PREDICTION OF EXTREME EVENTS BASED ON HETEROGENEOUS MULTIDIMENSIONAL TIME SERIES AND STATEMENTS OF EXPERTS[1]

G.S. Lbov, M.K. Gerasimov, M.G. Tretyakova
Institute of Mathematics (SB RAS), Novosibirsk State Technical University
Novosibirsk, RUSSIA
e-mail: lbov@math.nsc.ru, max_post@ngs.ru, martretj@mail.ru

Abstract

In this paper we present an approach of extreme situations (events) prediction based on multiple heterogeneous time series analysis. By assumption, extreme situations are described by several variables of different types.

1 Introduction

Let the some phenomenon observations represented by measurements of variables X_1,...,X_j,...,X_n taken at consecutive time moments $t_1, \ldots, t_\nu, \ldots, t_R$. In this paper we suppose that N different realizations $a^1, \ldots, a^i, \ldots, a^N$ of this phenomenon were observed (we shall say that a^i is an *object*, $i = \overline{1, N}$). Hence, we have N different n-dimensional time series for analysis. For each object a^i at the moment t_{R+1} a collection of variables Y_1,...,Y_l,...,Y_m is specified.

In this paper, we assume that extreme situations (events) are determined by certain combinations of values of variables Y_1,...,Y_l,...,Y_m.

2 Problem's Statement

We begin with some notation. Let $x_j^i(t_\nu)$ be the value of variable X_j for the object a^i at the moment t_ν, $y_l^i = Y_l(a^i)$ the value of variable Y_l for the object a^i. By D_j and D_l denote the sets of possible values of the variables X_j, $j = \overline{1, n}$; and Y_l, $l = \overline{1, m}$; respectively. $D_X = \prod_{j=1}^{n} D_j$, $D_Y = \prod_{l=1}^{m} D_l$, $x^i(t_\nu) = \Big(x_1^i(t_\nu), \ldots, x_j^i(t_\nu), \ldots, x_n^i(t_\nu)\Big)$, $x^i = \Big(x^i(t_1), \ldots, x^i(t_\nu), \ldots, x^i(t_R)\Big)$, $y^i = \Big(y_1^i, \ldots, y_l^i, \ldots, y_m^i\Big)$.

In this paper we suppose that extreme situations are represented by some subset $E^* \subset D_Y$ (defined by experts), $D = \prod_{l=1}^{m} C_l$, where $C_l = [\alpha_l, \beta_l]$, if Y_l is a quantitative variable, and C_l is some finite list of names if Y_l is a nominal variable.

Consider a new object a^*. As above, the variables $X_1, \ldots, X_j, \ldots, X_n$ were measured at the moments $t_1, \ldots, t_\nu, \ldots, t_R$. The problem consists in estimation of extreme event's

[1]This research was supported by RFBR-07-01-00331a.

probability for object a^*. This problem has some distinctive features. First, the set $\overline{X} = \{X_1, \ldots, X_j, \ldots, X_n\}$ simultaneously may contain binary, symbolic, and numerical variables. Secondly, a priori, few extreme situations were observed. Thirdly, extreme situations depend on combination of values of variables $Y_1, \ldots, Y_l, \ldots, Y_m$, hence we have to predict the whole collection $Y_1, \ldots, Y_l, \ldots, Y_m$ simultaneously.

By assumption, the distribution $P\Big(x(t_1), \ldots, x(t_\nu), \ldots, x(t_R), y\Big)$ is unknown, hence we have to estimate it on the base of observed multidimensional time series.

We suggest in this paper a forecasting method using class of logical decision functions of heterogeneous variables. As shown in theoretical and experimental investigations [1], this class is the most appropriate under previous conditions.

3 Extreme Events Prediction: Particular Solution

Let us introduce the method of extreme events forecasting. We first consider each moment t_ν, $\nu = \overline{1, R}$, separately.

A learning sample $V_\nu = \{x_1^i(t_\nu), \ldots, x_j^i(t_\nu), \ldots, x_n^i(t_\nu), y_1^i, \ldots, y_l^i, \ldots, y_m^i\}$, $i = \overline{1, N}$; is formed.

In the paper [2] a method of recognition of the heterogeneous multivariate variable was suggested. Consider some modification of this method.

Let $f : D_X \to \Theta(D_Y)$, where $\Theta(D_Y) = 2^{D_Y}$ be the power set of D_Y, $f(x) = E_y(x)$, $E_y(x) \subseteq D_Y$. Consider $F(f) = \int_{D_X} \mu(E_y(x))dP(x)$, where $\mu(E_y(x))$ be the measure of the set $E_y(x)$.

