Serving Object Information State Estimation

T. Dominyuk[1], Mgr., Post-Grad.St.; O.Ivakhiv[1], Ph.D., Doc; A. Kowalczyk[2], D.Sc.. Prof.; P. Mushenyk[1], St.; B.Stadnyk[1,2], D.Sc., Prof.

1) *National University «Lviv Polytechnic»* 12 Bandera str., Lviv, 79013, Ukraine
Phone: ***380-322-398619,*** Fax: ***380-322-744300,*** E-mail: ***oresti@polynet.lviv.ua***

2) *Rzeszow University of Technology,* 3B W.Pola str., Rzeszow, 35-959, Poland
Phone: ***48-17-8651231,*** Fax: ***48-17-8651575,*** E-mail: ***zmisp@prz.rzeszow.pl***

Abstract – **The multiplex analog sources time division system general informative effectiveness is investigated in this paper.**

Keywords – **multiplex measurement system, entropy, uncertainty, numeration, enumerative coding, permutation.**

Introduction

The investigated object or phenomena become more complex and functionally shared. The measurement system are used for their serving. There are some problems to need solving. The multichannel problem, the object state corrective estimation and the multiplex measurement system informative possibility researching.

The first one is solving by the numerous information technologies using such as the neuronal networks, expert system, fuzzy logic methodology, compression techniques etc. [1-4].

The second problem the best decision is corresponding with the object parameters entropy estimation [5]. The measurement system created for the observing object uncertainty decreasing. In the information theory sense, such uncertainty is described by the object entropy. Therefore, this entropy estimation mapping allows to describe the object information state as well. The expression for the multiplex system information capacity calculated depended on the measurement signals models and the measurement parameter precision measure using is found.

The third task realization is corresponding with the system operating information quantity calculation [6].

These problems are discussed in applications to the digital adaptive multiplex measurement system investigation. Such systems use the adaptive commutation or polynomial prediction compression [7]. Every one non-redundancy sample allows to receive all object parameters information activities estimation. It was found: the non-redundant samples identification marks sequences permutation coding give us the unique number corresponding with certain sequence. These numbers are the object parameters or phenomena information activities image convolutions and its entropy numerical estimations. Such compression estimation is reversible: the primary samples marks sequence may be presented back due to decoding procedure.

The totality system sources activities statistics is a priori well known.

In this case sources set, which shape sequence of the non-redundant samples and the amount isolated to any of them of positions in it are fixed. But on contrary to the regular type system the source signal sampling is random, corresponded with the signal current behavior. At the length $N >> 1$ the *i*-th non redundant selective values will occur in the sequence boundaries N_i times $\left(N_i = \alpha_i N\right)$. If it is compressed the information from *n* sources, then the sequence versions amount is $Q_0 = n^N$, as on the sequence any position can be a non redundant value of arbitrary of packed multiplexed sources. Nevertheless thanking to the convergence of the event frequency to its probability at a considerable experiences amount, for $N >> 1$. It is conditionally possible to arrange all sequences on two subgroups: typical, for which $N_i \approx \alpha_i N$ and atypical - for which this relation is variable [8]. Thus all typical sequences aggregate probability with the lengths N increasing will be guided to one, and atypical - to zero, so there will be only necessity of the typical sequences enumeration. From the mathematical point of view such *N*-positions sequences is the permutation with the repetitions [9, 10], in which each the *i*-th element address will occur in the sequences boundaries N_i times [11]. The different typical sequences number is as follow

$$Q = N!/\prod_i N_i! \tag{1}$$

and the corresponding each corresponding numbering we shall spend as

$$k_{frame} = \log_2 Q = \log_2 N! - \sum_i \log_2 M_i!$$

This number we shall term as a code of a disposition. Having taken advantage a Stirling formula for factorials $n! \approx r^n e^{-e}\sqrt{2\pi n}$, we shall note

$$ln(n!) = \frac{1}{2}ln 2\pi + \frac{1}{2}ln n + n(ln n - 1),$$

$$\lim_{n>>1}\{ln(n!)\} \approx n(ln n - 1)$$

and now

$$k_{frame} = k_Q = -N\sum_i \alpha_i \ln\alpha_i = \frac{-N\sum_i \alpha_i \log_2 \alpha_i}{\log_2 e} \approx$$

$$\approx -N\sum_i \alpha_i \log_2 \alpha_i = -NH(\alpha) \qquad (2)$$

On an disposition code enumeration of one sequence sample will come

$$k_Q = k_{frame} \,/ N = -\sum_i \alpha_i \log_2 \alpha_i \equiv H(\alpha)$$

The above obtained relation is by the upper estimation of number $\log_2 Q$, namely

$$-N\sum_{i=1}^{n} \alpha_i \,log_2\, \alpha_i > log_2 \frac{N!}{N_1!...N_n!},$$

As for the left-hand part of an inequality we have

$$-\sum_i N_i \,log_2 \frac{N_i}{N} = \sum_i log_2 (\frac{N}{N_i})^{N_i} = log_2 \prod_i \frac{(N)^{N_i}}{N_i} =$$

$$= log_2 \frac{\prod (N)^{N_i}}{\prod N_i} = log_2 \frac{(N)^N}{\prod_i N_i}. \qquad (3)$$

