4.2. Decisions trees and multivariate time series analysis.

In this section we will consider methods for the solvution of problems of the analysis and forecasting of multivariate heterogeneous time series.

In many practical problems, it is required to predict values of characteristics of an object on the basis of the analysis of their values at the previous moments of time.

Now the theory is developed and the large number of various methods of the analysis of multivariate numerical sequences is created. However, application of these results for the decision of a considered problem in case of heterogeneous characteristics is impossible (since for qualitative characteristics, arithmetic operations on set of their values are not defined).

Using decisions trees, we can solve the specified problems.

Let for the description of a object of research the set of stochastic characteristics $X(t)=(X_1(t),...,X_n(t))$ be used. The values of the characteristics are changed in the run of time. The characteristics can be both of the quantitative and qualitative type.

Let characteristics be measured at the consecutive moments of time $t^1,..., t^\mu,...$. For definiteness we will assume that measurements will be carried out through equal intervals of time. We will designate through $x_j(t^\mu)=X_j(t^\mu)$ the value of characteristic X_j at the moment of time t^μ. Thus, we have n-dimensional heterogeneous time series $x_j(t^\mu)$, $j=1,...,n$, $\mu=1,2,....$

Let us choose one predicted characteristic X_{j_0}, $1 \le j_0 \le n$.

We designate, for convenience, this characteristic through Y. The characteristic Y can be both quantitative, and qualitative type.

Let us consider the moment of time t^μ, and also a set of the previous moments of time $t^{\mu-1}$, $t^{\mu-2},..., t^{\mu-l}$, where l is a given size ("deep of history"), $1 \le l < \mu$.

We suppose that conditional distribution $Y(t^\mu)$, when all previous values $X(t)$ are given, depends only on values of series in l previous moments of time.

Besides, we suppose, that this dependence is the same for any value μ. The given assumption means, that the statistical properties of series determining dependence are stationary.

For any moment of time t^μ, it is possible to form a set $v^\mu =(X_j(t^{\mu-i}))$, $i=1,...,l$, $j=1,...,n$, representing the time series in l previous moments of time. We will call a set v^μ background of length l for the moment t^μ.

It is required to construct a model of dependence of characteristic Y from its background for any moment of time. The model allows to predict the value of characteristic Y at the future moment of time on values of characteristics for l last moments. In other words, the given model, using background, represents decision function for forecasting.

Depending on the type of characteristic Y, we will consider various types of forecasting:
1. Y is the qualitative characteristic.
In analogy to a usual PR problem, we will call a problem of the given type a problem of recognition of dynamic object. The analyzed object can change its class in the run of time.
2. Y is the quantitative characteristic.
In this case, we have a forecasting problem of quantitative characteristic of object.

We will represent a decision function for forecasting time series on its background as a decision tree. This decision tree differs from the described in §2 trees in statements concerning a characteristics X_j in some i-th moment of time back are checked. For convenience, we will designate these characteristics, with a glance to background, through X_j^i (figure 21). Thus, X_j^i means

characteristic X_j in i-th previous moment of time (concerning a present situation).

Let there be a set of measurements of characteristics $X=(X_1,...,X_n)$ at the moment of time $t^1,...,t^N$ and value l is also given. Thus, we have a multivariate heterogeneous time series of length N. We generate set of all histories of length l for the moments of time $t^{l+1},...,t^N$: $A = v^\mu$, $\mu=l+1,...,N$.

X^3_2

<0 >0

$[0,1]$

X^1_1 X^2_2

≤ 2 >2 $\hat{Y}(3)=1$ $\leq$ >1

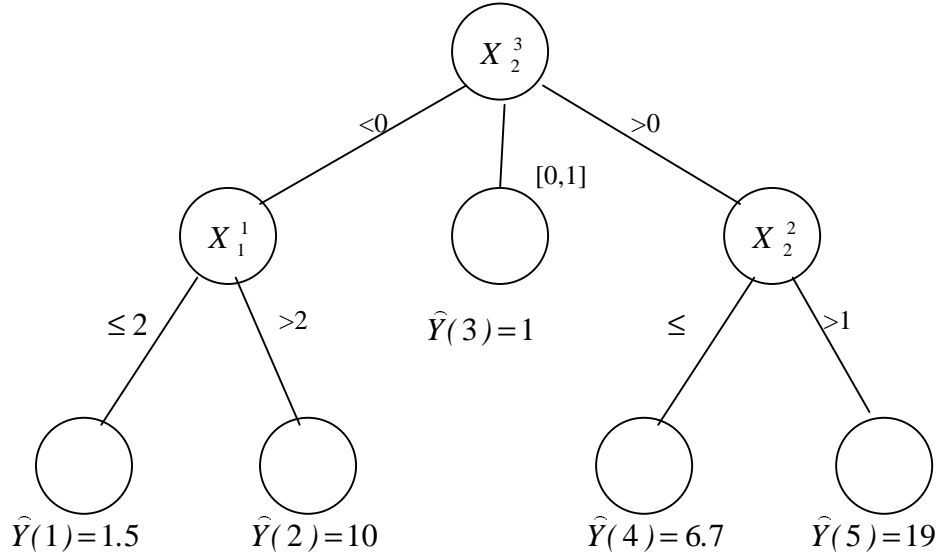$\hat{Y}(1)=1.5$ $\hat{Y}(2)=10$ $\hat{Y}(4)=6.7$ $\hat{Y}(5)=19$

