Bayesian Networks in the Classification of Multispectral and Hyperspectral Remote Sensing Images

Cristina Solares
University of Castilla-La Mancha
Department of Applied Mathematics
Ciudad Real
Spain

Ana Maria Sanz
University of Castilla-La Mancha
Department of Geological and Mining Engineering
Ciudad Real
Spain

Abstract: In this paper we study the application of bayesian network models to classify multispectral and hyperspectral remote sensing images. Different models of bayesian networks as: Naive Bayes, Tree Augmented Naive Bayes, Forest Augmented Naive Bayes and General Bayesian Networks, are applied in the classification of hyperspectral data. In addition, several bayesian multi-net models are applied in the classification of multispectral data. A comparison of the results obtained with the different classifiers is done.

Key–Words: Bayesian networks, Bayesian network classifiers, Multispectral image classification, Hyperspectral image classification, Bayesian multi-nets classifiers.

1 Introduction

Classification problems (see [7]) occur in a wide range of situations in real life such as disease diagnosis, image recognition, fault diagnosis, etc.

Probabilistic models, especially those associated with bayesian networks, are very popular as a formalism for handling uncertainty. The increasing number of applications developed these last years show that this formalism has practical value also.

In this paper we apply different models of bayesian networks to the classification of remote sensing images, considering multispectral and hyperspectral data sets. In the multispectral image the number of spectral bands for each pixel is less than 20, otherwise the image is called hyperspectral.

The paper is organized as follows. Section 2 introduces the bayesian networks and the bayesian networks as classifiers. Seven models of bayesian networks are introduced: General Bayesian Network (GBN), Naive Bayes (NB), Tree Augmented Naive Bayes (TAN), Forest Augmented Naive Bayes (FAN), GBN bayesian multinet, TAN bayesian multinet and the TAN-Based BCM multinet (tBCM). Section 3 presents the application of the above models to the classification of remote sensing images. In section 4 some conclusions are given.

2 Bayesian Networks

A Bayesian Network (BN) (see [1] for further details) over $\mathbf{X} = (X_1, \ldots, X_n)$ is a pair (D, P), where D is a directed acyclic graph with one node for each variable in X and $P = \{p_1(x_1|\pi_1), \ldots, p_n(x_n|\pi_n)\}$ is a set of n conditional probability distributions, one for each variable, given the values of the variables on its parent set Π_i (CP table). Each node in D represents a domain variable (eg, a dataset attribute) and each arc in D represents a probabilistic dependence between two variables quantified using the above CP table.

Here x_i and π_i denote realizations (instantiations) of X_i and Π_i, respectively. The joint probability distribution (JPD) of X can then be written as

$$p(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p_i(x_i|\pi_i). \quad (1)$$

2.1 Bayesian Network Classifiers

The application of bayesian network models to classification involves two sub-tasks: Learning the BN structure (the graphical structure D) and the BN parameters (CP table). It is trivial to learn the parameters for a given structure, simply use the empirical conditional frequencies from the data (see [5]). Constructing the BN structure can be performed using expert knowledge or directly from the data. There are different methods of learning a BN structure, as the score-based methods (see [5]) and the methods that learn the structure by identifying the conditional independence relation-ships among the nodes (CI-based methods). The score-based methods comprise a search procedure to find a network structure and a score is employed to evaluate each structure in the search space. The K2 algorithm, introduced in [5], is a search algorithm for finding a high quality bayesian network in a reasonable time. An example of CI-based method is the algorithm described in Cheng et al. [2]. Cheng et al. in [3] show that the CI-based learning algorithms are very efficient and the learned BN classifiers can give very good prediction accuracy.

