Support Vector Regression for Rice Age Estimation Using Satellite Imagery

Panu Srestasathiern, Siam Lawawirojwong, and Rata Suwantong
Geo-Informatics and Space Technology Development Agency (GISTDA)
Laksi, Bangkok, Thailand
Email: panu,siam,rata@gistda.or.th

***Abstract*—Rice age estimation is an process in crop management and monitoring. In this paper, we present an approach for estimating age of rice on satellite image using support vector regression technique. The advantage of using support vector regression is that it is non-parametric technique. Therefore, any age model is not required. Moreover, the support vector regression is robust to outlier. The input data is the satellite image features i.e., radiometric information. To tune the parameters of support vector machine, the k-fold cross validation technique is utilized. The experiment was conducted using Landsat-8 image. Comparing the estimated age result with ground truth, the proposed method showed expected performance.**

I. Introduction

The goal of remote sensing is to analyze or measure physical quantity of an object or phenomenon without physical contact. The advantage of using remote sensing, especially satellite imagery, is that the observation can be done over large area. Therefore, it can be cost effective comparing with ground surveying.

One of the main applications of remote sensing is for agriculture monitoring or management. For example, remote sensing technique was used to count the amount of oil palm trees [1]. In [2], satellite imagery was used to estimate the above-ground biomass of secondary forest in Brazil. By using remote sensing , agricultural monitoring over large are is possible and cost effective. In [3], multispectral satellite imagery i.e., FORMOSAT-2, was used to map its cultivation area and monitor its crop status on regional scales. Furthermore, remote sensing can also be used to detect plantation area [4] and yield prediction [5].

Rice is one of Thailand's important agricultural products and world's food crops. Monitoring the rice growing stage, age or seeding date is very crucial information for management. Remote sensing technology has been used for collecting spatial and temporal information about rice crops. The aim of this work is hence to use the satellite imagery for estimating rice age which is a parameter for yield estimation.

In literature, one of the a popular approach is to analyze the time series data of image feature extracted from satellite imagery. In [6], Suwantong et al. used time series of normalized vegetation index from MODIS imagery to estimate the starting date of rice growing (date of transplantation). The concept is to smoothing the noisy times series of the normalized vegetation index. Particularly, the moving horizon estimator was employed. Moreover, the time series was modeled as the triply modulated cosine function. To estimate the starting date, the threshold on the phase of cosine function.

Nuarsa et al. [7] used the relationship between the rice age and normalized vegetation index. Particularly, the relation was modeled as the quadratic function. Moreover, the total sum of normalized vegetation index was also used for yield estimation.

In [8], Choudhury et al. used radar satellite imageries from RADATSAT-1 and ENVISAT to estimate the height, date of transplantation and plant biomass of rice. The process begins with the extraction of back-scattering coefficient from satellite imagery. The height and date of transplantation can then be retrieved using parametric models i.e., polynomial functions, of which coefficients are determined a priori. The crop growth profile can then be computed using temporal back-scattering coefficient.

Similar to the work proposed in [8], Lopez-Sanchez etal. [9] proposed a method for estimating key dates and stage of rice crops from dual-polarization radar satellite imagery. To estimate the key date, crop growth profile was modeled as a dynamical system using state-space technique. The particle filtering approach was then adapted to perform key dates or phenological stage estimation.

To estimate rice's age or date of transplantation from optical satellite imagery, in this paper, we adopt the support vector machine for regression problems. That is, it is used for modeling the relation between input image features i.e., radiometric information, and rice age. The advantage of using the support vector machine is that it is non-parametric model. Therefore, the physical model of rice growing and the distribution of the input information do not need to be known a priori. Moreover, an interesting property of the support vector machine is the robustness to outlier. To tune the parameters of support vector machine, the cross-validation technique is utilized. To test the performance of the proposed method, Landsat-8 imagery was used.

