International Journal of Information Technology & Decision Making
Vol. 14 (2015)

DOI: 10.1142/S021962201550025X

When is a Decision-Making Method Trustworthy? Criteria for Evaluating Multi-Criteria Decision-Making Methods

Thomas L. Saaty
Graduate School of Business, University of Pittsburgh
Pittsburgh, PA 15260, USA
saaty@katz.pitt.edu

Daji Ergu
Southwest University for Nationalities
Chengdu 610041, China
ergudaji@163.com

Published 9 September 2015

Decision makers often face complicated decision problems with intangible and conflicting criteria. Numerous multi-criteria decision-making (MCDM) methods have been proposed to handle the measurement of the priorities of conflicting tangible/intangible criteria and in turn use them to choose the best alternative for a decision. However, the presence of many MCDM methods bewilders users. The existence of these methods becomes a decision problem in itself, and decision makers may be uncertain about which one to use. Thus the comparative analysis and evaluation of various MCDM methods has come under scrutiny by both researchers and practitioners in order to discover if there are logical, mathematical, social or practical reasons why one method is better than another. Criteria for their evaluation are the first important issue that needs to be resolved. In this paper, 16 criteria are introduced that may be used to judge and evaluate various MCDM methods. The criteria proposed and some guidelines for their evaluation are given to help readers evaluate these MCDM methods.

Keywords: Multi-criteria decision-making methods; conflicting criteria; decision problems; evaluation.

1. Introduction

The first author recently received the following message from a scholar in an Iranian university: "I'm an Iranian Operation Research PHD student, I'm very interested to continue my research and write my proposal on MCDM. But after many studies in this field and it's applications (such as AHP/ANP, ELECTRE, PROMETHE, TOPSIS, CHOQUET INTEGRAL, ...), I am confused and I feel that I don't know about the subject and think that other authors have worked on all the requirements of MCDM. Would you please help me to exit from this confusion and introduce me to some new research fields related to MCDM?"

Since the 1970s, multi-criteria decision-making (MCDM) research has developed rapidly and has become a hot research topic because many complex practical decision problems involve multiple and conflicting criteria as well as multiple objectives.[1] Over the past few decades a number of MCDM methods have been developed to deal with the measurement of tangible/intangible conflicting criteria and with the measurement of the alternatives of a decision with respect to these criteria. Some of the most popular methods include: the Analytic Hierarchy Process (AHP), the Analytic Network Process (ANP) (both are mathematical methods for measuring intangibles),[2–4] Additive Ratio Assessment (ARAS),[5] Complex Proportional Assessment of Alternatives (COPRAS), Compromise Programming (CP),[7] Decision Making Trial and Evaluation Laboratory (DEMATEL),[8] Dominance Based Rough Set Approach (DRSA),[9,10] Elimination Et Choice Translating Reality(ELECTRE),[11,12] Evidential Reasoning (ER),[13] GUESS method[14] Goal Programming (GP),[15] Grey Relational Analysis (GRA),[16] Inner Product of Vectors (IPV), Measuring Attractiveness by a Categorical Based Evaluation Technique (MACBETH).[17] Multi-Attribute Global Inference of Quality (MAGIQ),[18] Multi-attribute utility theory (MAUT),[19,20] Multi-attribute value theory (MAVT),[19] Maximal Entropy Ordered Weighted Averaging (ME-OWA).[21,22] New Approach to Appraisal (NATA), Nonstructural Fuzzy Decision Support System (NSFDSS)[23–25] Potentially All Pairwise Rankings of All Possible Alternatives (PAPRIKA),[26] Preference Ranking Organization Method for Enrichment of Evaluations (PROMETHEE),[27,28] Simple Additive Weighting (SAW),[29] Superiority and Inferiority Ranking Method (SIR),[30] Technique for Order Preference by Similarity to Ideal Solution (TOPSIS),[31] Utility Additive (UTA),[32] Value analysis (VA),[33] Value Engineering (VE),[34–36] VIKOR method,[37,38] Weighted Product Model (WPM),[39] Weighted Sum Model (WSM)[40,41] and on and on ad infinitum.

Many of the original MCDM methods have also been extended or adapted by the creators of those theories and by researchers on these methods. With these variations we have more than a hundred MCDM methods. This wide variety of available methods bewilders potential users (see the above message from the Iranian graduate student), resulting in the difficulty of selecting an appropriate method.[42] That is, the existence of many MCDM methods itself becomes a decision problem. In the next section, many studies that focus on the evaluation and comparison of different MCDM methods are reported on in order to analyze their similarities and differences or propose a procedure to guide users to select an appropriate MCDM method and explore if there are logical or mathematical, social or practical reasons why one method is better than another.

The question is how well does a MCDM method perform and how valid is its ranking and order? How do individuals provide judgment and how does a group combine the judgments of its members? Different MCDM methods should be evaluated and compared under a set of criteria since decision criteria are rules, measures and standards that guide decision making.

In this paper, 16 evaluation criteria are proposed to compare and evaluate MCDM methods. One may argue, as some have, that it is all a subjective matter, the judgments, the numbers used to represent them, why bother validating any method. It is precisely that the nearly 100 methods each claiming that its approach is the one to use that concerns the users who ask: which method is more reliable to yield the right answer? This is particularly relevant when the decision is complex and in addition involves Benefits, Opportunities, Costs, and Risks (BOCR). The arithmetic of each of the MCDM methods can be used in this evaluation process.

The rest of this paper is organized as follows. Section 2 provides a review of the literature on the comparative analysis of different MCDM methods. Evaluation criteria are defined and evaluation rules are proposed for them in Sec. 3. Section 4 concludes the paper.

