Paper Reference No.: PN-20

Title of the paper : Adaptive Local Hyperplane based Remote Sensing Classification

Name of the Presenter : Mahesh Pal

Author (s) Affiliation : NIT Kurukshetra

Mailing Address : Department of Civil Engineering, NIT Kurukshetra, 136119, Haryana, INDIA

Email Address : mpce_pal@yahoo.co.uk

Telephone number (s) : 01744-233356

Fax number (s) : 01744-238050

Adaptive local hyperplane based remote sensing classification

Mahesh Pal
Department of Civil Engineering
N. I. T. Kurukshetra, Harayana, 136119, INDIA
mpce_pal@yahoo.co.uk

Abstract: This paper explores the potential of a modified K-Nearest Neighbour classifier called adaptive local hyperplance (ALH) classifier for remote sensing data classification. ETM+ multispectral data set (England) was used to judge the suitability of this classifier and results were compared with a back propagation neural network in term of classification accuracy and computational cost. A classification accuracy of 88.32% was achieved by the ALH classifier in comparison to an accuracy of 87.75% provided by a backpropagation neural network with this dataset, suggesting a better performance by ALH classifier in term of classification accuracy. A comparison of computational cost also suggests a better performance by ALH classifier (6.45 second) in comparison to back propagation neural network (250.55 seconds).

1. Introduction

Past three decades have seen continuing developments in the area of pattern recognition procedures and users of remotely sensed data now have access to sophisticated statistical and neural algorithms for land cover classifications (Mather, 1999). Both the statistical and neural/connectionist approaches have their limitations. Statistical methods rely on the assumption that probabilities of class membership and the size of the training data sets on which estimates of the parameters of the multivariate normal distribution are based is a limiting factor, especially when working with hyperspectral data (Mather, 1999). On the other hand, neural methods work well with smaller training data sets than those required by statistical methods, but network training times can be lengthy, while choices in terms of the design of network architecture (in terms of numbers of hidden layers and neurones per layer) and the values of the learning rate parameters are not straightforward (Pal, 2009; Wilkinson, 1997).

Nearest neighbour classifiers have not been as widely used within remote sensing even being non-parametric in nature. Few studies reported the use of nearest neighbour classifier in remote sensing includes Ince (1987), Franco-Lopez et al. (2001), McRoberts et al. (2002), Haapanen et al. (2004), Gjertsen (2007) and Thessler et al. (2008). This work report the use of a modified nearest neighbour classifier called as Adaptive local hyperplane (ALH) classifier recently proposed Yang and Kecman (2008).

2.0 Adaptive local hyperplane classifier

Instance-based classifiers such as the K-NN classifier operate by relating the unknown pixel to the known according to some distance/similarity function. These classifiers do not derive any information from the training data during the training phase. Learning with these classifiers is merely a question of summarizing the training data. Classification using an instance-based classifier can be a matter of locating the nearest neighbour in instance space and labelling the unknown pixel with the same class label as that of the known neighbour, a reason of calling these classifiers as nearest neighbour classifiers. More robust nearest neighbour classifier can be achieved by keeping $K > 1$, neighbours and allowing the majority vote to decide the outcome of the class labelling on test dataset. A higher value of K will results in a smoother, less locally sensitive classifier but as the value of K approaches to the number of training samples, all test samples would be assigned to the most frequented class in the training data.

Keeping in view the problems in designing the K-NN classifier, several improved nearest neighbour classifiers have been proposed in literature. Li and Lu (1999) proposed an improved NN classifier called as nearest feature line (NFL) classifier whereas Chien and Wu (2002) proposed nearest feature plane (NFP) classifier. Both of these classifiers perform well in term of classification accuracy but requiring large computation cost and may not work well if the prototypes are far away from the new test sample. To improve upon the NFL and NFP, Zeng et al., (2004) proposed two linear combination based classification methods by using the locally linear embedding (LLE) method (Roweis and Saul, 2000). Vincent and Bengio (2002) proposed a modified NN classifier called as K-local hyperplane distance nearest neighbour classifier (HKNN), which builds a decision surface, separating different classes of data directly in input space. This classifier works by first selecting K-NNs of a query from each class as the class prototypes and then constructs a local hyperplane to approximate the local manifold of each class based on selected class prototypes. In spite of its better performance in comparison to K-NN classifier, Vincent and Bengio (2002) suggested that HKNN could work well for small values of K and assume that all features are equally relevant to the classification.