The rest of this paper is organized as follow. The concept of support vector machine for regression problem is discussed in Section II. The input image feature and parameter tuning technique will be discussed in Section III. The experimental result and performance evaluation will be presented in Section IV. The conclusion of this work will be discussed in Section

978-1-4673-9749-0/16/$31.00

V.

II. Regression by support vector machine

A. Support vector machine

Support Vector Machine (SVM) was first proposed for classification problem [10]. It is a supervised non-parametric statistical learning technique. Therefore, its major advantage is that the distribution of the data does not need to be known a priori [11], while other statistical technique e.g., maximum likelihood estimation usually assume that data distribution is known a priori.

To explain the concept of the support vector machine, a linear two class classification problem is used, see Figure 1. The aim of the suport vector machine technique is to find a hyperplane separating data into many classes, which is two classes in this case. Such hyperplane is called *decision boundary* or *SVM hyperplane*. To obtain a unique hyplane or optimal separation, a constraint that there is no data point in the margin of hyperplane is imposed, see Figure 1. The data points on the margin is called *suport vectors*. On other words, support vectors are used to define maximal margin hyperplane.

If the data is not distributed linearly, using hyperplane cannot separate data into many classes efficiently. To handle non-linear distribution of the data, the data is projected into higher dimensional space such that the data points distributed linearly in the new space. Using a proper projection function, the inner product in the higher dimensional space can be can be computed in the original space without mapping the data point into the feature space which possibly has infinite dimensionality via the use of *kernel function*.

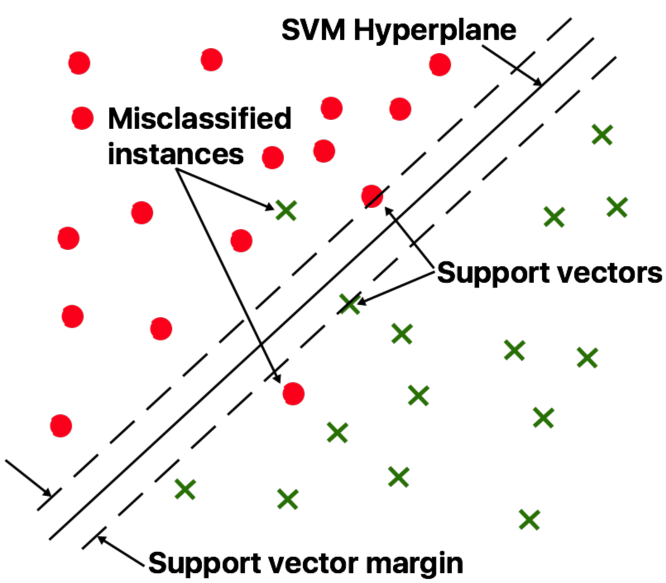

Fig. 1. Illustration of support vector machine's concept. Red dots and green cross are data in $\mathcal{R}^n$. The goal of the support vector machine is to find a hyperplane separating data such that there is no data in the hyperplane's margin.

B. Support vector machine for Regression

Support vector machine can also be applied for regression problem. That is, it is applied to find the continuous prediction output. In order to explain the support vector regression, the linear regression is used as an example. Given a linear function $f : \mathcal{R}^n \to \mathcal{R}$:

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b. \quad (1)$$

The goal of the linear regression is then to estimate the parameters $\mathbf{w}$ and $\mathbf{b}$. That is, the set of data $\mathcal{T} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_l, y_l)\}$, where $\mathbf{x} \in \mathcal{R}^n$ and $y \in \mathcal{R}$, is used to estimate the parameters of the linear function f.

