Proceedings of the 6th WSEAS Int. Conf. on NEURAL NETWORKS, Lisbon, Portugal, June 16-18, 2005 (pp62-67)

Bayesian Network Classifiers. An Application to Remote Sensing Image Classification

CRISTINA SOLARES[1] and ANA MARÍA SANZ[2]
[1] Department of Applied Mathematics
[2] Department of Geological and Mining Engineering
University of Castilla-La Mancha
13071 Ciudad Real
SPAIN

Abstract:- Different probabilistic models for classification and prediction problems are anlyzed in this article studying their behaviour and capability in data classification. To show the capability of Bayesian Networks to deal with classification problems four types of Bayesian Networks are introduced, a General Bayesian Network, the Naive Bayes, a Bayesian Network Augmented Naive Bayes and the Tree Augmented Naive Bayes. Finally, the novel application of bayesian networks in classification of spectral remote sensing images is shown.

Key-Words:- Classification, Bayesian Networks, Bayesian Network Classifiers, Naive-Bayes, Remote Sensing Image Classification, Prediction, Evidence Propagation.

1 Introduction

Classification and prediction problems 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 (see [1], [2], [4], [7] and [8]).

Several authors have been working with Bayesian Networks classifiers (see [6]). In this work we will do a formal study of the bayesian networks state-of-the-art in classification problems and some experimental results are compared. Different models of bayesian networks are applied to the classification of remote sensing spectral images.

2 Bayesian Networks

A Bayesian network (BN) 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 (see Castillo et al. [1]). 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

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

The importance of Bayesian networks relies in that the calculation of the marginal probabilities of the nodes $p(X_j = x_j)$ or the conditional probabilities $p(X_j = x_j|E = e)$, where E is a set of evidential nodes with known values e, can be easily calculated exploiting the JPD(1)

$$p(x_i) = \sum_{x_j \notin \{x_i\}} \prod_{k=1}^{n} p(x_k|x_1, \ldots, x_{k-1}) \qquad (2)$$

$$p(x_i|E = e) = \frac{\displaystyle\sum_{x_j \notin \{x_i\}, x_j \notin E} \prod_{k=1}^{n} p(x_k|x_1, \ldots, x_{k-1})}{\displaystyle\sum_{x_j \notin E} \prod_{k=1}^{n} p(x_k|x_1, \ldots, x_{k-1})}. \qquad (3)$$

The problem of updating the posterior probabilities of a set of variables of interest whenever a new evidence becomes available is known as

Proceedings of the 6th WSEAS Int. Conf. on NEURAL NETWORKS, Lisbon, Portugal, June 16-18, 2005 (pp62-67)

evidence propagation. Castillo, Gutiérrez, Hadi and Solares [2] and Castillo, Hadi and Solares [3], have been working with different algorithms for the exact and approximate propagation of evidence in Bayesian networks.

2.1 Bayesian Networks as Classifiers

A General Bayesian Network (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 taking values $\{c_1, \ldots, c_s\}$. The resulting model (1) can be used to classify a given set of attributes values $\mathbf{a} = (a_1, \ldots, a_n)$. The vector $\mathbf{a}$ belongs to class $c \in C$ that maximizes the posterior probability

$$\max_{c \in C} p(c|\mathbf{a}). \tag{4}$$

In this article the K2 algorithm (see [5]), search algorithm for finding a high quality Bayesian network in a reasonable time, has been used.

To improve the classification process different bayesian networks models have been developed, such as Naive Bayes, Tree Augmented Naive Bayes and a Bayesian Network Augmented Naive Bayes.

2.1.1 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 Figure 2(a)). The JPD (1) becomes

$$p(a_1, a_2, \ldots, a_n, c) = p(c) \prod_{i=1}^{n} p(a_i|c). \tag{5}$$

The Naive Bayes is one of the most effective classifiers, its predictive performance is competitive with other classifiers that have been developed these last years (see [6]).

2.1.2 Tree Augmented Naive Bayes (TAN)

The very strong assumption of independence of all the attributes $(A_1, \ldots, A_n)$ 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. Which means that the undirected graph is a tree, ie, there is one and only path between any pair of nodes. After that the TAN algorithm adds a link from the classification node C to the attribute nodes $A_i, i = 1, \ldots, n$ (see Figure 2(b)).

