TEXTURE CLASSIFICATION OF AERIAL IMAGE BASED ON BAYESIAN NETWORK AUGMENTED NAIVE BAYES

YU Xin [a,b] ZHENG Zhaobao [b] ZHANG Haitao [a] YE Zhiwei [c]

[a] Photogrammetry and Remote Sensing Center, Beijing Instititute of Surveying and Mapping,Beijing, 100038, China, - china_yuxin@yahoo.com.cn
[b] School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road,Wuhan 430079, China
[c] School of Computer Science, Hubei University of Technology, Wuhan 430068, China

KEY WORDS: Texture, Classification, Pattern, Recognition, Aerial, Image

ABSTRACT:

Classification is an open, old and basic problem in many domains. Recently, a lot of new methods come forth, such as Bayesian Networks. Bayesian Networks, one of probabilistic networks, are a powerful data mining technique for handling uncertainty in complex domains. In this paper, we apply Bayesian Networks Augmented Naive Bayes (BAN) to texture classification of aerial image and propose a new method to construct the network topology structure in terms of training accuracy based on the training samples. In order to validate the feasibility and the effectivity, we compare BAN to Naive Bayes Classifiers (NBC) and PCA-NBC. Thus six pieces of 23cm×23cm aerial image about Australia and ten pieces of 23cm×23cm aerial image about Wuhan city in China are used in the experiments. Experimental results demonstrate BAN outperform than NBC and PCA-NBC in the overall classification accuracy. Although it is time consuming, it will be an attractive and effective method in the future.

1. INTRODUCTION

Classification is a basic task in data mining and pattern recognition that requires the construction of a classifier, that is, a function that assigns a class label to instances described by a set of features (or attributes) (Friedman, N., 1997). Learning accurate classifiers from pre-classified data has been a very active research topic in the machine learning. In recent years, numerous approaches have been proposed such as Fuzzy sets, Rough sets, Neural Network, Support Vector Machine and Genetic Algorithms, Ant Behavior Simulation, Case-based Reasoning, Bayesian Networks etc.

Early in 1988, Pearl et al. had provided the concept of Bayesian Networks. It is a powerful tool for knowledge representation and inference under conditions of uncertainty. However, Bayesian Networks were not considered as classifiers until the discovery that Naive Bayesian Network, a very simple kind of Bayesian Networks that assumes the features are independent given the class attribute (node), are surprisingly effective (Langley, P ,1992). From that time on, it triggered experts to explore more deeply into Bayesian Networks as classifiers. Since the "Naive" independent assumption in Naive Bayesian Network can not be hold in many cases, researchers have wondered whether the performance will be better if we relax the strong independent assumption among features (or variables) (Yu Xin, 2005; D.Heckerman, 1995). Thus, this paper puts up a new method, to construct the topology structure of Bayesian Network Augmented Naive Bayes (BAN), and it can resolve the forenamed problem, because it allows arbitrary relation (arc) among features, which can be obtained in terms of training accuracy based on training data (samples). In addition, we apply BAN, to the texture classification of aerial image.

This paper is organized as follows. In section 2, we review some basic concept of Bayesian Networks and then we introduce the mathematic model and inference of BAN in detail in section 3. Then in section 4 we examine a straightforward application of BAN for texture classification of aerial image. Finally in section 5 we describe the experimental setup and results and draw some conclusions.

2. BAYESIAN NETWORKS FOR TEXTURE CLASSIFICATION OF AERIAL IMAGE

In this section, we briefly introduce some basic concepts related to Bayesian Networks and then apply it to texture classification of aerial image.

2.1 Bayesian Networks

Bayesian Networks have proved to be an effective knowledge representation and inference engine in artificial intelligence and expert systems. It is a graph, in which the features (nodes) represent the variables and the arcs represent a causal (or probabilistic) relationship among the connected variables. From Figure 1 we can see that Bayesian Network topology structure is a directed acyclic graph, provides a graphical view of variables' relationships and Bayesian Network parameters specify how a node depends on its parents with quantitative probability. If the node $X_i \to X_j$, it means that the node X_i has direct influence upon the node X_j. Further more, the node X_i is the parent node of the node X_j and by contraries the node X_j is the child node of the node X_i, as a rule we denote the parent set of the node X_i by $P_a(X_i)$. Consequently, we can get the joint probability of all the nodes (variables) based on the definition of Bayesian Networks.

