System analysis of monitoring ecological information using Fuzzy ART neural network

I L Kashirina, K A Fedutinov, T V Azarnova and Iu V Bondarenko

Abstract. The article considers the peculiarities of clustering of monitoring ecological information using a generalized model of ART family network. It formulates the approach to obtaining a comprehensive assessment of a region's ecological situation (by the example of Voronezh oblast) using Fuzzy ART network. The suggested method helps to identify the structural links between the monitoring indicators of the ecological system being investigated, which allows creating a basis for a logical and consistent approach to the decision-making problem. The clustering results and the analysis of the obtained cluster prototypes make it possible to select a grounded mathematical model of a comprehensive assessment of the region's ecological situation. Moreover, the constructed ecological situation analysis model and the developed software allow experts to keep track of trends in the region's ecological situation.

1. Introduction

The first paragraph after a heading is not indented. Today's data collection and storage possibilities resulted in the fact that within the ecological monitoring systems are formed large volumes of environmental information that can be used to search for correlations and patterns in information arrays. The need of introducing into the ecological monitoring software suite of subprograms that perform a system analysis of the ecological situation information results from the multidimensionality and multiconnectivity of environmental data. Many environmental processes are characterized by nonlinearity and uncertainty, which complicates the operator's assessment and prediction of the ecological situation. The region's ecological situation assessment methods that do not use the integral assessment methods are ineffective for the practical purposes of developing strategies for responding to the region's environmental situation, since they focus only on individual factors to the detriment of possible interaction of a complex of such factors [1]. The traditional method of consolidation (aggregation) of the ecological situation assessment results is the averaging of the monitoring indicator values. However, the performed studies indicate that such an approach not only fails to contribute to ranking of territories by the degree of the environmental problem intensity, but rather helps to level out the differences between the individual zones [2]. In order to improve the efficiency and accuracy of correct managerial decision-making, the intelligent data analysis technologies have increasingly been used recently, capable of working in the conditions of fuzzy initial information. In particular, in works [3] and [4], a scientific and methodological apparatus for a qualimetric approach to construction...
of a non-additive integral assessment of the ecological hazard of territories is proposed, this assessment being a weighted average "quasi-geometric" value. This integral assessment has a probabilistic sense, which allows competently measuring and meaningfully interpreting the overall eco-logical hazard of the territories on Harrington universal verbal-numerical scale based on the degree of non-compliance of the territories' ecological situation with the normative requirements to their quality. However, for the system analysis of problematic ecological situations and the construction of an integrated assessment of the geocological state, other approaches to establishing structural links between the monitoring indicators of the system under study should be considered, which will allow creating the basis for a logical and consistent approach to the decision-making problem.

The availability of a modern information and monitoring environment provides the possibility of cluster structuring of the ecological information throughout the totality of individual indicators. Cluster data structuring allows highlighting on the region's map the areas with a similar ecological situation. The results of complexing of alternative approaches to the cluster division of ecological information make it possible to proceed at the next stage of mathematical modeling to development of a grounded model for the region's ecological situation integrated assessment. Such a model should similarly assess the sets of indicators that have fallen into the same cluster. In addition, ecological information clustering methods are focused on the following objectives:

- "understanding" the data by defining its cluster structure: partitioning a region into clusters makes it possible to improve further decision-making and develop strategies to respond to the changes in the ecological situation by applying to each cluster its own analysis method;
- discovery of novelty - in this case, atypical territories are determined that do not fall in-to any of the clusters and require separate analysis.

Among the artificial intelligence technology application areas when developing software systems designed for ecological monitoring is the use of neural networks in the ecological situation analysis and assessment tasks. The neural networks can improve the efficiency and accuracy of the ecological assessment and, consequently, the effectiveness of the decisions taken. Presently, when solving clustering problems, the neural networks of ART (Adaptive Resonance Theory) family have shown themselves in the best way [5]. This is why, in order to solve the region's cluster structuring problem on the basis of the set of monitoring indicators, it is proposed to use Fuzzy ART [6] neural network in accordance with the generalized ART family network model proposed in [7].

2. Materials and methods

2.1. Fuzzy ART network
The ART family is a kind of artificial neural network based on the adaptive resonance theory by Grossberg and Carpenter [8,9]. It includes models (ART-1, ART-2, Fuzzy ART etc.) that use mainly unsupervised learning and are used for solving clustering problems. The Fuzzy ART model is capable of adjusting the size and number of clusters depending on the complexity and comprehensiveness of the incoming data set. The Fuzzy ART network allows: producing the result in real time; quickly adapting to changing data; scaling to the number of objects being received for processing; providing a data structure that does not expand too much when processing a large number of objects; detecting the presence of outliers (data abnormalities). It is proposed to use the Fuzzy ART network to perform primary clustering of initial ecological monitoring data, and to further use the obtained cluster prototypes for building an integrated assessment, as well as for environmental experts' compiling strategies for responding to the region's environmental situation. The formation of integrated assessments based on the data cluster structuring results is considered in [10,11].

