fram Mathematica system allows variants of the clustering method (Criterion Function): Automatic, Agglomerate, DBSCAN, GaussianMixture, JarvisPatrick, KMeans, KMedoids, MeanShift, NeighborhoodContraction, Optimize, SpanningTree, Spectral [10]. What segmentation methods are used in the calculations are written in the headings of the diagrams. Representative clustering options are shown, namely K Means (k-means clustering algorithm), k-medoids (partitioning around medoids), Optimal (Wolfram Mathematica method). The effects of the accepted clustering method (Possible settings for Method) are illustrated by the schemes in Figure 6. Clustering in the examples of this series was considered for three parameters, the FindClusters function was used, the norm in the examples of the series in Figure 6 was not set, but was determined by the default calculation module. These results are quite indicative. At the same time, taking into account the reference and the digital field of the original, we can consider the clustering options by the KMeans and Optimal methods as preferable

B. The impact of the metric

The issues of measuring the proximity of objects have to be solved with any interpretation of clusters and various classification methods, moreover, there is an ambiguity in choosing the method of normalization and determining the distance between objects. The influence of the metric (DistanceFunction) is illustrated by the diagrams in Figure 7. The results presented in this series are obtained by means of the corresponding software application included in the GeoBazaDannych from the Wolfram Mathematica, which allows different options for setting DistanceFunction.

The Wolfram Language provides built-in functions for many standard distance measures, as well as the capability to give a symbolic definition for an arbitrary measure. In particular, the following metric variant sare available for analyzing digital data [10]: EuclideanDistance, SquaredEuclideanDistance, NormalizedSquaredEuclideanDistance, ManhattanDistance, ChessboardDistance, BrayCurtisDistance, CanberraDistance, CosineDistance, CorrelationDistance, BinaryDistance, WarpingDistance, CanonicalWarpingDistance. What methods of DistanceFunction are used in calculations is recorded in the headers of the schemes. Representative variants are shown, namely: EuclideanDistance (the length of a line segment between the two points), ChessboardDistance, SquaredEuclideanDistance, BrayCurtisDistance, ChebyshevDistance (a metric defined on a vector space where the distance between two vectors is the greatest of their differences along any coordinate dimension), ManhattanDistance. It follows from the above results that for the considered configuration of data points, taking into account the digital field of the original, clustering options using Spectral EuclideanDistance methods can be considered preferable.

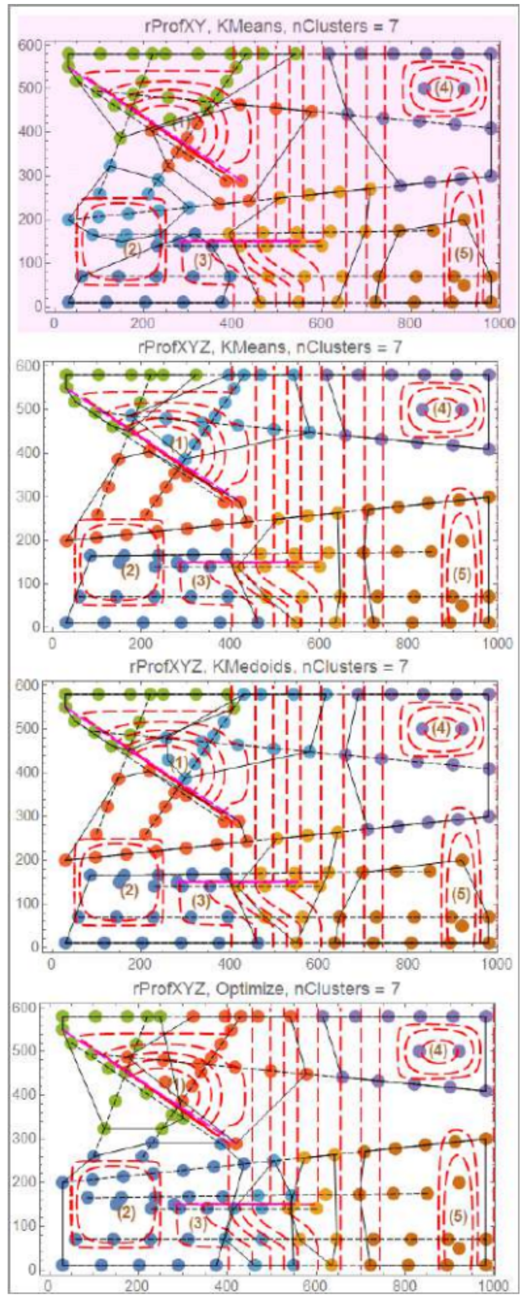

Figure 6. The effects of the accepted clustering method.

IV. Conclusion

The issues of instrumental filling and use of the interactive computer system GeoBazaDannych, expansion of its functionality through integration with the Wolfram Mathematica computer algebra system are considered. A modification of the typical clustering method is proposed, and computational experiments have confirmed the advantages in comparison with traditional methods.

