Soft Decision Analysis

Thomas Whalen

1. Soft Computing

Computing with words

“Satisficing.”

Barriers to humanization of decision processes.
press of too many trivial decisions;
information pathologies such as ignorance, misperception, or overload; and
compromising real perceptions to fit overly restrictive tools.

2. “Soft” Elements of Decision Analysis

From a standpoint of “computing with words,” standard methods of decision analysis begin by converting the natural perceptions of how usual or unusual the possible states are into Bayesian estimates of subjective probability measured on a ratio scale, and converting the natural perceptions of how acceptable or unacceptable the possible outcomes of state-action pairs are into utility numbers on an interval scale using Von Neuman-Morgenstern [1947] utility theory or some similar method. Thus, these two elements of traditional decision analysis can be considered to belong at least to the borderland of soft techniques.

2.1 Uncertain Probabilities

Bayes' Theorem was developed a very long time ago (before 1763) by the good Reverend Thomas Bayes, who was a talented dabbler in mathematics and statistics as a hobby. It was published shortly after his death in the Philosophical Transactions of the Royal Society, with a foreword by his friend Richard Price, who remarked that Bayes' theorem was important because it applies to a problem posed by De Moivre: "To show what reason we have for believing that there are in the constitutions of things fixt laws according to which events happen, and that, therefore, the future of the world must be the effect of the wisdom and power of an intelligent cause; and thus to confirm the argument taken from final causes for the existence of the Deity." Unfortunately, Bayes was so far ahead of his times that his article sat in the prestigious Transactions of the Royal Society virtually unread until the mid 1930's.

The reason we need Bayes Theorem is because the two quantities

P(A|B) and P(B|A) are not equal.

For example, consider an oil prospector who faces the decision:
to drill for oil on a site based on the information he already has,
to let the site go, or
to pay for a seismic test and drill if the seismologists' prediction is positive.

He expects a profit of $4 Million if he drills and finds a major oil strike, a profit of $1 Million if he drills and finds a Minor strike, and a loss of $1 Million if he drills and finds a Dry hole.

To be specific, suppose our prospector's current knowledge leads him to extimate P(Major)=.40, P(Minor) =0.20, and P(Dry) = 0.40.

Note that the prospector doesn't, and cannot, know the "true" probability of a major strike; after all, if the oil is down there it has been there for millions of years (probability = 1.0) and if it's not there it never has been and never will be (probability = 0).

The meaning of subjective probability is that the prospector would be indifferent between a side bet that pays off if there is a major oil strike at his site, and a side bet with an known probability of success of exactly 40%.

Based on these probabilities, his "expected" profit for drilling without doing a sesmic test is .4*4M + .2*1M - .4*1M = 1.4 Million dollars.

Geologists have been doing seismic experiments to support oil exploration work for many years; methods have continued to improve and standardize, and the petroleum exploration industry is very familiar with their track record. There are many instances on record of how drilling attempts have turned out, and on what the geological predictions submitted before drilling had said. That means that it is possible to have a very good idea how accurate the predictions typically turn out to be.

The information we are much more likely to be able to get is *P(Positive|Major)*, which is found by dividing

the total number of times a major oil strke was found in similar terrain after a positive seismic test had been reported

by

the total number of times a major oil strike was found in similar terrain after any seismic test had been reported.

This form, while it is more indirect, is more valuable because it focuses on the underlying ability of the geologists while abstracting away the particular characteristics of the unique mix of sites they've been presented with in the past.

The probability that the seismic test will be positive given that there is a major strike of oil waiting to be drilled is not the same as the probability of a major strike given that the seismic test was positive.

The thing that makes Bayes theorem possible is the fact that the following four quantities are all equal:

P(A|B)P(B) = P(A&B) = P(B&A) = P(B|A)P(A)

For example,

P(Major|Positive)xP(Positive)= P(Positive|Major)xP(Major).

We can use algebra to solve this for the probability of a major oil strike given that the seismic test was positive:

$$P(Major|Positive) = \frac{P(Positive|Major) \times P(Major)}{P(Positive)}$$

Our plan is now to get P(Positive|Major) from the historical record of the geological survey companies, calculate the overall probability of a positive test P(Positive), and use these to calculate the desired posterior probability P(Major|Positive) as well as the rest of the probabilities needed to fill out the decision tree.

We should also note that this implies that Major is *dependent* on Positive. If Major and Positive were *statistically independent* then we would find that P(Major|Positive) = P(Major), in which case Positive would *not* be a conditioning event. It doesn't matter whether Major causes Positive, Positive causes Major, they are both caused by a common factor (e.g., oil), or whether the situation is some complex mix of these, but there has to be some kind of connection in order to support inference.

