DOI: https://doi.org/10.15276/hait.07.2024.1
UDC 004.942

Assessment of the quality of neural network models based on a multifactorial information criterion

Oleksandr O. Fomin[1)]
ORCID: https://orcid.org/0000-0002-8816-0652; fomin@op.edu.ua. Scopus Author ID: 57103429400
Valentyn A. Krykun[1)]
ORCID: https://orcid.org/0000-0002-3764-9255; 9901053@stud.op.edu.ua. Scopus Author ID: 57968989300
[1)] Odessa Polytechnic National University, 1, Shevchenko Avenue. Odessa, 65044, Ukraine

ABSTRACT

The paper is devoted to the problem of assessing the quality of machine learning models in the form of neural networks in the presence of several requirements for the quality of intelligent systems. The aim of this paper is to develop a multifactorial information criterion that allows choosing a machine learning model in the form of a neural network that best meets the set of requirements for accuracy and interpretability. This goal is achieved through the development and adaptation of multifactorial information criteria for evaluating models in the form of neural networks and, in a particular case, three-layer time delay neural networks used to identify nonlinear dynamic objects. The scientific novelty of the work lies in the development of multifactorial information criteria for the quality of machine learning models that take into account the accuracy and complexity indicators, which, unlike existing information criteria, are adapted to the evaluation of models in the form of neural networks. The practical usefulness of the work lies in the possibility of automatic selection of the simplest machine learning model that provides suitable accuracy when used in intelligent systems. The practical significance of the obtained results lies in the application of the proposed criteria for selecting a machine learning model in the form of a time delay neural network for identifying nonlinear dynamic objects, which allows to increase the accuracy of modeling while ensuring the simplest architecture of the neural network.

Keywords: Information quality criteria; modeling accuracy; complexity of machine learning models; nonlinear dynamic objects; neural networks

***For citation*: Fomin O. O., Krykun V. A. "Assessment of the quality of neural network models based on a multifactorial information criterion". *Herald of Advanced Information Technology.* 2024; Vol. 7 No. 1: 13–23. DOI: https://doi.org/10.15276/hait.07.2024.1**

INTRODUCTION

Today, we can observe the development of intelligent systems to the point where they become an integral and organic part of our daily life. Examples of this are the successes of the using intelligent systems in many industries, from information search systems, voice assistants and recommendation systems to transportation management systems, high-tech manufacturing, which are widely used in medicine, biology, education, etc. [1]. Intelligent systems continue to penetrate various spheres of our society, demonstrating significant success in new areas of application and changing our perception and interaction with the world around us.

These systems make a significant contribution to the quality of human life and create new opportunities for people by saving resources, increasing productivity, and optimizing processes, which increases their importance in the modern world [1, 2].

The success of intelligent systems is determined by the simultaneous superposition of many factors, including the following: a significant increase in computing resources, which is ensured by the increase in hardware performance and the development of cloud technologies; development of machine learning algorithms, primarily neural network (NN) architectures [3, 4].

However, the quality of an intelligent system depends on the balanced development of all its components, including mathematical support [5, 6]. Machine learning models, being the mathematical basis, play a crucial role in the functioning of intelligent systems.

The task of comparing and selecting the best machine learning models is becoming a key issue in improving the quality of intelligent systems. A successful model selection not only increases the accuracy and reliability of the solution, but also improves the overall performance and efficiency of the system as a whole [6, 7].

However, today there is a noticeable gap between the potential of machine learning algorithms and their practical application due to limitations in the direction of mathematical models and methods [7].

This paper considers one of the aspects of mathematical support for intelligent systems – quality assessment of machine learning models.

The choice of a machine learning model depends on various factors, including goals and objectives, the nature of the data, the size of the data set, and the complexity of the model [6]. Different models have different strengths and weaknesses, and choosing the best model is not always easy.

Existing common model evaluation methods, such as statistical metrics, cross-validation, and information criteria, provide different approaches to evaluating certain properties of machine learning models [8]. However, the choice of an appropriate criterion for a particular case is complicated by the presence of modern quality requirements and limitations of the use of intelligent systems and requires careful analysis and understanding of the features of models and data.

Based on the analysis of the advantages and limitations of traditional methods for assessing the quality of models, the paper attempts to bridge the gap between the development of algorithmic and mathematical support for intelligent systems by developing criteria for assessing the quality of machine learning models.

LITERATURE REVIEW

The literature presents various approaches to comparing machine learning models,

Summarizing these approaches, we can distinguish two directions [8]. The first approach is to form an evaluation based on statistical tests [8, 9]. The second approach is to estimate the threshold value on a cross-validation or deferred sample (cross-validation or holdout) [8, 10].

In the specific case of comparing regression models, one of the simplest and most common approaches is to use statistical metrics such as mean absolute error (*mae*), mean square error (*mse*), and a family of relative scores derived from the considered criteria [9].

These metrics estimate the average error of the model on the test data and allow assessing the accuracy of the model. When it is necessary to select the model that produces the smallest number of large errors, the *mse* criterion is used to assess the accuracy of reproducing the original signal. When it is necessary to select the model that produces the smallest average error, the *mae* criterion is used to assess the accuracy of the output signal reproduction.

