FSS 4112
pp: 1-4 (col.fig.: Nil)
PROD. TYPE: COM
ED: DK
PAGN: Thanuja – SCAN: Nil

ARTICLE IN PRESS

Available at
www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT®

ELSEVIER

Fuzzy Sets and Systems III (IIII) III–III

FUZZY
sets and systems

www.elsevier.com/locate/fss

Book review

Evaluation and decision models: a critical perspective

D. Bouyssou, T. Marchant, M. Pirlot, P. Perny, A. Tsoukias, P. Vincke, (Ed.); Kluwer Academic Publishers, Boston/London/ Dordrecht, 2000, 274p. +viii

This is a rather unusual book on decision-making written by a group of scholars from Belgium and France. This is neither a monograph (that is, a research compendium) nor a tutorial volume. It is much more, a long discussion on the theoretical and practical aspects of decision sciences, with a view to emphasize two points: formal methods are needed to help decision-makers make rational choices; but any formal method has its own pitfalls and limitations, and should not be considered as a universal tool applicable to any decision problem. This point of view is the purpose of the introductory chapter.

The authors have not tried to present formal methods in detail. They systematically base their critical discussion on practical examples and elementary computations. The historical origin of methods is also pointed out. This approach makes this book attractive and easily read by non-mathematicians and especially by people involved in decision-making practice.

Chapter 2 is devoted to a critique of voting methods. It relies on the analysis of the French and the British voting systems. It shows paradoxical behaviors of both systems, and suggests that if a politician is elected using one method, it may be beaten using the other method.

Chapter 3 lays bare the arbitrary nature of grading methods in schools and universities. It explains that the process of defining a meaningful grading scale is far from obvious, let alone the issue of aggregating grades pertaining to distinct courses. The famous weighted sum method is not always capable of expressing any aggregation mode, and especially forbids interaction between criteria. The approach consisting in assigning qualitative grades (such as letters), translating them into arbitrary numbers, and performing their weighted average is especially pointed out as notoriously inconsistent. Namely, any change in the numerical encoding may lead to preference reversal effects, that is, altering the ranking of students obtained through numerical aggregation schemes.

Chapter 4 points out the limitations of statistical indicators such as I. Q., Dow Jones, air quality index, poverty index and the like. It demonstrates the fact that many of these indices are based on debatable measurement assumptions and arbitrary scale transformations, which severely hampers their meaningfulness. While they remain interesting indicators, they hardly capture all facets of the reality they are supposed to account for. The example of ranking athletes in decathlon competitions is especially enlightening, as a case where many adjustments to the performance indicator were introduced so as to tackle paradoxical behaviors of obvious measurement and aggregation schemes.

Chapter 5 points out the weakness of the famous cost–benefit analysis which is often claimed to be the only rational approach to investment decisions, such as budget allocations, building new

0165-0114/03/$ - see front matter © 2003 Published by Elsevier Science B.V.
doi:10.1016/S0165-0114(03)00038-1

roads, new hospitals and so on. It is shown that CBA takes a very narrow view on the problems they address. It is again based on a simplified view of reality. For instance parameters involved in the formal model, such as the social discount rate, are not so easy to determine. Even worse, the theoretical foundations of CBA reveal that it should be restricted to the study of marginal changes in the economy. In practice, it is however used for evaluating the consequences of important structural changes in the addressed system (hence a non-marginal change). Another tricky issue is the systematic use of money substitutes, like the cost of human life. Such a cost is for instance very much country-dependent, which is ethically questionable.

The longest chapter of the book is devoted to a comparative study of multicriteria decision-making methods. There is a very large body of literature devoted to rational decision based on several aspects or attributes. Roughly speaking there are two groups of approaches: those relying on numerical estimates and aggregation operations, and those based on the use of relations for the modeling and aggregation of preference information. Interestingly, the first group is popular in America, while the other group was basically developed in Europe. Various methods are compared on the basis of a single example of buying a sports car based on cost, acceleration pick up of the engine, brakes, and road-holding behavior. Methods considered are the weighted average of attribute-values, multicriteria utility theory, Saaty's analytic hierarchical process (AHP), and the outranking methods initiated by Roy and colleagues.

The car choice problem is first solved by intuitive considerations of a real person who did make a decision, and bought the car. Then, formal methods are used and their results compared to the choice made by this person. In the framework of numerical methods, it is shown that the meaning of criteria weights in weighted aggregations is problematic. Namely any change in measurement scales affects the order of magnitude of weights, even when restricting to weighted averages. This issue is especially critical if attribute values are aggregated without proper rescaling. The issue of numerical encoding of qualitative attribute values is again a tricky one. The merit of multicriteria utility theory is to properly address scaling problems, while at the same time distinguishing between values in the attribute scale, and decision-maker preference pertaining to these attribute values. The notion of utility function is tailored for laying bare this distinction, and partially solves difficulties of a blindly performed weighted average. The AHP method can be viewed as a systematic use of utility theory in a recursive way through a hierarchy of criteria. It proposes a technique for assessing weights in the weighted average of utility values. Unfortunately, it is shown that this techniques relies on an arbitrary translation of verbal levels of pairwise preference into numerical values (1,3,5,7,9). It creates significant differences between AHP and multicriteria utility theory. Especially, utility theory is based on an interval scale and leaves room for assessment methods based on indifference judgments. However the AHP is based on an absolute scale, and the debatable assumption that weights can be assessed by means of the same procedure at any level of the hierarchy. One weakness of multicriteria utility theory is its informational burden: the decision-maker is asked many questions and it is not clear that all answers can be provided. Also all attribute scales are supposed to be continuous (which is reasonable under the assumption that any relevant aspect of the decision-process can be expressed as a cost).

