Application of nonparametric Bayesian classifier to remote sensing data

Nina Jeliazkova, nina@acad.bg, +359 2 979 6606

Stela Ruseva , stela@acad.bg, +359 2 979 6606

Kiril Boyanov, boyanov@acad.bg

Institute of Parallel Processing, Bulgarian Academy of Sciences

25A "acad G.Bonchev str."

Sofia 1113

Hristo Nikolov , hriston@bitex.com

Doyno Petkov

The Solar and Terrestrial Influences Laboratory, Bulgarian Academy of Sciences

Sofia

Key Words: *Bayesian classification, kernel density estimation, remote sensing, neural network, cover type recognition.*

Abstract

A nonparametric Bayesian classification, based on a recently published very fast algorithm for multivariate density estimation is proposed. The classifier is applied to the problem of land cover type recognition of remote sensing data. A 7 channel satellite image of a region of North Bulgaria is used as input data. The procedure of object recording at a distance, forming the image by recording reflected light or radio waves, is known as remote sensing.

The performance of the nonparametric Bayesian classifier is analyzed and compared to the performance of a backpropagation neural network over the same data. The proposed probabilistic approach gives very promising results. The assigned class membership of a pixel is a "soft classification". These results could be regarded as a realistic interpretation of the world, where land covers intergrade gradually, and boundaries between classes are sometimes blurred. In addition to pixel-by-pixel classification of an image, the method allows a classification of predefined image regions as a whole.

1.Introduction

Every day earth-observing satellites produce vast quantities of image data, much of it multispectral. These images are used for a wide variety of applications, ranging from weather prediction, through agricultural usage monitoring, to making thematic maps of areas on Earth surface. The process is known as remote sensing, since the recording of objects is done at a distance, forming the image by gathering, focusing and recording reflected light from the sun, or reflected radio waves emitted by the spacecraft. A channel (band) is a slice of wavelengths from the electromagnetic spectrum, measured by the instrument onboard the satellite.

One of the core tasks in this image analysis is the identification of broad area features of interest: clouds, wheat fields, forests, etc. Recognition of broad area features can be considered as a problem of performing a pixel-by-pixel classification of a given image. Ideally we want to obtain a confidence of the class membership of each pixel (rather than producing a strict binary classification), in order to trade the detection rate against the false alarm rate by varying a confidence threshold.

Classification of broad area features in satellite imagery is one of the most important applications of the remote sensing. Many researchers have turned to techniques from the fields of statistics and machine learning to generate classifiers. Common techniques include statistical classifiers, neural networks and genetic algorithms.

The objective of classification methods is to determine the class that a given sample belongs to. The observation vector is usually obtained through some measurement process (not only spectral features, but also their combinations and some additional information such as DEMs, temperature, etc.), and serves as the input to a decision rule by which the sample is assigned to one of the given

classes. The simplest method for the classification is to compare observation points (in defined feature space), and decide on class membership based on closest distance between a point and a class. This family of methods requires a definition of the "distance" between points and user-defined classes, and is known as "lazy learning" or "model-free" approach. Another option are model based classifiers, which aim at deriving decision boundary between classes, and once derived, this "boundary" or "model" is sufficient for classification. Examples of model-based classifiers are linear classifiers (e.g. Linear Discriminant Analysis), nonlinear classifiers (quadratic classifiers, neural networks, support vector machines, etc.), hierarchical classifiers (e.g. decision trees) and probabilistic classifiers. Despite different names and techniques, the essential distinction between model based classifiers is the shape of the derived boundary [1].

Artificial neural networks (ANN) have been successfully applied to the remote sensing data classification. One serious advantage is that ANN can correctly classify non-continuos regions from the feature space. The same is true, however, for the probabilistic classification based on nonparametric density estimation, used in this paper. In fact it is known that ANN are equivalent to statistical classifiers, but provide more computationally effective procedures [2].

The probabilistic approach is based on Bayes Theorem and is well known for its theoretical optimality in the sense of minimum classification error. Its drawback is that the probability distribution of the data has to be known. In most cases the distribution is not known and there are two different approaches to its estimation:

- Assume that the probability distribution has a known shape (e.g. Gaussian) and estimate its parameters (e.g. mean and variance) – this is known as a parametric approach and is most

commonly used;

- Do not make an assumption about the shape of probability distribution, but estimate the distribution from data. These methods include semi parametric (mixtures) and nonparametric techniques like kernel density estimation [3,4].

