RVM-based multi-class classification of remotely sensed data

Foody, G. M.

The manuscript of the above article revised after peer review and submitted to the journal for publication, follows. Please note that small changes may have been made after submission and the definitive version is that subsequently published as:

RVM-based multi-class classification of remotely sensed data

G. M. Foody
School of Geography
University of Nottingham
University Park
Nottingham
NG7 2RD
UK

Abstract

The relevance vector machine (RVM), a Bayesian extension of the support vector machine (SVM), has considerable potential for the analysis of remotely sensed data. Here, the RVM is introduced and used to derive a multi-class classification of land cover with an accuracy of 91.25%, a level comparable to that achieved by a suite of popular image classifiers including the SVM. Critically, however, the output of the RVM includes an estimate of the posterior probability of class membership. This output may be used to illustrate the uncertainty of the class allocations on a per-case basis and help to identify possible routes to further enhance classification accuracy.

Keywords: image classification, relevance vector machine, posterior probability, uncertainty.

1. Introduction

Supervised classification is one of the most commonly undertaken analyses of remotely

sensed data. Despite the importance and long history of classification analysis within

remote sensing, the accuracy of classifications is often viewed negatively (Wilkinson,

2005). A variety of factors may be responsible for the low classification accuracies

sometimes achieved (Foody 2002). Considerable research has been directed at addressing

the various factors that may limit classification accuracy. Much of this research has

focused on the potential of new classifiers to accurately discriminate between classes.

The limitations of conventional and widely used parametric classifiers such as maximum

likelihood classification have been recognised and the potential of alternative approaches

evaluated (Foody and Mathur, 2004a). Recently, considerable attention has focused on

support vector machine (SVM)-based classification (e.g. Huang *et al*., 2002; Pal and

Mather, 2005; Bazi and Melgani, 2006). The SVM-based approach to classification has

many advantages over other approaches, notably a relative insensitivity to the

dimensionality of the data set (Melgani and Bruzzone, 2004; Pal and Mather, 2004), a

potential for accurate classification from small training sets (Foody and Mathur, 2004b;

Foody *et al*., 2006; Mathur and Foody, 2007) and, by focusing on maximising the margin

between classes, an avoidance of over-fitting problems (Chen and Tang, 2005). Although

originally designed for binary classifications the SVM approach may be used for multi-

class classification. The latter typically involves either breaking the multi-class problem

down into a series of binary analyses which can be addressed using a basic SVM (Huang

et al., 2002) or the adoption of a multi-class SVM (Hsu and Lin, 2002). Critically, multi-class classifications by SVM have often been found to be more accurate than those derived from a suite of popular alternative classifiers used in remote sensing (Huang *et al*. 2002, Foody and Mathur, 2004a).

Despite the current popularity of SVM-based classification there are some concerns with its use. The analyst must, for example, select appropriate values for the penalty term C and kernel specific parameter (e.g. gamma which controls the width of the widely used radial basis function kernel), often via cross-validation exercise that is wasteful of computational time and data (Tipping, 2001). The kernel function used must also satisfy Mercer's condition and the output of the analysis is just a class label prediction, conveying no information on the uncertainty of the class allocations predicted (Tipping, 2001; Chen and Tang, 2005). Additionally, the realisation of potential advantages such as the ability to use small training for accurate classification requires an ability to identify useful training sites in advance of the analysis (Foody and Mathur, 2004b; Mathur and Foody, 2007). In some circumstances it may be possible to address the concerns with SVM-based analyses. For example, it is possible to post-process the outputs of a SVM-based analysis to derive estimates of posterior probabilities but the reliability of this type of analysis can be questionable (Tipping, 2001). Alternative approaches to classification are, however, also worth exploring. A recent development of the SVM, the relevance vector machine (RVM), may offer an attractive alternative for image classification applications. The RVM is a Bayesian extension of the SVM. Key attractions of the RVM relative to the SVM are the removal of the need to define the parameter C, a reduced

sensitivity to the hyperparameter settings, an ability to use non-Mercer kernels, the provision of a probabilistic output and a typical requirement for considerably fewer basis functions (relevance vectors) for a given analysis (Tipping, 2001; Chen and Tang, 2005).

As with the SVM, the RVM was originally developed for binary applications. There are, however, extensions of the basic approach that may be used for multi-class classification. It is, for example, possible to undertake a one-against-all strategy in a manner similar to that used with binary SVMs or adopt a multi-class approach (Tipping, 2001; Zhang and Malik, 2005).

Like the SVM, the RVM may be used for regression and classification problems. Although both regression and classification problems are widely encountered in remote sensing the RVM has been very rarely used. Indeed, a search of the *ISI Web of Science* (on 21 May 2007) revealed only one previous publication using RVM in the remote sensing literature, and this was as a regression tool (Camps-Valls *et al*., 2006) although an embryonic literature base is emerging (e.g. Demir and Erturk, 2007). The aim of this article is to evaluate the potential of the RVM-based approach for multi-class classification.

