Handling uncertainty in DEX methodology

Martin Žnidaršič
Jožef Stefan Institute, Jamova cesta 39, martin.znidarsic@ijs.si

Marko Bohanec
Jožef Stefan Institute, Jamova cesta 39, marko.bohanec@ijs.si,
and University of Nova Gorica

ABSTRACT. The paper presents approaches to representation and use of uncertain knowledge in decision models of DEX methodology. A short introduction to this methodology is given, followed by a presentation of representation methods for uncertain input values and uncertain value functions. The emphasis is on the latter, since these were recently developed and adopted. Attention is given also to the representation method of higher level uncertainty in value functions, which represents a current response to the practical needs we identified, but remains open to further theoretical improvement.

KEYWORDS. Uncertainty; DEX methodology; Confidence; Higher-level uncertainty; MCDM.

1. INTRODUCTION

Uncertainty is an unavoidable companion of complex decision making problems. Dealing with uncertainty can be done in different ways, but it is limited by the representational capabilities of the decision support tools used. Enabling representation of uncertainties in such tools is therefore important and beneficial. However, uncertainty representational capabilities usually come at a price of higher workload during decision making and lower comprehensibility of decision support processes and results.

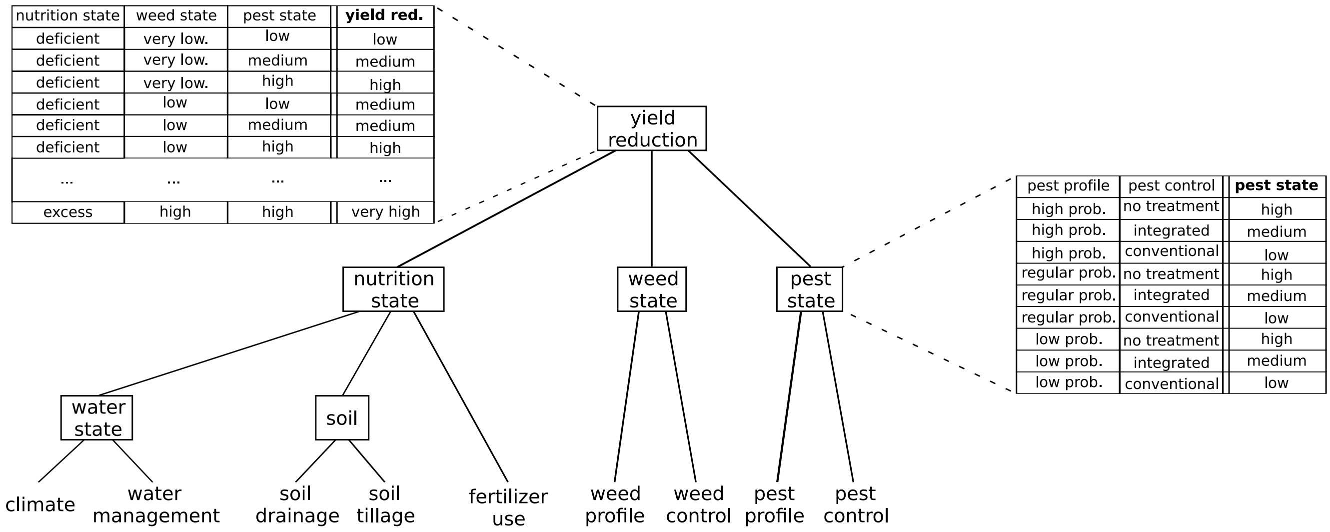
As decision analysts in several problems from a fresh field of research (impacts of agriculture with genetically modified crops) we encountered many situations when representation of uncertainty in decision models was beneficial or even asked for by domain experts. To cope with such demands of the problems and wishes of knowledge providers, we extended the DEX decision modeling methodology, which was used in these problems, with capabilities for uncertain knowledge representations.

The basic DEX methodology is briefly presented in section 2. The first extension, described in section 3.2, was aimed at enabling the model developer to express uncertain knowledge by providing values in form of probabilistic distributions. However, distributions did not solve the issue of marking parts of the model as based on knowledge of lower or higher confidence. A new parameter was therefore introduced for this purpose, along with approaches of its inference as a second extension that is described in section 3.3.

The presented approach covers most of the practical demands, but there are some details that have several proposed solutions or would need theoretical improvements. Section 4 is devoted to this discussion and ideas for further work.