2.1.3 General Bayesian Network Augmented Naive (GBAN)

The GBAN algorithm constructs a general bayesian network structure, with the K2 search algorithm, between the attribute nodes and after that adds a link from the classification node C to the attribute nodes $A_i, i = 1, \ldots, n$. The K2 algorithm is an iterative algorithm which adds to each variable parents set Π_i the node that is lower numbered than the variable A_i (all the nodes are ordered) and leads to a maximum increment in the quality measure of the network.

The JPD (1) becomes

$$\begin{aligned} &p(a_1, a_2, \ldots, a_n, c) = \\ &p(c) \prod_{i=1}^{n} p(a_i|c) \prod_{i=1}^{n} p(a_i|\pi_i - c). \end{aligned} \tag{6}$$

2.1.4 Application Example

Consider the "Car Evaluation" database from the UCI machine learning repository, that evaluates cars according to the following attributes: buying price, price of maintenance, number of doors, capacity in terms of persons to carry, size of luggage boot and estimated safety. In Figure 2(a)-(b) the Naive-Bayes and the TAN networks for classification and prediction in the "Car Evaluation" problem are shown. The accuracy of each classifier is based on the percentage of successful predictions. The prediction accuracy (%) and standard deviation of each classifier using CV10 (ten-fold cross validation) are: Naive Bayes (87.38 ± 0.28 (training) and 86.57 ± 2.18 (test)), Tree Augmented Naive Bayes (89.94 ± 0.27 (training) and 87.51 ± 1.9 (test)) and a General Bayesian Network (94.32 ± 0.18 (training) and 94.09 ± 1.71 (test)). In this example a General Bayesian Network with the

Proceedings of the 6th WSEAS Int. Conf. on NEURAL NETWORKS, Lisbon, Portugal, June 16-18, 2005 (pp62-67)

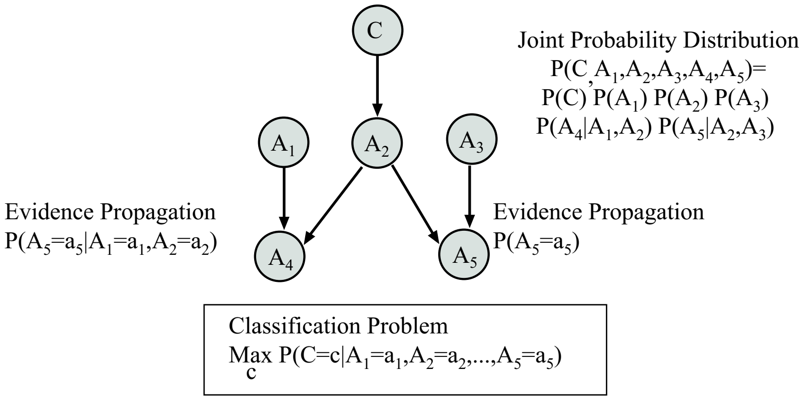

Fig. 1: Example of Bayesian Network for a classification problem.

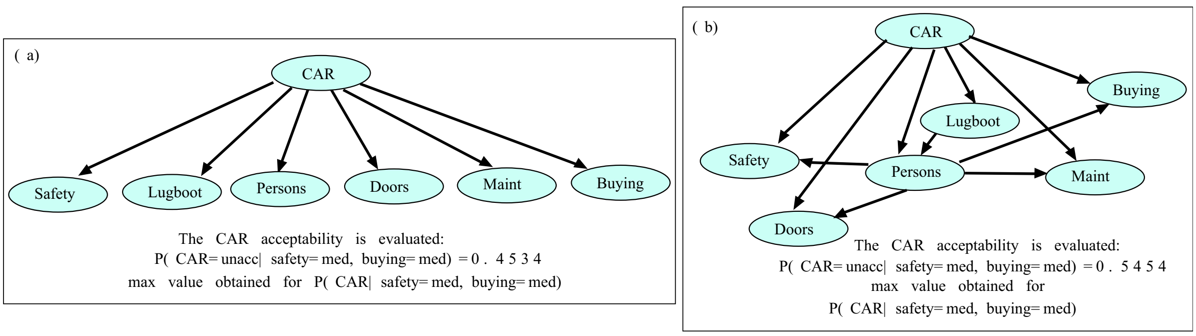

Fig. 2: Naive Bayes (a) and TAN (b) networks for the “Car Evaluation” problem.