We choose to use the nonparametric classification approach based on Bayes' decision rule, first because it provides a theoretically optimal classification [1, 3] and second, because in remote sensing data classification, higher order moments of probability distributions are more important for the classification [5]. According to statistical tests applied to our data (Jarque-Bera test using the corresponding Matlab function), the hypothesis of data normality is rejected and therefore we could expect more precise classification results if the data distribution is reflected more accurately by a nonparametric technique. As an illustration, normal probability plots are generated using Matlab and shown in Figure 1. If all the data points fall near the red line, the assumption of normality is reasonable. But, if the data is non normal, the plus signs may follow a curve. The plots below are clear evidence that the underlying distribution is not normal. The plots of the rest of the channels (not shown in the figure) are also non normal.
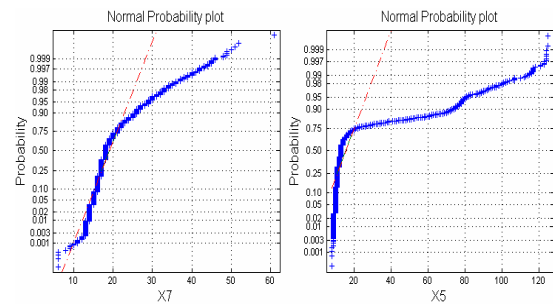

Figure 1. Normal probability plots of channels 5 and 7 of Landsat image.

Below we demonstrate the advantages of using Bayesian classifier, based on a recently proposed Very Fast Algorithm for Multivariate Kernel Density Estimation (Gray 2003) [6,7] to the problem of different land cover type recognition.

This algorithm achieves several orders of speed improvement by using computational geometry to organize the data. The implementation uses kind of "kd-trees", a hierarchical representation for point sets, which caches sufficient statistics about point locations in order to achieve potential speedups in computation. For kernels with infinite support (like Gaussian) [3,4,6,7] it provides an approximation tolerance level, which allows tradeoffs between evaluation quality and computation speed. The implementation of this algorithm is available as a Matlab toolbox (Ihler, 2004) [8].

2.Data

In the context of remote sensing, the observation vector (independent variables) consists of the spectral responses of image pixels. The class membership of homogeneous regions is identified by a topographic map, in-situ investigations or other knowledge about the region and is referred to as a "ground truth".

The input data consists of a 7-channel satellite image (667 x 663 pixels) of the central part of North Bulgaria. It is obtained by the Thematic Mapper instrument onboard the Landsat satellite [9,10]. The ground truth is based on a contemporary topographic map of the same region. GIS shape files, providing additional information about classes such as correct nomenclature, were introduced by the *Corine Land Cover 1994* project for Bulgaria [11,12]. An automated analysis was performed, and the 12 classes were identified. For our study we selected only 6 classes for classification as representatives for the broader categories of artificial surfaces, agricultural areas, forests and water bodies found in the selected area. The classes used are as follows (according to CLC1994): 112 – Urban fabric; 142 – Sport and leisure facilities; 221 - Vineyards; 231 - Pastures;

311 – Broad leaved forest; 512 – Water bodies.

Table 1. Statistics of the training and test set of the Landsat TM image.

No	Class ID	Image		Training set		Test set	
		Pixels	Obj	Pixels	Obj	Pixels	Objects
1	112	16254	13	10873	11	5381	2
2	121	1874	6	1302	5	572	1
3	142	1077	2	418	1	659	1
4	211	241637	44	146386	43	95251	1
5	212	2543	2	1069	1	1474	1
6	22	10357	14	5848	13	4509	2
7	231	31418	39	25841	38	5577	1
8	24	46076	62	40105	59	5962	4
9	311	26582	27	18564	26	8018	1
10	32	14258	25	8869	22	5389	4
11	411	580	1	285	1	285	1
12	512	10532	6	2367	5	8165	1

As an input to the classification procedure (model), every pixel is represented by 7 values for each spectral band and additional 8x7 values per each of its eight immediate neighbors, thus giving 63 input features per pixel. By using these additional features, some information about the texture is introduced. The texture information is necessary in order to make more accurate classifications of the regions with similar spectral response (for example deciduous forests and fruit trees).

