Title	Knowledge Integration and Creation for Solving Complex Problems
Author(s)	Makowski, Marek
Citation	
Issue Date	2007-11
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/4103
Rights	
Description	The original publication is available at JAIST Press http://www.jaist.ac.jp/library/jaist-press/index.html, Proceedings of KSS'2007 : The Eighth International Symposium on Knowledge and Systems Sciences : November 5-7, 2007, [Ishikawa High-Tech Conference Center, Nomi, Ishikawa, JAPAN], Organized by: Japan Advanced Institute of Science and Technology

Knowledge Integration and Creation for Solving Complex Problems

Marek Makowski
International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria
marek@iiasa.ac.at

Abstract

Complex problems cannot be rationally analyzed and solved without adequate integration of pertinent knowledge and its representation in form of the corresponding mathematical model. Such models are actually used for creation of knowledge about the problem through model analysis. Modeling complex problems for finding rational solutions demands not only novel modeling methods but also appropriate modeling technology. The paper presents selected issues of knowledge integration and creation aimed at model-based support for solving complex problems, and characterizes the structured modeling technology developed for supporting collaborative interdisciplinary modeling processes.

Keywords: Structured modeling, Knowledge, Creative environments.

1 Introduction

Everybody makes decisions all the time. In most cases we manage even complex problems by successfully making decisions based on experience and intuition. Consider driving a car, for example. Each driver controls a car subconsciously applying quite complex principles of adaptive control, typically without even understanding the dynamics of the car.[1] Moreover, in congested traffic each driver constantly monitors the behavior of the other drivers and every few seconds subconsciously predicts their behavior, assessing the risk related to various combinations of the predicted behavior. Given the complexity of this everyday activity, it is amazing how well (measured e.g., by the frequency of mistakes that leads to accidents) the problem of controlling cars is solved by drivers with very diversified backgrounds and experience. If every driver can do this, then one should ask why any formal methods may help solving problems that seem to be simpler.

The simplest answer to this question comes from the experience with science-based support for solving complex problems in policy-making, industry, and management. While it is possible to accumulate enough knowledge and experience to solve diverse problems, often even without understanding all the underlying mechanisms, in many other decision-making situations mathematical models and adequate methods of model analysis are necessary for finding and/or justifying rational decisions.[2] Such situations are characterized by at least one of the following issues:

- complex relations between the decisions and the corresponding outcomes (measures of consequences of their implementations);
- difficult to assess trade-offs between attainable goals (preferred values of outcomes);
- uncertainties and risks related to the decision-making situation;
- needs for supporting the transparency of the decision-making process and enhancing public understanding of problems and the considered solutions.

Rational decision-making[3] for such problems has to be based on an appropriate use of pertinent knowledge. Knowledge is typically understood as familiarity, awareness, or understanding gained through experience or study. Knowledge creation and integration is a rather complex process, which requires careful management. This field has recently been intensively researched, and the School of Knowledge Science at JAIST[4] is a leading center of this research. The research has a wide scope, therefore even an outline of its results is beyond the scope of this paper. Recent developments in this area including extensive bibliography can be found in [3; 4; 5].

This paper focuses on the two elements of such a process that are mostly related to knowledge management, namely knowledge integra-

[1] Control engineers could solve differential equations to optimize the way they drive a car, but they do not need to do so.

[2] A comprehensive discussion of model-based decision-making support is available in [1].

[3] Modern approaches to rational decision-making are discussed in more detail in [2].

[4] Japan Advanced Institute of Science and Technology.

tion and representation, and knowledge creation.

For many complex problems a large part of the pertinent knowledge can be represented by mathematical models. Knowledge grows very quickly, therefore even the best scholars can only master a tiny fraction of the knowledge available in their professional area. Consider, e.g. mathematical programming, which is on the one hand a rather specialized area of mathematics, but on the other hand is quite a broad area from the point of view of researchers working in a particular field (e.g. interior point methods for optimization, or wavelet-based approaches to analysis of time series). Therefore model development for solving complex problems requires collaboration of scientists and professionals who contribute (typically interdisciplinary and heterogeneous) knowledge. Such a collaboration is organized through virtual organizations, which for collaborative interdisciplinary modeling can be called virtual modeling laboratories. This important topic is beyond the scope of this paper, it is however discussed in [6].

