

IMPROVING THE SOM ALGORITHM TO ENSURE STABILITY AND REPRODUCIBILITY OF DATA CLUSTERING RESULTS

O.V. IVASHCHENKO, S.S. FEDIN

Abstract. The article proposes a method to improve the Kohonen Self-Organizing Map (SOM) learning algorithm to ensure the stability and reproducibility of clustering results, an urgent task when working with large amounts of data. SOM is widely used in clustering and visualization tasks, especially in applications that require analyzing multidimensional data structures, such as telecommunications billing systems and financial analysis. The standard SOM implementation, which includes random weight initialization and stochastic sample selection during training, leads to significant cluster variability even when using the same input data and identical network training parameters. This makes it difficult to apply this algorithm in cases where stability and reproducibility of results are required. To solve this problem, we propose modifying the algorithm to include its own random number generator and introducing a seed parameter to fix the initial training conditions. This reduces variability and ensures reproducible clustering results, thereby increasing the reliability of the analysis and the suitability of the SOM algorithm for real business tasks. The proposed method has been tested on data from billing systems, where the reproducibility of clustering results is critical for effective work with customer segments, the development of targeted marketing strategies, and the creation of personalized tariff plans.

Keywords: Kohonen self-organizing maps (SOM), data clustering, seed parameter, reproducibility of results, random number generator.

INTRODUCTION

In today's environment, telecommunications companies process large amounts of data on a daily basis that contain valuable information about subscriber behavior and service usage. This data plays an important role in making strategic decisions, developing personalized offers, and increasing the competitiveness of companies. One of the key tasks is to apply clustering to organize and analyze the data, allowing to draw useful analytical conclusions for further use.

Among modern approaches to clustering and visualization of multidimensional data, Kohonen's Self-Organizing Maps (SOM) [1] occupy a special place, preserving the topological structure of the data and allowing an intuitive understanding of its grouping through visualization. At the same time, algorithmic features such as random initialization of weights and stochastic samples selection during training lead to the fact that the results can differ significantly for the same input data. This creates difficulties in cases where stability and reproducibility of

clustering results are required, which is important for making informed decisions in business and analytics.

This opens up the possibility of improving the method to increase the reproducibility and accuracy of clustering, which is important for scientists and analysts using this tool.

LITERATURE ANALYSIS AND PROBLEM STATEMENT

Kohonen's Self-Organizing Map is a popular method for data analysis and clustering that allows to identify similar groups of objects in a data set and simplify their structure for further use. However, despite the widespread use of SOM in industries such as telecommunications, finance, and engineering, the problem of stability and reproducibility of clustering results remains unresolved. This limits the practical application of the algorithm in tasks requiring accurate and stable data grouping.

Researchers have already proposed some improvements to the SOM algorithm to increase the accuracy and reliability of clustering.

For example, Panu Somervuo and Teuvo Kohonen [2] discuss clustering large protein sequence databases using an extended SOM that allows the creation of clusters of protein sequences without converting the data into histogram vectors. In his study, Mark Van Hulle [3] analyzes the basic SOM algorithm, its properties, and the possibility of extending it to work with categorical data, time series, and tree structures. Jens Claussen [4] proposes the Winner-Relaxing Self-Organizing Maps (WRSOM) approach, which ensures the stability of the cluster location. Despite the success of these approaches, the problem of variability of results is still relevant.

The study by Melody Kiang, Michael Hu, Dorothy Fisher [5] considers the use of an extended version of Kohonen's self-organizing maps to segment the market of telecommunications companies based on behavioral and demographic factors, including the frequency of long-distance calls, household structure, and so on. By using the enhanced SOM algorithm, they were able to achieve better results than standard clustering methods such as factor analysis and k -means. However, the authors note that the stability of the resulting clusters could be improved, which opens up prospects for further improving the efficiency of this method. Similarly, in a study by Wei Wang, Shiwei Xu, Hong Ouyang, and Xinyu Zeng [6], where SOM was used to optimize the parameters of the power systems of unmanned electric drive chassis, improvements were proposed by combining SOM with an advanced genetic algorithm that uses isolated niches to improve the accuracy of the results. This increased the convergence rate and improved the global search capability of the algorithm, providing a more accurate clustering of the initial populations. This approach demonstrates the potential of combining SOM with other algorithms to solve complex multitask optimization problems, which emphasizes the importance of further research to improve the stability and accuracy of SOM results for various engineering and practical applications.

