Bayesian Network Structure Learning and Inference in Indoor vs. Outdoor Image Classification

Michael J. Kane
Space Systems
ITT
Rochester, New York
kane@csh.rit.edu

Andreas Savakis
Computer Engineering Department
Rochester Institute of Technology
Rochester, New York
savakis@mail.rit.edu

Abstract

Bayesian network model selection techniques may be used to learn and elucidate conditional relationships between features in pattern recognition tasks. The learned Bayesian network may then be used to infer unknown node-states, which may correspond to semantic tasks. One such application of this framework is scene categorization. In this paper, we employ low-level classification based on color and texture, semantic features, such as sky and grass detection, along with indoor vs. outdoor ground truth information, to create a feature set for Bayesian network structure learning. Indoor vs. outdoor inference may then be performed on a set of features derived from a testing set where node states are unknown. Experimental results show that this technique provides classification rates of 97% correct, which is a significant improvement over previous work, where a Bayesian network was constructed based on expert opinion.

1. Introduction

Scene classification is an important area of research in computer vision. Given an arbitrary digital image, we would like to automate detection of the type of semantic scene it depicts. Applications of this process include automatic albuming [7], and multimedia databases that are searchable based on high-level scene characteristics. Scene classification was initially approached using low-level features. Szummer and Picard [12] extracted color features based on the RGB histogram, Ohta histogram, and texture features using multiresolution simultaneous autoregressive models (MSAR) or discrete cosine transform (DCT) coefficients for given image sub-blocks and used a k-nearest neighbor classifier followed by majority decision to determine whether the image was indoor or outdoor. Serrano et al [11] used wavelet features to represent texture and support vector machine (SVM) classifiers to obtain improved low-level classification results with reduced computational complexity.

The use of semantic features was proposed by Paek [8] and Luo and Savakis [5, 6]. This work used näive Bayesian networks for knowledge representation and inference. In [6], features were formed by using a k-nearest neighbor classifier, and in the case of sky and grass semantic features, low-level features were used to derive these probabilities.

In this paper, we propose the use of network structure learning for scene classification. By specifying a Bayesian network for scene classification, a consistent method for inference is provided along with a framework which allows one to easily compensate for the addition of new features as opposed to other techniques, like neural network training, which requires relearning the structure with the entire training set.

Bayesian network structure learning methods generally fall into three categories: (i) asymptotically correct methods, (ii) gradient based methods, and (iii) model refinement methods combining both (i) and (ii). Asymptotically correct algorithms analyze dependency relationships between nodes and it can be shown that the derived structure approaches the optimal asymptotically, as the number events in the database increases. One such algorithm, proposed by Chen [1], uses mutual information between nodes as a criterion for parent-child relationships.

Gradient based methods use a scoring function to induce a search space on the space of possible networks. A local optimum is then found using techniques such as a gradient search [4], or particle swarm optimization [10]. In the selection of an appropriate scoring technique, it is desirable to select a fitness function that mitigates expectation maximization of a structure with the probability of the underlying event. One such method models state probabilities as

Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)
1051-4651/04 $ 20.00 IEEE

Dirichlet distributions and uses this information to score local structures [3, 9].

The model refinement approach to structure learning attempts to alleviate the computational difficulty involved in gradient based methods by using as an initial estimate a computationally efficient, although generally less accurate, asymptotically correct approach. This way the node conditional relationships are initialized, and a gradient approach is used to refine the initial network. This technique has been shown to reduce the computational complexity incurred by a bottom-up gradient search while retaining its potential accuracy [4].

In this paper, we use gradient based model selection to determine the structure of a Bayesian network that provides for the knowledge representation and inference of the indoor vs. outdoor scene classification problem. This approach significantly improves classification performance over using expert opinion for model determination. In the following sections, we present an overview of the model selection technique employed, an overview of the inference technique used, results using this technique, a comparison with past results, and we propose a system for accurate indoor vs. outdoor scene classification.

2. The Search Space

In this paper, feature selection and initial color and texture classification is based on the methods used in [5]. Four feature extraction techniques, namely, color and texture classification as well as blue sky and grass detection, were used on each image. Color features are based on the quantized color histogram in the Ohta color space. Texture features are based on three levels of Multiresolution Simultaneous Autoregressive Model coefficients [12]. Both sky and grass classification features are based on color and texture features [5].

After the feature extraction stage, k-nearest neighbor classification measures are extracted, normalized and quantized to the nearest 10% for a total of 11 possible states per feature node. The total number of possible feature instantiations is taken to be the number of nodes multiplied by the number of states per node, or 29,282 possible node instantiations.

The total number of possible directed acyclic graph given the number of nodes is given in [2] as:

$$f(n) = \sum_{i=1}^{n} (-1)^{(i+1)} \frac{n!}{(n-i)!i!} 2^{i(n-i)} f(n-i) \qquad (1)$$

where n is the number of nodes in the network. In the present case of a five node network, there is potential for 29,281 possible graph structures.

