

Advances in GMDH-based Predictive Analytics Tools for Business Intelligence Systems

Serhiy Yefimenko

Department for Information Technologies of Inductive Modelling, International Research and Training Center for Information Technologies and Systems, UKRAINE, Kyiv, 40, Ave Glushkov, email: syefim@ukr.net

Abstract: The paper analyzes approaches to prediction of economic processes in business intelligence systems. Contemporary tools of predictive analytics, used for effective making of business decisions, are considered. The concept of advanced GMDH-based predictive analytics tool is proposed.

Keywords: business intelligence, predictive analytics, GMDH, recurrent-and-parallel computing.

I. INTRODUCTION

Achieving success and ensuring competitiveness in today's fast changing economic conditions are impossible without the use of reliable and on-line information. Business data is becoming significant resource for knowledge acquisition and making important managerial decisions in different business fields. Up-to-date effective decisions require reliable and complete information, and it is impossible to do with the use of traditional information systems.

In our time, there is a rapid transformation of the global information area that affects society, market and business. There is a fast growth of the digital economy. 25% of the world's economy will be digital by 2020 [1], whereas this number was 15% in 2005. The Internet of Things (IOT) and Big Data, mobile and cloudy technologies contribute the economy digitization. Influence of these technologies on business will result in direct domain physical resources to become useless.

Business intelligence (BI) is a modern managerial tool in the digital economy. It contributes to the company's prosperity based on smart financial, business processes, and personnel management under considerable amount of information.

The purpose of the review is to consider modern approaches to prediction economic, production and financial processes in BI systems, as well as existing software tools for predictive analytics.

II. PREDICTIVE ANALYTICS & PREDICTIVE MODELLING

BI encompasses strategies and technologies used by enterprises to analyze business information [2]. BI refers to the management philosophy and toolkit used to help operate business information in order to make effective business decisions. BI technologies provide historical, current and predictive views on business operations.

The classification of technologies, used by business analytics, is given in [3]. Predictive modeling is one of the most effective.

Organizations of different types may be troubled by certain problems in the effectiveness of existing data using in their systems. In this regard, the quality and speed of information and analytical support for business management is of particular importance for companies. Most of them use BI analytical applications based on OLAP systems for planning, analyzing and controlling tasks. However, in new economic conditions, the functionality of such systems is not enough to solve new digital problems, since they oriented on retrospective analysis. Consequently, there is a need for predictive analytics, which complements and enhances BI capabilities in terms of predicting future events.

In general, there are several types of analytics that co-exist and supplement each other [4, 5]:

- *descriptive analytics* explores past facts in order to find the causes of previous successes or failures. It answers the question "What's up?". Descriptive analytics is still in use today. Most of the management reports for sales, marketing, finance use this kind of business analytics;

- *diagnostic analytics* goes further and gives an idea not only of the events that occurred, but also of their causes. It answers the question "Why something happened?";

- *predictive analytics* answers the question "What is likely to happen?". Historical data is combined with rules, algorithms and external data in order to determine the future value or the probability of an event;

- *prescriptive analytics* is the next stage in predicting future events, and offers a sequence of actions to gain most from predictions and shows the consequence of each decision. It answers the question "What should I do?".

Predictive analytics is defined in [6] as a variety of statistical techniques from predictive modelling, machine learning, and data mining that analyze current and historical facts to make predictions about future or otherwise unknown events. As a rule, big data arrays are used in the process of analysis. The main idea of predictive analytics is to determine one or more parameters that affect the predicted event. The process of predictive analysis can be represented as follows:

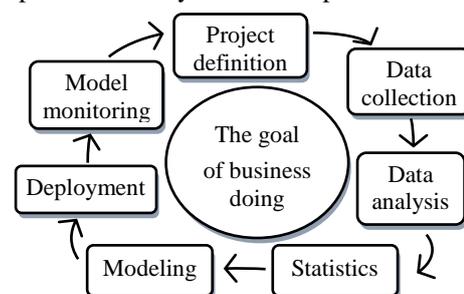


Fig.1. Predictive analytics process

Project definition. Definition of project results, components, scale of the work, business purpose, data set to be used.

Data collection. With the use of intelligent data analysis, data from different sources is prepared.

Data analysis. The process of data checking, clearing and modeling in order to identify useful information is performing.

Statistical analysis allows to confirm assumptions, hypotheses using standard statistical models.

Predictive modeling provides the ability to automatically build accurate predictive models.

Deployment of a predictive model provides the use of analytical results in the decision making process for obtaining reports.

Model monitoring. Models are tested to ensure the expected results.