The study by V. Dyachenko, O. Lyashenko, B. Ibrahim, O. Michal, and Y. Koltun [7] proposed a modification of Kohonen's self-organizing maps with a parallel learning algorithm that can significantly increase the speed of data

processing in multi-core processor systems. This approach demonstrates effectiveness in the tasks of clustering large amounts of data, ensuring the adaptation of the algorithm to a dynamic environment. The work of Rodrigo Cavalcanti, Bruno Pimentel, Carlos Almeida, and Renata Souza [8] presents a new variant of the Fuzzy Kohonen Clustering Network (FKCN), which uses the fuzzy c -means membership function instead of a fixed learning coefficient. This approach takes into account the intraclass and interclass variance, which allows to obtain better clustering results on real and synthetic data sets.

The works of N.I. Furmanova, O.Y. Farafonov, O.Y. Malyi, Y.O. Sitsilitsyn, V.O. Dyachenko, O.P. Mikhal, E.A. Egorova, V.G. Ivanov, E.S. Sakalo [9–11] consider various approaches to improving SOM, which demonstrate the wide application of this method in optimization, clustering, and data analysis. In particular, N.I. Furmanova, O.Y. Farafonov, and Y.O. Sitsilitsyn study the integration of SOM with genetic algorithms to reduce computational costs and avoid local minima in multidimensional optimization problems. V.O. Dyachenko and O.P. Mikhal proposed an improvement of SOM for work in distributed energy-critical sensor networks by parallel selection of several winning neurons, which reduces power consumption and optimizes computation time. E.A. Egorova, V.G. Ivanov, and E.S. Sakalo use the Kalman-Mayne filter to adapt SOM, providing accurate clustering even in the presence of noise in the data.

Despite the progress made, all authors note the importance of further research aimed at selecting optimal SOM parameters, such as the neighborhood function, initialization of the weights, and adaptation of the algorithm to dynamic data.

Given the need to increase the stability of SOM clustering results, the purpose of this study is to develop an improved approach to its application that ensures reproducibility of results.

The variability of SOM results is caused by several factors. First of all, the standard implementation of the algorithm randomly initializes the neural weights, resulting in different initial conditions even for the same input data. In addition, the learning process involves stochastic samples selection, which also introduces randomness at each stage of clustering. As a result, even if the algorithm is run repeatedly with the same data and parameters, SOM may generate different clusters.

The non-reproducibility of clustering complicates the practical application of SOM, especially in cases where stability of results is required. For example, in billing systems of telecommunication companies, the analysis of customer behavior data requires accurate and stable segmentation to generate personalized tariffs, etc. The variability of clustering results makes it difficult to accurately identify customer groups, which can lead to errors in understanding their needs and creating appropriate marketing strategies.

Thus, in order to eliminate the variability of SOM clustering results, it is necessary to develop a method that ensures reproducibility of the output data. In this study, we propose an approach that involves the introduction of a proprietary random number generator and a seed parameter for fixed initialization of network weights. This solution will eliminate the stochastic influence of the algorithm on the clustering results and increase the reliability of data analysis in telecommunication systems.

PURPOSE AND OBJECTIVES OF THE STUDY

The purpose of the study is to develop an improved algorithm for Kohonen's self-organizing maps that ensures stability and reproducibility of clustering results. This improvement is aimed at eliminating the variability of results arising from the random initialization of weight coefficients and stochastic selection of samples during training.

The following tasks were set to achieve this goal:

- identify the key factors that cause variability in SOM clustering results and assess their impact on the stability of the algorithm;
- to develop a method to improve the SOM algorithm by introducing its own random number generator and seed parameter to fix the initial training conditions;
- to test the effectiveness of the proposed method by analyzing data from telecommunications companies' billing systems and assess its impact on the stability and reproducibility of clustering results.