To solve the problems with HKNN, Yang and Kecman (2008) suggested an improved classifier called Adaptive local hyperplane (ALH) classifier that considers the feature weight, thus assigning different weights to different features. Their approach selects the prototypes by the adaptive nearest neighbor method and the feature weight is estimated by using the ratio of the between group to within group sums of squares. The resulting prototypes are then associated with the most discriminant feature dimensions (Yang and Kecman, 2008).

If the training set consists of N training samples with C classes. Each training sample consists of p input features $\mathbf{x}_i = (x_{i1}, \ldots\ldots\ldots, x_{ip})^T$ with class labels $y_i = c$, for $i = 1, \ldots\ldots\ldots, N \text{ and } c = 1, \ldots\ldots\ldots, C$. The aim of the proposed classifier is to predict the class label of a query with input data vector defined by ***q*.** Different steps in the design of ALH classifier involves in selecting nearest neighbours, constructing local hyperplane and finally query classification. ALH classifier uses weighted Euclidean distance metric to select the nearest neighbours of the query and feature weight is calculated based on the ratio of the between group to within group sums of squares. This ratio for a feature *f* is defined as:

$$r_f = \frac{\sum_i \sum_c I(y_i = c)\left(x'_{cf} - x'_f\right)^2}{\sum_i \sum_c I(y_i = c)\left(x_{if} - x'_{cf}\right)^2} \tag{1}$$

where $I(\bullet)$ denotes the indicator function, x'_{cf} is the *j*th component of centroid of class *c* and x'_f is the jth component of grand class centroid. Feature weight can then be defined:

$$w_f = \frac{\exp(\mathrm{T}(\mathrm{r_f}/\max(\mathrm{r_f})))}{\sum_{f=1}^{p} \exp(T(\mathrm{r_f}/\max(\mathrm{r_f})))} \tag{2}$$

where T is a positive parameter controlling the influence of $(\mathrm{r_f}/\max(\mathrm{r_f}))$ on w_f. Yang and Kecman (2008) suggested using the following equation to find local hyperplane of class *c*.

$$LH_c(\mathbf{q}) = \left\{ s \,\middle|\, s = \sum_{\mathrm{i}=1}^{\mathrm{n_c}} \boldsymbol{\alpha}_{\mathrm{i}} \mathbf{V}_{\mathrm{f}} + \mathbf{m} \right\} \quad \text{where } \mathbf{m} = \frac{1}{n_c}\sum_{i=1}^{n_c} \mathbf{pr}_i \tag{3}$$

$\mathbf{pr}_i$ is the *i*th nearest neighbour of class c, **V** is $p \times n_c$ matrix whose ith column is defined as: $\mathbf{V}_{.i} = \mathbf{pr}_i - \mathbf{m}$, n_c is the number of nearest neighbour in class c and $\boldsymbol{\alpha} = (\alpha_1, \ldots\ldots\ldots, \alpha_{n_c})^T$ are obtained by minimizing the following equation:

$$J_c(q) = \min_{\alpha} \sum_{f=1}^{d} w_f \left(\mathbf{V}_{f.}\boldsymbol{\alpha} + m_f - q_f\right)^2 + \lambda\,\boldsymbol{\alpha}^T\,\boldsymbol{\alpha}$$

$$= \min_{\alpha} (\mathbf{s} - \mathbf{q})^T\ \mathbf{W}\ (\mathbf{s} - \mathbf{q}) + \lambda\,\boldsymbol{\alpha}^T\,\boldsymbol{\alpha} \qquad (4)$$

where $\mathbf{V}_{f.}$ is the *f*th row of $\mathbf{V}$, $\mathbf{s} \in \mathbf{LH}_c(\mathbf{q})$, $\mathbf{W}$ is the diagonal matrix with $W(f,f) = w_f$ and λ is the regularization parameter.

ALH requires setting of three user-defined parameters: K (i.e. number of nearest neighbours), T and λ. A grid search method was used to find the optimal values of these three parameters using classification accuracy with the test dataset as the main criteria.

3. Data Sets and Methodology

The study areas used in this study is located near the town of Littleport in eastern England and the image was acquired on 19 June 2000. A sub-image consisting of 307-pixel (columns) by 330-pixel (rows) covering the area of interest was used for subsequent analysis and classification problem involved in identification of seven land cover types (i.e. wheat, potato, sugar beet, onion, peas, lettuce and beans). A total of 4737 pixels were selected for all seven classes using stratified random sampling. The pixels collected were divided into two subsets, one of which was used for training and the second for testing the classifiers, so as to remove any bias resulting from the use of the same set of pixels for both training and testing. Also, because the same test and training data sets are used for each classifier, any difference resulting from sampling variations was avoided. A total of 2700 training and 2037 test pixels were used.

