Intelligent Techniques In Business Games And Simulations – A Hybrid Approach

(International Simulation and Gaming Association 41st Annual Conference, 5-9 July 2010, Spokane WA, proceedings)

Mihail Motzev, Ph.D, MSc
Walla Walla University School of Business, College Place WA, USA
Mihail.Motzev@wallawalla.edu

Abstract. Developing a good, accurate model is very important element in business games and simulations. There are many problems in model identification, pattern recognition, approximation and extrapolation. To address these problems new techniques like artificial neural networks, genetic algorithms etc. have been developed and applied to analyze existing massive amounts of data and extract useful information. This paper presents a hybrid approach based on self-organizing data mining. The results show that it is able to develop even complex models reliably and achieves lower overall error rates than state-of-the-art methods.

1. INTRODUCTION

Models enable us to study various functions and behavioral characteristics of a particular system and its subsystems as well as how the system responds to given changes in inputs or reacts to changes in parameters or component characteristics. Models also enable us to study the extent to which outputs are directly related to changes in inputs - whether the system tends to return to the initial conditions of a steady state after it has been disturbed in some way, or whether it continues to oscillate within control limits. A good model can help us to understand which behavior is relevant to or to what extent the system is responsible for changes in environmental factors.

Models form the basis for any decision. They support and assist decision makers in many different ways. Models make it possible to recognize structure and function of the system (subject of identification). This leads to a deeper and better understanding of the problem. Usually, models can be analyzed more easily, faster and cheaper than the original problem. They help to find appropriate means for cause-and-effect influence on an object (subject of control) and to predict what the system has to expect in the future (subject of prediction). Eventually, models make possible to run experiments with the system of interest (subject of simulation), and apply "what-if" analysis.

2. MODELS AND BUSINESS GAMES

2.1 Model-based Business Games

The term 'model-based' is generally used to describe games that have a repetitive decision/result cycle moving onward through time – each cycle represents a certain length of real time, such as a month, a quarter or a year. The players' decisions establish the policies to be followed during that period, and the results are 'what happened as a result of them'.

Model-based business games were developed in business schools and universities, using the concept of economic modeling. Players receive a description of an imaginary business and an imaginary environment and make decisions – on price, advertising, production targets, etc. – about how their company should be run. The decisions are compared with a model, which determines how well they have fared. A 'model' in this sense is a set of mathematical rules which state that if a certain decision is taken then a certain result will follow. The detail of the rules is not made known to the players. The simplest example of such a rule is a price/demand relationship establishing a sales/demand figure for every possible asking price.

Chris Elgood, author of the bestselling Handbook of Management Games and Simulations, pointed out (Elgood, 2005): "Models are also core features of other types of game, but in this type – named for them – they have special prominence. Players submit their decisions to the same model time after time: its presence is ubiquitous." This paper is concentrated on models' development for any type business game.

A model of this sort can be very simple or very complex one. In spite of its complexity each model should be accurate enough. Unfortunately in economy, ecology, sociology etc. many objects are ill-defined systems that can be characterized by: inadequate a priori information about the system; big number of immeasurable variables; noisy and/or small data samples, and fuzzy objects with attributive variables.

Problems of complex objects modeling like systems identification, pattern recognition, approximation & extrapolation, and forecasting can be solved by deductive logical-mathematical modeling or by inductive sorting-out methods.

2.2 Theory-driven Modeling

In theory-driven approach (also known as theoretical systems analysis), models can be derived from existing theory. This approach leads to the hard systems method of modeling, based on the assumption that the world can be understood objectively and that knowledge about the world can be validated by empirical means. The elements of a model are represented by different variables described by numerical values. Then, the cause-and-effect relationships are formulated as mathematical equations or their equivalents.

A key assumption of this approach is that it is possible to construct and manipulate a model of the problem under study, i.e. the researcher has a priori knowledge and can describe the essential structure of the original system at least as far as it is necessary for the model purpose. In addition, the structure of relationships within the original system must be known and the environmental influences as well.

Unfortunately, in complex, ill-defined systems such as most business processes, researcher a priori has only insufficient knowledge about the relevant theory of the system under study. Thus, theory-driven modeling is affected considerably by the fact that the researcher is a priori uncertain regarding selection of the model structure due to insufficient knowledge about its variables and relationships.

The comprehensive application of theory-driven approach and theoretical systems research are hampered by several essential problems, which can be summarized by Zadeh's principle of incompatibility (Mueller & Lemke, 2003):

"As the complexity of a system increases, our ability to make precise and significant statements about its behavior diminishes until a threshold is reached beyond which precision and significance (or relevance) become almost mutually exclusive characteristics. A corollary principle may be stated succinctly as: the closer one looks at a real-world problem, the fuzzier becomes its solution."

2.3 Data-driven Approach

As outlined in the previous section, problems of complex object modeling cannot be solved in full by deductive logical-mathematical methods. In inductive sorting-out methods (known also as data-driven approach or experimental systems analysis) models can be derived from data. Knowledge extraction from data, i.e., to derive a model from experimental measurements, has advantages when a priori only a little knowledge or not well-define theory is on hand.

Data-driven approach generates a description of the systems behavior from observations of the real system evaluating how it behaves (output) under different conditions (inputs). This is similar to statistical modeling and its goal is to infer general laws from data samples. The mathematical relationship that assigns an input to an output and imitates the behavior of a real-world system using these relationships usually has nothing to do with the real processes running in the system. The systems details and relationships are not described at all. System is treated as a black box and this approach cannot be used to analyze cause-and-effects relationships in such fuzzy objects.

Another problem is that many other factors that are not observed or controlled may cause influence on the system's output, i.e. the knowledge of observed input values does not uniquely specify the output.

Problems exist in both groups and a possible solution is in unification of these methodologies. Knowledge discovery from data and in particular data mining techniques can help researchers analyzing the massive amounts of data and turning information located in the data into successful decisions.

3. SELF-ORGANIZING DATA MINING

3.1 Knowledge Discovery from Data

Knowledge Discovery in Databases (KDD) refers to the overall process of discovering useful knowledge from data (Fayyad & Shapiro & Smyth, 1996), and data mining refers to a particular step in this process (fig. 1). Data mining (DM) is the application of specific algorithms for extracting patterns from data, i.e. DM is a step in the KDD process that consists of applying data analysis and discovery algorithms that produce a particular enumeration of patterns (or models) over the data.

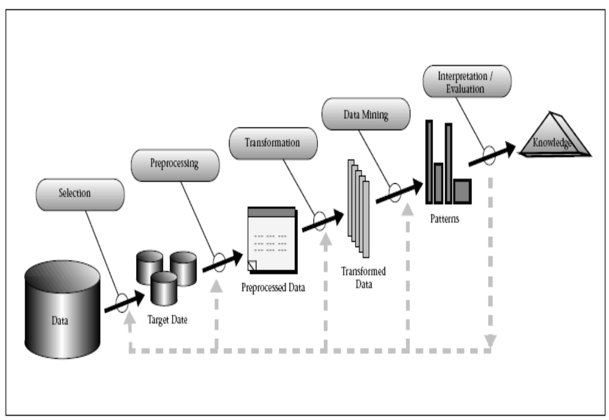

Figure 1: An overview of KDD process

The additional steps in the KDD process, such as data preparation, data selection, data cleaning, incorporation of appropriate prior knowledge, and proper interpretation of the results of mining are essential to ensure that useful knowledge is derived from the data. Blind application of DM methods (rightly criticized as data dredging in the statistical literature) can be a dangerous activity, easily leading to the discovery of meaningless and invalid patterns.

