inaccessible (e.g., private or no longer stored). Therefore, compared with data-level approaches, the proposed classifier-level approach is more efficient and more feasible.

2) Comparing Proposed Fine-tuning Algorithm to State-of-the-Art Algorithms:

The second step of the proposed method is fine tuning. There are two existing domain adaptation algorithms in the literature that are able to handle the same task: the Adaptive SVM (A-SVM) [28], [29] and the Projective Model Transfer SVM (PMT-SVM) [30]. The A-SVM enables the estimation of a fine-tuned target classifier by augmenting a perturbation term to the predicted classifier. However, it has been found that the A-SVM tends to generate a small margin space for the fine-tuned classifier [30]. The other algorithm, PMT-SVM, is able to estimate a fine-tuned classifier without penalizing its margin maximization. The limitation, however, is that it restricts the included angle between the predicted and fine-tuned classifiers to a value equivalent to or less than 90° in the feature space. This restriction leads to a decreased performance when the underlying optimal classifier is positioned at an angle greater than 90° from the predicted position.

The proposed fine tuning algorithm considers a maximum-margin-space term (i.e., the first term of the objective function in Eq. (10)), which can alleviate the small-margin-space problem that exists in the A-SVM. The proposed algorithm is also able to fine-tune a predicted classifier into a position larger than 90° as no restriction is applied on the fine-tuning angle, which is an improvement from the PMT-SVM. The proposed domain adaptation algorithm is named as TA-SVM for convenience, and it is compared with A-SVM and PMT-SVM in the experiments.

III. Experiment and Results

A. Study Area and Data Sets

The experimental area is located at the Coleambally Irrigation Area in the southwest of New South Wales, Australia (145°40′27″–146°08′12″ E, 34°40′37″–35°03′28″ S), as shown in Fig. 3. A high water-consuming crop, rice, is widely planted in this area. Its growing period lasts for about six months, roughly from October of the first year to April of the second year. The area has a semiarid climate (Köppen-Geiger BSk). The annual rainfall is about 400 mm, distributed evenly throughout the year. Due to the semi-arid climate and high water consumption of the crop industry, water supplement in this area largely depends on irrigation from the Murrumbidgee River. Therefore, timely and accurate mapping of land cover distributions is critical for the local Irrigation Company to make appropriate decisions on water budgeting and allocation, and to conduct environmental assessments.

Five cloud-free Sentinel-2A optical images were collected from the Google Earth Engine platform. The images were sensed on Jan. 28th, Feb. 07th, Feb. 27th, Mar. 09th, and Mar. 19th, respectively, in the year of 2017. Each image consists of 13 spectral bands covering the visible, near-infrared, and short-wave-infrared spectral regions. The images were provided in top of atmosphere (TOA) reflectance values with radiometric and geometric corrections applied (Level-1C data). The bands with 10 m (Bands 2, 3, 4, and 8) and 20 m (Bands 5, 6, 7, 8a, 11, and 12) spatial resolutions were reduced down to 60 m resolutions in order to make them consistent with the other bands (Bands 1, 9, and 10) that had 60 m resolutions.
There were five land cover classes identified in the study area: crops, forests, water, soil, and buildings. A total of 80 pixels were selected for each class in each image, and labelled via photointerpretation and with the help of the farmland boundary data provided by the Coleambally Irrigation Company. In the following experiments, the first four images were used as previous images to assist the training and classification of the fifth image.

B. Classifier Prediction Results

This section presents the classifier prediction results obtained with the proposed SCT-SVM approach. The temporal movements of classifiers, which were driven by the variations of class data, were firstly analyzed. Then fitting functions were used to fit the movement patterns of classifiers. Classifiers predicted with different fitting function were compared with the true classifiers to assess the prediction accuracies.