The remaining part of this paper is organized as follows.

Methods for selection of knowledge pertinent to the problem at hand and its representation is a kind of meta-knowledge that actually is tacit, i.e., experienced modelers know how to develop models to properly: (1) integrate and represent, and (2) create knowledge relevant to the modeled problem. This meta-knowledge is however fragmented and not documented, therefore can hardly be disseminated. Selected issues of knowledge integration are discussed in Section 2.

In the final step of model-based support for problem solving knowledge is created by model analysis, and used for supporting rational decision-making. More detailed discussion of the relations between knowledge management and decision support can be found in [7]. In this paper we concentrate in Section 3 on selected issues of knowledge creation through a model analysis.

Rational problem solving requires concerted handling of all pertinent elements of the decision-making support process, which is accompanied by the corresponding stages of the related modeling process, which in turn requires methods and tools appropriate for effective handling of complex models. These issues are discussed in Section 4.

2 Knowledge Integration

Probably the best way to integrate knowledge for problem-solving whenever it involves analysis of large amounts of data and/or non-trivial relations is to develop an appropriate mathematical model, or several models to provide diverse insights into the problem. In order to outline the knowledge integration let us consider a mathematical model as being composed of entities, and relations between them. Entities are of two types: (1) parameters, values of which represent pertinent information (i.e. a collection of data), and (2) variables, values of which are assigned during the model analysis. The model relations (conventionally called constraints or functions) represent knowledge about the relationships among the model entities.

A model integrates knowledge pertinent to solving a particular problem on two levels:

- symbolic model specification;
- the model instances (also called *substantive models* or *core models*) composed of the model specification and a selected set of data used for instantiation of relations (through assigning values to parameters of the relations).

In many situations, symbolic model specification can be based on commonly known rules of science. However, in other situations, knowledge pertinent to a particular relation is so diversified that a definition of the relation requires a dedicated study. To illustrate this problem let us recall that the relation between trophosperic ozone and its two precursors (nitrogen oxides and volatile organic compound) can be defined in very different ways, each having the corresponding diversified advantages and disadvantages depending on the content in which the relation is applied, see e.g., [8].

For complex models handling a representation of even very simple relations requires a rather sophisticated modeling technology. To illustrate this point let us consider a simple definition of an auxiliary variable:

$$y_{ipa} = \sum_{t \in T_{pa}} x_{iat}, \quad i \in I, p \in P, a \in A_{ip} \qquad (1)$$

where indices i, p, a, t denote a country, pollution type, economic activity, and technology, respectively. The complexity of this relation is caused by the fact that indices are members of sets which are indexed by the other indices. Therefore, even if the size of each set is not large, the structure of the corresponding indexed subsets is pretty complex and requires effective management. Such a management includes:

- efficient handling of the underlying data structures that in turn requires advanced use of DBMS,[5] which is inevitable for effective modeling process of large scale models;
- analysis of semantic correctness of indexing structures.

Note, that for a typical model[6] instance there are about 200 subsets A_{pa}, and the eq. (1) is represented in the corresponding optimization problem by about 40,000 constraints with about 200,000 corresponding non-zero elements of the Jacobian. The actual model is defined by dozens of relations, most of them more complex than eq. (1). This illustrates the need for qualitatively more efficient modeling tools than those traditionally used.

For large scale models relations for each subject (represented by a submodel) are defined in a close cooperation between specialists in the corresponding area and a team of modelers capable of:

- assessing the consequences of the considered relation types on the numerical complexity of the resulting computational tasks;
- ensuring consistency of the whole model to which the relation will be included.

Thus the development of symbolic model specification requires:

- analysis of the relevant (for the purpose of the model) knowledge about each modeled subject (submodel), and a selection of these elements of the knowledge which will be represented in the model;
- representation of the selected knowledge in a mathematical form consistent with relations defined for all other submodels;
- integration of all submodels into a consistent model that possibly best (in terms of both required accuracy and computational efficiency) represents the relations between the decisions and outcomes.

