Computational Experiments to Analyze the Effectiveness of the Algorithm for Decomposing a Formal Context into Fragments

Mongush Choduraa Mikhailovna[a], Ondar Sevil Kechil-oolovna[b], Bayyr Dolaan Borisovich[c] and Oorzhak Choygana Kamaevna[d]

Tuvan State University, 36 Lenina Street, Kyzyl, Russia

mongushchod91@yandex.ru, sevil.badyma@mail.ru, dolaan.bayyr@mail.ru, oorzhak-choygana@mail.ru

Keywords: analysis of formal concepts, fragmentation algorithm, formal context, combinatorial problem, the problem of finding the set of all formal concepts, computational experiments.

Abstract: The article presents the results of computational experiments. The following is evaluated: the effectiveness of the algorithm for fragmentation and defragmentation of the formal context with a single decomposition of the context; the number of fragments formed depending on the density of the original formal context; the number of possible iterations of the fragmentation process; the number of fragments and the running time of the algorithm for the context depending on the threshold value σ for the fragments being formed.

1 INTRODUCTION

This work continues the research started in the works (Mongush, 2019; Bykova, 2019), which are devoted to improving the performance of existing algorithms for solving the problem of finding all formal concepts by decomposing the formal context into fragments. In these works, to solve this problem, a method of fragmentation and defragmentation of the formal context without losing the sought-for formal concepts was proposed and proven. An algorithm implementing the proposed method has been developed. A detailed description of the algorithm is given in (Mongush, 2023). This paper presents the results of computational experiments to evaluate the effectiveness of the algorithm for fragmentation and defragmentation of the formal context.

2 BASIC CONCEPTS

We briefly present the basic concepts of formal concept analysis.

Formal Concept Analysis (FCA) is an applied direction of G. Birkhoff's lattice theory (Birkhoff, 1984). The main ideas of FCA were formulated in the works of R. Wille and B. Ganter in the early 80s of the 20th century (Ganter, 2011; Ganter, 2016). Based on the FCA method, the description of the term "concept" is formalized in the form of a pair (volume, content), object-attribute dependencies are visualized by constructing a lattice of formal concepts.

The formal context is an object-attribute table, represented as a triple $K = (G, M, I)$, where G is a set of objects, M is a set of features, and I is a binary incidence relation between G and M sets. The G, M sets are linearly ordered. Then the formal context K is uniquely defined by the (0, 1)-matrix T, where 0 means that $(g, m) \notin I$ object g does not have the attribute m, and in the matrix 1 denotes $(g, m) \in I$ – object g has the property m. An example of the formal context $K = (G, G, \neq)$ is presented in Table 1. This context corresponds to a (0, 1)-matrix, all elements of which are equal to 1, except for the diagonal elements.

Table 1: Formal context $K = (G, G, \neq)$.

0	1	1	1	1
1	0	1	1	1
1	1	0	1	1

[a] https://orcid.org/0000-0002-0006-4835
[b] https://orcid.org/0009-0003-8808-2505
[c] https://orcid.org/0009-0000-6904-7286
[d] https://orcid.org/0009-0009-0073-8770

1	1	1	0	1
1	1	1	1	0

In formal concept analysis, the problem of finding all formal concepts of a formal context is formulated as follows. The formal context $K = (G, M, I)$ is given, represented as a (0, 1)-matrix T. Find the set of all formal concepts of the formal context K.

In FCA, the formal concept is defined by a Galois map. Let $A \subseteq G$ and $B \subseteq M$. Then the Galois map for them is calculated using the formula:

$$A' = \cap_{g \in A} g' = \{m \in B \mid \forall g \in A\ (g, m) \in I\},$$
$$B' = \cap_{m \in B} m' = \{g \in A \mid \forall m \in B\ (g, m) \in I\}.$$

If for $A \subseteq G$ and $B \subseteq M$ the following is true

$$A' = B \text{ and } B' = A,$$

then the pair (A, B) is called the formal concept of the formal context $K = (G, M, I)$. In other words, the formal concept (A, B) of the context K with a (0,1)-matrix is a maximally complete submatrix of the matrix T, where the rows of this submatrix correspond to elements from A, and the columns correspond to elements from B. Here, a maximally complete submatrix is a submatrix, all elements of which are equal to 1 and which is not contained in other complete submatrices.

The set of all formal concepts of a formal context is denoted by FC. For the context $K = (G, G, \neq)$ from Table 1, the set FC contains exactly $2^{|G|}$ formal concepts. The number of formal concepts depends exponentially on the size of the original formal context. This problem belongs to combinatorial enumeration problems and has high computational complexity (Kuznetsov, 2004; Kuznetsov 2001).

