Applying the Methods of Data Mining and Knowledge Engineering to the SiMA Project Data Analysis

Dietmar Dietrich
Institute of Computer Technology
University of Technology
Vienna, Austria
e-mail: dietrich@ict.tuwien.ac.at

Matthias Jakubec
Institute of Computer Technology
University of Technology
Vienna, Austria
e-mail: jakubec@ict.tuwien.ac.at

Samer Schaat
Institute of Computer Technology
University of Technology
Vienna, Austria
e-mail: schaat@ict.tuwien.ac.at

A.I. Malakhova
Automated Management Systems Department
Ufa State Aviation Technical University
Ufa, Russia
e-mail: aimalakhova@gmail.com

Abstract[1]

This article is devoted to applying the Data Mining methods to the data analysis, received during the Simulation of the Mental Apparatus and Applications (SiMA) project simulation process. According to the principals of knowledge engineering the ontological knowledge base, containing the results of this analysis, was developed. Special attention is paid to the construction of SWRL rules on the base of developed ontology.

1. Simulation of the Mental Apparatus and Applications

The Simulation of the Mental Apparatus and Applications (SiMA) is a bionically inspired cognitive architecture in the field of Artificial General Intelligence, based on the metapsychologic theory of psychoanalysis 12. It has been developed by the team of Dietmar DIETRICH since the year 2000 at Vienna University of Technology together with research partners from all over the world 3. Until January 2015 the name of the project was ARS (Artificial Recognition System). It understands the human brain as the control unit of the human body as described by neuro-psychoanalysis 4 and based on that access it identifies three layers: the neuronal layer, that is the hardware on the bottom of the model, the psyche or mental apparatus on top, which covers the description of the mental activities, and the neurosymbolic layer in between, with the task to translate neural signals into mental symbols to be processed by the mental apparatus respectively the psyche as it is modeled in the topmost layer 5.

Based on the metapsychologic theory of FREUD in SiMA a number of some 50 sub functions of the psyche have been identified step by step in a top down design process which is not yet completely finished. The process has been named *Case Based Engineering*. It was introduced in 6 and is shown in Figure 1.

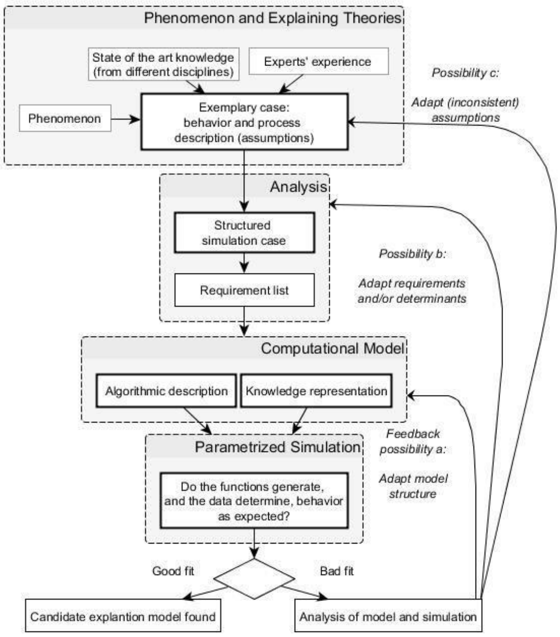

Fig. 1. Case based Engineering in SiMA 6

Based on the state of the art knowledge from different disciplines and the experience of experts a certain phenomenon gets described in an exemplary case based on assumptions about the behavior of the involved components (e.g. human or robot agents) and the observed process (e.g. an agent at a party who has to decide whether to get food from the buffet or to talk to an other agent). The case then gets structured in detail and described as a use case for the simulation. As a result the

[1] **Proceedings of the 3rd International Conference"Information Technologies for Intelligent Decision Making Support", May 18 - 21, Ufa, Russia, 2015**

requirements list for the computer model can be derived. Based on the requirements both of the following is developed: the description of the proper algorithms and the knowledge base of the agent. Now the agent gets tested in the described simulation case. Does it behave as expected? If so, a candidate for a model explaining the described behavior of an agent is found, otherwise the failing has to be analyzed. If the problem is merely in the technical realization of the artificial psyche an adaptation of the model structure is required (possibility a). The failure could also be in a misinterpretation of the exemplary case so that the identified requirements and determinants need revision (possibility b). But it might also be the result, that the machine is right. If the psyche really would work the way as described by the experts, this would lead to a behavior inconsistent with the exemplary case (possibility c). In such a case the complete theory of how the psyche works would have to be reinvestigated.