The coefficient of determination (R^2) is also used to assess the quality of models, which measures the proportion of variation in the dependent variable that is explained by the model.

However, certain problems arise when choosing these criteria for comparing regression models. For example, R^2 may be uninformative when comparing models with different numbers of parameters; *mae* and *mse* do not take into account model complexity.

To take into account several parameters in the model comparison criteria, information criteria such as Akaike's criterion (*aic*) and Schwartz's criterion (*bic*) have been developed [11]. These criteria are used to select the best model among several models built on the same data set. They take into account both model accuracy and complexity, which allows making a more balanced choice between simplicity and model accuracy.

Information criterion *aic* is a measure based on the estimation of the loss of information when reducing the number of model parameters using a likelihood function.

The information criterion *bic* is a statistical measure based on Bayesian principles for model selection. It is similar to the *aic* criterion, but emphasizes the simplicity of the model. The *bic* criterion imposes a more severe penalty for model complexity, which should lead to the selection of simplified models.

The *aic* criterion is used when the focus is on model selection and a compromise between model fit and model complexity needs to be considered. It is useful in a wide range of statistical analysis. The Schwartz *bic* criterion is particularly useful when complex models need to be penalized strictly, for example, in situations with limited data where simplicity is highly valued.

The choice of quality criteria for machine learning models largely depends on the specific context of the task and the requirements for the model.

Therefore, when comparing machine learning models, modern tasks require taking into account additional aspects of modeling, taking into account the purpose of the study the nature of the data, methods and algorithms, etc.

To overcome the disadvantages of the considered methods, this paper uses an approach to selecting the best machine learning models based on the development and adaptation of multifactorial information criteria for evaluating models in the form of NN and, in a special case, three-layer time delay NN used to identify nonlinear dynamic objects.

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

PROBLEM STATEMENT

When developing a machine learning model, researchers are faced with different types of models, each with its own strengths and weaknesses, as well as parameters and hyperparameters that can be tuned to achieve optimal performance in a certain sense.

The problem of evaluating the quality of models in machine learning is to develop a quantitative metric based on a set of factors determined by current trends and current standards for the development of intelligent systems. This metric is used to select the most suitable machine learning model for a particular task from the set of available models.

The formal statement of the problem of developing a metric for evaluating the quality of machine learning models is as follows.

Let there be a set of objects $X=\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ and a set of labels (answers) of these objects Y, such that $Y=\{y_1, y_2, \ldots, y_n\}$. Suppose there is also a training set $D_{train}=\{(\mathbf{x}_1,y_1), (\mathbf{x}_2,y_2), \ldots, (\mathbf{x}_n, y_n)\}$, where each pair $(\mathbf{x}_i, y_i)$ represents a description of object x_i and the corresponding label y_i (i=1, 2, …, n).

Suppose that there are p functions $F=\{f_1, f_2, ..., f_p\}$– machine learning models built on the training set D_{train}.

Then, the task of evaluating the quality of machine learning models is to build a certain metric $Q(\theta, f_k, D_{train})$, where θ is a set of hyperparameters (a set of factors) determined by the current requirements for the quality of an intelligent system, k=1, 2, ..., p.

At the same time, the minimum of the constructed metric is achieved on the model f^* from the set F, which best meets the current requirements for the quality of a particular intelligent system:

$$f^* = arg\ min\ Q(\theta, f_j, D_{train}),\ j=1,2,\ldots,k. \quad (1)$$

The developed metric Q must meet the following requirements [8, 9]:

1. The metric must be invariant with respect to the scale and shift of the features.

2. The metric should be sensitive to differences in model performance and allow detecting significant differences between them.

3. The metric should be easily interpretable by users, allowing them to understand which model best meets their needs.

Based on the demands of practice, the main attention in developing the Q metric should be paid to its adaptation for the case of evaluating regression models in the form of NN and, in a special case, three-layer time delay NN used to identify nonlinear dynamic objects.

Thus, the problem of assessing the quality of models in machine learning is reduced to solving the problem of constructing a metric Q that satisfies the above requirements and allows for an objective assessment of the quality of regression models of machine learning in the form of NN based on a set of factors determined by the requirements for intelligent systems.

The following factors are most often considered as the mentioned ones [12, 13].

– *Solution quality*: building a model that fits the target variable as accurately as possible on the test data.

– *Generalization ability*: selecting a model that summarizes the dependencies in the data and demonstrates a good fit to new data that has not been used before.

– *Interpretability*: the ability to explain how and why the model produces certain results.

– *Resources*: the consumption of computing resources or the requirement of a certain amount of training data for effective learning.

An effective solution to the problem of evaluating the quality of machine learning models requires a systematic comparison of different models based on a set of factors determined by the quality requirements for intelligent systems and existing standards, taking into account the specifics of the machine learning task and the purpose of the study.

PURPOSE AND OBJECTIVES OF THE STUDY

The aim of this paper is to develop a multifactorial information criterion that allows choosing a machine learning model in the form of a neural network that best meets the set of requirements for accuracy and interpretability.

To achieve this goal, we set the following objectives.

