

REMOTE SENSING CLASSIFICATION METHOD OF WETLAND BASED ON AN IMPROVED SVM

Yi Lin^{a,b}, Mingge Shen^{a,b}, Bing Liu^b, Qin Ye^{a,b}

^a College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092;

^b Research Center of Remote Sensing & Spatial Information Technology, Tongji University, Shanghai 200092

Commission VII WGVI/6

KEY WORDS: Multi-source remote sensing image, Wetland, Classification model, Support vector machine, Mixed kernel function

ABSTRACT:

The increase of population and economic development, especial the land use and urbanization bring the wetland resource a huge pressure and a serious consequence of a sharp drop in the recent years. Therefore wetland eco-environment degradation and sustainable development have become the focus of wetland research. Remote sensing technology has become an important means of environment dynamic monitoring. It has practical significance for wetland protection, restoration and sustainable utilization by using remote sensing technology to develop dynamic monitoring research of wetland spatial variation pattern. In view of the complexity of wetland information extraction performance of the SVM classifier, this paper proposes a feature weighted SVM classifier using mixed kernel function. In order to ensure the high-accuracy of the classification result, the feature spaces and the interpretation keys are constructed by the properties of different data. We use the GainRatio (feature_i) to build the feature weighted parameter h and test the different kernel functions in SVM. Since the different kernel functions can influence fitting ability and prediction accuracy of SVM and the categories are more easily discriminated by the higher GainRatio, we introduce feature weighted ω calculated by GainRatio to the model. Accordingly we developed an improved model named "Feature weighted & Mixed kernel function SVM" based on a series of experiments. Taking the east beach of Chongming Island in Shanghai as case study, the improved model shows superiority of extensibility and stability in comparison with the classification results of the experiments applying the Minimum Distance classification, the Radial Basis Function of SVM classification and the Polynomial Kernel function of SVM classification with the use of Landsat TM data of 2009. This new model also avoids the weak correlation or uncorrelated characteristics' domination and integrates different information sources effectively to offer better mapping performance and more accurate result. The accuracy resulted from the improved model is better than others according to the Overall Accuracy, Kappa Coefficient, Omission Errors and Commission Errors.

1. INTRODUCTION

Wetlands are considered as one of the most biologically diverse ecosystems, serving as critical habitat and productive intertidal zones to a wide range of wild plant and animal. They also play a number of roles in the environment, principally water purification, flood control, and shoreline stability which relate closely to human. The wetland take on the characteristics of a distinct ecosystem for a land area saturated with water, either permanently or seasonally. Biodiversity loss occurs in wetland systems because of land use changes, habitat destruction, pollution, exploitation of resources, and invasive species. With overexploitation of resources and neglect of environment protection, environmental degradation is more prominent within wetland systems than any other ecosystem on our planet which threatens human survival. How to control the degradation and produce sustainable wetlands becomes the crucial international research project.

Remote sensing technology becomes the important means of earth observation for superiority in extensive regional coverage; continuous acquisition of data; accurate and up-to-date information; comparability of a large archive of historical data and so on. It has been widely used in resources investigation, classification, change detection, landscape pattern change analysis and function assessment of wetland over recent two decades. And wetlands monitoring is the important element of

Geographical Conditions Monitoring, which makes research on wetlands based on remote sensing significant.

Consequently, wetlands information extraction using remote sensing data is primary step during wetland monitoring. When it comes to support vector machine, no one can deny that SVM made great contribution to classification and regression analysis. A support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which is a good separation achieved by the hyperplane that has the largest distance to the nearest training data point of any class (so-called functional margin), since the larger the margin the lower the generalization error of the classifier. The SVM algorithm is based on the statistical learning theory and the Vapnik-Chervonenkis (VC) dimension. Bernhard E. Boser, Isabelle M. Guyon and Vladimir N. Vapnik suggested a way to create nonlinear classifiers by applying the kernel trick to maximum-margin hyperplanes in 1992. This allows the algorithm to fit the maximum-margin hyperplane in a transformed feature space. The transformation may be nonlinear and the transformed space high dimensional; thus though the classifier is a hyperplane in the high-dimensional feature space, it may be nonlinear in the original input space. It simultaneously minimizes the empirical classification error and maximizes the geometric margin; hence they are also known as maximum margin classifiers. With the special properties, SVMs are used to solve various real world

problems such as medical science, hand-written characters reorganization, hypertext categorization and image classification.

