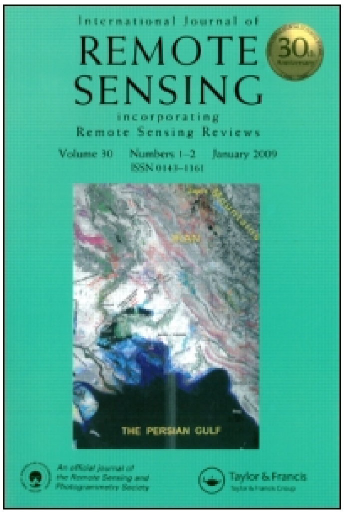


International Journal of Remote Sensing
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713722504

A maximum entropy approach to one-class classification of remote sensing imagery

Wenkai Li [a];Qinghua Guo [a]
[a] Sierra Nevada Research Institute, School of Engineering, University of California at Merced, California, USA

Online publication date: 28 April 2010

To cite this Article Li, Wenkai andGuo, Qinghua(2010) 'A maximum entropy approach to one-class classification of remote sensing imagery', International Journal of Remote Sensing, 31: 8, 2227 — 2235

To link to this Article: DOI: 10.1080/01431161003702245

URL: http://dx.doi.org/10.1080/01431161003702245

International Journal of Remote Sensing
Vol. 31, No. 8, 20 April 2010, 2227–2235

Taylor & Francis
Taylor & Francis Group

A maximum entropy approach to one-class classification of remote sensing imagery

WENKAI LI and QINGHUA GUO*
Sierra Nevada Research Institute, School of Engineering, University of California at Merced, California, USA

(*Received 11 November 2009; in final form 12 February 2010*)

In remote sensing classification there are situations when users are only interested in classifying one specific land type without considering other classes, which is referred to as one-class classification. Traditional supervised learning requires all classes that occur in the image to be exhaustively labelled and hence is inefficient for one-class classification. In this study we investigate a maximum entropy approach (MAXENT) to one-class classification of remote sensing imagery, i.e. classifying a single land class (e.g. urban areas, trees, grasses and soils) from an aerial photograph with 0.3 m spatial resolution. MAXENT estimates the Gibbs probability distribution that is proportional to the conditional probability of being positive. A threshold for generating binary predictions can be determined based on the omission rate of a validation set. The results indicate that MAXENT provides higher classification accuracy than the one-class support vector machine (OCSVM). MAXENT does not require other land classes for training. Its input is only a set of training samples of the specific land class of interest, as well as a set of known constraints on the distribution. Therefore, the effort of manually collecting training data for classification can be significantly reduced.

1. Introduction

In remote sensing applications, users are sometimes only interested in classifying one specific land type, without considering other classes (Byeungwoo and Landgrebe 1999, Foody *et al.* 2006). The classifiers seek to extract a specific land type from an image, given only the training sample of the class of interest. We refer to the specific land type of interest as positive and other land types as negative data. Traditional supervised classifiers require the availability of exhaustively labelled training sets for properly training the classification algorithms (Munoz-Marf *et al.* 2007). A non-exhaustively defined set of classes can cause major problems for traditional supervised classifications such as neural networks (Foody 2000, 2004a). However, manually collecting training data is labour-intensive and time-consuming (Foody *et al.* 2006), particularly when high spatial resolution images are used. Therefore, traditional supervised classifiers are inefficient in one-class classification, and it is necessary to develop classifiers to discriminate the single class of interest from the other classes without negative training data.

In the literature, one-class support vector machines (OCSVM) proposed by Schölkopf *et al.* (2001) have proved useful in dealing with the one-class classification

*Corresponding author. Email: qguo@ucmerced.edu

International Journal of Remote Sensing
ISSN 0143-1161 print/ISSN 1366-5901 online
http://www.tandf.co.uk/journals
DOI: 10.1080/01431161003702245

problem in various fields such as document classification (Manevitz and Yousef 2001), texture segmentation (Tax and Duin 2002) and ecological modelling (Guo *et al.* 2005). Recently, it has also been proposed in one-class classification of remote sensing imagery and has shown good results in some research (Foody *et al.* 2006, Munoz-Marf *et al.* 2007, Sanchez-Hernandez *et al.* 2007a,b). However, its output is sensitive to parameters that are difficult to tune (Manevitz and Yousef 2001).