2. RVM

The RVM was introduced in Tipping (2001), which includes a detailed discussion on the underlying mathematical basis of the technique. Further details and examples of its

application may be found in the literature (e.g. Bowd *et al*., 2005; Chen and Tang, 2005; Camps-Valls *et al*., 2006). This section aims to provide only a brief discussion focused on the salient features for a multi-class classification.

Like the SVM, the RVM was originally developed for binary analyses. In a two class classification by RVM the aim is, essentially, to predict the posterior probability of membership for one of the classes (0 or 1) for a given input **x**. A case may then be allocated to the class with which it has the greatest likelihood of membership. The basis of the RVM may be illustrated following Tipping's (2001) discussion. Using a Bernoulli distribution for $P(t|\mathbf{x})$ the likelihood function in the analysis is,

$$P(\mathbf{t}|\mathbf{w}) = \prod_{n=1}^{N} \sigma\{y(\mathbf{x}_n;\mathbf{w})\}^{t_n}[1 - \sigma\{y(\mathbf{x}_n;\mathbf{w})\}]^{1-t_n} \qquad (1)$$

where t defines the class labels, **w** are a set of adjustable weights, $\{\mathbf{x}_n\}_{n=1}^{N}$ is the set of training cases and σ_y is the logit function $\frac{1}{(1+e^{-y})}$. An iterative analysis is then followed to find the set of weights that maximises the function in which the hyperparameters, α, associated with each weight are up-dated. When completed, the set of non-zero weights defines the relevance vectors. The approach may be extended to multi-class classification by generalising (1) to the multinominal form:

$$P(\mathbf{t}|\mathbf{w}) = \prod_{n=1}^{N} \prod_{k=1}^{K} \sigma\{y_k(\mathbf{x}_n;\mathbf{w}_k)\}^{t_{nk}} \qquad (2)$$

where K is the number of classes, t_{nk} is the indicator variable for case n to be a member of class k and y_k is the predictor for class k (Tipping, 2001; Zhang and Malik, 2005). Class allocation may then be achieved following the one-against-all strategy sometimes used to derive multi-class SVM-based classification. A concern here is that the multi-class classification will require a series of binary classifications to be undertaken. An alternative based on the principles of multinominal logistic regression and in which y_k is not considered independently for each class is based on:

$$P(\mathbf{t}|\mathbf{w})=\prod_{n=1}^{N}\prod_{k=1}^{K}\sigma\{y_k;y_1,y_2,..,y_K\}^{t_{nk}} \tag{3}$$

where the predictors for each class y_k are coupled in the multinominal logit function (or softmax), $\sigma(y_k;y_1,...,y_K)=\dfrac{e^{y_k}}{(e^{y_1}+..+e^{y_K})}$ (Zhang and Malik, 2005). This approach forms the basis of the M-RVM software which may be used for classification with class-specific features. This software requires the specification of the priors associated with the hyperparameters, α, which have a Gamma distribution.

3. Data and methods

Remotely sensed data acquired by a Daedalus 1268 airborne thematic mapper (ATM) with a spatial resolution of ~5 m were used to classify crop types at an agricultural test site. To facilitate comparison against earlier work with this data set (e.g. Foody and

7

Mathur, 2004a) only the data acquired in three spectral wavebands that provided a high degree of class separability were used. These wavebands were located at 0.60-0.63, 0.69-0.75 and 1.55-1.75 μm.

The test site was the region of agricultural land adjacent to the village of Feltwell in Eastern England. At this site, the large fields that dominated the landscape were planted to a single crop. A map depicting the crop type planted in each field produced near the time of ATM data acquisition was used as ground data. Attention was focused on a region comprising mainly six classes: sugar beet, wheat, barley, carrot, potato and grass.

A training set comprising 100 randomly selected pixels of each class was obtained for the classification analysis. A further independent testing set was acquired for the purpose of accuracy assessment. This testing set comprised 320 randomly selected pixels. As a consequence of the sample design, the number of cases of each class in this testing set reflected the relative abundance of the classes at the time of data acquisition. The training and testing sets were the same as those used in an earlier comparison of contemporary image classification techniques (Foody and Mathur, 2004a). The use of these training and testing sets, therefore, facilitated the evaluation of the RVM-based approach relative to the classification approaches evaluated earlier: discriminant analysis, decision tree, neural network and multi-class SVM. Here, the RVM-based approach was evaluated for multi-class classification. Following, the literature (Tipping, 2001; Zhang and Malik, 2005) the priors on the hyperparameters were set to 0.

