Peer-reviewed & Open access journal
www.academicpublishingplatforms.com
The primary version of the journal is the on-line version

ATI - Applied Technologies & Innovations
Volume 8 | Issue 3| November 2012 |pp.45-55
http://dx.doi.org/10.15208/ati.2012.16

Methods and algorithms of selection the informative attributes in systems of adaptive data processing for analysis and forecasting

Olimjon Djumanov, Sunatillo Kholmonov
Faculty of Mathematics, Samarkand State University, Uzbekistan

corresponding e-mail: olimjondi@mail.ru
postal address: University blvd. 15, 104140, Samarkand, Uzbekistan

The principles, methods and algorithms of informative attributes selection were developed for optimization of description and representation for the objects in systems of adaptive data processing, where data are non-stationary by nature. The proposed algorithms of informative attributes selection for one-dimensional time series are based on the simplified ratings of correlation, mathematical expectation, dispersion of attributes. The algorithms have been developed using dynamic properties of information, i.e. control of conditional moments and parameters of regression model on confidential intervals. The methods and algorithms of adaptive data processing were tested on an example of extensive data represented for forecasting in systems of electro-supply.

Keywords: Non-stationary object, continuous information, informative attributes, representation of object, statistical parameters, dynamic properties, system of adaptive data processing, approximation, intellectual analysis, forecasting

Introduction

For today the actual problem is the construction of data processing systems on the basis of intellectual analysis technologies, particularly for decisions of pattern recognition, approximation and interpretation of time series, analysis and forecasting of non-stationary processes (Djumanov, 2008). The successful decision of these tasks requires development of methods, algorithms and tools of intellectualization for processing the data of various nature (Djumanov, 2008, 2011). Use of Data Mining concepts and methods represents the perspective technology for construction of systems of adaptive processing of continuous data. These methods are widely used in systems of artificial intelligence, recognition and classification of images, approximation of many variable functions, forecasting, optimization and management (Sankar and Sushmita, 1992; Haykin, 1996).

It is necessary to note, that in intellectual control systems the efficiency of preliminary data processing for representation of the sample trainees significantly depends on character of a studied object and process. Non-stationarity of processes is characterized by large uncertainty of the a priory information, limitation of statistics, difficulty of statistical parameters definition, tendentious change of parameters dynamics, complexity of mathematical description. All this proves the urgency of development of principles, methods and algorithms which adaptive process the data on the basis of use of statistical parameters and dynamic properties from information (Umbaugh, 1998).

The present work is dedicated to development of methods and algorithms for selection of informative attributes from dynamic objects for approximation, analysis and forecasting of processes, which describe, for example, change of production parameters of industrial-technical complex. The algorithms are developed for rating the statistical and information characteristics of data, for control of belonging the attributes on permitted borders, for classification of attributes by conditional probability characteristics. Proposed methods and algorithms are developed for description and representation of non-stationary objects.

Unconditional, conditional, average entropy; quantity of the information; the various factors of correlation can be used as criteria of informativeness rating in selection of informative attributes by informatic and statistical characteristics of data.

The following section of the paper describes results of algorithm development for selection of informative attributes by functions of correlation of times series.

Selection of informative attributes on the basis of simplified ratings of correlation functions

For obtaining the ratings of informative criteria we investigated the techniques of receiving coefficients of twin-, private-, auto-, multiple correlation and parameters of regression model for time series.

Let's show the technique for definition a rating of correlation functions. Let ζ and η be the random variables (RV), and $X_1, X_2, \ldots, X_i \ldots, X_m$ with $Y_1, Y_2, \ldots, Y_i \ldots, Y_n$ values of these RV, gotten accordingly from m and n independent tests. Then as a rating of correlation function $K_{m,n}(\zeta,\eta)$ we can use the known value

$$\tilde{K}_{m,n} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(X_i - \tilde{X})(Y_j - \tilde{Y}),$$

where, $\tilde{X} = \tilde{E}_m(X)$ and $\tilde{Y} = \tilde{E}_m(Y)$ are accordingly unbiased ζ and solvent η ratings of mathematical expectation determined by special technique.

