УДК 519.7

I. Didmanidze, R. Tkhilaishvili

CHOOSE OF THE COMPRESSION SCHEMA

The selection of a compression scheme shows that, in given limitations in open information systems, the lower limit at which compression occurs as a result of construction of iterative structures, is the four-digit coding system.

It suggests us to think that the fact that the codes constructed in the four-digit coding system will have a high value of the coefficient of compression. This result suggests that a four-letter alphabet coding, experimentally confirmed by the phenomenon of the genetic code. It is also shown its practical application value for the biological and natural systems, as well as – to explain, given the limitations of the phenomenon of short-term and long-term memory.

Keywords: compression schemes, iterative structures, the compression ratio.

Data compression – it is the technique of data size reducing on the different accumulators (on the magnetic discs or the magnetic stripe). It is carried out with different methods. It is used physical and logical compression, symmetrical and asymmetrical, adapter and half adapter codify, compression without and with losing and with minimal losing. Let's discus one exemplar of choose of the compression schema [1].

Let's suppose four-digit sequence.

$$v_{ri}^{i} = v_1 v_2 \ldots\ldots v_i \ldots\ldots v_r \quad (1)$$

Then f and A. re indication of binary sequence.

$$b_{ri}^{i} = b_1 b_2 \ldots\ldots b_i \ldots\ldots\ldots b_r \quad (2)$$

$v_i \in \{0,1,2,3\}; i = 1, r$

Next transformatio $v_{\mathrm{ri}}^{\mathrm{i}}$ happens by

$$S = v_i v_{i+1} v_{i+2} \ldots\ldots v_v^m + \frac{m}{2} v_{n+1} v_{n+2} \ldots\ldots v_{2^n} \quad (3)$$

As result of it we have such sequence

$$S_N = S_1 S_2 \ldots\ldots\ldots S_N \quad (4)$$

When $\mathrm{N} \geq \mathrm{r}$, n . ilength of the vector V_{kn} .

If (3) is used after each n . ement i.e. when step of the operation is t . ,then (1) between two vectors is equal of $\mathrm{N} = \mathrm{r}$.

But for further evaluation of compression coefficient of the information $\mathrm{K}_{\mathrm{comp}}$ most interesting problem is the definition of such t and n . en N . aches maximum.

It is evident, that the fastest growth of the value N . r every $\mathrm{n} > 1$ proceeds when $\mathrm{t} = 1$. More important is the question of the choice of the length of block vector V_{kn} when growth of N is $\mathrm{N}_{\max}$.

Let's introduce the notion of the expansion coefficient – K_n – which characterizes growth of the sequence N S_n^t with variations of the n i.e.

$$K_N = \frac{N(S)_{n>2}}{N(S)_{n=2}} \quad (5)$$

When $t = 1$, $N(S_n)_{n=2}$ and $N(S_n)_{n>2}$ amount of elements of the sequence when $n = 2b$ $n > 2$ in the vector V_{kn} .

LEMA1. When $t = 1$, $N = N_{max}$, n is equal or is close to

$$n = \frac{r+1}{4} \quad (6)$$

PROOF. Since $t = 1$, the amount of block vectors k V_{kn} in the sequence S_N^t will be $k = (r-n)-(n-1) = r-2n+1$, and the amount of letters $N = rn - 2n^2 + n$.

The choice of the optimal notion of n , providing the maximum $N = N_{max}$, must be found $\frac{dN}{dn} = \frac{d}{dn}\left(rn - 2n^2 + n\right) = 0$, when $n = \frac{r+1}{4}$.

THEOREMA:

For four-digit sequence $S_N^t = S_1 S_2 \ldots\ldots S_N$, $N \geq r$, minimal nomination of the N when $t = 1$ is determined as

$$n = n_{min} \sqrt{\frac{N}{2}} \quad (7)$$

PROOF. According to the lemma (1) we have

$$N = r \cdot n - 2n^2 + n \quad (8)$$

Or $r = \frac{N}{n} + 2n - 1$.

For definition of $\frac{dr}{dn} = r_{min}$, providing maximum $N = N_{max}$, let's see $\frac{dr}{dn} = 0$, here

$$n = n_{min}\sqrt{\frac{N}{2}}.$$

Example 1. Let's say $N = N_{max} = 8$. In the sequence b_r, we define r and the length of codding layer n_{min}, providing

$$n = n_{min} = \sqrt{\frac{N}{2}} = 2$$

$$r = \frac{N}{2} + 2n - 1 = 7.$$

So when number of bits in $b_r = 7$, then $t = 1$ maximal number of elements in the sequence S_N forms when $n = 2$.

Really, if $t = 1, r = 7, n = 2$ and is given some actual sequence $b_{r_i}^i$. So $b_{r_i}^i = 1110111 \to v_{r_i}^i \to 1233012 \to 30.03.31.10$ for the same $b_{r_0}^0$, $N = 7$ and $n = 7$, $N = 6$, $n = 3$. From the lemma 1 and theorem goes [2].

Conclusion 1. When $t = 1$ two adjacent codding layers of the length n have equal $n-1$ of equal elements.

As we already have seen algebraic sequence and feature of codding sequences let us to choose and to buy cyclical schema of the compression of data, built from these steps:

1) Function of the conversion $f: B \Rightarrow V$ transmits initial binary sequence $B_r^i = b_1^0 b_2^0 ......b_r^0$ into four-digit sequence $v_{r_i}^i = b_1^0 b_2^0 ......b_r^0$;
2) For $t = 1$ sequence v_r^0 is founded according the theorem transmits into four-digit sequence S_r^2 when the length of each codding word $n = 2$.

