Combining Configurational and Statistical Approaches in Image Retrieval

Huizhen Yu and Eric Grimson

Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: A key goal of image indexing systems is to create methods that can efficiently retrieve images from large collections. To achieve efficiency, two challenging problems are the representations for an image, and learning query concepts from users' interaction.

A powerful representation is the attributed graph [1]. It is a structure composed by attributed parts (their color, shape, for instance), and attributed relations such as relative brightness, relative texture change, and relative positions, etc. We will call subgraphs of these attributed graphs "configurations". This structured composition is convenient for representing contextual information in an image.

However, when the system gets only labeled images, learning the common structure becomes very hard, because the explicit representation of an image by a set of configurations lacks the simple structure of a vector space.

Previous Work: There is some work on learning a concept under a set representation (e.g., [4, 2]). In [4], an implicit set representation was used, but the method was computationally costly, and negative examples were not used in concept learning. In [2] an explicit set representation was used by fixing the structure of the configurations and representing images by a few of the major configurations they contain. This transforms concept learning into maximum likelihood parameter estimation. However, the structure of the configuration is not as flexible as in the implicit set representation.

Approach: Inspired by the analogy of words to documents, we are able to extend [4] using the attributed graph representation. The key is to derive a secondary vector representation after extracting candidate configurations.

In preprocessing of the database, we compute the attributed graph on color-based segmentation of each image. Upon query, the system extracts candidate configurations from example images and trains an inference module to softly select among these configurations, then predict over the dataset to retrieval relevant images.

The main idea of transforming into a vector representation is to map each image into a "secondary" feature vector where feature i corresponds to "whether configuration i has occurred in it." To allow uncertainty in the detection procedure, for each configuration, we model the associated graph checking algorithm as an autonomous agent [3] from whom we receive information, but have no clear knowledge how this information is gathered.

Figure 1 shows the graphical model in our current work. The *dummy node* e_i represents the virtual evidence "observed" from an image by the checking algorithm, who then reports us the likelihood ratio $\frac{P(e_i|X_i=1)}{P(e_i|X_i=0)}$. Denoting all evidence by e, and assuming uniform prior on class label Y,

$$P(Y \mid e) \propto \prod_i \sum_{x_i} P(e_i \mid x_i)\, P(x_i \mid Y)\,. \tag{7}$$

The parameters, $P(X_i \mid Y)$, are estimated in training using Expectation-Maximization (EM). The likelihood ratio $\frac{P(e_i|X_i=1)}{P(e_i|X_i=0)}$ suffices for calculations in both training and prediction.

The detail of this work and experimental results are presented in [5].

Impact: Our work exploits the richness in the structured description of visual contents as well as the simplicity of a vector representation. An advantage is that it separates decision from the inexactness in modeling and the uncertainty in measurements. This is beneficial when perfect detectors are not obtainable (e.g., due to inaccurate image segmentation), or not affordable due to efficiency reasons.

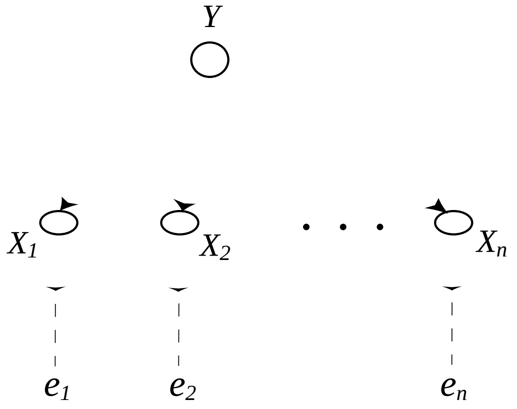

Figure 1: Naive Bayesian network with latent variables.

Future Work: The limitation of the current work is the conditional independence in the Naive Bayesian network. Though success has been shown in applying Naive Bayesian to some vision recognition tasks, the problem of being over-confidence in prediction becomes severe when examples are extremely few. We are working towards handling correlation between detectors, as well as improving both the representation and the learning algorithm by incorporating statistical features and prior information in a systematic way.

Research Support: This work is supported by Industrial Technology Research Institute.

References:

[1] R. Haralick and L. Shapiro. *Computer and Robot Vision.* Addison-Wesley, 1992.

[2] O. Maron and A. Lakshmi Ratan. Multiple instance learning for natural scene classification. *Proc. the 11th International Conference on Machine Learning,* 1998.

[3] J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference.* Morgan Kaufmann, 1988.

[4] A. Lakshmi Ratan and W.E.L. Grimson. Training templates for scene classification using a few examples. *Proc. IEEE Workshop on Content-based Access of Image and Video Libraries,* 1997.

[5] H. Yu and W.E.L. Grimson. Combining configurational and statistical approaches in image retrieval. *to appear The 2nd IEEE Pacific-Rim Conference on Multimedia,* 2001.