Therefore, scholars proposed a lot of improved methods on SVM classification. Lin and Zhang introduce fuzzy mathematics into the SVM models. Each sample weighted in SVM model (WeightedSVM) to overcome the gap of different sample sizes. Zhao Hui etc synthesized WSVM and FSVM to raise the Dual-Weighted SVM which considers both importance and size difference. Besides the advantages of SVMs - from a practical point of view - they have some drawbacks. The problems with SVMs are the high algorithmic complexity; selection of the kernel function parameters; extensive memory requirements of the required quadratic programming in large-scale tasks which need to be solved properly.

1.1 Study area

Study area locates in the east of Chongming Island at the estuary of Yangtze River; Chongming Dongtan (121° 50' E~122° 05' E, 31° 25' N~31° 38' N) wetland is an important habitat for international migratory birds as well as a key area of biological diversity of the coastal wetlands in China. Since the geographical location and the erosion condition, Dongtan becomes the important ecologically fragile area in China. Its ecological environment has been rapidly and constantly changing with economic development as well as human activities. So, it is essential to carry out the environment and resources investigation in Chongming Dongtan, which has important practical meaning. The study area, about 32610 hectares, is the largest wetland of Yangtze Estuary and extending seaward at a rate of 150~300 meter per year.

With humid subtropical climate characterized by hot, humid summers and generally mild to cool winters, study area's the average annual temperature is 15.3°. There is extremely limited precipitation during the winter, owing to the powerful anticyclonic winds from Siberia. Annual rainfall is over 1022mm. The climate and location make it to be the important region of migratory biology. It was designated as internationally important under the Ramsar Wetlands Convention in 2001 and a national nature reserve in 2005.

2. METHODS

2.1 Pre-processing

The testing data are Landsat TM images and Mapping Satellite-I multi-spectral images cover the study area. The Landsat images have the longest history for acquisition of satellite imagery of Earth, as a unique resource for global change research and applications in agriculture, cartography, geology, forestry, regional planning, surveillance and education. Mapping Satellite -I(also known as Tian Hui-1)is a Chinese earth observation satellite built by Dong Feng Hong, a China Aerospace Science and Technology Corporation (CASC). Tian Hui-1 was launched on 6 May 2012 into a sun-synchronous, polar orbit with an apogee of 490 km (300 mi) and perigee of 505 km (314 mi). Tian Hui 1 is equipped with two different camera systems in the visible and infrared range. The visible light camera is able to produce three-dimensional pictures in the spectral region between 510 to 690 nanometers with dissolution of approximately 5 meters and a field of view of approximately 25 degrees. The infrared camera reaches dissolution of approximately 10 meters and covers four wavelengths. The improved CCD cameras guarantee accurate data acquisition.

The images acquired on different time must keep the same spatial resolution in change detection, so we should resample the images to the same resolution.

In order to make a meaningful measure of radiance at the Earth's surface, the atmospheric interferences must be removed from the data. To insure a reliable use of this kind of data, a rigorous radiometric normalization step is required. However, obtaining an atmospheric characterization at a given acquisition date is difficult for every images. Normalization can be addressed by performing an atmospheric correction of each image in the time series. In this study, we chose histogram matching (HM) as method because different reference image yields different residual error. Histogram matching is a commonly used radiometric enhancement technique fully integrated in many image processing software packages called relative radiometric normalization. The histogram matching on other images of the given image obtained in October, 2002 to eliminate the influence of solar altitude or the atmosphere.