A logical decision function f_ν is constructed by learning sample V_ν such that $f_\nu = \arg\min_f F(f)$. This function corresponds to couple $\Big\langle \alpha_\nu, r_\nu(\alpha_\nu) \Big\rangle$, where $\alpha_\nu = \{E_\nu^1, \ldots, E_\nu^\mu, \ldots, E_\nu^{M_\nu}\}$ be the partition of space D; $r_\nu(\alpha_\nu) = \{r_\nu^1, \ldots, r_\nu^\mu, \ldots, r_\nu^{M_\nu}\}$, where $r_\nu^\mu = |E_y^\mu(\alpha_\nu) \cap E^*| \ / \ |E_y^\mu(\alpha_\nu)|$ be the estimation of extreme event's probability if $x(t_\nu) \in E_\nu^\mu$, $E_y^\mu(\alpha_\nu)$ be the set prescribed to set E_ν^μ, $E_y^\mu(\alpha_\nu) \subseteq D_Y$.

Thus, we have R different decision rules f_ν, $\nu = \overline{1, R}$, and R different estimations of extreme event's probability to each $x = \Big(x(t_1), \ldots, x(t_\nu), \ldots, x(t_R)\Big)$.

4 Extreme Events Prediction: General Solution

Now we introduce the method of general forecasting of extreme situations.

A new variable Z_ν is considered. Denote by D_{Z_ν} the set of possible values of variable Z_ν. By definition, put $D_{Z_\nu} = \{1, \ldots, \mu, \ldots, M_\nu\}$, and $z_\nu^i = Z_\nu(a^i) = \mu$ if $x^i(t_\nu) \in E_\nu^\mu$.

A learning sample $V(Z) = \{z_1^i, \ldots, z_\nu^i, \ldots, z_R^i, s^i\}$, $i = \overline{1, N}$; is formed, where $s^i = 1$ if $y^i \in E^*$, $s^i = 0$ otherwise. We see that initial problem of multidimensional extreme situations prediction is reduced to problem of two pattern recognition.

A logical decision function $\overline{f(Z)}$ is constructed by one of the algorithm for pattern recognition (e. g., LRP [1]). This function corresponds to couple $\Big\langle \alpha_{(Z)}, r_{(Z)}(\alpha_{(Z)}) \Big\rangle$, where $\alpha_{(Z)} = \{E_{(Z)}^1, \ldots, E_{(Z)}^m, \ldots, E_{(Z)}^M\}$ is the partition of space $D_{(Z)} = \prod_{\nu=1}^{R} D_{Z_\nu}$;

$r_{(Z)}(\alpha_{(Z)}) = \{r^1_{(Z)}, \ldots, r^m_{(Z)}, \ldots, r^M_{(Z)}\}$, where $r^m_{(Z)}$ is the estimation of extreme event's probability if $z \in E^m_{(Z)}$, $z \in D_{(Z)}$.

Therefore, the function $\overline{f(Z)}$ take each object a^* and its observations $\Big(x^*(t_1), \ldots, x^*(t_\nu), \ldots, x^*(t_R)\Big)$ to the single set $E^m_{(Z)}$. Thus, to each object a^* we can evaluate the extreme event's probability $r^m_{(Z)}$ and select most probable pattern number.

5 Conclusion

We stress that usually extreme situations are rare events resulted from unique combinations of cause-effect relations of complex phenomenon. Hence, the amount of corresponding precedents in empirical information is small with respect to the general volume of observations. In order to increase reliability of prediction we have to use as much as possible complete (complex) empirical information about this phenomenon.

Note that logical-and-probabilistic models are constructed by suggested method of time series analysis. These models are represented as list of logical statements. Thus, we can simultaneously analyse these results and experts' statements (knowledge) represented in similar form. However, statements of different experts may be partially contradictory. In the work [3], a method of coordination of several experts' statements was suggested. Finally, analysing time series and experts' statements, we can estimate extreme event's probability.

References

[1] Lbov G.S., and Starceva N.G. (1999). *Logical Decision Functions and Problems of Statistical Stability of Decisions.* Institute of Mathematics, Novosibirsk.

[2] Stupina T.A. (2006). Recognition of the Heterogeneous Multivariate Variable. *Proceeding of the International Conference "Knowledge - Dialogue - Solution"*, Varna, Bulgaria. Vol. **1**, pp. 199-202.

[3] Lbov G.S., Gerasimov M.K. (2004). *Constructing of a Consensus of Several Expert Statements. Proc. of XII International Conference "Knowledge - Dialogue - Solution"*, Varna, Bulgaria. Vol. **1**, pp. 193-195.