Next we describe different models of bayesian network classifiers:

General Bayesian Network (GBN) A GBN with JPD $p(A_1, A_2, \ldots, A_n, C)$, can be constructed to solve a classification problem (see Figure 1). The variables $\mathbf{A} = (A_1, \ldots, A_n)$ are the attributes of the problem and C is the class variable having k different states. The resulting model (1) can be used to classify a given set of attribute values $\mathbf{a} = (a_1, \ldots, a_n)$ (see [8]). The vector $\mathbf{a}$ belongs to class $c \in C$ that maximizes the posterior probability $p(c|\mathbf{a})$. The structure of the GBN can be learned using a score-based method as the K2 algorithm (see [5]) or a CI based method as the algorithm introduced in [2]. In this paper, we use the K2 search algorithm.

Naive Bayes (NB) A Naive Bayes is a simple structure of bayesian network with the classification node C as parent node of all other nodes. Not other connections are allowed in this type of networks (see [8]).

Tree Augmented Naive Bayes (TAN) The very strong assumption of independence of all the attributes given its parents set in the Naive Bayes, not always realistic, is relaxed in this type of network. The TAN algorithm constructs a tree structure between the attribute nodes and after that adds a link from the classification node C to the attribute nodes $A_i, i = 1, \ldots, n$ (see [8]). This model is based in the algorithm described by Chow et al. in [4], for learning tree-like bayesian networks.

Forest Augmented Naive Bayes (FAN) This classifier is very similar to the TAN one. In the FAN algorithm instead of a tree in the attribute space a set of disconnected trees is considered (see [14]).

GBN Bayesian Multi-net A GBN bayesian multi-net is a generalization of the GBN, a different GBN is built for each class and a collection of networks is used as a classifier (see Figure 2). For that, we partition the training data set by classes and for each class value we construct a GBN for the attribute variables.

TAN Bayesian Multi-net The TAN model forces the relations among attributes to be the same for all the different instances of the class variable C. A bayesian TAN multi-net is a generalization of the TAN, a different TAN is built for each class and a collection of networks is used as a classifier (see [8], [6]). This model allows the relations among the attributes to be different for the different values of the class. For that, we partition the training data set by classes and for each class value we construct a TAN for the attribute variables.

TAN-Based BCM Multinet (tBCM) The tBCM is a multinet classifier that learns each local network based on a detection-rejection measure (see [9]). The method searches for the structure maximizing a discrimination-driven score that is computed using training patterns of all classes. This model uses the SuperParent algorithm (see Keogh et al. in [10]) to learn each local network, a TAN model having only augmented edges that increase the classifier accuracy. In Gurwicz et al.[9] the average superiority of the tBCM model in comparison with other classifiers, as the TAN multinet, is shown.

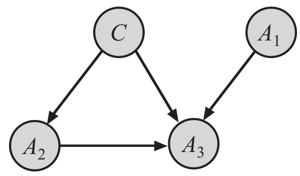

Figure 1: Example of General Bayesian Network (GBN). C is the class variable and A_1, A_2, A_3 are the attribute variables.

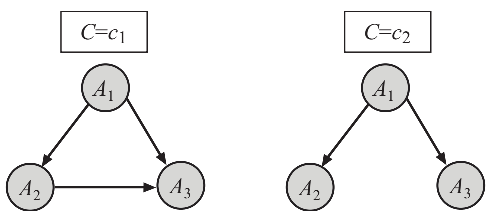

Figure 2: Example of Bayesian Multi-net. C is the class variable that takes two values c_1 and c_2, and A_1, A_2, A_3 are the attribute variables.

In the next section the above models of bayesian network classifiers are applied to the classification of remote sensing images.

3 Remote Sensing Image Classification

The models of bayesian networks introduced in Section 2 can be applied to classify remote sensing spectral images. For the implementation of the proposed models, we use the Bayes Net toolbox in matlab (see [12]) and the Leray's additional Structure Learning Package (see [11]).

A remote sensing spectral image consist of an array of multidimensional vectors assigned to particular spatial regions (pixel locations), reflecting the response of a spectral sensor at various wavelengths. Formally these images can be described as a matrix $V \equiv (\mathbf{v}_{11}(x^1, y^1), \ldots, \mathbf{v}_{nm}(x^n, y^m))$ where $\mathbf{v}_{ij}(x^i, y^j) \in R^l, i = 1, \ldots, n, j = 1, \ldots, m$ is the vector of spectral information associated with pixel location (x^i, y^j) and the vector components $v_{ijk}(x^i, y^j), k = 1, \ldots, l$ reflects the responses of a spectral sensor at various wavelengths.