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid Pa(X_i)) \quad (1)$$

Usually, Bayesian Network is accustomed to naming Bayesian Networks Classifiers in the classification domain (Cheng, J., 1999). In order to relax the independent assumption in Naive Bayes Classifiers (NBC), researchers proposed Tree Augmented Naive Bayes Classifiers (TAN), which extends Naive Bayes by allowing the nodes (features) to form a tree. Whereas the tree topology structure is still confined and can not describe the inherent relation among features. Thus in this paper, we apply Bayesian Network Augmented Naive Bayes (BAN) to classification and BAN classifiers extend Tree Augmented Naive Bayes (TAN) classifiers by allowing the features (attributes) to form an arbitrary graph rather than just a tree (Friedman, N., 1997).

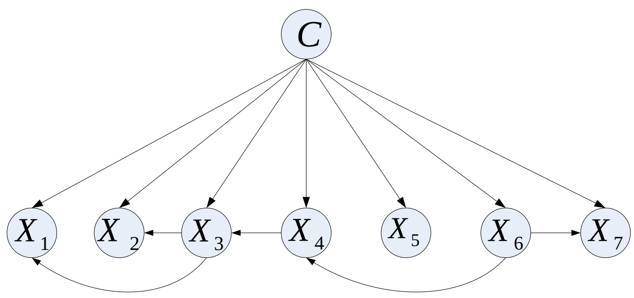


Figure 1. An Example of BAN Applied In the Classification

Figure 1 is an example of BAN applied in the classification. The node C denotes the class label (attribute) and $X_1, \ldots, X_6$ and X_7 denote the texture feature that are extracted from image classification unit. Thus, according to the above definition of Bayesian networks, we can get

$$P_a(C) = \varnothing, P_a(X_1) = \{C, X_3\}, P_a(X_4) = \{C, X_6\}.$$

Bayesian Networks learning involves topology structure learning and parameters learning (Cheng, J., 2001). In fact, learning the topology structure is to get some relationship among the features and the parameters learning is to estimate the parameters of the assumed probability density (distribution) from the training samples with known class label. However, structure learning is more difficult than parameters learning (D.Heckerman, 1997), which has been an open problem since Bayesian Networks occur. Therefore, in order to induct a Bayesian Network from data, researchers proposed a variety of methods. There are two kinds of methods to learn the topology structure. One is the scoring-based learning algorithm (Jiebo Luo, 2005), that find one certain structure that maximizes the Bayesian, MDL or Kullback-Leiber (KL) entropy scoring function (D.Heckerman, 1995) and the other is CI-based algorithm (Yu Xin, 2007) (the conditional independent test such as Chi-squared test and mutual information test). For the first type, the performance of learning algorithm depends not only on the score function but also on the search method. Moreover, the node ordering is necessary. Nevertheless, in the classification domain, it is extremely difficult or unmeaning to achieve the features (nodes) ordering in advance, because unlike in the diagnosis problem, there are distinct causal relationships among nodes. For example, cold (node) may bring on a fever (node), so in this case, as a matter of much experience in fact, the nodes ordering is easy to get.

This paper proposes a new method to acquire the topology structure of BAN in terms of the training accuracy based on training samples. Firstly one certain topology structure is given as initial structure (state) and then estimate the relevant parameters based on training samples. And we regard the training samples as test samples to test (or classify) and get the overall accuracy of the training samples, which is named the training accuracy. Then search all possible network topology structures and get the corresponding training accuracy. Among them, there exists the maximum training accuracy, whose topology structure is considered as the best one for fitting the training samples.

In Figure 1, we can notice that except the class node, the other nodes have more than one node as parent nodes. In this case, the probabilistic reasoning (computation of the joint probability) is more complicated and difficult than the case in the NBC (Yu Xin, 2006).

2.2 Mathematic Model and Inference of BAN

Suppose X_s is one arbitrary node in some BAN and for sake of simplicity the parent set of X_s is denoted as X_p (i.e. $Pa(X_s)$). Generally, we use a capital letter, like X to denote a random variable and a particular value of a random variable will be denoted with a lower case letter, in this case x.

