Notes on Probabilistic Reasoning

Stuart C. Shapiro

December 9, 2004

1 Introduction

These notes comment on, and are, in part, derived from Brachman & Levesque, *Knowledge Representation and Reasoning*, Chapter 12.

Three ways to relax categorical reasoning (*All sheep are white*):

Objective probability: *Most sheep are white*

Subjective probability: *I'm pretty sure that most sheep are white.*

Vague predicates: *I'm pretty sure that most sheep are rather white.*

2 Objective Probability

About frequency of events.

Derivation of Bayes' Rule

1. $\mathbf{Pr}(a \mid b) = \frac{\mathbf{Pr}(a \cap b)}{\mathbf{Pr}(b)}$
2. $\mathbf{Pr}(a \cap b) = \mathbf{Pr}(a \mid b) \times \mathbf{Pr}(b)$
3. $\mathbf{Pr}(a \cap b) = \mathbf{Pr}(a \mid b) \times \mathbf{Pr}(b) = \mathbf{Pr}(b \mid a) \times \mathbf{Pr}(a)$
4. (Bayes' Rule) $\mathbf{Pr}(a \mid b) = \frac{\mathbf{Pr}(b \mid a) \times \mathbf{Pr}(a)}{\mathbf{Pr}(b)}$

Why? For diagnosis, you might know

> **$\mathbf{Pr}(a)$:** the probability of some disease in the general population,
>
> **$\mathbf{Pr}(b \mid a)$:** the probability of people with that disease showing some symptom.
>
> **$\mathbf{Pr}(b)$:** the probability of people with that symptom in the general population.
>
> You can then calculate $\mathbf{Pr}(a \mid b)$: the probability that someone with that symptom has that disease. This is a kind of probabilistic *abductive* reasoning.

Where do the numbers come from? Statistics on the occurrence of the events in the world.

3 Subjective Probability

3.1 Introduction

About degree of confidence in a proposition.

Assume a set of atomic propositions, $p_1, \ldots, p_n$, each with an *a priori* probability, $\mathbf{Pr}(p_1), \ldots, \mathbf{Pr}(p_n)$.

Assume a set of literals, $P_1, \ldots, P_n$, where each P_i is either p_i or $\neg p_i$. (Note that $\mathbf{Pr}(\neg p_1) = (1 - \mathbf{Pr}(p_1)$.)

The probability that the world is described by $P_1 \wedge \cdots \wedge P_n$, assuming they are independent, is $\mathbf{Pr}(P_1) \times \cdots \times \mathbf{Pr}(P_n)$. This is the *joint probability,* $J(\langle P_1, \ldots, P_n \rangle)$.

Notice that a world description, $P_1 \wedge \cdots \wedge P_n$, is also a *situation*, or *interpretation* $\mathcal{I}$.

So, the probability that some conjunction of literals, α, is true is $\mathbf{Pr}(\alpha) = \sum\limits_{I \models \alpha} J(\mathcal{I})$, which is the probability of α over all worlds in which α is true.

And,

$$\mathbf{Pr}(\alpha \mid \beta) = \frac{\mathbf{Pr}(\alpha \wedge \beta)}{\mathbf{Pr}(\beta)} = \frac{\sum_{I \models \alpha \wedge \beta} J(\mathcal{I})}{\sum_{I \models \beta} J(\mathcal{I})}$$

3.2 Belief (Bayesian) Networks

Developed by Judea Pearl.

Suppose that $p_1, \ldots, p_n$ are not independent, but can be arranged in a dag (multi-rooted tree), such that any p_i is only dependent on its ancestors.

Let's let $\mathcal{I}^P$ be an interpretation over the set of atomic propositions P. So $\mathcal{I}^{\{a,b\}}$ would be one of $a \wedge b$, $\neg a \wedge b$, $a \wedge \neg b$, or $\neg a \wedge \neg b$.

Each node, p in a Bayesian network must be annotated with the following (conditional) probabilities:

- the a priori probability $\mathbf{Pr}(p)$ if p is a root.
- the conditional probabilities $\mathbf{Pr}(p \mid \mathcal{I}^{Parents(p)})$ if p is a non-root node, where $Parents(p)$ is the set of parent nodes of p, and $\mathcal{I}^{Parents(p)}$ ranges over all the interpretations of $Parents(p)$.

From those probabilities, the joint probability of any possible state of the world modeled by the Baysian network can be calculated. See the example in the text.

Where do the numbers come from? Subject Matter Experts (SMEs).

3.3 Dempster-Shafer Theory

One of a number of approaches in which intervals are used instead of specific numbers. Interval arithmetic is applicable, e.g., $[a, b] + [c, d] = [a + b, c + d]$.

4 Vague Predicates

Analysis in text comes from Lotfi Zadeh's theory of fuzzy sets and fuzzy logic.

Each vague predicate has a precise base function. E.g. tall has height as its base function. A vague predicate can be combined with a modifier, such as "not very tall", "fairly tall", "very tall", etc. These don't partition the base function, but each has a degree curve, so a particular individual might be fairly tall to one degree (between 0 and 1) and very tall to some other degree at the same time.

$$deg(\neg P) = 1 - deg(P)$$

$$deg(P \wedge Q) = min(deg(P), deg(Q))$$

$$deg(P \vee Q) = max(deg(P), deg(Q))$$

Text shows application of fuzzy logic to sets of production rules to make some decision.

Fuzzy logic controllers have been successful in a number of applications (elevators, rice cookers, etc.), but why is controversial.

The text gives a reconstruction in terms of Bayesian probability, but Zadeh insists that fuzzy logic is not probability.