2. Literature Review on the Comparative Analysis of MCDM Methods

MacCrimmon,[43] was probably the first to write about the need to compare MCDM methods and classify them. Subsequently some researchers proposed guidelines for selecting an appropriate MCDM method, while others applied several MCDM methods to the same decision problem, and compared the results obtained by those methods. For guidelines, Ozernoy[44] developed a conceptual framework for evaluating and comparing MCDM methods for a given decision situation. The proposed framework consists of five components: "*characteristics of different decision situations*", "*an extensive list of available MCDM discrete alternatives methods*", "*screening criteria and their criteria scales that can be used to eliminate those MCDM methods inappropriate in a particular decision situation*", "*evaluation criteria and their corresponding criteria scales which can be used to compare the resulting MCDM methods not eliminated by the screening criteria*", and "*a procedure for determining the MCDM method*(s) *most appropriate for the user in a given decision situation*". Again, Ozernoy[45] proposed a formal procedure to assist the decision maker to systematically select the most appropriate method in a particular application, whose steps are "*Development of the critical determinants of the selection process*" and "*Identification, acquisition, and representation of MCDM knowledge*". Guitouni and Martel[46] developed a conceptual framework for articulating seven tentative guidelines to choose an appropriate MCDM method, i.e., "*Determine the stakeholders of the decision process*", "*Consider the DM 'cognition' (DM way of thinking) when choosing a particular preference elucidation mode*", "*Determine the decision problematic pursued by the DM*", "*Choose the multicriteria aggregation procedure* (*MCAP*) *that can handle properly the input information available and for which the DM can easily provide the required information*", "*The compensation degree of the MCAP method is an important aspect to consider and to explain to the DM*", "*The fundamental hypotheses of the method are to be met* (*verified*)*, otherwise one should choose another method*", "*The decision support system coming with the method is an important aspect to be considered when the time comes to choose a MCDA method*".

Criteria describing and representing the characteristics of decision problems have been widely used to evaluate alternatives in MCDM problems. Raiffa and Keeney[19] suggested that the following five principles must be followed when formulating criteria: (1) completeness, (2) operationality, (3) decomposability, (4) non-redundancy, and (5) minimum size. Hobbs[42] proposed four criteria for choosing an MCDM method: *appropriateness, ease of use, validity and finally sensitivity of the results to the choice of a method*. He conducted an experiment that applied different methods to the same problem. Karni *et al.*[47] utilized three "real-life" cases to compare the rankings obtained by AHP, SAW, ELECTRE, and Weighted Linear Assignment Method (WLAM), and they found that the rankings obtained by AHP, SAW, and ELECTRE did not differ significantly, but WLAM did. Zanakis *et al.*[48] investigated the performance of eight methods: ELECTRE, TOPSIS, Multiplicative Exponential Weighting (MEW), Simple Additive Weighting (SAW), and four versions of AHP (original versus geometric scale and right eigenvector versus mean transformation solution) through a simulation experiment, and analyzed the results using 12 measures of similarity of performance. Salminen *et al.*[49] compared the results by applying ELECTRE II1, PROMETHEE I, II, and SMART to four different real applications to environmental problems in Finland, and they concluded that in many cases the results did not differ much, but in a particular problem the "best alternatives" may vary greatly, thus they recommended that "it is better to use several methods for the same decision problems when possible". Olson[50] applied SMART, PROMETHEE, and a centroid method to known baseball data over the period 1901–1991, and compared the results obtained by three MCDM methods. They found that PROMETHEE II with Gaussian preference functions and SMART were the most accurate. Cho[51] compared four methods by a simple example and gave two sets of criteria for users to evaluate the merits and shortcomings of each of these methods. Olson[52] compared TOPSIS with SMART method and centroid weighting schemes when different weighting schemes and different distance metrics were used in TOPSIS. Opricovic and Tzeng[53] analyzed the similarity and differences between TOPSIS and VIKOR methods. They extended the VIKOR method with weight stability intervals and trade-off analysis, and compared the results with TOPSIS, PROMETHEE, and ELECTRE. Peniwati[54] gave 16 criteria for evaluating group decision-making methods. Hajkowicz and Higgins[55] applied five multiple criteria analysis techniques, i.e., weighted summation, range of value, PROMTHEE II, Evamix, and compromise programming, to six water management decision problems in order to compare the results obtained. Peng *et al.*[56] proposed a novel fusion approach to produce a weighted compatible MCDM ranking of multiclass classification algorithms, which could resolve the conflicting rankings by obtaining a compromised solution. Geldermann and Schöbel[57] compared three Multi-Attribute Decision Making (MADM) methods, Simple Additive Rating (SAR), Simple Additive Weighting (SAW), and the outranking approach PROMETHEE and analyzed their similarities. They concluded that PROMETHEE can mimic other MCDM algorithms. Kou *et al.*[58] examined five MCDM methods and proposed an innovative approach to successfully resolve the conflicting MCDM rankings in the process of classification algorithms selection. Zolfani *et al.*[59] applied fuzzy AHP, SAW-G, and TOPSIS Gray to evaluate the performance of rural ICT centers (telecenters) in Iran. Mela *et al.*[60] studied six MCDM methods: the weighted sum method, the weighted product method, VIKOR, TOPSIS, PROMETHEE II, and a procedure based on the PEG-theorem in the context of building design. Kou *et al.*[61] investigated the judgment contradiction issue that occurred in MCDM decision matrices and developed innovative methods to improve the reliability of decision making. Velasquez and Hester[62] reviewed 11 MCDM methods and examined their advantages and disadvantages in order to provide a clear guide for how MCDM methods should be used in particular situations. Kou and Lin[63] proposed a flexible and efficient cosine maximization method for the priority vector derivation in the AHP. Anojkumara *et al.*[64] compared four MCDM methods (FAHP-TOPSIS, FAHP-VIKOR, FAHP-ELECTRE, and FAHP-PROMETHEE) for solving a pipes material selection problem in sugar industry. Tadić *et al.*[65] proposed a new hybrid MCDM model by combining fuzzy DEMATEL, fuzzy ANP, and fuzzy VIKOR methods. Kou *et al.*[66] developed a novel MCDM-based approach to evaluate the popular clustering algorithms and provide an extremely useful tool to evaluate clustering algorithms using a combination of validity measures. Chithambaranathan *et al.*[67] developed a gray-based hybrid framework using gray-based method ELECTRE and VIKOR approaches to evaluate the environmental performance of service supply chains.

