An Approach for Clustering of Seismic Events using Unsupervised Machine Learning

Markhaba Karmenova

D. Serikbayev East Kazakhstan Technical University
School of Information Technology and Intellectual Systems
A. K. Protazanov Str. 69, 070004 Ust-Kamenogorsk, Kazakhstan
e-mail: MKarmenova@ektu.kz

Aizhan Tlebaldinova

S. Amanzholov East Kazakhstan University
Kazakhstan Str. 55, 070004 Ust-Kamenogorsk, Kazakhstan
e-mail: atlebaldinova@vku.edu.kz

Iurii Krak

Taras Shevchenko National University of Kyiv
Volodymyrska Str. 60, 01033 Kyiv, Ukraine
e-mail: krak@unicyb.kiev.ua

Natalya Denissova, Galina Popova

D. Serikbayev East Kazakhstan Technical University
A. K. Protazanov Str. 69, 070004 Ust-Kamenogorsk, Kazakhstan
e-mail: {NDenisova, GPopova}@ektu.kz

Zheniskul Zhantassova

S. Amanzholov East Kazakhstan University
Kazakhstan Str. 55, 070004 Ust-Kamenogorsk, Kazakhstan
e-mail: zhantasova@vku.edu.kz

Elena Ponkina

Altai State University
Lenin Ave. 61, 656049 Barnaul, Russian Federation
e-mail: ponkina@math.asu.ru

György Györök

Óbuda University, Alba Regia Technical Faculty
Budai út 45, H-8000 Székesfehérvár, Hungary
gyorok.gyorgy@amk.uni-obuda.hu

Abstract: New and effective approaches for the analysis of seismic data make it possible to identify the distribution of earthquakes helping further to assess frequency of occurrence any associated risks. This paper proposes an effective approach for detecting areas with increased spatial density of seismic events and zoning territories on the map based on the Density-based Spatial Clustering of Applications with Noise algorithm (DBSCAN algorithm). The validity of the choice of this clustering algorithm is explained by the fact that the DBSCAN algorithm can detect clusters of complex shapes including geographical coordinates. This study uses seismic data from the seismic catalog of the Republic of Kazakhstan from 2011 to 2021 inclusive. Finally, the clusters detected over a certain period of time allowed for the presentation of a spatial model of the distribution of earthquakes and the detection of areas with increased spatial density on the map. In general, the results of the study were also compared and well associated with the general map of the seismic zoning of the Republic of Kazakhstan showing reliable results of clustering based on density. In addition, the architecture of intelligent information and the analytical system for analyzing seismic data is based on the proposed approach.

Keywords: data mining; machine learning; data analysis; clustering; DBSCAN algorithm; DBSCAN algorithm's parameters determination; intelligent information and analytical system

Introduction

Studies related to the prediction of earthquakes occurrence using large arrays of various types of seismic data are of considerable scientific and practical interest [1-3]. This is facilitated by the intensive development of intellectualization of data processing and analysis methods as well as the effective use of machine learning methods in solving such problems providing even more opportunities for their in-depth study and identification of hidden patterns. The significance and importance of these problems gives rise to interesting research problems divided into two large classes.

Problems of the first class are associated with the use of unsupervised machine learning methods [4-6]. For example, in work [4], for the concept of the evolution of the properties of the seismic wave field during an earthquake in L'Aquila used methods of uncontrolled machine learning to three-year continuous seismic data. As a result of the study, individual clusters were identified well correlating with different periods of seismic cycle. The authors of work [5] propose a new approach for detecting and clustering seismic signals in continuous seismic records obtained during the June 2017 of Nuugaatsiak landslide in Greenland. This approach combines a deep scattering network and a Gaussian mixture model for clustering seismic signal segments and detecting new structures. The effectiveness of algorithm application of uncontrolled machine learning by fuzzy clustering presented in work [6] where this type of cluster analysis helps to recognize the shapes of seismic waves in microseismic data or in earthquake data.

Problems of the second class are associated with the use of methods of controlled machine learning [7-9]. In works [10, 11], earthquake prediction approaches based on the use of deep learning methods such as neural networks. In work [10], a method based on a deep convolutional neural network (CNN) is described for predicting earthquake intensity measurements. In work [11], a classification system based on the support vector machine (SVR) and hybrid neural network (HNN) is constructed to obtain earthquake forecasts. The application of the neural network model is also presented in work [12] where one of the approaches is considered to estimate the stress field of the earth's crust based on the focal mechanisms of earthquakes. In work [13], an earthquake prediction system proposed combining seismic indicators with the AdaBoost ensemble method based on genetic programming (GP). Seismic indicators are calculated to obtain maximum information about the seismic state of the region and transmitted to the GP and AdaBoost-based ensemble classification methodology (GP-AdaBoost).