Ordinarily, it is reasonable to assume that X_s and X_p follow multi-normal distribution. In addition, X_s and X_p can constitute a new $(1+n)$-dimensional normal random vector X with corresponding mean vector μ_X and covariance matrix D_X.

$$X_s \sim N(\mu_s, D_{ss}), X_p \sim N(\mu_p, D_{pp}) \quad (2)$$

$$\underset{(n+1)\times 1}{X^{\mathrm{T}}} = [\underset{1\times 1}{X_s} \quad \underset{n\times 1}{X_p}] \quad (3)$$

$$\mu_X = \begin{bmatrix} \mu_s \\ \mu_p \end{bmatrix} = \overline{X},$$

$$D_X = \begin{bmatrix} D_{ss} & D_{sp} \\ D_{ps} & D_{pp} \end{bmatrix} = \frac{1}{n}(X - \mu_X)(X - \mu_X)^{\mathrm{T}} \quad (4)$$

The joint-normal probability density (also called the multi-normal or multivariate normal distribution) can be specified in following ways (Cui,Xizhang, 2001).

$$f(X)=(2\pi)^{-\frac{n+1}{2}}\times|D_X|^{-\frac{1}{2}}\times\exp\{-\frac{1}{2}\begin{bmatrix}X_s-\mu_s\\ X_p-\mu_p\end{bmatrix}^{\mathrm{T}}D_X^{-1}\begin{bmatrix}X_s-\mu_s\\ X_p-\mu_p\end{bmatrix}\} \quad (5)$$

According to Bayes' rule, we can get the conditional density of

X_s given $X_p=x_p$

$$f(X_s\mid x_p)=(2\pi)^{-\frac{n+1}{2}}\times|\tilde{D}_{ss}|^{-\frac{1}{2}}\times\exp\{-\frac{1}{2}\begin{bmatrix}X_s-\tilde{\mu}_s\end{bmatrix}^{\mathrm{T}}\tilde{D}_{ss}^{-1}\begin{bmatrix}X_s-\tilde{\mu}_s\end{bmatrix}\} \quad (6)$$

Where $\tilde{\mu}_s$ and $\tilde{D}_{ss}$ denote the conditional mean and covariance of X_s given $X_p=x_p$ respectively.

$$E(X_s\mid x_p)=\tilde{\mu}_s=\mu_s+D_{sp}D_{pp}^{-1}(x_p-u_p) \quad (7)$$

$$D(X_s\mid x_p)=\tilde{D}_{ss}=D_{ss}-D_{sp}D_{pp}^{-1}D_{ps} \quad (8)$$

Then the conditional probability of X_s given $X_p=x_p$ can be computed by

$$P(X_s\mid x_p)=f(X_s\mid x_p)\times dX_s \quad (9)$$

Where dX_s denotes extremely small step and is regarded as one constant in the computation.
Thus, based on formula (1) and (9), we can compute the joint probability of all variables by the following formula.

$$P(X_1,X_2,\ldots,X_n,C_i)=P(C_i)\prod_{s=1}^{n}P(X_s\mid X_p) \quad (10)$$

Where C_i denotes the class label (variable) and $i=1,\cdots,m$ represents the relevant class.
In terms of the conditional probability formula, we can have

$$P(C_i\mid X_1,\ldots,X_n)=\frac{P(X_1,\ldots,X_n\mid C_i)P(C_i)}{P(X_1,\ldots,X_n)}=\frac{P(X_1,\ldots,X_n,C_i)}{P(X_1,\ldots,X_n)} \quad (11)$$

As $P(X_1,X_2,\ldots,X_n)$ is a constant and is independent of C_i, so we have

$$P(C_i\mid X_1,X_2,\ldots,X_n)\propto P(X_1,X_2,\ldots,X_n,C_i) \quad (12)$$

In the final, when posterior probability $P(C_i\mid X_1,X_2,\ldots,X_n)$ is maximal, we can decide the class C^* will be $\max_i\{P(C_i\mid X_1,X_2,\ldots,X_n)\}$.

2.3 Texture Extraction and Description

Texture has played a role in image analysis and image classification for a long time. So far, many approaches have been put up in the classification domain. Commonly, these approaches can be split into two types: (1) structural (transform-based) texture features and (2) statistical texture features in the experiment.

(1) Skewness statistics, information entropy, and inverse difference moment based on the gray co-occurrence matrix, are denoted as X_1 、 X_2 and X_3 respectively.

