UDC 004.942:621

DOI: 10.20535/2523-4455.mea.344933

Clustering Method Using Cardinal Number Vectors

O. O. Abakumova, PhD, Assoc. Prof., 0000-0002-5467-2473
V. Yu. Zhuikov, Dr.Sc.(Eng.), Prof., 0000-0002-3338-2426
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" 00syn5v21
Kyiv, Ukraine

M. O. Lukianov, 0000-0001-8930-9992
Gdańsk University of Technology, 006x4sc24
Gdańsk, Poland

Abstract–In modern technical systems, clustering serves as a key procedure for structuring, analyzing, and interpreting large volumes of data, which in turn enhances decision-making efficiency and the optimization of system processes. This study presents a comparative analysis of the main groups of clustering methods and proposes a novel approach based on the powerful mathematical framework of cardinal numbers. The theoretical foundations for constructing cardinal number vectors are revealed, positioning them as a mathematical tool for data representation in clustering tasks. The proposed approach defines object distances within a selected orthogonal basis using the calculated cardinalities of abstract set sequences represented as vectors of cardinal numbers. The study explores the formation of these vectors and the computation of corresponding similarity metrics, followed by the generation of a distance matrix. A practical example illustrates the calculation of distances between three functions and a reference function based on their respective cardinal number vectors. It is demonstrated that altering the basis or projections according to the technical problem allows for the formation of different clusters, reflecting the flexibility and adaptability of the proposed method. The calculations are formalized, straightforward, and easily algorithmized, which enables the implementation of dynamic clustering. This approach holds significant promise for use in intelligent data analysis systems and information processing in electronic devices.

Keywords: clustering; vectors; cardinal numbers; similarity metric; local object; orthonormal basis.

I. INTRODUCTION

One of the current developments of energy today is creation of digital smart systems and networks [1] which combine traditional energy infrastructures with Internet of Things technologies. An important component of such systems is local objects, the processing of whose data affects the efficiency of control methods, which in turn is determined by the choice of clustering methods [2–4].

In local systems, clustering faces certain challenges, such as device heterogeneity, limited energy and information capacity of sensor nodes, the need for real-time data processing, and ensuring network scalability.

The clustering process uses various similarity metrics, such as Euclidean distance [5, 6], squared Euclidean distance [7], Manhattan distance [6, 8], Chebyshev distance [6], power distance [9], cosine similarity [10] or correlation [11], as well as feature sets that reflect key characteristics of objects, such as energy consumption level, geographical location, time parameters or technical specifications [12].

Clustering methods can be divided into several main groups, listed in Table 1.

In all clustering methods is considered use of information that directly describes a phenomenon with physical properties and dimensions, followed by the determination of distances.

A promising method may be one based on the use of the relational property used in relational models that use the concept of cardinal number [34].

II. PROBLEM STATEMENT. CARDINAL NUMBER VECTOR METHOD

The aim of the work is to develop a new clustering method using vectors of cardinal numbers, which allows to move from sequences of some specific quantities to abstract sequences while preserving the features of the sequences and significantly simplifying the subsequent transition to operations with vectors of cardinal numbers.

DOI: 10.20535/2523-4455.mea.344933

TABLE 1. CLUSTERING METHODS

Methods	Advantages	Problems	Field of application
Hierarchical [13, 14]	Visualization through dendrograms, flexibility in determining the number of clusters	High computational complexity, sensitivity to "noise"	Data compression, pattern recognition, taxonomy
Iterative [15, 16]	Easy to implement, efficient for big data	The need to determine the number of clusters, sensitivity to outliers	Customer segmentation, image processing, Internet of Things
Factorial [17]	Hidden variable detection, dimensionality reduction	Difficulty of interpretation, need for statistical assumptions	Data analysis (sociology, marketing, bioinformatics)
Modal density estimation [18, 19]	Free-form clusters, noise resistance	Dependence on parameters, problems with different densities	Anomaly detection, spatial data analysis
Using graph theory [20, 21]	Use of graph-based structures, flexibility for complex relations	High computational complexity, need for graph tuning	Social network analysis, bioinformatics, Internet of Things
Grouping [12, 22]	Ease of implementation, non-parametric, adaptability	Sensitivity to data scale, susceptibility to "noise"	Classification, regression, image processing, recommender systems
Grid-based [23, 24]	High speed, cluster shape independence	Accuracy depends on grid resolution	Real-time monitoring (Internet of Things, energy consumption)
Model-oriented [25, 26]	Modeling of complex distributions, high accuracy	Need to determine the number of clusters, high complexity	Prediction, classification, image processing
Fuzzy clustering [27, 28]	Flexibility for overlapping clusters, resistance to "noise"	The need to determine the number of clusters, the choice of the fuzziness parameter	Image segmentation, data analysis (biology, marketing)
Neural clustering (machine learning) [29–31]	Nonlinear data processing, multidimensional data visualization	Significant learning time, difficulty in interpretation	Complex data analysis, dimensionality reduction
Hybrid clustering [32, 33]	Combining the strengths of methods, flexibility for heterogeneous data	Increased complexity, need for careful tuning	Adaptive clustering in dynamic systems