1. Analysis of modern requirements for intelligent systems established by current standards.

2. Selection of a family of characteristics that have the greatest impact on the quality of intelligent systems.

3. Developing a metric for assessing the quality of machine learning models in the form of NN by developing multifactorial information criteria.

4. Investigating the effectiveness of the proposed metric for assessing the quality of machine learning models in the form of time delay NN in the task of modeling a test nonlinear dynamic object.

QUALITY REQUIREMENTS FOR INTELLIGENT SYSTEMS

1. Overview of modern quality standards in the field of system and software engineering

The quality requirements for intelligent systems should be based on existing standards [14]. The quality assessment of artificial intelligence systems is currently regulated by the international standards ISO/IEC 25059:2023 Software engineering, ISO/IEC TS 25058:2024 Systems and software engineering and ISO/IEC 2502n Quality Measurement.

The documents are part of the Systems and software Quality Requirements and Evaluation (SQuaRE) series of standards. They adapt the software and systems quality assessment methodology that has been successfully used in other well-known ISO/IEC standards to assess the quality of artificial intelligence systems.

The requirements listed in SQuaRE are primarily in the category of product requirements, which define quality characteristics that should be considered during the design, development, and testing of, a software product or system. If we look at this standard more broadly, it affects both the algorithmic part and the models used by the software.

The reason that complicates the direct use of the quality measurement standard is that it contains a framework model for measuring product quality and mathematical definitions of quality indicators, but does not provide practical guidance on building multifactorial quality criteria and adapting them to certain types of machine learning models [15]. In this regard, the task of building a multifactorial quality model for intelligent systems software becomes urgent, starting with the definition of quality indicators and metrics for their evaluation and ending with the integration of all measured indicators within a single model that can quantify the solution for further comparison of models and automation of the process of building the best machine learning model as part of an intelligent system.

In the next section, we present an algorithm for building the above-mentioned model of software quality of intelligent systems.

2. Algorithm for building a quality model of intelligent systems software

Defining and measuring the quality of intelligent systems software remains a difficult task that is usually solved with the help of quality models [16, 17]. Existing quality models of intelligent systems software are too abstract to be implemented in practice [16] and require additional instructions for their application. Therefore, in the following, we propose a quality model for intelligent systems software that allows turning an abstract and difficult to measure quality concept into a practical and effective tool for managing the quality of machine learning models and that can be conveniently implemented as part of intelligent system software.

The process of building a quality model for intelligent systems can be conveniently divided into three steps [17, 18].

1. Defining the quality meta-model. At this stage, the components of the basic structure of the model that affect the quality of the intelligent system software are determined. The components of the basic structure of the model are selected from the full list defined by existing standards, based on the context of the quality model application. This provides a basis for filling the quality model with the most relevant components, instead of trying to make a model from the complete list of all quality components defined by the existing standards.

2. Define metrics to quantify each component of the meta-model. Measuring the degree of quality of individual components is a complex task, as the concept of quality can have different aspects depending on the purpose and requirements.

3. Build a multifactorial quality model based on the meta-model defined in step one and the metrics adopted in step two.

3. Defining the components of the software quality meta-model

The quality model of intelligent systems software and metrics for evaluating its components are based on the analysis of existing requirements in the field of system and software engineering, as regulated by SQuaRE standards.

Thus, to build a quality model of intelligent systems software, first, significant quality components are identified: characteristics that determine the functionality, reliability and safety in solving a specific application task.

Taking into account the selected components, the quality model of intelligent systems software in general takes the following form:

$$Q = \langle Q_1, Q_2, ..., Q_s \rangle, \tag{2}$$

where Q_r is a quality component; r=1, 2, …, s; s is the number of selected quality components.

The quality model (2) allows for an adequate comparison of machine learning models with each

ISSN 2663-0176 (Print)
ISSN 2663-7731 (Online)

other, and therefore should address the existing conflicting requirements for intelligent systems, the most common of which are the following.

– *Functionality-complexity of model interpretation*: on the one hand, a model with fewer parameters is better in terms of resources and interpretation; on the other hand, such a model will be inferior in accuracy.

– *Functionality-complexity of deployment*: setting up a GPU server to deploy more functional models (e.g., neural network models) can be significantly more complex than setting up a regular service to train simpler models (e.g., logistic regression) using CPU resources.

– *Functionality-efficiency in the use of resources*: the choice between models with a significant difference in training time: for example, the training time of a gradient boosting model and a more functional neural network can differ by an order of magnitude.

– *Functionality-performance*: typically, more functional models are significantly slower.

– *Functionality-reliability*: complex models are vulnerable to adversarial attacks, when they add small noise to the original data, which can cause the model to change its decision; simple models are more resistant to such attacks.

The solution to one of the above contradictions, namely, "*Functionality-complexity of model interpretation*", which corresponds to the purpose set in this work, is carried out on the example of the task of modeling a test nonlinear dynamic object using time delay NN. Given the task, the analysis of ISO/IEC 25059:2023, ISO/IEC TS 25058:2024 standards is performed and the following significant characteristics are identified and metrics for their measurement are defined.