(2) the mean of LL sub-image, and standard deviation of LH sub-image and HL sub-image at the first decomposition level through the Symlets wavelet transform (Yang S, 2002)and fractal feature, are denoted as X_4 、 X_5 、 X_6 and X_7 respectively.

2.4 The Classification Scheme bstract

In order to explain BAN applied in the texture classification of aerial image in detail, the complete classification scheme is summarized below.

① Training samples (images) and testing samples of each class are randomly chosen from the whole database;
② Extract seven kinds of texture features from each classification unit;
③ Arbitrarily select some network topology structure (like Figure 1) as the initial structure (state) and estimate the parameters $\tilde{\mu}_c$ and $\tilde{D}_{cc}$ of each class (in the formula (6)) based on the training samples by the formula (7) and (8). And then the training samples are regarded as the testing samples to be classified and the initial training accuracy is obtained.
④ Search all possible network topology structure and learn relevant parameters based on the training samples. And then we can get the corresponding training accuracy of the different topology structures;
⑤ Single out the topology structure of the best training accuracy as the final training results;
⑥ In terms of the training results, we can compute the posterior probability of a new (unknown class label) sample X belonging to each class C_i, $P(C_i\mid X_1,\ldots,X_n),1\le i\le n$, by virtue of the formula (9) and (1), where n denotes the

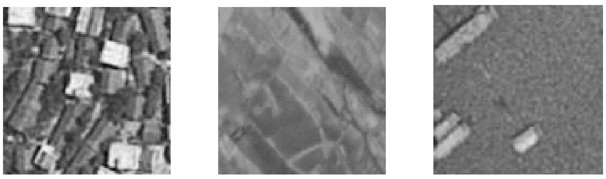
number of class. In the end, the sample X is labeled as belonging to the class C^* if $C^* = \max_i \{P(C_i \mid X_1, \cdots, X_n)\}$;

⑦ Statistical analysis of classification results.

3. EXPERIMENTS

3.1 Database for classification experiments

To validate the feasibility and effectivity of BAN applied in the texture classification of aerial image, six pieces of 23cm×23cm aerial image about Australia and ten pieces of 23cm×23cm aerial image about Wuhan city in China are used in the experiments. In fact, these images are segmented into 465 small areas (i.e. image classification unit) and grouped into three classes (or groups). The first class, residential area (a), has 167 images, including images D1 through D167; the second class, paddy field (b), contains 144 images, including images D168 through D311; the third class, water area (c), contains 154 images, including images D312 through D465. Whereas, among them the maximal area is 40×40 pixels and the minimum area is 16×16 pixels. One sample image of each class is shown in Figure 2.



(a) (b) (c)

Figure 2. A Sample Image of Each Class

3.2 Comparison experiments

The classification accuracy is calculated from the confusion matrix, which contains information about the correct classification and misclassification of all classes. To evaluate the efficiency of BAN, classification results were calculated based on BAN, PCA-NBC and Naive Bayes Classifiers (NBC) in terms of overall classification accuracy. The experimental results are shown in Table 1. And Figure 3 is the network topology structure of BAN when the training accuracy is best and the number of training samples is 50.

N	30	35	40	45	50	mean	std
BAN	86.0	85.4	85.6	88.2	88.6	86.6	0.01
NBC	66.5	66.3	67.5	67.1	67.1	67.6	0.03
PCA-NBC	84.1	84.3	85.8	86.0	86.0	85.1	0.01

Table 1 The Comparison results of Three Methods

Table 1 displays comparison of the accuracy among three methods in the condition of different training samples, and N denotes the number of training samples of each class. The best mean overall classification accuracy is 86.6% (BAN) and at the same time the fluctuation (std, standard deviation) of BAN is no better than that of PCA-NBC. As expected, BAN gives better classification results than PCA-NBC and NBC. For the sake of intuitionistic vision effect, it is also depicted in Figure 4, where the horizontal axis denotes the number of training samples of each class and the vertical axis denotes the overall classification accuracy. Obviously, BAN is the best one and NBC is the worst one among them in the experiments.