As can be seen from the literature reviewed above, the major problem of MCDM methods is that different techniques can yield different results when applied to the same problem, and none of the MCDM methods can be considered as the 'super method' appropriate for all decision situations despite the existence of a large number of refined methods. There is also the claim that it is difficult to classify, evaluate, and compare different MCDM methods since they are based on a variety of assumptions about the decision maker's preferences and use different types of preference information. We think that the process of MCDM itself should make it possible to judge the generality and effectiveness of the methods.

3. Evaluation Criteria and Principles of Evaluation

A decision is a quadruple: good understanding of a choice problem to minimize doubt and uncertainty; a complete structure to represent all the factors involving criteria and alternatives; a cardinal scale or scales of measurement to represent judgments; priorities derived from the numerical judgments.

No matter which kind of MCDM method is used, a decision involves definition of the criteria and evaluation of the alternatives for each criterion. In this section, 16 criteria are defined for comparing and evaluating the various MCDM methods. Specifically, an MCDM method should be simple to execute and provide the capability to build a comprehensive decision structure, with breadth, depth, and merits (BOCR) as subcriteria. Once the decision structure is defined, a scale of measurement needs to be introduced to rate or judge the performance of each alternative with respect to each of the criteria, and to quantify the criteria (or determine their weights), then the alternatives are ranked using the scale. The method should be capable of dealing with the ranking of tangibles as well as intangibles and with rank preservation and reversal. The different priorities of the alternatives under each criterion should be synthesized by a merging function in order to obtain the final priorities of the alternatives.

To validate the feasibility and robustness of an MCDM method, sensitivity analysis should be conducted. To evaluate the MCDM methodologies, we defined six subcriteria such as Mathematical logical procedure; Applicability to intangibles; Justification of the approach and its axioms; Validation of the decision process in the prioritization of both tangibles and intangibles; Prediction of outcomes with tangibles and intangibles; Generality of the process to decisions with dependence and feedback; Applicability to conflict resolution as a decision process; and finally, Trustworthiness and validity of the approach. These criteria will be described in greater detail below. Moreover, rating principles are proposed under each criterion to assist the reader to evaluate the popular MCDM methods. The proposed 16 evaluation criteria and the related popular MCDM methods referred to in the literature mentioned are shown in Table 1.

(1) Simplicity of execution

As reviewed above, Hobbs[42] proposed "*ease of use*" as one of the criteria for choosing a MCDM method, here the first criterion we proposed is "*Simplicity of execution*" since our biological heritage to make decisions uses a style of execution that is inescapable yet fully justified so that a person versed in mathematics can understand without question its underlying logic and procedure. When it is being used, no expert is needed to supervise the process. After all, people and animals are born with the ability to make choices and an appealing process needs to come as close as possible to exploit this biological ability.

A method is rated low if its logic is complicated and can only be used by the professional decision makers or experts; medium if one needs to put much efforts to learn it; high if it can be easily understood and implemented by most users in practice.

(2) Comprehensive structure: breadth and depth

A comprehensive structure should consider the breadth and depth of a decision. A structure is considered to be broad if there are many independent distinct elements (criteria and subcriteria), and be broader if it contains loops and feedback. A structure is deep if each element is further decomposed to the most detailed elements.

A method is rated low if it contains only a few criteria without decomposition to sub-criteria, sub-sub-criteria with the necessary detail; medium if its structure is neither broad nor deep; high if its structure is not only broad (contains a number of distinct criteria) but also deep (criteria can be broken down to sub-criteria).

Table 1. The 16 evaluation criteria and the various MCDM methods.

Evaluation Criteria	Simplicity of execution	Comprehensiveness of structure			Measurement Scale	Ranking of tangibles	Synthesis of priorities by merging functions	Sensitivity analysis	
		Breadth	Depth	Merits (BOCR)					
Evaluation Criteria (continued)	Methodology							Applicability to conflict resolution	Trustworthiness and validity of the approach.
	Logical, mathematical procedure	Generalization to ranking of intangibles	Justification of the approach-Justifiable axioms		Validation of decision problems	Prediction of the outcome of decision with intangibles	Generality to dependence and feedback		
Methods	AHP, ANP, ARAS, COPRAS, CP, DEA, DEMATEL, DRSA, ELECTRE, ER, GP, GRA, GUESS, IPV, MACBETH, MAGIQ, MAUT, MAVT, ME-OWA, NATA, NSFDSS, PAPRIKA, PROMETHEE, SAW, SIR, TOPSIS, UTA, VA, VE, VIKOR, Voting, WPM, WSM, etc.								