2. DEX METHODOLOGY BASICS

DEX is a decision modeling methodology that is well established in practice (Bohanec and Rajkovič, 1999; Bohanec et al., 2003), but without much attention in previews (Triantaphyllou, 2000; Figueira, 2005; Bouyssou et al., 2006) of decision modeling approaches. DEX combines the multi-criteria and rule-based approaches. Its main characteristic is its qualitative nature: the input values are expected to be qualitative or discretized and the value functions are rule-based, usually represented in tabular form as for example the right one in Fig. 1, which transforms the values of input attributes *pest control* and *pest profile* into values of an aggregate attribute *pest state*. A function of this kind would be defined for all aggregate attributes (boxed ones in Fig. 1).



nutrition state	weed state	pest state	yield red.
deficient	very low.	low	low
deficient	very low.	medium	medium
deficient	very low.	high	high
deficient	low	low	medium
deficient	low	medium	medium
deficient	low	high	high
...	...	...	...
excess	high	high	very high

pest profile	pest control	pest state
high prob.	no treatment	high
high prob.	integrated	medium
high prob.	conventional	low
regular prob.	no treatment	high
regular prob.	integrated	medium
regular prob.	conventional	low
low prob.	no treatment	high
low prob.	integrated	medium
low prob.	conventional	low

Figure 1. Example of a DEX model. Example is a simplification of a small part of a model of ecological and economical impacts of cropping systems that was developed for the project SIGMEA (2004-2006). Value function for *pest state* attribute and part of value function for *yield reduction* attribute are shown in the picture.

Models with defined hierarchical structure and value functions are used to evaluate alternatives and analyze (what if analysis, sensitivity analysis, etc.) the evaluation. An alternative is a vector of values of the lowest-level (input) attributes. If we follow the example depicted in Fig. 1, an alternative would be a vector of values for *climate*, *water management*, ..., *pest profile* and *pest control*. Values of input attributes get aggregated into values of higher-level attributes according to their value functions. For example, if *pest profile* has the value *high problem* and *pest control* has the value *no treatment*, then the value of the aggregated attribute *pest state* is *high* according to the first row in the value function of *pest state*. Values of all the other input attributes get aggregated into values of higher-level attributes in a similar way until we obtain the value for the highest level attribute, *yield reduction* in this example. Usually several alternatives are evaluated and analyzed with a model in order to find the most appropriate one, to rank them or simply to get an insight into the decision problem at hand.

Some aspects of value functions uncertainty, like imprecision for example, are covered already by the qualitative nature of the values used in the models, some other aspects demand extensions of the methodology and are covered in section 3.

The method DEX is supported by software DEXi (Bohanec, 2008), which is freely available from http://kt.ijs.si/MarkoBohanec/dexi.html. DEXi facilitates the development of a tree of attributes, definition of decision rules, evaluation and analysis of alternatives, and making reports and graphical presentations. DEXi has been used in many real-life decision problems in the areas such as selection and evaluation of computer hardware and software, evaluation of companies and business partners, personnel management, project evaluation, land-use planning, risk assessment in medicine and health-care. Recently, complex decision models and applications using DEXi and prototype software with methodological extensions (Žnidaršič et al., 2006; Žnidaršič et al. 2008) have been developed in European projects ECOGEN (2003-2006), SIGMEA (2004-2006) and Co-Extra (2006-2009). These projects addressed the impacts of using genetically modified crops in food and feed production and supply chains. The developed models included a model for the assessment of impacts of cropping systems on soil quality (Bohanec et al., 2007) and a model for economic and ecological assessment of cropping systems involving genetically modified maize (Bohanec et al., 2008).

3. UNCERTAIN KNOWLEDGE REPRESENTATION

Qualitative hierarchical decision analysis models have various components that the developer or user can be uncertain about. To start from the most fundamental, the *hierarchic structure* itself can be seen as uncertain. A common observation from practice is that the model developers see some parts of the hierarchy as obvious and unquestionable, but might feel less confident about some others. However, as various hierarchies can serve the purposes of decision analysis equally well and as representation and use of uncertainty in hierarchical structure in general holds little practical value, we do not elaborate on this aspect here. Other components that can be uncertain are the *value functions*, represented with groups of if-then rules in case of DEX methodology. The approach to representing uncertainty in these value functions is presented in sections 3.2 and 3.3. Finally, there are the *inputs*, which are not a part of the model, but a part of the problem representation. Approach to handling uncertain inputs is a part of DEX methodology for a long time and is briefly presented in the following section.