The Fuzzy ART network [4] includes layer F0 representing the current input vector; layer F1 that processes both the input from layer F0 and the feedback from layer F2, which represents the recognized category (Figure 1).
The output of layer $F0$ is the vector $I = (a_1, ..., a_M)$, each component of which belongs to segment $[0,1]$. Layer $F1$ at the output generates the vector $X = (x_1, ..., x_M)$, while layer $F2$ generates the vector $Y = (y_1, ..., y_N)$. The weight matrix $W$ interconnects layers $F1$ and $F2$, and to each neuron $j$ ($j = 1, ..., N$) of layer $F2$ corresponds the weight vector $W^j = (w_{j1}, ..., w_{jM})$. It is initially assumed that $w_{j1}(0) = ... = w_{jM}(0) = 1$.

The Fuzzy ART network has three setting hyperparameters: selection parameter $\alpha > 0$, influencing the selection of a category in layer $F2$ and preventing degeneration of the cluster prototypes; learning rate parameter $\beta \in [0, 1]$ and similarity parameter $\rho \in [0, 1]$, that controls the network resonance.

Figure 2 represents a generalized flowchart of ART family network functioning [7].

![Figure 1. Architecture of Fuzzy ART network](image.png)

Figure 1. Architecture of Fuzzy ART network

![Figure 2. Generalized model of ART family network functioning](image.png)

Figure 2. Generalized model of ART family network functioning.
Let us consider the realization of the stages shown in figure 2 as applied to the Fuzzy ART network. For the input vector \( I \) and each neuron \( j \) from layer \( F_2 \), a value of the selection function \( T_j(I) \) is calculated by the formula

\[
T_j(I) = \frac{I \wedge W_j}{\alpha + |W_j|}
\]

(1)

where the statement \( \wedge \) is one of the possible options of realization of "fuzzy AND": the \( j \)-th component of the vector \( (p \wedge q) \) has the form: \( (p \wedge q)_j = \min(p_j, q_j), j = 1, J \), while the norm \( | \cdot | \) is determined as \( |p| = \sum p_i \).

The Fuzzy ART network uses a reset mechanism: temporary cluster deactivation when the similarity condition is not met. This is why the cluster selection at the stage of the clustering category selection is determined as \( s = \arg \max_k T_k, k \in Q \), \( Q \) – the set of active (not deactivated) at this stage nodes, following which the selected prototype is checked against the input vector using the match function. If the maximum is achieved at more than one value, then a category \( s \) with the smallest index is selected.

The match function is determined as follows:

\[
M_s = \frac{|I \wedge W^s|}{|I|}
\]

(2)

The cluster is selected when the condition \( M_s \geq \rho \) is met, where \( \rho \) is the set similarity parameter. If the condition is violated, the cluster is marked as inactive and the selection function is called again. If there are no active clusters left, a new one is created whose weights are equivalent to the input vector.

If the condition \( M_s \geq \rho \) is met, the cluster is considered as selected, and the outputs of layer \( F_1 \) are calculated by the formula

\[
X = I \wedge W^s,
\]

(3)

and the outputs of layer \( F_2 \) take the form: \( y_s = 1, y_j = 0 \ (j \neq s) \).

The learning function is realized as follows: the weights of the prototype that has passed the check are modified at the next step \( t \) as

\[
W^s(t + 1) = (1 - \beta)W^s(t) + \beta X
\]

(4)

Since the weights of the prototypes are not normalized after learning, the problem arises associated with the learning function. The statement "fuzzy AND" used in formula (3) results in a constant trend of decreasing of the values of the \( X \)-vectors' coordinates. This decreasing can, under certain conditions, result in degeneration of the cluster prototypes. In order to solve the vector degeneration problem, at the data preprocessing stage it is recommended to apply to the input vectors the so-called complementary encoding – a \( M \)-dimensional vector turns into a \( 2M \)-dimensional one through its complementing with \( M \)-components:

\[
I = (a, a^c) = (a_1, a_2, ..., a_M, a_1^c, a_2^c, ..., a_M^c), \text{ where } a_i^c = 1 - a_i,
\]

(5)

which allows preserving its amplitude information \( (|I| = \sum_{i=1}^M a_i + M - \sum_{i=1}^M a_i = M) \).

3. Results and their discussion
3.1. Applying the Fuzzy ART network to the region's ecological situation assessment

The initial data of the region's ecological situation assessment task is a series of monitoring indicators taken in Voronezh oblast with a 500x500-meter resolution or interpolated from larger scale maps.