The preprocessing of the data included centering (subtracting the mean), scaling to unit variance and performing Principal Component Analysis (PCA). Based on PCA a set of first 12 principal components was selected to be used as most significant feature for the classification.

Other approaches of considering the texture information and feature selection will be a focus of further research.

The data was further separated into training and validation (or test) sets in a way that the validation data set for each class consists of a single, but the largest object from that class (Table 1).
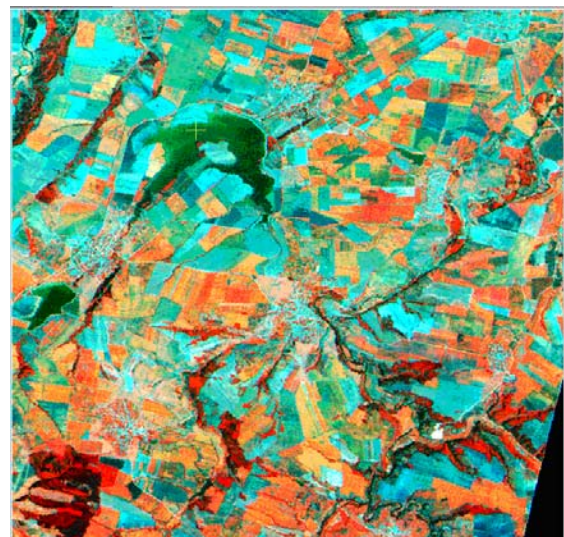

Figure 2. The satellite image used for the cover type recognition, RGB composite channels 5, 4, 3 of TM7.

3.Method

3.1.Nonparametric Bayesian classification

The Bayesian approach assumes that the observation vector **x** is a random vector whose conditional probability density depends on its class and is based on Bayes theorem:

$$p(\omega_i \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \omega_i)P(\omega_i)}{p(\mathbf{x})} = \frac{p(\mathbf{x} \mid \omega_i)P(\omega_i)}{\sum_{i=1}^{c} p(\mathbf{x} \mid \omega_i)P(\omega_i)} \qquad 1)$$

where $p(x/\omega_i)$ is the conditional probability density; $P(\omega_i)$ is the a priori probability of class ω_i, c is the number of classes. If the conditional density function for each class ω_1, ω_2,... ω_c is known, then according to the Bayes decision rule, we allocate the observation to the class with the highest posterior probability $p(\omega_i/x)$. This rule is the theoretically optimal decision rule, and guarantees lowest classification error [1,3]

$$p(\omega_i \mid x) > p(\omega_j \mid x), j=1..c, i \neq j \quad \rightarrow \quad x \in \omega_i \qquad 2)$$