Let us consider a method that allows us to get rid of the use of dimensions and, in a sense, to abstract from physical phenomena during clustering, and to determine the distance between objects in the selected orthogonal basis based on calculating the powers of abstract sequences of sets, which are presented in the form of vectors of cardinal numbers. If necessary, the properties of objects are taken into account by their weight coefficients.

For convenience and simplification of teaching, we will focus on the two-dimensional case.

First, let us move from dimensional quantities to dimensionless ones and consider the case of the dependence of some dimensionless function $P(t)$ from the argument t. Let us locate the values of function $P(t)$ in the range from $P(t)=0$ to $P(t)=P_{\max}$ in intervals $P_i < P(t) < P_{i+1}$, where $(P_i - P_{i-1}) = \Delta P_i$. Further we will count the number of cells h with dimension $\Delta P_i \times \Delta t_j$, where $\Delta t_j = t_j - t_{j-1}$, $\Delta t_j = \frac{T}{2^{m-1}}$, $i = 1, 2, \ldots, h$, $j = 1, 2, \ldots, 2^{m-1}$, where 2^{m-1} – maximum number of smallest intervals, m – dimension of the vector of cardinal numbers along the t axis, T – maximal interval. Value $k_{pi}(t_j)$, which is located in the cells $\Delta P_i \times \Delta t_j$, is defined as follows:

$$k_{pi}(t_j) = \begin{cases} 1, \text{ if } P(t) \in (\Delta P_i \times \Delta t_j) \\ 0, \text{ if } P(t) \notin (\Delta P_i \times \Delta t_j) \end{cases}, \quad (1)$$

To the first cell of the cardinal number vector K_p the number K_{p1} is written, which is defined as:

$$K_{p1} = \sum_{j=1}^{2^{m-1}} k_{p1}(t_j). \quad (2)$$

Then the vector of cardinal numbers K_p (transposed), will have the form:

$$K_p^T = \begin{bmatrix} K_{p1} & K_{p2} & \ldots & K_{pi} & \ldots & K_{pn} \end{bmatrix}. \quad (3)$$

We create intervals for the argument t_j, in which we will record the presence or absence of values $P(t)$ for each interval $t_j - t_{j-1}$. That is, the entire interval T is divided into several intervals, starting from the interval T itself by gradually decreasing the duration of the intervals – to $\Delta t_j = \frac{T}{2^{m-1}}$ with $j = 1, 2, \ldots, 2^{m-1}$. Value $k_{ts}(t_j)$, which determines the number of intervals to which any value corresponds $P(t)$, is defined as follows:

$$k_{ts}(t_{j-1}, t_j) = \begin{cases} 1, \text{ if } P(t) \in (t_{j-1} - t_j) \\ 0, \text{ if } P(t) \notin (t_{j-1} - t_j) \end{cases}, \quad (4)$$

where $t_{j-1}=t_{2^{m-s}\cdot(l-1)}$; $t_j=t_{2^{m-s}\cdot l}$, m – vector dimension K_t, s – vector cell number K_t, $l=1\div 2^{s-1}$ – ordinal number of the partition interval for the corresponding $k_{ts}(t_j)$.

Then the vector $K_t{}^T$, composed of cardinal numbers, will have the form (transposed):

$$K_t^T=\left[K_{t1}\;\;K_{t2}\;\;\ldots\;\;K_{tj}\;\;\ldots\;\;K_{tm}\right]. \qquad (5)$$

A certain law $\Delta t_j=f\left(\frac{T}{n}\right)$, according to which interval durations are defined, and the ratios between the intervals themselves are chosen under the conditions of a specific problem. The simplest law has the form $n=2^{m-1}$. The cardinal number corresponding to each interval is calculated using the formula:

$$K_{tj}=\sum_{j=1}^{m}k_{t1}(t_j).$$

In the same way, vectors of cardinal numbers are obtained for other time functions that are included in further clustering.

