On Construction of Inductive Modeling Ontology as a Metamodel of the Subject Field

Halyna Pidnebesna, Volodymyr Stepashko

International Research and Training Centre for Information Technologies and Systems of the NAS and MES of Ukraine, Glushkov ave., 40, Kyiv, 03680, Ukraine, e-mails: stepashko@irtc.org.ua, pidnebesna@ukr.net

Abstract: **The paper considers the constructing issue of ontology for the GMDH-based inductive modeling domain. It examines the main components of the GMDH algorithms in terms of their synthesis for designing the domain ontology to construct inductive modelling tools. Such approach significantly expands opportunities for construction of GMDH-based tools for building forecast models of complex processes of different nature.**

Keywords: **inductive modeling, GMDH, metamodel, domain ontology, model class, structure generator, selection criterion, computation tool**.

I. Introduction

With the development of technologies, the Internet, an extremely large and constantly growing number of information resources, there was the need to develop tools for automatic processing and analyzing data of different nature with taking into account the semantics (content) of this information. Some intelligent knowledge-based tools have been rapidly developed.

"Intelligent" computer systems (with artificial intelligence properties) may be understood as the ability to find ways to solve any task automatically, without (or with minimum) human intervention. The necessary features of such systems are adaptability, the ability to take into account the results obtained earlier, getting problem solutions by analogy with other cases, building valid (effective) algorithms, using the knowledge contained therein. That is, an intelligent computer system should be able to simulate the process of constructing an algorithm for solving the current problem, like human considerations.

In the design of computer modeling systems with artificial intelligence properties, the perspective direction of research is the use of an ontological approach to knowledge representation. This allows expanding computer capabilities, increases their "intelligence" and also simplifies the process of developing and modifying software products to solve specific tasks of constructing models and forecasts.

The advantage of an ontological approach is that ontology defines a conceptual structured environment in which the process of constructing a model of an object occurs [1]. This environment should be independent on the choice of a particular simulation object.

The automation task for intelligent computer modeling systems may be interpreted as modeling of the modeling process, which may be called as metamodeling. A metamodel is a model that describes the structure, principles of other models' operation.

The paper considers ways to increase efficiency of the development of inductive modeling software.

The Group Method of Data Handling (GMDH) [2] is one of the most effective inductive modeling methods having intelligent properties [3]. It is the method of building models with automatic selection of structure and parameters on the basis of a short data sample with incomplete and uncertain input information to identify unknown relationships of the object or process under study.

In this paper, an approach is considered for "intellectualization" of inductive modeling software tools by applying an ontological approach to representing the knowledge of the subject field to design a knowledge base, computing tools and intelligent interface.

The aim of this research is to build ontology of the inductive modeling subject area. For this purpose, the analysis of the modeling process is done and main its stages are characterized. The results of the analysis and structuring of this area are presented. The basic components and characteristics of them are determined and main principles of the GMDH ontology construction are outlined.

II. Structuring Knowledge of Inductive Modeling Domain

Modeling is a process of studying a real object, in which only some of its specific characteristics, description, and conditional image are used. We consider the mathematical modeling that is studying the properties of an object by analyzing and constructing its mathematical model.

There are two main approaches to constructing mathematical models of objects: the theory-driven (or *deductive*) and the data-driven (or *inductive*) ones (Fig. 1).

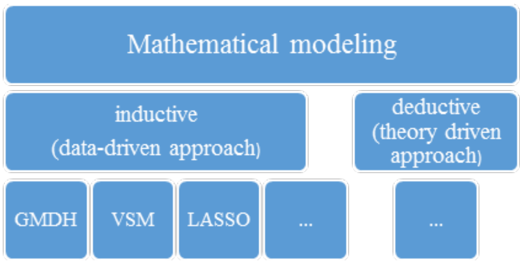

Fig. 1 Mathematical modelling approaches

Inductive modeling is the construction of a model based on the analysis and generalization of the statistical data about the object, obtained through observations or experiments.