1. *Functionality*: the system shall be able to reproduce the original signals on the test dataset.

Q_F metric: accuracy of reproduction of the original signal, e.g., *msa* and *mae*.

2. *Complexity*: the simpler the model structure, the fewer variables, equations and elements used to describe the system behavior, the greater the degree of its understanding (interpretability). In general, model complexity is often a compromise between accuracy, amount of information, and resources required to create, analyze, and use a model.

Q_C metric: the number of model parameters.

The most common software-oriented criteria for assessing model complexity [19, 20], [21]:

– number of valid multiplications of the model;

– the number of multiplication operations taking into account the bit depth (bit accuracy of input data, weights, and activation function);

– the number of fixed-point operations (taking into account shift operations and adders).

In this paper, to improve the interpretability of models, we use a simplified criterion based on the number of model parameters. Given that the structure of a machine learning model is determined with parameter accuracy, the number of valid multiplications in the model implementation can be replaced by the number of model parameters, which greatly simplifies the determination of the model complexity in general. This assumption is especially useful when determining the complexity of models in the form of NN.

For example, the criterion for assessing the complexity of a machine learning model in the form of a fully connected NN takes into account the weighting coefficients and shift coefficients of the model and takes the form:

$$Q_C = n_k + \sum_{i=1}^{k-1} n_i (n_{i+1} + 1), \quad (3)$$

where k is the number of NN layers; n_i is the number of neurons in the i-h layer.

For the special case of a three-layer fully connected NN with time delays, which has one output and is used to identify nonlinear dynamic objects, expression (3) takes the simplified form:

$$Q_C = (M+1)(K+1) + K, \quad (4)$$

where M is number of neurons in the input layer of the NN; K is number of neurons in the hidden layer.

Taking into account the selected components, the software quality model of intelligent systems takes the following form:

$$Q = \langle Q_F, Q_C \rangle, \quad (5)$$

where Q_F is accuracy of output signal reproduction; Q_C is number of parameters in the model.

The quality model (5) must meet two contradictory requirements: functionality-complexity – it must be able to reproduce the behavior of the object as best as possible and at the same time be user-friendly (have the simplest possible model structure). Increasing the model's fit to the data is usually associated with its complexity, and the more complex the model, the lower its interpretability. Therefore, when choosing between a simple and a complex model, the latter should significantly increase the model's fit to the data to justify the increase in complexity and the corresponding decrease in interpretability. If this condition is not met, the simpler model should be chosen.

Thus, in order to assess the extent to which a change in a certain quality indicator affects the quality of the entire system, it is advisable to develop criteria for comparing the quality of models. In this case, well-known and popular model quality metrics, such as *mse* and *mae*, cannot be applied because they do not take into account model complexity.

In the next section, it proposes a multifactorial quality model that takes into account several components of quality indicators.

BUILDING A MULTIFACTORIAL MODEL OF MACHINE LEARNING MODEL QUALITY

To solve the problem of assessing the quality of machine learning models, the *aic* and *bic* criteria based on the likelihood ratio are widely used. For models based on other indicators, there are no such criteria for model selection. In addition, the complexity indicator used in these criteria is too general and needs to be refined in accordance with the type of machine learning model used to build an intelligent system.

In general, *aic* is calculated by the formula [11]:

$$aic=2k-2ln(L), \tag{6}$$

where k is number of model parameters; L is the value of the model's likelihood function.

The best model is the one with the lowest *aic* value.

The expression (6) shows that the growth of the criterion is mainly due to an increase in the number of model parameters, not its error. In other words, the model is penalized more for increasing the number of parameters than for the share of unexplained error variance. Thus, the *aic* criterion is to select the model with the minimum number of parameters that explain the largest share of the error variance.

The *bic* criterion is based on the fact that as the number of model parameters increases, the value of the likelihood function increases, but there is a possible overfitting effect. When there are too many model parameters, the contribution of each of them to the value of the likelihood function becomes small, and they lose their significance.

Therefore, the task of choosing a model is to include a minimum of parameters that would nevertheless make the greatest contribution to the value of the likelihood function.

The value of the *bic* criterion is calculated by the formula [11]:

$$bic=k\cdot ln(v)-2ln(L), \tag{7}$$

where v is training sample size.

Both criteria are widely used to analyze time series and solve regression problems.

Developing this approach, we propose a metric for assessing the quality of machine learning models in the form of NN.

The metric consists of indicators of functionality and complexity and has the following form:

$$Q= ln(Q_c) -ln(1/l), \tag{8}$$

where l is value of the model loss function.

The best model is the one for which the value of Q is minimal.

The quality criterion for a machine learning model in the form of a fully connected neural network can be obtained by substituting the value of Q_c from expression (3) into expression (8):

$$Q = ln(n_k + \sum_{i=1}^{k-1}(n_i+1)n_{i+1}) - ln(1/l) \tag{9}$$

In a particular case, for a common architecture of a three-layer time delay neural network used to identify nonlinear dynamic objects, expression (9) takes the following form if we replace the value of Q_c with expression (4):

$$Q=ln((M+1)(K+1)+K)-ln(1/l), \tag{10}$$

To compare the two models Q_1 and Q_2, it can use the following relation:

$$q = Q_2 / Q_1. \tag{11}$$

If q is less than 1, then model Q_2 performs better than model Q_1, if q is equal to 1, then the models are equal in terms of quality, if q is greater than 1, then model Q_1 performs better than model Q_2.