Let $A_1, A_2 \subseteq G$ и $B_1, B_2 \subseteq M$. On the set of all formal concepts, we define the operations of union $\sqcup$ and $\sqcap$ intersection through the set-theoretic operations $\cap$ and $\cup$ as follows:

$$(A_1, B_1) \sqcup (A_2, B_2) = ((B_1 \cap B_2)', B_1 \cap B_2),$$
$$(A_1, B_1) \sqcap (A_2, B_2) = (A_1 \cap A_2, (A_1 \cap A_2)').$$

Then the ordered set of all formal concepts forms a lattice of formal concepts L. The lattice of formal concepts can be considered as a conceptual model of the domain. It defines classes of homogeneous objects and connections between them, which is the basis for solving applied problems of data analysis. Using FCA methods, it is possible to solve problems of classification, clustering, identification of dependencies between data, etc. ().

3 DECOMPOSITION ALGORITHM

Many algorithms have been developed to calculate the set FC and construct the lattice L (Qian, 2017; Simon, 2015; Poelmans, 2013). The running time of the algorithms in the worst case is $O(|FC| \cdot |G|^2 \cdot |M|)$. Since the value $|FC|$ exponentially depends on $|G|$ and $|M|$, the running time of these algorithms can also be exponential.

To date, research to reduce the computational complexity of the problem of finding all formal concepts is relevant. In the works (Mongush, 2019; Bykova, 2018), a method of fragmentation and defragmentation of the formal context is proposed. The idea of the method is that the original formal context is divided into various fragments. The fragments have different sizes and a nonempty intersection. Each fragment is further considered as a formal context and is decomposed again. As a result, a finite set of fragments is formed. Then formal concepts are found in each fragment using the well-known Close-by-One algorithm. The resulting formal concepts are combined and the sought-for set of all formal concepts of the formal context is formed. In the work (Mongush, 2019) it is proved that the method of fragmentation and defragmentation of the formal context is "non-distorting", i.e. each fragment contains at least one formal concept, when the context decomposes, the sought-for formal concepts do not disappear and new concepts do not appear. The process of context decomposition is performed iteratively. To stop this process, one must specify the number of iterations, and one can additionally enter a threshold for the density of fragments. Based on this method, an algorithm was developed, which is described in detail in the work (Mongush, 2023).

Theoretically, the computational complexity of the formal context fragmentation and defragmentation algorithm is high and amounts to

$$O(\sigma(K) \cdot |G|^2 \cdot |M|^2),$$

where $\sigma(K)$ is the density of the formal context, i.e. $\sigma(K) = ||T||/(|G| \cdot |M|)$ the ratio of the number of ones in the (0,1)-matrix T to the size of the formal context K.

The algorithm can be performed iteratively. If the number of iterations is equal to k, then the process of decomposing the original context into fragments by the algorithm takes $O(|G|^{2k} \cdot |M|^{2k})$ time. The iteration process can be stopped in two ways: fix the number of iterations k, and impose a limit on the density of the fragments formed σ. In this case, the running time of the algorithm will be polynomial relative to the size of the original formal context.

3 COMPUTATIONAL EXPERIMENTS

A number of computational experiments were conducted to analyze the effectiveness of the fragmentation and defragmentation algorithm of the formal context.

The first experiment is aimed at evaluating the effectiveness of the algorithm with a single decomposition of the context and without limiting the density of fragments. The results are presented in Table 2.

Table 2: Results of the first experiment.

	The size of the formal context	Time *t*, ms
Finding the *FC* set using the fragmentation algorithm	100×20	2,450
Finding the *FC* set in the simplest way		150,533
Finding the *FC* set using the fragmentation algorithm	200×30	91,200
Finding the *FC* set in the simplest way		810,530

The obtained result shows that finding the set of all formal concepts using the fragmentation and defragmentation algorithm of the formal context takes significantly less time compared to the algorithm without fragmentation.

The second experiment is aimed at estimating the number of fragments formed depending on the density of the original formal context at $|G|$=100, $|M|$=20, k=1. The results are shown in Table 3.

Table 3: Results of the second experiment.

The density of the original context σK	Number of fragments *N*	Time *t*, ms
0.1	137	11
0.2	261	108
0.3	466	506
0.4	677	1,760
0.5	826	3,234
0.6	828	6,262
0.7	883	10,974
0.8	402	14,438
0.9	478	15,004

From Table 3 it follows that the higher the density of the original context, the more time it takes to decompose the context once.

The third experiment is aimed at estimating the number of possible iterations of the fragmentation process and the time required to implement this process. The worst case is considered: $K = (G, G, \neq)$ at $|G|$= 7, k=6. The results are shown in Figures 1, 2.

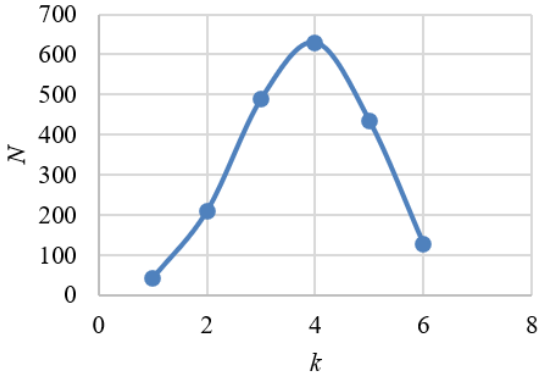

Figure 1: Dependence of the number of fragments *N* on the number of iterations *k*.