Thus, there are different approaches and methodologies for analyzing seismic catalogs with different amounts of earthquake data and attributes. However, as noted in work [5] most likely uncontrolled instruments are the most suitable for studying seismic data without using any explicit signal model and consequently suitable for discovering new classes of seismic signals.

It should be noted that machine learning methods, in contrast to deterministic approaches to seismic data processing [14-16] have opened new prospects in scientific research for the rapid analysis of daily accumulated and growing volumes of seismic data.

1 Problem Formulation

The purpose of this work is to develop new and more effective approaches for detecting and clustering seismic data on the example of the study of seismic activity of the Republic of Kazakhstan (RK). Also, it should be noted that scientific research does not sufficiently use datamining technologies and machine learning methods for processing seismic data recorded by stations on the territory of the Republic of Kazakhstan. The paper proposes an approach for detecting areas with an increased spatial density of seismic events based on cluster analysis methods.

The paper has the following structure: Section 2 describes the data set and presents the methodology of a new cluster analysis approach for detecting areas with increased spatial density of seismic events. Section 3 describes the results obtained using cluster analysis. Section 4 presents the architecture of an intelligent information and analytical system for analyzing seismic data and demonstration of operation process of the system. Section 5, in conclusion, provides the results concerning the effectiveness of the approach and possible directions for further research.

2 Methodology

2.1 Data Set and Preprocessing

This study uses a data set of 1668 records containing following structure: Data, GreenwichMeanTime, Latitude, Longitude, Depth (km), Energy. classandMagnitude. The data is obtained from the source of LLP "SOME", JSC "National Center for Seismological Observations and Research" where the stations conduct seismic observations (registration of earthquakes) in the territory of the Republic of Kazakhstan [17]. The dataset under study includes seismic events (earthquakes) with the attributes described above from 2012 to 2021 inclusive.

In the process of cluster analysis of spatial seismic data by the DBSCAN algorithm (Subsection 2.2, Figure 1) used geographical coordinates (longitude, latitude). During the preliminary data processing, the geographical coordinates were converted to radians for correct calculation of distance between two points on the earth's surface. In this regard, the 'Haversine' formula is used as the distance metric between two points. This formula is usually written in terms of the Haversine Function:

$$haversine(\theta) = sin^2(\theta/2) \tag{1}$$

(1) measures the arc distance between two points on the sphere taking into account their longitude and latitude, where the arc distance of the great circle is the shortest distance between two points on the surface (2):

$$d = 2rsin^{-1}\left(\sqrt{sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + +cos(\varphi_1)cos(\varphi_2)sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right) \tag{2}$$

where *d* is the distance between two points, where φ is the latitude and λ is the longitude and *r* is the radius of the Earth (mean radius = 6,371 km).

In most studies, the ‘Haversine ' formula is often used as a metric of the distance between two points as well as the Euclidean distance. This metric is used in the basis of machine learning algorithms for clustering data sets such as seismic data [18], data on volcanoes and data on tsunamis caused by earthquakes [19].

2.2 Density-based Clustering

Density-based clustering algorithms determine clusters of points located in close proximity based on a given search distance. A feature of density clustering algorithms is that clusters can have an arbitrary shape. The most popular density-based clustering algorithm is the DBSCAN algorithm [20]. The advantages of the algorithm are not requiring specification of the number of clusters in the data priori and is resistant to outliers. In [21], presented a description of an application called seiscloud designed for clustering and visualizing local seismic catalogs. In seiscloud structure, the procedure for cluster analysis of seismic data is performed using the DBSCAN clustering algorithm based on characteristics of epicentral position and time of origin. The performance of the algorithm in seiscloud system is tested on earthquake data in northern Chile. According to the authors, the possibilities of considered software implementation in presenting additional characteristics of earthquakes (location, depth, magnitude, time of occurrence, etc.) also can be used to obtain seismic clusters Z. Fan and others [22] for cluster analysis of seismic data, K-means and DBSCAN algorithms. Their characteristics are compared by fitting seismic belts with seismic data sets. A comparison of these algorithms for clustering seismic data has shown that the density-based DBSCAN algorithm is far superior to the K-means algorithm for studying the fit of seismic belts.

In fact, the control of algorithm is based on two parameters: the neighborhood of a given radius and the minimum number of objects. The DBSCAN algorithm identifies clusters detecting all the main points and expands each one to all the points achievable by density. If no more nodes are found in extended environment, the cluster is terminated. Further, remaining points continue to be checked to see if another central point can be found to start a new cluster. After processing all the points, non-main points assigned to cluster are considered noise.