To compensate for the distortion created by off-axis projector or screen placement or non-flat screen surface, image geometry correction is the imperative process. As for the same index of TM images, we took one of TM images as a given data and map from sources to images on ENVI software.

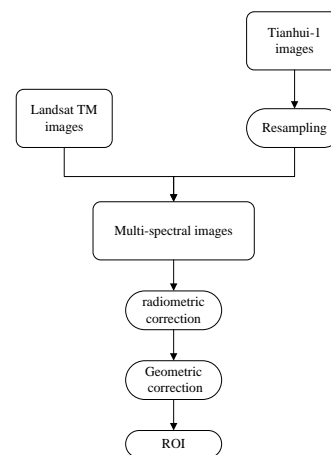


Figure 1. Pre-processing

2.2 Feature space

Interpretation and analysis of remote sensing images involves the identification or measurement of various targets in an image to extract useful information about them. Targets in remote sensing images may be any feature or object which can be observed in an image. Interpretation is the critical process of making sense of the data. Image analysis is the recently developed automated computer-aided application which is in increasing use. Firstly, analysis makes use of certain key elements of recognition.

The remote sensing data are digital which makes them most suitable for operations between bands. The indexes calculated within different bands can contribute to analysis. The Normalized Difference Vegetation Index (NDVI) is a simple graphical indicator that can be used to analyze remote sensing measurements. The NDVI is calculated from these individual measurements as follows:

$$NDVI = \frac{(NIR - VIS)}{(NIR + VIS)} \quad (1)$$

The Tasseled-cap transform is performed by taking “linear combinations” of the original image bands similar in concept to principal components analysis. Three feature bands are brightness, greenness and wetness, which are generated after operating on TM images. The third Tasseled-cap band is often interpreted as an index of “wetness”. We used wetness band which define the “soil plan” and represents the primary feature. The wetlands are characterized as having a water table that stands at or near the land surface for a long enough period each year to support aquatic plants. Therefore, the value of the index is between the values of soil and water.

After interpretation and analysis, we structure the 5-D feature space including the 4th band, the 3rd band, the 2nd band, NDVI, “wet” band (the third tasseled-cap band).

2.3 Feature weighted& Mixed kernel function SVM

2.3.1 Mixed kernel function

The Kernels are used in Support Vector Machines to map the learning data into a higher dimensional feature space where the computational power of the linear learning machine is increased. The kernel function may transform the data into a higher dimensional space to make it possible to perform the separation. SVMs are linear learning models that use a dual representation and operate in a kernel induced feature space. The research has a long history. Micchelli C.A., MacKay D., Girosi F., Schölkopf B. and other scholars raised different method on the function structure (MacKay D., 1999). Haussler D., Watkins C. proposed that kernel function could be defined on original set, not only Euclidean Space(Watkins C.,1999). With this theory, we support the new kernel function based on old ones.

Definition 1: K_1 and K_2 are kernels of $X \times X$ space. $X \subseteq R^n, a \in R^+$, The following functions are also kernel functions:

$$K(x, z) = K_1(x, z) + K(x, z) \quad (2)$$

$$K(x, z) = aK(x, z) \quad (3)$$

Proof: Fix a finite set of points $\{x_1, \dots, x_n\}$, K_1 and K_2 are matrixes in the data set. We must consider freedom vectors. Necessary and sufficient conditions to decide whether K is positive semi-definite matrix or not is: $\alpha K \alpha \geq 0$.

(1) $\alpha(K_1 + K_2)\alpha = \alpha K_1 \alpha + \alpha K_2 \alpha \geq 0$, $K_1 + K_2$ is positive semi-definite matrix and also kernel;

(2) Similarly, $\alpha a K_1 \alpha = a \alpha K_1 \alpha \geq 0$, $a K_1$ is also kernel.

