Object Recognition: Conceptual Issues

Issues in recognition

The statistical viewpoint
Generative vs. discriminative methods
Model representation
Generalization, bias vs. variance
Supervision
Datasets
Discrete Random Variables

\( X \) denotes a random variable. 

\( X \) can take on a countable number of values in \( \{x_1, x_2, \ldots, x_n\} \).

\( P(X=x_i) \), or \( P(x_i) \), is the probability that the random variable \( X \) takes on value \( x_i \).

\( P(.) \cdot \) is called probability mass function.

Continuous Random Variables

\( X \) takes on values in the continuum.

\( p(X=x) \), or \( p(x) \), is a probability density function.

\[ \Pr(x \in (a, b)) = \int_a^b p(x) \, dx \]

E.g.
Joint and Conditional Probability

\[ P(X=x \text{ and } Y=y) = P(x,y) \]

If \( X \) and \( Y \) are independent then
\[ P(x,y) = P(x) \cdot P(y) \]

\( P(x \mid y) \) is the probability of \( x \) given \( y \)
\[ P(x \mid y) = \frac{P(x,y)}{P(y)} \]
\[ P(x,y) = P(x \mid y) \cdot P(y) \]

If \( X \) and \( Y \) are independent then
\[ P(x \mid y) = P(x) \]

---

Law of Total Probability, Marginals

<table>
<thead>
<tr>
<th>Discrete case</th>
<th>Continuous case</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ \sum_x P(x) = 1 ]</td>
<td>[ \int p(x) , dx = 1 ]</td>
</tr>
<tr>
<td>[ P(x) = \sum_y P(x, y) ]</td>
<td>[ p(x) = \int p(x, y) , dy ]</td>
</tr>
<tr>
<td>[ P(x) = \sum_y P(x \mid y)P(y) ]</td>
<td>[ p(x) = \int p(x \mid y)p(y) , dy ]</td>
</tr>
</tbody>
</table>
Bayes Formula

\[ P(x, y) = P(x \mid y) P(y) = P(y \mid x) P(x) \]

\[ \Rightarrow \]

\[ P(x \mid y) = \frac{P(y \mid x) P(x)}{P(y)} = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}} \]

Normalization

\[ P(x \mid y) = \frac{P(y \mid x) P(x)}{P(y)} = \eta P(y \mid x) P(x) \]

\[ \eta = P(y)^{-1} = \frac{1}{\sum_x P(y \mid x) P(x)} \]

Algorithm:

\[ \forall x : \text{aux}_{x|y} = P(y \mid x) P(x) \]

\[ \eta = \frac{1}{\sum_x \text{aux}_{x|y}} \]

\[ \forall x : P(x \mid y) = \eta \text{ aux}_{x|y} \]
Bayes' Rule

Bayes Rule for point probabilities

\[ P(a \mid b) = \frac{P(b \mid a)P(a)}{P(b)} \]

or in distribution form

\[ P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)} \]

Useful for assessing diagnostic probability from causal probability:

\[ P(Cause \mid Effect) = \frac{P(Effect \mid Cause) P(Cause)}{P(Effect)} \]

E.g., let \( M \) be meningitis, \( S \) be stiff neck:

\[ P(m \mid s) = \frac{P(s \mid m)P(m)}{P(s)} = \frac{0.8 \times 0.001}{0.1} = 0.008 \]

Bayes' Rule

Bayes Rule for point probabilities

\[ P(m \mid s) = \frac{P(s \mid m)P(m)}{P(s)} = \frac{0.8 \times 0.001}{0.1} = 0.008 \]

\[ P(s) = P(s \mid m)P(m) + P(s \mid \neg m)P(\neg m) \]

Here we assumed that prior probability is known quantity – exactly 0.001 (point estimate)
Bayesian approach capture the uncertainty about prior as a distribution
Then the posterior will also be distribution
Simple Example of Recognition

Suppose we obtain measurement $z$
What is $P(zebra|z)$?

Causal vs. Diagnostic Reasoning

$P(zebra|z)$ is diagnostic.
$P(z|zebra)$ is causal.

Often causal knowledge is easier to obtain.
Bayes rule allows us to use causal knowledge:

$$P(zebra \mid z) = \frac{P(z \mid zebra)P(zebra)}{P(z)}$$

count frequencies!
Combining Evidence

Suppose we obtain another observation \( z_2 \).

How can we integrate this new information?

