PRACE NAUKOWE AKADEMII EKONOMICZNEJ WE WROCŁAWIU
Nr 1121 — 2006
Pozyskiwanie wiedzy i zarządzanie wiedzą

Mieczysław L. Owoc

Wrocław University of Economics

VALIDATION QUESTS OF HYBRID KNOWLEDGEBASES

1. Introduction

The main stream of knowledge validation techniques and tools is oriented at procedural representation of domain knowledge. There are at least two reasons of such setting. First, procedural knowledge seems to be very popular and efficient in intelligent systems supporting management. Second, rule-based systems can be verified relatively easy – there are many of developped tools that can be applied for validation of knowledge incorporated to these systems.

From the very beginning apart of procedural knowledgebases mentioned above we need other formalisms to represent knowledge about objects, relationships etc. (say declarative knowledge) but also we meet hybrid representations that mean mixture of knowledge artifacts sometimes stored as hidden forms. While many problems have been solved in the last decade, there remain many unresolved problems, and the evolving nature of KBS technology generates new V&V challenges. Many of the issues involved in V&V of rule-based systems are now well understood, but few modern KBS are purely rule-based [Knauf 2000; Ligęza 2005]. Current challenges lie in V&V techniques for complex, hybrid knowledge-representation and reasoning systems. Increasing adoption of distributed KBS technology (for example, multiple-agent systems, distributed enterprise knowledge bases, and web-based KBS) have raised new–and hard–V&V requirements.

The main goal of this paper is to specify critical validation problems that appear in knowledgebases composed from heterogenous components. Therefore the sorts of such knowledgebases must be discussed. In particular according to assumed criteria (forms of knowledge representation, artificial intelligence approaches applied, basic transformations existing in certain intelligent systems) the diversification of relevant knowledge is presented. Then specification of validation challenges is itemised including multidimensional aspects of applied techniques and properties. All of them are regarded in a concept of universal knowledge validator as an ultimate solution. The paper is concluded with remarks about usability of such validator in verification of hybrid knowledgebases.

2. Diversification of Knowledge in Hybrid Intelligent Systems

Validation as the key process assuring knowledgebase quality is mostly oriented on rule-based systems. There are huge intelligent systems containing knowledgebases represented in different ways. Very often knowledge engineers use hybrid way of knowledge representation and procedures suggested in such environments are expected.

Let us start from enumbering of knowledge forms that are potential existing in the discussed systems. At lest the following criteria can be formulated to present differentation of knowledge (compare [Ochmańska, Owoc 2001; Owoc 2001]):

– origin knowledge types – referring to procedural, declarative, heuristic etc, knowledge recognition,

– knowledge mapping formulaes – denoting whether knowledge is mapped in a symbolic or non-symbolic way

– techniques used for knowledge representation – covering particular formalisms, that are used to expressing knowledge, for example: propositional logic, object-attribute-value, semantic networks, rules frames and other notions,

– methods of knowledge base creation, regarding as ways of knowledge acquisition: developed or generated,

– areas of domain knowledge, focusing at sort of human activities; business, production, marketing, medicine and the like,

– the ultimate goal of application task, regarding as a type of generated solutions (classification, diagnosis, monitoring and interpretation etc.) and

– patency of topic knowledge, taking a shape of tacit and explicite knowledge.

In general, the criteria of presented knowledge differentiation should be regarded due to analyisis of composing knowledge bases in hybrid intelligent systems (HIS). Therefore any of the mentioned aspects of such defined systems seems to be important to define infrastructure of HIS. On surface modern HIS may consist of knowledge base components representing any possible mixture of primary recognised knowledge. Generallly speaking, among knowledge properties apply to knowledge in more general understanding **hetoregeneity** seems to be typical for more advanced applications.

