

COMPARING MACHINE LEARNING CLASSIFIERS FOR OBJECT-BASED LAND COVER CLASSIFICATION USING VERY HIGH RESOLUTION IMAGERY

Yuguo Qian^a, Weiqi Zhou^{a,*}, Jingli Yan^a, Weifeng Li^a, Lijian Han^a,

^a State Key Laboratory of Urban and Region Ecology, Research Centre for Eco-Environmental Sciences, Chinese Academy of Sciences – (ygqian, wzhou, li.wf, ljhan)@rcees.ac.cn, jlyan24@163.com

KEY WORDS: Object-based classification, Machine learning classifiers, Very high resolution image, Urban area, Tuning parameters

ABSTRACT:

This study evaluates and compares the performance of four machine learning classifiers—support vector machine (SVM), normal Bayes (NB), classification and regression tree (CART) and K nearest neighbour (KNN)—to classify very high resolution images, using an object-based classification procedure. In particular, we investigated how tuning parameters affect the classification accuracy with different training sample sizes. We found that: (1) SVM and NB were superior to CART and KNN, and both could achieve high classification accuracy (>90%); (2) the setting of tuning parameters greatly affected classification accuracy, particularly for the most commonly-used SVM classifier; the optimal values of tuning parameters might vary slightly with the size of training samples; (3) the size of training sample also greatly affected the classification accuracy, when the size of training sample was less than 125. Increasing the size of training samples generally led to the increase of classification accuracies for all four classifiers. In addition, NB and KNN were more sensitive to the sample sizes. This research provides insights into the selection of classifiers and the size of training samples. It also highlights the importance of the appropriate setting of tuning parameters for different machine learning classifiers and provides useful information for optimizing these parameters.

1. INTRODUCTION

Urban landscapes are extremely complex and heterogeneous. To adequately quantify the heterogeneity of urban land cover, high spatial resolution images are needed. A considerable amount of research has shown that object-based approaches are superior to traditional pixel-based methods in the classification of high spatial resolution data. Consequently, object-based approaches have been increasingly used for urban land cover classification (Duro, Franklin et al. 2012).

With object-based classification approaches, objects generated from image segmentation can be typically classified using a rule-based procedure (a set of rules) or using machine learning algorithms (MLA) based on training samples. While rule-based procedures, which use expert knowledge, have been increasingly used for classification, the majority of the studies have used supervised classifications (Laliberte, Rango et al. 2004; Mathieu and Aryal 2007). Many different kinds of MLA have been applied for supervised classifications. These algorithms are commonly categorized as parametric and non-parametric classifiers. The two widely-used types of parametric algorithms are the maximum likelihood classifier (MLC) and Bayes classifiers, and the frequently-used non-parametric classifiers include K nearest neighbour (KNN), decision tree (DT) and support vector machine (SVM).

Previous studies have shown that the use of different classifiers may lead to different classification results. Therefore, many studies have been conducted to investigate the effectiveness and efficiency of different classifiers (McInerney and Nieuwenhuis 2009; Song, Duan et al. 2012). However, these studies have been mostly conducted using pixel-based approaches. With the wide use of object-based approaches, there has been an

increasing interest in comparing different machine learning classifiers using object-based methods. When using these machine learning classifiers, we should consider at least four key factors that can dramatically affect the classification accuracy and efficiency. Specifically, these are image segmentation, training sample selection, feature selection and tuning parameter setting. While the first three factors have been investigated in many previous studies (Wieland and Pittore 2014), few studies have investigated the effects of the setting of tuning parameters. However, setting tuning parameters is the very first step, as well as one of the most important steps to appropriately use these machine learning classifiers. In addition, previous comparison studies of machine learning classifiers have been mostly focused on non-urban areas, such as grasslands, farmlands and coal mine area.

The overall objective of this study is to evaluate the four most frequently used MLAs for urban land cover classification, with an object-based approach, using very high spatial resolution imagery.

In particular, we aim to investigate how tuning parameters affect the classification results, especially with different training sample sizes. The four classifiers are: (1) normal Bayes (NB), a parametric algorithm; (2) SVM, a statistical learning algorithm; (3) KNN, an instance-based learning algorithm; and (4) the classification and regression tree (CART) classifier, a commonly-used DT algorithm. The results from this study can provide insights into classifier selection and parameter setting for high resolution urban land cover classification.

2. MAIN BODY

* Corresponding author

2.1 Study Area and Data

The study site is an urban area located in the Haidian District of Beijing, China, between latitudes $39^{\circ}58'30''$ and $40^{\circ}0'47''$ and longitudes $116^{\circ}17'55''$ and $116^{\circ}20'12''$. The study area is a complex urban area with many land use types, including parks, universities, construction sites and residential areas. Land cover types are mainly impervious surface, vegetation cover, bare soil and water, which are typical in urban areas. The dominant land cover in parks is vegetation and water, while in the universities and the residential areas, the primary land cover is impervious surfaces, mixed with dispersed small patches of greenspace. Bare soil is the dominant land cover type in construction sites (Figure 1)

We used WorldView-2 satellite imagery, acquired on 14 September 2012, for land cover classification. WorldView-2, launched in October 2009, is the first high resolution 8-band multispectral commercial satellite. Four land cover types were identified for the study area: (1) impervious surfaces; (2) vegetation; (3) water; and (4) bare soil. Impervious surfaces were mainly roads and building roofs. Vegetation included trees and grass. Water mostly occurred in parks and bare soil in construction sites. Shadows from buildings and trees are common in very high resolution images of urban areas. Therefore, we included the shadow class and separated shadows from unshaded land cover types.

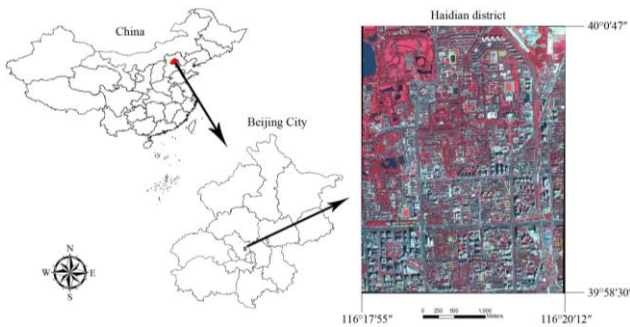


Figure 1. The study area, an urban area located in the Haidian District of Beijing, China.

2.2 Methods

The object-based classification procedure includes image segmentation, training sample selection, classification feature selection, tuning parameter setting and, finally, algorithm execution. We first segmented the image into land cover segments and then chose a certain amount of segments of different land cover types as training samples. After comparing the training sample characteristics of different land cover types, we selected certain object features for classification. Finally, we adjusted the tuning parameters of different classifiers to generate high classification accuracy. For all four classifiers, we used the same procedure for image segmentation and the selection of training samples and classification features. The optimal setting of tuning parameters, however, were determined separately for each classifier. Following the classifications, object-based accuracy assessment was applied to evaluate different classifiers.

Using Google Earth, we randomly chose 1500 object samples, 300 for each class, for the classifications and accuracy assessment. We then randomly divided the 1500 object samples into two sets: 1000 as training samples and 500 as testing

samples. To investigate the sensitivities of classifiers to the size of training samples, 8 training sample subsets were generated by randomly sampling from the total training sample set. The sizes of the training samples of those subsets were 125, 250, 375, 500, 625, 750, 875 and 1000, respectively. Within a training subset, the numbers of samples for each of the five classes were equal, and thus, the sample numbers of each class within the 8 training sample subsets were 25, 50, 75, 100, 125, 150, 175 and 200, respectively. Likewise, in the testing sample set, there were 100 samples per class.

For DT, we tested the value of “maximum depth” from 1 to 20. For SVM, we systematically tested 10 values for both C and gamma. Specifically, we tested the 10 values of C— 10^{-1} , 10^0 , 10^1 , 10^2 , 10^3 , 10^4 , 10^5 , 10^6 , 10^7 and 10^8 —and 10 values of gamma— 10^{-5} , 10^{-4} , 10^{-3} , 10^{-2} , 10^{-1} , 10^0 , 10^1 , 10^2 , 10^3 and 10^4 . For KNN, we examined K values from 1 to 20. All 8 training sample subsets have tested by these four classifier with different parameter settings.