Classification accuracy was assessed with the aid of a confusion matrix and expressed as the percentage of the testing cases correctly allocated. An assessment of the statistical significance of the difference in accuracy achieved by different classifiers was achieved using a McNemar test in recognition of the use of the same testing set in their evaluation (Foody, 2004).

4. Results and discussion

Of the 320 cases in the testing set, all except 28 were correctly classified (Table 1). The overall accuracy of the RVM-based classification was, therefore, 91.25%. Relative to results of earlier work with the same data set (Foody and Mathur, 2004a), this level of accuracy is larger than that achieved by classification with a discriminant analysis (90.00%) and decision tree (90.31%); the differences in accuracy were insignificant at the 95% level of confidence. The accuracy of the RVM-based classification was also only marginally, but insignificantly at the 95% level of confidence, below the accuracy of classification by a neural network (91.88%) and a multi-class SVM (93.75%). It was evident, therefore, that the RVM approach yielded a classification of high accuracy, comparable to that from a range of popular classifiers. In particular, the accuracy differed insignificantly from that of a SVM-based classification but was derived without the aforementioned limitations of such an analysis.

The probabilistic nature of the RVM-based classification output may be of considerable value. For example, the probabilistic output is valuable in providing an index of the

9

uncertainty in class allocation on a per-case basis. This feature has been found to be useful with other classifiers, notably as a means of providing a spatial representation of the uncertainty in class allocation, an important feature of classification quality. The potential value of this output from the RVM-based approach is indicated by Table 2 which shows the number of misclassified testing cases lying within quartiles defined on the magnitude of the posterior probability of membership to the allocated class for the testing set. It was evident that most of the 28 erroneously allocated cases displayed a relatively small posterior probability of membership to the allocated class with 17 (~60%) of the misclassified cases lying within the lowest quartile (Table 2). The posterior probability information derived could, therefore, be used to help highlight cases allocated with varying degrees of confidence. This information might perhaps be used to help direct fieldwork to refine the classification or to mask regions of high uncertainty from later analyses.

The posterior probabilities output may also be used to help in identifying possible routes to increase classification accuracy. It may, for example, highlight cases for which ancillary information is needed to increase classification accuracy. In particular, the output may be used to highlight some of the situations in which the classifier is unable to provide accurate discrimination. In helping to define the problematic cases the output may, therefore, help in the definition of enhancements that could be used to increase classification accuracy. For example, although most of the mis-classified cases displayed a relatively small posterior probability of membership to the allocated class, 7 of the mis-allocated cases lay within the upper two quartiles of posterior probability defined (Table

2). This information highlights that some of the errors in the classification were confident mis-allocations, in which cases were allocated with a large posterior probability to an incorrect class. Closer inspection of the output of the RVM classifier revealed that of the 8 most confidently mis-classified cases 6 were of cases of sugar beet being mis-allocated to the potato class. The errors arising from the confusion of these two classes was the largest source of classification error (Table 1) and the recognition that many of the errors were confident mis-allocations involving these classes indicated that further discriminatory variables (e.g. additional wavebands, textural information, acquisition of imagery at another time period) may be required to derive a more accurate classification with this particular classifier. Consequently, the analyst can be directed to focus efforts aimed at increasing classification accuracy on a major source of misclassification.

The results highlight the potential of the RVM-based approach for classification in remote sensing. The RVM-based approach is not, of course, without its problems. Although it may offer the potential for classification with very small training sets (Tipping, 2001; Bowd *et al*., 2005) it may be difficult to predict training sites likely to be appropriate relevance vectors. While the extreme nature of support vectors in a SVM-based analysis makes it reasonably easy to design a training data acquisition programme focused upon them (Foody and Mathur, 2004b; Mathur and Foody, 2007) the relevance vectors in a RVM analysis are more typical of the classes (Tipping, 2001) and possibly difficult to characterise in advance. The RVM may also be relatively unattractive as a classifier when training data are plentiful due to the computational complexity in learning (Tipping, 2001). Often, however, in remote sensing training data may be scarce or costly

to acquire and the RVM offers an attractive method of analysis in such circumstances, even if the data set is of high-dimensionality (Bowd *et al*., 2005). The attractive features of the RVM for the analysis of remotely sensed data should be further explored in future research.

5. Summary and conclusions

The RVM, a Bayesian extension of the SVM, was evaluated for multi-class image classification. The results highlighted that the RVM could be used to derive a very accurate multi-class classification (91.25%), a level insignificantly different to that from a SVM. However, the RVM has some major attractions over the SVM. In particular, the probabilistic nature of the RVM-based classification output may be of value from a variety of perspectives. It may, for instance, assist later users of the classification by indicating class allocation uncertainty on a per-case basis. In the example presented, most of the mis-allocations were associated with cases allocated on the basis of a small posterior probability of membership. The information on the posterior probabilities of class membership may, therefore, be used to provide a per-case guide to the confidence of class allocations. The probabilistic output information may also be of value to the analyst undertaking the classification, especially in helping to identify possible routes to refine the analysis to obtain further increases in classification accuracy. In the example presented, the output highlighted that some of the classification errors were the result of confident mis-allocations between two classes. This information should help focus efforts

to increase classification accuracy on the identification of means to enhance the separability of the problematic classes in the analysis.