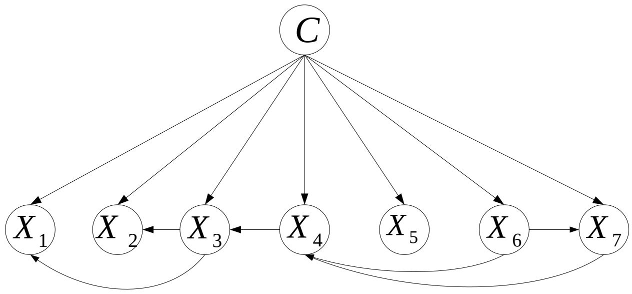


Figure 3. The Topology Structure of BAN

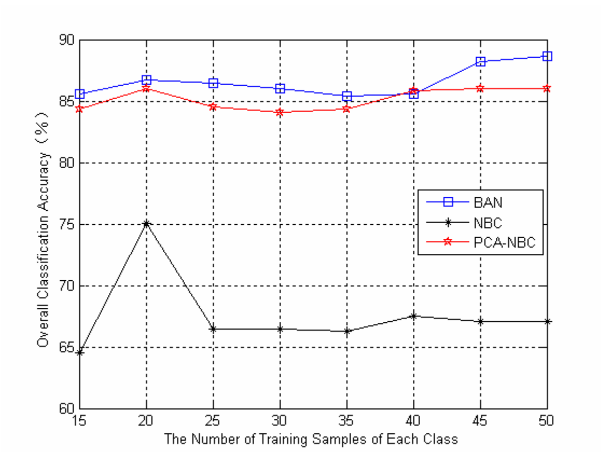


Figure 4. The Curve of the Overall Classification Accuracy

4. CONCLUSIONS AND FUTURE WORK

Bayesian Network is a directed acyclic graphic model and a powerful probabilistic representation. In this paper, Bayesian Network Augmented Naive Bayes (BAN) is used for texture classification of aerial image. Experimental results show that BAN outperforms than NBC and PCA-NBC.

However, learning the topology structure needs a great deal of time and furthermore how to explain the factual (causal or probabilistic) relationship like in the diagnosis domain is difficult, which will be our future work.

ACKNOWLEGEMENTS

This paper is financially supported by NSFC (No. 40571102 and 40271094). The authors wish to thank the anonymous reviewers for the comments and suggestions on this paper.

REFERENCES

Friedman, N., Geiger, D., Goldszmidt, M.,1997. Bayesian Network classifiers, Machine Learning, vol.29, pp.131~163.

Langley,P. Iba,W. and Thompson,K.,1992. An analysis of Bayesian classifiers. In proceeding of AAAI-92, pp.223-228.

Yu Xin, Zheng Zhaobao, LI Linyi, Ye ZhiWei.,2005. Texture Classification of aerial image based on PCA-NBC, Proceedings of SPIE - The International Society for Optical Engineering, v 6044, MIPPR 2005: Image Analysis Techniques.

D.Heckerman, D.Geiger, D.M.Chickering,1995. Learning Bayesian networks: the combination of knowledge and statistical data, Machine Learning 20, pp.197~244.

Cheng, J., Greiner, R.,1999. Comparing Bayesian Network Classifiers, In UAI-99.

Cheng, J. & Greiner, R.,2001. Learning Bayesian Belief Network Classifiers: Algorithms and System. Proc.14th Canadian Conference on Artificial Intelligence.

David Heckerman.,1997. Bayesian Networks for Data Mining. Data Mining and Knowledge Discovery 1, pp.79~119.

Jiebo Luo, Andreas E.Savakis, Amit Singhal.,2005. A Bayesian network-based framework for semantic image understanding. Pattern Recognition 38, pp.919~934.

Yu Xin, Zheng Zhaobao, YE Zhiwe, TIAN Liqia.,2007. Texture Classification Based on Tree Augmented Naive Bayes Classifier, Geomatics And Information Science of Wuhan University, Vol.32, No.4., p. 287~289.

Yu Xin, Zheng Zhaobao, Tang Ling, Ye Zhiwei.,2006. Aerial Image Texture Classification Based on Naive Bayesian Network, Geomatics And Information Science of Wuhan University, Vol.31, No.2., p. 108~111.

Cui,Xizhang, Yu,Zhongchou et al.,2001. Generalized adjustment of Survey Measurements. Wuhan: Wuhan University Press.

Yang S., Chang K.,2002. IEEE Transactions on Systems, Man and Cybernetics-Part A: Systems and Humans, Vol.32, No.3, pp.419~428.

Bart B.,Geert V.,2004. Bayesian network classifiers for identifying the slope of the customer lifecycle of long-life customers, European Journal of Operational Research, vol.156,pp.508~523.