3. Structure Learning

As we are dealing with a relatively small network, it is not necessary to resort to the model refinement technique. Instead, a gradient based method is adequate for the problem at hand. The fitness function used was the Data Given Model Probability (DGM) [3, 9] and is given by the equation:

$$p(D|M) = \prod_{i=1}^{I} \prod_{j=1}^{q_i} \frac{(c_i - 1)! \prod_{k=1}^{c_i} n(x_{ik}|\pi_{ij})!}{(c_i + n(\pi_{ij}) - 1)!} \qquad (2)$$

where c_i is the number of states of the ith node, x_{ik} is the ith node with instantiation k and π_{ij} denotes the parent nodes of the ith node with node instantiation j.

4. Inference

The inference technique used in this paper differs from the one used in previous work. In [5], the goal was to create an outdoor image detector. For a given image, an indoor classification is implied if it has a low probability of being outdoor. This was accomplished by choosing a threshold, such that outdoor node values below the threshold would correspond to an indoor image and values above the threshold would correspond to an outdoor image.

The technique employed in this paper uses Bayes' rule to compute the posterior probability of the indoor/outdoor states. This is accomplished by first realizing that the joint probability of the network is given by:

$$\begin{aligned} &Pr(IO, C, T, B, G) = \\ &\quad Pr(IO) \cdot Pr(C|IO) \cdot Pr(T|IO) \\ &\quad \cdot Pr(B|IO) \cdot Pr(G|IO) \end{aligned} \qquad (3)$$

Then an indoor vs. outdoor probability is computed using the equation:

$$\begin{aligned} &Pr(IO = o|C = c, T = t, S = s, G = g) = \\ &\quad \frac{Pr(IO = o, C = c, T = t, S = s, G = g)}{\sum_{io} Pr(IO = io, C = c, T = t, S = s, G = g)} \end{aligned} \qquad (4)$$

5. Results

The feature extraction techniques employed in this experiment were used on a Kodak database of 1308 consumer images. The feature probabilities for 654 randomly selected images were used for training and the remaining 654 were used in the inference experiment. The experiment was repeated 8 times to cross-validate the outcome.

The model derived using the Data Given Model Probability technique is shown in Figure 1. This is a single level

Bayesian network, which does not contain the intermediate nodes that are present in the network generated by expert opinion in [5]. Results the testing half of the database are shown in Table 1. These results are using the described inference technique on the model selection structure. For comparison, Table 2 shows results from previous work obtained using expert opinion. The results illustrate that the proposed model selection procedure for indoor vs. outdoor scene classification outperforms the previous method by approximately 12%.

6. Error Estimation

Along with inference, the posterior probabilities can also be used on the training set for error estimation. For a given event, we can use the error estimated by expectation maximization along with the frequency of the event to obtain an estimate of how often our inference will be incorrect. The expected error rate can then be given by:

$$\sum_{c,t,s,g} (1 - Pr(Inf(IO|...)|C = c, T = t, \\ S = s, G = g)) \cdot n(C = c, T = t, \\ S = s, G = g) \tag{5}$$

where $Inf(IO|...)$ is the inferred state of the network give the feature probabilities and $n(...)$ is the number of events in the training set where the states are as given by the arguments. It should be noted that this equation gives the expected rate per the number of events in the database. By normalizing by the number of events of the databases, we can obtain the error rate expressed as a percentage of the total number of training images.

By employing this technique, a resulting error rate of 18.6 per 654 images is found where the actual error rate is 16 per 654 images. Normalizing, we get an expected error rate of 2.8% where the actual is 2.7%. This error rate matches very well with the experimental results presented in Table 1.

7. Conclusion

We propose the use of model selection for model determination and posterior probabilities for inference of indoor vs. outdoor classification. The results obtained with the proposed method yield significant improvement in inference accuracy when compared with previous research. By using model selection, a structure is selected that accurately encodes the conditional dependencies between network nodes, given a training database.

At the same time, by computing posterior probabilities for inference, the need to select an arbitrary threshold for classification is alleviated. The model is able to justify the inference in a purely probabilistically manner.

One avenue not explored is the use of intermediate nodes. In the presented scene classification application, lack of intermediate nodes does not adversely affect performance or generalization. However, in some cases lack of intermediate nodes, when correlation exists between leaf nodes, may result in poor generalization. Therefore, this type of automatic model selection technique should be used prudently and the resulting network should be checked and cross validated before using it in general situations.

Another area of interest is the direct extraction of feature probabilities directly, rather than using k-nearest neighbor classification to determine the feature/semantic intersection probability. By removing the feature statistics from the semantic task, it seems likely that a single feature node will be useful for two, potentially independent, semantic tasks.