The proposed model (10) allows us to quantify the quality assessment of machine learning models, and expression (11) helps to compare several models with each other. Thanks to these models, it becomes possible to automate the process of optimizing machine learning models according to the target criterion.

In the next section, the effectiveness of the proposed criterion (10) for assessing the quality of machine learning models is investigated using a test training set as an example.

EXPERIMENT SETUP

1. Simulation model of the test object

The proposed multifactorial criterion for the quality of machine learning models is tested on the example of a test object with nonlinear dynamic characteristics. The test object is a structure with a first-order dynamic link and a nonlinear feedback link (Fig. 1) [22].

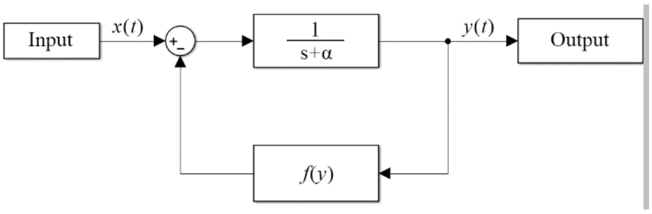

Fig. 1. **Simulation model of a test nonlinear dynamic object**
Source: **compiled by the authors**

The training sample obtained from the data of the input/output simulation experiment with the test object is a set of signals – object responses $y(t)$ to test input signals $x(t)$ in the form of pulse, step, linear, and harmonic functions with different amplitudes a.

When performing the input/output experiment, the following parameters of the simulation model were adopted:

$$\alpha=2.64,\ \beta=1.45,\ f(y)=\beta y^2(t).$$

2. Building a machine learning model

To identify the designated test object, a model in the form of a time delay NN is used. The most commonly used structure of such an NN consists of three layers: input, hidden, and output.

In this structure, the input layer includes M neurons, corresponding to the memory length of the object model. The number of neurons M is chosen to best reflect the dynamic properties of the object. The input layer receives data $\mathbf{x}(t_n)=[x(t_n), x(t_{n-1}), \dots , x(t_{n-M-1})]$, $t_n=n\Delta t$, $n=1, 2, \dots$

The hidden layer includes K neurons with a nonlinear activation function. The number of neurons K is chosen to best reflect the nonlinear properties of the object.

The activation function of the hidden layer neurons is a rectified linear unit (ReLu), which contributes to the sparsity of activations and leads to a significant reduction in the computational load during training.

The output layer of the time delay NN includes an adder with weights from the hidden layer neurons at the inputs and a multiplier by a constant coefficient at the output.

The signal $y(t_n)$ at the output layer at time t_n depends on the values of the input signal $\mathbf{x}(t_n)$ and is determined by the expression [23]:

$$y(t_n)=b_0+S_0\sum_{i=1}^{K} w_i\, S_i\left(b_i+\sum_{j=1}^{M} w_{i,j}\,x(t_{n-j})\right), \qquad (12)$$

where b_0, b_i are the bias of the neurons of the output and hidden layers, respectively; S_0, S_i are the activation functions of the neurons of the output and hidden layers, respectively; w_i, $w_{i,j}$ are the weighting coefficients of the neurons of the output and hidden layers, respectively.

Such a model can be trained for dynamic behavior taking into account nonlinear characteristics on the input-output data.

3. Choosing the best object model

To determine the best values of M and K in the adopted three-layer structure of the NN, a number of models with different numbers of neurons in the input and hidden layers were built.

The dependence of the averaged loss function determined by the results of 5 experiments for each combination of M and K on the number of neurons in the input and hidden layers is shown in Fig. 2. The *mse* criterion is used as the loss function.

Using expression (10), we calculate the quality criterion for each model built. The dependence of the criterion Q on the number of neurons in the input and hidden layers is shown in Fig. 3.

For comparison, Fig. 4 shows the dependence of the *aic* criterion on the number of neurons M and K in the input and hidden layers, respectively.

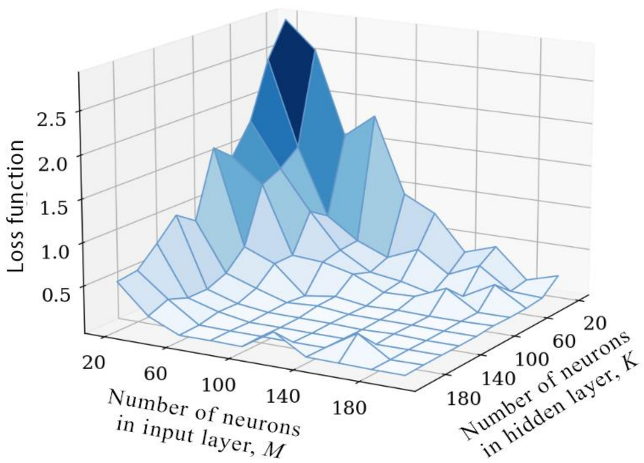

Fig. 2. **Dependence of the loss function on the number of neurons in the input and hidden layers**
Source: **compiled by the authors**