(3) Comprehensive structure consisting of merit substructures

A decision structure is said to be comprehensive if it represents a decision problem by considering comprehensive influence factors such as BOCR from political, social, economic, legal, environmental, technological, and military perspectives. Each of the BOCR contributes to the merit of a decision and must be rated individually on a set of criteria. These ratings are called the merits of a comprehensive structure.

A method is rated NA if it does not involve BOCR analysis; low if it only considers one to two of the BOCR; medium if it involves three parts of BOCR; high if it considers all the BOCR merits.

(4) Logical, mathematical procedure

Mathematical logical procedure means there is a formal mathematical representation of the logic and reasoning behind a theory. Interpretation of the meaning of a method needs reliable expert knowledge; for example, 5 dollars is often adequate for buying lunch, but is useless when buying a Mercedes Benz. Therefore, an MCDM method should provide a complete solution with logical mathematical justifications.

A method is rated low if it involves only a simple mathematical logical procedure (e.g., arithmetic value utility); medium if it uses references sequence or relative difference to rank alternatives; high if it uses a pairwise comparison technique to determine the dominance of one criterion over another.

(5) Justification of the approach-justifiable axioms

Justification plays an important role in the process of decision making, usually three justifying issues should be considered: justification of the procedure; justification of the consequences of the procedure and justification of the approaches.[68] The latter is used to evaluate the consequences and run the procedure. It is one of the key issues in selecting an MCDM method because, if it is not justified in one or more aspects, especially in its axioms, the decision made by such a method could be doubtful. Hanne[1] proposed the axioms of rational behavior as a criterion for judging the foundation of MCDM methods in which axioms were defined as "nondeduced basic assumptions which concern properties of the preferences of a decision maker or properties of the working of an MCDM method or its results". A method is justifiable if it has a theory based on meaningful axioms.

A method is rated low if it involves no mathematics with axioms to justify its use; medium if it involves axioms only in part; high if it involves complete and logical axioms.

(6) Scales of measurement

Scales of measurement are used to categorize or quantify the decision variables. The most commonly used scales of measurement are ordinal (nominal, ordinal) and cardinal (interval, ratio, and absolute). Not all are useful in combining measurements under different criteria but only interval, ratio, and absolute scales. Cardinal scales are essential in MCDM analysis because of the basic need to make tradeoffs. When an interval scale (invariant under linear transformations with positive leading coefficient) is used to represent judgments differences between pairs of values can be compared. However, interval scale values cannot be added or multiplied but only averaged. All four arithmetic operations can be performed on numbers from a ratio scale (invariant under positive similarity transformation). Ratio scales are widely used in science for measurement. Absolute scales (invariant under the identity transformation) are the strongest type of measurement scales and are a special case of a ratio scale with the constant multiplier being equal to one.

A method is rated low if it uses the nominal or ordinal scale; medium if it uses interval scales or ratio scales; high if it uses absolute scales.

(7) Synthesis of judgments with merging functions

In practical decision making, we obtain a number of individual judgments or measurements from different experts, and usually there is a need to combine them by some merging or aggregation or consensus procedure.[69] As far as a specific MCDM decision problem is concerned, the performance of the criteria or attributes on the alternatives also needs to be combined by some merging or aggregation function in order to obtain an overall ranking set of numbers, known as priorities. For the BOCR merits of a decision, the overall results need to be synthesized in order to capture the final outcome of the entire process.[70]

A method is ranked low if it is synthesized by averaging weights; medium if a simple weighted method is used;. high if a rigorous merging function with reasonable weights is used.

(8) Ranking of tangibles

In a decision, ranking is a relationship between a set of alternatives so that, for any two alternatives, the first is either ranked 'higher than', 'lower than', or 'equal to' the second. In mathematics, this is known as a weak order or total preorder of objects. To aid decision makers make a decision, ranking makes it possible to evaluate complex information according to certain criteria and produce the total preorder of alternatives by balancing the preferences on all the criteria.

A method is rated NA if it does not involve some kind of ranking of the alternatives; low if uses a nominal scale; medium if it uses an ordinal scale; high if it uses a cardinal scale to rank alternatives.

(9) Generalization to ranking of intangibles

Intangible factors by definition have no scales of measurement, therefore applicability to intangibles involves the inclusion, and measurement, of the multidimensionality of the factors involved.[54] MCDM problems usually involve tangible and intangible criteria, and intangible criteria need to be quantified to obtain numerical values, thus an MCDM method needs to be applicable to intangibles as it does to tangibles.

A method is rated low if it simply quantifies intangibles by assigning arbitrary ordinal numbers; medium if it transforms intangibles into cardinal numbers by using an interval scale or a ratio scale or an absolute scale; high if it assesses intangibles by using pairwise comparison technique to consider their relative importance, either with regard to an ideal held in memory or relatively and according to the user's wishes.

(10) Rank preservation and reversal

There are many instances in real life where rank reverses. There is no mathematical theorem which proves that rank must always be preserved and there cannot be in view of the counterexamples. Luce and Raiffa in their well-known book *Games and Decisions* worked hard to axiomatize (not prove as a theorem) rank preservation and gave up. After all, we have numerous counter examples and it is hard to know when rank should not be always preserved even if that seems unpalatable in some decisions. Rank preservation came to be accepted as a dogma because of techniques that could only rate alternatives one at a time as if they are independent of one another. They rate the alternatives by comparing each of them with an ideal alternative. But the ideal itself is derived from all the relevant alternatives that exist and thus indirectly each alternative depends in its rank on all those alternatives used to form the ideal. Thus while assuming independence, it turns out that interdependence lies hidden within the choice of the ideal. Thus all methods that only rate alternatives one at a time with an ideal may not be correct in their ranking, even if they are right at odd times. One way to more or less successfully rate alternatives one at a time when one is evaluating hundreds of alternatives is to compare some of the resulting top ones from the rating to produce a ranking that considers interdependence.

