EECS 3401 — AI and Logic Prog. — Lecture 19

Adapted from slides of Brachman & Levesque (2005)

Vitaliy Batusov
vbatusov@cse.yorku.ca

York University

November 25, 2020

- Today: **Uncertain Reasoning**
- Required reading: Russell & Norvig Ch.13 and Ch. 14.1, 14.2

- Ordinary common-sense knowledge quickly moves away from categorical statements like *a P is always a Q*
- Many ways to come up with less categorical information
  - things are usually (almost never; occasionally; seldomly; rarely; almost always) a certain way
  - judgments about how good an example something is ("barely rich", "not very tall", etc.)
  - imprecision of sensors
  - reliability of sources of information
  - strength/confidence/trust in generic information or deductive rules
- With information like this, conclusions will not "follow" in the usual sense

Weakening a Universal

There are at least three ways a universal can be made to be less categorical

$$\forall x\, P(x)$$

1. *Strength of quantifier*
   "95% of birds can fly"
   statistical interpretation, probabilistic sentences
2. *Degree of belief* in the whole sentence
   "80% confidence in this fact"
   uncertain knowledge, subjective probability
3. *Applicability of predicate / degree of membership*
   "fairly tall"
   flexible membershit, **vague predicates**

Objective probability

Statistical (frequency) view of sentences

- **Objective**: does not depend on who is assessing the probability
- Always applied to collections (as opposed to singleton random events)
- Can use probabilities to correspond to English words like “rarely”, “likely”, “usually”
- Generalized quantifiers
  Compare: *for most* x, $Q(x)$ vs. *for all* x, $Q(x)$

Basic postulates

- Real numbers between 0 and 1 representing frequency of an event in a large-enough random sample
- $0 =$ "never happens", $1 =$ "always happens"
- Start with set U of all possible occurrences. An event a is any subset of U. A **probability measure** is any function $\boldsymbol{Pr}$ from events to $[0, 1]$ satisfying
    - $\boldsymbol{Pr}(U) = 1$
    - If $a_1, \ldots, a_n$ are disjoint events, then $\boldsymbol{Pr}(\cup_i a_i) = \sum_i \boldsymbol{Pr}(a_i)$

Basic postulates

- Conditioning: the probability of one event may depend on its interaction with others

$$\boldsymbol{Pr}(a \mid b) = \frac{\boldsymbol{Pr}(a \cap b)}{\boldsymbol{Pr}(b)}$$

- Conditional independence: event a is judged *independent* of event b conditional on background knowledge s if knowing that b happened does not affect the probability of a

$$\boldsymbol{Pr}(a \mid s) = \boldsymbol{Pr}(a \mid b, s)$$

Some consequences

- Conjunction:

$$\begin{aligned}\boldsymbol{Pr}(ab) &= \boldsymbol{Pr}(a \mid b) \cdot \boldsymbol{Pr}(b) && \text{(in general)} \\ &= \boldsymbol{Pr}(a) \cdot \boldsymbol{Pr}(b) && \text{(conditionally indep.)}\end{aligned}$$

- Negation:

$$\begin{aligned}\boldsymbol{Pr}(\neg s) &= 1 - \boldsymbol{Pr}(s) \\ \boldsymbol{Pr}(\neg s \mid d) &= 1 - \boldsymbol{Pr}(s \mid d)\end{aligned}$$

Some consequences

- If $b_1, b_2, \ldots, b_n$ are pairwise disjoint and exhaust all possibilities, then

$$\boldsymbol{Pr}(a) = \sum_i \boldsymbol{Pr}(ab_i) = \sum_i \boldsymbol{Pr}(a \mid b_i) \cdot \boldsymbol{Pr}(b_i)$$

$$\boldsymbol{Pr}(a \mid c) = \sum_i \boldsymbol{Pr}(ab_i \mid c)$$

- Bayes' rule

$$\boldsymbol{Pr}(a \mid b) = \frac{\boldsymbol{Pr}(a) \cdot \boldsymbol{Pr}(b \mid a)}{\boldsymbol{Pr}(b)}$$

If a is a disease and b a symptom, it is usually easier to estimate numbers on the right-hand side

Subjective probability

- It is reasonable to have non-categorical beliefs even in categorical sentences
- E.g., to express confidence/certainty in the sentence as a whole
- *Someone's confidence* $\Rightarrow$ subjective
- **Prior probability** $\boldsymbol{Pr}(x \mid s)$: probability of x with respect to prior state of information s
- **Posterior probability** $\boldsymbol{Pr}(x \mid E, s)$: same, after acquiring new evidence E
- Need to combine evidence from various sources. How to derive new beliefs from prior beliefs and new evidence?