The one-class classification problem is also quite common in ecological niche modelling (Elith *et al.* 2006). It is often the case that only presence data are available, but information about species absence is difficult to obtain or is unreliable (Guo *et al.* 2005). To address this problem, Phillips *et al.* (2004) proposed the application of maximum-entropy (MAXENT) techniques to model the species geographic distributions based only on presence data. One major advantage of MAXENT is that it only requires positive data for training.

Techniques proposed to address the one-class classification problem in ecological niche modelling can also contribute to one-class classification in remote sensing. MAXENT is the current state-of-the-art method of modelling species distributions from only presence data (Phillips *et al.* 2004, Elith *et al.* 2006). However, to our knowledge, its applications in remote sensing classification are rarely studied. Therefore, we propose a maximum entropy approach to one-class classification of remote sensing imagery in this study. To evaluate the classification accuracy of the new approach, it was applied to classify the urban areas, trees, grasses and soils from a high spatial resolution image. The results were also compared to the commonly used OCSVM. The specific objective of this study is to evaluate the effectiveness of the maximum-entropy approach to one-class classification of remote sensing imagery.

2. Methods

2.1 *Maximum entropy (MAXENT)*

Entropy is a fundamental concept in information theory; it measures how much choice is involved in the selection of an event (Shannon 1948). The principle of maximum entropy indicates that the distribution model that satisfies any given constraints should be as uniform as possible (Phillips *et al*. 2004). This agrees with everything that is known, but carefully avoids assuming anything that is not known (Jaynes 1990). In this study, the MAXENT algorithm first proposed by Jaynes (1957) has already been modified and is ready to handle GIS raster data in the classification processes (Phillips *et al*. 2004).

The unknown probability distribution π is over a finite set X (the set of pixels in the study area); the elements of X are individual pixels x. The distribution π assigns a non-negative probability $\pi(x)$ to each pixel x, and these probabilities sum to one. The constraints on the unknown probability distribution π are represented by a set of features (real-valued functions) $f_1, \ldots, f_n$ on X. The information we know about π is the expectations (averages) of each feature f_j under π, which is defined as $\pi[f_j] = \sum_{x \in X} \pi(x) f_j(x)$.

A set of sample pixels $x_1, \ldots, x_m$ is drawn independently from X. The corresponding empirical distribution is denoted as:

$$\widetilde{\pi}(x) = \frac{|\{1 \leq i \leq m : x_i = x\}|}{m} \tag{1}$$

The empirical average of f_j under $\tilde{\pi}$ is defined as

$$\tilde{\pi}[f_j] = \frac{1}{m}\sum_{i=1}^{m} f_j(x_i) \tag{2}$$

We use $\tilde{\pi}[f_j]$ as an estimate of $\pi[f_j]$. The goal is to seek the probability distribution $\hat{\pi}$, an approximation of π, subject to the constraint that the expectation of each feature f_j under $\hat{\pi}$ is the same as its empirical average, stated formally as:

$$\hat{\pi}[f_j] = \tilde{\pi}[f_j] \tag{3}$$

There are many distributions satisfying these constraints. The maximum-entropy principle suggests that, among all distributions satisfying these constraints, we choose the one that has maximum entropy. The entropy of $\hat{\pi}$ is defined as:

$$H(\hat{\pi}) = -\sum_{x \in X} \hat{\pi}(x) \ln \hat{\pi}(x) \tag{4}$$

where ln is the natural logarithm.

The convex duality (Della Pietra *et al.* 1997) shows that the MAXENT probability distribution $\hat{\pi}$ is exactly equal to the Gibbs probability distribution that maximizes the likelihood of the m sample points, and the estimated Gibbs probability distribution is proportional to the conditional probability of being positive. More details about MAXENT can be found in Phillips *et al.* (2004, 2006).