The result of predictive analytics applying consists in the most effective business solutions making. An important requirement for a predictive model is to be as fit as possible and to be statistically significant. The predictive models may be [7]:

- *classification models.* They describe set of rules, according to which a new object can be assigned to the relevant class;

- *time series models.* They describe the functions that allow prediction of continuous numerical parameters and are based on information on the change of a certain parameter over the past time period.

According to Transparency Market Research [8], the market for predictive analytics will reach \$ 6.5 billion by 2019, while it was \$ 3.6 billion in 2015. The global market for predictive analysis systems will grow by an average of 17.8% annually. And as experience shows, the companies survive, that continue to invest in technology and innovation in the difficult economic times. And predictive analytics, of course, is one of such technology.

III. SOFTWARE TOOLS FOR PREDICTIVE ANALYTICS

Forrester Research has published in 2013 a report "Big Data Predictive Analytics Solutions, Q1 2013" in which market leaders for predictive analytics are contained [9]. According to it, SAS and IBM SPSS have the strongest position in the market and the best strategies among the largest developers of predictive analytics tools. The evaluation was carried out for 51 parameters - from the completeness of the functionality for the main analytical system to the size of the client base and the architectural advantages offered by the solutions developers.

SAS (Statistical Analysis System) *Enterprise Miner* [10] is leading in the segment of in-depth analytics, accounting for about a third of the market. It allows users to explore and analyze large amounts of data, to find patterns of relationships and to make well-informed decisions, based on facts and findings. Areas of effective use of the solution: banking sector, healthcare, oil and gas sector, insurance companies, telecommunications, transport, power system.

The main advantages of SAS Enterprise Miner include:

- advanced predictive modeling;

- convenient and clear interface allowing users to create predictive models on their own;

- automated process of routine application of models;

- possibility of batch processing;

- rapid data collection and preparation, aggregation and analysis;

- scalability and customization of the solution;

- high system performance when working with big data.

IBM SPSS (Statistical Package for the Social Sciences) [11] is a widespread intelligent tool for predictive analytics. SPSS's predictive analytics helps you analyze the patterns in historical and current transactions to predict potential future events.

A key component of the toolkit is *SPSS Modeler*, software environment for data mind allowing you to create intelligent predictive solutions by revealing the data patterns and relationships. *SPSS Modeler Server* supports integration with data mind and modelling tools provided by DBMS (database management system) developers, including IBM Pure Data System for Analytics. Using the SPSS Modeler, one can build and store models in the database. One can combine the analytical capabilities and ease of use of SPSS Modeler with the power and performance of the DBMS, using the built-in algorithms supplied by their developers. The models are built inside databases and are available for use with the convenient user interface of SPSS Modeler.

Dell Statistica (from 2017 -Tibco Software) in-depth data analysis platform [12] focuses on data professionals and organization needing to data process from a large number of IOT devices and heterogeneous sources. The functionality of the toolkit will help to prepare structured and unstructured data, deploy analytical tools on devices regardless of their location and use internal analysis functions on the MYSQL, Oracle, and Teradata platforms.

With Dell Statistica, companies are able to cope with the lack of data analysts and the complexity of today's IOT environments, as well as take into account new sources and data types.

Dell Statistica's features, simplifying predictive analytics, are as follows:

- dashboards with advanced visualization allowing users to easily see the results of the analysis at any stage;

- state-of-the-art web interface allowing users to share reports that can be opened in any browser;

- effective control of data, entered manually.

In addition to the represented (far from complete) developers of predictive analytics, there are also a large number of specialized firms providing business intelligence services. One of the most famous is Elder Research [13]. It has extensive experience in using many software tools (including all the above) for developing analytical solutions, programming, and personalized data visualization.

IV. GMDH-BASED PREDICTIVE ANALYTICS TOOLS

Among the various tools for predictive analytics, it should be emphasized several ones, the common feature of which is using of one of the most effective inductive modeling methods – Group Method of Data Handling (GMDH) [14].

Software tool *Insights* [15] is developed by German company *KnowledgeMiner Software* (created in 1993). Besides GMDH, it also uses Similar Patterns self-organizing modeling technology (also known as Analog Complexing) and fuzzy logic for modelling and prediction. It is possible to build linear and nonlinear, static and dynamic time series models, multi-input and one output models, many inputs and many outputs models. The outputs of the model can be represented both in analytical form (in the form of equations with estimated parameters) and graphically (using a system graph, which reflects the interconnections of the system structure).