Methods in the field include such algorithms for finding hidden patterns in data: GMDH, discovering of associative rules, sequence analysis, classification, regression, random forest, neural networks, support vector machine (SVM), genetic algorithms, least absolute shrinkage and selection operator (LASSO) etc.

The inductive modeling algorithms solve a range of tasks:

• building mathematical models of objects/processes;

• forecasting processes specified by time series;

• construction of classification rules (supervised learning) for attributing an object to a given class;

• clustering (unsupervised learning or self-training: identification of effective features, forms and rules of distinction); in GMDH this problem is called “Objective Computer Clasterization” (OCC);

• objective system analysis (OSA) when one need to find out which variables among the measured ones are independent (inputs), dependent (outputs) and irrelevant (uninformative) for building an appropriate model.

Inductive modeling based on statistical data is a process of sequential decision making, consisted of certain successive stages (Fig. 2). All methods of inductive modeling have standard components. This can be the basis of the metamodel of inductive modeling.

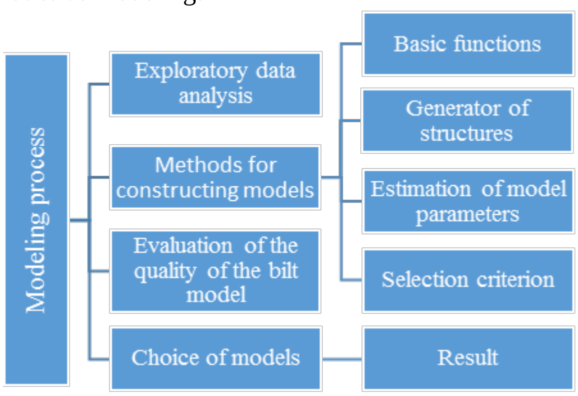

Fig.2 Components of inductive modeling process (fragment of the metamodel)

A metamodel provides the logical level of the domain and is interpreted dynamically at the application level. This adds additional flexibility to the system, since the domain logic can be changed without modifying the code. To allow or prohibit a particular type of communication at the logical level, it will suffice only to assign it to the formal terms of the metamodel.

In fact, the metamodel may be defined as a high level ontology, in terms of concepts of solution methods, key stages and constraints (Fig. 3). The ontological model of the subject domain of the lower level describes the algorithmic components of each particular modeling method in more details. To solve a practical task, an ontological model of a task is used having its own parameters, specific characteristics and areas of admissible values.

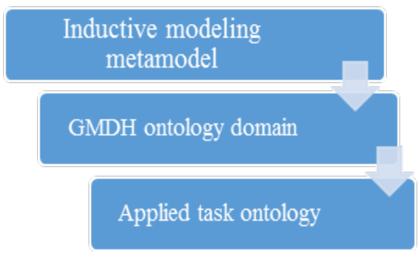

Fig.3 Hierarchy of GMDH domain ontologies

Any real problem can be characterized by the following main stages of the process of its solution: preparation; preliminary analysis; formulation of task; solving the task; analysis of results; their application. This upper level of structuring is supplemented by more detailed classifiers of subsequent hierarchical levels depending on the specificity of the problems under consideration.

The preparation of the task consists in determining the type of task (modeling of statics, time series or dynamics), modeling goals (approximation, interpolation, extrapolation, prediction, search for regularity), experiment planning (if the simulated system allows experimentation), obtaining a set of data (as a result of an active or passive experiment), their preliminary processing and organization of storage in the relevant database [4].

As a result of the field structuring, the principles of formation of algorithmic modules for solving a class of specific problem are determined. Depending on the type of tasks, adequate methods for solving them are selected.

Each of available methods corresponds to certain characteristics. According to them, it is possible to choice (may be automatically) a better method for a specific task. To do this, each of the set of solution methods should have some weight of importance to assess the adequacy of choosing this particular method at this stage. When choosing the appropriate (for a particular case) method at each stage of the modeling process, we get an algorithm (possibly the best) for solving a specific problem as a result of sequential synthesis in a structured set of possible options.

