STAKEHOLDER PREFERENCE ELICITATION

PROF. R.M. COOKE
Delft Institute for Applied Mathematics
Delft University of Technology
r.m.cooke@ewi.tudelft.nl
April 5, 2005-04-05

Abstract: I review arguments why stakeholder preferences cannot be modeled as utilities, multicriteria or otherwise. An approach to stakeholder preferences based on well known models for consumer preference in market research is proposed. Simple paired comparisons is used to represent group preferences on an affine unique scale, and regression is used to "explain" these preferences in terms of scores on a number of criteria. Using the rich body of standard regression techniques, we can analyse degree of fit, and we can deal with dependence in the "criteria". The tasks in stakeholder preference modeling can be apportioned between analysts, experts and stakeholders.

1. Introduction

Multi criteria methods are emerging in the area of risk analysis and risk management (Linkov et al, 2005). It is appropriate to recall the classical arguments why stakeholder preferences cannot be modeled as utilities, multicriteria or otherwise, in the sense of rational decision theory. An approach to stakeholder preferences based on well known models for consumer preference in market research is proposed (for a review see Cooke 1991). Simple paired comparisons is used to represent group preferences on an affine unique scale, and regression is used to "explain" these preferences in terms of scores on a number of criteria. Similar approaches to modeling valuations can be found in (McFadden 1974, Koop et al. 1994, Kind 1996, Saloman 2004, McCabe et al 2004). These approaches often use logit regression to model valuation of health states which can then be compared with "Standard Gamble" trade-off elicitation common in multi attribute utility theory (McCabe et al 2004, Torrance et al. 1996) Using the rich body of standard regression techniques, we can analyse degree of fit, and we can deal with dependence in the "criteria". The tasks in stakeholder preference modeling can be apportioned between analysts, experts and stakeholders.

2. Utility

Utility functions on consequences of choices/lotteries/actions emerge in the representation of rational preference, and are affine unique (unique up to choice of zero and unit). Modeling utility functions so as to be useful in social decision making has proven difficult, for inter alia the following reasons (French, 1988, Savage 1972, Arrow, 1963 , Arrow and Raynaud, 1986).

1) The set of consequences is large; suppose it is parametrized as a subset of real vectors of dimension n. Under certain stringent conditions the utility function can be represented as a function of "utilities on the coordinates". We then speak of *Multi Attribute Utility Theory (MAUT).* Unfortunately these conditions seldom apply in real decision problems.
2) Arrow's impossibility theorem shows that there is no way to model a group composed of rational individuals, as itself a rational individual. Thus, group decision making cannot be modeled as individual decision making for the "collective individual".

The above has been well known for many years, but has not prevented practitioners from re-making the same mistakes year after year, to wit:

1) ASSUMING without verifying that preference can be expressed in terms of preferences for criteria (coordinates) and scores on criteria (coordinate values), as in MAUT. The following preference pattern is eminently reasonable, yet inconsistent with the MAUT axioms:
    a. If unemployment is low and pollution is high, Prefer: *Close a dirty factory,*
    b. If unemployment is high and pollution is low, Prefer : *NOT close the dirty factory.*

2) ASSUMING that preference ratio's are meaningful at an individual or group level, without giving an operational definition. The presence or absence of operational definitions is illustrated as follows:
    a. John prefers *Close dirty factory* to *NOT close dirty factory* ⇒ OBSERVABLE BEHAVIOR: Vote to close
    b. John' preference ratio *Close dirty factory / NOT close dirty factory = 9* ⇒ OBSERVABLE BEHAVIOR?????

3) ASSUMING that a group can be treated as a rational individual. The following shows why this is NOT true. Consider a population with preferences

    1/3 of population: *Beethoven > Bach > Mozart*
    1/3 of population: *Mozart > Beethoven > Bach*
    1/3 of population: *Bach > Mozart > Beethoven*

    Then a 2/3 majority has the following pairwise preferences:

    Beethoven> Bach
    Bach > Mozart
    Mozart > Beethoven

While one may debate whether individual pairwise preferences should be transitive, there is no debate that majority preferences, are NOT transitive (May, 1952). Arrow and Raynaud (1986) summarize attempts to introduce axioms ensuring the transitivity of majority preference.

