

Method of cluster statistical coding of information resource data

Vladimir Barannik^a, Ivan Tupitsya^b, Anna Hahanova^b, Maksym Parkhomenko^c and Oleksandr Ihnatiev^c

^a V.N. Karazin Kharkiv National University, 4 Svobody Sq., Kharkiv, 61022, Ukraine

^b Ivan Kozhedub Kharkiv National Air Force University, Sumska Str. 77/79, Kharkiv, 61000, Ukraine

^c Kharkiv National University of Radio Electronics, Nauky avenue 14, Kharkiv, 61166, Ukraine

Abstract

A method of cluster statistical coding of information resource data using internal data restructuring is being developed. The analysis of transformations of the nature of the law of distribution of probabilities of the appearance of elements in a message as a result of clustering by the number of series of units is carried out. The factors that have a significant impact on the effectiveness of the developed method of coding an information resource are analyzed from the standpoint of additional reduction of the structural redundancy of the output code sequence.

Keywords 1

Transformation, restructuring, clustering, indication of the number of series of units, coding, information resource data.

1. Introduction

Existing data coding algorithms for information resources actively use methods of external data restructuring for a more advantageous presentation of the encoded data [1-10]. Analysis of methods of external restructuring indicates the presence of significant shortcomings [11-32]. Therefore, a fundamentally new approach was proposed - the internal restructuring of data, the essence of which is to identify patterns in the internal binary structure of message elements by quantitative criteria [33]. A conceptual model was developed to form a quantitative feature [34-37]. The analysis of this model indicates the advantages of using the combined approach in the process of forming a quantitative feature [38]. In works [39, 39] as a tool for data restructuring, the choice of the characteristic of the number of series of units was justified. Further, to assess the efficiency of data coding from the standpoint of additional reduction of the structural redundancy of the code sequence. it is proposed to develop a method for coding information resource data in the statistical space of sets that are formed in the process of clustering.

2. Method of cluster statistical coding of information resource data

As a result of clustering, the elements u_{ξ} original message $U(\theta)$ with the same quantity characteristic values λ_i series of units (SU) are combined into sets:

Proceedings Name, Month XX-XX, YYYY, City, Country

EMAIL: vvbar.off@gmail.com (A. 1); ivan20081982@gmail.com (A. 2); ann.hahanova@gmail.com (A. 3), maxpar76@gmail.com (A. 4); vika.iv55@gmail.com (A. 5)

ORCID: 0000-0002-2848-4524 (A. 1); 0000-0001-6806-4914 (A. 2); 0000-0002-1318-7973 (A. 3)

0000-0001-6062-7743 (A. 4); 0000-0003-1227-6840 (A. 5);



© 2020 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

$$U(\theta) \xrightarrow{f_{cl}} \{U(\lambda_1); \dots; U(\lambda_i); \dots; U(\lambda_n)\}, \quad (1)$$

where $f_{cl}(u_{\xi}, \lambda_i)$ - function of grouping elements u_{ξ} by sets $U(\lambda_i)$ by the value of the quantity attribute λ_i SU.

Set of different values for a quantity characteristic λ_i SU, is described by the following expression:

$$\Lambda = \{\lambda_1, \dots, \lambda_i, \dots, \lambda_n\}, \quad (2)$$

where Λ - a set of values for the quantity characteristic λ_i SU;

λ_i, λ_n - values i -th and n -th sign of the set Λ .

As a result of clustering, the elements u_{ξ} messages $U(\theta)$ with the same quantity characteristic values λ_i SU unite into sets $U(\lambda_i)$, i.e.:

$$u_{\xi,i} \in U(\lambda_i).$$

Further, it is proposed to carry out coding of elements $u_{\xi,i}$ in the statistical space of sets $U(\lambda_i)$.

So statistical coding of elements $u_{\xi,i}$ messages $U(\lambda_i)$ which consists of κ - elements ($\xi = \overline{1, \kappa}$) is given by the following expression:

$$U(u_{\xi,i}) \xrightarrow{f_{vlc}} L'(\kappa), \quad \xi = \overline{1, \kappa}, \quad (3)$$

where $f_{vlc}(u_{\xi,i}, P(u_{\xi,i}))$ - function of generating statistical code $\ell'_{\xi,i}$ variable length for elements $u_{\xi,i}$ multitudes $U(\lambda_i)$.

Here, we use overhead information about the distribution of probability values $P(u_{\xi,i})$ appearance of elements $u_{\xi,i}$ in message $U(\lambda_i)$. Given this information based on the function $f_{vlc}(u_{\xi,i}, P(u_{\xi,i}))$ code is generated $\ell'_{\xi,i}$. This function is described by the following formula:

$$\ell'_{\xi,i} = f_{vlc}(u_{\xi,i}, P(u_{\xi,i})), \quad (4)$$

where $\ell'_{\xi,i}$ - variable length code for elements $u_{\xi,i}$ sets $U(\lambda_i)$;

$P(u_{\xi,i})$ - the probability of the appearance of elements $u_{\xi,i}$ in sets $U(\lambda_i)$.

As a result of statistical coding of elements $u_{\xi,i}$ multitudes $U(\lambda_i)$ a sequence (set) is formed $L'(\kappa)$ codes $\ell'_{\xi,i}$ which looks like this:

$$L'(\kappa) = \{\ell'_{1,i}; \dots; \ell'_{\xi,i}; \dots; \ell'_{k,i}\}, \quad (5)$$

where κ - the number of elements in the sequence of codes $L'(\kappa)$.