The mesh step, formed for calculating cardinal numbers can be determined by the value of the derivative of the classified functions, which, when moving to discrete values, allows the use of, for example, a criterion of the type $(P_{n+1}-P_n)/(t_{n-1}-t_n)<D$, where D is a coefficient determined by the practically possible level of discrete regulation of such parameters as the amplitude of the output voltage of the microgrid system, the phase shift angle, non-linear distortion factor, and others used in the coordinated connection of microgrids to the general network.

To take into account the importance of individual cardinal numbers, it is advisable to use a matrix of weight coefficients, the values of the elements of which at the intersection of rows and columns of cardinal numbers indicate the influence of one cardinal number on another, and the diagonal elements are the own weights of cardinal numbers.

For clustering, it is desirable to have some basic or reference system and, accordingly, a basic set of vectors. For this desired reference functions $P(t)$ and Δt_j are given (in the two-dimensional case, or a set of functions - in the multidimensional case) and vectors of cardinal numbers are built.

The clustering process can be carried out in two ways:

1. with using the initial values of the vectors;
2. with using an orthogonal basis.

Next, using the Gram–Schmidt procedure, a transition to an orthogonal basis is made, which allows us to determine the distance between the vectors and their angles, the values of which are used for clustering, adding the volume of the region of their desired location.

Let us consider examples of the vectors formation of cardinal numbers and their application for clustering.

III. EXAMPLE OF FORMING VECTORS OF CARDINAL NUMBERS

Let us consider the formation of cardinal numbers for the function $P_1(t)$, which is shown in Fig. 1.

For a function $P_1(t)$ (this can be, for example, a function of the power of energy generation by some local object) the ordinate axis $P_1(t)$ is divided into 4 intervals P_1, P_2, P_3, P_4, and the abscissa axis t is partitioned into intervals according to the law $n=2^{m-1}$ ($m=1,2,3,4$).

According to (1) the numbers $k_{p1}(t_1)=0$, $k_{p1}(t_2)=1$, $k_{p1}(t_3)=1$, ..., $k_{p1}(t_8)=0$. Then according to (2) the first cardinal number $K_{p1}=3$. Similarly, the remaining cardinal numbers are defined: $K_{p2}=5$, $K_{p3}=5$, $K_{p4}=6$. Therefore, the transposed vector of cardinal numbers K_p^T will have the form:

$$K_p^T=\left[3\;\;5\;\;5\;\;6\right].$$

Let us find the vector $K_t{}^T$ for the case shown in Fig. 1. According to (4) for $s=1$ we have $l=1$, so the number $k_{t1}(0,t_8)=k_{t1}(T)=1$. Then, the first cardinal number is $K_{t1}=1$. For $s=2$ we have $l=1,2$, so the numbers $k_{t2}(0,t_4)=k_{t2}(0,T/2)=1$ and $k_{t2}(t_4,t_8)=k_{t2}(T/2,T)=1$. Then, the second cardinal number is $K_{t2}=2$. Similarly, the following two cardinal numbers are defined: $K_{t3}=4$ and $K_{t4}=6$.

Therefore, the transposed vector of cardinal numbers $K_t{}^T$ will have the form:

$$K_t{}^T=\left[1\;\;2\;\;4\;\;6\right].$$

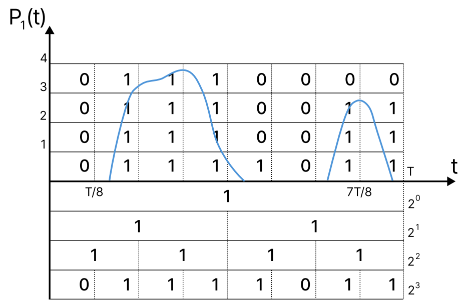

Fig. 1 The formation of cardinal numbers

DOI: 10.20535/2523-4455.mea.344933

DOI: 10.20535/2523-4455.mea.344933

Note that with such formation of the vector K_t the first cardinal number will always be equal to one.

Thus, two vectors of cardinal numbers are formed: K_p and K_t.