Regarding this study, the data points primarily represent the points of recorded seismic events or precise the earthquakes. In this regard, some isolated points detected by the algorithm will not be reachable by density from the main or node points, thus located in regions with low density and will not be assigned to any cluster. It should be noted that these points can also represent some areas of seismic zones according to historical data. Such patterns can be associated with local geological reliefs, the structure of rocks of the earth's crust and faults.

It was previously noted that the DBSCAN algorithm can find clusters of arbitrary shape detecting connection in various seismic data in terms of sample distribution density [23-24].

The DBSCAN algorithm uses a similarity matrix and as previously noted the algorithm is implemented based on the neighborhood of a given radius (Eps) and the minimum number of objects (MinPts, hereinafter M). Also, the DBSCAN algorithm determines the number of K clusters in operation. Let Eps be the neighborhood of object X.

$$D(X, Eps) = \{Y \in S: P(X, Y) \leq Eps\} \tag{3}$$

a core or base object of degree M with a given Eps value will be such an object where the Eps contains at least M of other objects. In the case where the value of M is pre-set, then the object Y is directly tightly reachable from the object X, if Y∈S (X, Eps) then the object X is the base one. Also, if object Y is densely reachable from object X and if there are such objects X_1,..., X_n, where X_1=X, X_n=Y, then object X_i+1 is directly densely reachable from X_i for all i=1,..., n-1.

The general sequence of actions of the DBSCAN algorithm:

Step 1. Set the values of the Eps and M, K=0 parameters.

Step 2. If all objects X∈S are viewed, then full stop. Otherwise, some of them selected and marked as viewed.

Step 3. If X is the base object, then a new cluster is created assuming that the value of K:=K+1 and proceeds to Step 4. In the opposite case, the point X is marked as "noise". It should be noted that during execution, this point may be in the Eps neighborhood of some other point and be included in another cluster. So, proceed to Step 2.

Step 4. The resulting cluster will include all objects densely reachable from the base object X, so proceed to Step 2.

The detailed implementation DBSCAN algorithm presented on Figure 1 [25].

```
Input Data: Data Set D
Output Data: A set of clusters C
Parameters: A set of points x, Eps, MinPts
Main function of the algorithm:
procedure DBSCAN(X, Eps, MinPts)
  for each unvisited point x∈X do
  mark x as visited
  N←dist(x, Eps)
  if |N| < MinPts then
    mark x as noise
  else
C←{x}
  for each point x'∈N do
  N← N\x'
    if x 0 is not visited then
    mark x' as visited
    N'← dist(x', Eps)
      if |N'| ≥ MinPts then
      N←N∪N'
      if x' is not yet member of any cluster then
      C← C∪{x'}
```

Figure 1
Pseudocode of the DBSCAN algorithm

Thus, the DBSCAN algorithm can combine two points which are far from each other and form an arbitrary cluster. These clusters may be located near the seismic faults and relative to their surrounding areas. Thus, displaying a highly active seismic zone, that connects the hypocenters of several earthquakes.

2.3 Determination of the Parameters and Clustering Results

As mentioned earlier (Section 2), the Eps and MinPts parameters are required for the DBSCAN algorithm. To determine the optimal value of the Eps parameter used the *k*-NN Distance method (*k*-nearest neighbor distance) [20]. The essence of the method is to calculate the distances in the matrix of points to the *k*-nearest neighbors. The average distance from each point to the *k*-nearest neighbors is calculated. The *k* value is set according to the MinPts values. The *k*-distance data is displayed in ascending order. The value corresponding to the threshold where a sharp change in the distance curve *k* occurs is selected. As can be seen from Figure 2, the optimal value for Eps is 0.005.

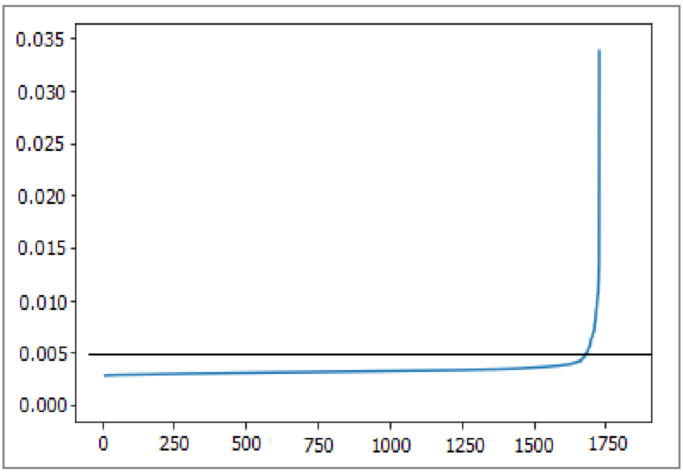


Figure 2
Determination of optimal value for Eps

3 Clustering Results

“Physically, seismic provinces should not have lots of changes in different periods of time. This implies that the number of clusters should have been almost the same in all periods of time [26]”. Based on the found Eps parameter and as a result of a number of experiments, the optimal MinPts values were selected, the results are presented in Table I.