III. Metamodel as the High Level Ontology

To significantly expand the scope of computer modeling systems, they must be independent of the particular simulation object and of the means of its implementation. That means there should be a high level of abstraction of the subject area.

Raising the level of abstraction is difficult, and developers are forced to take out part of the information model for some application. This means fixing this part of the model. At the same time, the setting flexibility for the subject area is lost. The solution of this problem is seen in the introduction of metamodels. Metamodels reduce uncertainty in the description of the subject area and allow to get rid of rigid fixation on the task specificity.

First of all the metamodel helps to determine the structure of the process and allows developers to show specific requirements of the process automation means. The metamodel defines "design details", from which a modeling system may be subsequently created.

Metamodels are closely related to ontologies, because they are used to structure information and to analyze the relationships between concepts. The ontology divides the variables needed for some set of computations and establishes the relationship between them [5].

Modeling can be considered as an explicit description (design and rules) of how a problem-oriented model is constructed. As a rule, metamodels are a strict set of rules. A real metamodel is an ontology, but not all ontologies are represented explicitly as metamodels.

The internal structure of an intelligent computer system is a reflection of certain knowledge that needs to be expressed explicitly, in a formal way. The use of ontologies can facilitate the description of the task of designing complex systems from components and implement a program that makes such a configuration independent of the product and the components itself, makes it possible to reuse.

Ontology is the exact specification of some field that contains a glossary of terms and a set of subject area links describing relations between these terms. It actually is a hierarchical conceptual skeleton of the subject area.

Formal ontology model (O) is an ordered triplet [6]

$$O=<T, R, F>,$$

where:

T is finite set of terms of the subject area being described by the ontology *O*;

R is finite set of relations between the given terms;

F is finite set of the interpretation functions given on the terms and/or relations of the ontology *O*.

The purpose of creating and using ontologies is support for activities to accumulate, distribute and reuse knowledge in a particular subject area.

Ontology allows one to specify a complex structure that can contain different types of data, provide a simple understanding of the presentation of structured knowledge and relatively easy updating.

In general case, the ontological model of the presentation contains a description of the situation/task (data, the purpose of the modeling) and the appropriate solution (algorithm for obtaining an adequate model). In most cases, in order to obtain an algorithm for solving a problem, it is sufficient a parametric representation in the form of a set of corresponding parameters given by the ontology, with specific values. In what follows, there is an example of an ontological representation of knowledge of the domain of inductive modeling.

IV. GMDH AS A METHOD OF MODEL BUILDING

The Group Method of Data Handling (GMDH) is one of the most effective methods of modeling from statistical data, which fully implements the essence of the inductive approach in modeling and has the intelligent properties.

GMDH is the method for constructing models with automatic determination of model structure and parameters from a data sample under conditions of incompleteness and uncertainty of input information in order to detect an unknown operation rule of an object or process under study.

GMDH characterizes by application of principles of automatic model generation with inductive complication of variants, non-definitive decisions and sequential selection according to external criteria for constructing models of optimal complexity. For comparison and selection of the best models, external criteria are used which are based on splitting the sample of input data into two or more parts. Estimation of parameters and quality assurance of models is carried out on different subsamples, which allows to automatically take into account different types of a priori uncertainty when constructing a model. These principles can be considered as metamodel characteristics for the process of building mathematical model of an object (process).

V. ONTOLOGICAL MODEL OF GMDH-BASED INDUCTIVE MODELING PROCESS

To structure knowledge in a domain, one needs to consider the following issues:

- define the main stages for solving typical problems in a specific domain to obtain the basis for constructing the metamodel of the inductive modeling process;
- identify the main methods for effective solving these problems to form the basis of the domain ontology;
- generalize the experience of applying these methods to develop relevant intelligent software tools.

Obviously, each of these problems has a complex multilevel structure. The results of analysis of these problems are used to create the ontology of the subject field.

GMDH as one of the methods of inductive modeling also has a standard sequence of stages to solving a specific problem, as discussed in [7].