By modified the concept of support vector machine, the regression by support vector is then to find the function having the most ϵ deviation (support vector margin) from y_i for all training data. The function f can then be estimated by solving the objective function:

$$\min_{\mathbf{w},\xi} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l}\left(\xi_i^+ + \xi_i^-\right) \quad (2)$$

subject to the constraint:

$$\begin{aligned} y_i - \langle \mathbf{w}, \mathbf{x}_i \rangle - b &\leq \epsilon + \xi_i^+ \\ \langle \mathbf{w}, \mathbf{x}_i \rangle + b - y_i &\leq \epsilon + \xi_i^- \\ \xi_i^+, \xi_i^- &\geq 0 \end{aligned} . \quad (3)$$

The trade-off between the flatness of the function f and the data deviation is controlled by the constant $C > 0$. That is, C is used to penalize the margin errors, when data points are outside support vector margin. The slack variables ξ^+ and ξ^- are introduced to cope with infeasible constraints of the optimization problem [12]. Namely, they are used for penalizing data points which violate the margin requirements.

In order to solve the primal problem of (3) efficiently, its dual formulation is utilized. In [13], it is shown that the final solution of (3) can be given by:

$$\mathbf{w} = \sum_{i=1}^{l}(\alpha_i + \alpha_i^*)\mathbf{x}_i, \quad (4)$$

and

$$f(\mathbf{x}) = \sum_{i=1}^{l}(\alpha_i + \alpha_i^*)\langle \mathbf{x}_i, \mathbf{x} \rangle + b, \quad (5)$$

where α and α^* are Lagrange multipliers.

The previous example is for the linear case. Similar to the support vector machine for classification problem, the kernel trick is also used in support vector regression in order to deal with non-linear problem. Some popular kernel function are:

- Gaussian radial basis function: $k(\mathbf{x}_i, \mathbf{x}_j) = e^{\gamma(\mathbf{x}-\mathbf{x}_j)^2}$,
- Power: $k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^\top \mathbf{x}_j)^d$,
- Polynomial: $k(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^\top \mathbf{x}_j + h_0)^d$.

These kernel can be used depending on the tasks. Replacing the inner product with the kernel function, the solution for the non-linear support vector machine can be formulated:

$$\langle \mathbf{w}, \mathbf{x} \rangle = \sum_{i=1}^{l}(\alpha_i + \alpha_i^*)K(\mathbf{x}_i, \mathbf{x}), \quad (6)$$

and

$$f(\mathbf{x}) = \sum_{i=1}^{l} (\alpha_i + \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b, \tag{7}$$

In this paper, the non-linear support vector regression is preferred because the aging process is complex. In the next Section, the multi-spectral satellite image features used for rice age estimation is discussed.

III. Rice age estimation

A. Satellite image features

In order estimate the rice age using a satellite image, an important task is to extract image feature that is related to the age of rice. Particularly, image information that is related to the complex aging process of rice must be selected. Texture is an commonly used image feature. However, it may not be appropriate because the resolution of the satellite image is very coarse e.g., 50 meters.

In this paper, we adopt to use the radiometric information as the image feature for rice age estimation. For example, green and yellow can be used for roughly estimate the rice age because the rice color change from green to yellow when it is growing. The commonly used satellite image bands consist of Red (R), Green (G), Blue (B) and Near-Infrared (NIR).

An important characteristic of plant's chlorophyll is that it absorbs energy in the wavelength of visible red band and is transparent in the wavelength of near-infrared. This characteristic was used to derive a vegetation index called Normalized Difference Vegetation Index (NDVI) which is defined as:

$$NDVI = \frac{NIR - R}{NIR + R}, \tag{8}$$

where NIR and R are extracted from near-infrared and red bands of satellite imagery. An alternative vegetation index is the Normalized Difference Index (NDI) which is defined as:

$$NDI = \frac{G - R}{G + R}. \tag{9}$$

Therefore, in this paper, the feature vector used for rice age estimation is:

$$\mathbf{x} = [R, G, B, NIR, NDVI, NDI]. \tag{10}$$

While the values of NDVI and NDI is in the range of $[-1, 1]$, the values of red, green, blue and near-infrared of a satellite imagery is not. For example, the value of the red band can be in the range $[0, 255]$. In the implementation, the feature vector $\mathbf{x}$ must be scaled:

$$\mathbf{x} = \left[\frac{R}{R_{max}}, \frac{G}{G_{max}}, \frac{B}{B_{max}}, \frac{NIR}{NIR_{max}}, NDVI, NDI\right]. \tag{11}$$

B. Parameter tuning by k-fold cross-validation

In practical, the data is separated into two sets. The first one is called *training set*, which is used for estimate the function, and the second *test set*, which is used to test the accuracy of the function estimate. In order to obtain optimal accuracy, the parameter for estimating the function must be tuned properly.