Code $\ell'_{\xi,i}$ is a variable length code $|\ell'_{\xi,i}|_2$,

$$|\ell'_{\xi,i}|_2 = VAR,$$

and consists of the sequence $[\ell'_{\xi,i}]_2$ binary digits $q_{\xi,\gamma}$, $\gamma = \overline{1, |\ell'_{\xi,i}|_2}$. It is set like this:

$$[\ell'_{\xi,i}]_2 = \{q_{\xi,1}; \dots; q_{\xi,\alpha}; \dots; q_{\xi,|\ell'_{\xi,i}|_2}\}, \quad (6)$$

where $q_{\xi,\alpha}$ - α -th digit of code $\ell'_{\xi,i}$.

3. Analysis of transformations of the nature of the probability distribution law for the appearance of elements in a message as a result of clustering by the number of series of units

Analysis of transformations of the nature of the probability distribution law for the appearance of message elements as a result of clustering based on the number of series of units.

As a result of clustering elements u_{ξ} messages $U(\theta)$ by sign λ_i SU, the following variants of change (transformation) of the nature of the law of distribution of the probabilities of the appearance of elements may occur (Fig. 1):

1. Sets can be formed that will have the nature of the distribution laws of elements similar to the original message, the difference will be observed in an increase in the dynamic range of values of the probability of occurrence of elements due to a decrease in the power of the encoded data. So, in the process of clustering elements u_ξ messages $U(\theta)$ by quantity λ_i SU, so many can sets $U(\lambda_i)$ elements $u_{\xi,i}$ for which the following expression will be valid:

$$P(u_{\xi,i}) \rightarrow P(u_\xi), \quad (7)$$

where $P(u_{\xi,i})$ - the probability of the item appearing $u_{\xi,i}$ in set $U(\lambda_i)$;

$P(u_\xi)$ - the probability of the item appearing u_ξ in the message $U(\theta)$.

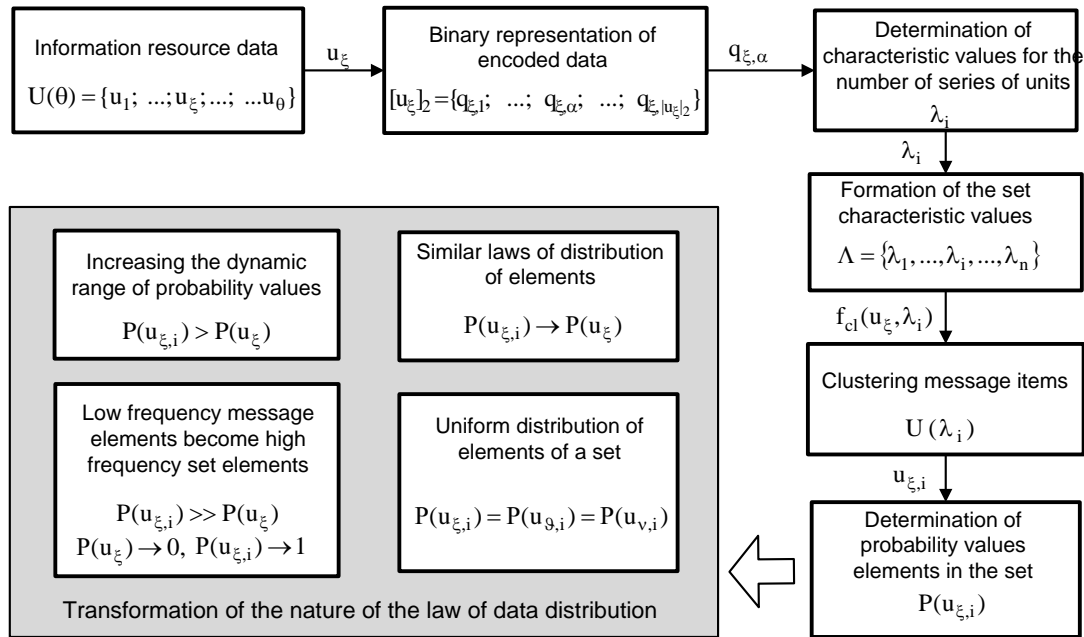


Figure 1: Structural and functional diagram of the transformation of the nature of the distribution law of message elements as a result of clustering by the number of series of units

This variant of transformation of the nature of the law of distribution of elements is possible under the following conditions:

- when in the process of clustering low-frequency elements u_ξ messages $U(\theta)$ (elements u_ξ with the lowest probabilities $P(u_\xi)$ appearing in the message $U(\theta)$) form separate sets $U(\lambda_i)$;
- the main part of the elements u_ξ messages $U(\theta)$ which includes high frequency elements u_ξ messages $U(\theta)$, forms one set $U(\lambda_i)$.

This means that if, during the clustering process, the elements u_ϑ and u_ν messages $U(\theta)$, the values of the probabilities of which in the message $U(\theta)$ tend to zero, form one set, then for the rest of the message elements $U(\theta)$, which form one set, the statement that is given by expression (7) will be true, i.e.:


$$\text{if } P(u_\vartheta) \rightarrow 0, P(u_\nu) \rightarrow 0 \text{ where } u_\vartheta, u_\nu \in U(\lambda_j) \text{ then } P(u_{\xi,i}) \rightarrow P(u_\xi), u_{\xi,i} \in U(\lambda_i),$$

where u_ϑ, u_ν - ϑ th and ν th message elements $U(\theta)$;

$P(u_\vartheta), P(u_\nu)$ - probability of occurrence ϑ -th and ν -th items in the message $U(\theta)$;

$U(\lambda_j)$ - the set into which they unite ϑ -th and ν -th message elements $U(\theta)$ in the process of clustering.

For the analyzed case, the transformation of the distribution law of elements u_ξ messages $U(\theta)$ in the process of clustering by the number of series of units is described by the following system of expressions:

 - elements $u_{\xi,3}$ that belong to the set $U(\lambda_3)$, $\lambda_3=4$.

