EXPERIMENTS IN CONSTRUCTING BELIEF NETWORKS FOR IMAGE CLASSIFICATION SYSTEMS

Seungyup Paek and Shih-Fu Chang

Columbia University, 1306 S.W. Mudd, 500 W. 120th St., Department of Electrical Engineering
New York City, N.Y. 10027, U.S.A.
syp@ee.columbia.edu, sfchang@ee.columbia.edu

ABSTRACT

We present procedures and experimental results in constructing belief networks for image classification systems based on probabilistic reasoning. In particular, we compare the performance of systems based on manually constructed and automatically constructed belief networks. The systems exploit existing image descriptions and also exploit interactions between multiple classifiers to improve classification performance. Performance evaluation results for the consumer photograph domain are presented.

1. INTRODUCTION

Digital visual information has become pervasive in personal computers, and it has become increasingly important to develop tools and techniques to improve the accessibility of visual information with specific content in personal computers.

Based on our experience in a previous Web based image retrieval system, WebSEEk, we found that subject hierarchy browsing was a popular user operation in interactive image retrieval [3]. Users usually first browsed through a subject hierarchy to get general ideas about the collection and then issued specific queries using keywords. Image classifiers can be used to automatically map images to specific classes to facilitate browsing.

Image classifiers can also be naturally extended for the automatic classification of videos since image classifiers can be used to classify the key frames of a video, which can be extracted using automatic scene change detection systems.

In [2], we presented the case for image classification systems based on probabilistic reasoning (ICPR systems), which was developed to satisfy two related objectives, as follows:

• First objective: exploit interactions between different classifiers to improve classification performance. This objective was motivated by the fact that a variety of image classifiers have been developed for consumer photographs [9, 1, 7, 5]. Examples of image classes are as follows: indoor, outdoor, sky, vegetation, city, nature, green vegetation, forest, mountain, sunset. These user-friendly classes can be effective for browsing consumer photographs [5].

• Second objective: exploit existing descriptions associated with an image to improve classification performance. This objective was motivated by the fact that consumer photographs and videos often have associated annotations. Annotations may be specified directly by users or derived indirectly from speech or text associated with the visual information.

The core of ICPR systems are belief networks for probabilistic reasoning. In [2], we presented experimental results for ICPR systems based on belief networks that were constructed automatically by using systems for belief network learning. In this paper, we present experimental results in manually constructing the topology of belief networks, which is necessary in a variety of scenarios, as described in section 4.

The paper is structured as follows. In section 2 we formulate the objectives described above. In section 3, we introduce a probabilistic reasoning system based on belief networks, which lies at the core of ICPR systems. In section 4, we present the steps to build ICPR systems. We present a case study of building an ICPR system for the domain of consumer photographs. Performance evaluation results for ICPR systems based on belief networks that are constructed both automatically and manually are reported in section 5.

2. PROBLEM FORMULATION

In this section we formulate the objectives presented in section 1. For a given domain, we define a set of N random variables X_1 ... X_N to represent the set of classification problems of the domain. A random variable X_i represents a classification problem, and can take one out of a set of discrete values, which correspond to a set of classes.

For a given domain, we define a set of M random variables Y_1 ... Y_M to represent the set of classifiers of the domain. A random variable Y_i takes one out of a set of discrete values, which depends on the output of the classifier.

Consider now the classification problem represented by the random variable X_i. Let us define X_i to take a value from a set of K classes $\{c_1, ..., c_K\}$. The objectives described in section 1 can be concisely formulated as the computation of the *maximum a posteriori* classification, c_{MAP}, as follows:

$$c_{MAP} = \underset{c \in \{c_1, \cdots, c_K\}}{argmax} \; P(X_i = c \mid Evidence) \tag{1}$$

$$Evidence \subseteq \{X_1, \cdots, X_N, Y_1, \cdots, Y_M\} \tag{2}$$

Notice that the set of evidence variables can be any set of all the random variables in the domain. This includes the set of random variables representing the classifiers and those representing the classification problems. Knowing the exact values of random variables that correspond to classification problems corresponds to exploiting existing description for classification. Knowing the exact values of random variables that correspond to multiple classifiers corresponds to exploiting multiple classifiers for classification. Examples are provided in section 5.