For simplification and acceleration of calculation $\tilde{K}_{m,n}$ we offered the recurrent formula

$$\tilde{K}_{m,n} = \tilde{K}_{m-1,n} + 2^{-[Q_1(m)+Q_2(n)]}[(X_m - \tilde{X})(Y_n - \tilde{Y}) - \tilde{K}_{m-1,n}],$$

where, $Q_1(m)$ and $Q_2(n)$ are the integer positive numbers dependent accordingly from m and n.

It is accepted that the rating $\tilde{K}_{m,n}$ is unbiased, when

$$E(\tilde{K}_{m,n}) = K_{m,n}(\zeta,\eta),\ m,n = 1, 2, \ldots$$

The rating is accepted as consistent when $\tilde{K}_{m,0} \to 0$, $m \to \infty$, $n \to \infty$.

For consecutive rating of auto correlation function for RV ζ we offered following recurrent formula

$$\tilde{K}_{m,m} = \tilde{K}_{m-1,m} + 2^{-2Q(m)}[(X_i - \tilde{X})(X_j - \tilde{X}) - \tilde{K}_{m-1,m}].$$

Calculation of ratings for moments and central moments of l - th order RV ζ is carried out by formulas

$$\tilde{\mu}_m = \tilde{\mu}_{m-1} + 2^{-Q(m)}[X_i^l - \tilde{\mu}_{m-1}],$$
$$\tilde{\nu}_m = \tilde{\nu}_{m-1} + 2^{-Q(m)}[(X_i - \tilde{X})^l - \tilde{\nu}_{m-1}].$$

For rating of reliability of getting theoretical techniques we carried out the experiment on 100 observations at 14 statistical samples of electro supply technical parameters. It is established, that values of mathematical expectation, dispersion, function of correlation and auto-correlations in comparison with values of same parameters, calculated on the classical formulas, give the error about 10-15 % with guarantee probability 0.9. However, the counting period is accelerated by 10-15 times and the size of used sample is reduced by 5-6 times.

Now let's state a principle of informative attributes selection by function of correlation for multidimensional time series. Only attributes which have significant coefficients correlation is chosen from space of object attributes. Hereinafter the attributes which have least coefficients of pair correlation are selected among chosen attributes. Further, the coefficients of linear regression and coefficients of multiple correlation are counted by set of selected attributes. Coefficient of multiple correlation is used for a rating of informative attributes. Statistical parameters of data of getting subsets (A) are compared with attribute of initial subset (B).

The measure of informativeness for set of attributes is estimated as:

$$J = 0.5tr[(v^A - v^B)] + 0.5tr[(v^{A^{-1}} + v^{B^{-1}})(\bar{x}^A - \bar{x}^B)(\bar{x}^A - \bar{x}^B)^T],$$

where, tr is the sum of diagonal elements from a matrix of correlation coefficients; J_j is private measure of x_j-th attribute informativeness; $\bar{x}_j$ is average value of x_j-th attributes; σ_{x_j} is mean-squared error of x_j-th attributes.

For calculation the coefficient of compression v by average values the covariance matrix V is defined as

$$V = M[(x_k - \bar{x}_k)(x_j - \bar{x}_j)],$$

where, M is operator of mathematic expectation, V^{-1} is matrix opposite to V; $\overline{X} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_j, \ldots, \bar{x}_n)$ is vector of average values of attributes; T is symbol of transpose.

We used the calculated values of coefficient v from set of n attributes, where $(1 \le v < n)$, for a finding admitted informative sample, which is allocated from set of $n-1, n-2, \ldots, 3, 2, 1$ attributes, where for each set informativeness of included in it attributes is calculated separately.