However, in addition there are some important special properties. Thus, technical knowledge has specifically strong links to physical artefacts (products, materials), to sciences, especially natural sciences and their methodologies, to purposeful functionalities, and to common systems for its operationalization and codification. In addition novel, useful and non-obvious managerial knowledge is patentable more or less worldwide. Apart from general qualitative conceptualizations and classifications of economic categories (labour forces, production, market, environmental infrastructure etc.) and features (such as effectiveness, productivity etc.), a number of more formalized representations of technology appear in management. This type of moregradual techno-economic improvements basically result from interaction between bodies of knowledge embodied in agents and products during a diffusion process, thus intertwined with an innovation process.

With a mapping (or rather a many-to-many correspondence) between technologies and realizable technical performance over time (absolute time or "production clock time") state-of-the-art and learning proxied by levels and changes in technical performance, enable a representation of knowledge in the technology space, in turn enabling introduction of concepts such as directions and trajectories of learning in the technology space.

Thus, in representing business knowledge and acommpanying its learning, the distinction and mapping between a technology space and a technical performance space for observable, measurable artefacts is important (although not necessary since growth of e.g. patents, publications and people might be used as indicators of knowledge and learning as well.) However, in representing managerial knowledge and learning, including organizational routines, the link to observable artefacts is far less clear, the knowledge is less codifiable and not patentable (apart from some software and business method embodiments). How to do this formally must be left an open issue here. There are attempts and possible approaches e.g. in the area of evolutionary algorithms or more specifically genetic algorithms. Nevertheless, it is important to recognize the impact of managerial learning on other types of evolution.

There are numerous models (e.g. simple ones as logistic, exponential or combinations thereof) of growth (of populations, capital stock, markets, sales, etc.) and diffusion (of information, innovations, diseases etc.) which have been applied to knowledge, then treated as a homogenous entity, typically linked to (embedded in) countable products or individuals.

Different pieces of bodies of knowledge are moreover interacting with each other. Growth in one knowledge area impacts growth in another and a diversity of interacting areas evolves. How could such growth and diversification of knowledge be modelled in an evolutionary (dynamic) way? This is largely raised as an open issue here with a brief exposé of some available formalized and therefore simplified approaches.

3. Crucial Problems of Validation of Heterogenuos Knowledge

Knowledge is intrinsically heterogeneous in the first place as mentioned. But knowledge heterogeneity represented by a conceptualized decomposition of knowledge into areas (bodies, fields, disciplines, classes etc.) is partly reflecting an exogenous reality (assuming it exists) being searched and researched, partly reflecting a cognitive or social construct where conceptualizations and reconceptualizations may change the heterogeneity (diversity) without necessarily changing the underlying "content" (compare: [Cantwell, Piscitello 1996]).

Before to itemise essential problems addressed to such understood heterogenous knowledge bases let us to setting more general infrastructure of the the validation process. The graphical shape of such environments is put in the fig. 1.

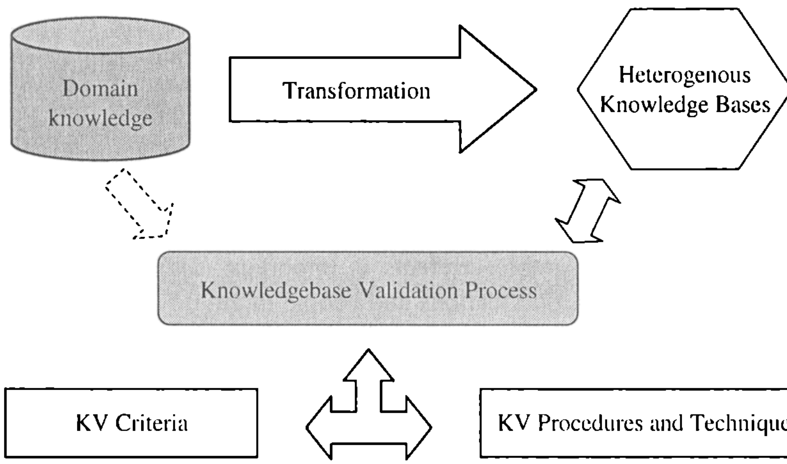

Fig. 1. Knowledge validation process environments

Source: own elaboration.