Note: choice of the notion $n = 2$, i.e. introducing extra amount, when in two adjacent vectors border elements are similar, is due the demand of simple decoding.

3) On the ground (7) of $N = N_{max}$ in S_r^2 defines minimal length r_i binary sequence $B_{r_i}^i$ and corresponding length of codding word n.
4) On the ground of (5) we transmit accounted notions r_i and n sequence S_r^2 in $S_{r_i}^{n-1}$, afterwards we can define concrete type of binary sequence b_1.

So finish first cycle of information compression.

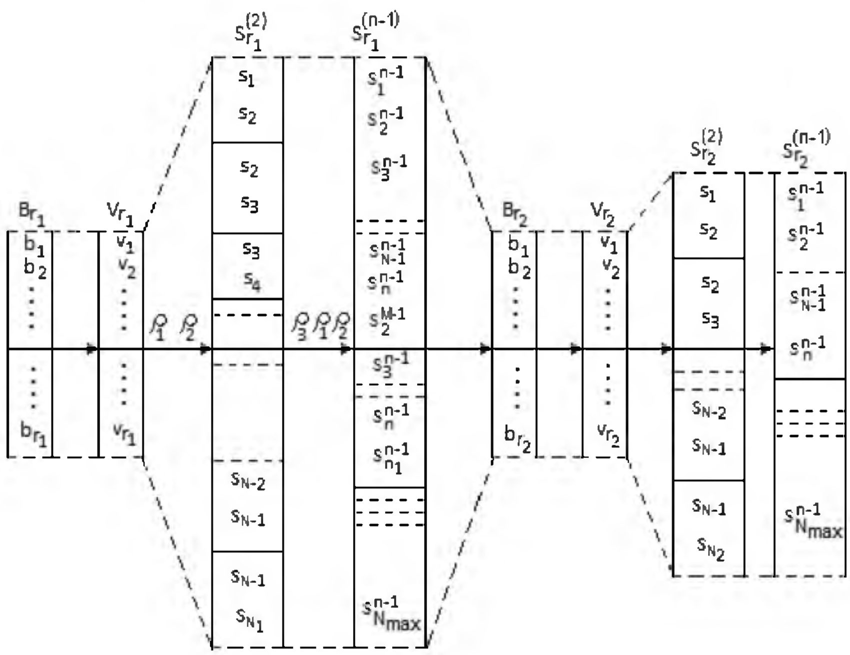
The coefficient of the compression for end of the first cycle is

$$K_{comp}^1 = \frac{(rn - 2n^2 + n)^2}{r}.$$

If we represent codding sequence $B_{r_i}^i$ as vertical matrix, size $r \times 1$,further conversion due the schema of compression within one cycle can be presented as such graphical schema (pic. 1).

From the graphical schema we can see than for formation of r_i elemental binary sequence B_{r_i} into shorter $B_{r_i}(r_2 < r_1)$ proceeds first by informatical – digital system [3].

We must say there are a lot of variations concerning the choice of construction of modification of information compression schemas.



Pic. 1. Schema of compression within one cycle

References

1. Nanobashvili N. Questions of the optimal compression of the discrete information within four-digit codding system. Cybernetics / N. Nanobashvili // Tbilisi University's works. – 1980. – T. 212 (2).
2. Shtarkov Y. M. General codes of Shannon / Y. M. Shtarkov // Problems of transmission of the information. – 1984. – T. 20. – № 3. – P. 3–16.
3. Witten I. H. «Arithmetic coding for data compression» Commun / I. H. Witten, R. M. Neal, J. C. Cleary // ACM. – 1987. – № 30 (6). – P. 520–540.

Дідманідзе І. Ш., Тхілаішвілі Р. Д.

ВИБІР СХЕМИ СТИСНЕННЯ

У роботі досліджено нижню межу схеми стиснення у відкритих інформаційних системах з використанням чотиризначної системи кодування. Доведено доцільність кодування в чотирисимвольному алфавіті.

Ключові слова: схеми стиснення, ітеративні структури, коефіцієнт стиснення.

Матеріал надійшов 31.07.2013

УДК 512.7

I. Didmanidze, G. Kakhiani

PREFORECAST TIME-SERIES ANALYSIS OF FINANCIAL DATA

The main result is the conclusion that the possibility of pre-forecast analysis of financial time series in order to prepare them for use in the prediction using neural networks.

Keywords: data series, neural networks, financial data, pre-forecast analysis.

In the article the pattern of use of fractal analysis to identify the basic characteristics of financial time series data, basic element of which is the ability to R/S-analysis.

According to the algorithm in Borland C++ Builder has developed a software product that allows you to identify and numerically evaluate the fundamental characteristics of the time series, such as the presence and depth of long-term memory, persistence or anti-persistence etc [2].

Fractal analysis is a new method to describe the evolutionary processes and forecasting of economic time series. The basic tool for the fractal analysis of time series analysis is an algorithm R/S-analysis. Methodology for R / S-analysis was developed in the mid XX century hydrologist Hurst during the study time series of river flow volumes. The inspection of the assumption that the data series are subject to the normal law, Hearst defined a new statistic – Hurst index (H). In the course of his research Hurst measure fluctuations of water in the reservoir relative to the average over time and introduced the dimensionless ratio by dividing the amplitude of R by the standard deviation S. This method of analysis has been called by the rescaled range (R / S-analysis). Hurst found that most natural phenomena, including river flows, temperatures, precipitation, sun spots should be "shifted to a random walk" – a trend with noise. The strength of the trend and the noise level can be measured by how the normalized amplitude with time, or in other words, as far as the value of H greater than 0,5.

We describe the algorithm for R / S-analysis in the form in which it is implemented in modern methods of fractal analysis [1; 2]. Given a time series: