System Representation of Technology Data Mining

Irina A. Brusakova
St. Petersburg Electrotechnical University "LETI"
St. Petersburg, Russia
brusakovai@mail.ru

Vladimir D. Chertovskoy
St. Petersburg Electrotechnical University "LETI"
St. Petersburg, Russia
vdchertows@mail.ru

***Abstract*— More and more wide circulation of technology of the data mining (DM) is marked. Need for streamlining of a large amount of the published materials is shown. The systematization option by means of systems analysis is offered and executed, what simplifies procedures of creation (synthesis) and a research of systems on the basis of IAD technology. One of synthesis algorithms is considered.**

Keywords*— technology system representation; data mining; streamlining

I. INTRODUCTION

The technology of data mining is widely adopted more and more. Now on data mining there is the significant amount of the publications covering different aspects of data domain. There is a set of implementation methods of technology that complicates understanding and digestion of material, the analysis and synthesis of systems using this technology.

Systematization, system reviewing of the materials which are saved up in this is necessary for the data domain which just began to develop.

This problem is solved in the real operation. The aim of system formation of theoretical and application-oriented knowledge for the subsequent creation (synthesis) and a research of difficult management systems is pursued by organizational objects. Owing to the initial stage of development of the considered class of technologies the important place is led out to informal methods process descriptions.

II. PROBLEM DEFINITION

It is known that the data mining (DM) – the technology of a research of data using methods of an artificial intelligence and oriented on giving to the system of properties of an artificial intelligence. This technology call also Data Mining which represents actually use of the existing methods of intellectual systems.

For systematization of the numerous formal methods [1–5] inherent in IAD technology, we use a method of systems analysis.

III. DECISION OF THE TASK

System streamlining of methods according to the diagram "intellectual systems-local technologies-the IAD procedures" is offered. This diagram (Fig. 1) defines stages of technology of creation of systems with IAD.

At the first stage the purpose of creation of system on the basis of which the choice of one of intellectual systems is carried out is created: multiagentny systems [2], intellectual management systems, expert systems, artificial neural networks [1], systems with genetic algorithms [4], systems in a natural language. The choice is defined by opportunities of systems. We will give their assignment.

Multiagentny system – set of interdependent agents both program, and the hardware capable to interact with each other and the environment, having certain mental abilities and potential of personal and combined actions. Intellectual management system – the system working in the mode of target adaptation. Expert system – the computing system using knowledge of the expert, the procedure of a logical output and allowing to offer explanations to the received results. Systems in naturally language – the search and algorithmic engines working with text information. Artificial neural networks – the systems imitating properties of neuron.

On the second step (Fig. 2) determination of the applied local technology is made [1]: real time, analysis of the text; extraction of knowledge from Web; visual analysis.

Technologies of real time is addition to set of statistical techniques. They use technologies of the accumulated training (new are permanently added to analysable data) in case of prediction perhaps using back coupling.

The technology of the analysis of the text allows to reveal new, useful templates in the unstructured data obtained from databases. At the same time preliminary text processing is required.

Technology extraction of knowledge from Web, being actually variety of technology of the analysis of specific text information. In case of a large number of information on the Internet normal search is connected to serious difficulties. Treat them: small percent of the necessary information; rare recurrence of calls; personalisation and setup of system on the specific user. The technology of obtaining information in this case includes search of resources, extraction of information on keywords, generalization and the analysis of results. Extraction of information are carried out with an involvement of the person.

The visual analysis is carried out during creation practically of any system with use of IAD. The purpose of the visual analysis to data representation in shape, the user, more evident and convenient for operation. The matter is that in case of a large number of data new knowledge can not always be found the simplest means of a graphics (decision trees, datagrams,

978-1-5386-5824-6/18/$31.00 ©2018 IEEE

two-dimensional diagrams) and more difficult means of display are required. The conversational mode of operation of the user which gives the chance to work with the non-uniform and strongly noisy data is necessary, without applying very difficult algorithms of the decision. At the same time data processing rate and confidentiality of results increases.