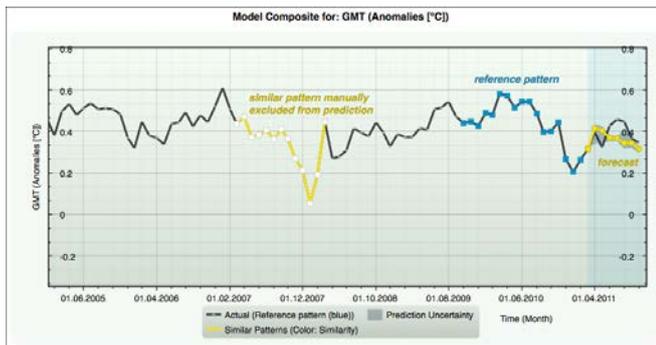


Fig.2. An example of Insights using for predictive analysis

Insights implements vector processing, multi-core and multiprocessor support for high-performance computing. It is scaled to the Apple Macintosh computer hardware. Regardless of which processor is used (dual-core or two six-core), the software automatically uses all the features of the PC.

GMDH Shell [16] is a contemporary software tool for predictive analytics. It is based on the classical GMDH algorithm and can be used for time series prediction, solving classification and clustering problems. GMDH Shell is a powerful solution for analyzing multidimensional data from various business fields. The software tool offers data mining algorithms – self-organized neural networks and combinatorial structural optimization of models. There is also the possibility of high-performance computing using a Linux-cluster.

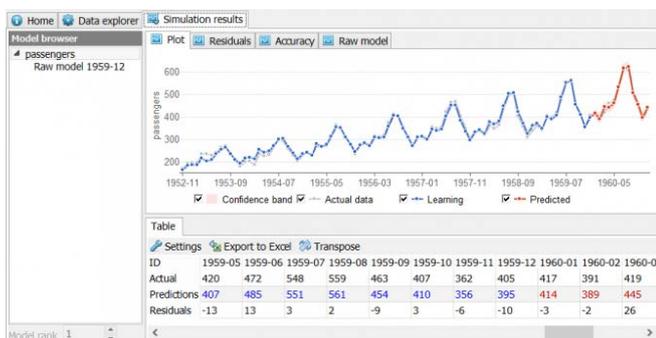


Fig.3. An example of GMDH Shell using for predictive analysis

It should be noted that GMDH Shell does not compete with KnowledgeMiner Insights in the sense that it is intended for use on the Windows operating system.

Software tool for modeling and prediction of complex multidimensional interrelated processes is developed in the

Department for information technologies of inductive modelling [17].

The tool is implemented for use on multiprocessor cluster systems. However, it can be embedded in any contemporary business intelligence system as an analytical tool for modeling and prediction of the dynamics processes in digital economy systems based on the detection and use of knowledge about the behavior and performance of such systems.

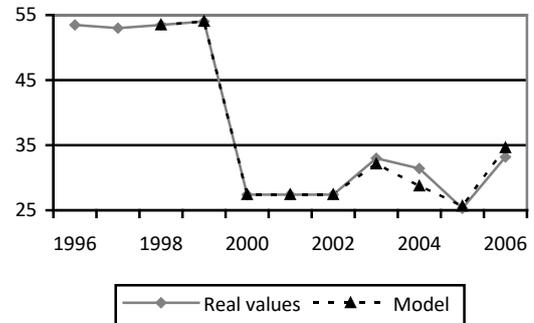


Fig.4. An example of software tool with recurrent-and-parallel GMDH algorithms using for predictive analysis

V. ON CONSTRUCTION OF ADVANCED GMDH-BASED PREDICTIVE ANALYTICS TOOL

Given that there is a considerable amount of predictive analytics tools, not fulfilling the whole range of problems, it may be concluded that there is still no single convenient solution on the market. To create accurate models and to obtain adequate predictions of various indicators, an advanced predictive analytics tool is required allowing to comprehensively reflect the relationships in models, being accessible and convenient for the user, allowing the user to customize the model and build reliable predictions. However, to date there is no such a complete solution. Developing of such a tool is an actual problem in the field of analytical business solutions.

The most important features of this advanced predictive analytics tool are:

- GMDH-based software tool [18, 19];
- recurrent-and-parallel computing [20];
- intelligent user interface [21];

GMDH-based software tool. The user does not need to have a thorough knowledge of the modeling principles when building models. He will be able to model with a convenient tool, knowing only the features of his domain. Built-in intelligent algorithms allow one to automatically build models on the available data set, which greatly facilitates the user's work.

Advanced predictive analytics tool, being constructed in the paper, is based on software for modelling and prediction of complex multidimensional interrelated processes in the class of vector autoregressive models.

Recurrent-and-parallel computing. Fundamentally new data-based solution for inductive modeling of complex processes has a high level of performance because of new concept, combining the efficiency of recurrent and parallel computing. The implementation of such solution provides

significant enhancing of efficiency and validity of managerial decisions.