MATERIALS AND METHODS OF RESEARCH

The Kohonen algorithm

Self-organizing maps are based on Kohonen neural networks and are designed to visualize multidimensional objects on a two-dimensional map, where the distances between objects correspond to the distances between their vectors in a multidimensional space, and the feature values themselves are displayed in different colors and shades [12].

The basic idea behind SOM is to create a two-dimensional mapping structure in which neighboring nodes on the map reflect the similarity between data. Each node on the map has weights that represent vectors in the feature space. During SOM training, these weights are changed to match the structure and distribution of the data [13; 14].

The network construction is based on competitive learning, where the output nodes (neurons) compete with each other for "victory". In the course of the competition, during the training process, neurons are selectively tuned for different input examples [15].

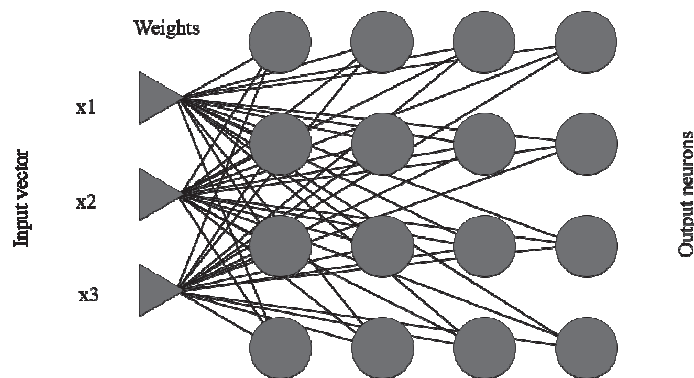


Fig. 1. Kohonen's network model

Input neurons form the input layer of the network, which contains one neuron for each input field. As in a regular network, input neurons do not participate in the training process. Their task is to transfer the values of the input fields of the initial sample to the neurons of the output layer. Each connection between neurons has a certain weight, which is randomly set in the interval [0;1] during initialization. The learning process consists in adjusting the weights. Unlike most neural networks, the Kohonen network has no hidden layers: the data from the input layer is sent directly to the output layer, whose neurons are arranged in a one- or two-dimensional grid of rectangular or hexagonal shape [16].

During SOM training, the following main stages are performed:

1. Competition: each output neuron calculates the distance between its weight vector and the input vector. The neuron with the smallest distance is declared the winner.
2. Cooperation: the winning neuron determines a group of neighboring neurons that also participate in the weight adjustment. This ensures the similarity of the weight vectors between neighboring neurons.
3. Adaptation: the weights of the winning neuron and its neighbors are adjusted to get closer to the input vector, promoting network self-organization and clustering.

The learning process of the Kohonen network involves a gradual decrease in the learning rate, which depends on the number of iterations. The training is divided into two phases: coarse tuning (with a larger influence radius and faster learning speed) and fine tuning (with a smaller radius and slower adaptation).

At the initial stage, if there is no a priori information about the distribution of data in the sample, the neuronal weights are initialized with random values. At the same time, the initial values of the learning rate and the learning radius R are set, which determines the number of neurons that are considered neighbors of the winning neuron and change their weights along with it. At the beginning of training, the radius R has a maximum value and gradually decreases with each iteration, which allows the network to accurately adapt to the data structure.

The Kohonen network training algorithm is based on the principles of unsupervised learning, i.e., without a teacher, and includes seven stages [15–18]:

1. Setting up the network structure (the number of neurons in the Kohonen layer).
2. Initialize the weight coefficients with random values according to the formula:

$$w_{ij} = (\text{random}(0;1) * (\max(x_{ni}) - \min(x_{ni}))) + \min(x_{ni}),$$

where x_{ni} is an input vector; w_{ij} is a vector of weight coefficients.

3. Competition. Supplying a random training example of the current training iteration to the network inputs and calculating the Euclidean distances from the input vector to the centers of all clusters:

$$D(W_j, X_n) = \sqrt{\sum_i (w_{ij} - x_{ni})^2},$$

where x_{ni} is an input vector; w_{ij} is a vector of weight coefficients.

The output neuron whose weight vector has the smallest distance to the object feature vector is declared the winner.