But what has all this to do with data mining? In principle the behavior of the agent and thus the particular working of his psychic functions depend on two aspects: on bodily conditions as for instance the reaction speed of his muscles and to an important amount of the intensity of his hormones and neurotransmitters on the one hand and on his conscious and unconscious memories on the other. Currently it still seems to be impossible to model these conditions in sufficient detail so that a number of abbreviations have to be made when programming the different sub functions of the artificial psyche. Thus a number of greater amount of personal parameters have been introduced and dependent on their values the functions in the SiMA model do their processing of the input. As one can imagine it is a task beyond human abilities to correlate the combined settings of these some hundred parameters with the particular behavior of the agent and therefore to do a proper analysis of the feedback possibilities sketched in figure 1. To gain this, the SiMA project requires improved methods of mass data analysis, for which purpose the work presented in this article has been done.

2. Data Mining

In the context of this work special attention is given to the approach, based on Data Mining (DM) of this great amount of data, received during the simulation process. Data Mining – is the a set of methods to detect in data previously unknown, nontrivial, practically useful and available interpretations of knowledge, required for the decision-making in various areas of human activity [7]. As applied to SiMA project the main purpose of Data Mining technology can be characterized as a way to find out in large quantities of received simulation data not clear, objective and useful for understanding the agent behavior patterns. The process of Data Mining is some kind of investigation. It provides the proper knowledge detection from the data, especially when the behavior of the agent is unexpected from the input parameters.

A base for Data Mining are different methods of classification, modeling and prediction, based on the usage of neural networks; genetic algorithms; evolution programming and fuzzy logic. As the methods of Data Mining are used different statistical methods, such as descriptive analysis, correlation and regression analysis, componential analysis.

At the first steps the main task is to find out in the simulation process log-file relations between the concrete parameters of the agent on the input and the agent behavior on the output. The input parameters are presented in the form of body or personality parameters with the appropriate modification range (for example, STOMACHSYSTEM.STARTLEVEL_STOMACH with the range from 0.0 to 0.7). Behavior of the agent is presented as the consequence of acts (for example, ACT:A13_GIVE_CAKE_L01). The certain scenario is also taken into the acount (fig.2).

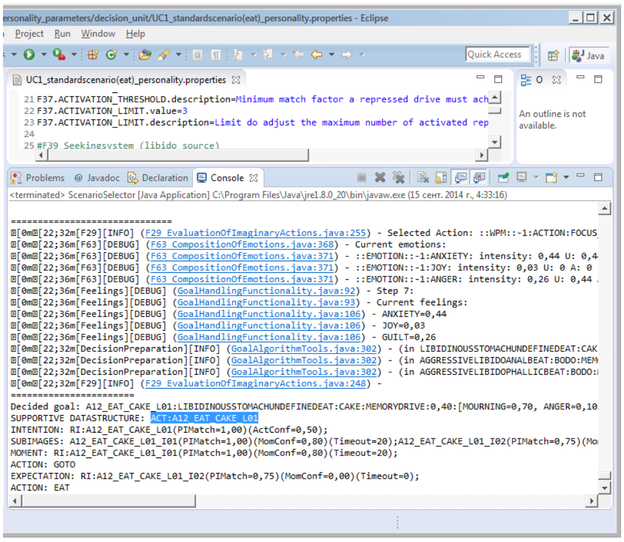

Fig. 2. Fragment of the log-file

The next step is to make the analysis of the received data according to the defined measures of similarity. On the base of obtained results we can make decisions about how the modifications of the parameters influence on agent's behaviour or does it have any influence or not. Meanwhile we have the task, connected with representing of the obtained knowledge. Within the DM approach the received knowledge can be presented in one of the following forms: associative rules; decision trees; clusters; mathematical functions. In the context of conducting research it is proposed to construct the ontological knowledge base where the received DM results are presented in the form of SWRL rules.

Ontology is a set of constraints (relations) between the elements (concepts) of the knowledge base, which is ordered, and different kinds of relationships allow organizing a complex hierarchical structure of domain concepts [8]. It is should be noted that in subject domain analysis and modeling a requirement of concepts allocation from a large number of descriptive information, which will be subsequently classes or attributes of these classes, arises. Ontology provides two main functions: defines the general terminological base for all decision-makers, allows formulating of rules and cases, using unambiguous domain concepts.

Consider a fragment of the developed DM ontology in Protégé 4.3., based on information analysis of simulation process (fig. 3).

