UDK 506+510

I.V. Baklan

LINGUISTIC MODELS OF FRACTAL DATA SERIES

Abstract. *The properties of construction linguistic models of dynamic processes time series. In the article the theoretical bases of linguistic models of dynamic processes. Attention is paid to the peculiarities of forming patterns with fractal properties.*

Keywords: *linguistic model, linguistic modeling, time series, fractal grammars.*

Formulation of problem

Standing detection task fractal properties of time series generated by dynamic processes, and building linguistic models using fractal formal grammars.

Analysis of recent research and publication

Linguistic modeling [1,2] based on the following mathematical theory known as the theory of time series, interval mathematics[3,4], mathematical tools of hidden Markov models, structural approach to pattern recognition, the theory of formal grammars[5].

Task of investigations

Research objective is offer linguistic modeling as a universal mechanism for time series modeling of dynamic processes with having fractal properties for further analysis and forecast.

Main part (the results of theoretical and experimental research)

We give a definition of linguistic modeling and linguistic model.

Linguistic modeling - a complex of methods, techniques and algorithms that use numerical arrays process of converting information into linguistic sequences on which resumed formal grammar. Linguistic model - is based on a set of linguistic modeling symbolic (linguistic) sequences for the chosen parameters of the process of transformation and restored on the basis of its formal grammar. Linguistic modeling should be considered as a specific kind of mathematical modeling to handle data in a character (not numerical) form.

In this paper we considerate main principles of new kind mathematical modeling, linguistic modeling, and formalisms, which released its.

At begin we will be done definition of linguistic model. The linguistic model is formal system that consists of four elements:

$$<D,I, L,G>$$

where D - set of input data(for example, time series);

I - rules of input data intervalization;

L – isomorphism from set interval to symbol set (alphabet);

G – formal grammar was restored by linguistic chains set given .

The linguistic modeling may be used to practice solving of different problems: time series forecasting; pattern recognition (sound, voice, emotion statement etc.); user's authentication by hand motions; early diagnose of skeleton-muscle diseases and other.

Now we will be discussed main procedures for realization of intervalization rules.

The linguistic model building process consist of next steps:

- data preprocession (first calculation of the time series of differences of various orders etc.) $\Delta^1 = x_{i+1} - x_i$;
- definition of the field values of the time series and its frequency characteristics;
- intervalisation values set (special kid of quantification);
- process of converting numerical series to symbolic form, which we named "linguistisation";
- formal grammar restoring with HMM.

Of course that the simplest linguistization scheme power corresponding alphabet must be much smaller than the coherence time series.

Let we have two partially ordered sets X and Y. The set, which asked the relevant procedure is called ordered if the order relation defined for any two of its elements, and partially ordered otherwise. Partially ordered set is called a structure, if any two-element subset that it has the exact upper and lower limit, and a complete structure, if each of its non-empty subsets has the exact boundaries.

Each of the sets and we assume conditional structures and mark S(X) and S(Y). The ratio of the order is denoted by ≤.

If we have a,b∈S(X) and a≤b, then the set $I(a,b) = [a,b] = \{x \in X, a \leq x \leq b\}$ set will be named an interval on S(X). The set of all intervals on the structure of S (X) denote $\mathcal{I}_{S(X)}$.

 ISSN 1562-9945

Thus, if $X = R^1$- set of real numbers, then $\mathcal{J}_{R^1}$- the set of closed intervals on the line of real numbers. In this case, $\mathcal{J}_{R^1}$ is called interval number.

Returning to the procedure of partitioning the set of values into intervals, we note that basically we consider intervals that are not degenerate. In this case, the simplest linguistization method could bring everything to the fact that all the values of the time series (or differences) is a degenerate interval and other intervals not.

We are interested in certain cases that reflect the following types intervalization: 1) when the intervals are equal; 2) logarithmic intervals; 3) when equal probability intervals; 4) intervals for a specific probability distribution (normal, beta distribution, Poisson, Dirichlet, etc.).

To construct intervals according to the selected distribution we use interval mathematics. For example, we have $[a_{3,}a_6] = [a_{3,}a_4] + [a_{4,}a_5] + [a_{5,}a_6]$ for $v_{3,6} = v_{3,4} + v_{4,5} + v_{5,6}$ (see fig.1).

When partitions on equal intervals of the N-level set X, we have $\omega(a_1, b_1) = \omega(a_2, b_2) = \cdots = \omega(a_N, b_N)$, where $\omega(a_i, b_i) = b_i - a_i$ is width of the interval $I[a_i, b_i]$.

When equally (or equifrequent) set partitioning values into intervals we have $dim\{I[a_1, b_1]\} = dim\{I[a_2, b_2]\} = \cdots = dim\{I[a_i, b_i]\} = \cdots = dim\{I[a_N, b_N]\}$.