4. Merge. All neurons located within the training radius relative to the winning neuron are identified.

5. Adjustment. According to the smallest of the values of R_j , the winning neuron j is selected, which is closest to the input vector in terms of values. For the selected neuron (and only for it), the weight coefficients are corrected:

$$w_{ij}^{new} = w_{ij}^{current} + l * (x_{ni} - w_{ij}^{current}),$$

where x_{ni} is an input vector; w_{ij}^{new} is a new vector of weight coefficients; $w_{ij}^{current}$ is the current vector of weight coefficients; l is the learning rate coefficient.

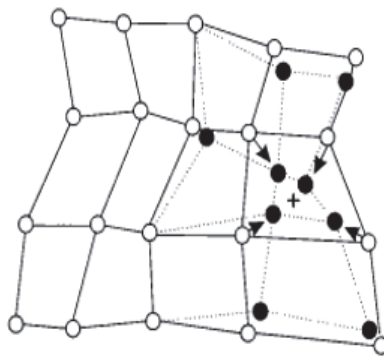


Fig. 2. Adjusting the weights of neurons

6. Correction. Changes the learning rate parameter according to the specified law.

$$l^{new} = l * \exp(i^{current} - i),$$

where l^{new} is the adjusted learning rate parameter; l is the initial learning rate parameter; $i^{current}$ is the current iteration; i is the total number of iterations.

7. The cycle is repeated from stage 3 (competition) until the end condition is met: stabilization of the neural network outputs or the specified number of iterations.

Technical implementation of the SOM algorithm improvement

In order to ensure reproducibility of clustering results, the standard implementation of the Kohonen Self-Organizing Maps algorithm was improved in this study. The main goal of the improvements was to eliminate the variability of results caused by random initialization of weights and stochastic samples selection during training. The proposed solution includes the implementation of a proprietary random number generator with the ability to fix initial conditions using the seed parameter.

At the stage of improving the SOM algorithm, a special random number generator was implemented that uses the mathematical function of sine to generate values. The code of this generator is implemented as a RandomGenerator class in C# (Fig. 3). The main feature of this generator is the

ability to fix the initial state using the seed parameter, which reduces the variability of the initial conditions and, as a result, stabilizes the clustering results.

Implementation features:

- The seed: parameter is set when creating an instance of the RandomGenerator class, which determines the initial state of the generator. This ensures the determinism of the sequence of pseudo-random numbers.
- Generation algorithm: a pseudo-random number is calculated as the residual of multiplying a sine by a scaling factor of 10000. This allows to get a uniform distribution of values within $[0; 1)$.
- Incrementing the seed: after each call to the Next() function, the seed value is incremented, ensuring a consistent change in the output values.

```

1  using System;
2
3  namespace SOM_Kohonen_WpfApp.Service
4  {
5      3 references | Oleksandr, 75 days ago | 1 author, 1 change
6      public class RandomGenerator
7      {
8          private int _seed;
9
10         1 reference | Oleksandr, 75 days ago | 1 author, 1 change
11         public RandomGenerator(int seed) => _seed = seed;
12
13         1 reference | Oleksandr, 75 days ago | 1 author, 1 change
14         public double Next()
15         {
16             var x = Math.Sin(_seed++) * 10000;
17             return x - Math.Floor(x);
18         }
19     }
20 }