In this application all variables (class variable and attributes of the problem) are assumed to be discrete, that is, each variable has a finite set of possible values.

3.1 An Example of Multispectral Data Set Analysis

In the present contribution we consider a LANDSAT TM image from Sierra de Gredos (Spain). This image has been taken from the GIS IDRISI 32 tutorial. LANDSAT TM satellite-based sensors produce images of the Earth in different spectral bands. In this work six bands (bands 1-5 and band 7) are strategically determined for optimal detection and discrimination of water, soil and four different forest type, these are the class values for the classification problem. Band 6 is often dropped from analysis because of the lower spatial resolution. The spectral information, associated with each pixel of a LANDSAT scene is represented by a vector $\mathbf{v}(x, y) \in R^6$, these vectors are the attribute values of the problem.

The GBN, NB and TAN models, have been previously applied by the authors, to the analysis of a multispectral data (see [15]). A GBN multinet model also has been previously applied to this problem (see Ouyang et al. in [13]). In this paper, we applied the GBN multinet, TAN multinet and the TAN-Based BCM multinet models to the classification of multispectral remote sensing images.

The dataset was randomly divided in a training and a test datasets. The above is a classification problem with six attributes and six class values. We apply the different models of bayesian multi-net classifiers to the above classification problem, the training and test accuracy obtained are shown in Table 1. The comparison shows only slight differences between the three bayesian multi-net models. All of them obtain almost 90% of test accuracy with a slight advantage of the TAN and tBCM multi-nets.

Table 1: Training and test accuracy obtained with each classifier in Section 3.1

Classifier	Training	Test
TAN Multi-net	90%	89%
GBN Multi-net	89%	88%
tBCM Multi-net	91%	90%

3.2 An Example of Hyperspectral Data Set Analysis

For some years, the above application has been limited to data of low dimensionality, less than 10 bands (multispectral data). Recent advances in sensor technology makes possible to work with several hundred bands (hyperspectral data). In this paper, we do the novel application of the NB, GBN, FAN and TAN models in the classification of hyperspectral data. The hyperspectral data used in our experiments is a section of a scene taken over northwest Indiana's Pines by the AVIRIS sensor in 1992. It can be downloaded from ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/. Although the AVIRIS sensor collects 224 bands of data, four of these bands contains only zeros and are discarded. The initial 220 bands are reduced to 200 by removing bands covering the region of water absorption: $[104 - 108], [150 - 163], 220$. In this work 200 bands are determined for optimal detection and discrimination of 9 different classes: Corn-no till, Corn-min till, Grass/Pasture, Grass/Trees, Hay-windrowed, Soybean-no till, Soybean-min till, Soybean-clean till and Woods. From the initial 16 land-cover classes, seven were discarded, since only few training samples were available for them. The above is a classification problem with 200 attributes and 9 class values.

We analyze the effectiveness of bayesian networks in classifying hyperspectral images directly in the original hyperdimensional feature space. The dataset was randomly divided in a training and a test datasets. We apply the different models of bayesian network classifiers (NB, TAN, FAN and GBN) to the above classification problem, the training and test accuracy obtained are shown in Table 2. The above problem also has been analyzed using 10-fold cross-validation, the results obtained are very similar to that ones on Table 2 (NB Test Accuracy: $63\% \pm (1.17)$, TAN Test Accuracy: $85\% \pm (1.36)$, FAN Test Accuracy: $86\% \pm (0.86)$, GBN Test Accuracy:

$86\% \pm (1.31)$). In Table 2 the superiority of the FAN, TAN and GBN bayesian network models, for the case of study, is shown. The comparison shows only slight differences between the FAN, TAN and GBN bayesian network models. All of them obtain almost 86% of test accuracy with a slight advantage of the FAN and GBN models.