3.1. Uncertainty of inputs

The inputs to the basic DEX methodology models are usually crisp qualitative values, but the models can operate also with probabilistic or fuzzy distributions. For this mode of operation, there are no adaptations of the model needed. The model stays crisp and probabilistic or fuzzy definitions are given only for the values of the alternatives. When the inputs are given as probabilistic or fuzzy distributions, the method of aggregation follows the appropriate probabilistic or fuzzy calculus. Each value from the input distribution is assessed through crisp rules and the corresponding values of higher level attributes become distributions since there are more rules active in each tabular function. Outputs of these rules are in the case of probability distributions weighted by probability products of input attributes' values and summed up into the resulting probability distribution that is further used as input to value functions on higher levels of the hierarchy. In case of fuzzy distributions, the approach is very similar with minimum used instead of a product and maximum instead of summation.

3.2. Uncertainty of value functions

Representation of uncertainty in value functions is the most recent major extension of DEX methodology towards representation of uncertain knowledge. By allowing uncertain representation of the value functions, we enable the model developers to express the softness of the rules, that is: whether the goal of the rule is a single crisp value or a distribution of higher or lower variance. The idea of extension is exactly that – extending the allowed types of rules to also the rules whose goals are probabilistic or fuzzy distributions. Examples of such rules are in Fig. 2 and Fig. 3, where the goal values of target attributes are probabilistic distributions.

pest profile	pest control	**pest state**
high prob.	no treatment	(low:0.0 medium:0.2 high:0.8) C:1.0
high prob.	integrated	(low:0.2 medium:0.6 high:0.2) C:0.3
high prob.	conventional	(low:0.7 medium:0.2 high:0.1) C:0.8
regular prob.	no treatment	(low:0.2 medium:0.6 high:0.2) C:0.8
regular prob.	integrated	(low:0.3 medium:0.6 high:0.1) C:0.6
regular prob.	conventional	(low:0.8 medium:0.1 high:0.1) C:0.6
low prob.	no treatment	(low:0.5 medium:0.3 high:0.2) C:0.3
low prob.	integrated	(low:0.7 medium:0.2 high:0.1) C:0.3
low prob.	conventional	(low:0.9 medium:0.1 high:0.0) C:0.9

Figure 2. Example of a value function for *pest state* with soft rules and confidence parameters.

nutrition state	weed state	pest state	**yield red.**
deficient	very low.	low	(no:0.2 low:0.7 medium:0.1 high:0.0 very high:0.0) C:0.9
deficient	very low.	medium	(no:0.0 low:0.3 medium:0.4 high:0.3 very high:0.0) C:0.5
deficient	very low.	high	(no:0.0 low:0.0 medium:0.2 high:0.6 very high:0.2) C:0.7
deficient	low	low	(no:0.1 low:0.3 medium:0.5 high:0.1 very high:0.0) C:0.7
deficient	low	medium	(no:0.0 low:0.1 medium:0.5 high:0.4 very high:0.0) C:0.5
deficient	low	high	(no:0.0 low:0.0 medium:0.1 high:0.7 very high:0.2) C:0.8
...	...	...	...
excess	high	high	(no:0.0 low:0.0 medium:0.0 high:0.2 very high:0.8) C:0.9

Figure 3. Example of a value function for *yield reduction* with soft rules and confidence parameters.

Reasoning in such a model is straightforward, the calculation of the output values at each level follows the probability or fuzzy calculus as with inputs (see section 3.1). The final output value is also a probabilistic or fuzzy distribution. For example, if the input to the *pest state* attribute are the values *low prob.* of *pest profile* and *conventional* for *pest control*, the *pest state* would be evaluated as (*low*:0.9 *medium*:0.1 *high*:0.0) and not only *low* as it would be according to crisp value function shown in Fig. 1. Let us consider that we obtain also evaluations (*deficient*:1.0 *optimal*:0.0 *overfertilized*:0.0 *excess*:0.0) for *nutrition state* and (*very low*:0.6 *low*:0.4 *medium*:0.0 *high*:0.0) for *weed state*. The value for *yield reduction* is then calculated according to its value function and the values/distributions of its three immediate descendants in the hierarchy. The simplest view of this calculation is to consider it as a weighted summation of all the distributions in the table from Fig.3. Each distribution is weighted according to the probabilities of the values that define it. Weight is a product of these probabilities. For example, the first distribution: (*no*:0.2 *low*:0.7 *medium*:0.1 *high*:0.0 *very high*:0.0) has a weight of 0.54, since this is a product of probabilities of *nutrition state* being *deficient* (1.0), *weed state* being *very low* (0.6) and *pest state* being *low* (0.9), which are exactly the values of immediate descendants that correspond to the first distribution – the first row in the table from Fig. 3. In the same way we obtain all the other weights and get to a final result. Showing only the non-zero weights, the calculation of resulting distribution is:

(*no*:0.2 *low*:0.7 *medium*:0.1 *high*:0.0 *very high*:0.0) * 0.54 +
(*no*:0.0 *low*:0.3 *medium*:0.4 *high*:0.3 *very high*:0.0) * 0.06 +
(*no*:0.1 *low*:0.3 *medium*:0.5 *high*:0.1 *very high*:0.0) * 0.36 +
(*no*:0.0 *low*:0.1 *medium*:0.5 *high*:0.4 *very high*:0.0) * 0.04

and the resulting distribution for *yield reduction* is:

(*no*:0.144 *low*:0.508 *medium*:0.278 *high*:0.07 *very high*:0.0).