Intelligent user interface. It is very important that predictive analytics tools are either too complicated for users or do not contain the necessary range of options. The intelligent user interface should be friendly and should allow building models without deep programming knowledge, which will significantly expand the range of users and increase their confidence in BI applications.

Advanced predictive analytics tool must include an intelligent shell allowing user (with any level of qualification) help to solve the data-based modelling problem (from data preprocessing to modelling algorithm choice). The intelligent shell provides the general use of automatic analysis and modeling procedures. It takes into account the user's wishes and a priori knowledge about the modeling object, and also provides decisions making control at every step of the modeling problem solving.

VI. CONCLUSION

Contemporary capabilities and advanced techniques of predictive analytics are becoming powerful way for increasing the company's productive efficiency. Predictive analytics is a new trend opening up broad prospects for the further development of companies.

Applying predictive analytics systems one should understand that the work of such systems is impossible without sufficient historical data and ineffective without the collection of current data. The less data will be used, the less accurate are predicted values.

The effectiveness of applying predictive analytics tools depends on both technologies used and the quality of such tools. And the advantage here will be on the side of the solutions, providing advanced methods of data mining. Such ones are just knowledge-oriented intelligent modeling software tools based on GMDH.

REFERENCES

- [1] https://www.accenture.com/t20160314T114937__w_/us-en/_acnmedia/Accenture/Omobono/TechnologyVision/pdf/Technology-Trends-Technology-Vision-2016.PDF.P. B. Johns, "A symmetrical condensed node for the TLM method," *IEEE Trans. Microwave Theory Tech.*, vol. MTT-35, pp.370-377, Apr. 1997.
- [2] https://en.wikipedia.org/wiki/Business_intelligence.
- [3] M. Goebel and L. Gruenwald, "A survey of data mining and knowledge discovery software tools", Volume 1, Issue 1 (June 1999), *Publisher ACM New York, NY, USA*.
- [4] https://en.wikipedia.org/wiki/Business_analytics.
- [5] <https://www.gartner.com/it-glossary/diagnostic-analytics>.
- [6] https://en.wikipedia.org/wiki/Predictive_analytics.
- [7] <http://www.globalcio.ru/workshops/968>.
- [8] <https://www.transparencymarketresearch.com/pressrelease/predictive-analytics-industry.htm>.
- [9] <https://www.forrester.com/report/The+Forrester+Wave+Big+Data+Predictive+Analytics+Solutions+Q1+2013/-/E-RES85601>.
- [10] https://www.sas.com/ru_ua/software/enterprise-miner.html.
- [11] <https://www.ibm.com/analytics/data-science/predictive-analytics/spss-statistical-software>.
- [12] <https://www.tibco.com/products/tibco-statistica>.
- [13] <https://www.elderresearch.com>.
- [14] Stepashko V. "Developments and Prospects of GMDH-Based Inductive Modeling" In: *Advances in Intelligent Systems and Computing II: Selected Papers from the International Conference on Computer Science and Information Technologies, CSIT 2017*, September 5-8, Lviv, Ukraine. N. Shakhovska, V. Stepashko Editors. *AISC book series, Volume 689*. – Cham: Springer, 2017, pp. 474-491.
- [15] <https://www.knowledgeminer.eu>.
- [16] <https://gmdhsoftware.com>.
- [17] <http://www.mgua.irtc.org.ua>.
- [18] Yefimenko S. "Building Vector Autoregressive Models Using COMBI GMDH with Recurrent-and-Parallel Computations" In: *Advances in Intelligent Systems and Computing II: Selected Papers from the International Conference on Computer Science and Information Technologies, CSIT 2017*, September 5-8, Lviv, Ukraine". N. Shakhovska, V. Stepashko Editors. *AISC book series, Volume 689*, Cham: Springer, 2017, pp. 601-613.
- [19] Stepashko V.S. and Yefimenko S.M. "Technologies of Numerical Investigation and Applying of Data-Based Modeling Methods" *Proceedings of the 2nd International Conference on Inductive Modelling ICIM 2008*, Kyiv, 2008, pp. 236-240.
- [20] Serhiy Yefimenko, Volodymyr Stepashko "Intelligent Recurrent-and-Parallel Computing for Solving Inductive Modeling Problems" *Proceedings of 16th International Conference on Computational Problems of Electrical Engineering (CPEE-2015)*, September 2-5, 2015, Lviv, Ukraine, 2015, pp. 236-238.
- [21] Stepashko V.S., Zvorygina T.F., Yefimenko S.M. "Problem of decision making intellectualization in tasks of models identification" (Problema intelektualizatsii pryiniattia rishen' u zadachakh identyfikatsii modelei), *Proceedings of ISDMIT-2005 Conference*, Kherson, Ukraine, 2005, Vol. 1, pp. 127-131.