Let's reconsider the probabilities for the oil prospector's situation. He really would have begun with 2 kinds of probability information:

Prior Probabilities represent his initial beliefs or knowledge about the presence or absence of oil. *P(Major)* is one such probability. These prior probabilities may be objective or subjective probabilities, but are often subjective.

Likelihoods are conditional probabilities that summarize the known performance characteristics of the tests. They are more typically objective, often based on relative frequencies. Whatever kind of testing may be involved in a problem, the conditional probabilities usually summarize how reliably such testing has performed in the past. For example, *P(Positive|Major)* should represent, for those past cases where the geologists did a study before a major oil find, the proportion of the time that study was positive.

That is the information available to us at the beginning. We want to use those probabilities to give us an indication of how testing will perform in *this particular case*. We want to use the above information to give us a revised view of the likelihood of the various outcomes of drilling after we know what the geologists prediction says. Just as *prior* means *before*, *posterior* means *after*. *P(Major|Negative)* is a ***Posterior Probability***, sometimes grandly referred to as a ***Bayesian Revised Probability***. In other words, once we know the outcome of the testing, it makes sense for us to revise what we believe to be the probability of a major oil find.

One useful way to look at the information involved in applying Bayes' Theorem is to describe the probabilities involved as:

P(Event) Our ***prior probability*** of an outcome that is our main interest. These probabilities, objective or subjective, reflect the circumstances of ***this particular situation***.

P(Prediction|Event) These ***likelihoods*** describe the expected performance of our source of additional information. They may arise from the "track record" of a market researcher, the past performance of a testing procedure, or from physical reality (as in counting black and white balls in the urns that statisticians so dearly love). In any case, these probabilities <u>do</u> <u>not</u> <u>reflect</u> the current situation at all. They describe the ***predictive power of the information source***.

P(Prediction & Event) These ***joint probabilities*** combine the above two kinds of probability information to reveal the probability that ***in this situation***, our information source will make a particular prediction ***and*** a particular outcome will happen.

P(Prediction) Sum the joint probabilities involving this prediction and you get this ***marginal probability*** that ***in this situation***, our information source will make this prediction.

P(Event|Prediction) The ***posterior probability*** that the event will occur, given a particular prediction. Given a prediction, this replaces our prior *P(Event)* in the decision process. These probabilities are what we seek in applying Bayes' Theorem.

Note that the prior probabilities and the marginal probabilities are unconditional. The likelihoods and the posterior probabilities ***are*** conditional probabilities, but the conditioning goes in opposite directions!

Our oil prospector's ***Prior Probabilities*** were:

P(Major) = 0.40

P(Minor) = .20

P(Dry) =.40

He got his ***Likelihoods*** from the performance record of the firm he expected to hire, *Valdez et Cie*. *Valdez et Cie*.'s track record was

***Valdez et Cie*.**	Cases where Site later turned out to be a Major strike	Cases where Site later turned out to be a Minor strike	Cases where Site later turned out to be a Dry hole	Total Cases
Positive Prediction	36	27	12	75
Negative Prediction	4	18	28	50
Total Cases	40	45	40	125

The next step is to convert these frequencies to likelihoods by appropriate divisions:

Valdez et Cie.

	Cases where Site later turned out to be a Major strike	Cases where Site later turned out to be a Minor strike	Cases where Site later turned out to be a Dry hole	Total Cases
Positive Prediction	$^{36}/_{40}$= .90	$^{27}/_{45}$=.60	$^{12}/_{40}$= .30	75
Negative Prediction	$^{4}/_{40}$= .10	$^{18}/_{45}$= .40	$^{28}/_{40}$= .70	50
Total Cases	40	45	40	125

Now we need to construct the ***Joint Probabilities*** associated with the kind of prediction our oil prospector will get and the results he will receive if he drills.

P(Pos&Major)= P(Pos|Major)P(Major) = .9*.4= .36	P(Pos&Minor)= P(Pos|Minor)P(Minor) = .6*.2= .12	P(Pos&Dry)= P(Pos|Dry)P(Dry) = .3*.4= .12
P(Neg&Major)= P(Neg|Major)P(Major) = .1*.4= .04	P(Neg&Minor)= P(Neg|Minor)P(Minor) = .4*.2= .08	P(Neg&Dry)= P(Neg|Dry)P(Dry) = .7*.4= .28

These results will let us construct the ***Marginal Probabilities*** for the results of the geological studies.

P(Pos) = P(Pos&Major) + P(Pos&Minor) + P(Pos&Dry) = .36 +.12 +.12 =.60
P(Neg) = P(Neg&Major) +P(Neg&Minor) +P(Neg&Dry) = .04 +.08 +.28 =.40

Now we have everything we need to apply Bayes' Theorem to find the ***Posterior Probabilities*** that we need to evaluate the worth of the seismic test.