The list of indicators used includes four exogenous indicators: territory zoning by the degree of hazard of development of such processes as gully erosion, landslides, suffosion-subsidence phenomena, swamping; three hydrosphere indicators: natural pollution protection of the main mining levels of groundwater, module of fresh groundwater operational resources (l/s•km2), local surface runoff value in thousands (m3/ km2); as well as a series of other indicators: soil-agrolandscape group, terrain landscape type, amount of harmful substance emissions into the atmosphere from stationary sources (thousand tons), integral criterion of climatic comfort, integral public health indicator, degree of change in the morphological characteristics of higher plants (%), man-caused load level.

The set of the processed data amounts to 206 thousand 676 values for each initial indicator. Each indicator's value is measured on an integer-value scale. The total number of indicators is 17 (including the geographic coordinates of the indicator values collection point). The data is provided by the Geological Ecology Department of Voronezh State University, where a large volume of factual materials for the analysis of natural and man-caused indicators of Voronezh oblast has been accumulated.

In order to solve the monitoring information cluster structuring problem, a software suite has been developed that implements the Fuzzy ART network, as well as its modification Fuzzy ART C, which allows limiting the maximum number of clusters being created (the C parameter sets this maximum number). The developed suite allows selecting any set from the available list of 17 indicators during the interactive work with the user, and assessing the natural environment situation on their basis.

It should be particularly mentioned that the Fuzzy ART network is very suitable for solving this problem since it ensures fast processing of large data arrays and stability of the obtained cluster structuring, yet no fine-tuning of the algorithm's hyperparameters for a specific data set is required in this case. The expert-developed strategies for adjusting the ecological environmental situation will be applied to significant areas rather than to individual points, and therefore a minor displacement of the cluster boundaries depending on the hyperparameter values is quite acceptable; it is important to define the common classes. Thus, working with the Fuzzy ART software module does not require any intervention of a neural network specialist for the hyperparameter setup.

Examples of the obtained clusters can be seen in figure 3 and 4. Figure 3 shows the example of partitioning of Voronezh oblast territory into 10 clusters by the 13 selected indicators (the territory zoning by the degree of hazard of development of such processes as erosion, landslide and swamping; natural pollution protection of the main mining levels of groundwater, module of fresh groundwater operational resources, local surface runoff value, soil-agrolandscape group, amount of harmful substance emissions into the atmosphere from stationary sources, integral criterion of climatic comfort, integral public health indicator, man-caused load level). An interesting peculiarity consists in that the clusters inherit hidden dependencies on the relief in the initial data, but some areas are distinguished that should be interpreted by experts.

Figure 4 shows the example of clustering by 4 separately selected man-caused and climatic factors (man-caused load level, integral criterion of bioclimatic comfort, integral public health indicator, degree of change in the morphological characteristics of higher plants). As one can see, there is no reliance on this map.

In clustering problems, the average intracluster distance is most frequently used as a quality metrics, which distance should be minimized, or the average intercluster distance, which is to be maximized. The silhouette coefficient metrics [12] in a certain sense unites both these approaches, this is why it is proposed to use it to assess the effectiveness of the Fuzzy ART network. The silhouette coefficient shows how the average distance between objects inside the clusters differs from the average distance to the objects of other nearby clusters. The silhouette coefficient silhouette lies in the range $[-1,1]$. 


Figure 3. The example of partitioning of Voronezh oblast territory into 10 clusters by the 13 selected indicators.

Figure 4. The example of partitioning of Voronezh oblast territory into 5 clusters by the 4 selected indicators.
The greater the coefficient value, the better the clustering performance. The values close to -1 correspond to bad (incorrect) clustering cases; the values close to zero indicate that the clusters intersect and overlap each other; and the values close to 1 correspond to densely grouped clusters.

The value of the silhouette coefficient for the clustering results shown in figure 3 is equal to \( s=0.91 \), and for clustering in figure 4 \( s=0.83 \) (which is not surprising since in figure 3 there are more clusters, and they are more densely grouped accordingly). However, both these values indicate that quite dense and separable clusters have been found in the data using the Fuzzy ART network.

For comparison, figure 5 shows the clustering results for the same 4 man-caused and climatic factors as figure 4, obtained using the standard k-means method. The clustering has been realized in Jupyter notebook environment using KMeans module of Scikit-learn library for Python language.

![Figure 5. The results of clustering using the k-means method.](image)

As one can see, the clustering results in fig. 4 and fig. 5 appeared to be quite similar, but there are still some differences. The silhouette coefficient values for the k-means method appeared to be equal to \( s=0.75 \), which indicates that the Fuzzy ART network was able to more accurately determine the boundaries of the clusters.

Let us consider the scheme of working with the proposed software suite.