While the task of verifying rule-based systems for logical properties is now well-understood, comparatively little work has been done validation methods for other kinds of KBS. Areas of concern which have been identified include: distributed KBS (for example, co-operative distributed problem-solvers), hybrid KBS (built using a combination of different knowledge representation formalisms and the like), non-monotonic KBS, and model-based KBS. It would seem that, while some validation issues are common across different types of system, some issues are quite different. A good example is the notion of consistency-of-knowledge: this is usually desirable in a monotonic rule-based system, but would we require consistency between the knowledge of different agents in a distributed KBS? This is arguable.

The general framework for the validation of heterogenuos knowledge bases and the same hybrid intelligent systems covers categories essential for the investigated phenomena: domain knowledge and knowledge bases as an input for the validation process and core components of this process: procedures, techniques and criteria. The basic problems arise in such setting of the scene can be categorised as follows:

– **granularity of knowledge** taking into account in domain knowledge as well as in a case of knowledgebases. The main challenge is how deeply knowledge should be mapped reflecting original concepts in the domain area,

– **a level of heterogeneity of knowledge** put into a particular hybrid intelligent system. The scale of the phenomena can be very different, from two relatively similar formalisms of knowledge representation (for example procedural knowledge and frames) up to composition of all mentioned before differences including: knowledge types, mapping formulaes, methods creation etc. that are present in the investigated solution. This factor seems to play decisive role in general evaluation of complexity hybrid intelligent systems,

– **setting of crucial relationships** among hybrid intelligent system components. Passing results between knowledge units and more general efficiency of the communication systems determines impact at a validation process as a whole. As a result establishing a common "platform" of understanding co-operated knowledgebases make easier validations of particular components.

The presented above version of knowledge validation problems seen in several dimensions in a case of heterogenous knowledgebases will involve in future but actually may be treated as initial stage of setting a scene for validation of the described systems.

4. Potential Scenarios of the Hybrid Knowledgebases Validation Process

Implementation of any particular hybrid knowledgebas needs to verify and evaluate the entire system, nevertheless of sorts and goals of the application. Therefore, validation and verification is an important stage to develop an intelligent system especially in the case of hybrid solution. One can observe a range of applications of various artificial intelligent techniques, such as artificial neuron networks, knowledge-based systems and expert systems that are applied in many areas. They have valuable experience in AI application and have sophisticated knowledge to cope with possible relevant problems, such as choosing the attributes, allocating matching weight, retrieve the learning depth and efficiency, training system, building conditional rules, and system validation/verification. Apparently the hybrid intelligent systems seem to be more attractive and effective than a traditional non-interrelated one.

Taking into account the numbered quests of knowledge validation oriented at hybrid intelligent systems on can define the future scenarios of knowledge validation in such systems. At least four options are available.

1. **Local validation of different components** relies on checking important for an application knowledgebase properties, for example: consistency and completeness for all determined knowledge units. This approach can be regarded as **analyti-**

cal, because we analyse very carefully quality of the separated parts of different knowledgebases.

2. **Transforming different knowledge representations** onto unified notion and then validation is performed in a traditional way. Usually these knowledgebases are expressed as rulebases because there are plenty of algorithms and tools supporting knowledge validation process. This approach can be recognised as **unified,** as a result of usage of unified platform for knowledge representation.

3. **Global validation of chosen properties** of a whole intelligent system. In this case validation may be applied for critical features of such a system. This approach can be termed as **synthetic** – one can get general opinion about the pointed out property. The most natural is evaluation of system reliability, however we should have conscious about simplified picture of the entire system.

4. **Selective validation of partially transformed components** allows for local aswell for global validation to verify and/or evaluate the chosen feature of the system. This approach can be termed as **mixed** because is composed from the previous options.

These options allow to validate knowledge base properties in different ways and are suggested toapply for more or less universal knowledge bases.