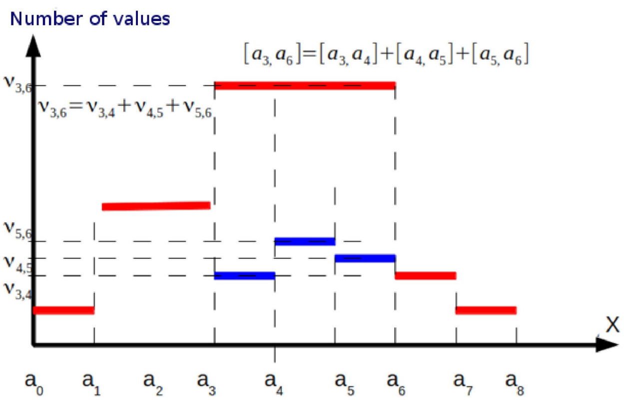

Fig. 3. Construction intervals using the rules of interval mathematics

Calculating frequency response range is in accordance with a distribution.

For example, the calculation parameters for beta distribution we can find the average m and variance s by difference time series. And then we have a beta distribution expression

$$a = -(m \cdot s + m^3 - m^2)/s$$

$$b = ((m-1) \cdot s + m^3 - 2 \cdot m^2 + m)/s$$

We have an alphabet $\mathcal{A} = \{e_i\}_{i=\overline{1,N}}$, the set of elements e_i, which also imposed ratio procedure $e_1 \ll e_2 \ll \cdots \ll e_N$, which we call lexicographic order for the alphabet $\mathcal{A}$. We have a time series $X = \{x_1, x_2, \dots, x_M\}$, which is consistent with the set of intervals I, ie, all the values of a number of values between $a_1$and b_N. The operation resulted intervalizatsiyi time series X is a sequence of intervals, which include elements of the time series, $(X) = I_X = \{I^1, I^2, \dots, I^M\}$, $I^j \in I, j = 1,2, \dots, M$. In fact, you can write $\mathcal{J}(x_j) = I^j, j = 1,2, \dots, M$.

Defined mapping $\mathfrak{L}: I \to \mathcal{A}$, morphism, which puts each interval $I_i \in I$ element alphabet $\mathcal{A}$.

At the stage of converting numerical values of the time series by using reflection $\mathfrak{L}$, it replaces elements x_i to $e_{j=}\mathfrak{L}(I_j)$, if $x_i \in I_j$.

As a result of this operation, we obtain a sequence of linguistic elements of a set of alphabet $\mathcal{A}$: $\mathfrak{L}(\mathcal{J}(X)) = \mathfrak{L}(I_X) = \{\alpha_1, \alpha_2, \dots, \alpha_M\}$, ie $\alpha_j = \mathfrak{L}(I^j), j = 1,2, \dots, M$.

In converting the results we obtained material (linguistic chain), which subsequently can be used to restore formal grammar (see fig.2,3).

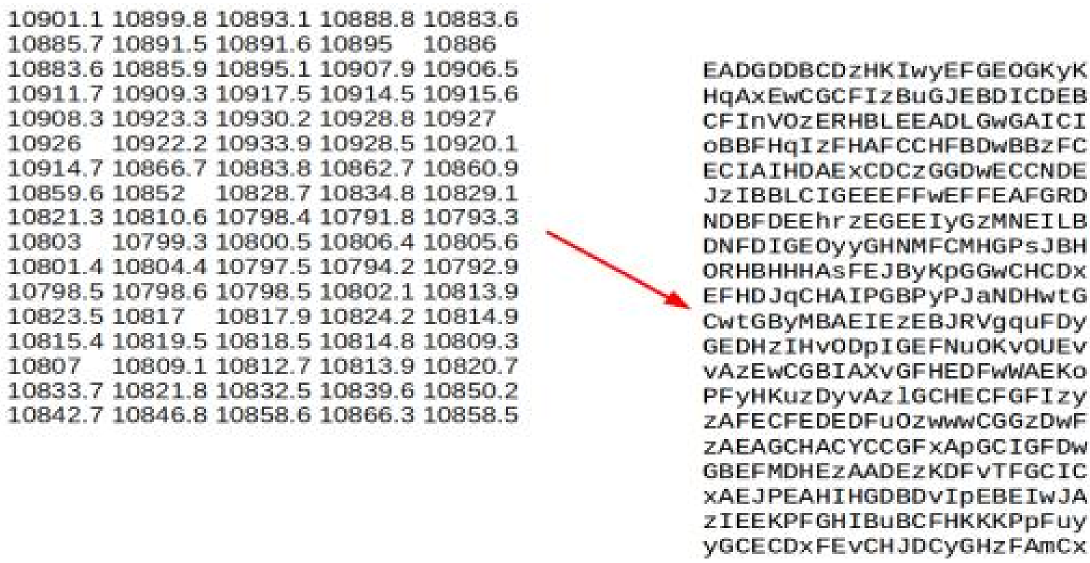

Fig. 2. Numerical series convert to a linguistic chain

ISSN 1562-9945

IEvCHuDhlxEyHCsQyEeGEMFExyXZvtgqtHGzOEIkfswAspvHGv FlqHNAHwsACwaKsBovlwtCmoxuFutAvEytwssyrMluxBpwzDowtsutsrx BtwyxtLBylyzvKBzyn