Enabling the model developer, more specifically the domain expert, to express the softness of rules is very useful, as some problems inherently imply some soft rules that would be awkward or imprecise, if expressed in another way. Dealing with distributions of values is practical also for some other purposes, like for the use of numerical inputs, which are much more naturally converted to distributions than to crisp values.

However, with positive features of such an approach, there come also some negative. More efforts must typically be put into definitions of soft value functions as well as into the analysis of results, since distributions are difficult to compare and rank. Easy and clear comparison or ranking of distributions is possible only in isolated special cases, in all others we must adopt a suitable procedure of transformation into a numerical value.

3.3. Higher order uncertainty of value functions

Introduction of distributions into the rules of value functions is useful. However, we have observed in practice that in many situations it did not ease the reluctance of domain experts with expressing uncertain knowledge. Using distributions they could express softness when necessary, but still this did not cover the phenomenon of uncertainty associated with these (either crisp or soft) assessments.

The constructor of the model, usually an expert in the problem domain, can be more certain or confident in his assessments (rules) in some parts of the model and less confident in others. Being less certain might prevent an expert to include any existing knowledge in the model, because of the fear of mistakes or being seen later as incompetent. This however hinders the possible uses of incomplete and uncertain knowledge in decision models. To enable the users to express which values are given with high certainty and which with a low one, we introduced a separated parameter named confidence. The naming follows the naming of Wang (Wang, 2001), whose approach inspired us for the use of higher level uncertainty measures in our methodology.

Such a parameter can be defined for every rule that corresponds to a combination of input values (every row in a tabular value function). Confidence parameters are shown in Fig. 2 beside distributions and are marked with a capital C. With a value from the interval [0,1] the parameter describes the confidence level of the goal value's assessment in each rule. Value of 1 represents a completely confident rule (a fact) and the value 0 a complete opposite (a guess) with the values in between representing intermediate levels accordingly. When defining the values of this parameter, the model developers can be assisted by the suggested interpretation: value of the parameter corresponds to the ratio among the number of (possibly hypothetical) observed cases of the given rule and the number of cases that are expected to affect the rule's distribution in the near future. As the parameter represents the stability of the initial assessment, but is not of the same kind (like a probability of a probability would be), it is not a second order uncertainty, but an undefined higher order uncertainty measure.

Assigning confidence levels to decision rules has more than one use in decision models. The simplest use case is when the model is used as a dissemination and discussion tool. A model with defined confidence parameters provides clear information about the confidence (and indirectly completeness) of encoded knowledge and provides clues which rules and concepts in the model need to be further studied. A more general use case is in situations when the information on levels of confidence in knowledge is used in the phase of analysis. For the purpose of analysis, the confidence parameters of aggregate attributes are calculated as products of confidence parameters of child attributes (immediate descendants in the hierarchy), the confidence parameter of the given rule and the weight of the given rule (the probability of this rule, given the alternative). Let us follow the example from section 3.2 to show a concrete calculation. Provided that the values of input attributes are certain and have the value C=1.0, the confidence of *pest state* distribution from the example would be 0.9. The confidence parameters of *nutrition state* and *weed state* would be calculated through the model, but let us assume that they are 0.2 and 0.5. The calculation of confidence for *yield reduction* is then again a weighted summation, but this time over its confidence values in rows from Fig. 3. However, the weights are a bit different now. Weight in each row corresponds to the product of probabilities of immediate descendants values (like in the inference of values) additionally multiplied by the confidence parameters of the descendants. The confidence value for *yield reduction* in our example would be therefore calculated as:

0.9 (confidence of first distribution from Fig. 3) * 0.54 (weight of this distribution) * 0.2 (confidence of *nutrition state* value) * 0.5 (confidence of *weed state* value) * 0.9 (confidence of *pest state* value) = 0.04374

Resulting aggregate attributes' parameters that are assessed this way can be used in analysis as a comparative (relative) indicator of background knowledge support of each alternative's assessment. The alternatives that get assessed through less certain parts of the model, i.e. that use rules with lower confidence parameters in their assessment, will consequently have lower confidence of the final result than the alternatives that get calculated through rules with higher confidence. This might not influence the evaluation procedure directly, but, for example it can warn the user to check the assessments of relevant alternatives more thoroughly when their confidence is comparatively much lower than the confidence of other alternatives.