A method is rated low if it simply assumes rank preservation; medium if it basically deals with both rank preservation and rank reversal; high if it is capable of interpreting the reasons for rank preservation and rank reversal and implements the idea in its procedures.

(11) Sensitivity analysis

Sensitivity analysis is a widely used method for quantitative model assessment in many disciplines in order to validate the feasibility, robustness, and reliability of a model or a method. Saltelli *et al.*[71] wrote that sensitivity is used "to ascertain how a given model (numerical or otherwise) depends on its input factors". They defined sensitivity analysis as "the study of how uncertainty in the output of a mathematical model or system (numerical or otherwise) can be apportioned to different sources of uncertainty in its inputs". To practitioners, the most important purpose is to understand the sensitivity of a model's outputs to simultaneous variations in several input parameters.[73] Sensitivity analysis is a way to predict the outcome of a decision if a situation turns out to be different compared to the key prediction(s).[70] Therefore, an MCDM method should be capable of assessing the stability and validity of outputs to imprecise values for the various parameters of the model.

A method is rated low if it involves only a single parameter; medium if it works on two to three parameters; high if it is capable of assessing more than three parameters.

(12) Validation of decision problems

A method that relies on judgments but without knowledge and experience is highly subjective in judging decision problems with intangibles and is tantamount to garbage in garbage out. Rationality, interpreted as "wanting more rather than less of a good", is widely used as an assumption of the behavior of individuals but there are numerous counter examples to this axiom. A rational decision is one that is not just reasoned, but is also optimal in achieving a goal or solving a problem. Any process of evaluation or analysis that may be called rational is expected to be highly objective, logical and meaningful. Thus, scientific theories must be verified against things that happen in the real world and must follow the guiding principles of science.[74] One of the best ways to validate a scientific theory is to show that the results predicted by the theory give correct answers, i.e., they match known results,[75] especially for decision problems with intangibles. Validation is a critical factor in a field where subjectivity tends to override other considerations.

A method is rated low if it does not involve justifiable validation of some kind; medium if it only uses examples with tangibles to validate its procedures; high if it uses real-world examples with intangibles to validate its effectiveness.

(13) Prediction of the outcome of decisions with intangibles (The capability to predict preferences)

According to the philosopher of science Karl Popper, a scientific theory must have specific testable predictions that can be used to falsify or validate it. A prediction is a statement about the way things will occur in the future. In science, a prediction is a rigorous, often quantitative, statement, forecasting what will happen under specific conditions. Prediction also uses some of the most important objective mathematical models. However, in decision making, the quality of prediction should be assessed in advance or the outcome should be insensitive to the uncertainty in prediction. Therefore, prediction has been regarded as a criterion for judging the reliability of an MCDM method. For instance, Khairullah and Zionts[76] proposed the capability to predict preference as a criterion for judging a method, that is, prediction of a complete rank ordering of alternatives and prediction of the best alternative. Park[77] applied the prediction criterion to judge an MCDM method. While Currim and Sarin[78] compared several consumer preference models with respect to the predictive accuracy criterion. Thus, we defined prediction of decision problems with intangibles as a criterion for evaluating various MCDM methods.

A method is rated low if it just plays with numbers by pretending what it does is correct and can be taken for granted; medium if it does a credible job without the need for much explanation or argument; high if it directly involves prediction of a decision (the predicted results matches the outcome of that decision).

(14) Generalizability to dependence and feedback

In the process of decision making, dependence and feedback are critical for making a rational decision. Feedback enables the decision maker to factor the future into the present to determine what he must do to attain a desired future. Also feedback can cause an unimportant element to become important.[79] An MCDM method should have the possibility for generalization to dependence and feedback in order to reflect the real state of a decision and adjust that decision accordingly.

A method is rated NA if it does not involve feedback in the process of making a decision; low if either dependence or feedback is used indirectly, medium if both partially influence the results of decision making; high if they are regarded as parts of the inputs of the decision itself that can affect the results of that decision.

(15) Applicability to conflict resolution

If individuals or groups with different objectives attempt to satisfy their own objectives, conflict of some kind will occur.[70] Applicability to conflict resolution means that the method should provide a way for each party in the conflict to evaluate the costs and the benefits to both sides to determine "fair" tradeoffs by giving up some of what it has, in return for getting what it wants from the other party.[54] An MCDM method needs to have an approach, and perhaps also normative standards, to find the best solution for a group conflict that is understandable, acceptable, practical, flexible, and has been demonstrated to work well in practice.

A method is rated low if it uses a simple mathematical compensation technique to make tradeoffs in a conflict; medium if it uses an analytical method for dealing with conflict resolution; high if it can provide the best solution for a group conflict that is understandable, acceptable, practical, and flexible.

(16) Trustworthiness and validity of the approach

For analyzing MCDM problems, the trustworthiness and validity of the chosen MCDM method is essential. Trustworthiness has been defined as that quality of a method and its findings that make it noteworthy to decision makers. There are four major concerns relating to trustworthiness: truth value (internal validity), applicability (external validity), consistency, and neutrality (objectivity).[80] Validity indicates that the method is likely to yield choices that reflect accurately the values of the user, which can be interpreted on a rational-axiomatic, empirical, or measurement theoretical basis.[81] Researchers often measure validity by checking how well a given method predicts the unaided decisions made independently of judgments used to fit the model. Validity of an assessment is the degree to which it measures what it is supposed to measure. Therefore, an MCDM method should generate an outcome that is valid and generally useful for different types of decisions.