```

Fig. 3. Implementation of a random number generator using the seed parameter

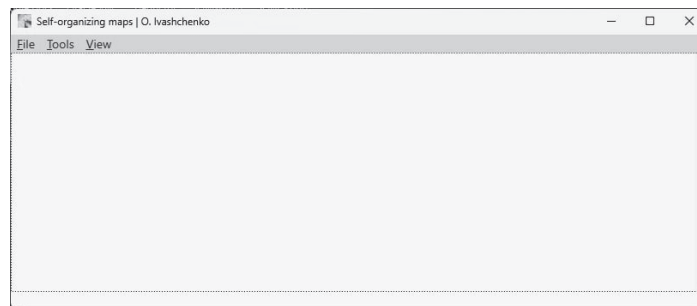


Fig. 4. The main window of the application

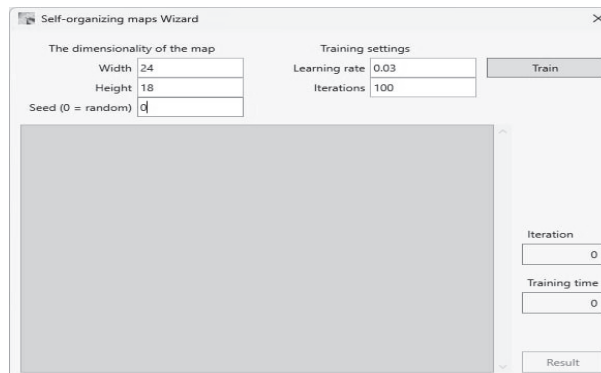


Fig. 5. Neural network training window, entering the seed parameter

The generated generator is used to initialize the weights of the SOM neurons. Each time the algorithm is run with the same value of the seed parameter, the neuronal weights receive the same initial values, making it impossible to vary clusters due to random initialization. In addition, the seed parameter controls the order in which training samples are selected during network training, which eliminates the stochastic influence on the clustering process.

To ensure the convenience of working with the advanced algorithm, an interface based on Windows Presentation Foundation (WPF) was created. The interface is designed to provide the user with easy access to all key functions. In particular, the user can upload input data, configure training parameters such as map dimension, number of iterations and seed value, which allows to control the initial training conditions. In addition, after training is complete, the user can view the clustering results in the form of a Kohonen map, which provides visualization of the results and allows detailed analysis of the cluster distribution.

RESEARCH RESULTS

Using the created application based on the improved Kohonen algorithm, the customer base of the telecommunications company was clustered with different values of the seed parameter, the results of which are shown in Figs. 6 and 7. The data contained information about the demographic characteristics of customers, their activity and the intensity of service use, which made it possible to create clear customer segments by behavioral characteristics.

Clustering with different seed values

Fig. 6 shows the clustering results for different values of the seed parameter (85690 and 368). It can be seen that changing this parameter leads to a change in the shape and location of the clusters. For example, the clusters highlighted in the figure change their boundaries significantly: the same customer segment can move around the map and change its shape and size. This makes it difficult to identify stable customer groups and can lead to difficulties in analyzing them accurately.

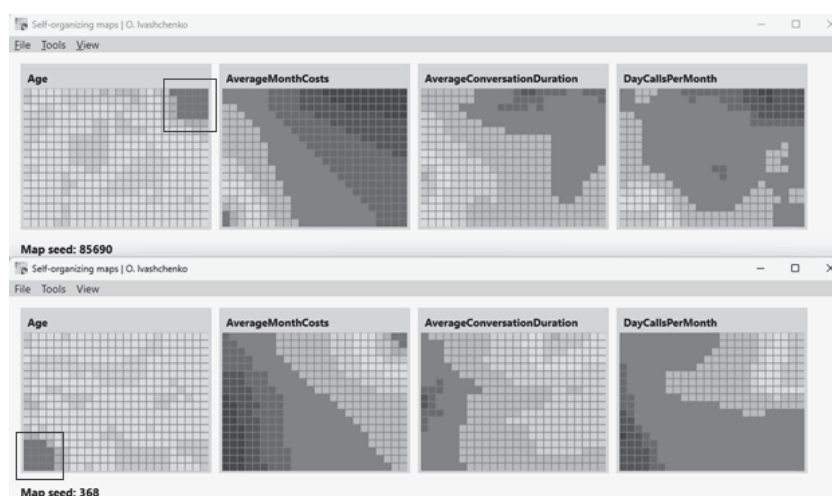


Fig. 6. Generated maps for different values of the seed parameter

Clustering with the same seed value

Fig. 7 shows the results of clustering using the same value of the seed parameter (6548). In this case, all resulting Kohonen maps are identical, regardless of the number of algorithm runs. The clusters highlighted in the figure have the same shape and location, confirming the stability of the algorithm. This shows that using the same value of the seed parameter guarantees not only the stability of individual clusters, but also the complete reproducibility of the entire Kohonen map.