We assume that the *a priori* probabilities of classes are equal, and estimate class-conditional probabilities by recently proposed Very Fast Algorithm for Multivariate Kernel Density Estimation [6,7]. As an illustration, only one dimensional class-conditional and posterior are shown on Figure 3, but for the classification 12-dimensional probability densities are estimated. The classification process for a single test region of class 231 is illustrated in Figure 4. First, for each pixel of the region, the multivariate posterior probability of each class is estimated, and is shown in Figure 4a. One can see that the posterior probability for the class 231 is higher for most pixels of the region (red pixels are with probability close to 1, and blue pixels are with probability close to 0). Then each pixel is assigned to the class with the maximum posterior probability (Figure 4b).
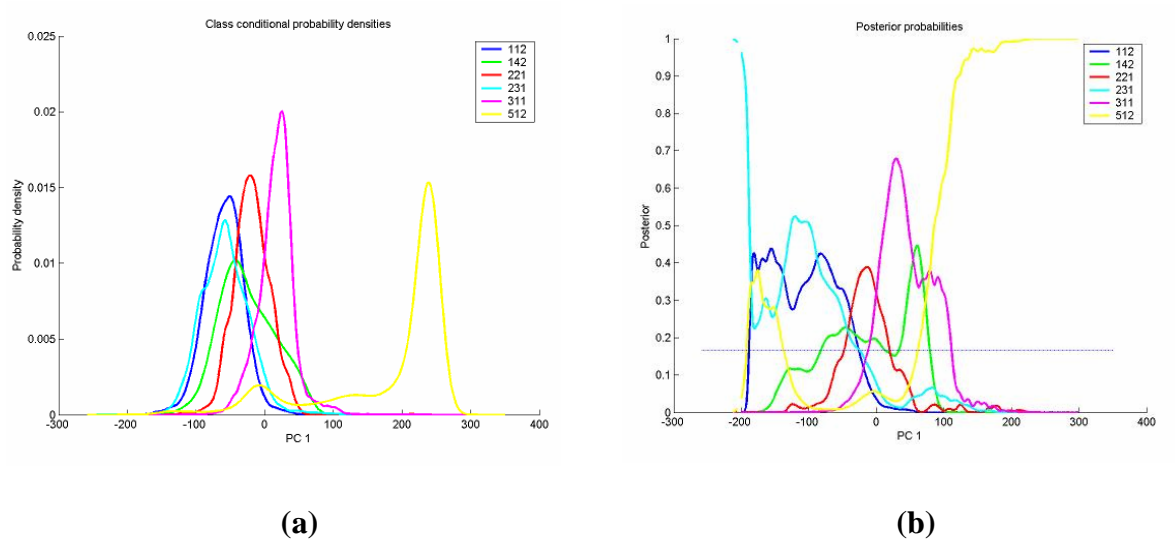
6

(a)   (b)

Figure 3. **The class-conditional (a) and posterior (b) probabilities for the first principal component. Only one-dimensional probability distributions are displayed here, but the classification is performed in the multivariate space of first 12 principal components.**
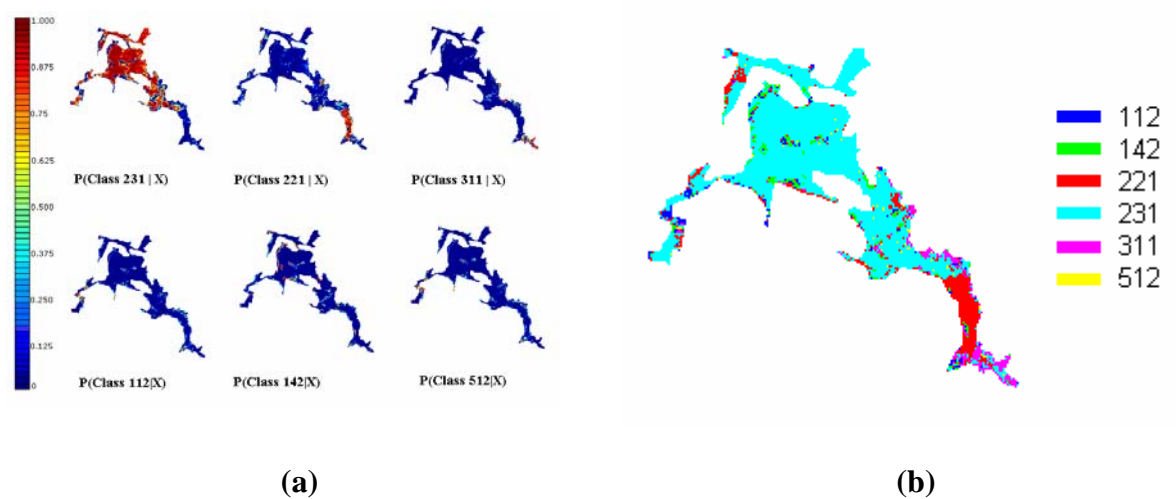

(a)   (b)

Figure 4. **(a) Posterior probabilities of a test object of class 231 (color means posterior probability; red is 1.0; blue is 0.0); (b) Resulting pixel by pixel classification**

3.2. Neural network

The performance of the proposed nonparametric Bayesian classification is compared to the performance of a neural network trained by backpropagation. The structure of ANN consists of seven neurons in the input layer, twenty five neurons in the hidden layer and six neurons in the target layer. The seven input neurons receive the seven spectral channels of the image. The activation functions are selected to be sigmoidal [13]. The Matlab Neural Network toolbox is used for the training and simulation of the neural network.