In practice, the available information supplied by decision-makers can be very poor and some attributes are more qualitative than quantitative, because they simply do not refer to any objectively measurable entity. In such a situation, it looks more appropriate to represent preference information along each attribute by means of

some sort of ordering relation comparing the merits of alternatives on a pairwise basis according to the decision-maker opinion. Local rankings of alternatives for each attributes are obtained. Global rankings are obtained using voting techniques. A decision outranks another one if it is better than the latter for a sufficient number of attributes.

This crude scheme is actually improved by means of weights attached to attributes, and possibly veto thresholds that prevent an alternative with a very bad evaluation on some attribute from outranking alternatives that are less attractive for other attributes. One difficulty of this approach is that it inherits all technical difficulties of voting methods, especially, local rankings are aggregated into a global outranking relation that does not necessarily provides a global ranking of alternatives in the end. This is due to the possibility of cycles in the final relation, due to Condorcet-effect, or the presence of incomparable alternatives. Whether such possibly poorly informed results are a good or a weak point actually depends on the application context.

Chapter 7 is of special interest for the readers of Fuzzy Sets and Systems as it proposes a discussion of fuzzy rule-based systems from the standpoint of decision making. While the authors acknowledge the merits of translating symbols appearing in rules by means of fuzzy intervals, they also point out that all the scaling difficulties encountered in multicriteria decision-making are met in fuzzy interpolative reasoning as well. The purpose of fuzzy sets is to make unrelated entities commensurate via membership functions. However this should be carried out with much care. Similar difficulties are pointed out in fuzzy versions of the k-nearest neighbor rule for classification purposes, even if the practical merits of these fuzzy methods are emphasized by the authors. They also report on an application to the control of ovens for cooking biscuits, where both fuzzy k-NN algorithms and fuzzy control techniques are conjointly used.

Chapter 8 discusses the importance of modeling and accounting for uncertainty in decision processes involving time and repeated decisions. The illustrative example is a case study in electricity production planning. It is shown that the classical expected utility criterion can be very difficult to apply in practice. However the use of other techniques may lead to new difficulties like dynamic inconsistencies and may fail to select non-dominated strategies.

Chapter 9 discusses a real-world case-study in software evaluation. The originality of this chapter is to adopt the point of view of the customer and to present the evaluation of the results of the study by the customer. It emphasizes the fact that the problem-formulation step is absolutely not a trivial matter, and that the evaluation model must take into account the value system of the customer. If the decision-maker cannot understand the decision-process he is involved in, the proposed method even if sophisticated and theoretically founded, will be rejected as unsuitable. This study also points out the necessity of exploiting the available information in a meaningful way, that is, one should avoid introducing arbitrary precision, and the method should correctly reflect the nature of the rating scales.

Overall, this book points out that despite their limitations, formal methods should be used in decision-making in order to better understand why an alternative is better than another. Formal methods force decision-makers to better justify their choices, and promote communication between various actors of the decision process. However by systematically pointing out pitfalls of the various existing approaches, the book strongly suggests that no approach is perfect, and that being aware of such limitations leads to a better, less naïve, use of these approaches.

As a matter of fact, it has been often assumed that the use of fuzzy sets can cope with the lack of objective attribute scale in decision-making

problems. Considering attributes such as comfort, trust-worthiness and the like, some authors have modeled linguistic terms pertaining to such attributes by fuzzy numbers on the unit interval, where a value in the unit interval represents a level of comfort, trust-worthiness, etc. It is often found that fuzzy number-extensions of weighted averages are used to perform aggregation of linguistic values pertaining to such complex, non-numerical attributes; fuzzy number ranking methods are then used to rank alternatives. These techniques look rather naïve and suffer from the same difficulties as some of the methods discussed and criticized in this book: arbitrary numerical encoding of linguistic levels, improper scaling, meaninglessness of weights and so on. The use of fuzzy numbers instead of precise value is somewhat delusive, all the more so as final rankings of alternatives are obtained via defuzzification schemes. In a nutshell, changing arbitrary numerical values, in an arbitrary numerical scale, into fuzzy intervals covering such a scale does not make obtained results more valid nor robust.

Reading this book could be beneficial to fuzzy decision-making scholars, because it would help them assess the merits and limitations of their fuzzy methods, often developed out of the main streams of decision research, and too seldom compared to established methodologies. The non-mathematical presentation of this book contrasts with the strength of its message, and recommends it to the attention of applied researchers in decision-making who used fuzzy set-based methods in case-studies. This book can be the bridge they need to cast their favorite fuzzy methods in the landscape of multicriteria decision-making methods. For this purpose, the impressive bibliography can be used as a starting point for further readings.

Didier Dubois
I.R.I.T.
Institut de Recherche en Informatique de Toulouse
France
E-mail address: didier.dubois@irit.fr