3. Stakeholder preferences

Group preferences do not satisfy the axioms which enable utility theory. On the other hand, there are techniques from consumer research which yield consumer preference functions. Suppose we wish to predict market share for a new product prior to bringing it on the market. We might try to "explain" observed market share of existing products by regressing observed market share on a number of explanatory variables, say Price, Size, Performance, etc. We could then score our new product on the explanatory variables and try to predict its future market share. A moment's reflection shows why this will not work. Market shares of products must add to one, if one goes up, others must go down. If we add a new product, the shares of the existing products must change.

Classical approaches (Thurstone, 1927, Bradley, Terry, 1952, Mosteller, 1952, Bradley 1953, David 1957, Torgerson 1958, McFadden 1974) attempt to deal with this by eliciting pairwise comparisons from experimental subjects, and "scaling" the data so as to yield a preference scale which can be related to market share. Of course, "market share" is just a convenient metaphor; "group preference" or "stakeholder preference" is a better designation. Examples of recent applications include patient preferences for medical treatments (Kind, 1996, McCabe et al. 2004, Salomon, J.A. 2004) and voter preference (Koop, G. and Poirier, D.J. 1994).

Most promising of these for the current problems are the Thurstone pair wise comparison methods. I sketch an implementation.

4. Implememtation

The implementation involves three groups of 'players'; an analysis team, an expert team and a set of stakeholders. Their roles are described briefly below.

Analysis team

The analysis team defines a set of alternatives (eg policies for costal defense) and a set of criteria (eg cost, expected number of fatalities, breach frequency, expected environmental impact , etc). A set of experts is also defined. The analysis team monitors and manages the entire process, and performs the mathematical analysis.

Expert team

The expert team scores each alternative with respect to each of the criteria. The experts may also provide feedback on the sets of alternatives and criteria, possibly iterating the definitions.

Stakeholders

The set of consumers are given the alternatives with their scores on the criteria. Each consumer expresses his/her pair wise preference for each pair of alternatives.

5. Analysis

The analysis team analyses the consumer preferences for consistency and significance, according to standard methods. This results in an affine unique consumer preference function which assigns a preference to each alternative. The preference values are regressed on the set of criteria. This yields a coefficient B_i for criteria C_i which optimally express the preference for each alternative as a linear function of the criteria scores. For alternative a and criteria $C_1 \ldots C_n$, we thus have:

$$PREF(a) = B_1 \times C_i(a) + \ldots B_n \times C_n(a) + \text{error}.$$

Ci(a) is the score of alternative a on criteria Ci. Unlike multicriteria analysis, the B's need not be positive and need not add to one; thus, they cannot be interpreted as “weights” for the criteria. However, they best explain the preference values in a lease squared sense. Standard tools are available to analyse the error and assess the degree to which the criteria scores explain the consumer preferences.

If the preferences are adequately explained by the model, the results are communicated to the problem owner. New alternatives can be evaluated using the regression model without iterating the consumer preference elicitation.

If the fit is not satisfactory, new criteria can be formulated and the regression step can be iterated. The regression model can also be extended to include interaction terms. This requires iterating the expert scoring and regression analysis, but does not require iterating the consumer elicitation. More details are illustrated in the following toy example.

6. Toy example

We consider a toy example for modeling group preferences for auto’s. We want to model stakeholder preferences for auto’s and use this model to predict future preference behavior and drive design improvements.

The analysis team selects 5 auto’s (policies) which cover the field parsimoniously, namely FOCUS, ASTRA, ROLLS, BMW, KA, XSRA. An expert team scores each auto on the criteria: PRICE, MONTHLY PAYMENTS, MILAGE, PASSENGER ROOM. Notice that a criteria like ROOM cannot be monotonic in value, whatever that may mean. A stakeholder will have an ideal size, and deviations above or below will be less desirable. Notice also that the criteria scores will evidently be stongly correlated.

These scores are passed to ten stakeholders, who evaluate the policies pairwise. Suppose the following preference matrix emerges (the first cell entry 6.0/10 means that 6 of the 10 stakeholders preferred the FOCUS to the ASTRA).

PREFERENCE MATRIX

Item	ASTRA	ROLLS	BMW	KA	XSRA

1 FOCUS	6.0/10	8.0/10	8.0/10	6.0/10	7.0/10
2 ASTRA		8.0/10	6.0/10	6.0/10	7.0/10
3 ROLLS			2.0/10	2.0/10	2.0/10
4 BMW				1.0/10	4.0/10
5 KA					6.0/10

The stakeholder preference summary matrix shows the number of times that each stakeholder preferred each auto to some other auto. The rightmost column shows the number of circular triads in each stakeholder's paired comparisons. With 3 or more circular triads, the null hypothesis that the stakeholder in question has given his preferences at random CANNOT be rejected. The analysis team might decide to re-elicit stakeholders Rock, Cliff, Ridge and Ruby.