A method is rated low if it uses cardinal measurement model with a simple structure; medium if it uses cardinal measurement model but does not provide rigorous mathematical axioms; high if it uses a cardinal measurement model with a mathematical logical procedure and mathematical axioms.

4. Conclusion

The numerous MCDM methods on the one hand have promoted and boosted the development of decision theories and helped decision makers make decisions. On the other hand, with increase in MCDM methods, decision makers face the challenge to determine which method is worthy of consideration. In this paper, we defined 16 criteria to be used to judge and evaluate the various MCDM methods.

We have intentionally avoided evaluating these methods ourselves due to the possibility of bias particularly because one of us is responsible for two of these methods. We have made a reasonably general designed approach for the evaluation of MCDM methods. We hope that along with what other authors have tried to do we have a meta-science for assessing scientific theories used in making decisions.

The AHP comes the closest to our intuition than any other theory we know about and its mathematics easily generalizes to interdependence and feedback. Also it does BOCR and combines them by using strategic criteria which to the best of our knowledge no other theory does. One would expect any theory claiming accuracy to share similar results for simple decisions. They cannot and would not for complex decisions because the AHP uses absolute scales of priorities that can be combined to also produce absolute numbers.

Data are not something sacred. They are measurement to be interpreted. The AHP can trivially use data of a ratio scale kind which, on normalization, lead to absolute numbers or appropriately transformed numbers from an interval scale which on forming differences become ratio scale numbers. However, in general, it is our mind that judges how useful and important data are by using judgment to interpret their significance under the appropriate criterion or criteria that need them. Even in science it is the experts who set the dogma by interpreting the information from data, and they are sometimes wrong and the theory changes. There is no theory that is true forever. We and our own satisfaction must be fulfilled. We are so complex that we are gradually learning how our electrical brains work. Everything depends on that. Consistency of our judgment is important for identifying the "truth". But consistency is only necessary but not sufficient to capture reality. A crazy person can be perfectly consistent about a nonexistent world.

Acknowledgments

The second author wishes to express his thanks for the grants from the National Natural Science Foundation of China #71373216, #71222108 and #91224001 to support his participation in this research.

References

1. T. Hanne, *Intelligent Strategies for Meta Multiple Criteria Decision Making* (Kluwer Academic Publishers, 2001).
2. T. L. Saaty, An Eigenvalue allocation model in contingency planning, University of Pennsylvania **19** (1972) 72.

3. T. L. Saaty, *The Analytical Hierarchical Process* (John Wiley, New York, 1980).
4. T. L. Saaty, The analytic hierarchy and analytic network processes for the measurement of intangible criteria and for decision-making. in *Multiple Criteria Decision Analysis: State of the Art Surveys* (Springer, 2005), pp. 345–405.
5. Z. Turskis and E. K. Zavadskas, A new additive ratio assessment (ARAS) method in multicriteria decision-making, *Technological and Economic Development of Economy* **16**(2) (2010) 159–172.
6. E. Zavadskas and A. Kaklauskas, *Multicriteria Evaluation of Building* (Pastatu sistemotechninis ivertinimas). (Technika, Vilnius, 1996).
7. M. Zeleny, Compromise programming, in *Multiple Criteria Decision Making* (1973), pp. 262–301.
8. A. Gabus and E. Fontela, World problems, an invitation to further thought within the framework of DEMATEL, DEMATEL Report NO. 1 Battelle Geneva Research Center, Geneva, Switzerland (1972).
9. S. Greco, B. Matarazzo *et al.*, Rough approximation of a preference relation by dominance relations, *European Journal of Operational Research* **117**(1) (1999) 63–83.
10. S. Greco, B. Matarazzo *et al.*, Rough approximation by dominance relations, *International Journal of Intelligent Systems* **17**(2) (2002) 153–171.
11. B. Roy, R. Benayoun *et al.*, ELECTRE Société d'Economie et de Mathématique appliquées, Paris, 1996 (in French.)
12. B. Roy, The outranking approach and the foundations of ELECTRE methods, in *Readings in Multiple Criteria Decision Aid* (Springer, 1990), pp. 155–183.
13. J. D. Lowrance and T. D. Garvey, Evidential reasoning: A developing concept, in *IEEE Int. Conf. on Cybernetics and Society* (Seattle, WA, 1982), pp. 6–9.
14. J. T. Buchanan, A naive approach for solving MCDM problems: The GUESS method, *Journal of the Operational Research Society* **48**(2) (1997) 202–206.
15. S. M. Lee, *Goal Programming for Decision Analysis* (Auerbach, Philadelphia, 1972).
16. J.-L. Deng, Introduction to grey system theory, *The Journal of Grey System* **1**(1) (1989) 1–24.
17. C. A. Bana e Costa and J.-C. Vansnick, MACBETH — An interactive path towards the construction of cardinal value functions, *International transactions in operational Research* **1**(4) (1994) 489–500.
18. J. D. McCaffrey, Using the Multi-Attribute Global Inference of Quality (MAGIQ) technique for software testing, *Sixth Int. Conf. Information Technology: New Generations, 2009. ITNG'09* (IEEE, 2009), pp. 738–742.
19. R. L. Keeney and R. L. Keeney, *Decisions with Multiple Objectives: Preferences and Value Tradeoffs* (John Wiley and Sons, New York, 1976).
20. P. Humphreys, Application of multi-attribute utility theory, *Decision Making and Change in Human Affairs* (Springer, 1977), pp. 165–207.
21. R. R. Yager, On ordered weighted averaging aggregation operators in multicriteria decisionmaking, *IEEE Transactions on Systems, Man and Cybernetics* **18**(1) (1988) 183–190.
22. D. Filev and R. R. Yager, Analytic properties of maximum entropy OWA operators, *Information Sciences* **85**(1) (1995) 11–27.
23. C. Tam, T. K. Tong *et al.*, Site layout planning using nonstructural fuzzy decision support system, *Journal of Construction Engineering and Management* **128**(3) (2002) 220–231.
24. C. Tam, T. K. Tong *et al.*, Non-structural fuzzy decision support system for evaluation of construction safety management system, *International Journal of Project Management* **20**(4) (2002) 303–313.