1) Temporal Variations of Classifiers:

The correlation between spectral features and sensing dates was firstly analyzed. Bands 4 (red) and 8 (near-infrared) were selected for the analysis. Linear correlation coefficients were calculated for the five land cover classes, with the results shown in Tab. 1. The highest temporal correlations were found for crops. The coefficient for the near-infrared band was -0.8306, indicating a strong negative correlation with time. The red band showed a positive but weaker correlation with a coefficient of 0.5213. Another vegetative class, forests, showed similar but less significant temporal correlations with coefficients of 0.4209 and -0.4699 for the red and infrared bands, respectively. The spectral features of the three non-vegetative classes, water, soil, and buildings, showed insignificant temporal correlations, with all the coefficients lower than 0.1.

More detailed analysis of the temporal correlations for crops and soil is shown in Fig. 6. By examining the class data of different images, it was found that the means of crops gradually moved from the top left towards the bottom right in the red/near-infrared feature space. In contrast, the means of soil remained relatively stable. Figure 7 explains the movement patterns of these classes data from a phenological perspective. It was found that the five images covered the mid-late growing periods of crops. During these periods, the crop leaves turned from green to yellow, accompanied by a gradual decrease in the value of Normalized Differential Vegetation Index (NDVI). The NDVI of soil remained stable with time, suggesting an insignificant temporal variation.
position moved consistently with time. Based on the first four
classifier positions, a predicted position was estimated for the
fifth classifier. Though the predicted classifier was not exactly
located at the true position, it was a good \textit{a priori} guess that
could provide advantageous help in seeking the true classifier.

2) Prediction Accuracies with Different Fitting Functions:

Different orders of fitting functions were tested for classifier
predictions. For each classifier separating two classes, the
Euclidean distance between the predicted and true positions
was calculated, in order to quantitatively measure the predic-
tion accuracy. The results are shown in Tab. II. The highest
accuracy for each classifier is shown in bold typeface. It was
found that, for classifiers that involved one or more vegetative
classes (crops and/or forests), second-order polynomial fitting
produced the best prediction results. For classifiers that sepa-
rated two non-vegetative classes, first-order polynomial fitting
was the best choice.

More detailed analysis of the prediction accuracies for
the crops-against-soil and soil-against-buildings classifiers are
shown in Fig. 9. For the crops-against-soil classifier, com-
pared with first-order polynomial, the second-order polynomial
was more capable of describing the non-linear movement of
classifier’s positions. The third-order polynomial generated
the worst result due to the overfitting effect. The non-linear
movement was less obvious for the soil-against-buildings
classifier, so the first-order fitting produced the best accuracy
in this case.

C. Classifier Fine-Tuning Results

The proposed SCT-SVM approach includes a new domain
adaptation algorithm, TA-SVM, to fine-tune classifiers from
the predicted positions to more accurate positions. In this
section, the performance of TA-SVM in classifier fine-tuning
was firstly compared with two existing algorithms, A-SVM

| Classifier       | First-Order | Second-Order | Third-Order |
|------------------|-------------|--------------|-------------|
| Crops/Forests    | 3.068       | 1.729        | 8.235       |
| Crops/Soil       | 2.515       | 1.324        | 7.142       |
| Crops/Water      | 2.613       | 1.771        | 6.851       |
| Crops/Buildings  | 1.791       | 1.271        | 5.982       |
| Forests/Soil     | 2.902       | 2.314        | 9.007       |
| Forests/Water    | 3.181       | 2.117        | 8.967       |
| Forests/Buildings| 3.452       | 2.209        | 8.437       |
| Soil/Water       | 0.912       | 1.127        | 5.426       |
| Soil/Buildings   | 0.814       | 1.477        | 3.251       |
| Water/Buildings  | 1.113       | 1.674        | 4.865       |

Fig. 8. Multitemporal positions of the crops-against-soil classifier in (a) the
feature space and (b) the classifier parameter space.

Fig. 9. Prediction accuracies for the crops-against-soil and soil-against-
buildings classifiers with different fitting functions.
and PMT-SVM. Then the impacts of two TA-SVM parameters, $F$ and $C$, on fine-tuning accuracies were analyzed and discussed.