And analyzing the polynomial formula [11]

$$\left(X_1 + X_2 + \cdots + X_n\right)^N =$$

$$\sum_{K_1+K_2+\cdots+K_n=N} \frac{\left(N_1 + N_2 + \cdots + N_n\right)!}{N_1!N_2!\cdots N_n!} X_1^{N_1} X_2^{N_2} \cdots X_n^{N_n} \quad (4)$$

(here K_i - manifold numbers sets, subordinated by the requirement $\prod_i K_i = N$, for $i = \overline{1,n}$), we observed, that the different sets mediums are the natural numbers K_i there is the set, in which $K_i = N_i$ (that it is possible to realize, as $\sum_i N_i = N$). And this combination gives only one of possible addend of this sum. Therefore bodily is obvious, that the sum (3) will be more than it's non-negative composite, namely:

$$\left(X_1 + X_2 + \cdots + X_n\right)^N > \frac{\left(N_1 + N_2 + \cdots + N_n\right)!}{N_1!N_2!\cdots N_n!} \times$$

$$\times X_1^{N_1} X_2^{N_2} \cdots X_n^{N_n}.$$

If the equation $X_i = N_i$ is also accepted, then we shall have as follow

$$\left(N_1 + N_2 + \cdots + N_n\right)^N > \frac{\left(N_1 + N_2 + \cdots + N_n\right)!}{N_1!N_2!\cdots N_n!} \times$$

$$\times N_1^{N_1} N_2^{N_2} \cdots N_n^{N_n} \text{ or } \frac{N^N}{\prod_i N_i^{N_i}} > \frac{N!}{\prod_i N_i!}. \qquad (5)$$

Having compared Eq. (2) and (4), at the logarithm basis more than one, there is true to note

$$\log_2 \frac{N^N}{N_1^{N_1}\ldots N_n^{N_n}} > \log_2 \frac{N!}{N_1!\ldots N_n!},$$

as it was necessary to prove.

Totality sources activity statistics is unknown

In this case the sequence set and its positions, distribution between separate source samples are not fixed. Therefore besides an arrangement code by a size k_Q , it is necessary to give also the information about its length by a size k_q and sequence sources set by a size $k_н$, as the activity distribution $\{\alpha_i\}$ a priori are not known. That is, the sequence service information

$$k_{frame} = k_Q + k_q + k_H. \qquad (6)$$

But in practice for the frame synchronization facilitation it is necessary to work with sequences of a stationary value of length, in particular from stationary values both information, and service parts. As against the parsed above structure it is possible, shaping a service part, thus to do without only by the set and disposition codes, which digit capacity $k_н$ and k_Q, accordingly, are fixed.

Let's assume, that the sequence shaping ceases at the sequence information part or the disposition code digit grid filling.

1) Let the first the sequence information part is filled. The set code length $k_н$ is determined by an information part length value N, which should be enough large, nevertheless from physical requirements, restricted. Its value should be such, that least fissile of totality sources has filled even in one position of an sequence information part – from one, and limits by a value practically implemented digit grids of a disposition code k_Q - on the other hand. So,

$$1/\alpha_{i\min} \le N_{\max} \le k_Q / H(\alpha). \qquad (7)$$

The probability of the of set code fast filling, than a disposition code, coincides probability that an sequence information part length exceeds selected one. It is neglectfully small, because a set code length k_Q optimality is close to an entropy. So:

$$Pr\{N > N_{max}\} = Pr\left\{\left[H(\alpha) - \frac{k_a}{N}\right] > \Delta\right\} < \varepsilon,$$

where ε and Δ - arbitrary, certainly given positive small values.

The disposition code binary digits number, which is necessary on one selective sample value, is determined by a totality sources activities entropy, so, will be maximum at the equiprobable activities distribution. From here at a given disposition code length is possible to receive the sequence information part selective values amount, at which there will be already its filling

$$N_{min} = \frac{k_Q}{log_2 n}. \quad (8)$$

2) The first the set code digit grid is filled. If a sequence information part positions thus yet are not exhausted, then the residual of positions can be allocated on the "shadow" interrogation or other additional information.