2.2 *Application of MAXENT in one-class classification*

We applied MAXENT to classify a specific land type from an aerial photograph with 0.3 m spatial resolution. The data were acquired in 2004 by a Leica ADS40 digital camera. Three bands are available in the image: red (610–660 nm), green (535–585 nm) and blue (430–490 nm). The study area, which includes houses, roads, trees, grasses and soils, is located in the city of El Cerrito, California (figure 1). Here we define the extraction of urban areas (including houses and roads), trees, grasses and soils as four different examples of one-class classification separately. We assume that only positive data are available for training, but both positive and negative data are available for testing. Hence, for each land type extraction, we randomly selected 3000 pixels from the aerial photograph, 2000 positive (the class of interest) and 1000 negative (other classes). The training set included 1000 positive pixels, whereas the testing set included 1000 positive and 1000 negative pixels. Both the training and testing data were obtained by manual interpretation. We then extracted 15 features for the image classification, including mean value, variance, homogeneity, contrast and second moment of the R, G and B bands. All features were calculated in ENVI software with a 3×3 pixel template, and then rescaled into the range [0,1].

We used the 'MAXENT' software that is freely available online at http://www.cs.princeton.edu/~schapire/maxent. The inputs are the locations of the positive training pixels and their corresponding feature values. We used the default output logistic that gives an estimate of probability between 0 and 1. Note the output is not the exact probability of being positive, but it is proportional to the conditional probability of being positive (Phillips *et al.* 2006). Therefore, a threshold is required to convert the probabilistic output to binary predictions. To avoid over-fitting, we randomly set aside 25% of the training set for validation. Normally, one should choose a threshold that maximizes the classification accuracy of the validation set.

Figure 1. Aerial photograph of El Cerrito, California (0.3 m resolution, 500×500 m).

However, because the validation set consists of only positive data, this approach can lead to over-predictions, particularly when outliers exist. To avoid over-predictions, we allow for a small omission rate to account for outliers in the validation set. In this study, we chose the logistic value corresponding to a 5% omission rate for the validation set as the threshold to make a binary prediction (Pearson *et al.* 2004). Other user-specified parameters were set to their default values: convergence threshold = 10^{-5}, maximum iterations = 500, regularization multiplier = 1 and maximum number of background points = 5000.

To evaluate the classification accuracy of MAXENT, we also compared it to OCSVM, a current state-of-the-art method in one-class classification of remote sensing (Foody *et al.* 2006, Munoz-Marf *et al.* 2007, Sanchez-Hernandez *et al.* 2007a,b). OCSVM also requires only positive data for training. It tries to find a hypersphere as small as possible to contain the training points in a multi-dimensional space. More detailed mathematical derivations of one-class SVMs can be found in Schölkopf *et al.* (2001). We implemented OCSVM by LIBSVM – a library for support vector machines developed by Chang and Lin (2001). The Gaussian RBF kernel function was used. We tuned the RBF kernel width in the range [0,1000] and the rejection fraction in the range [0,1] through 10-fold cross-validation.

Finally, classification results were evaluated using the confusion matrix, producer's accuracy (Pro), user's accuracy (Usr), overall accuracy (OA) and kappa coefficient (κ). Since these accuracy statements were derived from the same training samples, we used the McNemar's test to assess the statistical significance of differences in classification accuracy (Foody 2004b). This test is based on the Z statistic:

$$Z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \tag{5}$$

where f_{12} indicates the frequency of pixels correctly predicted by classifier one but incorrectly predicted by classifier two, whereas f_{21} is the frequency of pixels incorrectly predicted by classifier one but correctly predicted by classifier two (Foody 2004b). Two classifications differ at the 95% level of confidence if $|Z| > 1.96$ (Foody *et al.* 2006, Sanchez-Hernandez *et al.* 2007b).