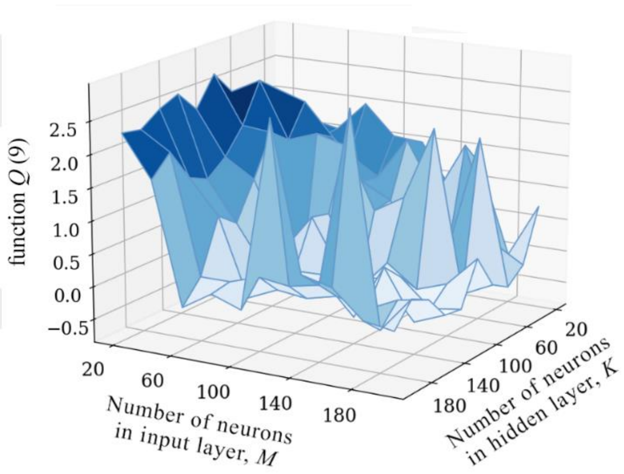

Fig. 3. **Dependence of the *Q* criterion on the number of neurons in the input and hidden layers**
Source: **compiled by the authors**

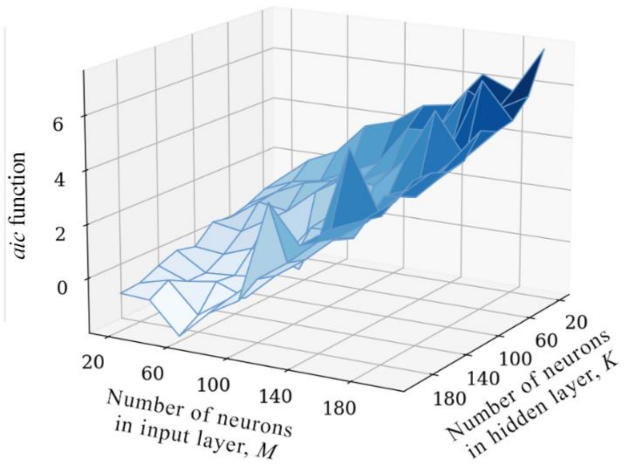

Fig. 4. **Dependence of the *aic* criterion on the number of neurons in the input and hidden layers**
Source: **compiled by the authors**

DISCUSSION OF THE RESULTS

According to the results of the experimental study of the proposed criterion (10), a neural network structure with the number of neurons in the input and hidden layers M=60 and K=80, respectively, was chosen as a model of a nonlinear dynamic object. This NN structure provides a compromise between accuracy (*loss* = 0.3) and complexity of the constructed NN – the smallest value of criterion (10).

In this case, the NN model of the test object selected by the *aic* criterion has a loss of 0.8 and a structure with the number of neurons in the input layer M=100 and the number of neurons in the hidden layer K=20 (in general, *aic* does not depend on *K*). The *bic* criterion has values close to those of the *aic* function, but tends to favor simpler models than *aic*.

The results of the computational experiment demonstrate the advantages of using the proposed criterion (10) exactly over the existing information criteria *aic* and *bic* with comparable complexity of the three-layer NN.

The advantage of criterion (10) is due to a more detailed consideration of the quality model component responsible for the complexity of the machine learning model.

As the complexity of the machine learning model increases, the sensitivity of the proposed criterion (10) decreases and it approaches the *bic* criterion, provided that the training sample size remains unchanged. In this case, the accuracy component of the quality model affects the quality assessment to a much greater extent.

The area of effective use of the proposed criterion is limited to a narrow class of machine learning models in the form of a fully connected three-layer NN. The proposed criterion (9) can be used to assess the quality of fully connected NN models with a different structure.

The practical utility of the work lies in the ability to automatically select the simplest machine learning model that provides suitable identification accuracy when used in intelligent systems.

The practical significance of the results obtained is the application of the proposed criterion for selecting a machine learning model in the form of a time delay NN for identifying nonlinear dynamic objects, which allows choosing the most accurate model of an object from a set of models of comparable complexity.

CONCLUSIONS

As a result of the work, the task of developing a multifactorial information criterion for selecting a machine learning model in the form of a neural network that best meets the set of requirements for the accuracy and interpretability of an intelligent system was successfully solved.

To achieve this result, the author analyzed modern standards for the quality of system and software engineering. Based on the results of the analysis, the algorithm for building a quality model of intelligent systems software is generalized, and the components of the meta-model of quality of intelligent systems software are defined: accuracy and complexity of machine learning models.

For the identified quality components, quality assessment metrics are presented. The expression for estimating the complexity of the model in the case of using a fully connected neural network as a machine learning model is further developed. A special case of estimating the complexity of a time delay neural network in the form of a fully connected three-layer structure is considered.

Based on the above metrics, a multifactorial information criterion for the quality of machine learning models in the form of a fully connected neural network is proposed.

The effectiveness of the proposed multifactorial criterion for assessing the quality of machine learning models is proved by solving the problem of identifying a test nonlinear dynamic object in the form of a time delay neural network, which allows choosing the most accurate model of the object from a set of models of comparable complexity compared to the known information criteria *aic* and *bic*.