IV. EXAMPLE OF SIMILARITY METRICS CALCULATING USING CARDINAL NUMBER VECTORS

Let us consider the calculation of the distances between functions $P_1(t), P_2(t), P_3(t)$ and the base function $P_b(t)=2$ (see Fig. 1, Fig. 2, Fig. 3 and Fig. 4) using the values of their cardinal number vectors.

First, we find the vectors K_{pb} and K_{tb} of cardinal numbers for the basic function $P_b(t)$:

$$K_{pb}^T=\begin{bmatrix}0 & 0 & 8 & 8\end{bmatrix},\ K_{tb}^T=\begin{bmatrix}1 & 2 & 4 & 8\end{bmatrix}.$$

Orthogonal normalized vectors $\hat{K}_{pb}$ i $\hat{K}_{tb}$ of cardinal numbers have the form:

$$\hat{K}_{pb}^T=\begin{bmatrix}0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix},\quad \hat{K}_{tb}^T=\begin{bmatrix}\frac{1}{\sqrt{13}} & \frac{2}{\sqrt{13}} & \frac{-2}{\sqrt{13}} & \frac{2}{\sqrt{13}}\end{bmatrix}. \tag{6}$$

For clustering, it is necessary to calculate the distance between the vectors themselves and between the vectors and the orthonormalized base space. To do this, we calculate the distances between the three vectors given by the functions $P_1(t), P_2(t), P_3(t)$, and the base space given by the vectors (6), the vectors of cardinal numbers of which have the form:

$$K_{p1}^T=\begin{bmatrix}3 & 5 & 5 & 6\end{bmatrix},\ K_{p2}^T=\begin{bmatrix}2 & 5 & 6 & 6\end{bmatrix},$$
$$K_{p3}^T=\begin{bmatrix}4 & 5 & 6 & 6\end{bmatrix}.$$

The vectors $K_{t_1}, K_{t_2}, K_{t_3}$ are the same for all three functions and have the form:

$$K_t^T=\begin{bmatrix}1 & 2 & 4 & 6\end{bmatrix}.$$

For clustering, it is advisable to calculate the distances between the projections of the vectors $K_{p1}^T, K_{p2}^T, K_{p3}^T, K_t^T, K_{pb}^T, K_{tb}^T$ and the distances between the projections of these vectors onto the orthonormalized base space.

The calculation of the projection of a certain vector K onto the base space is carried out by the expression:

$$proj_p=\frac{K\cdot\hat{K}_{pb}}{\hat{K}_{pb}\cdot\hat{K}_{pb}}\hat{K}_{pb}+\frac{K\cdot\hat{K}_{tb}}{\hat{K}_{tb}\cdot\hat{K}_{tb}}\hat{K}_{tb}. \tag{7}$$

Then the projection K_{p1}:

$$proj_{p1}=a\cdot\hat{K}_{pb}+b\cdot\hat{K}_{tb}=\frac{11}{16}\cdot\hat{K}_{pb}+\frac{15}{13}\cdot\hat{K}_{tb}.$$

Similarly, according to (7), we find the projection K_{p2}:

$$proj_{p2}=c\cdot\hat{K}_{pb}+d\cdot\hat{K}_{tb}=\frac{3}{4}\cdot\hat{K}_{pb}+\frac{12}{13}\cdot\hat{K}_{tb}.$$

The distance between the projections of vectors is calculated by the formulas:

$$proj_i-proj_j=(a-c)\cdot\hat{K}_{pb}+(b-d)\cdot\hat{K}_{tb}$$

and

$$\left\|proj_i-proj_j\right\|=\sqrt{(a-c)^2\cdot\hat{K}_{pb}+(b-d)^2\cdot\hat{K}_{tb}}. \tag{8}$$

Then the vector distance between the projections of vectors K_{p1} and K_{p2}:

$$proj_{p1}-proj_{p2}=-\frac{1}{16}\hat{K}_{pb}+\frac{3}{13}\hat{K}_{tb},$$

and the distance by the modulus:

$$\left\|proj_{p1}-proj_{p2}\right\|=\sqrt{\left(\frac{-1}{16}\right)^2\cdot 128+\left(\frac{3}{13}\right)^2\cdot 13}\approx 1{,}09.$$