3. Results

Figure 2 shows the prediction maps for each land type, and tables 1–4 show the comparison of classification accuracies for each land type between MAXENT and OCSVM. In general, MAXENT provides relatively higher accuracy in the extraction of a single land type from the aerial photograph. The prediction maps for each land type have good agreement with the original aerial photograph, especially for the

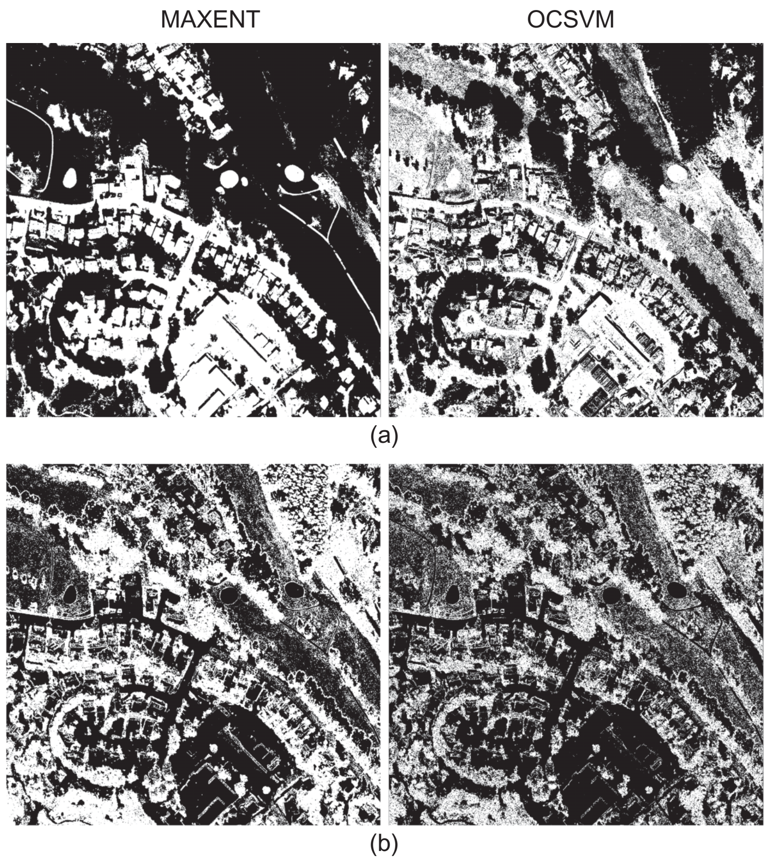
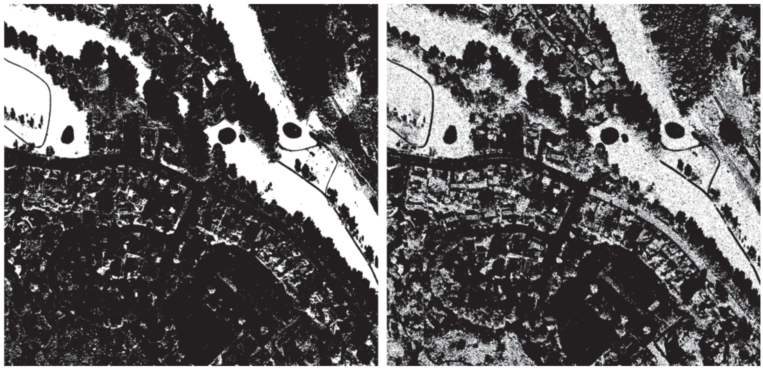
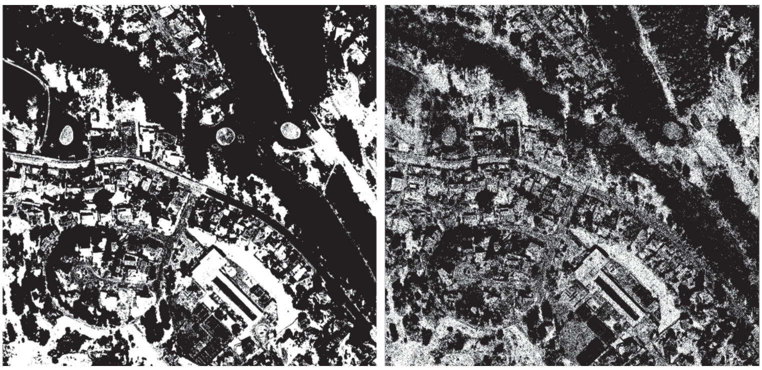


Figure 2. Prediction maps for each land type. (*a*) Urban, (*b*) tree, (*c*) grass, and (*d*) soil. White: positive; black: negative.