Fig. 21

For any given decision tree for forecasting by background, it is possible to define its quality similarly to how it was done for the usual tree: we will designate through $\hat{Y}(t^\mu)$ predicted value Y received with the help of a tree by background v^μ. The criterion of quality will be

$$Q = \frac{1}{N-l} \sum_{\mu=l+1}^{N} h(\mu),$$

where $h(\mu) = \begin{cases} 0, & if \ Y(t^\mu)=\hat{Y}(t^\mu) \\ 1, & otherwise \end{cases}$

for a recognition problem of dynamic object and

$$h(\mu) = (Y(t^\mu) - \hat{Y}(t^\mu))^2$$

for a forecasting problem of the quantitative characteristic.

The given series is used for learning to forecasting.

Let there be a series $x(t^\mu)$ of length N_c, $\mu=N+1,...,N+N_c$. It is then possible to compare the predicted values $\hat{Y}(t^\mu)$, received as a result of training with "true" values $Y(t^\mu)$ and to define an error of the forecast. We will say in this case, that the given series is used for control of qualities of forecasting.

How to construct a decision tree for forecasting by background on an available time series?

Some ways are described below. The initial problem of construction of a decision tree is divided into some more simple pattern recognition or regression analysis problems, depending on type of predicted characteristic Y.

We will present set v^μ as the *table* $v^\mu = (X_j(t^{\mu-i})$, $i=1,...,l$, $j=1,...,n$ containing l rows and n columns. Then, the initial information for forecasting is the set of tables v^μ, together with the values of predicted characteristic Y specified for each table $Y(t^\mu)$, $\mu=l+1,...,N$.

It is possible to present set $A = v^{l+1},..., v^N$ as the three-dimensional table of dimension $l \times n \times (N-l)$ to which the vector $(y^{l+1},...,y^N)$ corresponds (figure 22). However, available methods of recognition or regression analysis with using of decisions trees use bi-dimensional tables as input information.
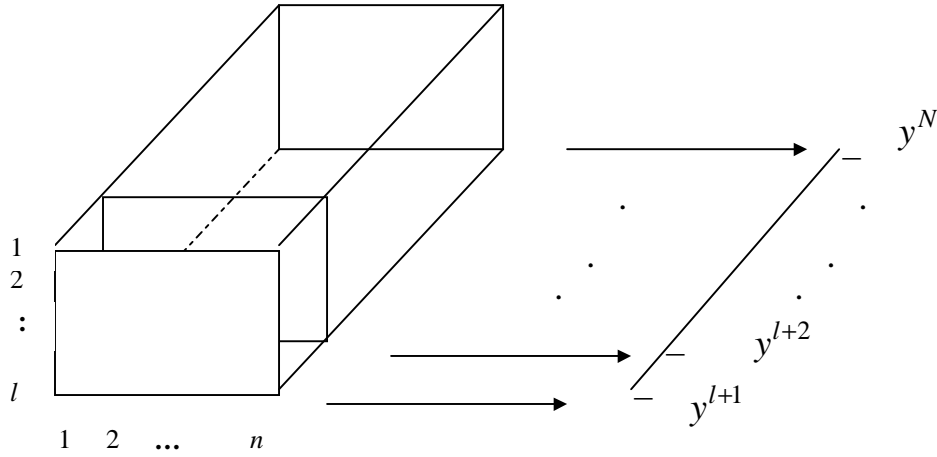

Fig. 22

There are various ways to use given methods for the analysis of three-dimensional data tables.

1. Each table v^μ is represented as a line of the appropriate values of characteristics $X_1^1, X_2^1,...,X_n^1, X_1^2, X_2^2,...,X_n^l$, $\mu=l+1,...,N$ (in other words, the table is stretched in line).

As a result, we receive the two-dimensional table of dimension $l \times n \times (N-l)$, for which the decision function as a decision tree is constructed. For this purpose one of the methods described in §2 can be used.

As researches show, the given way is very simple in realization, however the received decisions, based on conditions such as: the number of characteristics is large, the background is long and the length of rows is small, can be unstable, i.e. will give the large error on a control series. It is known, that the effect of instability appears when the sample size is small and the number of characteristics is large.