Ontology development is an integrated, sequential and iterative process. At the top level, the ontology contains a list of concepts and their general properties. In fact it is a thesaurus. A dictionary or a list of concepts is collected as a result of structuring knowledge domain. The next important step is to rank and organize the terms and build a hierarchy. In [8] a fragment of thesaurus of GMDH was given and general principles and main stages were described. The ideas given in [8] are substantially generalized in this paper.

The next step is more detailed study of GMDH algorithms, definition of the ontology structure and characteristics of the stages of the choice of a models class, structure generators, and model evaluation criteria. The result is the construction of corresponding ontological models.

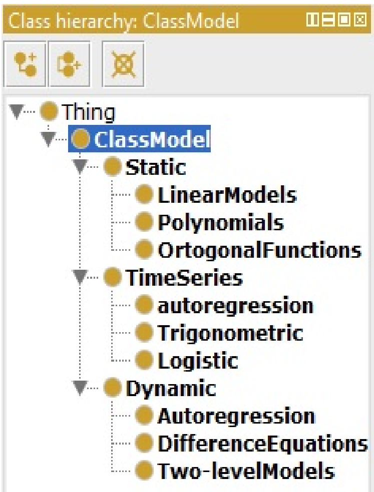

Fig. 4 The ontology of model classes

The ontology of *classes of models CM* (Fig. 4) is characterized by such key parameters as the number of input and output variables, and the number of past values (latencies) taken into account for input and output variables, respectively. Depending on the specific values of these parameters, one can obtain most of the variants of linear models that are used in practice to describe static objects, time series and dynamic objects and processes.

The *model generators* ontology *GS* (Fig. 5) contains two main types of GMDH structure generators: sorting-out and iterative ones. In turn, typical sorting-out algorithms to form different model structures are COMBI (exhaustive search) and multistage MULTI (directed search) [9]. Two main architectures of iterative structure generators are multilayer

MIA and relaxational RIA.

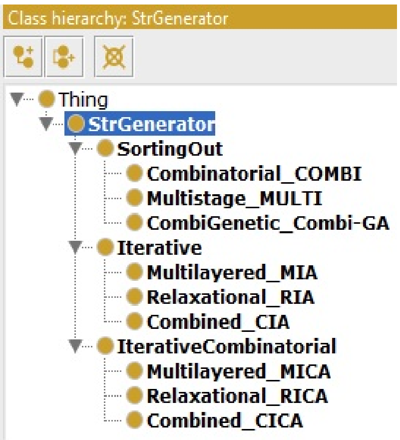

Fig. 5 The ontology of model generators

In recent years, new kinds of GMDH algorithms have been developed: generalized iterative algorithm GIA [10] and the hybrid combinatorial-genetic algorithm Combi-GA [11]. So, taking into account current trends, the ontology of generators of structures may be presented in the form of Fig. 5.

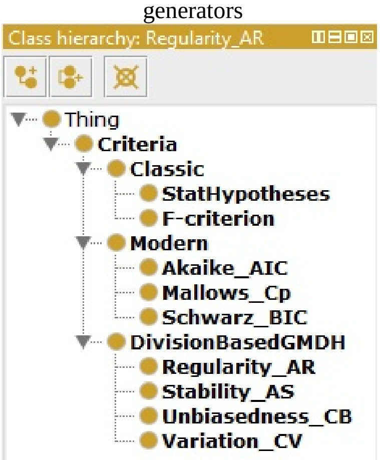

Fig.6 Ontology of model criteria

Ontology of *model criteria CR* (Fig. 6) may be defined by key parameters that describe the penalty functions for the model complexity, the model quality, estimations of the unknown variance. They characterize a set *CR* of criteria, which are applied in practice for tasks of structural identification of models of optimal complexity.

