European Journal of Operational Research 241 (2015) 458–468

Contents lists available at ScienceDirect

European Journal of Operational Research

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

Decision Support

Convergence properties and practical estimation of the probability of rank reversal in pairwise comparisons for multi-criteria decision making problems

Georgia Dede [a,b,*], Thomas Kamalakis [b,1], Thomas Sphicopoulos [a]

[a] Department of Informatics and Telecommunications, University of Athens, Panepistimiopolis, Ilisia, GR15784 Athens, Greece
[b] Department of Informatics and Telematics, Harokopio University, Eleftheriou Venizelou 70, Kallithea, GR17671 Athens, Greece

ARTICLE INFO

Article history:
Received 30 April 2013
Accepted 21 August 2014
Available online 28 September 2014

Keywords:
Uncertainty modeling
Multiple criteria analysis
Decision analysis
Pairwise comparisons

ABSTRACT

In this paper, we address the impact of uncertainty introduced when the experts complete pairwise comparison matrices, in the context of multi-criteria decision making. We first discuss how uncertainty can be quantified and modeled and then show how the probability of rank reversal scales with the number of experts. We consider the impact of various aspects which may affect the estimation of probability of rank reversal in the context of pairwise comparisons, such as the uncertainty level, alternative preference scales and different weight estimation methods. We also consider the case where the comparisons are carried out in a fuzzy manner. It is shown that in most circumstances, augmenting the size of the expert group beyond 15 produces a small change in the probability of rank reversal. We next address the issue of how this probability can be estimated in practice, from information gathered simply from the comparison matrices of a single expert group. We propose and validate a scheme which yields an estimate for the probability of rank reversal and test the applicability of this scheme under various conditions. The framework discussed in the paper can allow decision makers to correctly choose the number of experts participating in a pairwise comparison and obtain an estimate of the credibility of the outcome.

1. Introduction

Decision making (Bhushan & Rai, 2004; Yager, 2004) consists of choosing a specific course of action between several alternatives and is encountered in countless areas of human activity. In many circumstances, where complex decisions need to be made involving high stakes, it is desirable to proceed in a structured and methodological manner, rather than simply rely on the skills and intuition of a single decision maker. Multi-criteria decision analysis (MCDA) or multi-criteria decision making (MCDM) (Triantaphyllou, 2000) aim at facilitating decision makers in complicated situations where numerous and sometimes conflicting criteria or factors have to be taken into account.

MCDM is further classified into multi-objective decision making (MODM) and multi-attribute decision making (MADM) (Pohekar & Ramachandran, 2004). In MODM, a set of objective functions is optimized subject to constraints and hence the efficient solutions in a set of alternatives are sought. MODM typically requires the solution of a series of mathematical programming models in order to reveal implicitly defined efficient solutions. On the other hand, in MADM, a small number of pre-determined alternatives are to be evaluated under a common set of criteria and the best alternative is usually selected by making comparisons between alternatives with respect to each criterion.

A fundamental problem in decision making is to grade the importance of a set of alternatives and assign a weight to each of them. The importance of alternatives usually depends on several criteria which can be evaluated within the decision making framework in which pairwise comparisons (PWC) are an essential ingredient (Saaty & Vargas, 2001). In the context of MADM, PWC enables the ranking of alternatives by allowing the experts to compare the various criteria or alternatives in pairs instead of assigning their priorities in a single step (Saaty, 1977). This reduces the influence of subjective point of views, associated with eliciting weights directly. PWC is usually performed in MADM methods such as the Analytic Hierarchy Process (AHP) (Saaty, 2003), the Weighted Product Method (WPM) (Chang & Yeh, 2001), the preference ranking organization method for enrichment evaluation (PROMETHEE) (Brans, Vincke, & Mareschal, 1986), the Analytic Network Process (ANP) (Saaty, 2004) and so on. The aim

* Corresponding author at: National and Kapodistrian University of Athens, Department of Informatics and Telecommunications, Panepistimiopolis, Ilisia, GR15784 Athens, Greece. Tel.: +302107275313.

E-mail addresses: gdede@hua.gr, gdede@di.uoa.gr (G. Dede), thkam@hua.gr (T. Kamalakis), thomas@di.uoa.gr (T. Sphicopoulos).

[1] Tel.: +302109549286.

http://dx.doi.org/10.1016/j.ejor.2014.08.037
0377-2217/

of this paper is to consider the influence of uncertainty in PWC and how the credibility of the outcome can be assessed. Although the focus is on PWC alone, the results have some implications on the applications of MADM frameworks, particularly in the number of experts that are required and how the credibility of key parts of the framework can be ascertained.

In recent years, PWC has been used either as a stand-alone method or as part of complex MADM frameworks on several areas including government (Huanga, Chub, & Chiang, 2008), business (Lee & Kozar, 2006), industry (Chan, Lau & Ip, 2006), healthcare (Liberatore & Nydick, 2008), technology (Gerdsri & Kocaoglu, 2007), education (Zahedi, 1986), communications (Dede, Kamalakis, & Varoutas, 2011a, 2011b; Dede, Varoutas, Kamalakis, Fuentetaja, & Javaudin, 2010), agriculture (Abildtrup et al., 2006) and energy planning (Kok & Lootsma, 1985). The method itself has also been the focus of extensive research in the field of decision making. Recently in Fan and Liu (2010) , a form of uncertain preference information, called ordinal interval numbers, was used in pairwise comparisons in order to rank the alternatives. In Doumpos and Zopounidis (2004) , the issue of how pairwise comparisons can be used for the classification of alternatives in different classes of preference is discussed. The problem of deriving the weights from the pairwise comparison matrices using several alternative approaches is studied in Barzilai (1997) and Choo (2004) . In Kwiesielewicz and Van Uden (2004) several issues concerning the inconsistency of the pairwise comparison matrix and its impact on the decision making process are highlighted. In addition, fuzzy pairwise comparison for solving the decision making problems has been proposed in Boenderb, Graan, and Lootsma (1989), Deng (1999) and Mikhailov (2005) . In Marimin, Umano, Hatono, and Tamura (1998) linguistic labels are used in order to express fuzzy preference relations in pairwise decision problems. Moreover, Shiraishi, Tsuneshi, and Motomasa (1998) dealt with the properties of the principal eigenvector of PWC matrices.