For each of the 4 classifiers, we chose one thematic map with the highest overall classification accuracy for each of the 8 training sample subsets and then compared these 32 thematic maps based on overall accuracy, the kappa coefficient and the user’s and producer’s accuracy. In addition, we repeated the random selection of training and testing samples within the 1,500 samples 10 times and then compared average overall accuracy for all classifiers with the optimal parameter values.

2.3 Results and Discussion

The results showed that SVM generally had the best performance among the four classifiers (Table 3). With the optimal parameter setting, the minimum overall accuracy and kappa coefficient of SVM was 92.6% and 0.9075, which was higher than the maximum overall accuracy and kappa coefficient of DT (88.4% and 0.855) and KNN (86.8% and 0.835). Figure 2 shows the classification results with the highest overall accuracy for each classifier. Using the Z-statistics, we found that the overall accuracies of SVM were significantly greater than those of DT and KNN, regardless of the size of training samples. In addition, SVM had significantly higher overall accuracy than NB, when the size of training samples was relatively small. However, when training samples were greater than or equal to 100 per class, the classification accuracies of NB and SVM were similar and significantly higher than the accuracies of the other two classifiers.

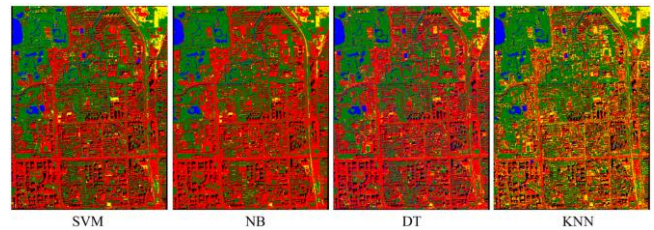


Figure 2. The classification results with the highest overall accuracy for each classifier.

In general, when the size of samples per class was less than 125, the accuracies of the four classifiers increased with increasing size of the training samples, and NB and KNN were more sensitive to sample sizes than SVM and DT. When the size of training samples increased from 25 to 125 per class, the classification accuracies of NB and KNN increased by 17% and

8.2%, while SVM and DT increased by 3.6% and 4.6%, respectively (Table 1; Figure 3). NB was the most sensitive to sample size. This may be because this parametric classifier used training samples to estimate parameter values for the data distribution, and thus, more training samples can lead to more accurate parameter estimation. In contrast, SVM is the least sensitive to sample sizes, because SVM only uses the support vectors instead of all training samples to build the separating hyperplane. Thus, adding more training samples may not significantly affect the classification accuracy. However, when the size of the samples is more than 125 per class, all four classifiers become insensitive to the increase of sample sizes. The classification accuracies of NB, KNN, SVM and DT fluctuated from 95% to 96.4%, 85.2% to 86.8%, 96.2% to 97.6% and 87.2% to 87.6%, respectively, with the sample size increasing from 125 to 200 per class. This result indicated that the training sample size of 125 per class might be a turning point, beyond which the increase of sample size does not necessarily lead to a significant increase in classification accuracies. These results have important implications for determining the appropriate sample size.

Table 1. The highest overall accuracy for the four classifiers using eight training sample sets and the corresponding parameter values. NB, normal Bayes; DT, decision tree.

Sample Size	C	Gamma	Accuracies	Kappa	Sample Size	Max Depth	Accuracies	Kappa
SVM					DT			
25	10,000	0.001	0.926	0.9075	25	3	0.838	0.7975
50	1,000,000	0.00001	0.94	0.925	50	5	0.866	0.8325
75	1,000,000	0.00001	0.958	0.9475	75	12	0.884	0.855
100	1,000,000	0.00001	0.944	0.93	100	6	0.884	0.855
125	10,000	0.0001	0.962	0.9525	125	8	0.874	0.8425
150	10,000	0.0001	0.964	0.955	150	5	0.876	0.845
175	1,000,000	0.00001	0.976	0.97	175	5	0.876	0.845
200	1,000,000	0.00001	0.964	0.955	200	6	0.872	0.84
NB					KNN			
25			0.794	0.7425	25	1	0.77	0.7125
50			0.836	0.795	50	1	0.796	0.745
75			0.912	0.89	75	3	0.81	0.7625
100			0.936	0.92	100	3	0.828	0.785
125			0.964	0.955	125	1	0.852	0.815
150			0.95	0.9375	150	1	0.852	0.815
175			0.958	0.9475	175	3	0.856	0.82
200			0.964	0.955	200	3	0.868	0.835