We should stress an important feature of a properly developed model: it integrates knowledge in a reliable way and thus provides an objective and justifiable way of analyzing the relations between the decisions and the consequences of their implementation. This objectivity can only be assured if:

- all model relations are actually based on knowledge, i.e. on verifiable facts and rules;
- the assumptions for these facts and rules are consistent with the assumptions agreed for the model;
- semantic correctness is enforced not only for each relation but also for the set of all relations (e.g., the units and the accuracy/precision of all entities are consistent);
- no representation of the preferential structure is included in the substantive model;
- data used for model instantiation is consistent with the model specification.

We briefly comment on one of the above summarized issues that is crucial for model correctness but is commonly neglected. Complex algebraic models are composed of diverse entities (variables, parameters, relations) that capture the physical dimensions of the corresponding entity (such as *mass*, or *volume*, or *a transfer coefficient*). Often such dimensions[7] have rather complex structure (e.g., $[Eq(S)/kWh] \times [GWh/year]$, where S stands for a pollution type). A model developer should ensure the semantic correctness of the model, which includes consistency across physical dimensions and their units of measure in the model relations. However, only one of the commonly used modeling environments[8] supports definition of units. Therefore there is practically no effective support for checking semantic correctness of models, which is clearly error-prone for development of non-trivial models. Effective support for dimensional consistency analysis is rather complex, see e.g., [10], and providing an effective support requires further research and the corresponding implementation effort.

Although a proper symbolic model specification is certainly the most challenging part of model building from the knowledge integration point of view, we have to stress that the data used for model instantiation also represents a necessary part of knowledge which needs to be integrated into the modeling process in a robust and efficient way. Data maintenance for a large complex model is by far the most risky element of

[5] Database Management Systems.

[6] RAINS model, described e.g., in [8].

[7] Also referred to as *units*.

[8] AIMMS, see e.g., [9].

any modeling process. Data for large models comes from different sources (as do the results from analysis of various models), and larger subsets of data are maintained by teams. To reliably maintain (collect, clean, verify, update) data for complex model a structured approach based on DBMSs is a must. A more detailed discussion of this topic can be found in [6].

3 Knowledge Creation

Knowledge can be created by diverse methods of model analysis. However, the methodology of model analysis is probably the least researched element of the modeling process. This is because each modeling paradigm has a specific type of analysis. However, the essence of model-based decision-making support is precisely the opposite; namely, to support diversified ways of model analysis, and to provide efficient tools for various comparisons of solutions. Such an approach can be called Integrated Model Analysis.

Actually, the primary goal of modeling for problem solving is to create knowledge about the modeled problem. In fact, model-based learning about the problem is typically even more important than finding *the best* solution, see e.g., [11]. Thus, a huge amount of knowledge has been created by various types of analyses of a countless number of models. Unfortunately, this knowledge is often difficult to use beyond the modeling process. The main reason for it the lack of semantic descriptions of model results. These are typically consumed for the analysis of the decision problem at hand, and are not documented sufficiently for re-use in different contents.

A typical model for supporting decision-making has an infinite number of solutions, and users are interested in analyzing trade-offs between a manageable number of solutions that correspond to various representations of their preferences, often called the preferential structure of the user. Thus, an appropriate integrated analysis should help users to find and analyze a small subset of all solutions that correspond best to their preferential structures that typically change during the model analysis. There are three types of problems that call for innovative research:

1. integration of various paradigms of model analysis;
2. extracting knowledge from large sets of solutions;
3. efficient solution of computational tasks (either resource demanding, or numerically difficult, or large sets of simple jobs).

We briefly summarize each of them below.

For a truly integrated problem analysis one should actually combine different methods of model analysis, such as: classical (deterministic) optimization (and its generalizations, including parametric optimization, sensitivity analysis, fuzzy techniques), multicriteria model analysis, stochastic optimization and Monte Carlo simulations, classical simulation, soft simulation, and several of its generalizations (e.g. inverse simulation, softly constrained simulation). However, no modeling tool supports such a complete analysis, and development of separate versions of a model with tools supporting different modeling paradigms is typically too expensive.

The second challenge is to develop and implement a methodology for a comprehensive analysis of large sets of solutions. One needs to explore applicability of various data mining and knowledge engineering techniques, and either adapt some of them, or develop new methods to extract and organize knowledge from large sets of solutions, and supply users with this knowledge in a form that will help further problem analysis.