(c)



(d)

Figure 2. (*Continued.*)

Table 1. Confusion matrices and accuracy assessment of urban.

	Prediction			
	MAXENT		OCSVM	
Reference	Positive	Negative	Positive	Negative
Positive	957	43	796	204
Negative	141	859	348	652
Pro (%)	95.70	85.90	79.60	65.20
Usr (%)	87.16	95.23	69.58	76.17
OA (%)	90.80		72.40	
κ	0.82		0.45	
Z	17.16			

urban areas and grasses. The classification accuracies of urban areas and grasses are relatively high, with overall accuracies of 90.80% and 92.65%, and kappa coefficients of 0.82 and 0.85, respectively. Compared to urban areas and grasses, more negative pixels are classified as positive for the types of trees and soils, and their classification accuracies are relatively lower, with overall accuracies of 82.80% and 86.65%, and kappa coefficients of 0.66 and 0.73, respectively.

Table 2. Confusion matrices and accuracy assessment of tree.

	Prediction			
	MAXENT		OCSVM	
Reference	Positive	Negative	Positive	Negative
Positive	958	42	754	246
Negative	302	698	216	784
Pro (%)	95.80	69.80	75.40	78.40
Usr (%)	76.03	94.32	77.73	76.12
OA (%)	82.80		76.90	
κ	0.66		0.54	
Z	5.91			

Table 3. Confusion matrices and accuracy assessment of grass.

	Prediction			
	MAXENT		OCSVM	
Reference	Positive	Negative	Positive	Negative
Positive	925	75	849	151
Negative	72	928	203	797
Pro (%)	92.50	92.80	84.90	79.70
Usr (%)	92.78	92.52	80.70	84.07
OA (%)	92.65		82.30	
κ	0.85		0.65	
Z	11.81			

Table 4. Confusion matrices and accuracy assessment of soil.

	Prediction			
	MAXENT		OCSVM	
Reference	Positive	Negative	Positive	Negative
Positive	959	41	725	275
Negative	226	774	194	806
Pro (%)	95.90	77.40	72.50	80.60
Usr (%)	80.93	94.97	78.89	74.56
OA (%)	86.65		76.55	
κ	0.73		0.53	
Z	9.67			

The results indicate that the classification accuracy of OCSVM was relatively low in this experiment. Compared with MAXENT, the predicted positive areas of each land type are larger, and the prediction maps show more 'salt and pepper' effect. Many pixels are misclassified, with relatively high rates of both false positives and false negatives for each land type. The classification accuracies of each land type, in terms of overall accuracy and kappa coefficient, are much lower than those of MAXENT. As with MAXENT, the accuracy of grasses is the highest of all land types, with an overall accuracy of 82.30% and a kappa coefficient of 0.65, but these are still much lower than those of MAXENT. According to the McNemar's test (tables 1–4), the differences in the classification accuracies are statistically significant at the 95% level of confidence ($|Z| > 1.96$), which indicates that MAXENT produced higher accuracy than OCSVM in this study.

4. Conclusion

In this study, the proposed MAXENT shows promise in one-class classification of remote sensing imagery. It provides better classification accuracy than OCSVM. One major advantage of MAXENT is that it does not require negative data for training. The input to MAXENT is only a set of positive samples from a target distribution, as well as a set of known constraints on the distribution. Hence, it can significantly reduce the effort of manually collecting training data for classification.

Acknowledgements

We thank Jinqing He and Otto Alvarez for their help with image classification and programming. We also thank the reviewers for their constructive comments that helped strengthen the paper.