Table1

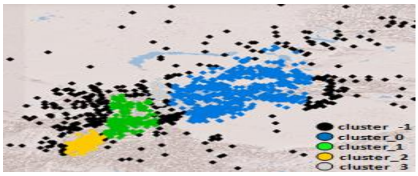
Selecting MinPts parameters

Eps	MinPts	Number of Clusters	Noise Ratio
0.005	20	7	23,3%
	21	8	26%
	22	8	25,8%
	23	8	27%
	24	**5**	31.1%
	25	**5**	**32**%
	26	**5**	**34,2**%
	27	**5**	**36,5**%
	28	**5**	**37,8**%
	29	**5**	**40,3**%
	30	6	41,2%
	31	6	43,1%

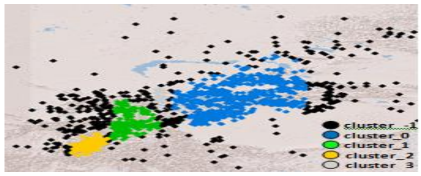
According to the data presented in Table 1, the number of clusters varies differently having the same Eps value and gradual increase in the MinPts values, as for the data identified as NoiseRatio noise gradually increases. Reducing the volume of noisy NoiseRatio data leads to a decrease in the number of clusters and to a decrease in accuracy of obtained results. As a result, the MinPts values were selected as the optimal values where the number of clusters have constant value. Seismic areas with increased spatial density with optimal values are shown in Figure 3.

The results of spatial clustering based on the density of points, identified 5 clusters and shows that the largest part of the dense spatial points (cluster_0) belongs to the Jungar-North Tien-Shan region, i.e. belongs to the seismically active belt of Kazakhstan and adjacent territories of other countries. It is well known that the Tien Shan region and the entire territory of Central Asia remain earthquake-prone regions due to thousands of active faults including the main Dzungarian fault. This fact, first of all, explains the very nature of occurrence of earthquake focus.

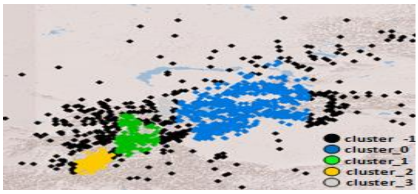
Eps=0.005, MinPts=24,Number of clusters=5



Eps=0.005, MinPts=25, Number of clusters=5



Eps=0.005, MinPts=26,Number of clusters=5



Eps=0.005, MinPts=27, Number of clusters=5

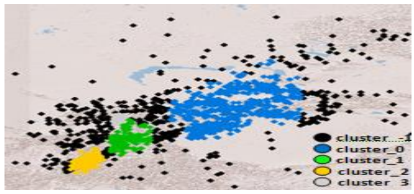


Figure 3

Outputs of the DBSCAN algorithm from optimal parameters

When comparing identified clusters with the general map of the seismogenerating zones of the territory of Kazakhstan [27], following zones were identified: South-Balkhash, Degeresskaya, Almaty, Zailiyskaya, Keminskaya, North-Kungeiskaya, Dubunskaya, Ketmenskaya, North-Ketmenskaya, Central-Ketmenskaya, Basulytauskaya, Bayankolskaya. Also, cluster_0 includes the borders of neighboring states. Cluster_1 and cluster_2 also refer to the regions of neighboring countries. Cluster_3, as well as Cluster_1 and Cluster_2 does not belong to the territory of the Republic of Kazakhstan. Points which are not associated with any cluster by the algorithm marked as noise points (cluster_-1), as well as individual distant noise points require additional research.

4 The Architecture of an Intelligent Information and Analytical System

Over the past decades, the growth of the flow of information requires a quick and high-quality assessment of the accumulated data. The rapid development of data science at the moment provides this opportunity. At the heart of many modern information systems, geographic information systems (GIS), algorithms and methods of machine learning are used to study and prevent various types of geological hazards [28-29]. The main task of such systems is to predict and prevent various risks including the prevention of seismic hazards [30]. In the future, for this kind of research, other research methods can also be applied by types of other earthquake data, such as pattern recognition [31].