5. Conclusions

There are more and more intelligent systems that use different components and are examples of hybrid knowledge bases. These systems should be validated to assure expacted quality of knowledge put to the entire system. The general findings of the paper can be formulated as follows:

1) there are several urgent quests that are necessary to solve in validation of hybrid knowledge systems.

2) as a result of analysis of activities and properties of knowledge bases treated as hybrid solution some options of performing knowledge validation are defined.

The results of the paper are converged to published in the previous papers, (compare: [Owoc 2001] or [Cornelissen 1997]) referring to compositional verification of knowledge-based systems. This way a concept of an universal evaluator of hybrid knowledgebases seems to be more real.

References

Bouali F., Loiseau S., Rousset M.C., *KBS Correction: a Proposal Based On Diagnosis Theory*, *Proceedings of the ECAI-94 Workshop onValidation of Knowledge-based Systems*, 1994, pp. 21-29.

Cantwell J., Piscitello L., *The Diversification of Internationalization of Corporate Technology: From a Strategic Choice Historically to a Complementary Combination Today*, mimeo, University of Reading, 1996.

Cornellissen F., Jonker C.M., Treur J., *Compositional Verification of Knowledge-based Systems: a Case Study for Diagnostic Reasoning*, [in:] *Proc. Fourth European Symposium on the Validation and Verification of Knowledge Based Systems*, eds. J. Vanthienen, F. van Harmelen, Katholieke Universitait Leuven, 1997.

Knauf R., *Validating Rule-based Systems. A Complete Methodology. Habilitation Thesis*, Shaker Verlag, Aachen 2000.

Ligęza A., *Logical Foundations for Rule-based Systems*, Springer 2005.

Menzies T., Gambetta W., *Exhaustive Abduction: a Practical Model Validation Tool, Proc. of the ECAI-94 Workshop on Validation of Knowledge-based Systems*, 1994, pp. 70-76.

Preece A., Shinghal R., Batarekh A., *Principles and Practice in Verifying Rule-based Systems*, "Knowledge Engineering Review" 1992, no 7, 115-141.

Ochmańska M., Owoc M.L., *Verification of Different Knowledge Bases*, [in:] eds. A.J. Baborski, R.F. Bonner, M.L. Owoc, Malardalens University Press 2001.

Owoc M.L., *Podejścia do weryfikacji heterogenicznych baz wiedzy*, [w:] *Pozyskiwanie wiedzy z baz danych*, red. A. Baborski, Prace Naukowe Akademii Ekonomicznej nr 891, AE, Wrocław 2001 [*Approaches to Heterogeneous Knowledgebases Verification*, [in:] *Knowledge Acquisition in Databases*, ed. A. Baborski A. (eds.) Wroclaw University Publishing Co.] (in Polish).

Winter S.G., *Knowledge and Competence as Strategic Assets*, [in:] ed. D.J. Teece, *The* Competitive *Challenge: Strategies for Industrial Innovation and Renewal*, Ballinger, Cambridge, MA 1987, pp. 159-184.

PROBLEMY WALIDACJI HYBRYDOWYCH BAZ WIEDZY

Streszczenie

Współczesne systemy inteligentne funkcjonują często w postaci niejednorodnych (pod względem stosowanych metod reprezentacji czy technik przetwarzania) baz wiedzy. Z drugiej strony mamy do dyspozycji bardzo zaawansowane (w przypadku np. systemów regułowych) techniki ich weryfikacji i oceny. Wartościowanie wiedzy zawartej w systemach inteligentnych, które charakteryzuje różnorodność formalizmów odwzorowania jest problemem złożonym i wymaga opracowania procedur skutecznych w tym kontekście, także z uwzględnieniem różnych formalizmów opisów wiedzy. W artykule podjęto próbę specyfikacji problemów pojawiających się podczas wartościowania wiedzy w odniesieniu do omawianej klasy systemów. Świadomość tych problemów umożliwiła przygotowanie kilku potencjalnych koncepcji wartościowania, znajdujących zastosowanie w różnych warunkach.