25. C. Tam, T. K. Tong *et al.*, Comparing non-structural fuzzy decision support system and analytical hierarchy process in decision-making for construction problems, *European Journal of Operational Research* **174**(2) (2006) 1317–1324.
26. P. Hansen and F. Ombler, A new method for scoring additive multi-attribute value models using pairwise rankings of alternatives, *Journal of MultiCriteria Decision Analysis* **15**(34) (2008) 87–107.
27. J.-P. Brans, P. Vincke *et al.,* How to select and how to rank projects: The PROMETHEE method, *European Journal of Operational Research*, **24**(2)(1986) 228–238.
28. J.-P. Brans and B. Mareschal, PROMETHEE methods, in Multiple Criteria Decision Analysis: State of the Art Surveys (Springer, 2005) 163–186.
29. C. W. Churchman and R. L. Ackoff, An approximate measure of value, *Operations Research* **2**(2) (1954) 172–187.
30. X. Xu, The SIR method: A superiority and inferiority ranking method for multiple criteria decision making, *European Journal of Operational Research* **131**(3) (2001) 587–602.
31. C.-L. Hwang and K. Yoon, *Multiple Attribute Decision Making* (Springer, 1981).
32. E. Jacquet-Lagreze and J. Siskos. Assessing a set of additive utility functions for multicriteria decision-making, the UTA method, *European Journal of Operational Research* **10**(2) (1982) 151–164.
33. P. G. Keen, Value analysis: Justifying decision support systems, *MIS Quarterly* **5**(1) (1981) 1–15.
34. L. W. Zimmerman and G. D. Hart, *Value Engineering: A Practical Approach for Owners, Designers, and Contractors* (Van Nostrand Reinhold New York, USA, 1982).
35. D. E. Parker, *Value Engineering Theory* (Lawrence D Miles Value Foundation, 1985).
36. A. E. Mudge, *Value Engineering: A Systematic Approach* (J. Pohl Associates, 1989).
37. S. Opricović, An extension of compromise programming to the solution of dynamic multicriteria problem, *Optimization Techniques* (Springer, 1980), pp. 508–517.
38. S. Opricovic and G.-H. Tzeng, Extended VIKOR method in comparison with outranking methods, *European Journal of Operational Research* **178**(2) (2007) 514–529.
39. E. Triantaphyllou and S. H. Mann, An examination of the effectiveness of multidimensional decision-making methods: A decision-making paradox, *Decision Support Systems* **5**(3) (1989) 303–312.
40. P. C. Fishburn, Letter to the editor — additive utilities with incomplete product sets: Application to priorities and assignments, *Operations Research* **15**(3) (1967) 537–542.
41. D. Von Winterfeldt and W. Edwards, *Decision Analysis and Behavioral Research* (Cambridge University Press, Cambridge, 1986).
42. B. F. Hobbs, What can we learn from experiments in multiobjective decision analysis?, *IEEE Transactions on Systems, Man and Cybernetics* **16**(3) (1986) 384–394.
43. K. R. MacCrimmon, An overview of multiple objective decision making, in Cochran J. L., Zeleny M. (Eds.), *Multiple Criteria Decision Making*, University of South Carolina Press (Columbia, 1973), pp. 18–44.
44. V. M. Ozernoy, A framework for choosing the most appropriate discrete alternative multiple criteria decision-making method in decision support systems and expert systems *in Toward Interactive and Intelligent Decision Support Systems* (Springer, 1987) 56–64.
45. V. M. Ozernoy, Choosing the "best" multiple criteria decision-making method, *Infor* **30**(2) (1992) 159–171.
46. A. Guitouni and J.-M. Martel, Tentative guidelines to help choosing an appropriate MCDA method, *European Journal of Operational Research* **109**(2) (1998) 501–521.
47. R. Karni, P. Sanchez *et al.*, A comparative study of multiattribute decision making methodologies, *Theory and Decision* **29**(3) (1990) 203–222.