3. PROBABILISTIC REASONING SYSTEMS

Probability theory shows us that the joint probability distribution of a given domain can be used to answer any *query* about a domain. However, the joint probability distribution can become intractably large as the number of random variables in the domain grows. Consider the case where there are n boolean random variables in a domain. The joint requires 2^n atomic probability specifications. For just $n = 20$ random variables, we need more than one million atomic probabilities to be specified.

To deal with this problem, we can use a data structure called a belief network to give a concise specification of the joint probability distribution for a set of random variables of a given domain [4]. Belief networks simplify the computation of *query results* and greatly reduce the number of conditional probabilities that need to be specified.

3.1. Semantics of belief networks

The joint probability distribution of a given domain can be used to answer any question about a domain. A generic entry in the joint is the probability of a conjunction of particular assignments to each random variable, such as $P(X_1 = x_1 \wedge \cdots \wedge X_n = x_n)$. We use the notation $P(x_1, \cdots, x_n)$ as an abbreviation for this. By using probability theory, we can rewrite the joint in terms of the following product [4]:

$$P(x_1, \cdots, x_n) = \prod_{i=1}^{n} P(x_i \mid x_{i-1}, \cdots, x_1) \tag{3}$$

For all the random variables of a given domain, consider structuring the random variables in a network in which each node of the network represents a random variable, and directed links connect pairs of nodes. If there is a directed link from node X to node Y, then we say that X is the parent of Y. Let us make the requirement that $Parents(X_i) \subseteq \{x_{i-1}, \cdots, x_1\}$. This condition is easily satisfied by labeling the nodes of the network in any order that is consistent with the partial order implicit in the network structure. For this network, let us assume that all the nodes in the network have been structured such that:

$$P(X_i \mid X_{i-1}, \cdots, X_1) = P(X_i \mid Parents(X_i)) \tag{4}$$

In other words, X_i is conditionally independent of its predecessor nodes, given the parents. For a network that meets these requirements, which we call a belief network, we have the following:

$$P(x_1, \cdots, x_n) = \prod_{i=1}^{n} P(x_i \mid Parents(X_i)) \tag{5}$$

Therefore, for a belief network, each entry in the joint is represented by the product of the appropriate elements of the conditional probability tables (CPTs) in the belief network. The CPTs provide a decomposed representation of the joint. For a belief network that is properly constructed for a domain, equation 5 provides a complete and concise description of the domain. Section 4 describes how belief networks can be constructed for a domain.

3.2. Inference in belief networks

Once a belief network has been constructed for a domain, in can be used as the basis for a probabilistic reasoning system that computes the posterior probability distribution for a query variable, given exact values for some evidence variables. That is, the system computes $P(Query \mid Evidence)$. Belief networks are flexible enough so that any node can serve as either a query or an evidence variable. Efficient inference mechanisms for answering queries given a belief network rely on applying Bayes' rule, standard methods for manipulating probability expressions, and the conditional independence relationships that are inherent in the network structure. These algorithms are discussed in detail in [4].

4. IMAGE CLASSIFICATION SYSTEMS BASED ON PROBABILISTIC REASONING

The steps to build an ICPR system for a given domain are outlined below. The system is based on the probabilistic reasoning system presented in section 3.

• First step: decide on the classification problems that are useful and interesting for the domain.

• Second step: define a set of random variables to represent the set of classification problems and the set of classifiers.

• Third step: construct a belief network for the domain.

A belief network learning system can be used to automatically construct the topology and learn the CPTs of a belief network, based on a set of training examples, as discussed in [2]. A training example consists of the values of the random variables defined in the previous step.

In developing ICPR systems, it may be necessary to manually construct all or part of the topology of a belief network for a domain. This is the case when it is not possible to have access to a large set of training examples with all the values of the random variables specified.

In a domain in which there are relatively few image classification problems, the following heuristic procedure can be used to manually construct the topology of a belief network for ICPR systems.

• First step: for a set of random variables representing the classification problems of a domain, list all the possible ways to order the random variables.

• Second step: incrementally construct a belief network for each ordering of random variables, by inserting nodes into the network according to the ordering. Specify the parents of each node such that equation 4 holds.