The third set of issues is related to efficient and robust organization of computational tasks typically needed for large-scale models, and includes:

- efficient support for handling of a large number of results, possibly coming from various types of analyses of large models;
- adaptation of specialized optimization algorithms for badly conditioned problems;
- support for exploiting the structure of huge optimization problems that need to be solved on computational grids.[9]

There is a lot of experience related to these three types of problems. However, the relevant knowledge is fragmented and distributed amongst modelers that concentrate rather on the modeling processes than on documenting and sharing the model-analysis knowledge. The latter is not only due to the lack of time but also due to the complexity of diverse paradigms of model analysis.

[9] This item is contingent on additional resources, and on availability of external collaboration with partners having suitable experience.

To illustrate the complexity of model analysis we mention here the type of analysis which appears to be simple for not experienced modelers, namely multicriteria analysis of large sets of discrete alternatives characterized by a large number of criteria. There is a large number of books and articles dealing with analysis of discrete alternatives however actually none of the documented methods is methodologically correct for analysis of the class of problems described in [12]. In other words, even experienced modelers face new challenges also in the area of model analysis, including types of problems that are considered standard. Moreover, some problems require rather specific methods of model analysis, see e.g., [8; 13]. This observation is another justification of the statement presented in [14]: modeling has been, and will remain to be a combination of science, art, and craft.

4 Modeling for Knowledge Integration and Creation

The complexity of an appropriate knowledge integration, and of creation of knowledge that possibly best supports problem understanding and solving, requires an adequate support for the corresponding modeling process, which in turn determines the requirements for modeling technology that are substantially different from the technologies successfully applied for modeling well-structured and relatively simple problems. In most publications that deal with modeling, small problems are used as an illustration of the presented modeling methods and tools. Often, they can also be applied to large problems. However, as discussed above, the complexity is characterized not primarily by the size, but rather by: the requirements of integrating heterogeneous knowledge, the structure of the problem, and the requirements for the corresponding modeling process. Moreover, efficient solving of complex problems requires the use of a variety of models and modeling tools; this in turn will require even more reliable, re-usable, and shareable modeling resources (models, data, modeling tools). The complexity, size, model development process, and the requirements for integrated model analysis form the main arguments in justifying the needs for the new modeling methodology.

Structured Modeling Technology (SMT) has been developed for meeting such requirements. SMT supports distributed modeling activities for models with a complex structure using large amounts of diversified data, possibly from different sources. A description of SMT is beyond the scope of this paper, therefore we only summarize its main features here:

- SMT is Web-based, thus it supports *any-where, any-time* collaborative modeling.
- It follows the principles of Structured Modeling proposed by Geoffrion, see e.g., [15]; thus it has a modular structure which supports the development of various elements of the modeling process (model specification, processing (subsets of) data, integrated model analysis) by different teams.
- It provides automatic documentation of all modeling activities.
- It uses a DBMS for all persistent elements of the modeling process, which results in efficiency and robustness; moreover, the capabilities of DBMSs allow for the efficient handling of huge amounts of data.
- It ensures the consistency of: model specification, meta-data, data, model instances, computational tasks, and the results of model analysis.
- It automatically generates a Data Warehouse with an efficient (also for large amounts of data) structure for:
  - ⋆ data, and the tree-structure of data updates,
  - ⋆ definitions of instances,
  - ⋆ definitions of preferences for diversified methods of model analysis,
  - ⋆ results of model results,
  - ⋆ logs of all operations during the modeling process.

  This conforms to the requirement for the persistency of all elements of the modeling process.
- It exploits computational grids for large amounts of calculations.
- It also provides users with easy and context-sensitive problem reporting.

More detailed arguments (including an overview of the standard modeling methods and tools) supporting this statement are available in [16].

Mathematical modeling of a complex problem is actually a network of activities involving interdisciplinary teams collaborating closely with experts in modeling methods and tools. Dantzig summarized in [17] the opportunities and limitations of using large-scale models for policy making. Thanks to the development of algorithms and computing power today's large-scale models are at least 1000-times larger; thus, large-scale models of the 1970s are classified as rather small today. This, however, makes Dantzig's message relevant to practically all models used today, not only for policy-making but also in science and management. Today's models are not only large, but the modeled problems are more and more complex (e.g., by including the representation of knowledge coming from various fields of science and technology), and many models are developed by interdisciplinary teams. Moreover, the modeling processes supporting policy-making have to meet strict requirements of: credibility, transparency, replicability of results, integrated model analysis, controllability (modification of model specification and data, and various views on, and interactive analysis of, results), quality assurance, documentation, controllable sharing of modeling resources through the Internet, and efficient use of resources on computational Grids.