At a set code shaping similarly to Eq (1) here it is necessary to enter one more numeral [13] – boundary, if only to designate the unused sequence information part positions, that is

$$k_H = log_2 \frac{(N+n)!}{N!n!} \approx N log_2 \frac{N+n}{N} + n log_2 \frac{N+n}{n} \quad (9).$$

The limiting set code length is determined by the greatest possible sequence information part selective samples values $N_{\max}$

$$k_{H\ \max} = N_{\max} \log_2 \frac{N_{\max}+n}{N_{\max}} + n\log_2 \frac{N_{\max}+n}{n}.$$

Expenditures, which have on one selective sampling value are as follow

$$k_H = \frac{k_{H\ \max}}{N} < \frac{k_{H\ \max}}{N_{\min}}.$$

Taking into account relations Eq. (6) and (7), we shall show an asymptotic optimality of the given approach Eq. (8), i.e.

$$\lim_{k_Q\to\infty} k_H = \lim_{k_Q\to\infty} \frac{log_2 n}{k_Q}\left\{\frac{k_Q}{H(\alpha)} log_2\left[1+\frac{nH(\alpha)}{k_Q}\right] + n\, log_2\left[1+\frac{k_Q}{nH(\alpha)}\right]\right\} = 0.$$

Thus, this an enumeration method is asymptotic effective at a sequence considerable size of N, as the minimum possible expenditures on an enumeration can not be less than entropy value [5, 14]. Such approach may be used with intelligent devices designing [15].

Serving object entropy estimation uncertainty evaluation

The exactitude of entropy estimation depends on exactitude of the measuring information source activity determination. A relative error value of entropy estimation may be found [16]

$$\delta_{H(a_i)} = \frac{\Delta H(a_i)}{H(a_i)} = \frac{1}{H(a_i)}\left[\sum_{j=1}^{n} \frac{\partial}{\partial a_j} H(a_i) a_j \delta_{aj}\right].$$

At first we shall consider partial derivatives, namely:

$$\sum_{j=1}^{n} \frac{\partial}{\partial a_j} H(a_i) = \frac{\partial}{\partial a_j}\left(-\sum_{i=1}^{n} a_i \ln a_i\right) = -\frac{\partial}{\partial a_j}\left[\ln\left(\prod_{i=1}^{n} a_i^{a_i}\right)\right] = -\left(1+\ln a_j\right).$$

So, now

$$\delta_{H(a_i)} = \frac{(-1)}{H(a_i)}\left\{\sum_{j=1}^{n}\left[1+\left(\ln a_j\right)\right] a_j \delta_{\alpha j}\right\} = \frac{(-1)}{H(a_i)}\left\{\sum_{j=1}^{n} a_j \delta_{aj} + \sum_{j=1}^{n}\left(a_j \ln a_j\right)\delta_{aj}\right\}$$

If relative errors of a source activity index estimation (i. e. activities numbers detection) to accept identical to all system sources, then

$$\delta_{H(a_i)} = \frac{(-1)}{H(a_i)}\left\{\delta_a \sum_{j=1}^{n} a_j - \delta_a H(a_j)\right\} = \frac{\delta_a\left[H(a_j)-1\right]}{H(a_i)},$$

here $\delta_{\alpha j} = \delta_a$ for all sources ($i = \overline{1,n}$).

If a relative error of an activity index estimation for any of system sources to accept inverse to this index, that is $a_j \delta_{aj} = c$ (for all $i = \overline{1,n}$), then

$$\delta_{H(a_i)} = \frac{(-1)}{H(a_i)}\left\{c\left[n\left(1+\log_2 c\right) - \sum_{j=1}^{n} \log_2 \delta_{aj}\right]\right\}.$$

In turn, exactitude of an activity estimation is connected with an amount of probes, that is amount of the samples in the sequence *N*. The amount of the *i*-th sources activity demonstration is random and submits to a binomial distribution law [7]

$$Pr(N_i) = \binom{N}{N_i} a_i^{N_i} \left(1-a_i\right)^{N-N_i}$$

With mean and variance value, accordingly $\overline{N_i} = N a_{иi} = N_{иi}$ and $d_i^2 = N a_{иi}\left(1-a_{иi}\right)/N^2 = N_{иi}\left(1-a_{иi}\right)/N^2$.

If an amount of the in sequence samples considerable (near hundred or more), a source activity index a_i is not rather large and not rather small, namely: $N(1-a_i)$ and Na_i is more 4, then behind the Moivre-Laplace theorem the binomial distribution can be approximated by normal.

And then the activity index estimation relative guaranteed error value may be found as

$$\delta_{ai} = \frac{e}{a_i} = \frac{t_p}{a_i}\sqrt{\frac{a_i(1-a_i)}{N}} = t_p\sqrt{\frac{(1-a_i)}{a_i N}}$$

here e and t_p – the absolute error value and the guaranteed coefficient, accordingly.