3.2 Data Mining Process

There are three main phases in DM process (Mueller & Lemke, 2003):

A. Data Selection and Preprocessing - data selection is a form of domain knowledge. Only a domain expert can formulate the task and figure out what relevant information is available or can be collected. Moreover, data have to be cleansed, selected and preprocessed by someone with a good deal of knowledge about the problem domain under study. Without the guidance and knowledge of a domain expert, these steps are impossible. One approach to automate this step in DM is to execute an automatic sensitivity analysis that detects which variables should be used. It has even advantages, because what may look as an outlier sometimes, and deleted from the analyst's viewpoint can actually be a key data point worth focusing on. Standard preprocessing and data representation steps such as scaling, symbol mapping, and normalization, necessary in some DM tools can be automated using rules and core statistical techniques.

B. Choice and Application of DM Algorithms - many techniques have been developed yet. This set of methods contains data visualization, tree-based models, artificial neural networks (ANN), methods of mathematical statistics and artificial intelligence. They can be applied to perform activities like associations, clustering, classification, modeling, sequential patterns, and time series forecasting. In this paper we'll present a hybrid algorithm, developed for building complex multi-input to multi-output models.

C. Analysis of Extracted Knowledge - in this phase, the DM output must be evaluated to figure out if new and useful knowledge of the domain was discovered. Data are what DM tools create, store and provide. Information (i.e. data in business context) is what the business needs. The researcher has to decide the relative importance of the facts generated by DM algorithms. The extracted information is valuable to a business only when it leads to actions which create value or market behavior that gives a competitive advantage.

3.3 Group Method of Data Handling

The Group Method of Data Handling (GMDH) is a heuristic self-organizing modeling method (Madala & Ivakhnenko, 1994). This method is particularly useful in solving the problem of modeling multi-input to single-output data. In GMDH-type self-organizing modeling algorithms, models are generated adaptively from data in the form of networks of active neurons in a repetitive generation of populations of competing models of growing complexity, corresponding validation, and selection model until an optimal complex model that is not too simple and not too complex have been realized.

The modeling approach grows a tree-like network out of data of input and output variables (seed information) in a pair-wise combination and competitive selection from a simple single individual (neuron) to a desired final solution that does not have an overspecialized behavior (model). In this approach, neither the number of neurons and the number of layers in the network, nor the actual behavior of each created neuron is predefined. The modeling is self-organizing because the number of neurons, the number of layers, and the actual behavior of each created neuron are adjusting during the process of self-organization.

It should be noted that self-organization does not replace a good domain theory. Inclusion of some well-known a priori information widens the basic scheme of self-organizing modeling by knowledge extraction from data and scientific theory (fig. 2). However, very often self-organization provides the only way to get any knowledge from a complex system or, to add some new aspects to existing theoretical fragments.

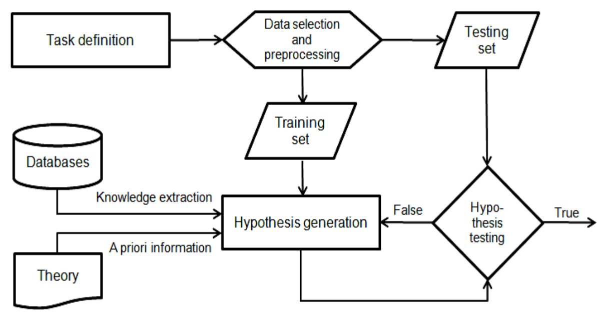

Figure 2: Self-organizing modeling with a priori information

GMDH is one of the most successful methods in Statistical Learning Networks (Mueller & Lemke, 2003), which have been developed to overcome the common problems of ANN – ANN are implicit models with no explanation component by default, designing ANN topology is a trial-and-error process, no rules how to use the theoretical a priori knowledge in ANN design etc.

4. A HYBRID ALGORITHM

Different techniques and algorithms based on GMDH approach have been developed with thousands of successful implementations since its introduction in 1968 and new ones come out almost every year (Onwubolu, 2008). Here, we'll describe a hybrid algorithm designed for synthesizing models of Simultaneous Equations (SE) – multi-input to multi-output. It contains three main parts:

Part one covers activities in DM phase A:

- Using a given data set variety of hypotheses is generated, including nonlinear transformations and taking into account data history (entered by researcher or added automatically by computer program).