In the next section of paper we discuss results of development of algorithms which select the informative attributes of non-stationary object, where the simplified ratings of mathematical expectation, dispersion, functions of times series distribution are accepted as statistical parameters.

Selection of informative attributes on the basis of control of statistical parameters belonging to permitted values borders

Let attributes of non-stationary object represent random variables $\alpha \in \Omega$ with function of probabilities density distribution $\omega(\alpha)$ given in area Ω, with dispersion σ_α^2 and mathematical expectation a_α. RV α_i, wrongly transformed when information processing, we represent through β_j.

For selection of informative attributes by mathematical expectation we suggested the rule which control belonging of attributes in limits of the bottom $a_\alpha - x$ and top $a_\alpha + y$ thresholds. The thresholds divide set of attributes values $\{\beta_j\}$ on a subset of permitted $\{\beta_p\}:\{a_\alpha - x \le \beta_p \le a_\alpha + y\}$ and forbidden $\{\beta_f\}:\{$ $\alpha_{\min} \le \beta_f < a_\alpha - x$; $a_\alpha + y \le \beta_f < \alpha_{\max}$ $\}$ values.

At the control, the attribute is considered informative if β_j belongs to a subset of permitted values $\{\beta_p\}$ and non-informative if β_j belongs to a subset of the forbidden values $\{\beta_f\}$.

The algorithm of selection of informative attributes leaves undetected errors of two kinds. The errors of first kind P_1 ("passing of errors") arise when non-informative values gets in a subset of permitted values $\{\beta_p\}$ with probability

$$\Re\{\alpha_\mathrm{i} \ne \beta_j;\ \ \beta_j \in \{\beta_p\}\}.$$

The errors of second kind P_2 ("false alarm") arise, when informative value of attribute is outside the subset of permitted values with probability

$$\Re\{\alpha_\mathrm{i} = \beta_j;\ \ \beta_i \in \{\beta_f\}\}.$$

We accepted a hypothesis about equal probabilities for transformations $\alpha_i \to \beta_j$, proceeding from what probability of undetected errors of first kind we wrote down as:

$$P_1 = \Re\{\alpha_\mathrm{i} \ne \beta_j;\ \ \beta_j \in \{\beta_p\}\} = P\frac{x+y}{B-\delta} \tag{5}$$

and accordingly for the second kind:

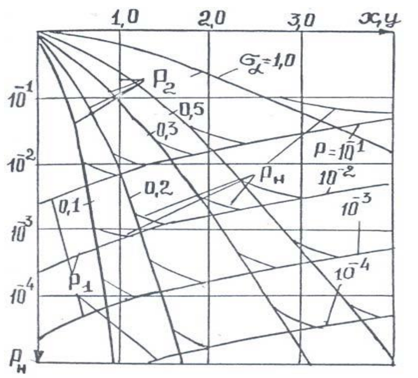
$$P_2 = \Re\{\alpha_i \neq \beta_i;\ \beta_i \in \{\beta_f\}\} = (1-P)\left[1 - \sum_{\alpha_i = a_{\alpha_i} - x/\delta}^{a_{\alpha_i} + y/\delta} f(\alpha_i)\right]. \qquad (6)$$

Changing discrete values to continuous sizes essentially simplifies the further calculations and makes results more evident and convenient for practical application. After such changing, the formula of undetected errors probability is written down as:

$$P_H = P\frac{x+y}{B} + (1-P)\left[1 - \int_{a_\alpha - x}^{a_\alpha + y} w(\alpha)d\alpha\right] \qquad (7)$$

In Figure 1 the diagrams illustrate depending of errors of first P_1 and second P_2 kind on borders of information control (x, y) when for law of distribution is normal for RV α probabilities distribution function, probability of errors is $P = 10^{-3} \div 10^{-5}$ and dispersion is $\sigma_\alpha = 0.2$.