For the analyzed example, the elements u_{ξ} block U_{τ} the following values of the quantity characteristic correspond λ_i SU:

$$\Lambda = \{2,3,4\}$$

Accordingly, as a result of clustering by the quantity λ_i SU, elements u_{ξ} block U_{τ} combine into 3 sets $U(\lambda_i)$, i.e.:

$$N(U(\lambda_i)) = 3.$$

Thus, the result of clustering elements u_{ξ} analyzed block U_{τ} by quantity λ_i SU is the formation of three sets $U(\lambda_i)$, which are described by the following expression system:

$$U_{\tau} \xrightarrow{f_{cl}} \begin{cases} U(\lambda_1), \lambda_1 = 2, |U(\lambda_1)| = 4; \\ U(\lambda_2), \lambda_2 = 3, |U(\lambda_2)| = 59; \\ U(\lambda_3), \lambda_3 = 4, |U(\lambda_3)| = 1. \end{cases} \quad (10)$$

Block diagram of the transformation (transformation) of the nature of the probability distribution law $P(u_{\xi})$ appearance of elements u_{ξ} in the block U_{τ} , as a result of clustering on a quantitative basis λ_i , is shown in Fig. 3.

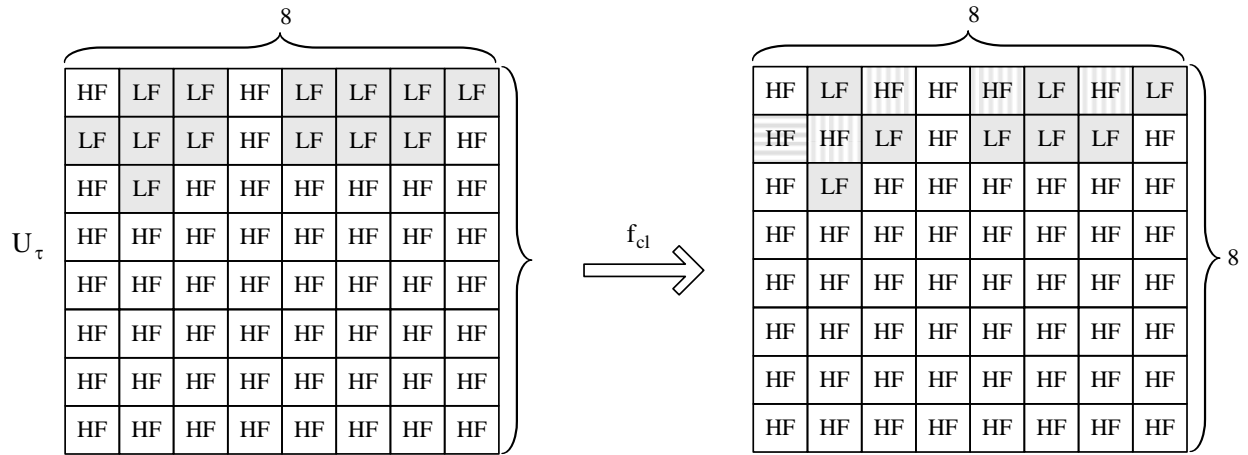






Figure 3: Block diagram of the transformation of the nature of the probability distribution law $P(u_{\xi})$ appearance of elements u_{ξ} in the block U_{τ}

In Fig. 3, the following designations are adopted:

 low-frequency elements u_{ξ} block U_{τ} ;

 high frequency elements u_{ξ} block U_{τ} ;

 elements u_{ξ} block U_{τ} , which, after applying clustering, become high-frequency elements of the set $U(\lambda_i)$, $\lambda_i = 2$.

 elements u_{ξ} block U_{τ} , which, after applying clustering, become high-frequency elements of the set $U(\lambda_i)$, $\lambda_i = 3$.


 elements u_{ξ} block U_{τ} , which, after applying clustering, become high-frequency elements of the set $U(\lambda_i)$, $\lambda_i = 4$.

Table 1. the results of transformation of the probability distribution law are presented $P(u_{\xi})$ appearance of elements u_{ξ} in the block U_{τ} after clustering by quantity λ_i SU.

Table 1

Results of transformation of the nature of the probability distribution law $P(u_\xi)$ appearance of elements u_ξ block U_τ after clustering

ξ	u_ξ	$[u_\xi]_2$	λ_i	$f(u_\xi)$	$P(u_\xi)$		ℓ_ξ		$ \ell_\xi _2$, bit		$\Delta \ell_\xi _2$, bit
					$U(\lambda_i)$	U_τ	$U(\lambda_i)$	U_τ	$U(\lambda_i)$	U_τ	
1	117	01110101	3	1	0.016949	0.015625	011100	1011110	6	7	1
2	125	01111101	2	1	0.25	0.015625	11	101001	2	6	4
3	136	10001000	2	1	0.25	0.015625	10	100110	2	6	4
4	138	10001010	3	1	0.016949	0.015625	011101	100101	6	6	0
5	140	10001100	2	1	0.25	0.015625	01	100100	2	6	4
6	144	10010000	2	1	0.25	0.015625	00	1011111	2	7	5
7	146	10010010	3	1	0.016949	0.015625	01000	101011	5	6	1
8	148	10010100	3	4	0.067797	0.0625	0110	1000	4	4	0
9	150	10010110	3	1	0.016949	0.015625	011111	101010	6	6	0
10	157	10011101	3	1	0.016949	0.015625	011110	100111	6	6	0
11	162	10100010	3	1	0.016949	0.015625	01001	101100	5	6	1
12	164	10100100	3	1	0.016949	0.015625	01010	101101	5	6	1
13	169	10101001	4	1	one	0.015625	0	101000	1	6	5
14	172	10101100	3	1	0.016949	0.015625	01011	101110	5	6	1
15	175	10101111	3	20	0.338983	0.3125	11	11	2	2	0
16	177	10110001	3	17	0.288136	0.265625	10	01	2	2	0
17	180	10110100	3	10	0.169492	0.15625	00	00	2	2	0