The k-fold cross validation method is a popular and practical approach for parameter tuning (or model selection) in machine learning applications. Particularly, in this paper, it is adopted to tuning the parameter of the support vector regression. The concept of k-fold cross validation is to divide the data into k (randomly selected) disjoint subsamples of roughly equal size i.e., $\frac{n}{k}$. The commonly used k values are 5 and 10. When $k = n$, this case is called *leave-one-out*.

Given a set of training dataset S and the set of support vector regression's parameters Θ, the algorithm for k-fold cross-validation can be summarized as following:

1) Divide the training dataset S into k equal size disjoint subset i.e., $S_1, S_2, S_3, \ldots, S_k$.
2) for each $\theta \in \Theta$
    a) For each validation sample S_i, the function f_i is estimated with the parameter θ using the union of the other validation samples . $\bigcup_{j \neq i} S_j$
    b) Compute the empirical risk for the validation sample S_i: $r_i = \frac{k}{n} \sum_{\mathbf{x} \in S_i}^{k} |y - f_i(\mathbf{x})|$.
    c) Compute the average empirical risk from all validation sample: $error(\theta) = \frac{1}{k} \sum_{i=1}^{k} r_i$.
3) Find $\theta^* = \min_{\theta} error(\theta)$.
4) Estimate the function f with the parameter θ^* using the entire dataset S.

In this paper, the empirical risk is defined as the absolute error in order. Therefore, the cross validation score will be consistent with performance measure used in the experiment.

IV. Experiment

To test the performance of the proposed methods, we first did a ground survey for collecting rice age in Ang-Thong district, Thailand. The shape file of paddy fields in Ang-Thong district was overlaid on Landsat-8 image for planing the field survey. Rice age from about 300 sampled sites were recorded and GPS receiver was used to locate the sampled sites. To extract feature from satellite imagery i.e., Landsat-8, the radiometric information at the sampled sites and their neighbors in 3×3 windows were used. That is we assume that the rice age at the sampled sites and their neighbor were equal. Therefore, the size of the data is about 2,700.

To estimate the relation between the input image feature (11) and rice age, in this experiment, the Gaussian radial basis function is used as the kernel for support vector machine.. Therefore, the parameters of support vector machine and kernel are C, ϵ and γ. These parameters need to be tuned for optimal performance. However, γ is fixed at 10 in order to avoid data overfitting problem [14]. To tune C and ϵ, we used 5-fold cross-validation technique. To perform cross-validation, parameter grid of C and ϵ was created such that the parameter range is the power of 2 i.e., 1, 2, 4, 8, 16 and 32.

The cross-validation score of each parameter pair is reported in Table I. It can be observed that the optimal parameters for support vector machine are $\epsilon = 1$ and $C = 32$. According to Table I, it can be noticed that cross-validation score increases when C increase. The is because, with a specific ϵ, C gives penalties to the points outside support vector margin, see (3). With a large C value, most of the data point should therefore deviate from the support vector hyperplane less than ϵ in order to minimize the objective function (3). Conversely, using large ϵ value increase the estimation error because outlier can be in the support vector margin. That is, the error from outlier will not appear in the objective function (3), if the ϵ value is too large.