At the third stage the IAD specific procedures are selected: classification, clustering, separation of the sequences and associations, forecast. Each of procedures can be executed by different mathematical methods. Their Classification can be such.

Forecast
- Functions of a matrix argument
- Multivariate normal law
- Optimum mean squared linear forecast
- Simulation of multidimensional accidental data

Visualization
- Principal components and factor analysis
- Discriminant analysis
- Non-linear principal components
- Visual data analysis – Visual Mining

Classification and cluster
- Stochastic distances and measures of closeness
- Cluster analysis
- Naive Bayes method
- Decision trees
- Machines of reference vectors (Support Vector Machines)
- The adaptive methods of a clustering

Sequences
- Nonparametric methods
- Parametric models
- Prediction
- Statistical investigation of dependences
- Least-squares method
- Orthogonal regression
- Multiple regression
- Non-linear methods
- Data handling with passes

Associations
- Search of the associative rules

Digitization of not quantitative data
- Expert methods
- Method of the analysis of hierarchies Saati
- Metric and nonmetric scaling

Analysis of text information
- Extraction of key concepts
- Classification and cluster
- Annotation

The special class is made by specific implementation methods of IAD.

Development of the concept of databases
- Decision making support systems
- Data stores
- OLAP-systems
- Data mining standard

IV. EXAMPLE OF USE OF IAD TECHNOLOGY

Most often the IAD technology is illustrated using the multiagentny of systems [1] Further we will consider use for the called purposes of intellectual management systems [2].

We will consider the automated control system for production consisting of two subsystems: technical and economic planning and operational control of the main production. The system has three-level structure of the controlling part: "the principal – the dispatcher – chiefs of shops". For the mathematical description of such system [2] homogeneous method constructed on the basis of the task of the dynamic linear programming (DLP) are offered and approved. The task DLP is made of widely known task of static linear programming and the system of the difference equations. On the basis of the mathematical description the mathematical model working in the optimum mode and allowing to carry out process of planning both in traditional option is built and upon operational transition to release of new production.

The last property allows to study the procedure of prediction on continuous intervals of time according to the diagram "what will be if?", i.e. to use IAD technology on the basis of intellectual control system. As local technology the technology of real time, and as procedures – prediction is selected.

