Learning Image Attributes using the Indian Buffet Process

Soravit Changpinyo
Department of Computer Science
Brown University
Providence, RI 02912
schangpi@cs.brown.edu

Erik B. Sudderth
Department of Computer Science
Brown University
Providence, RI 02912
sudderth@cs.brown.edu

Abstract

In the domain of object recognition and image classification, a recent trend is to use image properties or attributes to represent the images. Most of the proposed models in the past require that the number of attributes and attribute semantics be specified in advance. In this paper, we propose a generative model for image attributes that combine attribute-based vision models and feature-based nonparamatric models. We learn the model using Gibbs sampling. Qualitatively, we demonstrate the learned attributes of images in three categories. Quantitatively, we show that our model outperforms simple baseline methods in image retrieval and transfer learning tasks.

1 Introduction

Object recognition is an important area in computer vision. Most work in object recognition traditionally focuses on finding object categories. However, simply naming a specific object does not generalize well across categories. Some recent work [6, 26] proposes transitioning the goal of object recognition from *naming to describing*. In particular, objects can also be described in terms of certain properties or attributes. For example, we can describe a dog with attributes "brown," "hairy" and "has legs," etc. This attribute-centric approach attracts researchers for several reasons. First, it improves the degree to which machines perceive visual objects. The ability of machines to *describe* objects provides useful applications such as image search engines that perform well on specific queries. Additionally, attributes are often shared by different objects. Therefore, those attributes provide useful information for organizing collections of images. Moreover, in recognition tasks, the knowledge about attributes allows part of the learning task to be shared amoung categories. This cross-category generalization has a computational advantage. For example, Torralba et al. [24] employ a joint boosting procedure by finding common features that can be shared across object classes. By training classifiers jointly rather than independently, they discover that the required number of features grows sublinearly with the number of classes. Another advantage of having generic features is that those features are necessary when few or no training examples are available. Some recent work on attribute-based transfer learning [17] shows that attibutes can be used to build a learning object detection system that requires one or no training images of the target classes.

In almost all proposed attributed learning algorithms and attributed-based object detection frameworks, attributes must be defined a priori. Some other work [2, 27] takes advantage of natural language descriptions associated with images as a means to define their attributes. In this paper, we attempt to learn attributes in a completely unsupervised way from any given set of images. We borrow a data-driven, nonparametric Bayesian statistical method called the infinite sparse factor analysis, which is a linear transformation method in which the desired representation of multivariate data is the one that minimizes the statistical dependence of the components of the representation. In a sparse implementation, we allow the choice of whether a component is active for a data point. In