References

[1] J. Chen, D. Bell, and W. Liu. An Algorithm for Bayesian network construction from data. *Proc. of the Sixth ACM International Conference on Information and Knowledge Management*, 1997.

[2] R. Cowell et. al. *Probabilistic Networks and Expert Systems*. Springer-Verlag, New York, NY, 1999.

[3] D. Heckerman A tutorial on learning with Bayesian networks. *Technical Report MSR-TR-09-06*, 1997.

[4] M. J. Kane, F. Sahin, A. Savakis. A Two Phase Approach to Bayesian Network Model Selection and Comparison between the MDL and DGM Scoring Heuristics. *IEEE Int. Conf. on Systems, Man and Cybernetics*, October 2003.

[5] J. Luo and A. Savakis. Indoor vs. Outdoor Classification of Consumer Photographs. *Int. Conf. Image Proc. ICIP'01*, October 2001.

[6] J. Luo, A. Savakis, S. Etz, and A. Singhal. On the Application of Bayes Networks to Semantic Understanding of Consumer Photographs. *Int'l Conf. Image Proc. ICIP'00*, 2000.

[7] A. C. Loui and A. Savakis. Automatic Image Event Segmentation and Quality Screening for Albuming Applications. *Int'l Conf. on Multimedia and Expo*, 2000.

[8] S. Paek et. al. Integration of Visual and Text Based Approaches for the Content Labeling and Classification of Photographs. *ACM SIGIR'99 Workshop on Multimedia Indexing and Retrieval*, August, 19, 2000.

[9] M. Ramoni and P. Sebastiani. Learning Bayesian Networks from Incomplete Databases. *Conf. on Uncertainty in Artificial Intelligence*, 1995.

[10] F. Sahin, J. C. Tillet, R. M. Rao. A Particl Swarm Optimization Approach to Structural Bayesian Network Learning for Large Datasets.

[11] N. Serrano, A. Savakis, and J. Luo. A Computationally Efficient Approach to Indoor/Outdoor Scene Classification. *Int'l Conf. on Pattern Recognition*, August, 1995.

[12] M. Szummer and R. W. Picard. Indoor-Outdoor Image Classification. *IEEE Int'l Workshop on Content-Based Access of Image and Video Databases, ICCV '98*, 1998.

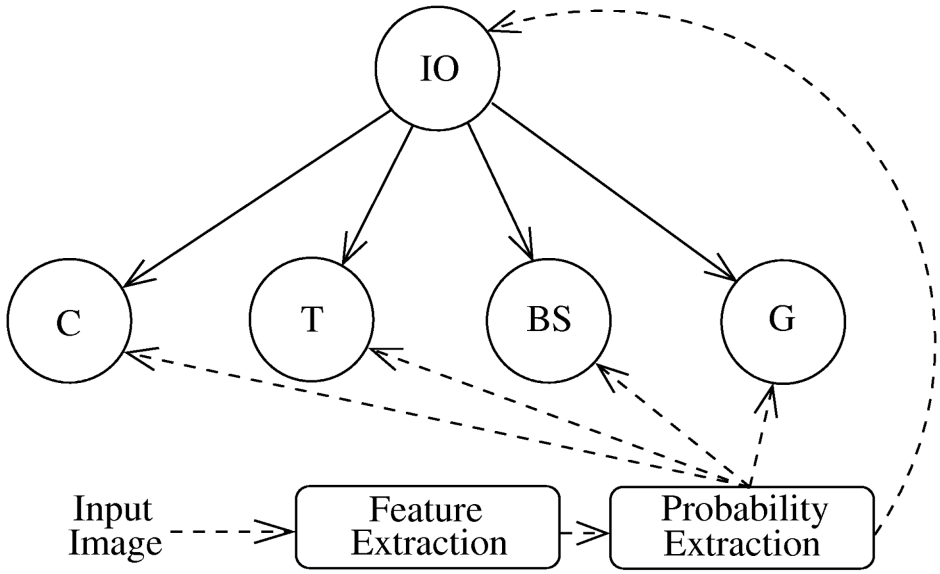

Figure 1. Flow Graph for Indoor vs. Outdoor Scene Classification where: IO is the indoor/outdoor node; C is the result of color classification; T is the result of texture classification; BS is the detection of blue sky; and G is the detection of grass.

Table 1. Indoor vs. Outdoor Classification using Computed Semantic Feature Model Selection

Indoor vs. Outdoor Classification using Computed Semantic Feature Model Selection			
	Correct	Incorrect	Percent Correct
Indoor	288	9	97.0%
Outdoor	350	7	98.0%
Overall	638	16	97.3%

Table 2. Indoor vs. Outdoor Classification using Expert Opinion

Indoor vs. Outdoor Classification using Expert Opinion			
	Correct	Incorrect	Percent Correct
Indoor	519	96	84.4%
Outdoor	589	104	85.0%
Overall	1108	200	84.3%