K2 search algorithm provide us the best prediction. The running time is less than 3 minutes with all the classifiers. These results have been obtained with the Matlab Bayes net Toolbox [11].

3 Application to Remote Sensing Image Classification

The models of bayesian networks studied in previous sections can be applied on the supervised classification of remote sensing spectral images, the aim of the work is to subdivide the data space into subsets where each subset corresponds to a specific surface covers such as forest, industrial regions, etc.

Several authors have been applying Neural Networks on the classification of remote sensing spectral images (see [9], [10] and [12]).

Remote sensing spectral images 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 $\mathbf{v}_{ij}(x^i, y^j) \in \mathbb{R}^l, i = 1, \ldots, n, j = 1, \ldots, m$ 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. The element $v_{ijk}(x^i, y^j)$ is called the $k-$th image band.

In the present contribution we consider a portion of a LANDSAT TM image (see Figure 3) from a Massachusetts region called Howe Hill, acquired on September 1987. This image has been taken from the GIS IDRISI 32 tutorial. The studied area contains 72×86 pixels (about 557 Ha). LANDSAT TM satellite-based sensors produce images of the Earth in different spectral bands. In this work four bands are strategically determined for optimal detection and discrimination of water, agriculture, urban soil, deciduous and coniferous forest, these are the class values of the classification problem (see Figure 3). The spectral information, associated with each pixel of a LANDSAT scene is represented by a vector $\mathbf{v}(x, y) \in \mathbb{R}^4$, these vectors are the attribute values of the problem, 6192 instances in this example.

Four different models of Bayesian networks (General Bayesian Network with the K2 search algorithm, Naive Bayes, General Bayesian Network Augmented Naive Bayes and Tree Augmented Naive Bayes) have been used to classify the image. The attributes of the classification problem are represented by the bands B_1 to B_4 and the attribute values are the vector $\mathbf{v}(x, y) \in \mathbb{R}^4$ components with the spectral information associated with pixel location (x, y). In Figure 4 the NB network for the remote sensing image classification is shown. The prediction accuracy and standard deviation of each classifier using CV10 (ten-fold cross validation) are shown in Table 1. From Table 1 we found out that the best results are obtained with the GBN and TAN models. In all experiments the running time is less than 7 minutes. An example of evidence propagation using the GBN model is shown in Table 2. The posterior probability $p(C = c|E = e)$, for each value of C, is obtained when new evidence becomes available. In particular, the posterior probabilities $p(C = c|B_1 = 5)$, $p(C = c|B_1 = 5, B_2 = 3)$ and $p(C = c|B_1 = 5, B_2 = 3, B_3 = 3)$ are calculated. When the band values $B_1 = 5$, $B_2 = 3$ and $B_3 = 3$ are known we can conclude that the class value is "water" with probability one.

4 Conclusion

In this article, the Bayesian Networks (General Bayesian Network, Naive Bayes, General Bayesian Network Augmented Naive Bayes and Tree Augmented Naive Bayes) as classifiers, are formally introduced and experimental results are compared. Bayesian Networks appear as powerfull tools in remote sensing image classification.

References:

[1] Castillo, E., Gutiérrez, J.M., *Expert Systems and Probabilistic Network Models*, Springer-Verlag, 1997.

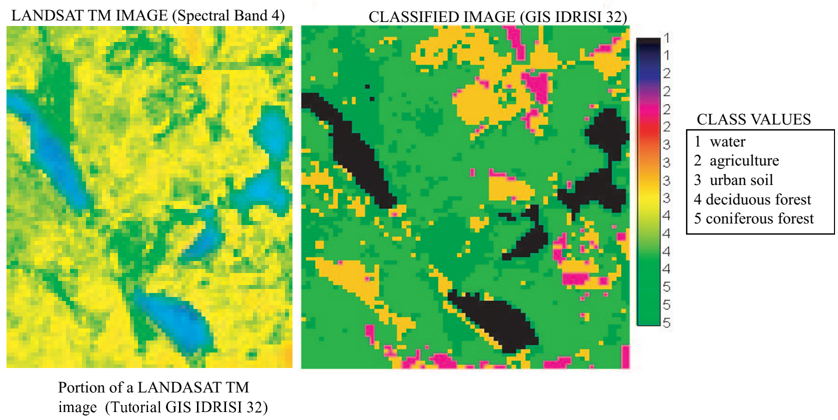

Fig. 3: Fourth band and classified LANDSAT TM image with IDRISI 32.