In the next step, we calculate the projections of the base vectors K_{pb}^T and K_{tb}^T onto the orthonormal base space:

$$proj_{pb}=(8\sqrt{2},0)\ \text{and}\ proj_{tb}=\left(6\sqrt{2},\frac{5}{\sqrt{13}}\right).$$

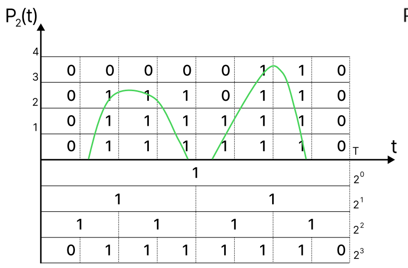

Fig. 2

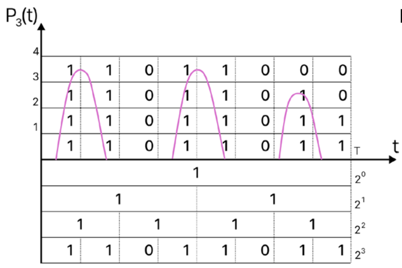

Fig. 3

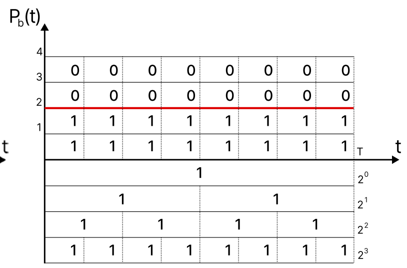

Fig. 4

DOI: 10.20535/2523-4455.mea.344933

Then, according to (8), the distance by the modulus between the projections $proj_{p1}$ and $proj_{pb}$:

$$\left\|proj_{p1} - proj_{pb}\right\| = = \sqrt{\left(\frac{11\sqrt{2}}{2} - 8\sqrt{2}\right)^2 + \left(\frac{15}{\sqrt{13}} - 0\right)^2} \approx 5{,}46,$$

and the distance by the modulus between the projections $proj_{p1}$ and $proj_{tb}$:

$$\left\|proj_{p1} - proj_{tb}\right\| = = \sqrt{\left(\frac{-1}{16}\right)^2 \cdot 128 + \left(\frac{10}{13}\right)^2 \cdot 13} \approx 2{,}77.$$

Similarly, we calculate the distances by the modulus between other projections and compose a matrix D of similarity metrics (Table 2), which illustrates the data, using of which allows us to carry out various possible options for clustering by the projections of some vectors onto others.

For example, in the cell D_{12} the distance between the projections $proj_{p1}$ and $proj_{p2}$ is equal to 1,09, which is close to 1, and indicates that these vectors are practically the same and can be assigned to the same cluster. However, the projections $proj_{p1}$ and $proj_{p2}$ of vectors on the projections $proj_{pb}$ та $proj_{tb}$ of vectors differ significantly. Therefore, clustering depends on the technical task. In this case, despite the fact that the vectors are close in power, they differ in time. The projections $proj_{p1}$ and $proj_{p2}$ of vectors on $proj_t$ are close to each other and by projection $proj_t$ they can also be assigned to the same cluster. The projections $proj_{p2}$ and $proj_t$ of vectors on the projection $proj_{tb}$ are also close, and the projection $proj_{p1}$ is significantly distant from the projection $proj_{p2}$, $proj_{p3}$, $proj_t$ on the projection $proj_{pb}$.

Such a variety of variants allows us to form different clusters depending on the selected projections and the distances between them.

Thus, the proposed clustering method provides variations in the selection of clusters by projections, by defining vectors whose projections differ from others, and, that is, provides various options for choosing projections, by a predetermined distance threshold for clustering. And the transition by the matrix of similarity metrics from one option to another makes it possible to take into account various features of systems in time for dynamic clustering.

It should be noted that the vector of weight coefficients, which is typically formed based on expert assessment, should be selected based on the practical conditions of microgrid operation in the general grid. For example, the weight of the energy share generated during hours of peak energy consumption may be several times greater than the weight, related to quiet periods. For the given example, the vectors of weight coefficients for functions $P_2(t)$ and $P_3(t)$ could be $\begin{bmatrix}2 & 1 & 1 & 1\end{bmatrix}$, and $\begin{bmatrix}4 & 1 & 1 & 1\end{bmatrix}$, respectively. Multiplying the original vectors by the vectors of weight coefficients significantly increases the impact of peak loads.