REFERENCES

1. Maglogiannis, I., Iliadis, L. & Pimenidis, E. (Eds.). "Artificial intelligence applications and innovations". *16th IFIP WG 12.5 International Conference, AIA. Proceedings, Part I.* Neos Marmaras: Greece. 2020. DOI: https://doi.org/10.1007/978-3-030-49161-1.

2. Wirtz, B. W., Weyerer, J. C. & Geyer, C. "Artificial intelligence and the public sector – applications and challenges". *International Journal of Public Administration.* 2019; 42 (7): 596–615, https://www.scopus.com/record/display.uri?eid=2-s2.0-85050557949&origin=resultslist.
DOI: https://doi.org/10.1080/01900692.2018.1498103.

3. Hamm, P. & Klesel, M. "Success factors for the adoption of artificial intelligence in organizations: A literature review". *AMCIS Proceedings*. 2021.

4. Komarov, O. V. "Reducing the search area of genetic algorithm using neural network autoencoder". *Herald of Advanced Information Technology*. 2020; 3 (3): 113–124. DOI: https://doi.org/0.15276/hait 03.2020.1.

5. Chisty, N. M. A. & Adusumalli, H. P. "Applications of artificial intelligence in quality assurance and assurance of productivity". *ABC Journal of Advanced Research.* 2022; 11 (1): 23–32. DOI: https://doi.org/10.18034/abcjar.v11i1.625.

6. Kharchenko, V, Fesenko, H. & Illiashenko, O. "Quality models for artificial intelligence systems: characteristic-based approach, development and application". *Sensors.* 2022; 22 (13): 4865, https://www.scopus.com/record/display.uri?eid=2-s2.0-85132828611&origin=resultslist.
DOI: https://doi.org/10.3390/s22134865.

7. Kutyniok, G. "The mathematics of artificial intelligence". *Proc. Int. Cong. Math.* 2022; 7: 5118–5139. DOI: https://doi.org/10.4171/ICM2022/141.

8. Tian, J. "Quality-evaluation models and measurements". *IEEE Software.* 2004; 21 (3): 84–91. DOI: https://doi.org/10.1109/MS.2004.1293078.

9. Trzęsiok, M. "Measuring the quality of multivariate statistical models". *Acta Universitatis Lodziensis. Folia Oeconomica*. 2019; 6 (339): 99–100. DOI: https://doi.org/10.18778/0208-6018.339.06.

10. Pal, K. & Patel, B. V. "Data classification with K-fold cross validation and holdout accuracy estimation methods with 5 different machine learning techniques". *Fourth International Conference on Computing Methodologies and Communication*. Erode: India. 2020. p. 83–87. DOI: https://doi.org/10.1109/ICCMC48092.2020.ICCMC-00016.

11. Agiakloglou, C. & Tsimpanos, A. "Evaluating the performance of AIC and BIC for selecting spatial econometric models". *J Spat Econometrics*. 2023; 4 (2): 1–35. DOI: https://doi.org/10.1007/s43071-022-00030-x.

12. Ghallab, M. "Responsible AI: requirements and challenges". *AI Perspect*. 2019; 1 (3): 1–7. DOI: https://doi.org/10.1186/s42467-019-0003-z.

13. Kästner, C. & Kang, E. "Teaching software engineering for AI-enabled systems". *The 42nd International Conference on Software Engineering. Software Engineering Education and Training.* 2020. Available from: https://arxiv.org/abs/2001.06691. – [Accessed: Feb. 2023].

14. ISO/IEC 25010. "Systems and software engineering. Systems and software quality requirements and evaluation (SQuaRE). System and software quality models". 2011. Available from: https://www.iso.org/standard/35733.html. – [Accessed: Feb. 2023].

15. Liubchenko, V. V., Komleva, N. O., Zinovatna, S. L. & Pysarenko, K. O. "Framework for systematization of data science methods". *Applied Aspects of Information Technology*. 2021; 4 (1): 80–90. DOI: https://doi.org/10.15276/aait.01.2021.7.

16. Siebert, J., Joeckel, L., Heidrich, J. et al. "Construction of a quality model for machine learning systems". *Software Qual J.* 2022; 30: 307–335. DOI: https://doi.org/10.1007/S11219-021-09557-Y.

17. Hamada, K. et al. "Guidelines for quality assurance of machine learning-based artificial intelligence". *SEKE: the 32nd International Conference on Software Engineering & Knowledge Engineering.* 2020. p. 335–341. DOI: https://doi.org/10.1142/S0218194020400227.

18. Ishikawa, F. "Concepts in Quality Assessment for Machine Learning – From Test Data to Arguments". In: Trujillo, J. C. et al. (eds.) Conceptual modeling. *37th International Conference*. Xi'an: China. *Springer,* Cham: Switzerland. 2018. p. 536–544. DOI: https://doi.org/10.1007/978-3-030-00847-5_39.