TABLE I
CROSS VALIDATION SCORE OF EACH PARAMETER PAIR ($\gamma = 10$)

	$\epsilon = 1$	$\epsilon = 2$	$\epsilon = 4$	$\epsilon = 8$	$\epsilon = 16$	$\epsilon = 32$
$C = 1$	7.474	7.488	7.504	7.654	8.083	14.234
$C = 2$	7.394	7.407	7.433	7.545	8.027	14.234
$C = 4$	7.350	7.353	7.374	7.483	7.981	14.233
$C = 8$	7.316	7.315	7.336	7.454	7.924	14.236
$C = 16$	7.281	7.279	7.308	7.426	7.860	14.244
$C = 32$	**7.240**	7.246	7.282	7.407	7.837	14.251

To evaluate the performance of our algorithm for rice age estimation, two different measures i.e., Mean Absolute Error (MAE) and Cumulative Score (CS), were used. The mean absolute error is defined as the mean of absolute error between ground truth and estimated one:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |\widehat{y}_i - y_i| \,, \tag{12}$$

where $\widehat{y}_i$ is the predicted rice age using input $\mathbf{x}_i$ i.e., $\widehat{y}_i = f(\mathbf{x}_i)$. The Cumulative Score (CS) is defined as:

$$CS = \frac{N_{e<th}}{N} \times 100\%, \tag{13}$$

where $N_{e<th}$ is the amount of prediction errors which is less than the threshold th. Therefore, the cumulative score profile monotonically increases and converges to 100.

The distribution of absolute error using optimal parameter to train the support vector machine is illustrated in Figure 2. One can see that the distribution of absolute error decreases exponentially when absolute error increases.The integration of the distribution is then the profile of cumulative score according to (13).

In Figure 3, the cumulative score (13) is plotted as a function of threshold value in the range $[0, 60]$. Particularly, the cumulative scores from support vector machines using optimal parameters i.e., $C = 32$ and $\epsilon = 1$, and sub-optimal parameters i.e., $C = 32$ and $\epsilon = 32$, are compared. The It can be observed that, using optimal parameters almost 80% of

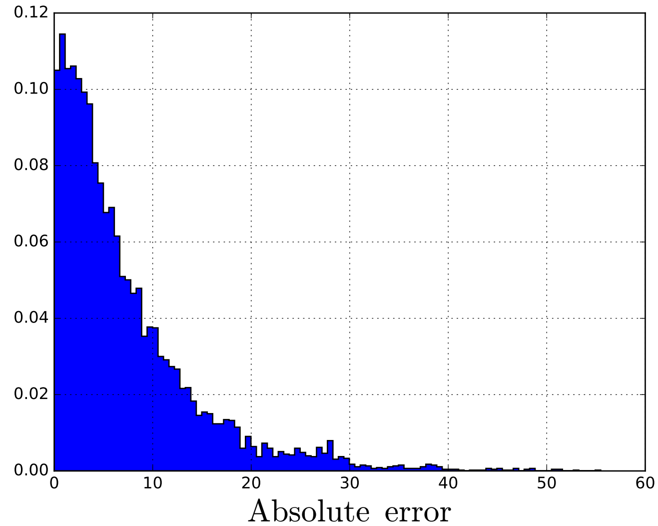

Fig. 2. Distribution of absolute error using optimal parameters i.e., $C = 32$ and $\epsilon = 1$, for training the support vector machine.

the error is less than 10, while, using sub-optimal parameters, 80% of the errors are greater than 20 days. This observation is hence prove the importance of optimal parameter tuning for training the support vector machine.