One application of this approach is related to load analysis in the general power grid to which individual microgrids with solar panels and wind turbines are connected. The proposed approach allows for varying the choice of parameters used to determine the distances between parameter vectors and for selection for comparison and clustering.

CONCLUSION

In the proposed clustering method, the transition to some abstract sequences allows preserving the features of the original data while significantly simplifying the subsequent transition to operations with vectors of cardinal numbers that are formed.

At the same time, the method allows you to use different options for clustering with orientation towards

TABLE 2. MATRIX D OF SIMILARITY METRICS

		1	2	3	4	5	6
		$proj_{p1}$	$proj_{p2}$	$proj_{p3}$	$proj_t$	$proj_{pb}$	$proj_{tb}$
1	$proj_{p1}$	0	1,09	0,76	1,66	5,46	2,77
2	$proj_{p2}$	1,09	0	0,55	1,63	4,37	1,94
3	$proj_{p3}$	0,76	0,55	0	1,97	4,80	2,50
4	$proj_t$	1,66	1,63	1,97	0	4,92	1,79

DOI: 10.20535/2523-4455.mea.344933

a particular base vector or to vectors depending on the selected projections.

For any clustering options, the calculations are formalized, simple and easily algorithmized, which makes it possible to implement dynamic clustering.

REFERENCES

[1]. O. Kyrylenko, B. Stognii, S. Denysiuk, and M. Sopel, "SMART-MONITORING OF ELECTRICAL POWER SYSTEMS", Tekhnichna Elektrodynamika, vol. 2024, no. 5, pp. 48–62, Aug. 2024. DOI: 10.15407/techned2024.05.048.

[2]. G. J. Oyewole and G. A. Thopil, "Data clustering: application and trends", Artificial Intelligence Review, vol. 56, no. 7, pp. 6439–6475, Nov. 2022. DOI: 10.1007/s10462-022-10325-y.

[3]. B. S. Everitt, S. Landau, M. Leese, D. Stahl, Cluster Analysis, London, Wiley, 2011, 352 p. DOI: 10.1002/9780470977811

[4]. P.-N. Tan, M. Steinbach, A. Karpatne, V. Kumar, Introduction to Data Mining. 2nd ed. Boston, MA: Pearson, 2021, 856 p.

[5]. A. Ultsch and J. Lötsch, "Euclidean distance-optimized data transformation for cluster analysis in biomedical data (EDOtrans)", BMC Bioinformatics, vol. 23, no. 1, Jun. 2022. DOI: 10.1186/s12859-022-04769-w

[6]. S. A. Afghani and W. Y. M. Putra, "Clustering with Euclidean Distance, Manhattan - Distance, Mahalanobis - Euclidean Distance, and Chebyshev Distance with Their Accuracy", Indonesian Journal of Statistics and Its Applications, vol. 5, no. 2, pp. 369–376, Jun. 2021. **doi**: 10.29244/ijsa.v5i2p369-376

[7]. Y. Shim, S.-W. Choi, M.-G. Yang, K.-Y. Chung, and K.-H. Baek, "Energy Efficient Distance Computing: Application to K-Means Clustering", Electronics, vol. 11, no. 3, p. 298, Jan. 2022. **doi**: 10.3390/electronics11030298.

[8]. M. Sari and A. Armansyah, "Manhattan Metric Technique in K-Means Clustering for Data Grouping", Journal of Information Systems and Informatics, vol. 6, no. 3, pp. 1945–1961, Sep. 2024. **doi**: 10.51519/journalisi.v6i3.841

[9]. D. Al-Shammary, E. Hakem, A. M. Mahdi, A. Ibaida, and K. Ahmed, "A novel brain EEG clustering based on Minkowski distance to improve intelligent epilepsy diagnosis", Informatics in Medicine Unlocked, vol. 47, p. 101492, Jan. 2024. **doi**: 10.1016/j.imu.2024.101492.

[10]. S. Biswas, A. Fole, N. Khare, and P. Agrawal, "Enhancing correlated big data privacy using differential privacy and machine learning", Journal of Big Data, vol. 10, no. 1, Mar. 2023. **doi**: 10.1186/s40537-023-00705-8

[11]. N. Wiroonsri, "Clustering performance analysis using a new correlation-based cluster validity index", Pattern Recognition, vol. 145, p. 109910, Jan. 2024. **doi**: 10.1016/j.patcog.2023.109910.