Analyzing the results shown in Table 1 the following conclusions can be drawn:

1. As a result of applying element clustering u_ξ block U_τ by quantity λ_i SU bulk of elements u_ξ block U_τ which includes high frequency elements u_ξ , forms a separate set $U(\lambda_i)$, which is given by the following expression:

$$U(\lambda_2) \in \{u_\xi | \lambda_2 = 3\}.$$

In this case, the result of using clustering is an insignificant decrease in power $|U_\tau|$ block U_τ . So the cardinality of the set $U(\lambda_2)$ elements $u_{\xi,i}$ with feature value λ_i equal to 3 has the following meaning:

$$|U(\lambda_2)| = 59.$$

Considering the factor that the power $|U_\tau|$ block U_τ as a result of clustering elements u_ξ quantitatively λ_i decreased slightly (by 7.8% for many $U(\lambda_2)$) a situation arises when the probability distribution law $P(u_{\xi,i})$ appearance of elements $u_{\xi,i}$ in set $U(\lambda_i)$ has a similar (identical) character with the probability distribution law $P(u_\xi)$ appearance of elements u_ξ in the block U_τ , i.e.:

$$P(u_{\xi,i}) \rightarrow P(u_\xi).$$

It should be noted that the dependence of the probability $P(u_{\xi,i})$ element appearance $u_{\xi,i}$ in many $U(\lambda_i)$ from power changes $|U_\tau|$ block U_τ as a result of clustering is given by the following relationship:

$$\frac{|U_\tau|}{|U(\lambda_i)|} = \frac{P(u_\xi)}{P(u_{\xi,i})}, \quad (11)$$

where $|U_\tau|$ - block power U_τ ;

$|U(\lambda_i)|$ is the cardinality of the set $U(\lambda_i)$.

That is, the less power $|U(\lambda_i)|$ sets $U(\lambda_i)$, which is formed during the clustering of elements u_ξ by quantity λ_i SU, the greater the probability value $P(u_{\xi,i})$ occurrence of elements in a set $U(\lambda_i)$. As a consequence, with increasing probability $P(u_{\xi,i})$ shortening $|\ell'_{\xi,i}|_2$ code $\ell'_{\xi,i}$ which is assigned to the element $u_{\xi,i}$.

Thus, the result of the application of the developed coding method is the reduction in the length of code structures that are assigned to message elements as a result of coding in the statistical space of sets:

1. For items with a quantity characteristic value λ_i SU equal to 4 at 83.3%.
2. For items with a quantity characteristic value λ_i SU equal to 3 - from 14.3 to 16.7%.
3. For items with a quantity characteristic value λ_i SU equal to 2 - from 66.7 to 71.4%.
4. Accordingly, on average for the analyzed message by 25.6%.

Expression (11) will be valid not only for the case when the probability distribution law $P(u_{\xi,i})$ appearance of elements $u_{\xi,i}$ in many $U(\lambda_i)$ has a similar (identical) character with the probability distribution law $P(u_\xi)$ appearance of elements u_ξ in the block U_τ , and in all other cases of transformation (transformation).

So, for the analyzed example, high-frequency elements u_ξ block U_τ have the following probability values $P(u_\xi)$:

$$P(u_8) = 0,0625, P(u_{15}) = 0,3125, P(u_{16}) = 0,2656, P(u_{17}) = 0,1562.$$

The same elements u_ξ in many $U(\lambda_i)$ correspond to such values of probabilities $P(u_{\xi,i})$:

$$P(u_{8,2}) = 0,067797, P(u_{15,2}) = 0,338983, P(u_{16,2}) = 0,288136, P(u_{17,2}) = 0,169492.$$

Thus, we can conclude that the value of the probability $P(u_{\xi,i})$ occurrence of an element in a set $U(\lambda_i)$ tends to the value of probability $P(u_\xi)$ element appearance u_ξ in the block U_τ , i.e. expression (11) becomes valid for this case.

2. As a result of applying element clustering u_ξ block U_τ by quantity λ_i SU part of low-frequency elements u_ξ (elements u_ξ with the lowest probabilities $P(u_\xi)$) block U_τ formed separate sets $U(\lambda_i)$ which are given by the following expressions:

$$U(\lambda_1) \in \{u_2, u_3, u_5, u_6 \mid \lambda_1 = 2\}, U(\lambda_3) \in \{u_{13} \mid \lambda_3 = 4\}.$$

3. Power $|U(\lambda_i)|$ sets $U(\lambda_i)$ low-frequency elements u_ξ much lower power $|U_\tau|$ block U_τ , i.e.:

$$|U(\lambda_1)| \ll |U_\tau|, |U(\lambda_3)| \ll |U_\tau|.$$

Thus, power reduction $|U_\tau|$ block U_τ led to the transformation of the probability distribution law $P(u_\xi)$ appearance of elements u_ξ .

So, for the analyzed example, low-frequency elements u_ξ block U_τ - elements u_ξ with the following probabilities $P(u_\xi)$:

$$P(u_2) = P(u_3) = P(u_5) = P(u_6) = P(u_{13}) = 0,015625,$$

became high-frequency elements of sets $U(\lambda_i)$:

$$P(u_{2,1}) = P(u_{3,1}) = P(u_{5,1}) = P(u_{6,1}) = 0,25, P(u_{13,3}) = 1.$$

Thus, the next variant of changing the nature of the probability distribution law $P(u_\xi)$ appearance of elements u_ξ in the message $U(\theta)$, as a result of the application of clustering, is its essential transformation for individual sets $U(\lambda_i)$.