In general, an intelligent information system is a separate set of tools, methods and personnel that has the ability to store, process and issue information, as well as independently adjust its parameters depending on the state of the external environment (i.e. initial data) and the specifics of the problem being solved [32]. In this regard, in the developed architecture of the intelligent information and analytical system, the processes of collection, processing, analysis of data sets and the use of machine learning methods are implemented. Also, when processing and analyzing seismic data, the DBSCAN algorithm, depending on the accumulated data, adjusts the clustering results and determines the number of earthquake clusters.

Thus, based on the results of spatial clustering by the DBSCAN algorithm, one of the modules (Module 2) of the intelligent information and analytical system for the analysis of seismic data was designed. The complete architecture of the intelligent information and analytical system is shown in Figure 4.

In addition, the intelligent information and analytical system also analyzes data related to the assessment of the earthquake resistance of urban objects. Other machine learning methods, such as clustering (kmeans/hkmeans) and classification methods (decision tree and random forest), are used as algorithms for evaluating an urban object for earthquake resistance [33-34]. The results of the study of the data set for Module 1 and the implementation of the assessment of urban objects for earthquake resistance are described in other scientific papers [35-36].

In the intelligent information and analytical system architecture, the designed modules have such properties as functional integrity and completeness. Each module implements one function and independently performs a complete set of operations to implement its function. At the input, the program module receives a certain set of source data and performs meaningful processing, thereby returning one set of result data. Also in the intelligent information and analytical system architecture there is a logical independence, the result of the program module depends only on the source data, but does not depend on the work of other modules.

As previously noted, in the intelligent information and analytical system architecture in Module 2, seismic data from storage 2 is processed and analyzed.

The intelligent information and analytical system architecture uses a virtual data warehouse model and an ETL method for loading data into the storage. Next, the data is structured and transformed and loaded into the storage for further analysis.

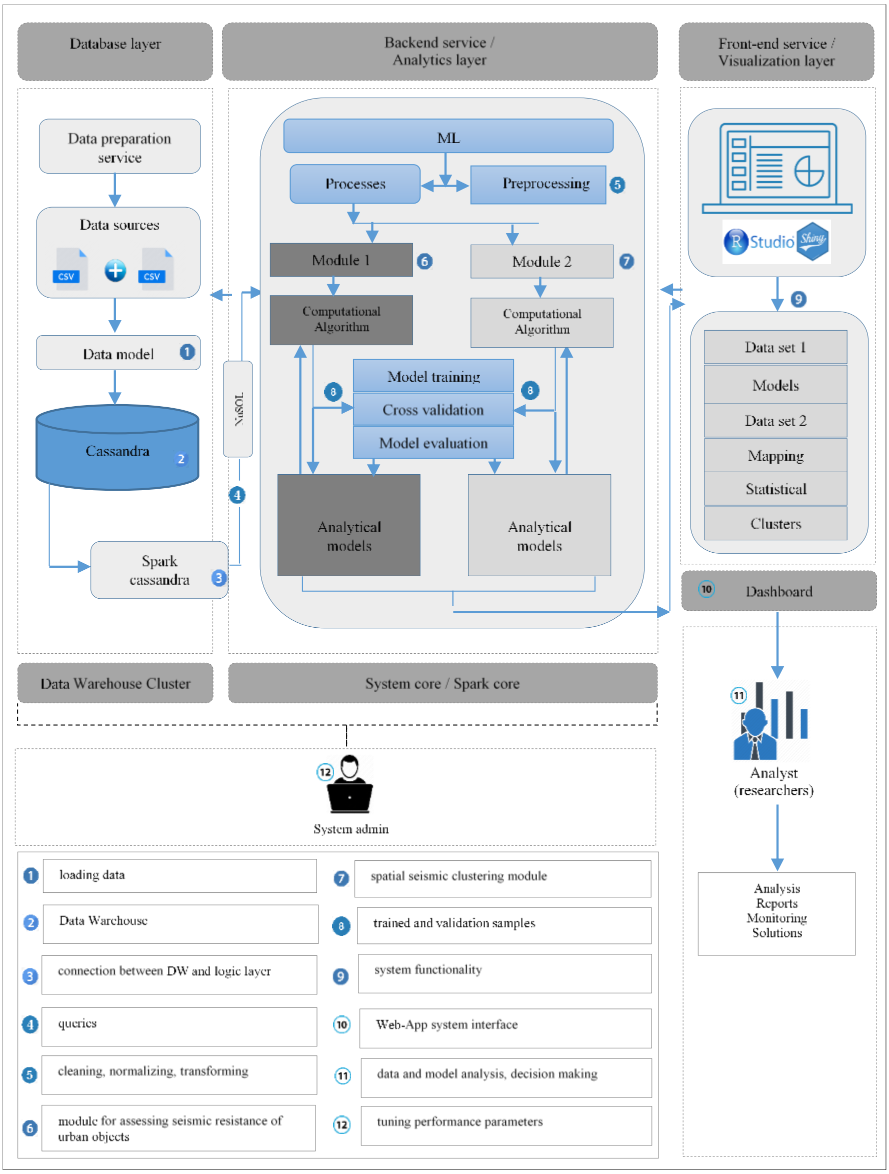


Figure 4

The intelligent information and analytical system architecture

Module 2 also contains sections on statistical analysis of seismic data and their spatial clustering. The web interface of the system is developed using the Shiny toolkit in the Rstudio environment (Figure 5 shows one of the main pages of the system).