More generally, how can we estimate
\[
P(x \mid z_1 \ldots z_n)\]

Example: Second Measurement

\[
P(z_2 \mid \text{zebra}) = 0.5 \quad P(z_2 \mid \neg \text{zebra}) = 0.6
\]

\[
P(\text{zebra} \mid z_1) = \frac{2}{3}
\]

\[
P(\text{zebra} \mid z_2, z_1) = \frac{P(z_2 \mid \text{zebra}) P(\text{zebra} \mid z_1)}{P(z_2 \mid \text{zebra}) P(\text{zebra} \mid z_1) + P(z_2 \mid \neg \text{zebra}) P(\neg \text{zebra} \mid z_1)}
\]

\[
= \frac{\frac{1}{2} \cdot \frac{2}{3} \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{2}{3} + \frac{3}{5} \cdot \frac{1}{3}} = \frac{5}{8} = 0.625
\]

\* \( z_2 \) lowers the probability that the picture is zebra.
Object categorization: the statistical viewpoint

- MAP decision: \( p(\text{zebra} \mid \text{image}) \) vs. \( p(\text{no zebra} \mid \text{image}) \)

\[
P(x \mid y) = \frac{P(y \mid x) P(x)}{P(y)} = \frac{\text{likelihood} \cdot \text{prior}}{\text{evidence}}
\]

- Bayes rule:

\[
p(\text{zebra} \mid \text{image}) \propto p(\text{image} \mid \text{zebra}) p(\text{zebra})
\]

\[
\begin{align*}
\text{posterior} & \quad \text{likelihood} & \quad \text{prior} \\
\end{align*}
\]
Object categorization: the statistical viewpoint

\[ p(\text{zebra} \mid \text{image}) \propto p(\text{image} \mid \text{zebra}) p(\text{zebra}) \]

- **Discriminative methods**: model posterior
- **Generative methods**: model likelihood and prior

Discriminative methods

- Direct modeling of \( p(\text{zebra} \mid \text{image}) \)
Generative methods

- Model $p(\text{image} \mid \text{zebra})$ and $p(\text{image} \mid \text{no zebra})$

<table>
<thead>
<tr>
<th>$p(\text{image} \mid \text{zebra})$</th>
<th>$p(\text{image} \mid \text{no zebra})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Middle</td>
</tr>
<tr>
<td>High</td>
<td>Middle → Low</td>
</tr>
</tbody>
</table>

Generative vs. discriminative methods

- Generative methods
  - Interpretable
  - Can be learned using images from just a single category
  - Sometimes we don’t need to model the likelihood when all we want is to make a decision

- Discriminative methods
  - Efficient
  - Often produce better classification rates
  - Can be hard to interpret
  - Require positive and negative training data
Spectrum of supervision

Learning approaches proceed in supervised way: need some labeled data

Unsupervised  "Weakly" supervised  Supervised

Definition depends on task

What task?

- Classification
  - Object present/absent in image
  - Background may be correlated with object

- Localization / Detection
  - Localize object within the frame
  - Bounding box or pixel-level segmentation

http://pascallin.ecs.soton.ac.uk/challenges/VOC/

2008 Challenge classes:

Person: person
Animal: bird, cat, cow, dog, horse, sheep
Vehicle: aeroplane, bicycle, boat, bus, car, motorbike, train
Indoor: bottle, chair, dining table, potted plant, sofa, tv/monitor

http://pascallin.ecs.soton.ac.uk/challenges/VOC/

• Main competitions
  – **Classification**: For each of the twenty classes, predicting presence/absence of an example of that class in the test image
  – **Detection**: Predicting the bounding box and label of each object from the twenty target classes in the test image

http://pascallin.ecs.soton.ac.uk/challenges/VOC/

- “Taster” challenges
  - **Segmentation:**
    Generating pixel-wise segmentations giving the class of the object visible at each pixel, or “background” otherwise

  - **Person layout:**
    Predicting the bounding box and label of each part of a person (head, hands, feet)

Labeling with games

http://www.gwap.com/gwap/

Figure 1: Partners agreeing on an image in the ESP Game. Neither player can see the other’s guesses.

Figure 2: peekaboom. “Peek” tries to guess the word associated with an image closely revealed by “Boom.”

LabelMe
http://labelme.csail.mit.edu/

**Summary**

- Recognition is the “grand challenge” of computer vision
- History
  - Geometric methods
  - Appearance-based methods
  - Sliding window approaches
  - Local features
  - Parts-and-shape approaches
  - Bag-of-features approaches
- Issues
  - Generative vs. discriminative models
  - Supervised vs. unsupervised methods
  - Tasks, datasets