2. Sets $U(\lambda_i)$ can form, for which the nature of the probability distribution law changes significantly $P(u_{\xi,i})$ appearance of elements $u_{\xi,i}$.

This type of transformation becomes possible when the low-frequency element u_ξ messages $U(\theta)$ becomes a high frequency element $u_{\xi,i}$ sets $U(\lambda_i)$. Thus, if the probability value $P(u_\xi)$ element appearance u_ξ in the message $U(\theta)$ tends to zero i.e.:

$$P(u_\xi) \rightarrow 0,$$

then as a result of clustering by the quantity λ_i SU element u_ξ can form such a set $U(\lambda_i)$, in which the probability value $P(u_{\xi,i})$ the appearance of the element will tend to one, i.e.:

$$P(u_{\xi,i}) \rightarrow 1.$$

A visual representation of a significant transformation of the nature of the probability distribution law $P(u_\xi)$ appearance of elements u_ξ (element transition u_ξ from the low-frequency component of the block U_τ to the high-frequency component of the set $U(\lambda_i)$) is shown in the third version of Fig. 4.

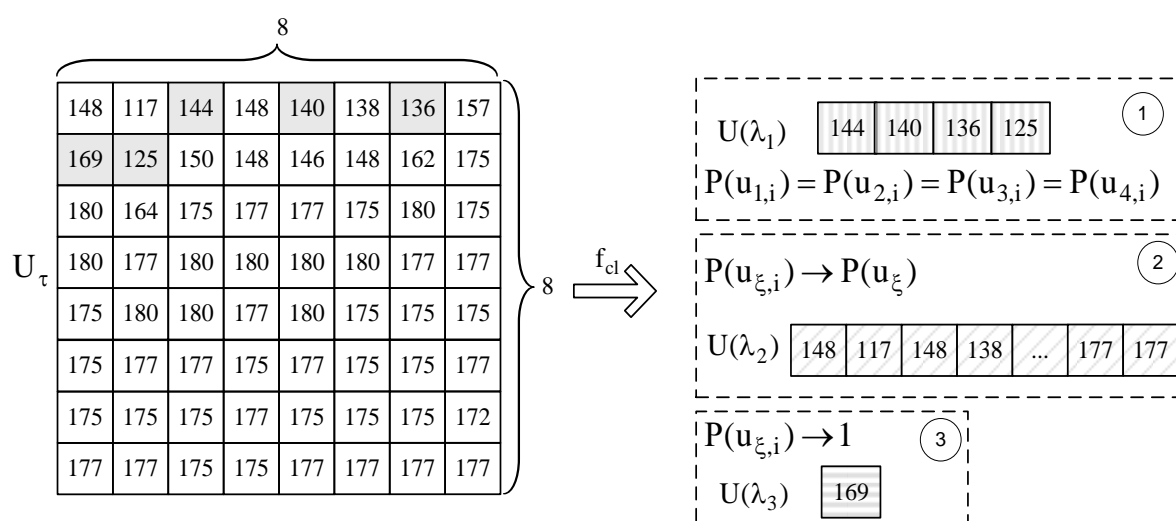


Figure 4: Schematic illustration of possible transformations of the nature of the probability distribution law $P(u_\xi)$ appearance of elements u_ξ

In Fig. 4, the following designations are adopted:

- low-frequency elements u_ξ block U_τ , for which the nature of the probability distribution law $P(u_\xi)$ in the process of clustering, it undergoes significant transformation;

- elements u_ξ block U_τ , for which the nature of the probability distribution law $P(u_\xi)$ does not change during clustering;

① - option of equiprobable redistribution of elements u_ξ block U_τ in the process of clustering;

② - option to redistribute elements u_ξ block U_τ into a set with identical nature of the probability distribution law $P(u_\xi)$;

③ - a variant of a significant transformation of the nature of the probability distribution law $P(u_\xi)$;

- elements u_ξ that belong to the set $U(\lambda_1)$, $\lambda_1 = 2$;

- elements u_ξ that belong to the set $U(\lambda_2)$, $\lambda_2 = 3$;

- elements u_ξ that belong to the set $U(\lambda_3)$, $\lambda_3 = 4$.

Figure 4 shows possible variants of change (transformation) in the nature of the probability distribution law $P(u_\xi)$ appearance of elements u_ξ messages $U(\theta)$ as a result of using clustering by quantity λ_i SU, namely:

- option of equiprobable redistribution of elements u_ξ block U_τ in set $U(\lambda_i)$, which is formed in the course of clustering (the first option in Fig. 4).

Thus, the elements u_ϑ and u_ν messages $U(\theta)$ (which appear in the message $U(\theta)$ with different or the same frequency), provided that in the process of clustering they are combined into one set $U(\lambda_i)$ (i.e. have the same values of the quantity λ_i SU), an equiprobable distribution law in the set $U(\lambda_i)$, i.e.:

$$\text{if } P(u_\vartheta) \neq P(u_\nu) \text{ and } u_\vartheta, u_\nu \in U(\lambda_i), \text{ then } P(u_{\vartheta,i}) = P(u_{\nu,i}),$$

where $P(u_\vartheta), P(u_\nu)$ - probability of occurrence ϑ -th and ν -th items in the message $U(\theta)$;

$$P(u_{\vartheta,i}), P(u_{\nu,i}) - \text{probability of occurrence } \vartheta\text{-th and } \nu\text{-th elements in the set } U(\lambda_i).$$