These requirements demand a qualitative jump in modeling methodology: from supporting individual modeling paradigms to supporting a *Laboratory World*[10] in which various models are developed and used to learn about the modeled problem in a comprehensive way. The truth is that there are no simple solutions for complex problems, thus learning about complex problems by modeling is in fact more important than finding an *"optimal"* solution. Such a Laboratory World requires integration of various established methods with new (either to be developed to properly address new challenges, or not yet supported by any standard modeling environment) approaches needed for appropriate (in respect to decision-making process, and available data) mathematical representation of the problem and ways of its diversified analyses. Therefore, to be able to adequately meet the demand for advanced modeling support one indeed needs to develop and apply novel modeling methodologies.

5 Conclusions

The traditional approach to modeling is based on the assumption that a small team can organize and document a modeling process. However, this approach is neither reliable nor efficient for complex models developed by several (or more) teams working intensively[11] at distant locations. Despite the countless number of successful applications, there is also well-justified criticism of various critical aspects of modeling, e.g., in [20; 21; 22; 8; 23]. The role of models in modern decision-making, a view shared by the author of this paper, is discussed in detail in [1], which also presents the methodology and tools for model-based decision-making support, and illustrates them with several applications to complex environmental policy-making problems. A more focused discussion of selected elements of modeling for decision support is provided in [11], which also includes an updated bibliography on modeling for decision support.

A number of methods have been developed for dealing with each of the above listed issues. The craft of decision-making support consists of adopting an appropriate approach to support each element of the decision-making process while remembering that the strength of a chain is determined by its weakest link. Given the heterogeneity and amount of the pertinent modeling knowledge, effective collaboration between scientists working on various fields (including knowledge creation and integration, knowledge engineering, operations research) is a necessary condition for effective and truly knowledge-based problem-solving support.

Acknowledgment

Many colleagues and friends have helped in various ways to understand different elements of key modeling and knowledge paradigms, and to develop ideas, concepts, and tools. There is no way to adequately characterize individual contributions. Therefore the author thanks here only colleagues whose impact on the research outlined in this paper has been most significant: Y. Nakamori and A. Wierzbicki, as well as many IIASA colleagues with whom the author has been collaborating in developing and analyzing various models since over 20 years. The challenging modeling problems have been the best motivation for developing novel methods and tools, and the recent achievements in knowledge science have provided complementary (to the traditional mathematical modeling

[10] Originally proposed by Dantzig, see e.g. [17; 18].

[11] This implies that diversified elements of the model are developed/modified at practically random times.

paradigms) understanding of the relations between integration and creation of knowledge, and model-based problem-solving support.

References

[1] A. Wierzbicki, M. Makowski, and J. Wessels, editors. *Model-Based Decision Support Methodology with Environmental Applications*. Series: Mathematical Modeling and Applications. Kluwer Academic Publishers, Dordrecht, 2000. ISBN 0-7923-6327-2.

[2] A. Wierzbicki and J. Wessels. The modern decision maker. In Wierzbicki et al. [1], pages 29–46. ISBN 0-7923-6327-2.

[3] Y. Nakamori. Systems methodology and mathematical models for knowledge management. *Journal of Systems Science and Systems Engineering*, 12(1):49–72, 2003.

[4] A. Wierzbicki and Y. Nakamori, editors. *Creative Space: Models of Creative Processes for Knowledge Civilization Age*, volume 10 of *Series: Studies in Computational Intelligence*. Springer, Berlin, Heidelberg, New York, 2006. ISBN 3-540-28458-3.

[5] A. Wierzbicki and Y. Nakamori, editors. *Creative Environments: Issues of Creativity Support for Knowledge Civilization Age*, volume 59 of *Series: Studies in Computational Intelligence*. Springer, Berlin, Heidelberg, New York, 2007. ISBN 978-3-540-71466-8.

