European Journal of Operational Research 000 (2016) 1–7

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Decision Support

Road to robust prediction of choices in deterministic MCDM

Tommi Pajala, Pekka Korhonen, Jyrki Wallenius*

Department of Information and Service Economy, Aalto University School of Business, P.O. Box 21220, Aalto 00076, Finland

ARTICLE INFO

Article history:
Received 29 December 2015
Accepted 1 October 2016
Available online xxx

Keywords:
Multi objective programming
Preference prediction
Multiple criteria decision making
Linear value function
Bayesian model

ABSTRACT

We compare five different prediction methods (linear estimated weights, AHP weights, equal weights, logistic regression, and a lexicographic method) in their success rate for predicting preferences in pairwise choices. Students were asked to make pairwise comparisons between student apartments on four criteria: size, rent, travel time to the university and travel time to a (hobby) location of their choice. First ten choices were used to set up the estimation model, and subsequent ten choices are used for prediction. We find that the linear estimation method has the highest prediction success rate. Furthermore, the probability of predicting a choice correctly differs only slightly (by 0.1) between linear consistent and inconsistent subjects, ie. subjects whose preferences were consistent or inconsistent with a linear value function. This shows that in the absence of other preference information, a linear value function is suitable for prediction purposes.

1. Introduction

The use of linear models in decision making has been quite common (Dawes, 1979; Dawes & Corrigan, 1974; Saaty, 1980; Steuer, 1986; Yu & Zeleny, 1975; Zionts & Wallenius, 1976). Dawes (1979) and Dawes and Corrigan (1974) focused on the use and performance of regression models in predicting external phenomena. However, in this article we focus on predicting the choices of decision makers themselves by using a linear value function model.

It is well known that DMs have a tendency not to be perfectly rational when making decisions (Einhorn & Hogarth, 1981; Payne, Bettman, & Johnson, 1992; Simon, 1956; Tversky, Sattath, & Slovic, 1988; Weber & Johnson, 2009). This makes predicting choices a difficult task.

In an exploratory study, Korhonen, Silvennoinen, Wallenius, and Öörni (2012) showed that a linear model succeeded rather well in predicting the choices of subjects. Korhonen et al. (2012, 2013) tested the linear estimation model (the so-called max ϵ formulation) in prediction in a two-criteria setting. With two criteria, the model proved to be robust, predicting choices well both for subjects who were consistent with a linear value function, and also for those who were inconsistent with a linear value function. In this paper, we refer to these two classes of subjects as linear and nonlinear subjects. With maximum likelihood estimation the prediction success rate was 87,6% and 81,2% for linear and nonlinear subjects, respectively (Korhonen et al., 2012). Moreover, the success rate of those predictions was relatively independent both of the errors that the subjects made, and of the linear consistency of the subjects. From a robustness perspective this is a desirable quality for a prediction model.

Our aim in this paper is three-fold:

1. to study how the increase of criteria from two to four changes the results,
2. to provide a Bayesian comparison of the different prediction models, and
3. to study some additional factors which might influence linear consistency vs. inconsistency.

The use of a linear model that outperforms human judgment forms a stream of literature that has existed for decades, with perhaps the most famous example being Dawes' papers from the seventies (Dawes, 1979; Dawes & Corrigan, 1974). What is different in our case is that we are actually trying to predict the preferences of the human decision maker herself. This should be a case where the decision maker has an advantage: after all, the DM should be the most knowledgeable expert regarding her own preferences. What statistical models can do, is just to rely on the information provided by the DM – whether tradeoffs, exemplar choices or preference judgments – and make inferences based on that. Obviously, the DM has much more information than just what he or she has provided to the statistical model. In this sense, we are making the task harder for statistical models to do well.

In this paper, we compare five methods of prediction: (1) a linear function with weights estimated from a subset of the subject's choices, (2) a linear function with equal weights, (3) a

* Corresponding author.
E-mail address: jyrki.wallenius@aalto.fi (J. Wallenius).

http://dx.doi.org/10.1016/j.ejor.2016.10.001

linear function with weights based on ex ante judgments of importance elicited with AHP, (4) logistic regression, and (5) a lexicographic choice model. Based on past research (notably Korhonen et al., 2012) we formulate and test the following three hypotheses:

1. The max epsilon method will be the most successful of the five prediction methods.
2. Consistency with a linear value function will not be related to a higher rate of prediction success.
3. The higher the value difference of alternatives A and B, the higher the chance of a successful prediction.

2. Compared prediction methods

2.1. Max epsilon model

The max epsilon method is one of the five methods tested. We describe it here at some length before the other models, since of the tested models it is less known. In this model, we assume that the subject's answers can be represented by a linear value function.

In short, we assume the subject's preferences are represented by: $v(X) = \sum_{j=1}^{p} \lambda_j x_j$, where

- p is the number of criteria, in our case $p = 4$,
- N is the number of alternatives,
- x_j refers to the value of the jth attribute from the set {price, size, distance to university, distance to leisure location}, and
- λ_j refers to the weight of the jth attribute.