STAKEHOLDER PREFERENCE SUMMARY MATRIX

Stakeholder	FOCUS	ASTRA	ROLLS	BMW	KA	XSRA	Triads
Rock	4.0	4.0	1.0	1.0	2.0	3.0	4
Cliff	3.0	2.0	1.0	4.0	4.0	1.0	4
Ridge	4.0	2.0	0.0	2.0	3.0	4.0	3
Crystal	4.0	5.0	1.0	1.0	2.0	2.0	2
Jade	4.0	1.0	1.0	2.0	5.0	2.0	2
Pebble	2.0	3.0	5.0	4.0	1.0	0.0	0
Shale	3.0	2.0	0.0	1.0	5.0	4.0	0
Flint	4.0	4.0	0.0	1.0	2.0	4.0	1
Opal	4.0	4.0	0.0	1.0	4.0	2.0	1
Ruby	3.0	4.0	1.0	2.0	3.0	2.0	6
Total	35.0	31.0	10.0	19.0	31.0	24.0	

Preference values are shown below. Three common models for analying paired comparison data are shown. Thurstone C is the model used here. (For a discussion see next section.) The others are shown for the sake of comparison. For both Thurstone models a Chi square statistic tests the hypothesis that the model assumptions hold. The value 7.0349 is far from significant, thus the data do not lead to rejecting the model (the NEL (Bradely Terry) model would also be unrejected at 4.9667).

PREFERENCE VALUES

Item name	NEL(Bradley-Terry)	Thurstone C	Thurstone B
1. FOCUS	0.2998	0.4525	0.8546
2. ASTRA	0.2193	0.2699	0.5098
3. ROLLS	0.0416	-0.7015	-1.3273
4. BMW	0.0896	-0.2981	-0.5278
5. KA	0.2193	0.3117	0.5570
6. XSRA	0.1304	-0.0345	-0.0663
Goodness of fit :	4.9667	7.0349	
(Chi-square, DF=	10	10	

It is convenient to put the preferences from the Thurstone C model and the criteria scores in one matrix. We also add a column of "1's". The effect of adding this column is to enable the criteria scores to reflect deviations from the mean, when the criteria scores are standardized.

CRITERIA SCORES

PREF	CONST	PRICE	MONTHLY	MILAGE	ROOM
0.4525	1.0000	20.0000	0.2000	14.0000	4.0000
0.2699	1.0000	25.0000	0.1800	12.0000	6.0000
-0.7015	1.0000	40.0000	0.5000	9.0000	8.0000
-0.2981	1.0000	45.0000	0.4800	8.0000	7.0000
0.3117	1.0000	12.0000	0.1200	20.0000	3.0000
-0.0345	1.0000	15.0000	0.1500	16.0000	4.0000

The regression analysis finds criteria scores which yield the best linear fit to the preferences, in the sense of least squares. These scores and the resulting error are shown below:

```
PREF = CRITERIA SCORES × REG.COEFF’S + ERROR

= CRITERIA SCORES × 1.6908  + ERROR
                     0.0409
                    -4.0369
                    -0.0403
                    -0.2126

ERROR =     0.1654
            0.0423
            0.0539
           -0.0808
            0.0577
           -0.2385
```

We see that the fit is not spectacularly good. Of course the number of alternatives is small for the number of regression coefficients to be estimated. The point is to illustrate the analytical tools which can be applied. We can analyse the covariance matrix of the regression coefficients and derive confidence bounds for the coefficient estimates. The diagonal terms are the variances of the criteria, the off-diagonal terms the covariances. We see that the standard deviation of the criterion PRICE is $\sqrt{(0.0009)} = 0.03$. This large deviation reflects of course the small number of alternatives evaluated in this toy example. In this case, the problem owner would be told that the model of stakeholder preferences was not very good; we could not claim that the coefficient for price was significantly different from zero.

REG COEFF’S COVARIANCE MATRIX

	CONST	PRICE	MONTHLY	MILEAGE	ROOM
C	1.8396	-0.0270	1.1116	-0.0758	-0.0812
Pr	-0.0270	0.0009	-0.0389	0.0012	-0.0001
Mnth	1.1116	-0.0389	2.4437	-0.0435	-0.0347
Mi	-0.0758	0.0012	-0.0435	0.0032	0.0029
RM	-0.0812	-0.0001	-0.0347	0.0029	0.0103

The off diagonal terms indicate that the fluctuations in the criteria values are not independent.

REGRESSION COEFFICIENT PRODUCT MOMENT CORRELATION

CONST	PRICE	MONTHLY	MILEAGE	ROOM
1.0000	-0.6749	0.5243	-0.9873	-0.5894
-0.6749	1.0000	-0.8426	0.6907	-0.0257
0.5243	-0.8426	1.0000	-0.4918	-0.2185
-0.9873	0.6907	-0.4918	1.0000	0.4967
-0.5894	-0.0257	-0.2185	0.4967	1.0000

Suppose that, given scores on the criteria, the preferences were sampled from a distribution in accord with the assumptions of the regression model. Then on repeated samples, the criteria coefficients found by the least squares algorithm would fluctuate with variances on the diagonal of the covariance matrix, and the scores would be correlated as in the correlation matrix. We could obtain uncorrelated estimates of the coefficients by carefully choosing our options (cars) in such a way that the criteria scores are uncorrelated. This would be a so-called orthogonal design. Having an orthogonal design is convenient but by no means necessary.

On the Thurstone C model, each consumer's utility for the 6 cars is modeled as 6 samples from independent normal variables $X_1 \dots X_6$ with unit standard deviation and with means given by the preference scores. The probability that X_j is most preferred is modeled as the probability that $X_j = \max(X_1 \dots X_6)$., and this probability is the predicted market share. It can be computed by simulation, and we find:

CAR	PREDICTED MARKET SHARE
1. FOCUS	0.28
2. ASTRA	0.22
3. ROLLS	0.04
4. BMW	0.09
5. KA	0.23
6. XSRA	0.14

Note the similarity to the Bradely-Terry scale values, which solves for market share directly from the paired comparison data (see below). We cannot extract a preference ordering from the stakeholders in this toy example, owing to intransitivities. If we could extract such an ordering, we could compare it to the predicted market shares and derive an additional check on the modeling assumptions. Pairwise comparisons are intended to deal with the volatility of consumer preference by allowing each alternative to be judged several times, in combination with all other alternatives.

7. Discussion

We briefly summarize the assumptions underlying the models Thurstone C, Thurstone B and (NEL) Bradely Terry (there is a Thurstone A model, but it is not tractable). These are compared briefly with the logit regression approach.   For a fuller discussion see (Cooke 1991).

Thurstone C: We assume that each stakeholder has an affine unique utility value for each alternative. We can arbitrarily choose 2 alternatives and scale them equal to 0 and 1. We do this for all stakeholders, using the same alternatives. The utility values for all stakeholders are now on a common scale. Each expert's values for the remaining alternatives are modeled as a sample from a random vector. We assume that the components of this vector are independent normals with unit variance and means which are solved from the paired comparison data.

Thurstone B: This is identical to Thurstone C, except that the components of the random vector of stakeholder values are not independent, but the correlation of values is constant. The solution algorithm for either model yields values determined on an *interval scale*; that is, the values are uniquely determined when a "zero" and "unit" are fixed.

The Thurstone models make assumptions which are compatible with the theory of rational preference at the individual level, and yield preference values which are affine unique.

NEL / Bradely Terry: NEL denotes Negative Exponential Life model. In the NEL model each stakeholder performs a thought experiment to determine which of two independent components with exponential life distributions outlives the other. The Bradely-Terry model assumes that values of alternatives are determined on a ratio scale, and that the probability that a stakeholder prefers alternative i over j is $V_i/(V_i+V_j)$, where V_i is the value of alternative i. Note that this ratio is invariant under linear transformations, but not under affine transformations. The computational algorithms of these two models are identical, and yield estimates of values V_i determined on a ratio scale. The Bradely Terry solution algorithm assumes that each choice event is modeled as an independent coin-tossing experiment. The estimates of the values V_i are obtained by maximal likelihood.

The coin-tossing model is inconsistent with the assumption that each stakeholder's preferences are represented by utilities. If a person prefers Bach to Mozart, then the probability is one that he will choose Bach over Mozart. However, if we interprete each choice event as a random choice of one stakeholder from the population, we could analyse the data under the Bradely Terry assumptions. Notice that we then ignore the knowledge that the *same* stakeholder made many pairwise comparisons. If we consider only the subpopulation of stakehoders who prefer either i or j above all other alternatives, then the probability of drawing a stakeholder for whom $(i > j)$ from this subopulation is indeed $V_i/(V_i + V_j)$, where V_i and V_j are the relative sizes of the subpopulations preferring i and j respectively, above all other alternatives. Bradely Terry assume that the same probability hold for all other stakeholders. This is similar to the assumption in the Thurstone C model that the experts utility values are independent normal variables.

The logit models, like the Bradely Terry model, assume that $P(i > j) / P(j > i) = V_i / V_j$, and find the scale values by log linear regression. The probability of choosing alternative i over j is $\exp(u_i) / (\exp(u_i) - \exp(u_j))$ where u_i is a "population utility" about which the individual utility values are distributed according to an extreme value distribution. Note the similarity with the Bradely Terry model by putting $u_i = \ln(V_i)$.

In conclusion we remark that the stakeholder preference method sketched above has no problem dealing with intransitive preferences. With reference to the Bach, Beethoven, Mozart example, these alternatives would all receive the same scale value.

Perhaps the main virtue of the stakeholder preference approach is that it enables standard checks on model fit and model adequacy.

8. References

1) Arrow, K.J. (1963) “Social choice and individual values” 2end edition, Wiley, New York,
2) Arrow, K.J. and Raynaud, J. (1986) "Social choice and Multicriterion Decision Making", MIT press.
3) Bradely, R. (1953) “ Some statistical methods in taste testing and quality evaluation” *Biometrica*, vol. 9, 22-38.
4) Bradley, R. and Terry, M. (1952) “Rank analysis of incomplete block designs *Biometrica*, vol. 39, 324-345.
5) Cooke, R.M. (1991) *Experts in Uncertainty* Oxford University Press.
6) David, H.A. (1957) The Method of Paired comparisons, Charles griffin, London, 1963.
7) French, S. (1988) *Decision Theory; an Introduction to the Mathematics of rationality* Ellis Horwood, Chichester.
8) Kind, P.(1996) "Deriving cardinal scales from ordinal preference data: the analysis of time trade-off data using pairwise judgement models", Paper presented to HESG, Brunel University.
9) Koop, G. and Poirier, D.J. (1994) "Rank-ordered logit models: an empirical analysis of Ontario voter preferences. *Journal of Applied Econometrics* 9, 369-388.
10) Linkov, I., Sahay, S. Kiker, G. Bridges, T. Belluck, D. and Meyer, A. (2005) “Multi-criteria decision analysis; comprehensive decision analysis tool for risk management of contaminated sediments”
11) McCabe, C. Brazier, J. Gilks, P. Tauchiya, A. Roberts, J. O'Hagan, A. and Stevens, K. (2004) " Estimating population cardinal health state valuation models from ordinal (rank) health state preference data"Sheffield Health Economics Group, Discussion Paper Series, Ref. 04/2.
12) McFadden, D. (1974) "Conditional logit analysis of qualitative choice behavior" in Zarembka, P (ed) *Frontiers in Econometrics*, New York Academic Press, USA, 105-142.
13) Mosteller, F. (1952) “Remarks on the method of paired comparisons: the least squares solution assuming equal standard deviations and equal correlations” *Psychometrica*, vol. 16, no. 1 3-9.

14) May, K.D. (1952) 'A set of necessary and sufficient conditions for simple majority decisions' *Econometrica* 20, 680-684.
15) Salomon, J.A. (2004) "The use of ordinal ranks in health state valuations" IHEA Conference, USA, San Francisco.
16) Savage, L.J. (1972) *Foundations of Statistics* 2end edition, Dover, New York.
17) Thurstone, L. (1927) “A law of comparative judgment” *Pschyl. Rev*. vol. 34, 273-286.
18) Torgerson, W. (1958) *Theory and Methods of Scaling*, Wiley, New York, 1958.
19) Torrance, G.W., Feeny, D.H., Furlong, W.J. Barr, R.D. Xhang, Y., and Wang, Q. (1996) "A multi-attribute utility functijon for a comprehensive health status classification system: Health Utilities Mark 2. *Medical Care* 34 (7) 702-722.