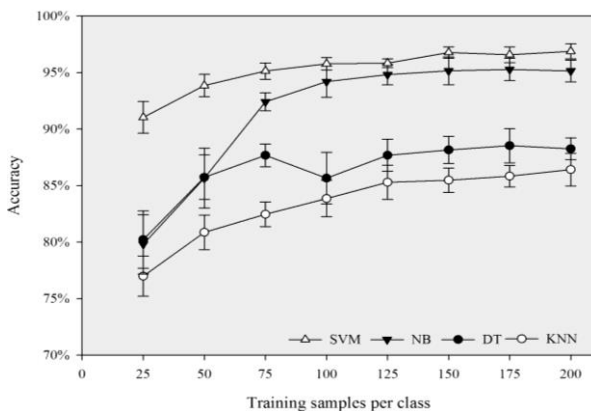


Figure 3. Overall accuracies of the four classifiers with increasing size of training samples.

3. CONCLUSION

SVM and NB were superior to CART and KNN in urban land classification. Both SVM and NB could achieve very high classification accuracy, with appropriate setting of the tuning parameters and/or enough training samples. However, each of the two classifiers has its advantages and disadvantages, and thus, the choice of the appropriate one may be case dependent. SVM could achieve relatively high accuracy with a relatively small amount of training samples, but the setting of tuning parameters could be subjective and time-consuming. In contrast, NB does not need the setting of any tuning parameter, but generally requires a large amount of training samples to achieve relatively high accuracy.

Both the size of training samples and the setting of tuning parameters have great impacts on the performance of classifiers. When the size of training samples is less than 125 per class, increasing the size of training samples generally leads to the increase of classification accuracies for all four classifiers, but NB and KNN were more sensitive to the sample size. Increasing the size of training samples does not seem to significantly improve the classification once the size of training samples reaches 125 per class. The tuning parameters of the classifier had a great impact on the classification accuracy. While the optimal settings of tuning parameters varied with the size of training samples, some general patterns occurred. For SVM, setting C between 1,000,000 and 100,000,000, and gamma between 0.00001 and 0.001 usually achieved the best overall accuracy. With DT, the best classification accuracy was generally achieved when the max depth of classification tree was between five and eight. For KNN, the optimal K value was either one or three. These findings provide insights into the selection of classifiers and the size of training samples when implementing an object-based approach for urban land classification using high resolution images.

REFERENCES

- Duro, D. C., S. E. Franklin and M. G. Dube (2012). "A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery." *Remote Sensing of Environment* **118**: 259-272.
- Laliberte, A. S., A. Rango, K. M. Havstad, J. F. Paris, R. F. Beck, R. McNeely and A. L. Gonzalez (2004). "Object-oriented image analysis for mapping shrub encroachment from 1937 to 2003 in southern New Mexico." *Remote Sensing of Environment* **93**(1-2): 198-210.
- Mathieu, R. and J. Aryal (2007). "Object-based classification of Ikonos imagery for mapping large-scale vegetation communities in urban areas." *Sensors* **7**(11): 2860-2880.
- McInerney, D. O. and M. Nieuwenhuis (2009). "A comparative analysis of kNN and decision tree methods for the Irish National Forest Inventory." *International Journal of Remote Sensing* **30**(19): 4937-4955.
- Song, X., Z. Duan and X. Jiang (2012). "Comparison of artificial neural networks and support vector machine classifiers for land cover classification in Northern China using a SPOT-5 HRG image." *International Journal of Remote Sensing* **33**(10): 3301-3320.
- Wieland, M. and M. Pittore (2014). "Performance Evaluation of Machine Learning Algorithms for Urban Pattern Recognition from Multi-spectral Satellite Images." *Remote Sensing* **6**(4): 2912-2939.