

PREDICTION IN BIG DATA TECHNOLOGY



S. BOROVNIKOV, PhD
*Associate Professor,
Department of Informa-
tion and Computer
Systems Design
BSUIR*



E. SHNEIDEROV
M.Eng.
*Department of Informa-
tion and Computer
Systems Design
BSUIR*



M.I. TSYRELCHUK
*Student of the 5th
course of Faculty of
Radioengineering and
Electronics BSUIR*



S.S. DZIK
*Student of the 4th
course of Faculty of
Engineering and Eco-
nomics BSUIR*

*The Belarusian State University of Informatics and Radioelectronics, Republic of Belarus
E-mail: bsm@bsuir.by, shneiderov@bsuir.by*

Abstract. Authors offer the variant (scheme) of the classification of forecasting methods that can be demanded in the design of analytical systems that use Big Data. Authors consider in article a brief explanation of the forecasting methods and indicate their possible use.

One of the tasks for large data processing technology is the creation of different analytical reports and obtaining prediction, which will be used by companies and organizations (hereinafter companies) in their activities. Efficiency of activity of product manufacturing companies and provision of services (commercial, educational, health, etc.) can be described by a target variable, one or more. In many cases of predicting tasks with the use of Big Data target variable can be predicted on the basis of a set of features. Moreover this set should be exhaustive in terms of predicting the target variable with high reliability [1].

For the successful design of analytical systems that use big data analysts must have a clear idea of predicting as a subject area of scientific prediction of the behavior of the business and other processes in the future. In some cases, the absence of a clear classification of predicting methods may cause the complexity of designing of analytical systems and create problems with the choice of application software (packages) for them, which allow high efficiency of predicted values of the target variable, based on big data. Therefore, the classification of predicting methods is a topical and important task.

To predict - means to forecast the future state of the object or process behavior based on the consideration of factors and events, associated with the object or process in some way.

The authors offer one of classification schemes of predicting methods, which in one degree or another may be in demand for the design of analytical systems based on Big Data (Figure 1).

During the heuristic prediction, the prognosis is given by an expert analyst on the basis of subjective weighting of a combination of factors, most of which may be qualitative in nature. In this case the result of the forecast depends largely on experience and even intuition of the expert who authorize the prognosis. Used quantitative and qualitative information about the country of interest (or group) is usually prepared on the basis of data (Big Data), which was taken from different sources. For more information about the method (in relation to technology, but easy to carry on the activities of companies) see [2].

During mathematical prediction to form results (forecast) experts use the information about business or other process received by data taken from different sources. This forecast is obtained by the subsequent processing of this information by means of formal, pre-selected mathematical methods. The forecast depends largely on the information about a business or other process, which is obtained on the basis of Big Data, as well as on the selected mathematical methods of data processing (analysis software tools).