[12]. J. AlShaqsi, W. Wang, O. Drogham, and R. S. Alkhawaldeh, "Quantitative and qualitative similarity measure for data clustering analysis", Cluster Computing, vol. 27, no. 10, pp. 14977–15002, Aug. 2024. **doi**: 10.1007/s10586-024-04664-4.

[13]. X. Ran, Y. Xi, Y. Lu, X. Wang, and Z. Lu, "Comprehensive survey on hierarchical clustering algorithms and the recent developments", Artificial Intelligence Review, vol. 56, no. 8, pp. 8219–8264, Dec. 2022. **doi**: 10.1007/s10462-022-10366-3.

[14]. M. Jasiński, T. Sikorski, Z. Leonowicz, K. Borkowski, and E. Jasińska, "The Application of Hierarchical Clustering to Power Quality Measurements in an Electrical Power Network with Distributed Generation", Energies, vol. 13, no. 9, p. 2407, May 2020. **doi**: 10.3390/en13092407

[15]. A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data", Information Sciences, vol. 622, pp. 178–210, Apr. 2023. **doi**: 10.1016/j.ins.2022.11.139.

[16]. M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means Algorithm: A Comprehensive Survey and Performance Evaluation", Electronics, vol. 9, no. 8, p. 1295, Aug. 2020. **doi**: 10.3390/electronics9081295.

[17]. Y. Terada, "Strong consistency of factorial $$K$$ K -means clustering", Annals of the Institute of Statistical Mathematics, vol. 67, no. 2, pp. 335–357, Mar. 2014. **doi**: 10.1007/s10463-014-0454-0.

[18]. E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu, "DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN", ACM Transactions on Database Systems, vol. 42, no. 3, pp. 1–21, Jul. 2017. **doi**: 10.1145/3068335.

[19]. D. K. Kotary and S. J. Nanda, "A Distributed Neighbourhood DBSCAN Algorithm for Effective Data Clustering in Wireless Sensor Networks", Wireless Personal Communications, vol. 121, no. 4, pp. 2545–2568, Aug. 2021. **doi**: 10.1007/s11277-021-08836-y.

[20]. X. Yang, C. Xie, K. Zhou, S. Song, J. Yang, and B. Li, "Towards attributed graph clustering using enhanced graph and reconstructed graph structure", Artificial Intelligence Review, vol. 57, no. 11, Sep. 2024. **doi**: 10.1007/s10462-024-10958-1.

[21]. A. S. Tarasenko, V. B. Berikov, I. A. Pestunov, S. A. Rylov, and P. S. Ruzankin, "A fast consistent grid-based clustering algorithm", Pattern Analysis and Applications, vol. 27, no. 4, Nov. 2024. **doi**: 10.1007/s10044-024-01354-0.

[22]. I. Strelkovska, I. Solovska, Yu. Strelkovska, V. Paskalenko, "Kompleksna splain-aproksymatsiia v zadachakh pozytsionuvannia", *Visti vyshchykh uchbovykh zakladiv. Radioelektronika*, № 65 (7), pp. 445-456, 2022. **doi**: 10.20535/S0021347022100028.

[23]. M. Du and F. Wu, "Grid-Based Clustering Using Boundary Detection", Entropy, vol. 24, no. 11, p. 1606, Nov. 2022. **doi**: 10.3390/e24111606.

[24]. M. Al Jreidy, J. Constantin, F. Dornaika, and D. Hamad, "Clustering using graph convolution networks", Progress in Artificial Intelligence, Jan. 2024. **doi**: 10.1007/s13748-023-00310-z.

[25]. C. Weber, D. Ray, A. Valverde, J. Clark, and K. Sharma, "Gaussian mixture model clustering algorithms for the analysis of high-precision mass measurements", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol. 1027, p. 166299, Mar. 2022. **doi**: 10.1016/j.nima.2021.166299.

[26]. S. Sarkar, V. Melnykov, and R. Zheng, "Gaussian mixture modeling and model-based clustering under measurement inconsistency", Advances in Data Analysis and Classification, vol. 14, no. 2, pp. 379–413, May 2020. **doi**: 10.1007/s11634-020-00393-9.

