10 Automating interpretation systems

Interpretation

Plausible inference of hidden semantic structure from observable inputs

E.g.

input		hidden structure
word sequence	→	meaning
pixel matrix	→	object, relations
speech signal	→	phonemes, words
words in e-mail Subject:	→	Is message spam? Yes/No
symptoms	→	illness

How to combine ambiguous, incomplete and conflicting evidence to draw reasonable conclusions?

Distinct from logical reasoning

- plausible inference:
  - non-monotonic: might change conclusions given more evidence
  - uncertain: conclusions are not guaranteed to be correct
    (but still want to do as well as possible)
- logical inference:
  - monotonic: once a conclusion is drawn it can never be retracted
  - certain: conclusions are certain given assumptions

10.1 How to build an interpretation system?

observables → ? → hidden semantic structure

Two key problems

1. need to represent facts about process that connects evidence to truth
2. need principles of evidence combination

In this course

We will represent uncertain knowledge using *probability theory*

Some alternatives we will not cover are

- fuzzy logic, fuzzy sets
- default logic
- rule-based systems
- Dempster-Shafer theory
- rough sets, ...

10.2 Probability theory

We will cover this in depth for the next several lectures. To get started, consider of some simple examples and basic properties of probability

- example: rolling a dice
- example: random variable

Independent random variables

$$\mathrm{P}(X_1 = x_1, X_2 = x_2) = \mathrm{P}(X_1 = x_1)\, \mathrm{P}(X_2 = x_2)$$

Alternative definition

$$\mathrm{P}(X_1 = x_1 | X_2 = x_2) = \mathrm{P}(X_1 = x_1)$$

It is easy to prove these two definitions are equivalent (prove it!)

Conditional probability

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A, B)}{\mathrm{P}(B)}$$

Bayes' Theorem

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(B|A)\, \mathrm{P}(A)}{\mathrm{P}(B)}$$

Conditionally independent random variables

X_1 and X_2 are *conditionally independent given* X_3 if

$$\begin{aligned} &\mathrm{P}(X_1\!=\!x_1, X_2\!=\!x_2 | X_3\!=\!x_3) \\ &\quad = \ \mathrm{P}(X_1\!=\!x_1 | X_3\!=\!x_3)\mathrm{P}(X_2\!=\!x_2 | X_3\!=\!x_3) \quad \text{for all } x_1, x_2, x_3 \end{aligned}$$

Equivalently, if

$$\begin{aligned} &\mathrm{P}(X_1\!=\!x_1 | X_2\!=\!x_2, X_3\!=\!x_3) \\ &\quad = \ \mathrm{P}(X_1\!=\!x_1 | X_3\!=\!x_3) \quad \text{for all } x_1, x_2, x_3 \end{aligned}$$

Prove these definitions are equivalent

10.3 Forward generative models

Now, to apply this to building interpretation systems

1. Represent knowledge with probability: forward generative models

↓	world
truth	↓
↓	sensor
evidence	↓
	observed measurement

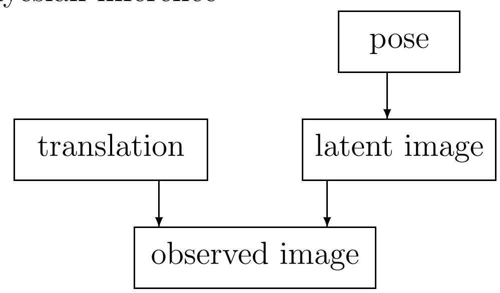
2. Principle of evidence combination: Bayesian inference

$$\begin{aligned} \text{conclusion} &= \arg \max_{\text{possible truth}} P(\text{possible truth}|\text{evidence}) \\ &= \arg \max_{\text{possible truth}} \frac{P(\text{evidence}|\text{possible truth})P(\text{possible truth})}{P(\text{evidence})} \\ &= \arg \max_{\text{possible truth}} P(\text{evidence}|\text{possible truth})P(\text{possible truth}) \end{aligned}$$

10.4 Demos

Demo 1: Image normalization

Bayesian inference



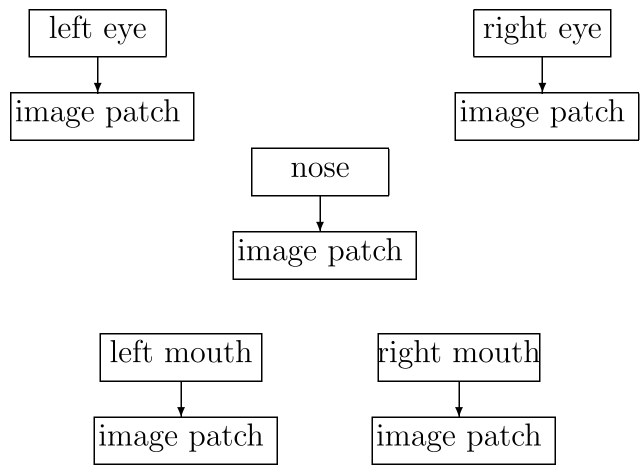
A time component is included to model image stabilization

Demo 2: Independent object tracking

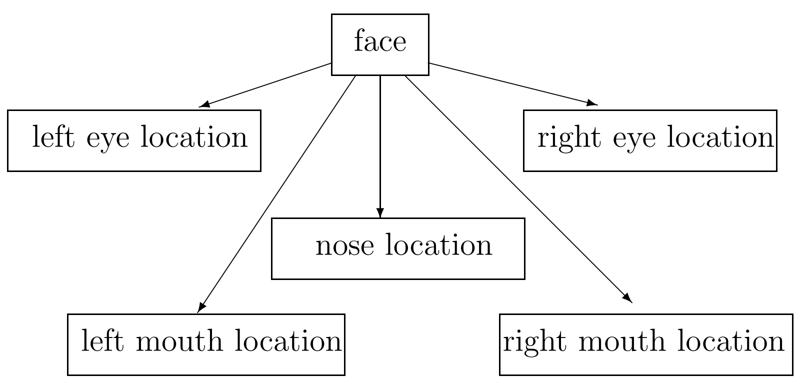
Demo 3: Independent object tracking and object removal

Demo 4: Face tracking

Two models combined: low-level model



High-level model



Readings

Dean, Allen, Aloimonos: Section 3.7
Russell and Norvig: Section 14.7