1. The initial problem is solved in two stages. At the first stage are considered l two-dimensional tables of a kind $(X_j(t^{\mu-i}), Y(t^\mu))$ where $j \in \{1,2,...,n\}, \mu \in \{l+1, l+2,...,N\}, i \in \{1,2,...,l\}$
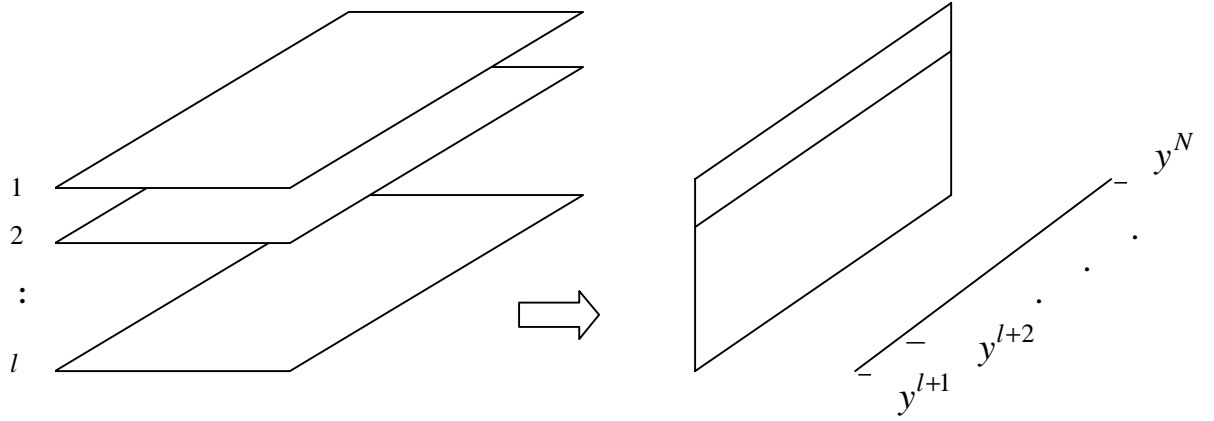
Fig. 23

We construct *l* various decisions trees for a prediction of value of *Y* on each of the given tables. Each of the tables, with the help of the constructed tree, transfer into an one-dimensional row of symbols (each symbol is coded by the number of the corresponding leaf of a tree). Thus, we get a two-dimensional table for which at the second stage the decision tree is again constructed. Then each symbol transforms back into the appropriate chain of statements.

3. The problem is solved in some stages.

1) We consider *l* tables $X_j(t^{\mu-i}),Y(t^{\mu})$, where

$j \in \{1,2,...,n\}, \mu \in \{l+1,l+2,...,N\}, i \in \{1,2,...,l\}$ and then *n* tables $X_j(t^{\mu-i}),Y(t^{\mu})$, where

$j \in \{1,2,...,n\}, \mu \in \{l+1,l+2,...,N\}, i \in \{1,2,...,l\}$. Thus, we have *l* horizontal and *n* vertical cuts of the initial table (figure 24).

For each received two-dimensional table, the decision tree is constructed. In a result we receive a set of trees $T_1,T_2,..., T_{l+n}$.

We will denote the best of these trees as T^*.

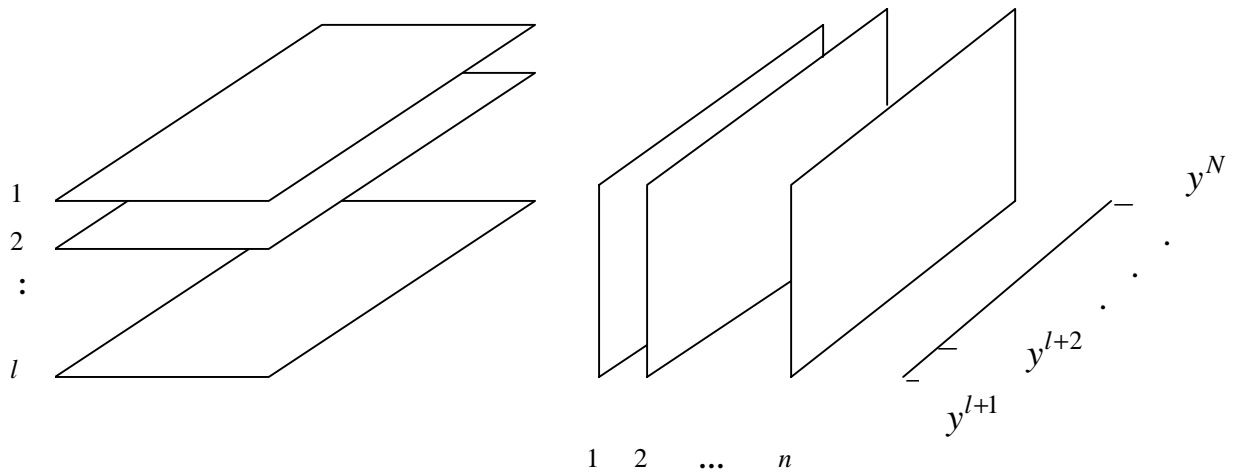2) All terminal nodes of a tree T^* are considered.

Fig. 24

The trees for which the error of forecasting exceeds the given size are selected. For each of

4

these nodes the appropriate set of backgrounds $A' \subset A$ is formed and then we repeat the process of tree building for A' with the first step.

3) The process stops when the received tree will satisfy to a termination condition (i.e. when the number of leaves will be equal to given size M_{max} or when the error of forecasting will be less than the given value).

The given way differs from previous one in step-by-step tree building.