4. Results

The purpose of the present study was to evaluate the performance of ALH classifier for land cover classification and comparing its performance with a back propagation neural network classifier. The optimal values of user-defined parameters with both ALH and neural network classifier are provided in Table 1. For further details about user-defined parameters for back propagation neural networks readers are referred to Pal (2009). Similar to neural network classifier, the best values for hyperparameters K, T and λ with ALH classifier are data dependent. Variation in classification accuracy with the change in value of K and T using ALH classifier is provided in Figure 1.

Table1. Classification accuracy, user-defined parameters and computational cost with both classifiers.

Classifier used	User-defined parameters	Accuracy (%)	Computational cost (seconds)
ALH classifier	K = 11 T = 2.5 lambda = 0.5	88.32	6.45
Back propagation neural network	Learning rate =0.25, Momentum = 0.2, nodes in hidden layers =26, number of iterations = 2200, number of hidden layers =1	87.75	250.55

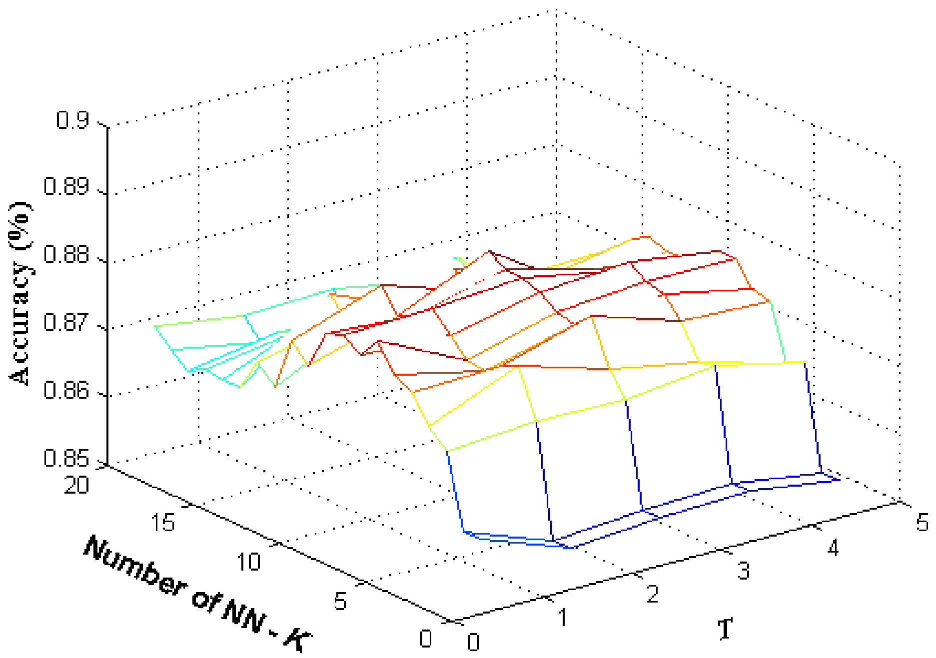

Figure 1. Variation in the accuracy of the ALH classifier with different values of K and T .

Table 1 also provides the results obtained by using ALH and neural network classifier using ETM+ (England) data set. Results suggest that ALH works well to the back propagation neural network in terms of classification accuracy. A classification accuracy of 88.32% is achieved by ALH classifier in comparison to 87.87% by the back propagation neural network. Computational cost (i.e. training and test time) of a classifier often represents a significant proportion of cost in remote sensing classifications. For all experiments in this study, a personal computer with a Pentium IV processor and 4GB of RAM was used. Table 1 also provide the computational cost with ALH and back propagation neural network. Results (table 1) suggest a better performance by ALH classifier in term of computational cost also. Tables 2 and 3 provide the confusion matrices with ALH and neural network classifier respectively.