FIGURE 1. GRAPHS OF ERRORS OF THE FIRST AND SECOND KIND



As diagrams show, regulation of the control thresholds narrows or expands a range of values for probabilities of errors of the first and second kinds. Hence, is more optimal the border of information control, the less probability of undetected errors.

It is necessary to note, that the research of efficiency for offered algorithm can be carried out also by criterion of minimal mean-squared error of informative attributes selection. Such criterion allows to take into account a degree of data variations about average value and allow to use principles of smoothing of non-informative attributes by reduction to dispersion or to control border values or to average values.

For selection of informative attributes on the basis of control by dispersion we developed the algorithm $A^0_{\sigma_n}$ for conditions, when the control borders are limited to a range of practical scope of RV β_j.

Algorithm $A^0_{\sigma_n}$. Let's designate a range of practical scope of RV β_j as $\alpha_{\min}$ and $\alpha_{\max}$, and their difference as $\alpha_{\max} - \alpha_{\min} = B$. Theoretically $\alpha_{\min} \to -\infty$; $\alpha_{\max} \to +\infty$. All values of RV β_j in limits $\alpha_{\min}$ and $\alpha_{\max}$ are considered informative if:

$$\gamma = \beta_j, \quad when\left(\frac{\alpha_{\min}}{\delta} \le \beta_j \le \frac{B}{\delta} - 1\right),$$

i.e. the subarea of non-informative values $\{\beta_f\}$ is considered empty.

Further, for simplicity of calculations, we combined $\alpha_{\min}/\delta$ with the beginning of coordinates and used the known formulas for sums of natural series of integers and their squares. As results we note expression of mean-squared error of informative attributes selection as:

$$\sigma^2 = \sum_{\alpha_i=0}^{B/\delta-1} f(\alpha_i) \sum_{\substack{\alpha_i=0 \\ (\alpha_i \ne \beta_j)}}^{B/\delta-1} (\alpha_i\delta - \beta_j\delta)^2 \frac{P\delta}{B-\delta} + Q(\delta) \tag{8}$$

where, $Q(\delta)$ is summand with maximum order of minuteness, dependent on size of quantization step δ.

If the step of quantization is small enough, then we can write formula (8) as follow:

$$\sigma^2 = \sum_{\alpha_i=0}^{\frac{B}{\delta}-1} f(\alpha_i) \sum_{\alpha_j=0;\alpha_i\ne\alpha_j;}^{\frac{B}{\delta}-1} (\alpha_i - \beta_j)^2 \frac{P\delta^2}{B-\delta}. \tag{9}$$

After some transformations we found

$$\sigma^2 = \sum_{\alpha_i=0}^{\frac{B}{\delta}-1} f(\alpha_i) P(\alpha_i^2\delta^2 - \alpha_i\delta B + \frac{1}{3}B^2 + \frac{\alpha_i^2\delta^2}{B-\delta} - \frac{B-\delta}{6}), \tag{10}$$

where, $M\{a_i\delta\} = \sum_{a_i=0}^{\frac{B}{\delta}-1} \alpha_i f(\alpha_i) = a_{\alpha_i}$; $M\{\alpha_i^2\delta^2\} = \sum_{a_i=0}^{\frac{B}{\delta}-1} \alpha_i^2 f(\alpha_i) = \sigma^2_{\alpha_i} + a^2_{\alpha_i}$.

In limit $\delta \to 0$ the formula (8) is written down as:

$$\sigma^2 = \int_0^B \omega(\alpha)d\alpha \int_0^B (\alpha-\beta)^2 \frac{P}{B} d\beta = \int_0^B \frac{P}{3B}\left[\alpha^3 - (\alpha - B)^3\right]\omega(\alpha)d\alpha =$$
$$= \int_0^B \frac{P}{B}\left[\alpha - (\alpha - B)\right]\left[\alpha^2 + \alpha(\alpha - B) + (\alpha - B)^2\right]\omega(\alpha)d\alpha = \quad (11)$$
$$= P\left(M\{\alpha^2\} - BM\{\alpha\} + \frac{1}{3}B^2\right) = P\left(\sigma_\alpha^2 + a_\alpha^2 - Ba_\alpha + \frac{1}{3}B^2\right).$$

The results of developing the algorithms of control by mathematical expectation and dispersion of random attributes serve as basis for construction of algorithms which select the informative attributes on the basis of use dynamic properties of information.