Fig. 3. Linguistic chain of time series

One of the possible options to solve this problem - the use of mathematical tools hidden Markov models for constructing probabilistic formal grammars. Example transition probability matrix will be show in fig.4.

	r	s	t	u	v	w	x	y
p	0	0	0	0	0,007874015	0,007874015	0	0
q	0	0	0,007874015	0	0	0	0	0
r	0	0	0	0	0	0	0,007874015	0
s	0,007874015	0,007874015	0	0,007874015	0	0,007874015	0	0,007874015
t	0	0,015748031	0	0	0	0,015748031	0	0
u	0	0	0,015748031	0	0	0	0,007874015	0
v	0	0	0,007874015	0	0	0	0	0
w	0	0,015748031	0,015748031	0	0	0	0	0,007874015
x	0	0	0,007874015	0,007874015	0	0	0	0,007874015
y	0,007874015	0	0,007874015	0	0	0	0,007874015	0
z	0	0	0	0	0,007874015	0	0	0,007874015
A	0	0,007874015	0	0	0,007874015	0	0	0
B	0	0	0,007874015	0	0	0	0	0,007874015
C	0	0,007874015	0	0	0	0,007874015	0	0
D	0	0	0	0	0	0	0	0
E	0	0	0	0	0,007874015	0	0,007874015	0,015748031

IEvCHuDhlxEyHCsQyEeGEMFExyXZvtgqtHGzOEIkfswAspvHGv FlqHNAHwsACwaKsBovlwtCmoxuFutAvEytwssyrMluxBpwzDowtsutsrx BtwyxtLBylyzvKBzyn

Fig. 4. Transition probability matrix of linguistic chain HMM

Fractal dimension in the classic sense is a number that describes quantitatively how an object, process fills the space. There are many ways of calculating fractal dimension. They are based on the calculation of volume or area of fractal formation in the same space where there is education.

Consider the time series $\{x_i\}_{i=\overline{1,N}}$ that describes a dynamic process. If the levels of a series x_i of independent, it is easy to see the lack of bright trends and dynamic behavior of the process will be more likely to resemble "white noise". In this case, the fractal dimension will be directed to the magnitude of topological dimension plane, that is. If the value of the time series are not independent, the fractal dimension is much smaller than 2. Contents of this is that our time series dynamic process has memory, that there are certain time intervals varying quality trends that change the times of uncertainty.

A completely different situation when we deal with linguistic links.

At the ordered $A=\{\alpha_1,\dots,\alpha_M\}$ alphabet introduced a measure $\mu(\alpha_i)=f(i)$ where f - a function of the elements of the alphabet (the Index).

The notion of distance between elements of the alphabet $R[\alpha_i,\alpha_j] = \begin{cases}\mu(\alpha_i)-\mu(\alpha_j), & if\ i>j\\ \mu(\alpha_j)-\mu(\alpha_i), & if\ j>i\end{cases}$.

While similar numerical series, introducing indicators fractal linguistic series.

$$\mathcal{L}_H = 2 - \mathcal{H},$$

$$\mathcal{H} = \frac{log\left(\frac{\mathcal{R}}{\mathcal{S}}\right)}{log\left(\frac{N}{2}\right)},$$

where $\mathcal{R} = max\{\mu(\alpha_i)|\alpha_i \in \mathcal{A}\} - min\{\mu(\alpha_i)|\alpha_i \in \mathcal{A}\}$ and $\mathcal{S}$ is the mean linguistic deviations of linguistic series.

Conclusions and prospects for further research

The article was the analysis to identify fractal properties of linguistic dynamic process chains. Features of the application of linguistic modeling to build models of dynamic processes with fractal properties.

REFERENCES

1. Baklan I.V. Linguistic modeling for solving of different problems // Intellectual System for Decision Making and Problems of

Computational Intelligence: Conference Proceeding. – Kherson: KNTU, 2013. – P.398-400.

2. Baklan I.V., Stepankova H.A. Classification of Markov models like: scientific monograph. - Kyiv: National Academy of Management, 2012. - 84 p. (in Ukrainian).
3. Moore R.E. Interval analysis. – Englewood Cliffs: Prentice Hall, 1966.
4. Sunaga T. Theory of an interval algebra and its application to numerical analysis // RAAG Memoirs. – 1958. – Vol. 2, Misc. II. – P. 547-564..
5. Michael A. Harrison. Introduction to formal language theory. - Addison-Wesley Pub. Co., 1978 - 594 p.