Table 2: Training and test accuracy obtained with each classifier in Section 3.2

Classifier	Training	Test
NB	64%	62%
TAN	93%	84%
FAN	93%	86%
GBN	89%	86%

4 Conclusion

Bayesian networks appear as powerful tools in hyperspectral remote sensing image classification. Different models of bayesian networks as: Naive Bayes (NB), Tree Augmented Naive Bayes (TAN), Forest Augmented Naive Bayes (FAN) and General Bayesian Networks (GBN), have been applied in the classification of an hyperspectral image. In addition, several bayesian multi-net models: TAN multinet, GBN multi-net and the model developed by Gurwicz and Lerner, TAN-Based Class-Matched multinet (tBCM) are applied in the classification of multispectral data. Feature selection is an important task in remote sensing data processing, particularly in case of hyperspectral images. Actually, we are studying the application of bayesian network models to the classification of hyperspectral data, combined with a band selection method to reduce the dimensionality of the feature space

Acknowledgements: The authors are indebted to the Spanish Ministry of Science and Technology (Project MTM2006-15671) and to the Junta de Comunidades de Castilla-La Mancha (Project PAI-05-044) for partial support.

References:

[1] C. Castillo, J.M. Gutiérrez and A.S. Hadi, *Expert Systems and Probabilistic Network Models,* Springer –Verlag, New York 1997.

[2] J. Cheng, D.A. Bell and W. Liu, An algorithm for Bayesian belief network construction from data, *In proceedings of AI & STAT'97*, 1997, pp. 83–90.

[3] J. Cheng and R. Greiner, Learning Bayesian belief network classifiers: Algorithms and system, *In proceedings of 14th Canadian Conference on Artificial Intelligence*, 2001, pp. 141–151.

[4] C. Chow and C. Liu, Approximating discrete probability distributions with dependence trees, *IEEE Transactions on Information Theory*, 1968, pp. 462–467.

[5] G.F. Cooper and E. Herskovitz, A Bayesian method for the induction of probabilistic networks from data, *Machine Learning* 9, 1992, pp. 309–347.

[6] D. Gacquer, F. Delmotte, S. Delcroix and S. Piechowiak, Comparison of several classifiers for the detection of polluting smokes, *International Conference CIMCA-IAWTIC*, 2006.

[7] R.O. Duda, P.E. Hart and D.G. Stork, *Pattern Classification,* John Wiley & Sons, New York 2001.

[8] N. Friedman, D. Geiger and M. Goldszmidt, Bayesian network classifiers, *Machine Learning* 29, 1997, pp. 131–163.

[9] Y. Gurwicz and B. Lerner, Bayesian class-matched multinet classifier, *Lecture Notes in Computer Science* 4109, 2006, pp. 145–153.

[10] E.J. Keogh and M.J. Pazzani, Learning the structure of augmented bayesian classifiers, *International Journal on Artificial Ingelligence Tools* 11(4), 2002, pp. 587–601.

[11] P. Leray and O. Francois, BNT, Structure learning package: documentation and experiments, *Technical Report. Laboratoire PSI-INSA Rouen-FRE CNRS 2645*, 2004.

[12] K.P. Murphy, BNT, The Bayes Net toolbox for Matlab, *Computing Science and Statistics* 33, 2001.

[13] Y. Ouyang, J. Ma and Q. Dai, Bayesian multinet classifier for classification of remote sensing data, *International Journal of Remote Sensing* 27, 2006, pp. 4943–4961.

[14] J.P., Sacha, *New Synthesis of Bayesian Network Classifiers and Cardiac SPECT Image Interpretation,* Ph.D. Thesis, The University of Toledo 1999.

[15] C. Solares and A.M. Sanz, Bayesian network classifiers. Some engineering applications, *Proceedings of the Ninth IASTED International Conference Artificial Intelligence and Soft Computing*, 2005, pp. 331–335.