From Statistics to Belief

- Would like to go from statistical information (*probability that a bird chosen at random will fly*) to a degree of belief (*a particular bird named Tweety will fly*)
- Traditional approach: find a *reference class* for which we have statistical information and use the statistics for that class to compute an appropriate degree of belief for an individual
- Imagine trying to assign a degree of belief to the proposition *"Eric, who is an American man, is tall"* provided the facts
  - A 20% of American men are tall
  - B 25% of Californian men are tall
  - C Eric is from California
- **Direct inference** — this kind of move from statistics to concrete belief

Many classes

- **Problem**: individuals belong to many classes
- With just (A): probability is 0.2
- A + B + C: prefer more specific classes, probability is 0.25
- A + C: no stats for more specific class — probability 0.2?
- B: are Californians a representative sample?

Basic Bayesian Approach

- Would like a more principled way of calculating subjective probabilities
- Assume we have n atomic propositions $p_1, \ldots, p_n$ that we care about
- A logical interpretation $\mathcal{I}$ can be though of as a specification of which p_i are true and which are false.

  Notation: for $n = 4$, we use $\langle \neg p_1, p_2, p_3, \neg p_4 \rangle$ to mean the interpretation where only p_2 and p_3 are true.

Joint Probability Distribution

- A **joint probability distribution** J is a function from interpretations to $[0, 1]$ satisfying

$$\sum_{\mathcal{I}} J(\mathcal{I}) = 1$$

- $J(\mathcal{I})$ is the degree of belief that the world is as $\mathcal{I}$ describes it

Degree of Belief

- The **degree of belief** in any sentence ϕ:

$$\boldsymbol{Pr}(\phi) = \sum_{\mathcal{I} \models \phi} J(\mathcal{I})$$

- Example:

$$\begin{aligned}\boldsymbol{Pr}(p_2 \land \neg p_4) = {} & J(\langle \neg p_1, p_2, p_3, \neg p_4 \rangle) + \\ & J(\langle \neg p_1, p_2, \neg p_3, \neg p_4 \rangle) + \\ & J(\langle p_1, p_2, p_3, \neg p_4 \rangle) + \\ & J(\langle p_1, p_2, \neg p_3, \neg p_4 \rangle)\end{aligned}$$

Problem with this approach

- To calculate the probabilities of arbitrary sentences involving the propositions p_i, we would need to know the full joint distribution function
- For n atomic sentences, this requires knowing 2^n numbers — insane
- Would like to make some plausible assumptions to cut down on what needs to be known about the world
- In the simplest case, all the atomic sentences are independent. This gives us that

$$J(\langle P_1, \ldots, P_n \rangle) = \boldsymbol{Pr}(P_1 \wedge \ldots \wedge P_n) = \Pi_i \boldsymbol{Pr}(P_i)$$

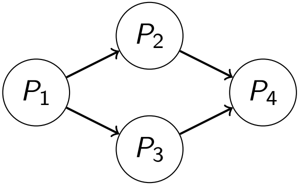
  (where P_i is either p_i or $\neg p_i$), so only n numbers are needed.
- **Too strong an assumption**

A better assumption

- A better assumption:
  The probability of each P_i only depends on a small number of P_j, and the dependence is acyclic

Belief Networks

- Represent all atoms in a **belief network** (Bayes' network)



- Assume: $J(\langle P_1, \dots, P_n \rangle) = \Pi_I \boldsymbol{Pr}(P_i \mid parents(P_i))$,
  where $parents(P)$ denotes the set/conjunction of **parents** of node P.
- Under this assumption, we get for network above:

$$J(\langle P_1, \dots, P_n \rangle) = \boldsymbol{Pr}(P_1) \cdot \boldsymbol{Pr}(P_2 \mid P_1) \cdot \boldsymbol{Pr}(P_3 \mid P_1) \cdot \boldsymbol{Pr}(P_4 \mid P_2, P_3)$$

Belief Networks

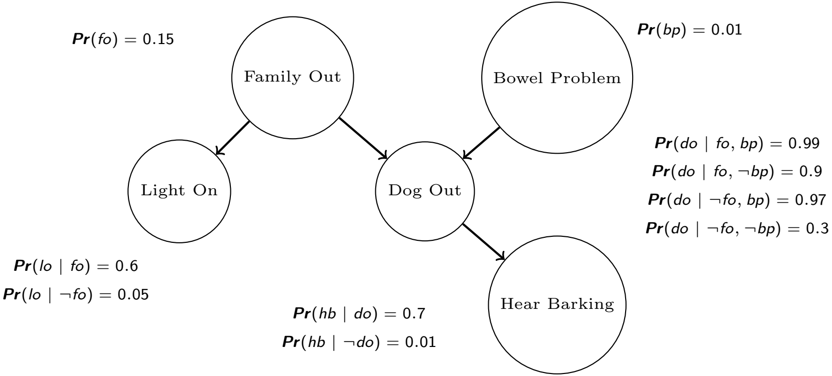
- For a particular interpretation:

$$
\begin{aligned}
&J(\langle p_1, \neg p_2, p_3, \neg p_4 \rangle) = \\
&\qquad \boldsymbol{Pr}(p_1) \cdot \boldsymbol{Pr}(\neg p_2 \mid p_1) \cdot \boldsymbol{Pr}(p_3 \mid p_1) \cdot \boldsymbol{Pr}(\neg p_4 \mid \neg p_2, p_3) \\
&\qquad \boldsymbol{Pr}(p_1) \cdot [1 - \boldsymbol{Pr}(p_2 \mid p_1)] \cdot \boldsymbol{Pr}(p_3 \mid p_1) \cdot [1 - \boldsymbol{Pr}(p_4 \mid \neg p_2, p_3)]
\end{aligned}
$$

- To fully specify the joint distribution (and therefore probabilities over any subset of variables), we only need to know $\boldsymbol{Pr}(P \mid parents(P))$ for every node P
- E.g., if node P has parents $Q_1, \ldots, Q_m$, then we need to know the values of $\boldsymbol{Pr}(p \mid q_1, q_2, \ldots, q_m)$, $\boldsymbol{Pr}(p \mid \neg q_1, q_2, \ldots, q_m)$, $\boldsymbol{Pr}(p \mid q_1, \neg q_2, \ldots, q_m)$, $\boldsymbol{Pr}(p \mid \neg q_1, \neg q_2, \ldots, q_m)$, $\ldots$, $\boldsymbol{Pr}(p \mid \neg q_1, \neg q_2, \ldots, \neg q_m)$
- $n \cdot 2^m \ll 2^n$ for large n

Example

- Assign a node to each variable in the domain
- Draw arrows toward each node P from a select set $parents(P)$ of nodes perceived to be “direct causes” of P



- DAG: directed acyclic graph

Example

- Then

$$J(\langle FO, LO, BP, DO, HB\rangle) =$$
$$\boldsymbol{Pr}(FO) \cdot \boldsymbol{Pr}(LO \mid FO) \cdot \boldsymbol{Pr}(BP) \cdot \boldsymbol{Pr}(DO \mid FO, BP) \cdot \boldsymbol{Pr}(HB \mid DO)$$

- Can calculate the full joint distribution using this formula and ten values above

Example calculation

- What are the chances the family is out if the light is on, but can't hear any barking?

$$\boldsymbol{Pr}(fo \mid lo, \neg hb) = \frac{\boldsymbol{Pr}(fo, lo, \neg hb)}{\boldsymbol{Pr}(lo, \neg hb)}$$

$$\boldsymbol{Pr}(fo, lo, \neg hb) = \sum_{BP,DO} J(\langle fo, lo, BP, DO, \neg hb\rangle)$$

$$\boldsymbol{Pr}(lo, \neg hb) = \sum_{FO,BP,DO} J(\langle FO, lo, BP, DO, \neg hb\rangle)$$

Example

Substituting values:

$$J(\langle fo, lo, bp, do, \neg hb\rangle) = 0.15 \cdot 0.6 \cdot 0.01 \cdot 0.99 \cdot 0.3 = 0.0002673$$
$$J(\langle fo, lo, bp, \neg do, \neg hb\rangle) = 0.15 \cdot 0.6 \cdot 0.01 \cdot 0.01 \cdot 0.99 = 0.00000891$$
$$J(\langle fo, lo, \neg bp, do, \neg hb\rangle) = 0.15 \cdot 0.6 \cdot 0.99 \cdot 0.9 \cdot 0.3 = 0.024057$$
$$J(\langle fo, lo, \neg bp, \neg do, \neg hb\rangle) = 0.15 \cdot 0.6 \cdot 0.99 \cdot 0.1 \cdot 0.99 = 0.0088209$$
$$J(\langle \neg fo, lo, bp, do, \neg hb\rangle) = 0.85 \cdot 0.05 \cdot 0.01 \cdot 0.97 \cdot 0.3 = 0.000123675$$
$$J(\langle \neg fo, lo, bp, \neg do, \neg hb\rangle) = 0.85 \cdot 0.05 \cdot 0.01 \cdot 0.03 \cdot 0.99 = 0.0000126225$$
$$J(\langle \neg fo, lo, \neg bp, do, \neg hb\rangle) = 0.85 \cdot 0.05 \cdot 0.99 \cdot 0.3 \cdot 0.3 = 0.00378675$$
$$J(\langle \neg fo, lo, \neg bp, \neg do, \neg hb\rangle) = 0.85 \cdot 0.05 \cdot 0.99 \cdot 0.7 \cdot 0.99 = 0.029157975$$

So, $\boldsymbol{Pr}(fo \mid lo, \neg hb) = \frac{0.03316}{0.06624} = 0.5$