The polynomial kernel shows better extrapolation abilities at lower orders of the degrees, but requires higher orders of degrees for good interpolation. On the other hand, the RBF kernel (a local kernel) has good interpolation abilities, but fails to provide longer range extrapolation. Therefore, Smits G.F. and Jordan E.M. presented the mixtures of with these two kernels. It is shown that, where the RBF kernels fails to extrapolate and a very high degree Polynomial kernel is needed to interpolate well, the mixture of the two kernels is able to do both. The model makes use of the character of mixture kernel by more global and local ability and the influence of difference kernels which can be turned by weight factor in the determination of the kernels.

There are several ways of mixing kernels. What is important though, is that the resulting kernel must be an admissible kernel.

One way to guarantee that the mixed kernel is admissible, is to use a convex combination of the two kernels K_{poly} and K_{rbf} , for instance

$$K_{mix} = \lambda K_{poly} + (1 - \lambda) K_{rbf} \quad (4)$$

where the optimal mixing coefficient λ has to be determined. The value of λ is a constant scalar. The scholars prove that the degree of polynomial and width of RBF were fixed to 1 and 0.15, respectively. The mixing coefficient was varied between 0.5 and 0.95 only.

Another possibility of mixing kernels is to use different values of λ for different regions of the input space. Through this approach, the relative contribution of both kernels can be varied over the input space. In this paper, a uniform λ over the entire input space is used. We select 500 for samples for training and testing. It shows that the results are optimal when λ equals 0.9 and degree of polynomial $d=4$ (Table 1).

values of parameter	number of vector	accuracy
$\lambda=0.5$ $d=4$	25	97.60%
$\lambda=0.6$ $d=4$	23	97.60%
$\lambda=0.7$ $d=4$	22	98.00%
$\lambda=0.8$ $d=4$	22	98.80%
$\lambda=0.9$ $d=4$	22	99.60%
$\lambda=0.9$ $d=3$	25	97.60%
$\lambda=0.9$ $d=2$	40	97.60%

Table 1. Accuracy of different values of parameter

2.3.2 Feature weighted

In SVM learning, the data is mapped non-linearly from the original input space X to a high-dimensional feature space F and separated by a maximum-margin hyperplane in that space F . The per-feature distance multiplied by the weight of the feature effected accuracy of model to a certain extent. Notice that an inner product is in fact one of the most simple similarity measures between vectors as it gives much information about the position of these vectors in relation to each other. In that sense the learning process can benefit a lot from the use of special purpose similarity or dissimilarity measures in the calculation of K . As a weighting scheme we used a quantity called information gain ratio which calculates for every feature the amount of information it contains with respect to the determination of the class label. To remedy this bias information gain can be normalized by the entropy of the feature values, which gives the gain ratio:

$$GainRatio(A) = \frac{I(t_1, t_2, \dots, t_n) - E(A)}{-\sum_{i=1}^v p(a_i) \log(p(a_i))} \quad (5)$$

For a feature with a unique value for each instance in the training set, the $entropy = -\sum_{i=1}^v p(a_i) \log(p(a_i))$ of the feature

values in the denominator will be maximally high, and will thus give it a low weight. It is applied to characteristics correlation analysis, which will set higher weight to strong correlated feature and lower to weak correlated ones. Thus the model can avoid the domination of the weak correlated or uncorrelated feature. Information gain ratio biases the decision tree against considering attributes with a large number of distinct values. So it solves the drawback of information gain—namely,

information gain applied to attributes that can take on a large number of distinct values might learn the training set too well. For example, suppose that we are building a decision tree for some data describing a business's customers. Information gain is often used to decide which of the attributes are the most relevant, so they can be tested near the root of the tree. One of the input attributes might be the customer's credit card number. This attribute has a high information gain, because it uniquely identifies each customer, but we do not want to include it in the decision tree: deciding how to treat a customer based on their credit card number is unlikely to generalize to customers we haven't seen before.