References

BAZI, Y. and MELGANI, F., 2006. Toward am optimal SVM classification for hyperspectral remote sensing images, *IEEE Transactions on Geoscience and Remote Sensing*, **44**, 3374-3385.

BOWD, C., MEDEIROS, F. A., ZHANG, Z., ZANGWILL, L. M., HAO, J., LEE, T-W., SEJNOWSKI, T. J., WEINREB, R. N. and GOLDBAUM, M. H., 2005. Relevance vector machine and support vector machine classifier analysis of scanning laser polarimetry retinal nerve fiber layer measurements, *Investigative Opththalmology and Visual Science*, **46**, 1322-1329.

CAMPS-VALLS, G., GOMEZ-CHOVA, L., MUNOZ-MARI, J., VILA-FRANCES, J., AMOROS-LOPEZ, J. and CALPE-MARAVILLA, J., 2006. Retrieval of oceanic chlorophyll concentration with relevance vector machines, *Remote Sensing of Environment*, **105**, 23-33.

CHEN, Z. and TANG, H., 2005. Sparse Bayesian approach to classification, *IEEE Networking, Sensing and Control Proceedings*, 914-917.

DEMIR, B. and ERTURK, S., 2007. Hypersepctral image classification using relevance vector machines, *IEEE Geoscience and Remote Sensing Letters*, 4 (in press).

FOODY, G. M., 2002. Status of land cover classification accuracy assessment, *Remote Sensing of Environment*, **80**, 185-201.

FOODY, G. M., 2004. Thematic map comparison: evaluating the statistical significance of differences in classification accuracy, *Photogrammetric Engineering and Remote Sensing*, **70**, 627-633.

FOODY, G. M. and MATHUR, A, 2004a, A relative evaluation of multiclass image classification by support vector machine. *IEEE Transactions Geoscience Remote Sensing,* **42**, 1335-1343.

FOODY, G. M. and MATHUR, A, 2004b, Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification. *Remote Sensing of Environment,* **93**, 107-117.

HSU, C. W. and LIN, C. J., 2002, A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks,* **13**, 415-425.

HUANG, C., DAVIS, L. S. and TOWNSHEND, J. R. G., 2002, An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing,* **23**, 725-749.

MATHUR, A. and FOODY, G. M., 2007. Crop classification by a SVM with intelligently selected training data for an operational application, *International Journal of Remote Sensing* (in press).

MELGANI, F. and BRUZZONE, L., 2004. Classification of hyperspectral remote sensing images with support vector machines, *IEEE Transactions Geoscience Remote Sensing,* **42**, 1778-1790.

PAL, M. and MATHER, P. M., 2004. Assessment of the effectiveness of support vector machines for hyperspectral data, *Future Generation Computer Systems*, **20**, 1215-1225.

PAL, M. and MATHER, P. M., 2005. Support vector machines for classification in remote sensing, International Journal of Remote Sensing, **26**, 1007-1011.

TIPPING, M. E., 2001. Sparse Bayesian learning and the relevance vector machine, *Journal of Machine Learning Research*, **1**, 211-244.

WILKINSON, G. G., 2005. Results and implications of a study of fifteen years of satellite image classification experiments, *IEEE Transactions on Geoscience and Remote Sensing*, **43**, 433-440.

ZHANG, H. and MALIK, J., 2005. *Selecting shape features using multi-class relevance vector machine*, Technical Report UCB/EECS-2006-6, Electrical Engineering and Computer Sciences, University of California at Berkeley, USA, 2005.

	Sugar beet	Wheat	Barley	Carrot	Potato	Grass	$\sum$
Sugar beet	87	3	0	0	7	0	97
Wheat	3	90	2	1	0	0	96
Barley	1	5	45	0	0	0	51
Carrot	0	1	0	32	0	0	33
Potato	1	2	0	0	22	1	26
Grass	0	0	0	1	0	16	17
$\sum$	92	101	47	34	29	17	320

Table 1. Confusion matrix from the RVM-based classification. In the matrix, rows represent the actual class of membership while columns the predicted class of membership. The highlighted main diagonal indicates correctly allocated cases.

	Posterior probability		Number	Number
Quartile	Min	Max	misclassified	classified correctly
1	0.9311	1.0000	2	78
2	0.8504	0.9307	5	75
3	0.7241	0.8497	4	76
4	0.3161	0.7241	17	63

Table 2. The number of testing cases lying within quartiles defined on the posterior

probability of membership to the allocated class (*n*=320).