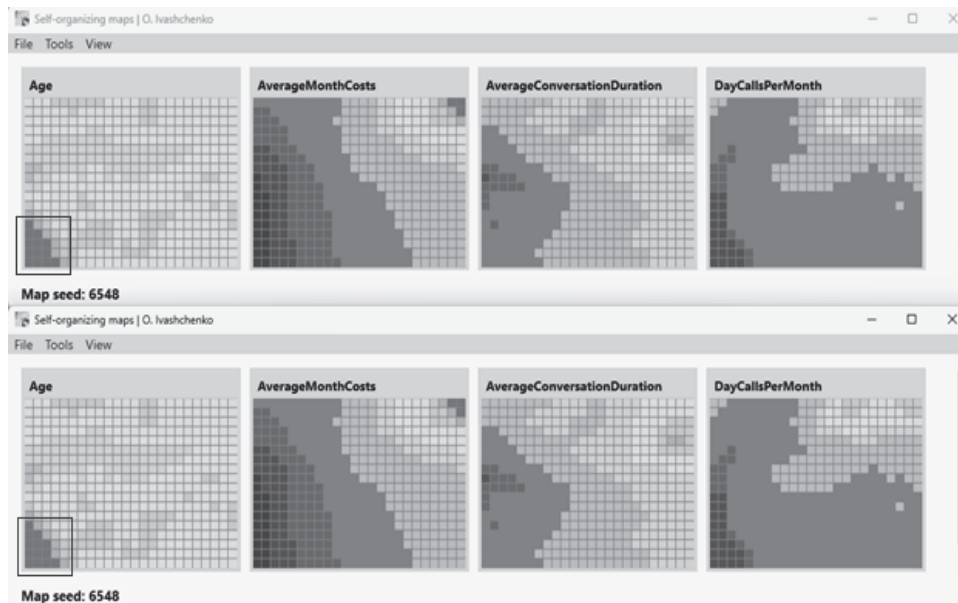


Fig. 7. Generated maps for the same values of the seed parameter

Therefore, the use of a fixed seed parameter allows to achieve full reproducibility of the results, which is impossible in the case of random initialization of the weights.

The seed values (85690, 368, and 6548) used in this study were chosen randomly. However, each identical set of seed values guarantees identical clustering results, ensuring that the shape and location of the clusters remain consistent. This improvement eliminates the variability in results that occurred due to the random initialization of weights and samples selection during the algorithm. This emphasizes the reproducibility of results, which is an important aspect for scientific research and practical use of the SOM algorithm.

DISCUSSION OF THE OBTAINED RESULTS

The results of the study confirmed that the improved Kohonen Self-Organizing Map algorithm with the implemented seed parameter for fixed initialization of network weights ensures stability and reproducibility of clustering results. Using the same seed value eliminated the problem of variability caused by random initialization of weights and stochastic samples selection during network training. This is especially important in tasks where accurate group identification is critical for decision making.

A comparative analysis of the clustering results showed that with different values of the seed parameter (Fig. 6), there was significant variation in the size, shape, and location of the clusters. The same customer segment could change its boundaries or move around the map, making it difficult to interpret and analyze the data consistently. Instead, clustering with a fixed seed value (Fig. 7) ensured complete identity the Kohonen maps on each run of the algorithm, confirming the stability and reliability of the results.

The practical significance of the results obtained is particularly relevant for telecommunications companies, where the stability of customer segmentation plays a key role in the development of personalized tariff plans and marketing strategies. The improved algorithm provides reliable analysis of customer behavior data, which enables more accurate marketing budget calculations, minimizes customer churn, and increases the efficiency of customer base management.

Compared to other approaches to improve SOM, such as combining it with genetic algorithms or using modifications of WRSOM, the proposed method is simple to implement and does not require additional computational resources. This makes it an effective solution for tasks that require stable results at minimal technical costs.

Some limitations of the proposed approach should also be noted. The value of the seed parameter needs to be adapted for different data sets, which may require additional testing to achieve optimal results. In future research, it is advisable to consider automating the selection of the seed parameter or adapting it to work with dynamic data. This will increase the flexibility and versatility of the algorithm for a wider range of tasks.