Let's do just one ***Posterior Probability*** the long way, then we'll see how to shortcut the rest.

$$P(\text{Major}|\text{Pos}) = \frac{P(Pos|Major)\times P(Major)}{P(Pos|Major)\times P(Major)+P(Pos|Minor)\times P(Minor)+P(Pos|Dry)\times P(Dry)}$$

$$P(\text{Major}|\text{Pos}) = \frac{.90\times.40}{.90\times.40 \quad + \quad .60\times.20 \quad + \quad .30\times.40}$$

$$P(\text{Major}|\text{Pos}) = \frac{.36}{.36 \quad + \quad .12 \quad +.12} = \frac{.36}{.60} = .60$$

Looking at that last calculation, you should be able to recognize that there is really an easy shortcut available to us, since we have already calculated all the ***Marginal*** and ***Joint Probabilities***. We could have done it like this:

$$P(\text{Major}|\text{Pos}) = \frac{P(Major\&Pos)}{P(Pos)} = \frac{.36}{.60} = .60$$

Now let's do all the rest of the ***Posterior Probabilities*** using the shortcut:

$P(Major\|Pos) = \dfrac{P(Major\&Pos)}{P(Pos)} = \dfrac{.36}{.60} = .60$	$P(Minor\|Pos) = \dfrac{P(Minor\&Pos)}{P(Pos)} = \dfrac{.12}{.60} = .20$	$P(Dry\|Pos) = \dfrac{P(Dry\&Pos)}{P(Pos)} = \dfrac{.12}{.60} = .20$
$P(Major\|Neg) = \dfrac{P(Major\&Neg)}{P(Neg)} = \dfrac{.04}{.40} = .10$	$P(Minor\|Neg) = \dfrac{P(Minor\&Neg)}{P(Neg)} = \dfrac{.08}{.40} = .20$	$P(Dry\|Neg) = \dfrac{P(Dry\&Neg)}{P(Neg)} = \dfrac{.28}{.40} = .70$

As you can hopefully now see, Bayes' Theorem helps us in Decision Analysis in a fairly special set of circumstances:

- You have the opportunity, usually at a price, to get additional information before you must commit to an Alternative Action.
- You have likelihood information that describes how well that source of information should be expected to perform (often based on how it has performed in the past).
- You wish to revise your prior probabilities (the probabilities you would be forced to use if you couldn't get the additional information).

Don’t be misled by the atmosphere of mathematical precision attending these calculations; remember that their foundation is on the subjective prior probabilities that ultimately express the decision maker’s propensity to bet. Thus, for example, P(Major|Neg) = .10 means that, if (and only if) the prospector would be indifferent between a side bet on a major strike and a side bet with a known probability of 40% when he didn’t have seismic information, then he should rationally be indifferent between a side bet on a major strike and a side bet with a known probability of 10% when he possesses the results of a negative seismic test.

Finally, with these probabilities in hand we ccan calculate the prospector's expected return given a Positive or a Negative Test.

If the test is Positive, his expected return is .6*4M + .2*1M - .2*1M = $2.4 Million.

If the test is negative, his expected return for drilling is .1*4M + .2*1M - .7*1M = -$0.1 Million.

What the latter number really means is that, if he gets a Negative seismic test, he will not drill at all, for a return of zero. (Note that this ignores any sunk costs which are irrelevant to decision making because they're the same for all outcomes currently under consideration.)

Value of Imperfect Information

We looked earlier (Churchill & Whalen, 2001) at the EVPI (Expected Value of Perfect Information), a convenient fiction. Here we have something more realistic, sometimes called Imperfect Information, and sometimes called Sample Information. The Expected Value of Sample Information (EVSI) represents the difference between our best EMV for the entire tree if sample information had not been available and the EMV we actually got.

If the cost of the information is known, then we usually measure the value in excess of the cost; this is the Net EVSI. I

f the cost of the information is unknown (or negotiable!), then we just measure the value as though we could get the information for free; the Gross EVSI. That gives us a standard against which we can later compare the price at which someone might offer us the information.

The value of information almost always varies as a function of time. The correct time (that is, the correct point in the decision tree) at which to determine the real value of information such as the seismic test is at the time that we must decide whether to purchase the information or not.

The oil prospector's expected return if he could get the seismic test for free would be

P(Pos)*[expected return for best bet given Positive]

+ P(Neg)*[expected return for best bet given Negative]

= .6*$2.4M + .4*$0 = $1,440.000.

We have already seen that, without any seismic information, he could just drill now for an expected return of $1,400,000. Thus, the value of the information contained in the seismic test is $40,000. If he can get the test performed for a cost of $30,000 he should do it, but if the test costs $50,000 he should say "no thank you" and just take his chances drilling.