So in the analyzed example, the elements u_2, u_3, u_5, u_6 corresponds to the following probability value $P(u_\xi)$ appearances in the block U_τ :

$$P(u_2) = P(u_3) = P(u_5) = P(u_6) = 0,015625.$$

As a result of clustering elements u_ξ block U_τ on the basis of λ_i equal to 2 (i.e. $\lambda_i = 2$) elements u_2, u_3, u_5, u_6 combine into one set $U(\lambda_i)$, i.e.:

$$\{u_2, u_3, u_5, u_6 \mid \lambda_i = 2\} \in U(\lambda_i).$$

The law of distribution of the probabilities of the appearance of elements u_2, u_3, u_5, u_6 in many $U(\lambda_i)$ is equally likely. This means that the elements $u_{\xi,i}$ in set $U(\lambda_i)$ appear with the same probability values $P(u_{\xi,i})$, i.e.:

$$P(u_{2,i}) = P(u_{3,i}) = P(u_{5,i}) = P(u_{6,i}) = 0,25, \quad i = 1.$$

- a variant of a significant (significant) transformation of the nature of the probability distribution law $P(u_\xi)$ appearance of elements u_ξ (element transition u_ξ from the low-frequency component of the block U_τ to the high-frequency component of the set $U(\lambda_i)$) is shown in Fig. 4 (third option). This option becomes relevant in the following cases:

- if in the case of clustering by quantity λ_i SU low frequency element u_ξ block U_τ forms a separate set $U(\lambda_i)$.

This means that if the probability value $P(u_\xi)$ element appearance u_ξ in the message $U(\theta)$ tends to zero, then as a result of clustering by a quantitative criterion λ_i it can form a separate set $U(\lambda_i)$, probability $P(u_{\xi,i})$ the appearance of an element in which it will tend to unity, i.e. the following condition will be met:

$$\text{if } P(u_\xi) \rightarrow 0 \text{ and } \{u_\xi\} \in U(\lambda_i) \text{ then } P(u_{\xi,i}) \rightarrow 1. \quad (12)$$

This variant of transformation of the nature of the law of probability distribution $P(u_\xi)$ appearance of elements u_ξ shown in Fig. 4 (option 3). So in Fig. 4 the element u_ξ block U_τ with a value equal to $u_\xi = 169$, the following probability value corresponds $P(u_\xi)$ appearances in the block U_τ :

$$P(u_\xi) = 0,015625.$$

Since the probability value $P(u_\xi)$ element appearance u_ξ in the block U_τ tends to zero, i.e.

$$P(u_\xi) \rightarrow 0,$$

we can conclude that this element u_ξ is the low-frequency component of the block.

As a result of applying clustering based on the quantity λ_i CE equal to 4, element u_ξ forms a separate set $U(\lambda_i)$ (the third option in Fig. 4), which looks like this:

$$\{u_\xi \mid \lambda_i = 4\} \in U(\lambda_i) \dots$$

This allows you to significantly increase the value of the probability $P(u_\xi)$ element u_ξ appearance block U_τ . So for the analyzed example, the element u_ξ multitudes $U(\lambda_i)$ corresponds to the probability value $P(u_\xi)$, which is equal to one, i.e.:

$$P(u_{\xi,i}) = 1.$$

For this scenario, the transformation of the nature of the probability distribution law $P(u_{\xi})$ appearance of elements u_{ξ} , as a result of applying clustering by the sign λ_i , is given by the following expression:

$$P(u_{\xi,i}) \gg P(u_{\xi}), \quad (13)$$

where $P(u_{\xi})$ - value of the probability of occurrence of a low-frequency element u_{ξ} block U_{τ} , i.e. $P(u_{\xi}) \rightarrow 0$;

$P(u_{\xi,i})$ - value of the probability of occurrence of a high-frequency element u_{ξ} sets $U(\lambda_i)$, i.e. $P(u_{\xi,i}) \rightarrow 1$.

Thus, the result of clustering the elements u_{ξ} block U_{τ} by quantity λ_i SU is the following transformation of the probability distribution law $P(u_{\xi})$ - element transition u_{ξ} from the low-frequency component of the block U_{τ} to the high-frequency component of the set $U(\lambda_i)$.

• if, as a result of clustering by a quantitative criterion λ_i low frequency elements u_{ξ} block U_{τ} form sets $U(\lambda_i)$, power $|U(\lambda_i)|$ which is much lower power $|U_{\tau}|$ block U_{τ} , i.e.:

$$\text{if } P(u_{\xi}) \rightarrow 0 \text{ and } \{u_{\xi}\} \in U(\lambda_i), |U_{\tau}| \gg |U(\lambda_i)|, \text{ then } P(u_{\xi,i}) \gg P(u_{\xi}).$$

This variant of transformation of the nature of the law of probabilities $P(u_{\xi})$ appearance of elements u_{ξ} block U_{τ} , as a result of applying clustering by the attribute λ_i , is given by expression (13), i.e.:

$$P(u_{\xi,i}) \gg P(u_{\xi}).$$

Such a change in the nature of the probability distribution law is evident $P(u_{\xi})$ appearance of elements u_{ξ} shown in Figure 4 (first option), where the use of clustering by quantitative λ_i allows to reduce power $|U_{\tau}|$ block U_{τ} 16 times, i.e.:

$$\frac{|U_{\tau}|}{|U(\lambda_i)|} = 16.$$

As a result, the probability value $P(u_{\xi,i})$ appearance of elements $u_{\xi,i}$ in set $U(\lambda_i)$ 16 times the probability $P(u_{\xi})$ appearance of elements u_{ξ} in the block U_{τ} , i.e.:

$$\frac{P(u_{\xi,i})}{P(u_{\xi})} = 16.$$

Thus, the results of evaluating the effectiveness of the application of the developed coding method from the standpoint of further reducing the structural redundancy of the code sequence indicate that for the analyzed example, the length of the code structures that are assigned to the message elements as a result of coding in the statistical space of sets is reduced by an average of 25.6%. This is due to the fact that the use of clustering of message elements based on the number of series of units, for a more advantageous presentation of the encoded data, leads to significant transformations of the nature of the distribution law of elements in sets.