19. Nistala, P., Nori, K. V. & Reddy, R. "Software quality models: A systematic mapping study". *Proceedings IEEE/ACM International Conference on Software and System Processes ICSSP*. Montréal: Canada. 2019. p. 125–134. DOI: https://doi.org/10.1109/ICSSP.2019.00025.

20. Marselis, R. & Shaukat, H. "Machine intelligence quality characteristics. How to measure the quality of Artificial Intelligence and robotics". *Sogeti*. Paris: France. 2018.

21. Nakamichi, K. et al. "Requirements-driven method to determine quality characteristics and measurements for machine learning software and it's evaluation". *IEEE 28th International Requirements Engineering Conference (RE)*/ Zurich: Switzerland, 2020. p. 260–270, DOI: https://doi.org/10.1109/RE48521.2020.00036.

22. Fomin, O. et al. "Mathematical models of software quality assurance for interpretation of dynamic neural networks". *Scientific notes of Taurida National V.I. Vernadsky University. Series: Technical Sciences.* 2023; 34 (73) (5): 250–256. DOI: https://doi.org/10.32782/2663-5941/2023.5/39.

23. Fomin, O. et al. "Interpretation of dynamic models based on neural networks in the form of integral-power series". In: Arsenyeva, O., Romanova, T., Sukhonos, M., Tsegelnyk, Y. (eds). *Smart Technologies in Urban Engineering. Lecture Notes in Networks and Systems*. Springer. Cham. 2022; 536: 258–265, https://www.scopus.com/record/display.uri?eid=2-s2.0-85050654902&origin=resultslist.
DOI: https://doi.org/10.1007/978-3-031-20141-7_24.

Conflicts of Interest: the authors declare no conflict of interest

Received 20.12.2023
Received after revision 28.02.2024
Accepted 15.03.2024

DOI: https://doi.org/10.15276/hait.07.2024.1
УДК 004.942

Оцінювання якості нейромережевих моделей на основі багатофакторного інформаційного критерію

Фомін Олександр Олексійович[1)]
ORCID: https://orcid.org/0000-0002-8816-0652; fomin@op.edu.ua. Scopus Author ID: 57103429400
Крикун Валентин Андрійович[1)]
ORCID: https://orcid.org/0000-0002-3764-9255; 9901053@stud.op.edu.ua. Scopus Author ID: 57968989300
[1)] Національний університет «Одеська політехніка», проспект Шевченка, 1. Одеса, Україна

АНОТАЦІЯ

Робота присвячена проблемі оцінювання якості моделей машинного навчання у вигляді нейронних мереж при наявності декількох вимог до якості інтелектуальних систем. Метою роботи є розробка багатофакторного інформаційного критерію, що дозволяє вибирати модель машинного навчання у вигляді нейронної мережі, яка найкраще задовольняє сукупності вимог до точності та інтерпретованості. Ця мета досягається шляхом розвитку та адаптації багатофакторних інформаційних критеріїв для оцінки моделей у вигляді нейронних мереж та, в окремому випадку, тришарових нейронних мереж із часовими затримками, що використовуються для ідентифікації нелінійних динамічних об’єктів. Наукова новизна роботи полягає у розробленні багатофакторних інформаційних критеріїв якості моделей машинного навчання, що враховують показники точності та складності, які на відміну від існуючих інформаційних критеріїв адаптовані до оцінки моделей у вигляді нейронних мереж. Практична користь роботи полягає у можливості автоматичного вибору найпростішої моделі машинного навчання, що забезпечує придатну точність при використанні в інтелектуальних системах. Практичне значення одержаних результатів полягає у застосуванні запропонованих критеріїв для вибору моделі машинного навчання у вигляду нейронної мережі з часовими затримками для ідентифікації нелінійних динамічних об’єктів, що дозволяє підвищити точність моделювання при забезпеченні найпростішої архітектури нейронної мережі.

Ключові слова: інформаційні критерії якості; точність моделювання; складність моделей машинного навчання; нелінійні динамічні об'єкти; нейронні мережі

ABOUT THE AUTHORS

Oleksandr O. Fomin - Doctor of Engineering Sciences, Professor. Department of Computerized Systems and Software Technologies, Odessa Polytechnic National University, 1, Shevchenko Ave. Odessa, 65044, Ukraine
ORCID: https://orcid.org/0000-0002-8816-0652; fomin@op.edu.ua. Scopus Author ID: 57103429400
***Research field*:** Math modeling; information systems; intelligent control systems

Фомін Олександр Олексійович - доктор технічних наук, професор кафедри Комп’ютеризованих систем та програмних технологій. Національний університет «Одеська політехніка», проспект Шевченка, 1, Одеса, 65044, Україна

Valentyn A. Krykun - graduate student, Department of Computerized Systems and Software Technologies. Odessa Polytechnic National University, 1, Shevchenko Ave. Odessa, 65044, Ukraine
ORCID: https://orcid.org/0000-0002-3764-9255; 9901053@stud.op.edu.ua. Scopus Author ID: 57968989300
Research field: Modeling; intelligent information technologies

Крикун Валентин Андрійович - аспірант кафедри Комп’ютеризованих систем та програмних технологій. Національний університет «Одеська політехніка», проспект Шевченка, 1, Одеса, 65044, Україна