[6] M. Makowski and A. Wierzbicki. Virtual laboratories. In A. Wierzbicki and Y. Nakamori, editors, *Creative Environments: Issues of Creativity Support for Knowledge Civilization Age*, volume 59 of *Series: Studies in Computational Intelligence*, pages 233–254. Springer, Berlin, Heidelberg, New York, 2007. ISBN 978-3-540-71466-8.

[7] M. Makowski, Y. Nakamori, and A. Wierzbicki. Decision support versus knowledge creation support. In A. Wierzbicki and Y. Nakamori, editors, *Creative Space: Models of Creative Processes for Knowledge Civilization Age*, volume 10 of *Series: Studies in Computational Intelligence*, pages 219–250. Springer, Berlin, Heidelberg, New York, 2006. ISBN 3-540-28458-3.

[8] M. Makowski. Modeling techniques for complex environmental problems. In M. Makowski and H. Nakayama, editors, *Natural Environment Management and Applied Systems Analysis*, pages 41–77. International Institute for Applied Systems Analysis, Laxenburg, Austria, 2001. ISBN 3-7045-0140-9, available from `http://www.iiasa.ac.at/~marek/pubs/prepub.html`.

[9] J. Bisschop and M. Roelofs. AIMMS, *The User's Guide*. Paragon Decision Technology, Haarlem, The Netherlands, 2004.

[10] V. Nastase, M. Makowski, and W. Michalowski. Dimensional consistency analysis in complex algebraic models. Interim Report IR-07-29, International Institute for Applied Systems Analysis, Laxenburg, Austria, 2007. (to appear).

[11] M. Makowski and A. Wierzbicki. Modeling knowledge: Model-based decision support and soft computations. In X. Yu and J. Kacprzyk, editors, *Applied Decision Support with Soft Computing*, volume 124 of *Series: Studies in Fuzziness and Soft Computing*, pages 3–60. Springer-Verlag, Berlin, New York, 2003. ISBN 3-540-02491-3, draft version available from `http://www.iiasa.ac.at/~marek/pubs/prepub.html`.

[12] A. Wierzbicki, J. Granat, and M. Makowski. Discrete decision problems with large number of criteria. Interim Report IR-07-25, International Institute for Applied Systems Analysis, Laxenburg, Austria, 2007. (to appear).

[13] M. Makowski. Model-based decision making support for problems with conflicting goals. In *Proceedings of the 2nd International Symposium on System and Human Science, March 9-11, 2005, San Francisco, USA*. Lawrence Livermore National Laboratory, Livermore, USA, 2005. CD edition of the Proceedings available from LLNL.

[14] J. Paczyński, M. Makowski, and A. Wierzbicki. Modeling tools. In Wierzbicki et al. [1], pages 125–165. ISBN 0-7923-6327-2.

[15] A. Geoffrion. An introduction to structured modeling. *Management Science*, 33(5):547–588, 1987.

[16] M. Makowski. A structured modeling technology. *European J. Oper. Res.*, 166(3):615–648, 2005. draft version available from `http://www.iiasa.ac.at/~marek/pubs/prepub.html`.

[17] G. Dantzig. Concerns about large-scale models. In R. Thrall, R. Thompson, and M. Holloway, editors, *Large-Scale Energy Models. Prospects and Potential*, volume 73 of *AAAS Selected Symposium*, pages 15–20. West View Press, Boulder, Colorado, 1983.

[18] C. Holling, G. Dantzig, W. Clark, D. Jones, G. Baskerville, and R. Peterman. Quantitative evaluation of pest management options: The spruce budworm case study. Technical report, US Department of Agriculture: Washington Forest Service, Washington, USA, 1979.

[19] M. Makowski and H. Nakayama, editors. *Natural Environment Management and Applied Systems Analysis*. International Institute for Applied Systems Analysis, Laxenburg, Austria, 2001. ISBN 3-7045-0140-9.

[20] R. Ackoff. Management misinformation systems. *Management Science*, 14(4):43–89, 1967.

[21] R. Ackoff. The future of operational research is past. *Journal of OR Society*, 30(2):93–104, 1979.

[22] C. Chapman. Science, engineering and economics: OR at the interface. *Journal of Operational Research Society*, 39(1):1–6, 1988.

[23] J. Sterman. All models are wrong: reflections on becoming a systems scientist. *Systems Dynamics Review*, 16(4):501–531, 2002.