48. S. H. Zanakis, A. Solomon *et al.*, Multi-attribute decision making: A simulation comparison of select methods, *European Journal of Operational Research* **107**(3) (1998) 507–529.
49. P. Salminen, J. Hokkanen *et al.*, Comparing multicriteria methods in the context of environmental problems, *European Journal of Operational Research* **104**(3) (1998) 485–496.
50. D. L. Olson, Comparison of three multicriteria methods to predict known outcomes, *European Journal of Operational Research* **130**(3) (2001) 576–587.
51. K. T. Cho, Multicriteria decision methods: An attempt to evaluate and unify, *Mathematical and Computer Modelling* **37**(9) (2003) 1099–1119.
52. D. L. Olson, Comparison of weights in TOPSIS models, *Mathematical and Computer Modelling* **40**(7–8) (2004) 721–727.
53. S. Opricovic and G.-H. Tzeng, Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS, *European Journal of Operational Research* **156**(2) (2004) 445–455.
54. K. Peniwati, Criteria for evaluating group decision-making methods, *Mathematical and Computer Modelling* **46**(7) (2007) 935–947.
55. S. Hajkowicz and A. Higgins, A comparison of multiple criteria analysis techniques for water resource management, *European Journal of Operational Research* **184**(1) (2008) 255–265.
56. Y. Peng, G. Kou, G. X. Wang and Y. Shi, FAMCDM: A fusion approach of MCDM methods to rank multiclass classification algorithms, *Omega-International Journal of Management Science* **39**(6) (2011) 677–689.
57. J. Geldermann and A. Schöbel, On the similarities of some multicriteria decision analysis methods, *Journal of Multi-Criteria Decision Analysis* **18**(3–4) (2011) 219–230.
58. G. Kou, Y. Lu, Y. Peng and Y. Shi, Evaluation of classification algorithms using MCDM and rank correlation, *International Journal of Information Technology & Decision Making* **11**(1) (2012) 197–225, doi: http://dx.doi.org/10.1142/S0219622012500095.
59. S. H. Zolfani, M. Sedaghat and E. K. Zavadskas, Performance evaluating of rural ICT centers (telecenters), applying fuzzy AHP, SAW-G and TOPSIS Grey, a case study in Iran, *Technological and Economic Development of Economy* **18**(2) (2012) 364–387.
60. K. Mela, T. Tiainen *et al.*, Comparative study of multiple criteria decision making methods for building design, *Advanced Engineering Informatics* **26**(4) (2012) 716–726.
61. G. Kou, D. Ergu and J. Shang, Enhancing data consistency in decision matrix: Adapting hadamard model to mitigate judgement contradiction, *European Journal of Operational Reseach* **236**(1) (2013) 261–271.
62. M. Velasquez and P. T. Hester, An analysis of multi-criteria decision making methods, *International Journal of Operations Research* **10**(2) (2013) 56–66.
63. G. Kou and C. Lin, A cosine maximization method for the priority vector derivation in AHP, *European Journal of Operational Reseach* **235** (2014) 225–232.
64. L. Anojkumara, M. Ilangkumarana and V. Sasirekha, Comparative analysis of MCDM methods for pipe material selection in sugar industry, *Expert Systems with Applications* **41**(6) (2014) 2964–2980.
65. S. Tadić, S. Zečević *et al.*, A novel hybrid MCDM model based on fuzzy DEMATEL, fuzzy ANP and fuzzy VIKOR for city logistics concept selection, *Expert Systems with Applications* **41**(18) (2014) 8112–8128.
66. G. Kou, Y. Peng and G. Wang, Evaluation of clustering algorithms for financial risk analysis using MCDM methods, *Information Sciences* **27** (2014) 1–12.
67. P. Chithambaranathan, N. Subramanian, A. Gunasekaran and P. L. K. Palaniappan, Service supply chain environmental performance evaluation using grey based hybrid MCDM approach, *International Journal of Production Economics* **166** (2015) 163–176.

68. R. Piroozfar, *Fuzzy Logic: A Rule-Based Approach, in Search of a Justified Decision-Making Process in Urban Planning* (Universitätsbibliothek, 2010).
69. J. Aczel and F. S. Roberts, On the possible merging functions, *Mathematical Social Sciences* **17**(3) (1989) 205–243.
70. T. L. Saaty, Relative measurement and its generalization in decision making-why pairwise comparisons are central in mathematics for the measurement of intangible factors the analytic hierarchy/network process (To the memory of my beloved friend Professor Sixto Rios Garcia), *Revista De La Real Academia De Ciencias Exactas Fisicas Y Naturales Serie a-Matematicas* **102**(2) (2008) 251–318.
71. A. Saltelli, S. Tarantola *et al.*, A quantitative model-independent method for global sensitivity analysis of model output, *Technometrics* **41**(1) (1999) 39–56.
72. A. Saltelli, M. Ratto *et al.*, *Global Sensitivity Analysis: The Primer* (Wiley-Interscience, 2008).
73. H. M. Wagner, Global sensitivity analysis, *Operations Research* **43**(6) (1995) 948–969.
74. L. L. LeShan and H. Margenau, *Einstein's Space and Van Gogh's Sky: Physical Reality and Beyond* (Macmillan, New York, 1982).
75. T. L. Saaty and N. Begicevic, The scope of human values and human activities in decision making, *Applied Soft Computing* **10**(4) (2010) 963–974.
76. Z. Y. Khairullah and S. Zionts, An experiment with some algorithms for multiple criteria decision making, in *Multiple Criteria Decision Making Theory and Application*, (Springer, 1980), pp. 150–159.
77. C. W. Park, A conflict resolution choice model, *Journal of Consumer Research* **5**(2) (1978) 124–137.
78. I. S. Currim and R. K. Sarin, A comparative evaluation of multiattribute consumer preference models, *Management Science* **30**(5) (1984) 543–561.
79. T. L. Saaty, Fundamentals of the analytic network process — Dependence and feedback in decision-making with a single network, *Journal of Systems Science and Systems Engineering* **13**(2) (2004) 129–157.
80. E. G. Guba, Criteria for assessing the trustworthiness of naturalistic inquiries, *ECTJ* **29**(2) (1981) 75–91.
81. B. F. Hobbs, V. Chankong *et al.*, Does choice of multicriteria method matter? An experiment in water resources planning, *Water Resources Research* **28**(7) (1992) 1767–1779.