- Each competing hypothesis is a hypothesis of entering one additional factor in the table of observations (i.e. that this factor is potentially important in the particular model).
- The variety is limited by criteria “correlation with dependent variable”, but can be supplemented with heuristic solutions by the researcher. In all cases “protection of the variables” is guaranteed.

The *second part* is a multi-stages selection procedure (MSSP) of each model equation (fig. 3):

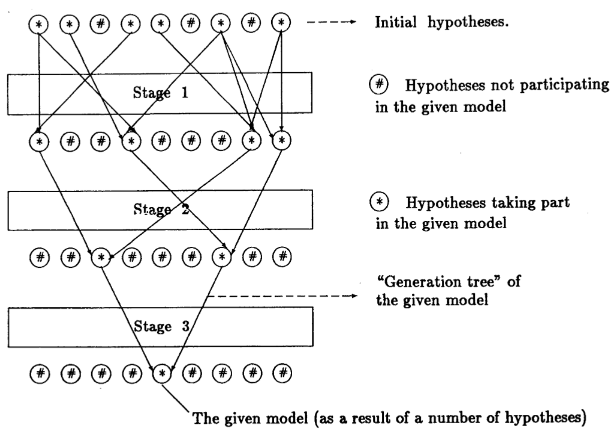

Figure 3: An overview of the MSSP

- Variety of hypotheses is generated by introducing new intermediate variables at each stage and a new generation of intermediate equations (function of two variables), which include indirectly more and more complex combinations of initial factors is created (fig. 4).
- Each generated hypothesis is a potential equation of the model, which competes with others “fighting for survival”.
- At the end of each selection stage, there is a choice of a predefined number of good equations (principle of non-finalized solutions). This gives the researcher a set of alternative good equations.
- Estimation of the coefficients in each intermediate equation is done using criteria of the mean squared error (MSE). Here, existing set of data is divided into two parts: a “teaching” set used for estimating the coefficients, and a testing set used for evaluating the adequacy of each equation (“Cross validation” principle) and choosing the good ones.
- Chosen equations in a given generation are used for generating new, more complex equations at the next stage of the MSSP. From them a predefined number is chosen as good ones etc.
- MSSP ends when satisfactory results are achieved (minimum MSE for the generation, reaching the maximum number of selections and others).

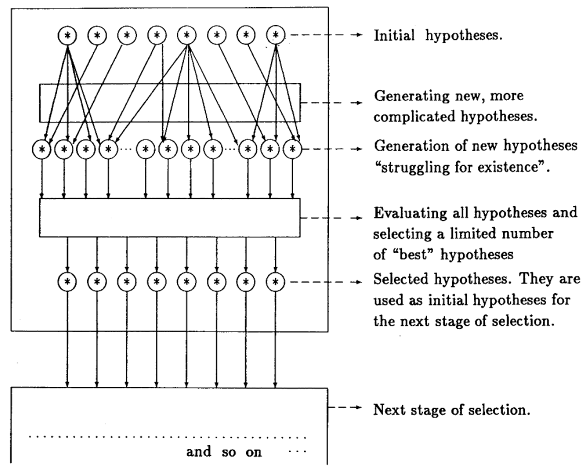

Figure 4: One stage of the MSSP

- At the end of the second part the full form of selected equations is restored using an automated backward tracking algorithm and each equation of the model has a set of alternative “good” versions.

It should be noted, that in this algorithm structural and parametrical identification of each equation is done in one, highly automated procedure. Researcher has the opportunity to apply some a priori knowledge as in theory-driven methods and then model identification is done using the MSSP. In this way, both the structure of each equation is identified and an estimation of the unknown coefficients is done.

The procedure described above is a multilayer GMDH algorithm for multi-input to single-output models identification. It can be used in many different cases, but cannot be directly applied for synthesizing complex models in the form of SE - multi-input to multi-output models. For this and some other reasons, we developed an additional third part as an iterative procedure (Motzev & Marchev, 1988). Here, the equations from the previous part are used to synthesize a model in the form of SE:

- The variety of alternative variants of the model is generated combining already chosen, good equations (because of this, similar procedures are called combinatorial algorithms).
- Each of the competing hypotheses is a hypothesis of the significance of entering a given version of a single equation into the system of SE.
- Each generated SE is considered as a potential model for the system of interest, which competes with others “fighting for survival”.
- The evaluation of these competing models is done, using a variety of criteria – MSE, coefficient of determination, MAPE, and others.
- If the results are unsatisfactory after solving the structural form of the system (biased values of the

coefficients, low accuracy of the equations, etc.) the procedure returns to part two.

- There, researcher can apply some new a priori knowledge and/or add fresh data observations (if any available), or change the selection criteria. Then a new synthesis of the structural equations is done and with so-obtained new set of equations the third part begins again. The iterations end with achieving satisfactory results.
- The final choice of the "best" model is made by researcher, who has one final option to apply additional, qualitative information/knowledge, but after having the guarantee that a big number of possible models have been evaluated and the final choice is based on a small number of good ones.

5. APPLICATIONS

There are many successful applications of Self-Organizing Data Mining Algorithms. Table 1 presents a summary of the basic type algorithms developed so far. Most of them have been used to address existing problems in simulation modeling.

Table 1: Self-Organizing Data Mining Algorithms

Variables	**Parametric**	**Non-parametric**
Continuous	- Combinatorial (COMBI) - Multilayered Iterative (MIA) - Objective System Analysis (OSA) - Harmonical - Two-level (ARIMAD) - Multiplicative-Additive (MAA)	- Objective Computer Clusterization (OCC); - "Pointing Finger" (PF) clusterization algorithm; - Analogues Complexing
Discrete or binary	- Harmonical Rediscretization	- Based on Multilayered Theory of Statistical Decisions

The first working prototype of the algorithm described above was used to develop a series of macroeconomic simulation models of the Bulgarian economy (Motzev & Marchev, 1985 & 1991). As one can see from Table 2, which presents the summary of model's characteristics, the accuracy of the most complex model, containing 39 equations and more than a hundred variables, with a time lag of up to 5 years, is very high – average mean squared error (MSE) < 1%.

It is worth noting that the algorithm was used also in time-series analysis to build autoregressive models for 24 macroeconomic variables with a time lag of up to 5 years. The average MSE for all models is 4.74% and the coefficient of determination is > 0.9 for most of them (average $R^2 = 0.9339$).

Another example presented in Table 3 shows the results from the same prototype, applied to develop a model-based business game. The game "National Economy" had been used for many years at the Economic University in Sofia, Bulgaria (Motzev & Marchev, 1984). The original version contains a model developed with theory-driven methods (multiple regression analysis). Same data and set of variables were used to build a new model, using the hybrid algorithm described above.

Table 2: Macroeconomic simulation models

Year of design	Main purpose & Accuracy	Model characteristics
1981	Analysis of the possibilities for automated model building in the form of SE using MSSP. Accuracy = 2.7%.	A one-product macroeconomic model in the form of 5 SE. Contains 5 endogenous, 5 lag and 1 exogenous variables
1985	Software development for simulation experiments with SE. Analysis of different criteria for model evaluation & selection. Accuracy = 2.0%.	Aggregated macroeconomic model in the form of 12 SE. Contains 12 endogenous, 5 exogenous and 26 lag variables with lag of up to 3 years.
1987	Improving MSSP for synthesis of a big number of SE. Simulation and prediction of the main macro-economic indexes. Accuracy < 1%.	Complex macro-economic model of 39 SE, with 39 endogenous, 7 exogenous and 82 lag variables with a time lag of up to 5 years.