No.	Classification problem	Classifiers	Descriptions
1	indoor-outdoor	indoor-outdoor, sky, vegetation	none
2	sky-no sky	indoor-outdoor, sky, vegetation	none
3	vegetation-no vegetation	indoor-outdoor, sky, vegetation	none
4	indoor-outdoor	indoor-outdoor	sky-no sky
5	sky-no sky	sky	indoor-
			outdoor
6	vegetation-no vegetation	vegetation	indoor-outdoor

Table 1: Scenarios for performance evaluation

• Third step: eliminate the belief networks for which any of the required CPTs cannot be specified either subjectively or from a database of training examples.

• Fourth step: select the belief network that minimizes the number of links. Minimizing the number of links minimizes the number of CPTs that need to be specified.

• Fifth step: add the nodes representing the classifiers. A node representing a classifier is always a child of the node representing the corresponding classification problem. The node representing a classifier is not used as a parent node of any other random variables.

5. PERFORMANCE EVALUATION

In this section we present performance evaluation results for ICPR systems that were built for the consumer photograph domain. ICPR systems were built to integrate the following classification problems and their associated classifiers: indoor-outdoor, sky-no sky, green vegetation-no green vegetation. These user-friendly classes are intuitive and can be highly effective for browsing consumer photographs [5].

The classifiers built for each of these classification problems was based on a block matching approach first proposed in [9, 7]. In our experiments, we used a system for belief network learning and inferencing from the Knowledge Media Institute [10]. The experiments used a database of 1,708 consumer photographs, courtesy of Kodak. Examples of these images are shown in [8].

The experiments compare the performance of the individual classifiers with the ICPR systems, for the scenarios described in Table 1. These only represent a subset of all the possible scenarios the system can accomodate. Scenarios 1 through 3 are examples of using multiple classifiers for a given image classification problem. For example, scenario 1 classifies images as indoor or outdoor, given the outputs of the indoor-outdoor, sky, and green vegetation classifiers. Scenarios 4 through 6 are examples of using existing descriptions for image classification. For example, scenario 4 classifies images as indoor or outdoor, given the output of the indoor-outdoor classifier and an indoor-outdoor label. The evaluation procedure consists of the following steps:

• The set of all images in the database were divided into two sets, set_a and set_b.

• The images in set_b were used as the training set for the block matching classifiers. The block matching classifiers were used to classify the images in set_a. This is repeated by using set_a as the training set and set_b as the testing set. At the end of this step, all the image in the database have been classified by the block matching classifiers.

• The manual labels and outputs of the block matching classifiers for images in set_b were used as the training set to generate belief network bn_b. Belief network bn_b is then used to classify the images in set_a.

• The manual labels and outputs of the block matching classifiers for images in set_a were used as the training set to generate belief network bn_a. Belief network bn_b was used to classify the images in set_b.

For each scenario, we compared the block matching classifiers with the ICPR systems. We also compared the ICPR systems based on automatically constructed belief networks with ICPR systems based on manually constructed belief networks. McNemar's test was used to determine the statistical significance of the differences in classification performance [6].

To apply McNemar's test, we divide an available data set S into a training set R and a test set T. We use the training set R to learn two classifiers, C_A and C_B (note that for manually constructed belief networks, training data is still required to learn the CPTs). These classifiers are used to classify the test set T, and the following are computed:

• n_{00}: number of examples misclassified by both C_A and C_B.

• n_{01}: number of examples misclassified by C_A but not by C_B.

• n_{10}: number of examples misclassified by C_B but not by C_A.

• n_{11}: number of examples misclassified by neither C_A nor by C_B.

The null hypothesis to be tested is that the two learned classifiers have the same error rate, which means that $n_{01} = n_{10}$. McNemar's test is based on a χ^2 test for goodness-of-fit that compares the distribution of counts expected under the null hypothesis to the observed counts.

$$\frac{(\mid n_{01} - n_{10} \mid -1)^2}{n_{01} + n_{10}} \tag{6}$$

The distribution of the statistics in equation 6 is approximated as a continuous χ^2 distribution. The constant 1 in the numerator is used to compensate for the fact that n_{01} and n_{10} are discrete values. Note that the larger the test size, the smaller the continuous-discrete approximation error.

If the null hypothesis is correct, then the probability that this quantity is greater than ${\chi^2}_{1,0.95} = 3.841459$ is less than 0.05. So we may reject the null hypothesis in favor of the hypothesis that the two algorithms have different performance when trained on the particular training set R. The assumptions and details of the test are described in detail in [6].