The influence of uncertainty due to the imperfect and subjective expert judgments is of paramount importance when considering the credibility of the outcome of a decision making process. Several studies have attempted to shed some light on this issue in the context of PWC. For example, in Carmone, Karab, and Zanakis (1997), Monte Carlo simulations were performed to study the Incomplete Pairwise Comparisons (IPC) algorithm and investigate the effect of missing information in pairwise comparisons. Furthermore, in Aull-Hyde, Erdogan, and Duke (2006) it has been shown that given a sufficiently large group size, the consistency of the aggregate comparison matrix is guaranteed regardless of the measures used to estimate the consistency of the individual matrices, if the geometric mean method is used to estimate the weights. Moreover, in Hahn (2003) a stochastic characterization of the pairwise comparison judgments is provided, while statistical models for deriving the weights of the alternatives using Markov chain Monte Carlo are also presented. Furthermore, Farkas (2007) theoretically studied the conditions for rank reversal on perturbing the PWC matrices, while Chen and Kocaoglu (2008) also studied the rank reversal problem in this particular context, and came up with an algorithm to analyze the sensitivity of hierarchical decision models.

The main purpose of our work is to provide a suitable characterization of the impact of uncertainty in PWC. A first step in order to characterize the impact of uncertainty in PWCs, is to identify a suitable measure for quantifying its effects. Assume for instance that N different alternatives are pairwise compared by M experts, each with possibly a different view on the ranking of the alternatives. As discussed further below in Section 2.1, PWC aims at providing an average ranking, encompassing all these diverse opinions of the experts. It is of course natural to expect that the credibility of the overall process will be increased as the size of the expert group increases. Therefore, one possible way of measuring the trustworthiness of the results is to define the probability of rank reversal (P_{RR}) (Saaty & Vargas, 1984) as follows: Let $W_1, \ldots, W_N$ be the weights calculated by the PWC in the case of a very large group of experts ($M \to \infty$). In a practical situation where M is finite, uncertainty may undermine the PWC and the calculated weights w_k may turn out different than W_k. Uncertainty can be due to the difference of opinion among the experts or inconsistent pairwise comparisons. The probability of rank reversal is formally defined as:

$$P_{RR} = P\{\text{the ranking obtained by } w_i,\ 1 \le i \le N, \text{ is different than that of } W_i\} \quad (1)$$

A high P_{RR} implies that the outcome of the PWC in question is not trustworthy and could therefore lead to incorrect decision making. There are two important issues that need to be addressed concerning P_{RR}:

a) How does P_{RR} relate to the number of experts M? An obvious way to reduce the effect of uncertainty is simply to increase M, but from a practical point of view, this is not a trivial task. It is usually difficult to locate many experts within a single organization or even in the wider public with sufficient expertise that would be willing to participate in the PWC surveys. On the other hand, there is no clear answer to the question of how many more experts need to participate in order to considerably reduce the uncertainty of the outcome. Say for example, that there are already $M = 10$ experts participating in the endeavor. How much is there to be gained in terms of P_{RR} by doubling or even quadrupling the size of the group of experts?
b) How can P_{RR} be estimated from actual expert judgments in practice? When applying PWC in a specific decision making problem, one has access to the elements of a limited number of pairwise comparison matrices $\mathbf{P}^{(m)}$ (where $1 \le m \le M$). So the question becomes whether one can extract any kind of information regarding P_{RR} and hence the credibility of the results based on just the elements of $\mathbf{P}^{(m)}$.

The present paper attempts to deal with both points above. We first discuss a model for incorporating uncertainty in PWC and consider a suitable measure for quantifying the uncertainty level. We then discuss how P_{RR} varies with the group size M depending on the uncertainty level and extract several interesting conclusions from this variation. It is shown that there is not much sense in using more than $M = 15$ experts in the decision making process because the rate of decrease of P_{RR} is already small for $M > 15$. We then address the issue of how P_{RR} can be estimated from just the values of the pairwise comparison matrices $\mathbf{P}^{(m)}$ obtained by the experts. Given this information, we discuss a numerical method for estimating P_{RR} based on Monte Carlo simulation. The results indicate that for a sufficiently large group of experts, one can obtain a reasonable approximation to the actual value of P_{RR}.

The rest of the paper is organized as follows: In Section 2, we briefly summarize the PWC method laying the theoretical foundations for incorporating uncertainty and present the model used in our simulations. In Section 3, the results obtained when applying the proposed model are presented and the convergence of the P_{RR} is examined, considering the impact of various aspects which may affect the estimation of probability of rank reversal, such as varying uncertainty level among the experts, alternative preference scales and weight estimation methods. We also consider the case where the judgments are determined in a fuzzy manner. In Section 4 a numerical method is proposed in order to estimate the P_{RR} from the actual user judgments. Finally, some concluding remarks are presented in Section 5.

2. Uncertainty modeling in PWC

In this section, we introduce the model used for incorporating uncertainty in the pairwise comparison matrices. Section 2.1