In [8] an example is given for ontological model of sorting-out GMDH algorithm COMBI. The same way can be defining ontological models for other GMDH algorithms. For instance, let us consider the following set of parameters:

- an element of the model classes set *CM* is k_i^* = < linear regression models >,
- an element of structure generators set *GS* is g_i^*=< sorting-out algorithm:: directed search >,
- an element of parameter estimators set *EP* is p_i^*=< least-squares method >,
- an element of selection criteria set *CR* is r_i^*=< regularity criterion >.

This set of parameters of the inductive modeling ontology defines the sorting-out GMDH algorithm MULTI [9].

In case when element of structure generators set *GS* is g_i^*=< iterative algorithm:: relaxational >, it defines the Relaxational iterative GMDH algorithm RIA.

In case element of structure generators set *GS* is g_i^*=< iterative algorithm:: multilayered >, it defines the Multilayered iterative GMDH algorithm MIA.

These are examples of ontology models as part of GMDH-based domain ontology. Preliminary analysis of the subject field enables the generalization of many different methods, identifying the key parameters. Ontology allows defining both general rules for constructing the algorithm and specific parameters when making an application.

VI. Conclusion

The way to generalization of software tools of inductive modeling means by applying an ontological approach as metamodel representing the knowledge of GMDH-based domain is considered. This enables substantial simplification of developing specifications and software tools for solving various applied tasks.

The paper presents the results of structuring of the inductive modeling domain. The examples of main components of the modeling process defining their basic characteristics for building ontologies are provided. Some fragments of the ontology constructed using Protégé are presented as significant modules of the domain metamodel.

References

[1] T. Gruber, “Toward principles for the design of ontologies used for knowledge sharing,” *International Journal Human-Computer Studies, 43(5-6)*, 1995, pp. 907-928

[2] H.R. Madala, A.G. Ivakhnenko, Inductive Learning Algorithms for Complex Systems Modeling. New York: *CRC Press*, 1994, 384 p.

[3] V.S. Stepashko, Conceptual fundamentals of intellectual modeling. *Control Systems and Computers*. – Kyiv: IRTC ITS, #4, pp. 3-15 (In Russian)

[4] V.S. Stepashko, On the problem of structuring the expert's knowledge in the field of modeling by empirical data. ISSN 0454-9910. *Kibernetika i vychisl. tekhnika*. 1991, Issue 92, pp. 80-83. (in Russian)

[5] Metamodeling, [cited 2017 Oct. 16]. Available from: https://en.wikipedia.org/wiki/Metamodeling.

[6] T.A. Gavrilova, V.P. Khoroshevsky, Knowledge Base Intelligent Systems, SPb.: *Piter*, 2000, 384 p. (in Russian)

[7] V. Stepashko, G. Pidnebesna, “Generalized Multifunctional Modules Concept for Construction of Inductive Modeling Tools,” *Proc. of the 4th Int. Conf. on Inductive Modelling ICIM-2013*, Kyiv, Ukraine, Kyiv: IRTC ITS NASU, 2013, pp. 225-230.

[8] H. Pidnebesna, On Constructing Ontology of the GMDH-based Inductive Modeling Domain, *Proc. of 8th International Workshop on Inductive Modeling IWIM 2017*, Lviv, Ukraine, 2017, pp.511-513.

[9] V.S. Stepashko, "A Finite Selection Procedure for Pruning an Exhaustive Search of Models," *Soviet Automatic Control*, 1983, vol. 16, no. 4, pp. 84-88.

[10] V. Stepashko, O. Bulgakova, V. Zosimov Construction and Research of the Generalized Iterative GMDH Algorithm with Active Neurons. – In: *Advances in Intelligent Systems and Computing II. CSIT 2017 /* Shakhovska N., Stepashko V. (eds). – *Advances in Intelligent Systems and Computing, vol 689. Springer, Cham, 2018.* – P. 492-510.

[11] V. Stepashko, O. Moroz, “Hybrid Searching GMDH-GA Algorithm for Solving Inductive Modeling Tasks,” *IEEE Int. Conf. on Data Stream Mining & Processing*, Lviv, Ukraine, pp. 350-355, August 2016.