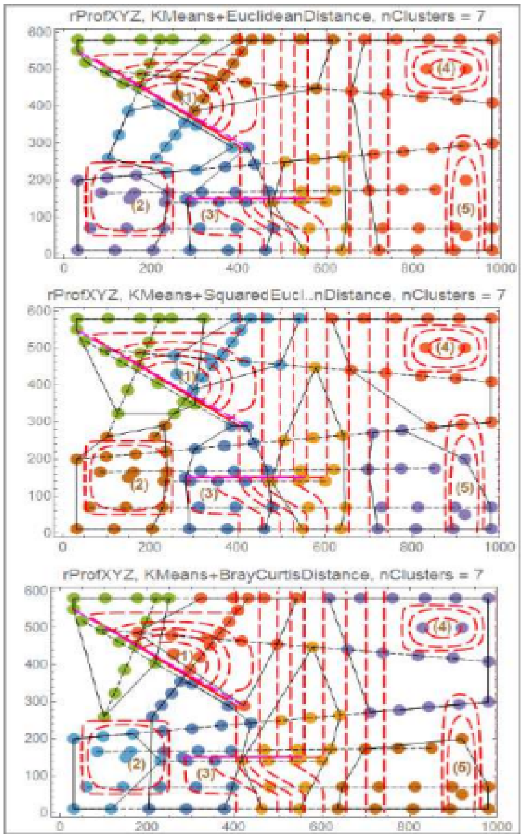

Figure 7. The effects of the accepted metric (EuclideanDistance, SquaredEuclideanDistance,BrayCurtisDistance).

Referenses

[1] . P. Savinyh, V. Ya. Cvetkov, “Geodannye kak sistemnyi informacionnyi resurs”, Vestnik Rossiiskoi akademii nauk, vol. 84, no.9, pp. 826–829, 2014, (in Russian).

[2] . Golenkov, N. Guliakina, and D. Shunkevich, Otkrytaja tehnologija ontologicheskogo proektirovanija, proizvodstva i jekspluatacii semanticheski sovmestimyh gibridnyh intellektual'nyh komp'juternyh sistem [Open technology of ontological design, production and operation of semantically compatible hybrid intelligent computer systems], V. Golenkov, Ed. Minsk: Bestprint [Bestprint], 2021. 690 p.

[3] . Taranchuk, Komp'yuternye modeli podzemnoi gidrodinamiki. Minsk: BGU, 2020. 235 p., (in Russian)

[4]V. Taranchuk, “Interactive Adaptation of Digital Fields in the System GeoBazaDannych”, Communications in Computer and Information Science. Book series Springer, vol. 1282, –2020, P. 222–233.

[5]M. Abdideh, A. Ameri, “Cluster Analysis of Petrophysical and Geological Parameters for Separating the Electrofacies of a Gas Carbonate Reservoir Sequence. Nat Resour Res 29, 1843–1856 (2020). https://doi.org/10.1007/s11053-019-09533-1

[6]M. A. Tkachenko, Y. V. Karelina “A cluster analysis as a method of identifying potential chromite mineralisation within the Voykar-Synyinsky massif”. International Research Journal. — 2023. — 10 (136). DOI: 10.23670/IRJ.2023.136.63

[7]V. Taranchuk, “Integration of computer algebra tools into OSTIS applications”. Otkrytye semanticheskie tekhnologii proek V. Taranchuk, “Integration of computer algebra tools into OSTIS applications”. Otkrytye semanticheskie tekhnologii proek

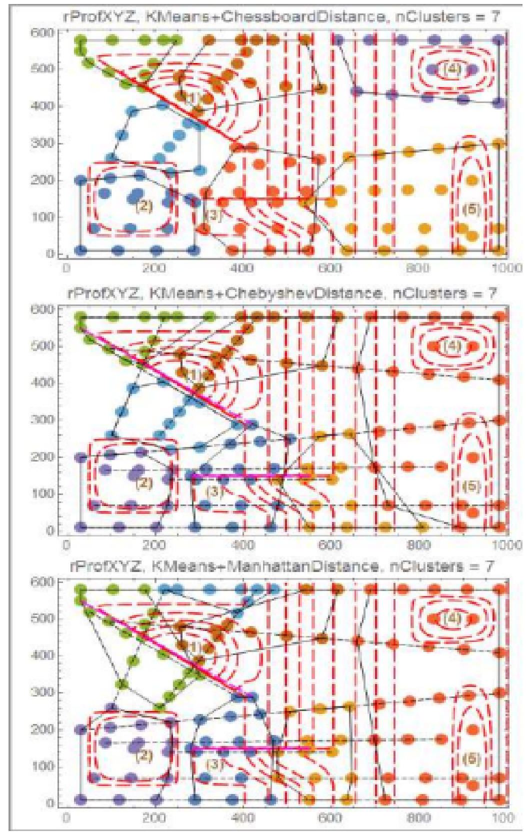

Figure 8. The effects of the accepted metric (ChessboardDistance, ChebyshevDistance, ManhattanDistance).

tirovaniya intellektual'nykh system [Open semantic technologies for intelligent systems], 2022, no 6, pp. 369-374.