1) Comparisons with A-SVM and PMT-SVM Algorithms:

The three fine-tuning algorithms, A-SVM, PMT-SVM and the proposed TA-SVM, were compared. In this analysis, the Number of Training Samples ($N_T$) from the fifth image ranged from 5 to 50, in order to analyze and compare the algorithms under different levels of training data insufficiency. Classification accuracies achieved with the fine-tuned classifiers on the rest of the labelled samples were calculated as a measure to evaluate the three algorithms.

The results are shown in Fig. 10. It was found that the accuracies of A-SVM and TA-SVM increased with $N_T$, but the TA-SVM generated relatively higher accuracies especially when training data were insufficient ($N_T<10$). The other algorithm, PMT-SVM, produced low accuracies for all $N_T$ levels. The results indicated that the TA-SVM outperformed the other two algorithms in classifier fine-tuning.

2) Impacts of Regularization Parameters:

The two regularization parameters adopted in TA-SVM, $F$ and $C$, control the contributions of information from previous images and current training samples, respectively. Here experiments were conducted to analyze the impacts of these parameters on the fine-tuning results. In the experiments, the Number of Training Samples ($N_T$) from the fifth image was set to 5, 20, and 50, corresponding to training sample cases that were insufficient, moderate and sufficient, respectively. The rest of the labelled samples were used to assess the accuracies of the classifiers fine-tuned with different settings of regularization parameters. The results are presented in Fig. 12.

Fig. 12a shows the results for $F$. It was found that, if training samples were limited ($N_T = 5$), higher classification accuracy was achieved with a larger $F$ value. Low classification accuracy was observed (76.4%) when $F$ was set to a small value of 0.01, as the information from previous images played an insignificant role. When $F$ was set to 10 or higher, it was able to provide enough weight for the information from previous images, and considerably increased classification accuracies were achieved (>92.6%). The importance of $F$ was less critical if training samples were sufficient ($N_T = 50$). When $F$ was increased from 0.01 to 10, the increase in classification accuracy was from 91.9% to 96.0%. This increase was less significant compared with the training sample insufficiency case (i.e., $N_T = 5$).

Fig. 12b shows the results for $C$. When $C$ was set to
Fig. 12. Impact of the regularization parameters $F$ and $C$ on overall classification accuracy with different number of training samples ($N_T$) in the incoming image.

D. Classification Results

We used the proposed SCT-SVM approach to classify the fifth image. The one-against-one strategy was used to extend the binary classifiers to multiple classes. According to the results presented above, regularization parameters $F$ and $C$ were set to 20 and 50, respectively. The second-order polynomial fitting function was selected for classifiers that involved one or more vegetative classes, while for the rest of the classifiers the first-order polynomial was selected. With the assistance of the first four images, the fifth incoming image was trained and classified under the training sample insufficiency situation ($N_T = 5$). The training and classification were repeated 10 times with different sets of training samples selected from the labelled data. The classification accuracies were compared with those obtained when SVM was applied directly to the fifth image without using the a priori knowledge from previous images (Dir-SVM).

The results are shown in Fig. 13. It was found that, with the a priori knowledge provided by previous images, the overall classification accuracy was improved from 76.2% to 93.8%. The greatest improvements were found for crops and forests, two vegetative land cover types that share similar spectral characteristics and have relatively high temporal dynamics. The remaining three land cover types showed improved classification accuracies as well, though the improvements were comparatively lower. These results demonstrate that, when the training samples are limited, the utilization of a priori knowledge from previous images is advantageous to enhance classification accuracy.

Fig. 13. Comparison of classification accuracies with and without the a priori knowledge from previous images. Error bars indicate the standard deviations of classification accuracies obtained with different sets of training samples.