[27]. Y. Jie, X. Li, M. wang, F. Zhou, and H. Tan, "Medical image fusion based on extended difference-of-Gaussians and edge-preserving", Expert Systems with Applications, vol. 227, p. 120301, Oct. 2023. **doi**: 10.1016/j.eswa.2023.120301.

[28]. H. Huang, "Fuzzy K-means clustering with reconstructed information", International Journal of Machine Learning and Cybernetics, vol. 16, no. 1, pp. 43–53, Jun. 2024. **doi**: 10.1007/s13042-024-02167-7.

[29]. A. Javed, D. M. Rizzo, B. S. Lee, and R. Gramling, "Somtimes: self organizing maps for time series clustering and its application to serious illness conversations", Data Mining and Knowledge Discovery, vol. 38, no. 3, pp. 813–839, Oct. 2023. **doi**: 10.1007/s10618-023-00979-9.

[30]. M. R. Basarab and K. O. Ivanko, "Deep Learning for the Detection and Classification of Diabetic Retinopathy Stages", Microsystems, Electronics and Acoustics, vol. 29, no. 2, Aug. 2024. **doi**: 10.20535/2523-4455.mea.309642.

[31]. D. S. Zaruba, M. Y. Shvets, and Y. V. Khokhlov, "Machine Learning for a Power Consumption and Generation Prediction", Microsystems, Electronics and Acoustics, vol. 24, no. 6, pp. 17–21, Dec. 2019. **doi**: 10.20535/2523-4455.2019.24.6.186996.

[32]. B. Zhou, B. Lu, and S. Saeidlou, "A Hybrid Clustering Method Based on the Several Diverse Basic Clustering and Meta-Clustering Aggregation Technique", Cybernetics and Systems, vol. 55, no. 1, pp. 203–229, Aug. 2022. **doi**: 10.1080/01969722.2022.2110682.

[33]. H. Nabli, R. Ben Djemaa, and I. Amous Ben Amor, "Improved clustering-based hybrid recommendation system to offer personalized cloud services", Cluster Computing, vol. 27, no. 3, pp. 2845–2874, Aug. 2023. **doi**: 10.1007/s10586-023-04119-2.

[34]. C. J. Date, An Introduction to Database Systems, 8th ed. Boston, MA, Addison-Wesley Professional, 2003, 1024 p.

Надійшла до редакції 08 листопада 2025 року
Прийнята до друку 23 грудня 2025 року

DOI: 10.20535/2523-4455.mea.344933

УДК 004.942:621

Метод кластеризації з використанням векторів кардинальних чисел

О. О. Абакумова, канд. філос. наук, доц., 0000-0002-5467-2473
В. Я. Жуйков, д-р техн. наук, проф. 0000-0002-3338-2426
Національний технічний університет України
"Київський політехнічний інститут імені Ігоря Сікорського" 00syn5v21
Київ, Україна

М. О. Лук'янов, 0000-0001-8930-9992
Політехніка Гданська, 006x4sc24
Гданськ, Польща

Анотація–У сучасних технічних системах кластеризація є ключовою процедурою для структурування, аналізу та інтерпретації великих обсягів даних, що забезпечує підвищення ефективності прийняття рішень і оптимізації системних процесів. В роботі проведено порівняльний аналіз основних груп методів кластеризації та запропоновано новий підхід до кластеризації з використанням потужного математичного апарату кардинальних чисел. Розкрито теоретичні засади побудови векторів кардинальних чисел як математичного інструменту для представлення даних у задачах кластеризації. Запропоновано визначення відстаней між об'єктами здійснювати в обраному ортогональному базисі на основі підрахунку потужностей абстрактних послідовностей множин, поданих у вигляді векторів кардинальних чисел. Розглянуто формування векторів кардинальних чисел та обчислення відповідних метрик схожості з подальшим формуванням матриці відстаней. Наведено приклад розрахунку відстаней між трьома функціями та базовою функцією на основі аналізу відповідних векторів кардинальних чисел. Показано, що зміна базису чи проекцій залежно від технічної задачі дозволяє формувати різні кластери, що свідчить про гнучкість і адаптивність запропонованого підходу. Розрахунки формалізовані, прості та легко алгоритмізуються, що дає можливість для втілення динамічної кластеризації. Такий підхід є перспективним для застосування в інтелектуальних системах аналізу даних та обробки інформації в електронних пристроях.

Ключові слова: кластеризація; вектори; кардинальні числа; метрика схожості; локальний об'єкт; ортонормований базис.