4. Results

4.1. Pixel-by-pixel

classification by a nonparametric Bayesian classifier

A pixel is assigned to the class with the maximum posterior probability. The mean accuracy of the training and test sets as well as the accuracy per each class are shown in Table 2. The ground truth and predicted classes are shown in Figure 5.

Table 2. Classification accuracy of a pixel-by-pixel classification by nonparametric Bayesian approach

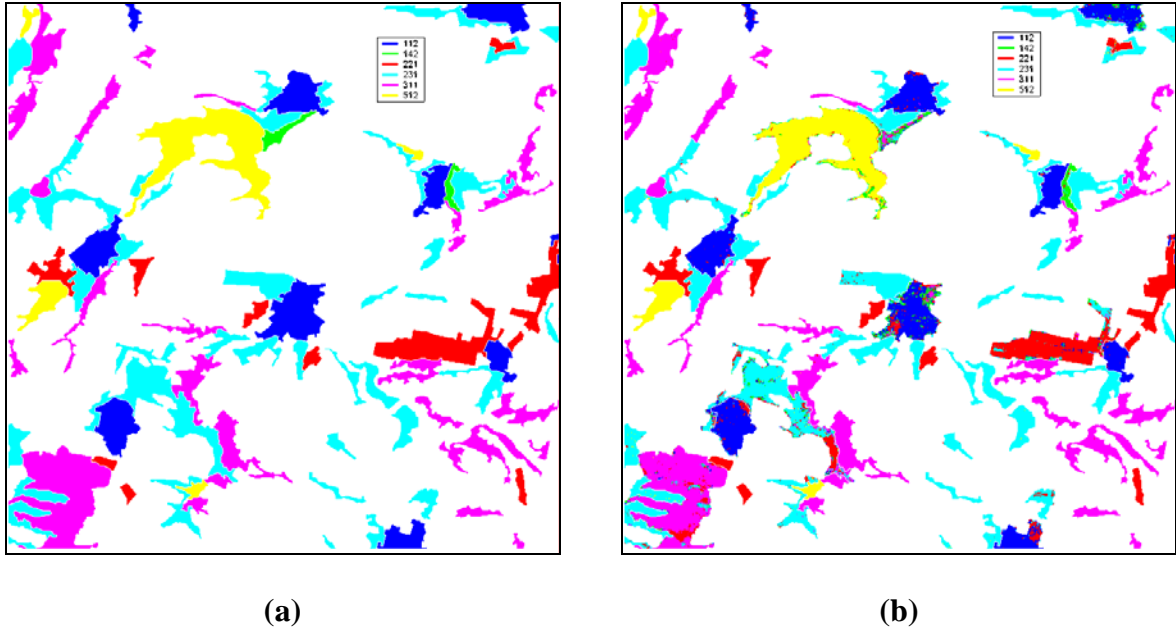| Class ID	Class name	Accuracy per class	
		Training set	Test set
112	Urban fabric	92%	72%
142	Sport and leisure facilities	92%	41%
221	Vineyards	99%	84%
231	Pastures	99%	74%
311	Broad-leaved forest	99%	89%
512	Water bodies	100%	91%
All	**Mean over classes**	**97%**	**82%**

Figure 5. The "ground truth" (a) and predicted classes (b) (color coded).

4.2. Object classification by a nonparametric Bayesian classifier

In the land cover type recognition we are more interested in recognizing continuous regions than single points. On the chart below (Figure 6) the mean probability over pixels in the test regions is shown. If the region is classified to the class with a maximum mean probability, then all classes are correctly recognized by this criterion.
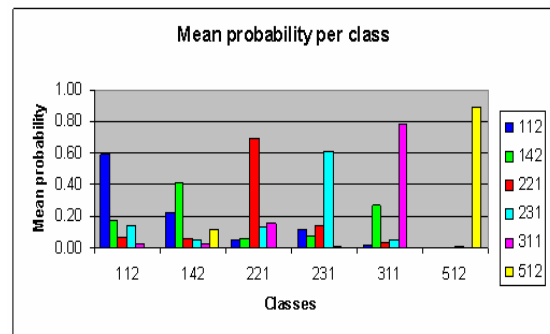

Figure 6. Mean probability per class for test regions. For all classes the largest mean probability is for the correct class, therefore all regions are correctly recognized.