Thus, the results of the study showed that the introduction of the seed parameter allows to achieve stable and reproducible clustering, which is important for scientific research and practical use in business intelligence, especially for telecommunication systems.

CONCLUSIONS

1. The key factors of variability in the results of SOM clustering are identified. The main reasons for the instability of the results are the random initialization of neuronal weights and the stochastic selection of training samples during network training. These factors lead to different locations and shapes of clusters for the same input data, making stable analysis impossible.

2. A method to improve the SOM algorithm is developed. It is proposed to introduce its own random number generator with the ability to fix the initial conditions using the seed parameter. This allows to set the same initial neuronal weights and a deterministic sequence of training samples selection, which eliminates variability and ensures the stability of the clustering results.

3. The effectiveness of the proposed approach is tested. The results of clustering the customer base of a telecommunications company have shown that using the same value of the seed parameter ensures full reproducibility of Kohonen maps. This allows for stable identification of customer groups, simplifying data analysis for the development of targeted marketing strategies and personalized tariff plans.

REFERENCES

1. “Teuvo Kohonen, Timo Honkela, Kohonen Network,” *Scholarpedia*. 2007. Accessed on: 12 June 2024. Available: http://www.scholarpedia.org/article/Self-organizing_feature_map
2. Panu Somervuo, Teuvo Kohonen, “Clustering and Visualization of Large Protein Sequence Databases by Means of an Extension of the Self-Organizing Map,” *Lecture Notes in Computer Science*, vol. 1967, 2000. doi: https://doi.org/10.1007/3-540-44418-1_7
3. Marc M. Van Hulle, “Self-Organizing Maps,” *Handbook of Natural Computing*. Springer, Berlin, Heidelberg, pp. 585–622, 2012.
4. Jens Christian Claussen, “Winner-Relaxing Self-Organizing Maps,” *Neural Computation*, 17(5), pp. 996–1009, 2005. doi: <https://doi.org/10.1162/0899766053491922>
5. Melody Y. Kiang, Michael Y. Hu, Dorothy M. Fisher, “An extended self-organizing map network for market segmentation — a telecommunication example,” *Decision Support Systems*, vol. 42, issue 1, October 2006, pp. 36–47. doi: <https://doi.org/10.1016/j.dss.2004.09.012>
6. W. Wang, S. Xu, H. Ouyang, X. Zeng, “Parameter Optimization of the Power and Energy System of Unmanned Electric Drive Chassis Based on Improved Genetic Algorithms of the KOHONEN Network,” *World Electric Vehicle Journal*, 14(9), 260, 2023. doi: <https://doi.org/10.3390/wevj14090260>
7. V. Diachenko, O. Liashenko, B.F. Ibrahim, O. Mikhal, Yu. Koltun, “Kohonen network with parallel training: Operation structure and algorithm,” *Int. J. Adv. Trends Comp. Sci. Eng.*, vol. 8, no. 1.2, pp. 35–38, 2019. doi: <https://doi.org/10.30534/ijatcse/2019/0681.22019>
8. Rodrigo B. de C. Cavalcanti, Bruno Pimentel, Carlos W.D. de Almeida, Renata M.C.R. de Souza, “A Multivariate Fuzzy Kohonen Clustering Network,” *IEEE Transactions on Neural Networks*, vol. 31, no. 4. pp. 75–82, 2020. doi: <https://doi.org/10.1109/IJCNN.2019.8852243>
9. N.I. Furmanova, O.Y. Farafonov, O.Y. Malyi, Y.O. Sitsilitsyn, “Improvement of the method of searching for solutions to solve the optimization problem using a genetic algorithm by preliminary clustering,” (in Ukrainian), *Instrumentation Technology*, no. 2, pp. 6–9, 2017. Available: <https://elar.tsatu.edu.ua/server/api/core/bitstreams/f045c4ca-7d17-4c9c-a1cb-7b2df9aafe7e/content>
10. V.O. Dyachenko, O.F. Mikhal, “Prospects for the use of of the classical Kohonen algorithm in distributed energy-critical sensor networks,” (in Ukrainian), *Control, Navigation and Communication Systems*, no. 4, pp. 75–79, 2023. doi: <https://doi.org/10.26906/SUNZ.2023.4.075>
11. E.A. Egorova, V.G. Ivanov, E.S. Sakalo, “Optimization of process of the Kohonen self-organizing map based on the Kalman-Mayne filter,” (in Russian), *Control, Navigation and Communication Systems*, issue 4(8), pp. 52–55, 2008. Available: https://dspace.nlu.edu.ua/bitstream/123456789/6713/1/Ivanov_52-55.pdf
12. Achraf Khazri, “Self-Organizing Maps (Kohonen’s maps),” *Medium*, [website]. 2019. Available: <https://medium.com/data-science/self-organizing-maps-1b7d2a84e065>
13. T. Kohonen, “The self-organizing map,” *Proceedings of the IEEE*, 78(9), pp. 1464–1480, 1990.
14. E.O. Kaminsky, *Forecasting system for supporting activities using deep learning. Bachelor’s thesis*. NTUU “KPI”, Kyiv, 2023, p. 78.
15. *Kohonen Networks*, [website]. Accessed on: 13.11.2022. Available: <https://ppt-online.org/46514>
16. T. Kohonen, *Self-Organizing Maps*. Springer-Verlag Berlin Heidelberg, 2001.