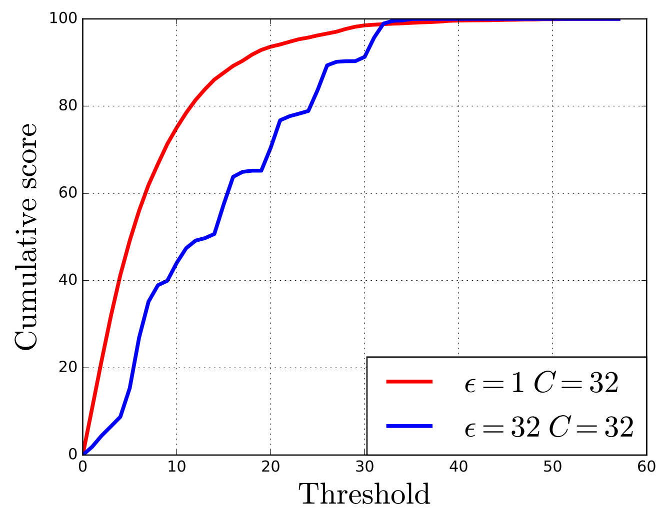

Fig. 3. Cumulative score of support vector machine trained using optimal parameters (red) and sub-optimal parameters (blue).

V. CONCLUSIONS AND FUTURE WORK

In this paper, we present an approach for estimating rice age using satellite imagery. Particularly, the support vector regression technique was utilized to establish the relationship between the image features and rice age i.e., R, G, B, NDVI and NDI. The performance of the proposed method wasted using Landsat-8 imagery of Ang-Thong district. To obtain an acceptable age estimation result, the parameters of the support vector machine was tuned using 5-fold cross validation technique. From the experiment, the estimation error is about 7 days. Since all estimations always have error, in the future work, the time series of satellite imagery will be taken in to account in order to reduce the estimation or checking whether the estimated age is correct or not.

REFERENCES

[1] P. Srestasathiern and P. Rakwatin, "Oil palm tree detection with high resolution multi-spectral satellite imagery," *Remote Sensing*, vol. 6, no. 10, pp. 9749–9774, 2014.

[2] M. K. Steininger, "Satellite estimation of tropical secondary forest above-ground biomass: data from brazil and bolivia," *International journal of remote sensing*, vol. 21, no. 6-7, pp. 1139–1157, 2000.

[3] Q. Zhao, V. I. Lenz-Wiedemann, F. Yuan, R. Jiang, Y. Miao, F. Zhang, and G. Bareth, "Investigating within-field variability of rice from high resolution satellite imagery in qixing farm county, northeast china," *ISPRS International Journal of Geo-information*, vol. 4, pp. 236–261, 2015.

[4] M. K. Mosleh and Q. K. Hassan, "Development of a remote sensing-based boro rice mapping system," *Remote sensing*, vol. 6, pp. 1938–1953, 2014.

[5] M. K. Mosleh, Q. K. Hassan, and E. H. Chowdhury, "Application of remote sensors in mapping rice area and forecasting its production: A review," *Sensors*, vol. 15, pp. 769–791, 2015.

[6] R. Suwantong, P. Srestasathiern, S. Lawawirojwong, and P. Rakwatin, "Moving horizon estimator with pre-estimation for crop cultivation date estimation in tropical area," in *American Control Conference*, 2016, accepted conference paper.

[7] I. W. Nuarsa, F. Nishio, and C. Hongo, "Relationship between rice spectral and rice yield using modis data," *International Journal of Agricultural Sciences*, vol. 3, no. 2, pp. 80–88, 2011.

[8] I. Choudhury, M.Chakraborty, and J. Parihar, "Estimation of rice growth parameter and crop phenology with conjunctive use of radarsat and envisat," in *Envisat Symposium 2007*, April 2007.

[9] C. De Bernardis, F. Vicente-Guijalba, T. Martinez-Marin, and J. Lopez-Sanchez, "Estimation of key dates and stages in rice crops using dual-polarization sar time series and a particle filtering approach," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 8, no. 3, pp. 1008–1018, March 2015.

[10] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *5th Annual ACM Workshop on COLT*, D. Haussler, Ed. ACM Press, 1992, pp. 144–152.

[11] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, pp. 247–259, 2011.

[12] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, pp. 199–222, 2004.

[13] V. N. Vapnik, *Statistical Learning Theory*. John Wiley, 1998.

[14] A. Ben-Hur and J. Weston, "A users guide to support vector machines," Tech. Rep.