These results could be regarded as a realistic interpretation of the world, where land covers intergrade gradually, and boundaries between classes are

sometimes blurred (e.g. for the class 142 – "Sport and leisure facilities", the second highest probability is that of class 112 - "Urban fabric" or for the class 311 – "Broad leaved forests", the second highest probability is that of class 142, and it is highly probable that parks consisting of broad leaved trees are recognized as broad leaved forests).

4.3.Pixel-by-pixel classification by a backpropagation neural network

The accuracy of neural network classification is summarized in Table 3, where rows represent classes as observed (ground truth), and columns represent predicted classes. The cell *(i,j)* contains the percent of pixels from *class i*, predicted as *class j*. One could easily see that even the class with the lowest accuracy (112) is correctly recognized with more than 50%. The comparison between the two proposed methods leads to the conclusion that both

non-linear techniques perform well in data classification.

Table 3. Neural network classification accuracy of test sites

Class	112	142	221	231	311	512
Accuracy, %						
112	**65.28**	8.87	6.7	2.4	3.51	13.95
142	5.26	**72.32**	5.61	0.79	7.95	6.37
221	5.39	8.26	**79.81**	1.19	1.28	4.92
231	7.42	0.65	2.11	**87.87**	0	2.59
311	1.09	6.45	3.86	1.82	**84.46**	5.51
512	15.62	6.45	4.91	6.75	1.79	**65.66**

5.Conclusions

The proposed probabilistic approach gives very promising results. The assigned class-membership of a pixel is a "soft classification", i.e. a probability of a pixel class membership is provided instead of a"yes"/"no" answer. This could be very helpful in the context of classification of remote sensing imagery, since it is useful to predict the degree of membership to a given class. In addition to pixel-by-pixel classification of an image, it allows classification of predefined regions of the image as a whole. The classifications of regions as a whole are accurate even in a

hard scenario like using only spectral information as features and discriminating among 12 classes. The high error rate for some of the classes is the result of insufficient data and failure to use the relevant features (texture), which will be a focus of further research. The method could be applied not only to remote sensing data, but to any data classification problem.

Together with the probabilistic approach ANN was investigated. The obtained results proved that if not superior this method is at least a good alternative for data classification. We propose future research to include the investigation of different ANN structures, more features for the classes, and refine training data.

References:

1. R. Duda, P. Hart, D. Stork. Pattern Classification, 2nd ed., John Wiley & Sons, 2000.

2. Haykin, S. Neural Networks: A Comprehensive Foundation , Prentice Hall, 1998

3. L. Devroye, L. Gyorfi, G. Lugosi. A probabilistic Theory of Pattern Recognition, Springer, 1996.

4. Nikolova N., Tourlakov H., Boyanov K., Knowledge Data Base search Methods for Big Volume of Data, 3/2002, pp. 3-9 (in Bulgarian)

5. Landgrebe D., Multispectral Data Analysis: A Signal Theory Perspective, http://dynamo.ecn.purdue.edu/~biehl/MultiSpec/Signal_Theory.pdf

6. Gray A. and A. Moore, Nonparametric Density Estimation:Toward Computational Tractability, Proc. SIAM International Conference on Data Mining , San Francisco, USA, 2003.

7. Gray A and A. Moore, Very Fast Multivariate Kernel Density Estimation using via Computational Geometry, in Proceedings, Joint Stat. Meeting 2003.

8. MATLAB KDE class, http://ssg.mit.edu/~ihler/code/kde.shtml.

9. The Landsat project,

http://landsat7.usgs.gov

10. Landgrebe, D., The Evolution of Landsat Data Analysis, Photogrammetric Engineering and Remote Sensing, Vol. LXIII, No. 7, July 1997, pp. 859-867.

11. http://ptl.gisat.cz/clc_dirb1.shtml

12.   http://europa.eu.int/comm/eurostat /ramon/corine/corine_en.html

13. Kanellopoulos, I. and Wilkinson, G. G. Strategies and best practice for neural network image classification. International Journal of Remote Sensing, 18, 711-725, 1997.