17. S. Kaski, "Self-Organizing Maps," in C. Sammut, G.I. Webb (Eds.) *Encyclopedia of Machine Learning*. Springer, Boston, 2011, pp. 886–888.
18. "Self-Organizing Maps, SOM," *Studfile.net*. Accessed on: 15 Nov. 2022. Available: <https://studfile.net/preview/3021431/>

Received 25.12.2024

INFORMATION ON THE ARTICLE

Oleksandr V. Ivashchenko, ORCID: 0009-0007-5470-9137, National Transport University, Ukraine, e-mail: alexander.ivashchenkoo@gmail.com

Serhii S. Fedin, ORCID: 0000-0001-9732-632X, National Transport University, Ukraine, e-mail: sergey.fedin1975@gmail.com

УДОСКОНАЛЕННЯ АЛГОРИТМУ SOM ДЛЯ ЗАБЕЗПЕЧЕННЯ СТАБІЛЬНОСТІ ТА ВІДТВОРЮВАНІСТІ РЕЗУЛЬТАТІВ КЛАСТЕРИЗАЦІЇ ДАНИХ / О.В. Іващенко, С.С. Федін

Анотація. Запропоновано метод удосконалення алгоритму навчання самоорганізаційних карт Кохонена (Self-Organizing Maps, SOM) для забезпечення стабільності та відтворюваності результатів кластеризації, що є актуальним завданням у ході роботи з великими обсягами даних. SOM широко застосовується у задачах кластеризації та візуалізації, особливо у сферах, де необхідно аналізувати багатовимірні структури даних, зокрема у білінгових системах телекомунікаційних компаній, фінансовому аналізі тощо. Стандартна реалізація SOM, яка включає випадкову ініціалізацію ваг і стохастичний вибір зразків під час навчання, призводить до значної варіативності кластерів навіть за умови використання однакових вхідних даних та ідентичних параметрів налаштування тренування мережі. Це ускладнює застосування цього алгоритму у випадках, коли потрібна стабільність та відтворюваність результатів. Для вирішення цієї задачі запропоновано модифікацію алгоритму, що включає власний генератор випадкових чисел і введення параметра seed для фіксації початкових умов навчання. Це дає змогу знизити варіативність і забезпечити відтворюваність результатів кластеризації для підвищення достовірності аналізу та придатності алгоритму SOM за використання в реальних бізнес-завданнях. Запропонований метод протестовано на даних білінгових систем, де відтворюваність результатів кластеризації має критичне значення для ефективної роботи з клієнтськими сегментами, розроблення таргетованих маркетингових стратегій, персоналізованих тарифних планів тощо.

Ключові слова: самоорганізаційні карти Кохонена (SOM), кластеризація даних, параметр seed, відтворюваність результатів, генератор випадкових чисел.