Table 2. Confusion Matrix with ALH classifier

	class1	class2	class3	class4	class5	class6	class7	Sum	Users
class1	288	3	3	3	0	0	0	297	96.97
class2	5	251	13	1	18	9	4	301	83.39
class3	4	11	245	22	0	6	2	290	84.48
class4	2	12	27	262	3	0	2	308	85.06
class5	1	13	3	2	275	1	0	295	93.22
class6	0	9	8	7	4	269	20	317	84.86
class7	0	1	1	3	0	15	209	229	91.27
sum	300	300	300	300	300	300	237		
Producers	96.00	83.67	81.67	87.33	91.67	89.67	88.19		

Table 3. Confusion matrix with Neural network classifier

	class1	class2	class3	class4	class5	class6	class7	Sum	Users
class1	290	5	4	2	0	0	0	301	97.64
class2	2	239	11	4	9	5	3	273	79.40
class3	4	17	241	16	1	4	4	287	83.10
class4	2	5	26	266	5	10	3	317	86.36
class5	2	22	5	1	282	0	0	312	95.59
class6	0	11	12	9	3	271	26	332	85.49
class7	0	1	1	2	0	10	201	215	87.77
sum	300	300	300	300	300	300	237		
Producers	96.67	79.67	80.33	88.67	94.00	90.33	84.81		

5.0 Conclusions

The main aim of this study was to assess the usefulness of ALH classifier, an alternative to K-NN classifier for land cover classification using multispectral data. The results presented in this study suggest that ALH works well in comparison to the neural network classifier both in terms of classification accuracy and computational cost. Further, results also suggest that it is easier to identify the user-defined parameters of ALH classifier than the back propagation neural network classifier.

References

Chien, J.T. and Wu, C.C., 2002, Discriminant wavelet faces and nearest feature classifiers for face recognition, *IEEE Trans. Pattern Anal. Mach. Intell.*, 24 (12), 1644–1649.

Franco-Lopez, H., Ek, A. R. and Bauer, M. E., 2001, Estimation and mapping of forest stand density, volume, and cover type using the k-nearest neighbours method, *Remote Sensing of Environment*, 77(3), 251-274.

Gjertsen, A. K., 2007, Accuracy of forest mapping based on Landsat TM data and a kNN-based method, *Remote Sensing of Environment*, 110(4), 420-430.

Haapanen, R., Ek, A. R., Bauer M. E. and Finley, A. O., 2004, Delineation of forest/nonforest land use classes using nearest neighbor methods. *Remote Sensing of Environment*, Volume 89, Issue 3, 265-271.

Ince, F., 1987, Maximum likelihood classification, optimal or problematic? A comparison with the nearest neighbour classification, *International Journal of Remote Sensing*, 8(12), 1829 – 1838.

Li, S.Z. and Lu, J.W., 1999, Face recognition using the nearest feature line method, *IEEE Trans. Neural Networks*, 10 (2), 439–443.

Mather, P. M., 1999, *Computer Processing of Remotely-Sensed Images: An Introduction*. Second Edition, Chichester: John Wiley and Sons.

McRoberts, R. E., Nelson, M. D. and Wendt, D. G., 2002, Stratified estimation of forest area using satellite imagery, inventory data, and the k-Nearest Neighbors technique, *Remote Sensing of Environment*, 82(2-3), 457-468.

Pal, M. 2009, Extreme learning machine based land cover classification, *International Journal of Remote Sensing*, 30(14), 3835–3841.

Roweis, S.T. and Saul, L.K., 2000, Nonlinear dimensionality reduction by locally linear embedding, *Science*, 290, 2323–2326.

Thessler, S., Sesnie, S., Bendaña, Z. S. R., Ruokolainen, K., Tomppo, E. and Finegan, B., 2008, Using k-NN and discriminant analyses to classify rain forest types in a Landsat TM image over northern Costa Rica, *Remote Sensing of Environment*, 112(5), 2485-2494.

Vincent P. and Bengio, Y., 2002, K-Local Hyperplane and Convex Distance Nearest Neighbor Algorithms, in *Advances in Neural Information Processing Systems*, Eds. by T. G. Dietterich, S. Becker, and Z. Ghahramani (MIT Press, Cambridge, MA), 14, 985–992.

Wilkinson, G.G., 1997, Open questions in neurocomputing for Earth observation. *Neuro-Computational in Remote Sensing Data Analysis*, New York: Springer-Verlag, 3-13.

Yang, T. and Kecman, V., 2008, Adaptive local hyperplane classification, *Neurocomputing*, 71, 13-15, 3001-3004.

Zheng, W., Zhao, L. and Zou, C., 2004, Locally nearest neighbor classifiers for pattern classification, *Pattern Recognition*, 37, 1307 – 1309.

Mahesh Pal obtained his PhD degree from University of Nottingham and currently working in the Department of Civil Engineering, NIT Kurukshetra, Haryana. He has published about 55 Paper in international/national journals and conferences. His current interests include kernel based approaches for land cover classification and feature selection with Hyperspectral data. He is in the editorial board of *Remote Sensing Letters*, an international journal to be published by Taylor and Francis from 2010.