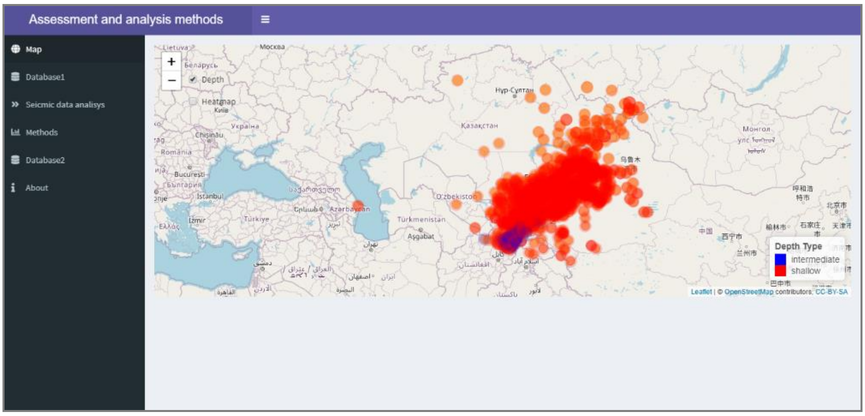


Figure 5
Module 2 of IIAS– a fragment of the visualization section on the map of seismic data on the magnitude and depth of the earthquake source

Conclusion

Nowadays, the natural hazard in the form of the earthquake is still impossible to accurately predict, but the analysis of seismic data based on data analysis technology and machine learning methods makes possible, the study of this phenomenon in greater depth. This paper proposes a new approach to detecting and clustering seismic data. According to the results of the study and cluster analysis, clusters with increased spatial density were found in the studied set of seismic data. The study of these clusters, in accordance with the general map of the seismogenerating zones of Kazakhstan allowed for the determination of appropriate seismic regions.

It should be noted that most of the earthquakes do not belong to the territory of the Republic of Kazakhstan based on the cluster analysis. Also, based on the obtained results one of the modules for statistical and cluster analysis of seismic data was designed in the intelligent information and analytical system architecture. The module for assessing urban objects for earthquake resistance and designed module for analyzing seismic data is also one of the main modules in the integrated architecture of the system. In general, the cluster analysis approach under consideration is also applicable to other data sets in this kind of research. The effectiveness of the proposed approach is proved on the data of the Republic of Kazakhstan. For future work, the results of density-based cluster analysis may be useful to further identify seismogenerating zones and determine the seismic potential of the earth's crust in Kazakhstan, as well as, for generally improving the quantitative assessment of the spatiotemporal distribution of the earthquakes.

Acknowledgements

We are grateful to LLP "SOME" and JSC "National Center for Seismological Observations and Research" of the Republic of Kazakhstan for providing information on earthquakes in an open Internet source. We also thank funding Goszadanie of Altai State University "Turk World of the Greater Altai: unity and diversity in the past and modern times" 748715F.99.1. BB97AA00002.

References

[1] R. Jena, B. Pradhan, Gh. Beydoun, A. M.Alamrid, Ardiansyah, Nizamuddin, H. Sofyan. Earthquake hazard and risk assessment using machine learning approaches at Palu, Indonesia. Science of the Total Environment, Pub Date: 2020-08-11, Vol. 749, p. 141582, DOI: 10.1016/j.scitotenv.2020.141582

[2] D.-H. Yang, X. Zhou, X.-Y. Wang, J.-P. Huang. Mirco-earthquake source depth detection using machine learning techniques. Information Sciences, Pub Date: 12 January 2021, V. 544, pp. 325-342, DOI: 10.1016/j.ins.2020.07.045

[3] R. Shcherbakov. Statistics and Forecasting of Aftershocks During the 2019 Ridgecrest, California, Earthquake Sequence. Journal Of Geophysical Research Solid Earth. 13 January 2021, V. 126(2), DOI:10.1029/2020jb020887