The brief comparison shows that the new version has much better accuracy (almost 7 times smaller average MSE) and thus provide more reliable base for simulations and further analysis of the system of interest.

Increasing model accuracy provides many other benefits. For example it helps researcher to analyze more precisely the problem, which leads to its deeper and better understanding. Also, a model with high accuracy will generate better predictions and support managers making better decisions etc.

Table 3: Business game "*National Economy*"

Original Version	New Version
A one-product macro-economic model developed as a system of five SE. Contains five endogenous, one exogenous, and five lag variables.	A one-product macroeconomic model with the same structure. Contains same set of variables.
Indirect OLS used to estimate unknown coefficients in equations.	Model synthesized using the hybrid algorithm.
Model accuracy - mean squared error (MSE) = 14%	Model accuracy MSE = 2.7%

6. CONCLUSIONS

This paper presents a new perspective in simulations and model-based business games — using a hybrid algorithm of MSSP for synthesizing the "best" model. The proposed approach provides opportunities for shorten the time, and reducing the cost and the efforts in model building. Also, the results show that it is able to develop even complex models reliably with low overall error rates.

One element in this approach that needs further analysis is how to split existing set of data into two parts — a teaching set used for estimating the coefficients, and a testing set used for evaluating the adequacy of each equation. The so called "Cross validation" principle doesn't have yet one optimal strategy and sometimes requires additional experiments to clarify what structure in data set should be used.

Another future improvement to be made is to update the software, which had been designed for mainframes and mini computers (Motzev & Marchev, 1989). In this regard a good solution is the KnowledgeMiner program (Mueller & Lemke, 2003), one of the leading software platforms in self-organizing data mining.

With this software new models and business games could be developed fast and easy. One model-based management game, which is under construction, is "Inventory Management" game, designed for students in Production & Operations Management class at the Walla Walla University in College Place, WA.

REFERENCES

Elgood, C. (2005). *Using Management Games.* Burlington VT, USA: Ashgate Publishing.

Fayyad, U., & Shapiro G., & Smyth P. (1996). "From Data Mining to Knowledge Discovery in Databases", *AI Magazine*, vol. 17, no.3, Fall, pp. 37-54.

Madala, H., & Ivakhnenko, A. (1994). *Inductive Learning Algorithms for Complex Systems Modeling.* Boca Raton, FL, USA: CRC Press.

Motzev, M., & Marchev, A. (1984). "Applications of Management Simulation Games for Student Training in Business Education", *X Internationales Seminar Uber Rechnergestutzte Planspiele*, Berlin.

Motzev, M., & Marchev, A. (1985). "Computer Macro-economic Models for Simulation Experiments", *Systems Analysis and Simulation*, vol. 28, 1985 II, pp.145-150.

Motzev, M., & Marchev, A. (1988). "Multi-Stage Selection Algorithms in Simulation", *Proceedings of XII IMACS World Congress.* Paris, France: vol. 4, July, pp. 533-535.

Motzev, M., & Marchev, A. (1989). "Principles of Multi-Stage Selection in Software Development in Decision Support Systems", *Methodology and Software for Interactive Decision Support* (Lecture Notes in Economics and Mathematical Systems). Springer-Verlag, 337 IIASA, pp. 181-189.

Motzev, M., & Marchev, A. (1991). "Macroeconomic Models for Simulation of the Bulgarian Economy", *Systems Analysis, Models and Simulation*, vol. 8, Fall, pp. 145-150.

Mueller, J.A., & Lemke, A. (2003). *Self-Organizing Data Mining: An Intelligent Approach to Extract Knowledge from Data.* Bloomington IN, USA: Trafford Publishing.

Onwubolu, G. (2008). "Design of hybrid differential evolution and GMDH networks for modeling and prediction", *Information Sciences*, vol. 178, no.18, May, pp. 3616–3634.