IV. CONCLUSIONS

A SVM-based Sequential Classifier Training (SCT-SVM) approach is proposed for multitemporal remote sensing image classification. The approach leverages the classifiers of previous images to reduce the required number of training samples for the classifier training of a new image. Based on the temporal trend of previous classifiers, classifiers are firstly predicted for the new image. Then, a newly developed domain adaptation algorithm, Temporal-Adaptive Support Vector Machines (TA-SVM), is used to fine-tune the predicted classifiers into more accurate positions. Experimental results showed that the TA-SVM outperformed state-of-the-art algorithms in classifier tuning. Compared with those obtained without the assistance from previous images, the overall classification accuracy was improved from 76.2% to 93.8%, with the proposed SCT-SVM, which demonstrated that the leverage of a priori information from previous images could provide advantageous assistance for the classification of later images.

ACKNOWLEDGMENT

The authors would like to thank Mr. Graham Parton and Mr. Bernard Star at the Coleambally Irrigation Co-operative Ltd for providing farmland boundary data of the Coleambally Irrigation Area.
REFERENCES

[1] J. Verbesselt, R. Hyndman, A. Zeileis, and D. Culvenor, “Phenological change detection while accounting for abrupt and gradual trends in satellite time series,” Remote Sensing of Environment, vol. 114, no. 12, pp. 2970–2980, 2010.

[2] C. Gómez, J. C. White, and M. A. Wulder, “Optical remotely sensed time series data for land cover classification: A review,” ISPRS Journal of Photogrammetry and Remote Sensing, vol. 116, pp. 55–72, 2016.

[3] J. A. Richards, Remote Sensing Digital Image Analysis (Fifth Edition). Springer-Verlag Berlin Heidelberg, 2013.

[4] L. Bruzzone and B. Demir, “A review of remote sensing and pattern recognition methods for land-cover mapping in Europe,” in Land Use and Land Cover Mapping in Europe, vol. 3. Springer, 2014, pp. 127–143.

[5] Y. Shao and R. S. Lunetta, “Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points,” ISPRS Journal of Photogrammetry and Remote Sensing, vol. 70, pp. 78–87, 2012.

[6] K. Bahirat, F. Bovolo, L. Bruzzone, and S. Chaudhuri, “A novel domain adaptation bayesian classifier for updating land-cover maps with class differences in source and target domains,” IEEE Transactions on Geoscience and Remote Sensing, vol. 50, no. 7, pp. 2810–2826, 2012.

[7] B. Demir, F. Bovolo, and L. Bruzzone, “Classification of time series of multispectral images with limited training data,” IEEE Transactions on Image Processing, vol. 22, no. 8, pp. 3219–3233, 2013.

[8] J. T. Morgan, “Adaptive hierarchical classification with limited training data,” Ph.D. dissertation, the University of Texas at Austin, 2002.

[9] T. Sakamoto, M. Yokozawa, H. Toritani, M. Shihayama, N. Ishitsuka, and H. Ohno, “A crop phenology detection method using time-series MODIS data,” Remote Sensing of Environment, vol. 96, no. 3, pp. 366–374, 2005.

[10] J. G. Moreno-Torres, T. Raeder, R. Alair-Rodríguez, N. V. Chawla, and F. Herrera, “A unifying view on dataset shift in classification,” Pattern Recognition, vol. 45, no. 1, pp. 521–530, 2012.

[11] S. Vanonckelen, S. Lhermitte, and A. Van Rompaey, “The effect of atmospheric and topographic correction methods on land cover classification accuracy,” International Journal of Applied Earth Observation and Geoinformation, vol. 24, pp. 9–21, 2013.

[12] M. Xu, M. Pickering, A. J. Plaza, and X. Jia, “Thin cloud removal based on signal transmission principles and spectral mixture analysis,” IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 3, pp. 1659–1669, 2016.

[13] M. Xu, X. Jia, M. Pickering, and A. J. Plaza, “Cloud removal based on sparse representation via multitemporal dictionary learning,” IEEE Transactions on Geoscience and Remote Sensing, vol. 54, no. 5, pp. 2998–3006, 2016.