In next section of paper we analyze the results of developing the algorithms based on control an increment and difference between initial and predicted values of attributes in time series representing principles of use the dynamic properties of data.

Algorithm of informative attributes selection by control the increments of time series

Let's consider random process $\alpha(t)$ with two-dimensional probability distribution function $\omega(\alpha_1, \alpha_2, \tau)$, mathematical expectation a_α, dispersion σ_α^2 and coefficient of correlation $R(\tau)$ ($\tau = t_k - t_{k-1}$ is a step of quantization by time).

We designated through $\Delta\alpha_k$ $(0 \le \Delta\alpha_k \le B)$ the difference (increment) $\alpha_k - \alpha_{k-1}$ between k-th and $(k-1)$-th discrete levels in a sequence of time series. Through $\Delta\beta_k$ we designated a difference (increment) $\beta_k - \beta_{k-1}$ $(0 \le \Delta\beta_k \le B)$ of transformed data. On an axis of set $\{\Delta\beta\}$ we established negative $(-\Delta x)$ and positive (Δy) borders of control, which divide set of increments $\{\Delta\beta\}$ on permitted $\{\Delta\beta_p\}$: $(-\Delta x \le \Delta\beta_p \le \Delta y)$ and forbidden $\{\Delta\beta_f\}$: $(-B \le -\Delta\beta_f \le -\Delta x,\ \Delta y \le \Delta\beta_f \le B)$ subsets.

The principle of the informative attribute control is based on accepting the attribute α_k as informative if $\Delta\beta_k \in \{\Delta\beta_p\}$ and non-informative if $\Delta\beta_k \in \{\Delta\beta_f\}$.

Accordingly to algorithm the values of non-informative attributes are accepted as

$$\gamma_k = \begin{cases} \beta_k, & if \quad \Delta\beta_k \in \{\beta_\rho\}, \\ \beta_{k-1}, & if \quad \Delta\beta_k \in \{\beta_f\}, \end{cases}$$

i.e. non-informative attribute is identified with value of previous attribute β_{k-1}.

The demonstrated rule is used at construction of rules for informative attributes selection on the basis of control of error for smoothing, approximating models and other various models of statistical prediction.

Common expression for estimation of minimal mean-square error of informative attributes selection is written down as:

$$\sigma_n^2 = M[(\gamma_k - \beta_k)^2] = M_{\alpha_k \neq \beta_k;\, \Delta\beta_k^* \in \{\Delta\beta_\rho^*\}}[(\alpha_k - \beta_k^*)^2] + M_{\alpha_k = \beta_k;\, \Delta\beta_k^* \in \{\Delta\beta_3^*\}}[(\alpha_k - \beta_k^*)^2] + M_{\alpha_k \neq \beta_k;\, \Delta\beta_k^* \in \{\Delta\beta_3^*\}}[(\alpha_k - \beta_k^*)^2], \quad (12)$$

In this formula we taken into account that selection is carried out by methods of prediction in view of probabilities of errors of the first and second kind.

First summand in right part of expression (12) gives ratings of errors caused by first kind errors probability. In second and third parts of (12) the probabilities of second kind errors are taken into account. As a result of mathematical transformations we got:

$$\sigma_n^2 = \int_\Omega \omega(\alpha) \frac{P}{B}\left[\int_{0(-\infty)}^{B} (\alpha - a_\alpha)^2 d\beta + \int_x^y (\alpha - \beta)^2 d\beta + \int_y^{B(+\infty)} (\alpha - a_\alpha)^2 d\beta\right] d\alpha + (1-P)\left[\int_\Omega (\alpha - a_\alpha)^2 \omega(\alpha) d\alpha - \int_x^y (\alpha - a_\alpha)\omega(\alpha) d\alpha\right],$$

where $\int_\Omega (\alpha - a_\alpha)^2 \omega(\alpha) d\alpha = \sigma_\alpha^2$.