[8]D. Tupper Charles, “Concepts of Clustering, Indexing, and Structures”, Data Architecture, 2011, pp. 241–253, https://doi.org/10.1016/B978-0-12-385126-0.00013-9.

[9]B. S. Everitt, S. Landau, M. Leese, D. Stahl. Cluster Analysis. 5th Edition, John Wiley Sons, 2011, 360 p.

[10]Distance and Similarity Measures. https://reference.wolfram.com/language/ guide/ DistanceAndSimilarityMeasures.html/ (accessed 2024, Feb).

ПРИМЕРЫ ИНТЕГРАЦИИ МОДУЛЕЙ ИНТЕЛЛЕКТУАЛЬНЫХ ВЫЧИСЛЕНИЙ И СИСТЕМЫ ГЕОБАЗАДАННЫХ

Таранчук В.Б.

Обсуждаются методические и технические решения интеграции модулей интеллектуальных вычислений системы Wolfram Mathematica и инструментов программного комплекса ГеоБазаДанных

в задачах формирования, интерпретации, обработки, визуализации цифровых полей при компьютерном моделировании объектов геологии, геоэкологии. Предложена модификация типового способа кластеризации, расчетами на представительных данных подтверждены преимущества.

Received 04.04.2024

Designing Intelligent Systems with Integrated Spatially Referenced Data

Sergei Samodumkin
Belarusian State University of Informatics and Radioelectronics
Minsk, Belarus
Email: Samodumkin@bsuir.by

***Abstract*—The paper is devoted to the issues of representation, integration and processing of spatially referenced data in intelligent systems built on the principles of ostissystems.**

***Keywords*—OSTIS, intelligent system with integrated spatially referenced data, sema**

I. Introduction

Large sets of spatially referenced data accumulated by mankind, their representation and storage through the created cartographic services [1]–[3], the development of remote sensing technologies [4] have contributed to the creation and development of applied geoinformation systems for various purposes

Modern geographic information systems are computer systems that provide input, manipulation, analysis and output of spatially referenced data about the territory, social and natural phenomena in solving tasks related to inventory, analysis, modeling, forecasting and management of the environment and territorial organization of society

Since the above tasks are intelligent, such systems belong to the class of intelligent systems with integrated spatially referenced data (ISRD)

The currently offered ISRD development tools are not sufficiently interoperable due to the lack of unification of knowledge of subject areas for the benefit of which application systems are designed, ontologies of terrain objects and phenomena, and temporal components.

It is obvious that for a fixed territory the same spatially related data are used in different application areas: epidemiology, construction, environmental protection and nature management, land relations, creation of digital twins of enterprises, mobile robotics systems, etc., which determines that it is necessary to harmonize the ontologies of subject areas with the objects of terrain and phenomena inherent in a given territory, thus ensuring a vertical (subject-oriented) level of design. On the other hand, when designing ISRD for a new territory, the basic functional requirements are preserved and it is necessary to take into account not only the previous experience of system design, but also to use previously designed functional components, i. e. we are talking about the horizontal level of ISRD design when the territorial area is expanded and systems are designed for new territories.

The third aspect is the temporal component, relevant for retrospective analysis and modeling, thus ensuring the creation of dynamic ISRD that can deal with terrain objects and phenomena within a specific time period

Therefore, the constant evolution of models and tools for ontological description of subject areas using spatial and temporal components, heterogeneity of spatial components and ambiguity of temporal components, poses new challenges in terms of interaction, integration and interoperability of different types of knowledge used in ISRD by integrating subject area ontologies (vertical level), extending systems at the horizontal level to new territories and time intervals, re-using

The necessity of solving the indicated problems determines the demand for intelligent systems with integrated spatially referenced data and indicates the existence of scientific and technical problem of intellectualization of systems with integrated spatially referenced data, and also becomes relevant to create the development tools themselves, which ultimately provides information support and automation of the activities of developers of application systems.

Within this article, fragments of structured texts in the SCn-code [5] will often be used, which are simultaneously fragments of the source texts of the knowledge base, understandable to both human and machine. This allows making the text more structured and formalized, while maintaining its readability. The symbol ":="in such texts indicates alternative (synonymous) names of the described entity, revealing in more detail certain of its features

II. Knowledge-based approach in the tasks of representation, integration and processing of spatially related data

The importance and necessity of analyzing data in space and time is primarily to discover hidden connections and patterns that may not be obvious at first glance. Representation of spatial data, correlating them