[14] F. Zhao, X. Gu, T. Yu, W. Verheof, Y. Guo, Y. Du, H. Shang, and H. Zhao, “Bidirectional reflectance effects over flat land surface from the charge-coupled device data sets of the HJ-1A and HJ-1B satellites,” Journal of Applied Remote Sensing, vol. 7, no. 1, pp. 073466–073466, 2013.

[15] Y. Guo, X. Jia, and D. Paull, “Superpixel-based adaptive kernel selection for angular effect normalization of remote sensing images with kernel learning,” IEEE Transactions on Geoscience and Remote Sensing, 2017.

[16] D. Tuia, E. Pasolli, and W. J. Emery, “Using active learning to adapt remote sensing image classifiers,” Remote Sensing of Environment, vol. 115, no. 9, pp. 2232–2242, 2011.

[17] D. Tuia, C. Persello, and L. Bruzzone, “Domain adaptation for the classification of remote sensing data: An overview of recent advances,” IEEE Geoscience and Remote Sensing Magazine, vol. 4, no. 2, pp. 41–57, 2016.

[18] H. L. Yang and M. M. Crawford, “Domain adaptation with preservation of manifold geometry for hyperspectral image classification,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 9, no. 2, pp. 543–555, 2016.

[19] L. Bruzzone and M. Marconcini, “Towards the automatic updating of land-cover maps by a domain-adaptation SVM classifier and a circular validation strategy,” IEEE Transactions on Geoscience and Remote Sensing, vol. 47, no. 4, pp. 1108–1122, 2009.

[20] B. Banerjee, F. Bovolo, A. Bhattacharya, L. Bruzzone, S. Chaudhuri, and K. M. Buddhhiraju, “A novel graph-matching-based approach for domain adaptation in classification of remote sensing image pairs,” IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 7, pp. 4045–4062, 2015.

[21] G. Matasci, M. Volpi, M. Kanevski, L. Bruzzone, and D. Tuia, “Semi-supervised transfer component analysis for domain adaptation in remote sensing image classification,” IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 7, pp. 3550–3564, 2015.

[22] B. Demir, F. Bovolo, and L. Bruzzone, “Updating land-cover maps by classification of image time series: A novel change-detection-driven transfer learning approach,” IEEE Transactions on Geoscience and Remote Sensing, vol. 51, no. 1, pp. 300–312, 2013.

[23] G. Mountrakis, J. Im, and C. Ogole, “Support vector machines in remote sensing: A review,” ISPRS Journal of Photogrammetry and Remote Sensing, vol. 66, no. 3, pp. 247–259, 2011.

[24] Y. Guo, X. Jia, and D. Paull, “A domain-transfer support vector machine for multi-temporal remote sensing imagery classification,” in 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, 2017.

[25] J. Nocedal and S. Wright, Numerical Optimization (Second Edition). Springer Science & Business Media, LLC, 2006.

[26] L. Bruzzone and M. Marconcini, “Domain adaptation problems: A DASVM classification technique and a circular validation strategy,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 5, pp. 770–787, 2010.

[27] L. Duan, I. W. Tsang, D. Xu, and S. J. Maybank, “Domain transfer SVM for video concept detection,” in IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009). IEEE, 2009, pp. 1375–1381.

[28] J. Yang, R. Yan, and A. G. Hauptmann, “Cross-domain video concept detection using adaptive SVMs,” in Proceedings of the 15th ACM International Conference on Multimedia. ACM, 2007, pp. 188–197.

[29] J. Yang, “Learning to adapt across multimedia domains,” Ph.D. dissertation, Columbia University, 2008.

[30] Y. Aytar and A. Zisserman, “Tabula rasa: Model transfer for object category detection,” in 2011 IEEE International Conference on Computer Vision (ICCV). IEEE, 2011, pp. 2252–2259.