[4] P. Shi, L. Seydoux, P. Poli. Unsupervised Learning of Seismic Wavefield Features: Clustering Continuous Array Seismic Data During the 2009 L'Aquila Earthquake. Journal of Geophysical Research: Solid Earth, Pub Date: 2020-12-09, DOI: 10.1029/2020jb020506

[5] L. Seydoux, R. Balestriero, P. Poli, M. de Hoop, M. Campillo, R. Baraniuk. Clustering earthquake signals and background noises in continuous seismic data with unsupervised deep learning. Nature Communications, Pub Date: 2020-08-07, DOI: 10.1038/s41467-020-17841-x

[6] Y. Chen. Automatic microseismic event picking via unsupervised machine learning. Geophysical Journal International, Volume 212, Issue 1, January 2018, pp. 88-102, https://doi.org/10.1093/gji/ggx420

[7] P. Debnath, P. Chittora, T. Chakrabarti, P. Chakrabarti, Z. Leonowicz, M. Jasinski, R. Gonoand E. Jasińska. Analysis of Earthquake Forecasting in India Using SupervisedMachine Learning Classifiers. Sustainability, MDPI, Open Access Journal, January 2021, 13(2), pp. 1-13, DOI: 10.3390/su13020971

[8] Sh. Qu, Zh. Guan, E. Verschuur, Y. Chen. Automatic high-resolution microseismic event detection via supervised machine learning. Geophysical Journal International, V. 222, Issue 3, September 2020, pp. 1881-1895, https://doi.org/10.1093/gji/ggaa193

[9] Barmak, O., Krak, I., Manziuk, E.Diversity as the basis for effective clustering-based classification CEUR Workshop Proceedings, 2020, 2711, pp. 53-67

[10] D. Jozinović, A. Lomax, I. Štajduhar, A. Michelini. Rapid prediction of earthquake ground shaking intensity using raw waveform data and a convolutional neural network, Geophysical Journal International, Volume 222, Issue 2, August 2020, pp. 1379-1389, https://doi.org/10.1093/gji/ggaa233

[11] K. M. Asim, A. Idris, T. Iqbal, F. Martínez-Álvarez. Earthquake prediction model using support vector regressor and hybrid neural networks. PloS One. 2018 Jul 5, 13(7):e0199004. doi:10.1371/journal.pone.0199004. PMID: 29975687; PMCID: PMC6033417

[12] T. Uchide. Focal mechanisms of small earthquakes beneath the Japanese islands based on first-motion polarities picked using deep learning, Geophysical Journal International, Volume 223, Issue 3, December 2020, pp. 1658-1671, https://doi.org/10.1093/gji/ggaa40

[13] K. M. Asim, A. Idris, T. Iqbal, F. Martínez-Álvarez. Seismic indicators based earthquake predictor system using Genetic Programming and AdaBoost classification. Soil Dynamics and Earthquake Engineering, Volume 111, 2018, pp.1-7, ISSN 0267-7261, https://doi.org/10.1016/j.soildyn.2018.04.020

[14] G. Liu, R. Li, J. He, W.-le Li, J. Lu, W. Long, P. Gao, G. Cai, M. Tang. Scaling relation of earthquake seismic data. Physica A: Statistical Mechanics and its Applications V. 492, 15 February 2018, pp. 2092-2102, DOI: 10.1016/j.physa.2017.11.126

[15] N. J. Lindsey, E. R. Martin, D. S. Dreger, B. Freifeld, S. Cole, S. R. James, B. L. Biondi, J. B. Ajo-Franklin. Fiber-Optic Network Observations of Earthquake Wavefields. Geophysical Research Letters. V. 44, Issue 23, 16 December 2017, pp. 11792-11799, https://doi.org/10.1002/2017GL075722

[16] C. Langenbruch, M. Weingarten & M. D. Zoback. Physics-based forecasting of man-made earthquake hazards in Oklahoma and Kansas. Nat Commun 9, 3946 (2018) https://doi.org/10.1038/s41467-018-06167-4

[17] http://some.kz/index.php?p=card&outlang=1&oid=9

[18] N. A. Karri, M. Y. Ansari and A. Pathak. Identification of Seismic Zones of India using DBSCAN. IEEE 2018 International Conference on Computing, Power and Communication Technologies (GUCON), Greater Noida, Uttar Pradesh, India, 2018, pp. 65-69

[19] C. E. Yoon, K. J. Bergen, K. Rong, H. Elezabi, W. L. Ellsworth, G. C. Beroza, P. Bailis, Ph. Levis. Unsupervised Large-Scale Search for Similar Earthquake Signals. Bulletin of the Seismological Society of America, 2019, 109(4), pp. 1451-1468, https://doi.org/10.1785/0120190006