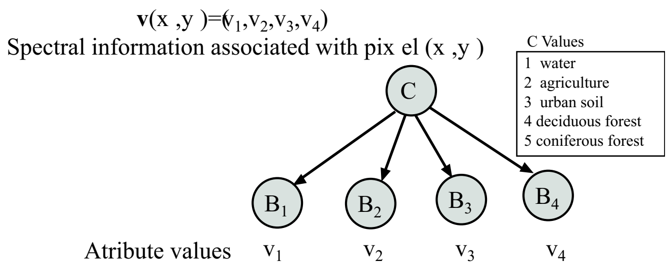

Fig. 4: NB network for the remote sensing image classification.

Table 1: Prediction accuracy (%) and standard deviation with each classifier,

Classifier	Training Acc.	Test Acc.
NB	90.0875 ± 0.11	89.2927 ± 1.09
TAN	93.9994 ± 0.23	93.2026 ± 2.57
GBAN (K2)	92.2409 ± 0.17	87.9035 ± 1.14
GBN (K2)	91.3365 ± 0.18	90.6863 ± 0.39

Table 2: Marginal and posterior probabilities of class C when some band values are known (evidence propagation), the bands are denoted as $B1, B2, B3$,

Class Value:c	$p(C = c)$	$p(C = c \mid B1 = 5)$	$p(C = c \mid B1 = 5, B2 = 3)$
Water	0.0917	0.0917	0.9846
Agriculture	0.0378	0.0010	0.0
Urban Soil	0.1449	0.0078	0.0008
D. Forest	0.5157	0.1204	0.0032
C. Forest	0.2099	0.0987	0.0115

Class Value:c	$p(C = c \mid B1 = 5, B2 = 3, B3 = 3)$
Water	1
Agriculture	0.0
Urban Soil	0.0
D. Forest	0.0
C. Forest	0.0

[2] Castillo, E., Gutiérrez, J.M., Hadi, A.S., Solares, C., Symbolic Propagation and Sensitivity Analysis in Gaussian Bayesian Networks with Application to Damage Assessment, *Journal of Artificial Intelligence in Engineering*, Vol. 11, 1997, pp.173–181.

[3] Castillo, E., Hadi, A.S., Solares, C., Learning and Updating of Uncertainty in Dirichlet models, *Machine Learning*, Vol. 26, 1997, pp.43–63.

[4] Castillo, E., Solares, C., Gómez P., Tail Uncertainty Analysis in Complex Systems, *Artificial Intelligence*, Vol. 96, 1997, pp.395–419.

[5] Cooper, G.F., Herskovitz, E., A Bayesian Method for the Induction ol Probabilistic Networks from Data , *Machine Learning*, Vol. 9, 1992, pp.309–347.

[6] Friedman, N., Geiger, D., Goldszmidt, M., Bayesian Network Classifiers, *Machine Learning*, Vol. 29,1997, pp.131–163

[7] Gámez, J.A., Moral, S.,Salmerón, A., Advances in Bayesian Networks, *Springer*, 2004.

[8] Gutiérrez, J.M., Cano, R.,Cofiño, A.S., Sordo, C.M., Redes Probabilísticas y Neuronales en las Ciencias Atmosféricas, *Series Monográficas, Ministerio de Medio Ambiente*, 2004.

[9] He, H., Collet, C., Combining Spectral and Textural Features for Multispectral Image Classification with Artificial Neural Networks, *International Archives of Photogrammetry and Remote Sensing* , Vol. 32, Part 7-4-3 W6, 1999

[10] Liu, J., Shao, G., Zhu, H., Liu, S., An Neural Network Approach for Information Extraction from Remotely Sensed Data, *Proceedings of the International Conference on Geoinformatics* , 2004, pp.655–662

[11] Murphy, K.P. The Bayes net Toolbox for Matlab, *Computing Science and Statistics*, Vol. 33, http://www.cs.berkeley.edu/murphyk/Bayes/bnt.html

[12] Villmann, T., Erzsébet, M., Hammer, B., Neural maps in remote sensing image analysis, *Neural Networks*, Vol.16, 2003, pp.389–403

Acknowledgments:
The authors are indebted to the Spanish Ministry of Science and Technology (Project BFM2003-05695) for partial support.