4. Conclusion

In turn, the analysis of possible transformations of the nature of the law of distribution of the probabilities of the appearance of elements in the message as a result of the use of clustering by quantitative criteria indicates the following:

• transformation of the nature of the probability distribution law $P(u_{\xi})$ appearance of elements u_{ξ} messages $U(\theta)$ as a result of clustering, it leads to a decrease in the power of the encoded data, which makes it possible to increase the efficiency of entropy coding from the standpoint of further reducing the structural redundancy of the code representation of information resource data. It should

be noted that the more significant the transformation, the shorter the length of individual code structures that are assigned to elements and, accordingly, the output code sequence as a whole.

- transformation of the nature of the probability distribution law $P(u_\xi)$ appearance of elements u_ξ in the message $U(\theta)$ as a result of clustering by quantity λ_i CE has a significant impact on the positioning strategy of individual code constructs that are assigned to elements u_ξ messages $U(\theta)$ in the process of coding in the statistical space of sets. This means that in the process of coding message elements in the statistical space of sets, a situation arises when identical codes can be assigned to elements of different values, or a code structure that is assigned to one message element can be the beginning of a variable length code that is assigned to another element.

Therefore, the goal of further research is a detailed analysis of the effect of transformation of the nature of the probability distribution law $P(u_\xi)$ appearance of elements u_ξ in the message $U(\theta)$ as a result of the application of clustering by quantitative criteria λ_i on the strategy of positioning individual code structures in a common code sequence.

5. References

- [1] S. Ramakrishnan, et al., *Cryptographic and Information Security Approaches for Images and Videos*. CRC Press, 2018, 962 p.. doi: 10.1201/9780429435461.
- [2] Announcing the ADVANCED ENCRYPTION STANDARD (AES). Federal Information Processing Standards Publication, 197 (2001).
- [3] DSTU 7624:2014: Information Technology. Cryptographic protection of information. Symmetric block transformation algorithm. Order of the Ministry of Economic Development of Ukraine № 1484 (29.12.2014).
- [4] DSTU GOST 28147:2009: Information processing system. Cryptographic protection. Cryptographic transformation algorithm GOST 28147-89 (22.12.2008).
- [5] F. Dufaux, T. Ebrahimi, Toward a Secure JPEG. *Applications of Digital Image Processing XXIX*, Vol. 6312, 2006. doi: 10.1117/12.686963.
- [6] M. Farajallah, Chaos-based crypto and joint crypto-compression systems for images and videos, 2015. URL: <https://hal.archives-ouvertes.fr/tel-01179610>.
- [7] T. Honda, Y. Murakami, Y. Yanagihara, T. Kumaki, T. Fujino, Hierarchical image-scrambling method with scramble-level controllability for privacy protection, in: *IEEE 56th International Midwest Symposium on Circuits and Systems (MWSCAS)*, 2013, pp. 1371-1374. doi: 10.1109/MWSCAS.2013.6674911.
- [8] Information technology – JPEG 2000 image coding system: Secure JPEG 2000. International Standard ISO/IEC 15444-8; ITU-T Recommendation T.807, 2007, 108 p.
- [9] Sh. Ji, X. Tong, M. Zhang, Image encryption schemes for JPEG and GIF formats based on 3D baker with compound chaotic sequence generator, 2012. URL: <https://arxiv.org/abs/1208.0999>.
- [10] JPEG Privacy & Security Abstract and Executive Summary, 2015. URL: https://jpeg.org/items/20150910_privacy_security_summary.html.
- [11] R. L. Rivest, A. Shamir, L. M. Adleman, A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, (2) 21, 1978, pp. 120–126. doi: 10.1145/359340.359342
- [12] R. Sharma, S. Bollavarapu, Data Security using Compression and Cryptography Techniques. *International Journal of Computer Applications*, Vol. 117, No. 14, 2015, pp. 15-18. doi: 10.5120/20621-3342.
- [13] V. B. Vasiliev, I. N. Okov, Yu. N. Strezhik, A. A. Ustinov, N. V. Shvetsov, Video data compression and protection in UAV information exchange radio channels, in: *Scientific and practical conference on Prospects for the development and use of complexes with unmanned aerial vehicles*, 924 State Center for Unmanned Aviation of the Ministry of Defense of the Russian Federation, 2016, pp. 202–204.