5.1. Results

Table 2 summarizes the results. The belief network topologies that were manually constructed and automatically learned

No.	Belief network topology	ICPR system accuracy	Improvement
1	Automatic	86.24%	+0.70%
	Manual	86.48%	+0.94%
2	Automatic	79.33%	-0.23%
	Manual	80.56%	+1.00%
3	Automatic	77.22%	-0.29%
	Manual	78.22%	+0.74%
4	Automatic	88.64%	+3.10% *
	Manual	88.89%	+3.34% *
5	Automatic	82.32%	+2.75% *
	Manual	85.30%	+5.74% *
6	Automatic	83.78%	+6.26% *
	Manual	83.55%	+6.03% *

Table 2: Results for performance evaluation (* denotes statistical significance)

can be viewed in [8]. For each scenario, the improvement in performance refers to the improvement of using the ICPR system compared to the individual block matching classifiers. Results are reported for ICPR systems based on belief networks that are constructed both manually and automatically. For manual construction of belief networks, we used the procedure described in section 4.

The ICPR systems led to statistically significant improvements over the block matching classifiers for the scenarios in which existing descriptions were used for classification. ICPR systems based on manually constructed belief networks only had statistically significant improvements over those based on automatically constructed belief networks for scenarios 2 and 5. Note that table 2 only reports the results of McNemar's test for comparing ICPR systems and block matching classifiers.

6. CONCLUSIONS AND FUTURE WORK

We presented experimental results in constructing belief networks for image classification systems based on probabilistic reasoning (ICPR systems). In particular, we compared the performance of ICPR systems based on manually constructed and automatically constructed belief networks.

On the one hand, ICPR systems are designed on a principle of separation, in which we use individual image classifiers that have been optimized for different image classificaiton problems. One the other hand, ICPR systems are designed on a principle of integration, in which we use probabilistic inference to integrate multiple image classifiers and image classification problems.

In the future we plan on extending the system for the classification of consumer photographs by incorporating more classifiers into the system. We also plan on building ICPR systems for the news photograph domain. For this system, we plan on incorporating classifiers based on both textual and visual information [1], since news photographs have accompanying captions and articles.

7. REFERENCES

[1] S. Paek, C. Sable, V. Hatzivassiloglou, A. Jaimes, B. Schiffman, S.-F. Chang and K. McKeown *Integration of Visual and Text Based Approaches for the Content Labeling of Photographs,* ACM SIGIR Conference '99. Special Session on Multimedia Indexing and Retrieval. Berkeley, CA. August 1999.

[2] S. Paek and S.-F. Chang *A Knowledge Engineering Approach for Image Classification Based on Probabilistic Reasoning Systems* IEEE International Conference on Multimedia and Expo 2000. New York, NY. July, 2000.

[3] S.-F. Chang, J. R. Smith, M. Beigi and A. Benitez *Visual Information Retrieval from Large Distributed On-Line Repositories,* Communications of the ACM, vol. 40, no. 12, 1997. Special issue on Visual Information Management.

[4] S. Russell and P. Norvig *Artificial Intelligence: A Modern Approach,* Prentice Hall, 1995.

[5] A. Vailaya, A. Jain and H. J. Zhang *On Image Classification: City vs. Landscape,* Proceedings of the IEEE Workshop on Content Based Access of Image and Video Libraries. June 21, 1998. Santa Barbara, California.

[6] T. Dietterich *Approximate statistical tests for comparing supervised classification learning algorithms,* Journal of Neural Computation, vol. 10, no. 7, pages 1895-1924. 1998.

[7] M. Szummer and R. W. Picard *Indoor-Outdoor Image Classification* IEEE Workshop on Content Based Access of Image and Video Databases (CAIVD-98), 1998.

[8] http://www.ee.columbia.edu/advent *Web demo of image classification systems based on probabilistic reasoning*

[9] M. Gorkani and R. W. Picard *Texture orientation for sorting photos at a glance,* International Conference on Pattern Recognition. Oct. 1999. Jerusalem, Israel.

[10] M. Ramoni and P. Sebastiani *Learning Bayesian Networks from Incomplete Databases,* Proceedings of the Thirteenth Conference on Uncertainty in Artificial Intelligence. 1997. San Mateo, California. http://kmi.open.ac.uk/projects/bkd