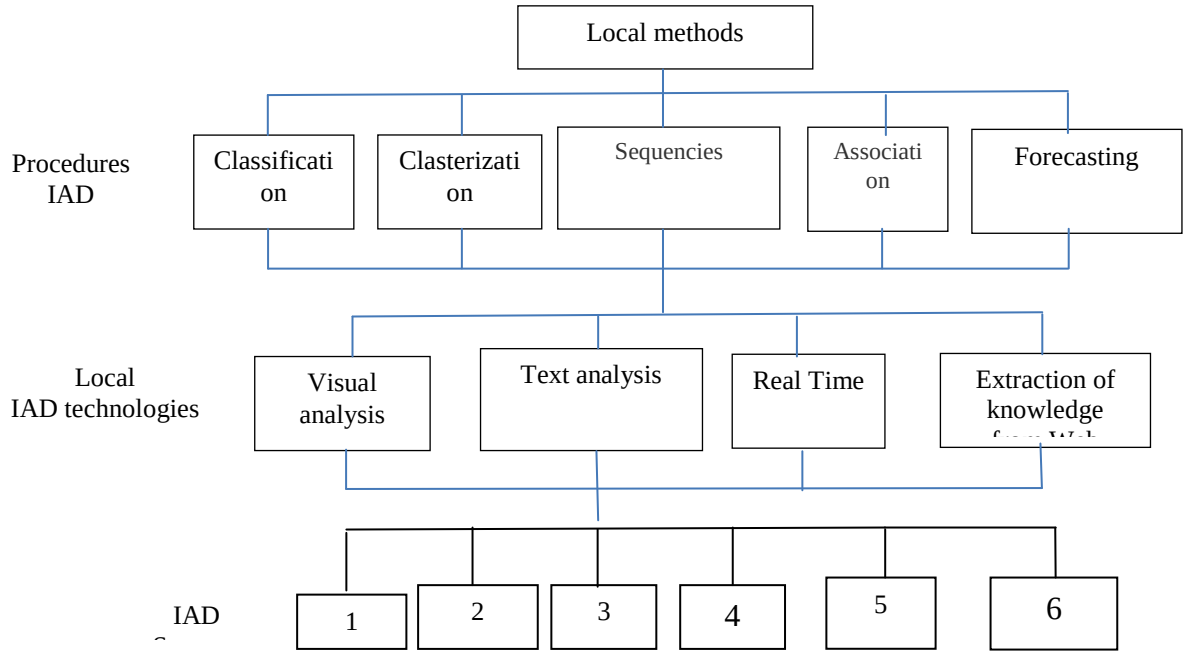

Рис. 1. System of components of IAD technology: 1 - SGA to system with genetic algorithms; 2 - An IAU to multigoagentny systems; 3 - SOWING to system in a natural language; 4 - ES to expert systems; 5 - INS to artificial neural networks; 6 - ISU to intellectual management systems

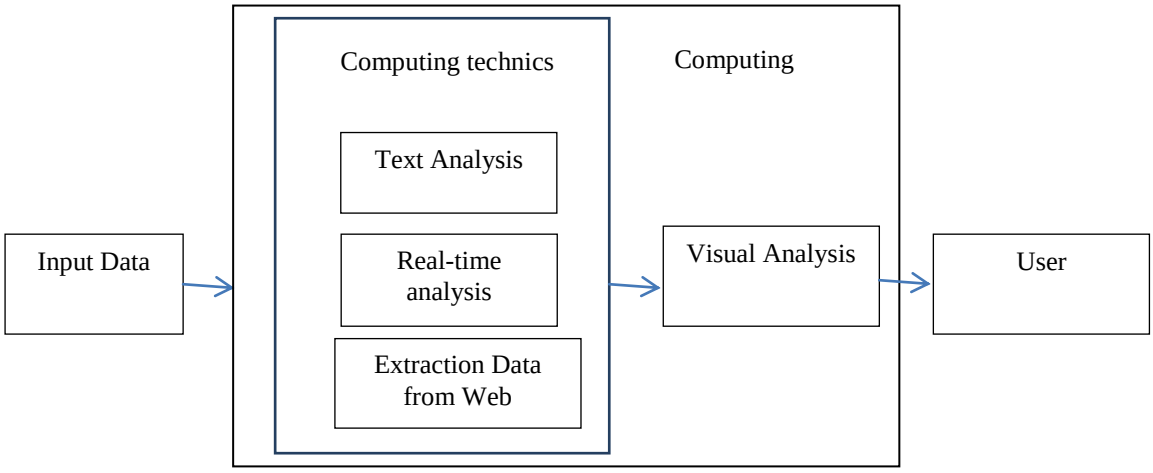

Рис. 2. Diagram of use of the IAD local technologies

V. CONCLUSION

Now the most developed classes of systems are the IAD systems on a basis the multiagently of systems, intellectual management systems, partly – expert systems. At the same time use of technology on artificial neural networks, systems with genetic algorithms, systems in a natural language is of interest.

Further development of theoretical aspects of data mining is connected as to enhancement of propositional logic, and the logician, knowledge, more close to psychology of framing, by the person.

REFERENCES

[1] Barsegyan A.A. Data analysis and processes: studies. manual /A.A. Barsegyan, M.S. Kupriyanov, I.I. Holod, M.D. Tess, S.I. Yelizarov. SPb.: BHV, 2009. 512 pages.

[2] Sovetov B.Ya., Tsekhanovsky V.V., Chertowskoy V. D. Intellectual systems and technologies. M.: Academy, 2013. 320 pages.

[3] Shabanov V.I., Andreyev A.M. Method of classification of text documents based on full-text query search//Works of the first Russian seminar according to methods of information search. Under the editorship of I.S. Nekrestyanov. SPb.: Scientific Research Institute Himii St.Petersburg State University, 2003.

[4] Mkhitaryan V.S. Data analysis. M.: Jurait, 2016.

[5] Olifer N., Olifer A. Network operating systems. SPb.: St. Petersburg, 2001. 525 pages.