Definition 2: K are kernels of $X \times X$ space, $X \subseteq R^n$, P is linear transformation matrix in n . And we can get feature weighted kernel function:

$$K_p(x_i, x_j) = K_p(x_i^T P, x_j^T P) \quad (6)$$

Therefore, we take $\text{GainRatio}(\text{feature}_i)$ calculate feature weight vector ω . In the training process, the main algorithm follows.

$$\begin{cases} \max : L(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j K(x_i^T P, x_j^T P) \\ S.T. \\ C \geq \alpha_i \geq 0, i = 1, 2, \dots, l \\ \sum_{i=1}^l \alpha_i y_i = 0 \end{cases}$$

3. EXPERIMENT

In the experiments we want to evaluate the performance of the new model. So we also selected polynomial kernel, RBF kernel and Minimum Distance to do our experiment for comparison. Our experiment processed on matlab software. The main procedures of new model are following.

Step 1: Mark each unnamed entity as $\text{feature}_i (1 \leq i \leq n)$ to calculate GainRatio to weigh feature. Diagonal matrix P constructed from feature weighted vector ω ($P = \text{diag}(\omega)$). Linear transformation on feature space (2.2) to achieve all feature weighted;

Step 2: Train the Feature Weighted of multi- dimensions data set with mixtures kernel;

Step 3: Make the optimal classification decision according to SVM algorithm;

Step 4: Classification base on the decision from step 3.

Classification process shows on

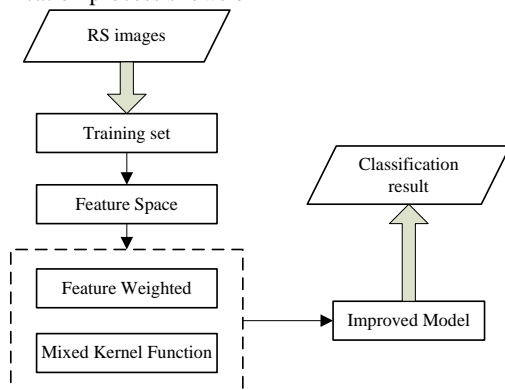


Figure 2. Classification process

4. ACCURACY

For comparison, four methods must be tested on the same data set on our experiment. The result shows on the Fig. 3.

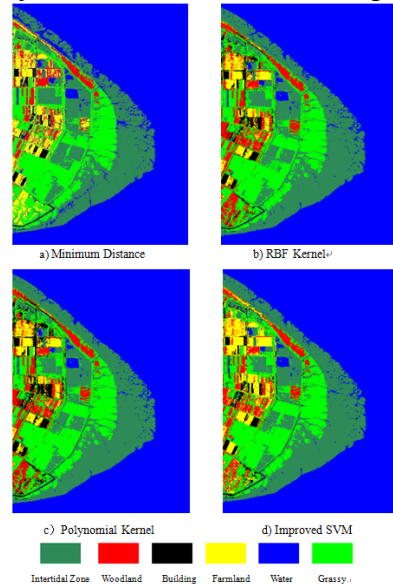


Figure 3. Classification Results

	intertidal zone	woodland	building	farmland	water	grassy	Omission errors (%)	User's accuracy (%)	
intertidal zone	79	0	0	0	2	0	0.00	97.53	
woodland	0	94	0	2	0	0	3.09	96.91	
building	0	0	53	2	0	1	3.64	96.36	
farmland	0	9	0	106	0	0	7.83	92.17	
water	2	0	0	0	106	0	0.00	98.14	
grassy	25	0	0	0	0	103	19.53	80.47	
Omission errors (%)	23.58	8.74	0.00	3.64	0.00	0.96			
Producer's Accuracy (%)	76.42	91.26	100.00	96.36	98.14	99.04			
Overall Accuracy = (545/584) 93.3219%		Kappa Coefficient = 0.9192							