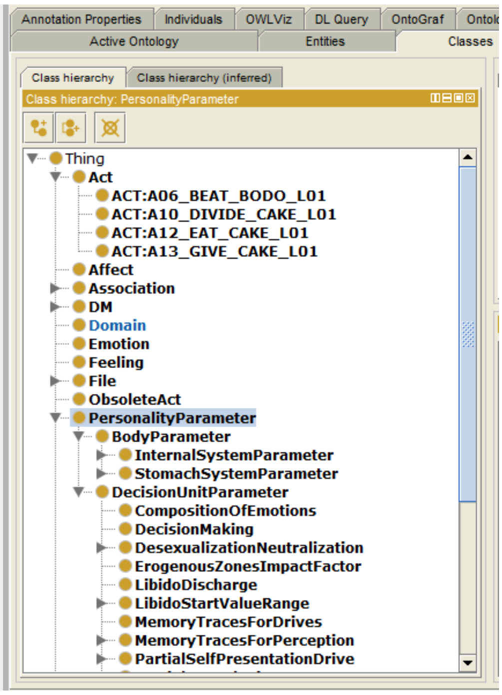

Fig. 3. Fragment of DM ontology

Agent's body and personality parameters are defined in the ontology as classes with appropriate properties and values. Ontological models are used in SiMA project for representing the agent's memory, so it is possible for the project members to work easily with the Data Mining ontology. Ontological approach has several advantages that justify its usage in environments like SiMA project with large amounts of information and the need for rapid retrieval of its parts: collection of theoretical knowledge, its representation as a semantic network of concepts and relations between concepts, as well as improving the efficiency of information search based on structuring and classification of the stored knowledge. Ontology solves the problem of sharing and reuse of the knowledge by different users and computer programs. Knowledge becomes visible and useful for all members of the project team [9].

Thus ontology can be presented as the set of the following components (1):

$$Onto = \langle C, Pr, V, I, R, A, D \rangle, \quad (1)$$

where C – a set of classes $\{C_1, C_2, \ldots, C_n\}$ and their interpretations in certain knowledge domain;
Pr – properties of classes;
V – values of properties; in OWL exists division of the properties on two classes: object properties (individuals of *owl:ObjectProperty*) and data type properties(individuals of *owl:DatatypeProperty*);
I – a set of individuals $\{I_1, I_2, \ldots, I_n\}$, defines with the help of axioms and definition of concrete class properties (called facts);
R – a set of relations $\{R_1, R_2, \ldots, R_n\}$;
A – a set of axioms $\{A_1, A_2, \ldots, A_n\}$;
D – a set of ontology inference algorithms $\{D_1, D_2, \ldots, D_n\}$.

Results of data analysis are presented in ontology as the rules. The DM ontology, added by a set of rules, forms the ontological knowledge base (fig.4).