[20] E. Schubert, J. Sander, M. Ester, H.-P. Kriegel, X. Xu. DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN. ACM Transactions on Database Systems, July 2017, 42(3), pp. 1-21, https://doi.org/10.1145/3068335

[21] Cesca, S. Seiscloud, a tool for density-based seismicity clustering and visualization. J Seismol 24, 2020, pp. 443-457, https://doi.org/10.1007/s10950-020-09921-8

[22] X. Shang, X. Li, A. Morales-Esteban, L. Dong, K. Peng. K-Means Cluster for Seismicity Partitioning and Geological Structure Interpretation, with Application to the Yongshaba Mine (China). Shock and Vibration, Pub Date: 2017-01-01, DOI: 10.1155/2017/5913041

[23] Y. Yu, Z. Jia, L. Cao, et al. Fast density-based clustering algorithm for location big data.Journal of Software, 2018, 29(8), pp. 2470-2484

[24] Z. Fan, X. Xu. Application and visualization of typical clustering algorithms in seismic data analysis. Procedia Computer Science, 2019, V. 151, pp. 171-178, https://doi.org/10.1016/j.procs.2019.04.026

[25] Md. M. A. Patwary, D. Palsetia, A. Agrawal, W.-k. Liao, F. Manne, A.Choudhary. A new scalable parallel DBSCAN algorithm using the disjoint-set data structure. SC '12: Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis, November 2012, pp. 1-11, DOI:10.1109/SC.2012.9

[26] M. Kazemi-Beydokhti, R. A. Abbaspour & M.Mojarab. Spatio-Temporal Modeling of Seismic Provinces of Iran Using DBSCAN Algorithm. Pure Appl. Geophys., 174, 1937-1952 (2017) https://doi.org/10.1007/s00024-017-1507-0

[27] http://neotec.ginras.ru/neomaps/M050_Kazakhstan_2003_Seism-region_Karta-obshchego-seysmicheskogo-rayonirovaniya-respubliki-kazahstan.html

[28] P. Yariyan, H. Zabihi, I. D. Wolf, M. Karami, S. Amiriyan. Earthquake risk assessment using an integrated Fuzzy Analytic Hierarchy Process with Artificial Neural Networks based on GIS: A case study of Sanandaj in Iran. International Journal of Disaster Risk Reduction, 2020, 50, 101705, https://doi.org/10.1016/j.ijdrr.2020.101705

[29] D. Balla, M. Zichar, A. Kozics, T. Mester, T. Mikita, J. Incze1, T. J. Novák. A GIS Tool to Express Soil Naturalness Grades and Geovisualization of Results on Tokaj Nagy-Hill. Acta Polytechnica Hungarica. 2019, 16(6), pp. 191-205

[30] M. J. Somodevilla, A. B. Priego, E. Castillo, I. H. Pineda, D. Vilariño, A. Nava. Decision support system for seismic risks. JCS&T, August 2012, Vol. 12, No. 2, pp. 71-77

[31] A. Tlebaldinova, N. Denissova, O. Baklanova, I. Krak, G. Györök. Normalization of vehicle license plate images based on analyzing of its specific features for improving the quality recognition. Acta Polytechnica Hungarica, 2020, 17(6), pp. 193-206

[32] Kozlov, A. N. Intellektual'nyye informatsionnyye sistemy: uchebnik /A. N. Kozlov; Min-vo s-kh. RF, FGBOU VPO Permskaya GSKHA. - Perm': Izd-vo FGBOU VPO Permskaya GSKHA, 2013.– 278 s.

[33] S. Mangalathu, H. Sun, C. C. Nweke, Zh. Yi, H. V. Burton. Classifying earthquakedamage to buildings usingmachine learning. Earthquake Spectra. September 2019, 36(1), DOI:10.1177/8755293019878137

[34] S.-H. Hwang, S. Mangalathu, J. Shin, J.-S. Jeon. Machine learning-based approaches for seismic demand and collapse of ductile reinforced concrete building frames. Journal of Building Engineering, Pub Date: 2020-10-22, DOI: 10.1016/j.jobe.2020.101905

[35] M. Karmenova, A. Nugumanova, A. Tlebaldinova, A. Beldeubaev, G. Popova, A. Sedchenko. Seismic assessment of urban buildings using data mining methods. Proceedings of the 6th International Conference on Computer and Technology Applications, April 14-16, 2020, DOI:10.1145/3397125.3397152

[36] Wójcik W., Karmenova M., Smailova S., Tlebaldinova A., Belbeubaev A. Development of data-mining technique for seismic vulnerability assessment. //International Journal of Electronics and Telecommunications, 2021, Vol. 67, No. 2, pp. 261-266, DOI: 10.24425/ijet.2021.135974