- [14] K. Wong, K. Tanaka, DCT based scalable scrambling method with reversible data hiding functionality, in: 4th International Symposium on Communications, Control and Signal Processing (ISCCSP), 2010, pp. 1-4. doi: 10.1109/ISCCSP.2010.5463307.
- [15] L. Yuan, P. Korshunov, T. Ebrahimi, Secure JPEG Scrambling enabling Privacy in Photo Sharing, in: 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 2015, pp. 1-6. doi: 10.1109/FG.2015.7285022.
- [16] K. M. Faraoun, A parallel block-based encryption schema for digital images using reversible cellular automata. *Engineering Science and Technology*, Vol. 17, 2014, pp. 85–94. doi: 10.1016/j.jestch.2014.04.001.
- [17] S. Auer, A. Bliem, D. Engel, A. Uhl, A. Unterweger, Bitstream-based JPEG Encryption in Real-time, in: *International Journal of Digital Crime and Forensics* (2013). doi: 10.4018/jdcf.2013070101.
- [18] H. Kobayashi, H. Kiya, Bitstream-Based JPEG Image Encryption with File-Size Preserving, in: *IEEE 7th Global Conference on Consumer Electronics (GCCE)*, 2018, pp. 1-4. doi: 10.1109/gcce.2018.8574605.
- [19] K. Minemura, Z. Moayed, K. Wong, X. Qi, K. Tanaka, JPEG image scrambling without expansion in bitstream size, in: 19th IEEE International Conference on Image Processing, 2012, pp. 261-264. doi: 10.1109/ICIP.2012.6466845.
- [20] A. Phatak, A Non-format Compliant Scalable RSA-based JPEG Encryption Algorithm. *International Journal of Image, Graphics and Signal Processing*, Vol. 8, No. 6, 2016, pp 64-71. doi: 10.5815/ijigsp.2016.06.08.
- [21] Ch.-L. Tsai, Ch.-J. Chen, W.-L. Hsu, Multi-morphological image data hiding based on the application of Rubik's cubic algorithm, in: *IEEE International Carnahan Conference on Security Technology (ICCST)*, 2012, pp. 135-139. doi: 10.1109/CCST.2012.6393548.
- [22] K.-W. Wong, Image encryption using chaotic maps. *Intelligent Computing Based on Chaos*, Vol. 184, 2009, pp. 333–354. doi: 10.1007/978-3-540-95972-4_16.
- [23] Yu. Wu, S. Aгаian, J. Noonan, Sudoku Associated Two Dimensional Bijections for Image Scrambling, in: *IEEE Transactions on multimedia*, 2012, 30 p. URL: <https://arxiv.org/abs/1207.5856v1>.
- [24] Y. Yang, B. B. Zhu, S. Li, N. Yu1, Efficient and Syntax-Compliant JPEG 2000 Encryption Preserving Original Fine Granularity of Scalability. *EURASIP Journal on Information Security*, Vol. 2007, Article ID 56365, 2008, 13 p. doi: 10.1155/2007/56365.
- [25] V. Barannik, N. Barannik, Yu. Ryabukha, D. Barannik, Indirect Steganographic Embedding Method Based On Modifications of The Basis of the Polyadic System, in: 15th IEEE International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET'2020), 2020, pp. 699-702. doi: 10.1109/TCSET49122.2020.235522.
- [26] V. Barannik, V. Barannik, Binomial-Polyadic Binary Data Encoding by Quantity of Series of Ones, in: 15th IEEE International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET'2020), 2020, pp. 775-780. doi: 10.1109/TCSET49122.2020.235540.
- [27] V. Barannik, T. Belikova, P. Gurzhii, The model of threats to information and psychological security, taking into account the hidden information destructive impact on the subconscious of adolescents, in: 2019 IEEE International Conference on Advanced Trends in Information Theory (ATIT), 2019, pp. 656-661. doi: 10.1109/ATIT49449.2019.9030432.
- [28] V. V. Barannik, M. P. Karpinski, V. V. Tverdokhleб, D. V. Barannik, V. V. Himenko, M. Aleksander, The technology of the video stream intensity controlling based on the bit-planes recombination, in: 2018 IEEE 4th International Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS), 2018, pp. 25-28. doi: 10.1109/IDAACS-SWS.2018.8525560.
- [29] V. V. Barannik, Yu. N. Ryabukha, O. S. Kulitsa, The method for improving security of the remote video information resource on the basis of intellectual processing of video frames in the telecommunication systems, *Telecommunications and Radio Engineering*, Vol. 76, No 9, 2017, pp. 785-797. doi: 10.1615/TelecomRadEng.v76.i9.40.

- [30] V. Barannik, S. Shulgin, The method of increasing accessibility of the dynamic video information resource, in: 2016 13th International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET), 2016, pp. 621-623. doi: 10.1109/TCSET.2016.7452133.
- [31] V. Barannik, D. Tarasenko, Method coding efficiency segments for information technology processing video, in: 2017 4th International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), 2017, pp. 551-555. doi: 10.1109/INFOCOMMST.2017.8246460.
- [32] Ch.-Ch. Chen, W.-J. Wu, A secure Boolean-based multi-secret image sharing scheme. Journal of Systems and Software, Vol. 92, 2014, pp. 107-114. doi: 10.1016/j.jss.2014.01.001.
- [33] T.-H. Chen, Ch.-S. Wu, Efficient multi-secret image sharing based on Boolean operation. Signal Processing, Vol. 91, Iss. 1, 2011, pp. 90-97. doi: 10.1016/j.sigpro.2010.06.012.
- [34] M. Deshmukh, N. Nain, M. Ahmed, An (n, n)-Multi Secret Image Sharing Scheme Using Boolean XOR and Modular Arithmetic, in: IEEE 30th International Conference on Advanced Information Networking and Applications (AINA), 2016, pp. 690-697. doi: 10.1109/aina.2016.56.
- [35] M. Naor, A. Shamir, Visual Cryptography, in: Proceedings of the Advances in Cryptology – EUROCRYPT'94. Lecture Notes in Computer Science, Vol. 950, 1995, pp. 1-12. doi: 10.1007/bfb0053419.
- [36] Ch.-N. Yang, Ch.-H. Chen, S.-R. Cai, Enhanced Boolean-based multi secret image sharing scheme. Journal of Systems and Software, Vol. 116, 2016, pp. 22-34. doi: 10.1016/j.jss.2015.01.031.
- [37] P. Korshunov, T. Ebrahimi, Using warping for privacy protection in video surveillance, in: 18th International Conference on Digital Signal Processing (DSP), 2015, pp. 1-6. doi: 10.1109/ICDSP.2013.6622791.
- [38] G. Nattress, Chroma Sampling: An Investigation, 2008. URL: http://www.nattress.com/Chroma_Investigation/chromasampling.htm.
- [39] D. A. Kerr, Chrominance Subsampling in Digital Images, 2012. URL: <http://dougkerr.net/Pumpkin/articles/Subsampling.pdf>.