Let's note that the expressions of rating the mean-squared error of informative attributes selection are gotten for normal law of attributes probabilities distribution. We got results of similar researches at lognormal, exponential, Weibull and Rayleigh laws of distributions for one-dimensional series.

Alongside with it the large interest represents optimization of data processing for multi-dimensional time series. So, in next section we offered the algorithm of informative attributes selection by control of conditional statistical and probability characteristics of multidimensional time series. This algorithm extends the application area of mechanisms which use the data dynamic properties at data processing.

Algorithm of informative attributes selection by conditional moments

We assumed, that the following statistical characteristics of the data are known: two-dimensional function of probabilities distribution $\varpi_2(\xi, \eta, t_2, t_1)$; unconditional dispersions σ_ξ^2, σ_η^2; mathematical expectations α_ξ, α_η; coefficients of correlation $R(\tau)$ (where τ is difference of time between numbers of sampling). Let's enter designations $m(\eta/\xi)$ and $m(\xi/\eta)$ for conditional mathematical expectations of RV, dependent from variable ξ and η, with which numbers of a discrete sequence are conditionally designated. Conditional dispersions of examined processes are designated through $D(\eta/\xi)$ and $D(\xi/\eta)$. Ranges of practical scope of RV η and ξ are designated through $B_\eta = \eta_{\max} - \eta_{\min}$ and $B_\xi = \xi_{\max} - \xi_{\min}$. The true values of input variable η and ξ can be transformed into values of some output two-dimensional variable α and β with probabilities $P_{\eta\alpha}$ and $P_{\xi\beta}$. For clarity representation of analytical decision it is lawful to consider probabilities of errors $\eta \to \alpha$ and $\xi \to \beta$ equiprobable, and to designate as P_{α} and P_{β} the average a priory probability of errors for transformed sizes.

Rule of informative attributes selection by control of conditional moment is written down as:

- attribute informative, if $m_{\acute{I}}\ (\beta/\alpha) \le m(\beta/\alpha) \le m_B(\beta/\alpha)$;
- attribute non-informative, if $\{ m(\beta/\alpha) < m_{\acute{I}}\ (\beta/\alpha),\ m(\beta/\alpha) > m_B(\beta/\alpha) \}$,

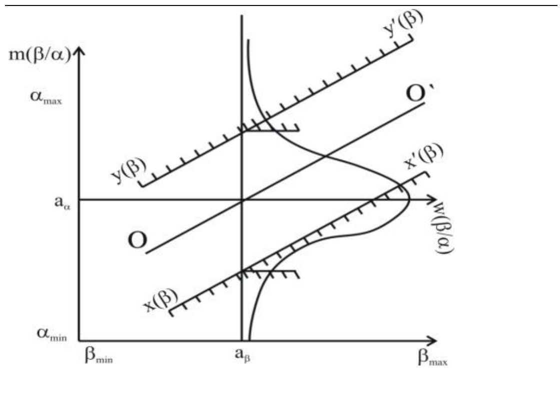
where, $m_H(\beta/\alpha)$, $m_{\acute{I}}\ (\alpha/\beta)$ and $m_B(\beta/\alpha)$, $m_B(\alpha/\beta)$ are allowable bottom and top borders of conditional average values.

Let's note that rating of informativeness also is determined by conditional moment for sizes α and β.

In Figure 2 the principle of informative attributes selection on conditional average values is graphically represented. The direct line OO' is the a line of regress β relative to α, constructed on the basis of assumption, that between RV α and β is linear correlation. Straight lines $x(\beta)x'(\beta)$ and $y(\beta)y'(\beta)$, carried out in parallel lines to regress OO' mean borders of control of conditional average values. According to rule, the coordinates of any point laying inside this shaded area, are considered permitted.