Generally for arbitrary index value a_i we should remember, that a binomial distribution asymmetric and discrete. Therefore amount of developments of activity demonstrations the *i*-th source with the probability p is determined as $\Pr\left\{\left|N_i - \overline{N_i}\right| \le e\right\} = p$ or $\Pr\{N_{i1} < N_i < N_{i2}\} = p$, here $e = 0.5(N_{i2} - N_{i1})$ – neglectfully small positive number.

A probability value *p*, the sequence length *N* and the activity index a_i determine boundaries of an uncertainty interval of indeterminacy for an activity manifestation amount. For selected probability and admissible uncertainty value e the necessary sequence length *N* can be established behind equations [17,18]

$$\begin{cases} \sum_{m=N_i}^{N} C_N^m a_i^m (1-a_i)^{N-m} = \frac{1}{2}(1-p) \\ \sum_{m=1}^{N_i} C_N^m a_i^m (1-a_i)^{N-m} = \frac{1}{2}(1-p) \end{cases},$$

For the activity manifestation estimation a_i^* the mean and the variance, accordingly, are as follow a_i and $\left(s_i^*\right)^2 = \frac{a_i(1-a_i)}{N}$. In particular, $N_i = a_i^* N$ and for an activity index $a_i = 0.48$, probability value $p = 0.9$ and admissible uncertainty $e = 0.1$, the necessary sequence length *N* makes 100.

Conclusions

The binary digit rate demands at the sequence length number $N \to \infty$ is guided to the group source entropy productivity. The code number is carried out a uniform unique code. At the receiver side the sequence notes in the register, and the decoded number is used for the samples distribution from the memory register on the relevant channels. Thus the sequence number can serve also generalized an object information condition performance, or as the separate channel element estimation for the hierarchy higher degree next representation.

References

1. Holland J.H. Adaptation in Natural and Artificial System. – Cambrige, Massachuset: MIT Press, 1992. – 257 p.
2. Zadeh L.A. Fuzzy sets // Information and Control. – August 1965. – № 8. – P. 338–353.
3. Negoita C.V., Ralescu D.A. Application of fuzzy sets to systems analysis. –Basel: Birkhauser, 1975. – 321 p.
4. Brunuk S. and Lautrup B. Neural Networks. Computers with Intuition. – Singapore: World Scientific, 1990. – 540 p.
5. *Cover T., Thomas J. Elements of Information Theory. John Wiley & Sons.Inc.- New York/Chichester/Brisbane/Toronto/Singapore.-1991.- 542 p.*
6. *Kolesnik V., Poltyrev G. Information theory introduction (sources coding). - Leningrad: Leningrad University Press, 1980, 129 p.*
7. Kalashnikov I.,Stepanov V., Churkin A.: Adaptive Data Gathering and Transmitting Systems. Energiya, Moscow, 1975 (in Russian), 240 p.
8. *Oliver B. M. Efficient Coding // Bell System Technical Journal. - 31, 4, 1952. - P. 724-750.*
9. *John K.Karlof. "Permutation Codes for the Gaussian Channel" // IEEE Transactions on Information Theory, 1989, vol. 35, no. 4, pp. 726-732.*
10. Slepian D. Permutation modulation, Proc. IEEE, 1965, vol. 53, no. 3, pp. 228–236.
11. *Vilenkin N. Combinatorics. - Moscow: Nauka Press, 1969, 197 p. (in Russian).*
12. *Ivakhiv O. Measurement source enumerative coding // Kontrolno-vymiriuvalna technika, № 50, 1993, p.12-22 (in Ukrainian).*
13. *US Patent, №31369.222, 340-172.5. Data Compressor. Webb James Administration of the National Aeronautics and Space Administration with respect to invention of E. C. Posner, I. Eisenberger, T. O. Anderson and W. A. Lushbaugh..*
14. *Behota R., Ivakhiv O., Kushnir Z., Tolopko Y. Enumerative source coding asymptotic optimality // Kontrolno-izmeritelnaya technika, № 46, 1989, p. 3-9. (in Russian).*
15. O.Ivakhiv, A.Kowalczyk, R. Velgan. "Intelligent Programmable Measurement System" Proceedings of the XVI IMEKO World Congress (September 25-28, 2000, Vienna-Wien, Austria). Volume IX, topic 30 - Artificial Intelligence in Measurement Techniques, p.341-345, Vienna, 2000.
16. Ivakhiv O. Measurement Experiment Results Compressed Representation // Measurement instrumentation and Metrology № 58, 2001, p. 20-25.
17. *Ventsel E. Probability theory . - Moscow: Nauka Press, 1968, 576 p. (in Russian).*
18. Dunin-Barkovsky I., Smirnov H. Probability Theory and Mathematical statistics technique. – Moscow: Gostechizdat, 1955. – 556 p.