Assume preferences can be represented by a preference set $P = \{(X_r, X_s) | X_r \succ X_s, r, s \in N\}$. Hence for each pair $(X_r, X_s) \in P$, X_r is preferred to X_s. Both alternatives are defined according to the aforementioned four criteria $X_r = [x_{r1}, x_{r2}, x_{r3}, x_{r4}]$.

Using the DM's responses, constraints for the LP weight estimation problem are constructed. For each "I prefer A over B" response, the following inequality restriction is constructed:

$$\sum_{j=1}^{p} \lambda_j x_{Aj} - \epsilon \geq \sum_{j=1}^{p} \lambda_j x_{Bj}$$

A similar constraint is generated conversely, for each "I prefer B over A" response. When it comes to indifference, we argue that the answer "I don't know" is ambiguous: we cannot know whether the respondent means that A and B are incomparable, or perhaps exactly equally good, or that he or she just doesn't know. However, an indifference statement can be interpreted to mean that the value from each option is "close" to each other. Naturally, what this "close" means is debatable. We have used an arbitrary value of σ that equals 56th percentile of the value difference in the data set. What this means is that σ is set so that 56% of the epsilon values in the data are lower. Choosing a lower value would mean we have to exclude one or two subjects, since the weight optimization routine does not compute. However, robustness checks showed that the results of the Bayesian hierarchical model do not change when these outliers are excluded from the data set. To include all the data, we use the 56th percentile sigma. We only use the indifference statements for estimating the weights. We do not try to predict indifference, but merely ignore these statements in the prediction phase. Accordingly, our model to estimate the weights is:

max ϵ subject to:

$$\sum_{j=1}^{p} \lambda_j x_{rj} - \epsilon \geq \sum_{j=1}^{p} \lambda_j x_{sj}, \quad \text{for all } (X_r, X_s) \in P$$

$$\sum_{j=1}^{p} |\lambda_j x_{rj} - \lambda_j x_{sj}| \leq \sigma, \quad \text{for all } (X_r, X_s) \notin P \;\&\; (X_s, X_r) \notin P$$

$$\sum_{j=1}^{p} \lambda_j = 1$$

$$\lambda_j \geq \theta, j = 1, 2, \ldots, p, \quad \text{where } \theta > 0 \text{ is non-Archimedean.}$$

As $\theta > 0$ is non-Archimedean, clearly $\lambda_j > 0 \; \forall j = 1, 2, \ldots, p$.

It ought to be noted that in our model getting an epsilon value over zero means that the subject has answered in a fashion consistent with a linear value function. Also note that in the opposite case of a negative max epsilon, the model does produce criterion weights λ_j, but this does not mean the subject is consistent with a linear value function – the obtained function just tries to approximate the choices as close as possible.

2.2. The other prediction methods

We compared the following prediction methods with the max epsilon method. Models 1–2 are all based on the assumption of a linear value function – they only differ in the weights they use. In short, a linear value function just assumes that the value from choosing an alternative A is $v(X_A) = \sum_{j=1}^{p} \lambda_j x_{Aj}$. If $v(X_A) > v(X_B)$, then we expect A to be chosen.

1. AHP weights (AHP)
   We set the weights equal to the judgments of importance that subjects made in the beginning of the study. This means that the weights are set equal to the normalized values in the AHP importance of criteria matrix.
2. Equal weights (EQ)
   The equal weights method is also a linear value function method, but instead of using estimated weights, we set all the weights to $\lambda_j = 0.25$.
3. Lexicographic heuristic (LEX)
   The lexicographic method is a non-compensatory heuristic, which compares options one criterion at a time, starting from the most important criterion. In each round, only the option or options with the best value continue on to the next round, all other options are removed from consideration. The process is repeated until only one winning option remains. We estimated the best lexicographic order for each subject from the first 10 questions, and used that order to try to predict the next 10 questions.
4. Logistic regression (LOGREG)
   Logistic regression is quite common in preference prediction, so it makes sense to include it as a baseline model.[1] In this model, we estimate the model $p(X_A \text{ is chosen}) = \text{logit}^{-1}(\sum_{j=1}^{p} w_j(x_{Aj} - x_{Bj}))$. Note that without loss of generality, in all pairs the chosen alternative could be labeled as "A", this would make the dependent variable a constant. Thus, results of the logistic regression may depend on the labeling of alternatives.

The criteria of the apartment options had very different scales, ranging from tens of square meters to thousands of euros. To make the scales more comparable, we divided each criterion by the maximum value of that criterion for a subject. For example, if 1000 euros/month was the highest rent encountered by a subject, we divided the rent values for her options by this value.

3. The experiment

147 students from the Aalto University School of Science participated in the experiment. 20 subjects (13,6%) dropped out during the filling out of the questionnaire. 100 randomly selected subjects who completed the full experiment were rewarded with a

[1] We thank one anonymous reviewer for this suggestion.