References

Byeungwoo, J. and Landgrebe, D.A., 1999, Partially supervised classification using weighted unsupervised clustering. *IEEE Transactions on Geoscience and Remote Sensing*, **37**, pp. 1073–1079.

Chang, C.-C. and Lin, C.-J., 2001, LIBSVM: a library for support vector machines. Available online at: http://www.csie.ntu.edu.tw/~cjlin/libsvm/.

Della Pietra, S., Della Pietra, V. and Lafferty, J., 1997, Inducing features of random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **19**, pp. 1–13.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S. and Zimmermann, N.E., 2006, Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, pp. 129–151.

Foody, G.M., 2000, Estimation of sub-pixel land cover composition in the presence of untrained classes. *Computers & Geosciences*, **26**, pp. 469–478.

Foody, G.M., 2004a, Supervised image classification by MLP and RBF neural networks with and without an exhaustively defined set of classes. *International Journal of Remote Sensing*, **25**, pp. 3091–3104.

Foody, G.M., 2004b, Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. *Photogrammetric Engineering & Remote Sensing*, **70**, pp. 627–633.

Foody, G.M., Mathur, A., Sanchez-Hernandez, C. and Boyd, D.S., 2006, Training set size requirements for the classification of a specific class. *Remote Sensing of Environment*, **104**, pp. 1–14.

Guo, Q., Kelly, M. and Graham, C.H., 2005, Support vector machines for predicting distribution of Sudden Oak Death in California. *Ecological Modelling*, **182**, pp. 75–90.

JAYNES, E.T., 1957, Information theory and statistical mechanics. *Physical Review*, **106**, pp. 620–630.

JAYNES, E.T., 1990, Notes on present status and future prospects. In *Maximum Entropy and Bayesian Methods*, W.T. Grandy and L.H. Schick (Eds.) (Dordrecht: Kluwer).

MANEVITZ, L.M. and YOUSEF, M., 2001, One-class SVMs for document classification. *Journal of Machine Learning Research*, **2**, pp. 139–154.

MUNOZ-MARF, J., BRUZZONE, L. and CAMPS-VAILS, G., 2007, A support vector domain description approach to supervised classification of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, **45**, pp. 2683–2692.

PEARSON, R.G., DAWSON, T.P. and LIU, C., 2004, Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography*, **27**, pp. 285–298.

PHILLIPS, S.J., ANDERSON, R.P. and SCHAPIRE, R.E., 2006, Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, pp. 231–259.

PHILLIPS, S.J., DUDÍK, M. and SCHAPIRE, R.E., 2004, A maximum entropy approach to species distribution modeling. In *Proceedings of the Twenty-First International Conference on Machine Learning*, Banff, Alberta, Canada. 1015412, pp. 655–662 (New York, NY: ACM).

SANCHEZ-HERNANDEZ, C., BOYD, D.S. and FOODY, G.M., 2007a, Mapping specific habitats from remotely sensed imagery: support vector machine and support vector data description based classification of coastal salt marsh habitats. *Ecological Informatics*, **2**, pp. 83–88.

SANCHEZ-HERNANDEZ, C., BOYD, D.S. and FOODY, G.M., 2007b, One-class classification for mapping a specific land-cover class: SVDD classification of Fenland. *IEEE Transactions on Geoscience and Remote Sensing*, **45**, pp. 1061–1073.

SCHÖLKOPF, B., PLATT, J.C., SHAWE-TAYLOR, J., SMOLA, A.J. and WILLIAMSON, R.C., 2001, Estimating the support of a high-dimensional distribution. *Neural Computation*, **13**, pp. 1443–1471.

SHANNON, C.E., 1948, A mathematical theory of communication. *The Bell System Technical Journal*, **27**, pp. 379–423, 623–656.

TAX, D.M.J. and DUIN, R.P.W., 2002, Uniform object generation for optimizing one-class classifiers. *Journal of Machine Learning Research*, **2**, pp. 155–173.