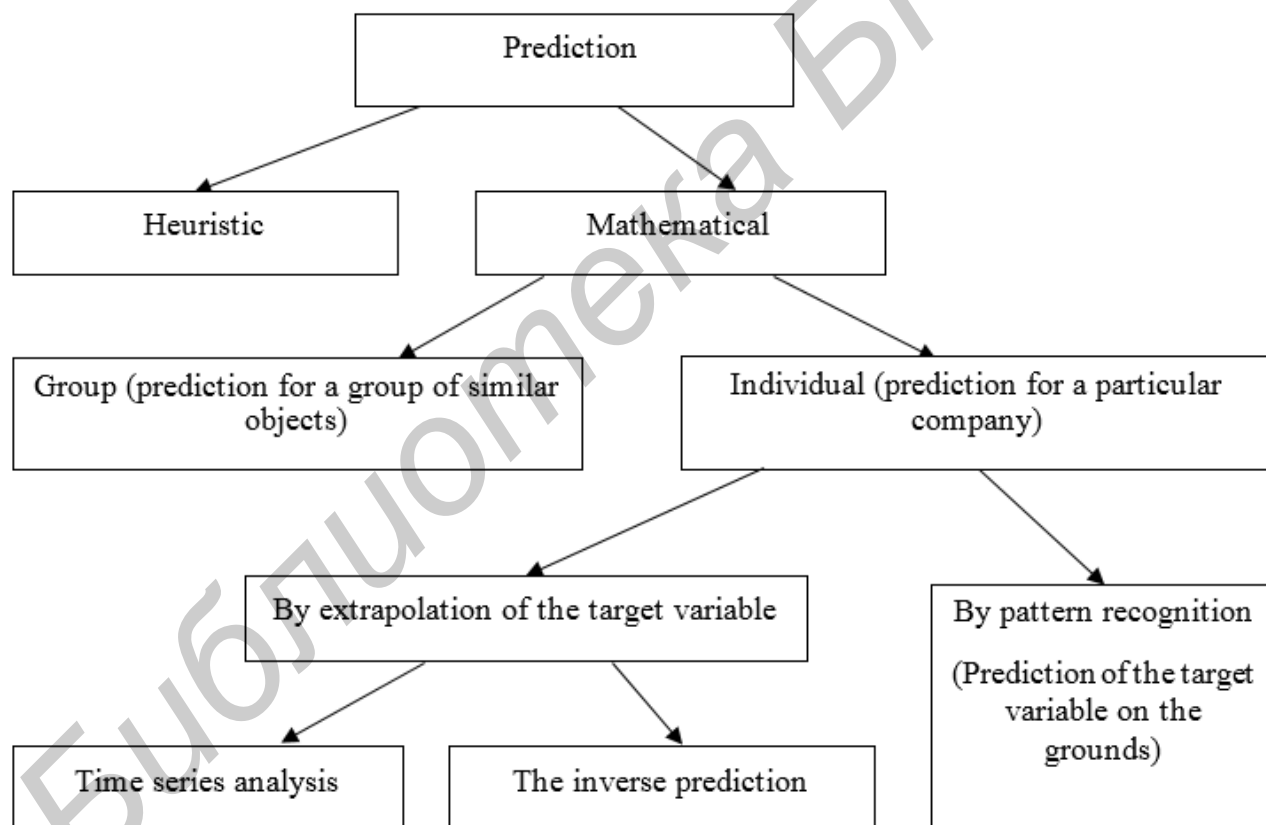


Fig. 1. A generalized scheme of classification of prediction methods, demanded in the Big Data technology

Prediction for a group of objects of business or other fields of activity (group prediction) of the same type - it is a prediction, in which the forecast is equally applicable to any object of this group of similar objects, such as shops of «Euroopt»-network. In practice, the result of such prediction can be obtained in the form of the

probability that the target variable of interest (for example, sales volume) at a given future point in time will be not less than a certain value.

When individually prediction is used, data about the activities of a particular company, obtained from different sources, and based on the processing of these data give a forecast for the value of the target variable for future points in time in relation to the same company. At the same time, values of target variable found at different times or features, which determine the values of the target variable for the past and future points in time, are used for the forecast. During prediction method of extrapolation of target variable, we use directly the values of target variable found at different times in the past on the basis of the use of data from different sources. Time series analysis can be considered as a practical method of extrapolation of the target variable in order to find its predictive value, which has a high reliability for the future points of interest in time.

Table 1. Possible applications of prediction methods in Big Data

Prediction method (see. Figure 1)	Possible applications
1. Heuristic	1. Evaluation of the importance (valuables) of big data obtained from different sources. 2. Evaluation of reliability of used big data obtained from different sources.
2. Mathematical	Evaluation of predicted values of the target variable, which describes the activities of companies and business organizations, education, transportation, medicine, etc.
3. Group (prediction for a group of similar objects)	Cases where effect of corporate activities, communities as a whole (a network of shops, industrial associations, educational institutions of the same type, etc.) is a point of interest.
4. By extrapolation of the target variable (time series analysis)	Cases of direct evaluation of values of the target variable for the time points in the past and present times based on Big Data.
5. By extrapolation of the target variable (the inverse prediction)	The tasks that need to restructure the activities of companies and organizations, without bringing their work into losses (economic, technical, social, etc.).
6. Prediction of the target variable on the grounds	Estimation of the value of the target variable, one or more, describing the activities of companies and organizations in different areas of business, public and social spheres. This is the most typical tasks using Big Data.

The inverse prediction is the case of extrapolation of the target variable, when we

are interested not in the value of the variable for a given future point in time, but interested in the future point of time when the target variable for the company is going to get worse than the maximum permissible level.

Prediction of the target variable on the grounds (otherwise by pattern recognition) is based on the use of a set of features for a description of the target variable of the business process of the company. This feature sets correspond to the points of time in the past and the present. the prediction task is to predict the target values for the future points in time by the data obtained from different source, i.e., to get the grounds of the target variable, and then to give the forecast for its value for the point in time by the grounds of this variable.

Below the Table 1 shows the various possible applications of prediction methods that may be claimed in analytical systems that use Big Data.

Authors presented a possible classification of prediction methods and tried to explain their essence and indicate possible applications while creating analytical systems based on big data.in a brief but systematic manner. We hope that this article in some way will help analysts and IT-specialists in the creation of analytical systems using Big Data.

References

- [1]. Franks, B. Taming the big data. Finding Opportunities in Huge Data Streams with Advanced Analytics / B. Franks ; 2014. – p. 352
- [2]. Sergey M. Borovikov. Theoretical bases of design, technology and reliability: textbook for universities' engineering specialties / S.M. Borovikov. – Minsk : Design PRO, 1998. – p. 336