COS 424: Interacting with Data

Lecturer: Fei-Fei Li
Scribe: Yun Zhang

Lecture # 21
April 26, 2007

1 Introduction to Computer Vision

What is computer vision?

Vision is the task of "see". It is seeing with understanding other than seeing with camera.When we "see" things, our eyes (sensing device) capture the image, then pass the information to brain (interpreting device). The brain interprets the image, gives us meanings of what we see. Similarly, in computer vision, camera serves as sensing device, and computer acts as interpreting device to interpret the image the camera captures.

Computer vision is related to many areas, including biology, psychology, information engineering, physics, maths, and of course computer science. This lecture will focus on the machine learning methods adopted in computer vision area, specifically object recognition, which is the hardest domain in computer vision research.

2 Object Categorization

Question:How many kinds of objects there are in this world?

Answer: It is still an open question because it is hard to define objects.

2.1 Challenges in object categorization

In object categorization research, there are several main challenges, which are:

- **view point variation** When we look at Michalangelo's sculpture from different view points, we get toally different pictures. Yet they are the same object. How can we categorization the following three pictures?
- **illumination** When we take a picture for a person under different illuminations, we can get quite different images too. How can we know they belong to the same object class?
- **occlusion** Human tend to complete an incomplete image automatically. When we can only see the top half of a person, we still think that is an object of "person" because we complete the image by occlusion. How can computers do that correctly?
- **scale** How can computers know two objects belong to the same category if they dont have the same scale?
- **deformation**
- **background clutter** To pick an object out from its background is not an easy task for computers
- **intra-class variation** Objects in the same class may differ from each other greatly. For example, there are many kinds of chairs, some have arms some dont, some have legs some dont, some have square shape some are round..... How could computer know they are in the same category?

2.2 Statiscal Viewpoint

If we want to identify whether a picture contains zebra or not, from the statistical point of view, actually we want to know:

$p(zebra|image) vs. p(nozebra|image)$

Applying Bayes rule:

$$\frac{p(zebra|image)}{p(nozebra|image)} = \frac{p(image|zebra)}{p(image|nozebra)} \cdot \frac{p(zebra)}{p(nozebra)}$$

in which

$\frac{p(zebra|image)}{p(nozebra|image)}$ is posterior ratio

$\frac{p(image|zebra)}{p(image|nozebra)}$ is likelihood ratio

$\frac{p(zebra)}{p(nozebra)}$ is prior

We can use two methods in statistical learning for this problem.

- **Discrimitive Method** It directly models posterior, separate zebra pictures from non-zebra pictures.
- **Generative Method** It models likelihood and priors

3 Discriminative vs. Generative Methods

3.1 Three Main Issues

- **Representation** How to represent an object category. The method we choose can directly affects how we represent the images.
- **Learning** Learn how to distinguish objects rather than manually specify the differences. We can choose the level of supervision. Learning is different for discriminative and generative methods.
- **Recognition**

3.2 Generative Models

3.2.1 Bag-of-words Models

The bag-of-words model chops an image into patches. We dont care where the patches come from (their original position in the image). In this model, we lose all the geometric information.

Why the bag-of-words is a good way to approach? When we are browsing a piece of article, we see the highlighted words. We will probably recognize which kind of object this image coud be. The patches are already giving meanings of this image. In another word, the distribution of visual words can give lots of information on which object it is.

The question is: how to chop the image?

- **grid cutting**
- **feature detector** look for busy parts of the image. There are corner detector, edge detector etc.

In the learning phase, we chop images using feature detection & representation. We put patched into the codewords dictionary, then build a classifier. In the recognition phase, we use the same dictionary to make predictions.

Case Studies

Notations:

w_n: each patch in an image

w: a collection of all N patches in an image

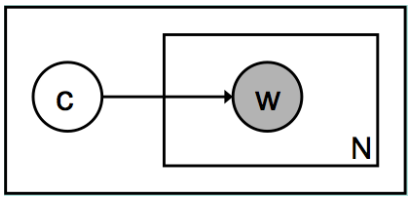
d_j: the $j^(th)$ image in an image collection

c: category of the image

z: theme or topic of the patch

1 Naive Bayes Classifier

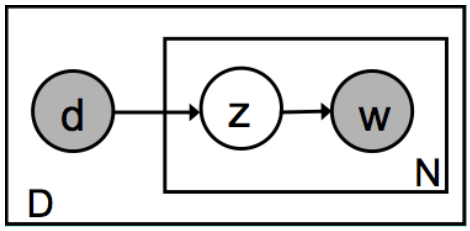
We can take each of the codewords. For each category, we look at how these codewords are distributed.



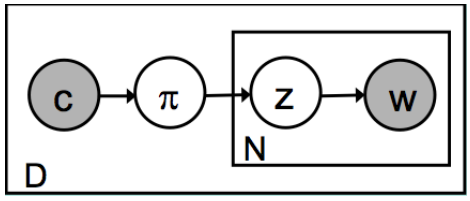
$$c^* = argmax_c p(c|w) \propto p(c)p(w|c) = p(c)\Pi_{n=1}p(w_n|c)$$

2. Hierarchical Bayesian Text Models

Probabilistic Latent Semantic Analysis (pLSA)

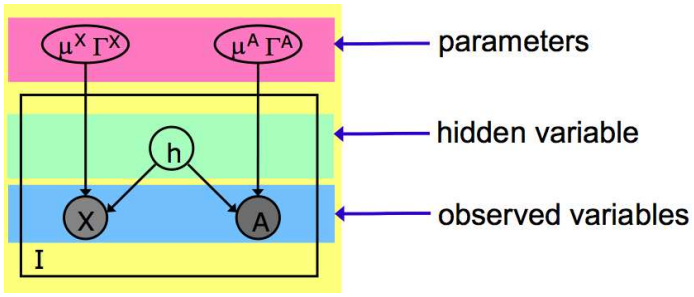


Latent Dirichlet Allocation (LDA) is more advanced than pLSA by applying a prior



3.2.2 Part-based Models

Part-based models puts a lot of geometric information to image. One-shot learning is one example. It uses lots of prior knowledge. By applying priors to parameters of the images, we can learn the distribution of the parameters (how they may shift around), therefore the correctness of the generative model is secured.



Here $\mu^X\tau^X$ is the shape model, and $\mu^A\tau^A$ is the appearance model.

3.3 Discriminative Methods

Discriminative methods is different from generative model in a sense that we dont care how the image looks like exactly. We only care which class this image is in rather than the exact shape of the image.

In another word, generative model cares about exactly what the image is, while discriminative model cares about the different between classes. Discriminative methods include nearest neighbor, neural learning, SVM etc. Boosting is one of the mostly used method and the most efficient method too. To use boosting, we should first define weak classifiers ("weak detectors"). Various weak detectors have been proposed by man researchers.