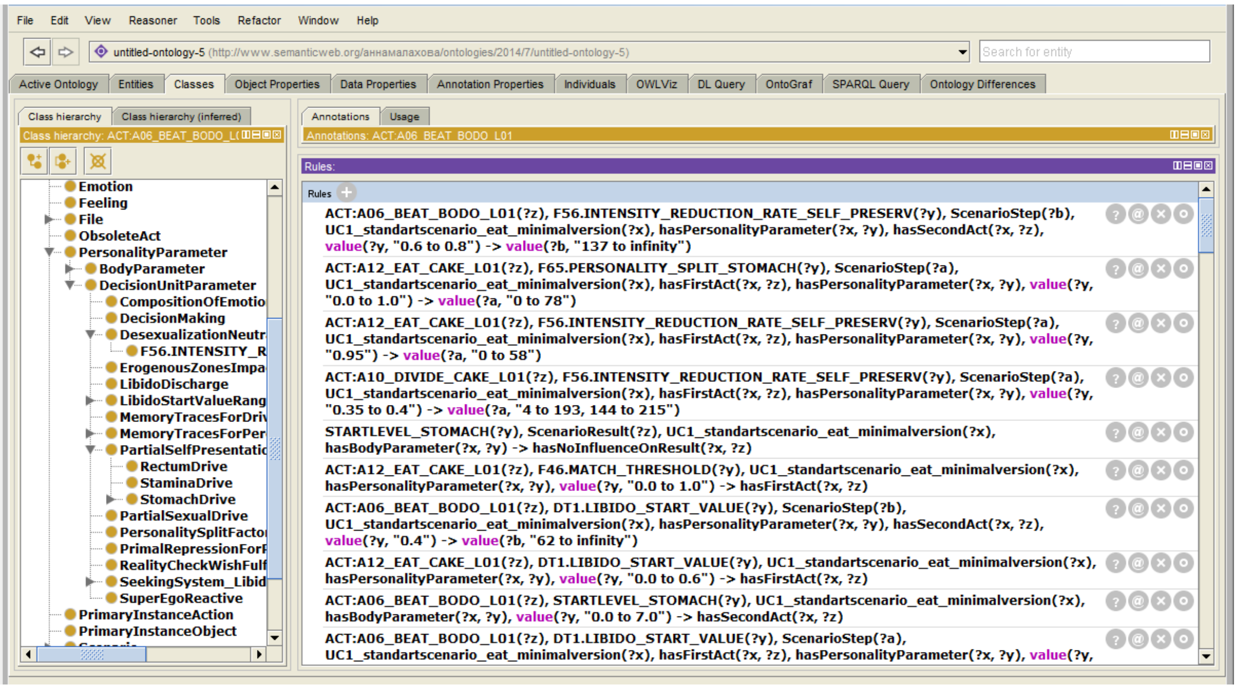

Fig. 4. Ontological knowledge base

Rules are constructed on the basis of DM ontology on SWRL language (Semantic Web Rule Language). SWRL is an extension of OWL DL as in a form of abstract syntax, and for the expression of the theoretical model semantics for writing rules. Obtained SWRL rules are written in the form of clauses of Horn (2). Rules operate the notions of concepts and roles. Concepts describe the domain classes, and roles describe relationships between them:

$$C_1(?x) \wedge C_2(?y) \wedge P_1(?x,?y) \wedge C_3(?x,?z) \rightarrow C_2(?z,?y) \quad (2)$$

where $(C_1,C_2,C_3) \in C$; $P_1 \in P$; x, y – instances or variables; z – variables or values.

Rules are constructed in such a way to show how the agent acts on the concrete step of the simulation process scenario relative to the defined value of the input parameter. For searching the rules in the ontological knowledge base in Protégé the information search module can be used. The module allows extracting the information from the ontology, based on data queries on language SPARQL. Formulated query can be saved by adding it to the ontology query library.

3. Conclusion

Applying the methods of Data Mining and knowledge engineering allows to implement the mass analysis of the SiMA project experimental data. Results of such analysis are reflected in the developed ontological knowledge base in Protégé 4.3in the form of SWRL rules. It can be effectively used for decision making in project SiMA. At the current step of the research the main difficulties are connected with automation of the Data Mining process.

References

1. Samer Schaat, Alexander Wendt, Matthias Jakubec et al.; "ARS: An AGI Agent Architecture"; in Ben Goertzel, Laurent Orseau, Javier Snaider (Eds.); "Artificial General Intelligence"; *Proceedings of the 7th International Conference, AGI 2014*; Springer; New York u. a.; 2014; S. 155 – 164.
2. *Simulation of the Mental Apparatus and Applications;* http://sima.ict.tuwien.ac.at/ ; accessed 2015-01-09.
3. Dietmar Dietrich, Georg Fodor, Gerhard Zucker, Dietmar Bruckner (eds.); *Simulating the Mind*; Springer; Wien; 2009.
4. Mark Solms; *What is the 'Mind'? A Neuro-Psychoanalytical Approach*; in Dietmar Dietrich, Georg Fodor, Gerhard Zucker, Dietmar Bruckner (eds.); *Simulating the Mind*; Springer; Wien; 2009; S. 115 – 122.
5. Dietmar Dietrich et al.*; Naturwissenschaftliches, psychoanalytisches Modell der Psyche – Technical Report II*; Institut für Computertechnik, Technische Universität Wien; 2014.
6. Samer Schaat, Dietmar Dietrich; *Case-Driven Agent-Based Simulation for the Development and Evaluation of Cognitive Architectures*; 26th Benelux Conference on Artificial Intelligence; pp. 73-80; November 2014.
7. Duke V.A. Samolenko A.P. Data Mining: Studing course. – St-Pb.: Piter, 2001. – 368 pp.
8. *Spinning the Semantic Web. Bringing the World Wide Web to Its Full Potential.* Edited by Dieter Fensel, James A. Hendler, Henry Lieberman and Wolfgane Wahlster. Foreword by Tim Berners-Lee. The MIT Press, Cambridge, Massachusetts, London, England, 2003, – 503 pp.
9. Decision support in strategic management based on knowledge engineering / under the editorship of L.R. Chernyakhovskaya – Ufa: Republic Bashkortostan Academy of Sciences, Gilem, 2010. – 128 pp.