The stated principles and rules of informative attributes selection by conditional statistical characteristics are advanced by use of techniques for choice adequate regression model.

FIGURE 2. PRINCIPLE OF INFORMATION SELECTION



Algorithm of informative attributes selection on basis of the adequate regress model

The model is considered adequate, if it reduces significantly the mean-squared error of information in informative attributes selection. For check of adequacy i-th and j-th models we defined variables:

$$y^* = 2y - f_i(\overline{X}, \overline{b}_i) - f_i(\overline{X}, \overline{b}_j);\; x^* = f_i(\overline{X}, \overline{b}_i) - f_j(\overline{X}, \overline{b}_j).$$

where $f_i(\overline{X}, \overline{b}_i)$ are the regress models constructed on the basis of specified estimations.

Then for calculation of coefficient $\hat{\lambda}_{ij}$ estimation from equation of regress we offered the formula

$$\lambda_{ij} = \frac{\sum_{l=1}^{N} y_l^* x_l^*}{\sum_{l=1}^{N} (x_l^*)^2}.$$

The bottom and top borders of confidential intervals for informative attributes selection is calculated by formulas:

$$T_H = \hat{\lambda}_{ij} - t_{kp} \sqrt{\frac{\sum_{l=1}^{N} (y_l^* - \hat{\lambda}_{ij} x_l^*)^2}{\gamma \sum_{l=1}^{N} (x_l^*)^2}};$$

$$T_{\hat{A}} = \hat{\lambda}_{ij} + t_{kp} \sqrt{\frac{\sum_{l=1}^{N} (y_l^* - \hat{\lambda}_{ij} x_l^*)^2}{\gamma \sum_{l=1}^{N} (x_l^*)^2}},$$

where γ is the number of degrees of freedom calculated on the formula $\gamma = N - k$, $k = \inf\{k_i, k_j\}$; t_{kp} are the values of Student's t-criterion when degrees of freedom are γ and significance value is α.

The algorithm determines the controllable attribute as informative if the conditional mathematical expectation is in limits

$$T_H \le m(\beta / \alpha) \le T_B,$$

and as non-informative, when value of conditional mathematical expectation are outside of these borders.

Conclusion

The developed methods and algorithms of informative attributes selection project mechanisms of use the simplified estimations of mathematical expectation, correlation function and distribution function. Specialty and novelty of the offered algorithms for selection of informative attributes at multidimensional time series is using the dynamic properties of non-stationary objects on the basis of control of conditional average values, parameters of adequate regress model and model on confidential intervals.

The reliability of theoretical techniques is proved to extensive experiments by testing algorithms of adaptive data processing for approximation and forecasting of technical parameters of electro supply companies. It is established, that the acceleration of accounts is achieved 10-15 time at the considerably reduced sizes of data samples.

References

Djumanov, O., 2008. "The system of image recognition for dynamic processes objects on a basis of neural technologies", 9th International Conference on Electronics, Information and Communication [9-ya Mezhdunarodnaya Konferenciya po Elektronike, Informacii i Kommunikacii], in Russian, Tashkent, pp.286-90

Djumanov, O., 2011. "Adaptive designing for neuronetworking system of processing the data with non-stationary nature," ATI - Applied Technologies & Innovations, Vol.4(1), pp.48-57

Haykin, S., 1996. Adaptive Filter Theory, 3rd edition, Englewood Cliffs, NJ: Prentice-Hall

Sankar K. Pal, Sushmita, M., 1992. "Multilayer perceptron, fuzzy sets and classification," IEEE Transactions on Neural Networks, Vol.3(5), pp.683-97

Umbaugh, S., 1998. "Computer vision and image processing," Englewood Cliffs, NJ: Prentice Hall