4. OPEN ISSUES

There are several aspects of the presented approach that might be improved or would benefit from further work and research on human perception of uncertainties in decision problems. The most evident issue is the approach of calculation of confidence parameters during assessment of alternatives. Currently, the default is the presented weighted product, but the product causes parameters of the attributes on high levels of hierarchy to become very small. Although the relative comparison among the alternatives is what is aimed at, the small values cause discomfort of the users, as the results of the model seem weakly supported. Instead of a product, we could use any of the t-norms (Klement et al., 2004), for example the minimum, the maximum or the Lukasiewicz's t-norm to name only the most known and interesting ones. Even better, an empirical experiment or a series of experiments could be conducted to find out which norms are best suited to developers' and users' perception of the inference of confidence.

ACKNOWLEDGEMENTS

The work presented here was supported by the ECOGEN project, funded by the Fifth European Community Framework Programme under contract QLK5-CT-2002-01666, by the SIGMEA project, funded by the Sixth European Community Framework Programme under contract SSP1-2002-502981, and by the Slovenian Research Agency under grant Knowledge Technologies.

REFERENCES

Bohanec M, DEXi: Program for multi-attribute decision making, User's manual, Version 3.00. IJS Report DP-9989, Jožef Stefan Institute, Ljubljana, 2008, http://kt.ijs.si/MarkoBohanec/pub/DEXiManual30r.pdf.

Bouyssou D, Marchant T, Pirlot M, Tsoukiàs A, Vincke P, *Evaluation and decision models with multiple criteria: Stepping stones for the analyst*. International Series in Operations Research and Management Science, Volume 86. Springer, Boston, 1st edition, 2006.

Bohanec M, Rajkovič V, Multi-Attribute Decision Modeling: Industrial Applications of DEX, *Informatica* 23, 487-491, 1999.

Bohanec M, Rajkovič V, Cestnik B, Five decision support applications, In: Mladenić D, Lavrač N, Bohanec M, Moyle S, (Eds.), *Data Mining and decision support : integration and collaboration*, Kluwer Academic Publishers, 177-189, 2003.

Bohanec M, Cortet J, Griffiths B, Žnidaršič M, Debeljak M, Caul S, Thompson J, Krogh PH, A qualitative multi-attribute model for assessing the impact of cropping systems on soil quality, *Pedobiologia* 51, 207-218, 2007.

Bohanec M, Messéan A, Scatasta S, Angevin F, Griffiths B, Krogh PH, Žnidaršič M, Džeroski S, A qualitative multi-attribute model for economic and ecological assessment of genetically modified crops, *Ecological Modelling* 215, 247-261, 2008.

Co-Extra, 2006-2009. GM and non-GM supply chains: Their co-existence and traceability. FP6 EU project, FP6-FOOD-2005-7158: http://www.coextra.eu/

ECOGEN, 2003-2006. Soil ecological and economic evaluation of genetically modified crops. FP5 EU project, QLK5-CT-2002-01666: http://www.ecogen.dk

Figueira J, Greco S, Ehrgott M, *Multiple Criteria Decision Support Software*, Springer Verlag, Boston, Dordrecht, London, 2005.

Klement E, Mesiar R, Pap E, Triangular norms. position paper I: basic analytical and algebraic properties, *Fuzzy Sets and Systems* 143, 5-26, 2004.

SIGMEA, 2004-2006. Sustainable introduction of genetically modified crops into European agriculture. FP6 EU project, SSP1-2002-502981: http://sigmea.dyndns.org/

Triantaphyllou E, *Multi-Criteria Decision Making Methods: A Comparative Study*, Kluwer Academic Publishers, 2000.

Wang P, Confidence as higher-order uncertainty, In: *Proceedings of the Second International Symp. on Imprecise Probabilities and Their Applications*, 352-361, 2001.

Žnidaršič M, Bohanec M, Zupan B, proDEX – A DSS tool for environmental decision-making. *Environmental Modelling & Software* 21, 1514-1516, 2006.

Žnidaršič M, Bohanec M, Zupan B, Modelling impacts of cropping systems: Demands and solutions for DEX methodology, *European Journal of Operational Research* 189, 594-608, 2008.