

MATHEMATICAL AND ALGORITHMIC SUPPORT FOR THE INTEGRATION OF MLP AND SOM NEURAL NETWORK MODELS FOR MULTIDIMENSIONAL GEOPHYSICAL DATA ANALYSIS

E. Oshkodorov¹

¹Institute of Telecommunications and Global Information Space of the NAS of Ukraine, Kyiv, Ukraine

oshkodorov.e@gmail.com

The paper proposes a hybrid algorithmic approach to solving the binary classification problem of multidimensional geophysical data. The method is based on the integration of a multilayer perceptron (MLP) and a Kohonen self-organizing map (SOM), which makes it possible to combine the probabilistic approximation of the separating hypersurface with topological cluster analysis.

Keywords: machine learning, multilayer perceptron, Kohonen self-organizing map, reservoir classification, decision support systems, well logging.

Modern challenges of environmentally safe natural resource management require the development of accurate mathematical methods for analyzing complex, nonlinear, and multidimensional datasets [1]. Traditional approaches based on heuristic threshold values are often unable to account for hidden correlation and topological patterns in geophysical signals [1]. Therefore, an urgent task is the development of formalized neural network algorithms that ensure reproducibility of results and minimize the risks of erroneous decisions. The aim of this study is to develop and substantiate a mathematical model for the combined application of MLP and SOM neural networks for geological layer classification.

Problem Statement and Feature Space Normalization

Let a set of objects be given, each represented by a feature vector $X \in R^5$. The components of the vector correspond to five petrophysical parameters: averaged electrical resistivity values (shallow and deep probes), natural gamma activity, signal standard deviation, and probe contrast. The target classification function takes discrete values $Y \in \{0, 1\}$, where the value 1 identifies a productive reservoir layer. To ensure the convergence of gradient-based optimization algorithms, linear scaling of the feature space to the interval $[0;1]$ was applied using the min-max normalization rule [4]:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min} + \varepsilon}, \quad (1)$$

where ε is an infinitesimally small positive value introduced to prevent division by zero.

Mathematical classification models

Multilayer Perceptron (MLP). The discriminative function is implemented as a feedforward neural network with a single hidden layer. The use of a sigmoid activation function allows the output signal to be interpreted as the posterior probability of the vector belonging to a given class. The computational process is described by a system of nonlinear equations [4]:

$$h = \sigma(W_1x + b_1), \quad \hat{y} = \sigma(W_2h + b_2). \quad (2)$$

Model training reduces to the problem of minimizing the log loss function (binary cross-entropy):

$$L(y, \hat{y}) = -[y \ln \hat{y} + (1 - y) \ln(1 - \hat{y})]. \quad (3)$$

Kohonen Self-Organizing Map (SOM). For unsupervised structural learning, the SOM model is applied, which maps a multidimensional feature space onto a two-dimensional topological grid. In the process of competitive learning, a best-matching unit is determined for each vector x based on the minimum Euclidean distance criterion [3]: A neural network model has been developed, implemented as a separate module that interacts with the SCADA system via API. The processed results are displayed to the operator on an information panel with recommendations for further actions [5].

$$i^* = \arg \min_i \|x - w_i\|_2. \quad (4)$$

The adaptation of weight coefficients is performed iteratively, taking into account the topological neighborhood function $h_{i,i^*}(t)$:

$$w_i(t+1) = w_i(t) + \alpha(t) \cdot h_{i,i^*}(t) \cdot (x - w_i(t)). \quad (5)$$

To transform the SOM into a classifier, a neuron calibration procedure based on the majority voting rule has been developed.

Computational experiments and results

The performance evaluation of the proposed algorithmic framework was carried out on labeled geophysical datasets using standard metrics: accuracy, precision, recall, and F1-score. During the experiments, both models demonstrated complete convergence with the reference labels. Specifically, the MLP correctly classified 300 out of 300 samples, and the calibrated SOM map achieved perfect alignment with 52 out of 52 examples on the corresponding sub-sample. From the standpoint of pattern recognition theory, such a result serves as rigorous confirmation of the high degree of class separability within the established 5-dimensional feature space. At the same time, to statistically validate the generalization ability of the models and prevent overfitting, the methodology entails subsequent cross-validation on independent test sets.

Conclusions

The scientific novelty of this study lies in the mathematical substantiation and algorithmic implementation of an integrated decision-support framework in which nonlinear probabilistic classification (MLP) is complemented by topological clustering (SOM). The hybridization of these approaches significantly expands the analytical capabilities of the system: SOM performs the function of detecting structural anomalies and uncertainty regions within the feature space, while MLP provides the formation of the final decision together with a quantitative estimation of the confidence interval. The implementation of such models in SCADA systems represents a promising tool for minimizing environmental risks in geophysics [4].

References

1. Asquith G., Krygowski D., Basic Well Log Analysis, AAPG Methods in Exploration Series, Tulsa, 2004, 244 pp.
2. Kohonen T., Self-Organizing Maps, Springer, Berlin, 2001, 501 pp.
3. Bishop C. M., Pattern Recognition and Machine Learning, Springer, New York, 2006, 738 pp.
4. Kotsyuba I. H., Babiichuk S. M., Intelligent Methods for Geological Data Analysis in Decision Support Systems, *Environmental Safety and Natural Resources* **42** (2022), 2, 67–80.