From these graphs it follows that an increase in the number of iterations leads to an increase in the number of fragments to be further decomposed, and for large values of *k*, the number of fragments and the time of decomposition of the formal context grow rapidly. Therefore, the value of k should be small and significantly less than $k << n/2$.

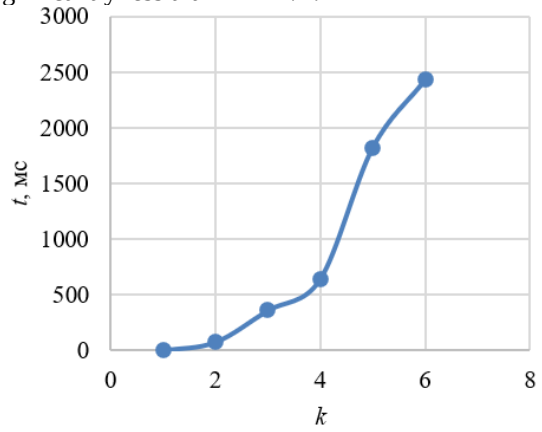

Figure 2: Dependence of the running time t on the number of iterations k.

The fourth experiment is aimed at estimating the number of fragments and the running time of the algorithm for the context, depending on the threshold value σ for the fragments being formed. Let's consider the context $K = (G, G, \neq)$. The evaluation was carried

out at $|G|= 7$, k=6, $\sigma = 1$; $\sigma = 0.9$; $\sigma = 0.87$; $\sigma = 0.865$; $\sigma = 0.86$. The results are presented in Table 4.

From Table 4 it follows that the threshold value for the density of fragments to be further decomposed must be selected from the interval $\sigma_K < \sigma < 1$. If σ is close to or equal to one, for example, $\sigma = 1$ and $\sigma = 0.9$, then the number of fragments formed increases and, accordingly, the decomposition time of the formal context increases. If σ is close to σ_K, for example, $\sigma = 0.86$, then the formal context is decomposed into fragments once.

Table 4: Results of the fourth experiment.

$\sigma = 1$						
k	1	2	3	4	5	6
N	42	210	490	630	434	126
t, ms	1	75	360	640	1,820	2,430
$\sigma = 0.9$						
k	1	2	3	4	5	6
N	42	210	490	630	714	
t, ms	1	75	350	940	1,620	
$\sigma = 0.87$						
k	1	2	3	4	5	6
N	42	210	490			
t, ms	1	60	380			
$\sigma = 0.865$						
k	1	2	3	4	5	6
N	42	210				
t, ms	1	60				
$\sigma = 0.86$						
k	1	2	3	4	5	6
N	42					
t, ms	3					

4 CONCLUSIONS

Thus, the results of computational experiments showed the effectiveness of the developed algorithm if certain recommendations are followed.

REFERENCES

Mongush, Ch.M., Bykova, V.V., 2019. On decomposition of a binary context without losing formal concepts. Journal of Siberian Federal University. *Mathematics and Physics.* Vol. 12. № 3. P. 323–330.

Bykova, V.V., Mongush, CH.M., 2019. Dekompozicionnyj podhod k issledovaniyu formal'nyh kontekstov. *Prikladnaya diskretnaya matematika.* № 44. S. 113–126.

Mongush, CH.M., 2023. Algoritm fragmentacii i defragmentacii formal'nyh kontekstov. *Inzhenernyj vestnik Dona.* № 11.

Birkgof, G., 1984. Teoriya reshetok. M.: Nauka, 568 s.

Ganter, B., Wille, R., 2012. Formal Concept Analyses: Mathematical Foundations. Springer Science and Business Media.

Ganter, B., Obiedkov S.A., 2016. Conceptual Exploration. Berlin Heidelberg: Springer, 2016. 315 p.

Kuznetsov, S. O., Ganter, B., Eklund, B., 2004. Sertkaya Machine Learning and Formal Concept Analysis. 2th International Conference on Formal Concept Analysis. Berlin: Heidelberg. Vol. 2961. P. 287–312.

Kuznecov, S.O., 2001. Avtomaticheskoe obuchenie na osnove analiza formal'nyh ponyatij. *Avtomatika i telemekhanika*. № 10. S. 3–27.

Qian, T., Wei, L., Qi, J., 2017. Decomposition methods of formal contexts to construct concept lattices. *International Journal of Machine Learning and Cybernetics.* Vol. 8. No. 1. P. 95–108.

Simon, A. A., 2015. Best-of-Breed approach for designing a fast algorithm for computing fixpoints of Galois Connections. *Information Sciences.* Vol. 295. No. 2. P. 633–649.

Poelmans, J., Ignatov, D. I., Kuznetsov, S. O., Dedene, G., 2013. Formal concept analysis in knowledge processing: a survey on applications. *Journal of Mathematical Sciences.* Vol. 40. No16. P. 6538-6560.