1) Determining the purpose of the ecological situation analysis and selecting the required criteria.
2) Performing software-assisted clustering of the data array by the selected components. Primary visualization of the results in the projection on the terrain model.
3) Adjusting the sought-for number of clusters and repeating item 2.
4) Performing an expert assessment of the obtained cluster prototypes and developing action strategies for each of them.
5) Possible additional training of the network after receiving new data and repeating item 4.

Note that although the data can be grouped by a geographical sign (Left and right banks of the Don river, Oka-Don and Kalachskaya lowlands), or by their origin (group of exogenous indicators, group of hydrological indicators, group of man-caused and social indicators), the developed complex allows carrying out clustering without combining the initial indicators or any additional stage of formation and selection of significant signs, since the ART networks are initially designed for processing of poorly structured data, and allow clustering the data with interdependent input parameters.

The obtained region's ecological situation assessment model can be applied in the long run as well. In this case, one can immediately identify two possible analysis directions. The first one consists in using the network for the primary clustering and constructing a scheme of the region's ecological situation management strategies, while the second one consists in tracking the ecological situation
Having once analyzed the global state of the region, the model allows using the new incoming data for additional training of the network, and its plasticity will ensure the ability to track the trends. The model also allows one to see the migration of the prototypes and to timely adjust the strategy schemes for responding to the regional environmental situation. The possibility to the number of clusters allows performing another type of analysis – sequential partitioning into 2, 3, ..., N clusters. The analysis of the corresponding prototypes’ splitting may allow the expert to make more accurate conclusions about the nature of the certain clusters and their accompanying characteristics of the real environmental ecology, which will lead to more effective management strategies.

4. Conclusion

The developed neural network system allows performing a comprehensive assessment of the region's ecological situation by a selected set of factors depending on the objectives of the required analysis. For example, one set of factors may be used for a preliminary environmental analysis when constructing industrial facilities, while for agricultural facilities it can be completely different. Medical institutions may be interested in analyzing exclusively man-caused, social and climatic factors, and so on. The clustering results and the obtained cluster prototype analysis make it possible to select a grounded mathematical model of the integral assessment of the region's ecological situation. Moreover, the constructed ecological situation analysis model and the developed software allow experts to keep track of trends in the region's ecological situation.

Thus, as a result of the conducted research, the scientific and methodological apparatus for the neural network approach to complexing of ecological information based on the data cluster structuring throughout the entire set of monitoring indicators has been developed and tested by the example of Voronezh oblast. On the whole, such a method is applicable to an arbitrary set of geological, hydrogeological, geophysical and ecological data. It allows investigating the spatial interaction of the elements of the system being studied and identifying the key factors influencing this set.

References

[1] Ivanov D 2005 Interaction of environmental components on the territory of the Voronezh region ArcReview vol 3 pp 1-3
[2] Karlin L N and Abramov V M 2006 Environmental and ecological risk management (Saint-Petersburg: RSHU publishing house) p 332
[3] Perelet R A, Umyvakin V M and Shevchuk A V 2014 Trudy Instituta sistemnogo analiza Rossiijskoj akademii nauk (Proceedings of Institute of system analysis of Russian Academy of Sciences) vol 64 pp 59-63
[4] Ledeneva T M, Umyvakin V M and Shvets A V 2016 Voprosy nauki (Science Questions) vol 1 pp 58-73
[5] Xu R and Wunsch D 2005 Survey of Clustering Algorithms IEEE transactions on neural networks vol 16 pp 645-678
[6] Keskin G A and Ozkan C 2009 An Alternative Evaluation of FMEA: Fuzzy ART Algorithm Quality and reliability engineering international vol 25 pp 647-661
[7] Kashirina I L and Fedutinov K A 2018 Sistemy upravleniya i informacionnye tekhnologii (Management systems and information technology) vol 71 pp 33-39
[8] Carpenter G and Grossberg S 2003 Adaptive Resonance Theory The Handbook of Brain theory and Neural Networks (Cambridge, MA: MIT Press) pp 87-90
[9] Grossberg S 2013 Adaptive Resonance Theory: How a brain learns to consciously attend, learn, and recognize a changing world Neural Networks vol 37 pp 1-47
[10] Kashirina I L, Lvovich J E and Sorokin S O 2015 Ehkonomika i menedzhment sistem upravleniya (Economics and management of control systems). vol 15 pp 330-337
[11] Kashirina I L, Lvovich J E and Sorokin S O 2017 Informacionnye tekhnologii (Information technologies) vol 23 pp 228-232
[12] Rousseeuw P 1987 Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis *Computational and Applied Mathematics* vol 20 pp 53–65

[13] Skovajsova L and Rojcek M 2011 Soft clustering algorithms based on neural networks *IEEE 12th International Symposium on Computational Intelligence and Informatics (CINTI)*, (Budapest) pp 439-442