Table 2. Confusion Matrix

	Overall Accuracy (%)	Kappa Coefficient (%)
Minimum Distance	78.4247	0.7404
RBF Kernel	93.1507	0.9173
Polynomial Kernel	92.9795	0.9152
Improved SVM	93.3219	0.9192

Table 3. Overall Accuracy and Kappa Coefficient of 4 methods

To assess the accuracy of an image classification, we create the confusion matrix on improved model (Table 2). Compared the overall accuracy of improved method to others' on tables (Table 3), we can conclude:

1) The accuracy of three different SVM classifiers is much better than the traditional minimum distance classification. The Feature weighted & Mixed kernels SVM proposed in the paper has the highest accuracy above 92% which closes but higher than the one using RBF kernel.

2) The study adds feature weight vector to mixed kernels of SVM which make more precise classification on study wetland, such as water area. Besides water information extraction, it can classify the pure texture objects such as grassy area.

3) The method has the advantage on small plaque from remote sensing image, like building lands.

The improved model combining RBF kernel function and polynomial kernel function advantage brings the higher accuracy sometime when data set is tested separately on each traditional means.

5. CONCLUSION

The improved SVM classifier focused on feature weight and kernel function which can avoid the domination of the weak correlated or uncorrelated feature; make up for shortage of traditional models. And tests on a series of images separately by improved SVM, polynomial kernel, RBF kernel and Minimum Distance show that the new model has better ability on extrapolation and interpolation. It has the advantage on fusing different kind of information. The higher values of overall accuracy, kappa coefficient, omission errors and commission errors compared to other methods' results also indicate its improvement on classification.

ACKNOWLEDGEMENTS

The objective of information extraction in study is to facilitate wetland conservation. And the accurate classification plays a critical role on further change detection. What we want to study more is about the fast and precise information extraction and monitoring from remote sensing images.

REFERENCES

- Mitra P., Shankar B. U., Pal S. K. 2004, Segmentation of multispectral remote sensing images using active support vector machines. *Pattern Recognition Letters*, 25(9),pp.1067-1074.
- Damoulas T., Girolami M.A. 2009. Combining feature spaces for classification. *Pattern recognition*, 42(11),pp.2671-2683.
- Smits, G.F. Jordaan, E.M. 2002, Improved SVM regression using mixtures of kernels. *Proceedings of the International Joint Conference on Neural Networks*, 3, pp.2785-2790.
- Lin C. F., Wan Sh D.. Fuzzy 2002, Support vector machines. *IEEE Transactions on Neural Networks*, 13(2), pp.464-471.
- Barbara J. Kent, Joy Nystrom Mast. 2005, Wetland change analysis of San Dieguito Lagoon, California, USA: 1928–1994. *Wetlands*, 25(3), pp.780-787.
- T Schmid, M Koch, J Gumuzzio, et al., 2004, A spectral library for a semi-arid wetland and its application to studies of wetland degradation using hyperspectral and multispectral data. *International Journal of Remote Sensing*, 25(13), pp.2485-2496.
- Jessika Tboyro, Alain Pietroniro. 2005, Towards operational monitoring of a northern wetland using geomatics-based techniques. *Remote Sensing of Environment*, 97(2), pp.174-191.
- Augusteijn MF, Warrender CE, 1998, Wetland classification using optical and neural network classification. *International Journal of Remote Sensing*, 19(8), pp. 1545-1560.
- R S Dwivedi, B R M Rao. 1999, Mapping wetlands of the Sundaban Delta and its environs using ERS—1 SAR data. *International Journal of Remote Sensing*, 20(11), pp.2235-2247.

Girosi F. 1998, An equivalence between sparse approximation and support vector machines. *Neural Computation*, 10(6), pp. 1455-1480.

ACKNOWLEDGEMENTS

This study is a part of internationally collaborative research project by Ministry of science and technology (Project ID: 2009DFA92310). The authors acknowledge the support of MOST, Shanghai Committee of Science and Technology (STCSM), Fudan University and Queen's University.