

The Role and Place of Artificial Neural Network Architectures Structural Redundancy in the Input Data Prototypes and Generalization Development

Conrad Onésime Oboulhas Tsahat, Ngoulou-A-Ndzeli, Béranger Destin Ossibi

Ecole Nationale Supérieure Polytechnique, Université Marien Ngouabi, Brazzaville, Republic of Congo

Email: oboulhas@yahoo.fr, becker20000@yahoo.fr, destinberanger2002@yahoo.fr

How to cite this paper: Oboulhas Tsahat, C.O., Ngoulou-A-Ndzeli and Ossibi, B.D. (2024) The Role and Place of Artificial Neural Network Architectures Structural Redundancy in the Input Data Prototypes and Generalization Development. *Journal of Computer and Communications*, 12, 1-11. <https://doi.org/10.4236/jcc.2024.127001>

Received: April 10, 2024

Accepted: June 25, 2024

Published: June 28, 2024

Copyright © 2024 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Neural Networks (NN) are the functional unit of Deep Learning and are known to mimic the behavior of the human brain to solve complex data-driven problems. Whenever we train our own neural networks, we need to take care of something called the generalization of the neural network. The performance of Artificial Neural Networks (ANN) mostly depends upon its generalization capability. In this paper, we propose an innovative approach to enhance the generalization capability of artificial neural networks (ANN) using structural redundancy. A novel perspective on handling input data prototypes and their impact on the development of generalization, which could improve to ANN architectures accuracy and reliability is described.

Keywords

Multilayer Neural Network, Multidimensional Nonlinear Interpolation, Generalization by Similarity, Artificial Intelligence, Prototype Development

1. Introduction

Knowledge representation and search are two fundamental problems that occupy artificial intelligence (AI) developers. The first relates to the new knowledge acquisition and AI “strong” level degree is achieved depending on its solution [1]. At the same time, many cybernetic systems task is to imitate the cellular architecture’s structure and nervous systems functionality to imitate the automata intelligent behavior (the so-called bionic direction), as much as in the individual functions implementation inherent in “strong” AI. One of these functions is the

ability to make reliable deductive conclusions based on automatic generalization of input data.

Within the machine learning paradigm, “strong” AI desired function is realized through reliable new samples classification not included in the training samples. Although the term “classification” is poorly applicable here because classification is a division (taxonomy) and assignment to a given thing. In our case, the desired AI function implements automatic synthesis of a new sample description based on an input previously unobserved signals tuple. “Not previously observed” refers to both absolute values at the information system input and their combinations.

Modern ideas about AI systems ability (not necessarily based on machine learning) to generalize adhere to the point of view according to which the classification reliability (or more precisely the new images or concepts synthesis) is determined [2]-[4]:

- 1) Statistical precedent data representativeness;
- 2) Architecture (parameter values structural complexity and adequacy) of the AI system;
- 3) Recognizable patterns topology in the feature space (if we are talking about the images metric representation).

The first and second factors are related and poorly controlled by the developer.

Most modern artificial neural network (NN) architectures make it possible to generalize new input samples with high but finite reliability due to the fact that information-intensive multilayer networks make it possible to approximate a very large number of training samples examples (if you do not take into account the convolution technology, which allows you to significantly increase the reliability due to recognition invariants). Architectures such as FixResNeXt-101 (2020), Vit-H/14 (2021), RepVGG-B1 (2021), etc., combined with the ability to decompose learning procedures (DeepLearning variations, displace classical learning algorithms) and significant computing power use determine in many ways, an extensive path to the success such cloud projects as ChatGPT-X, Claude-X, Geminu, YandexGPT, Kandinsky 2.1 a generative architectures wide range, etc. [1] [5]-[7].

But any training sample is finite which means it has a limited examples variety and does not cover the entire precedents set which can be used to reliably use the trained model [8] [9]. There are different approaches to solving the problem. Even though for machine learning algorithms theorems have been proven about the direct (in general case nonlinear) recognition reliability dependence on the training sample size and the information network capacity [10]-[12], a simple increase in the training data volume is often not enough for reliable examples recognition that are not included in the training data sample.

There are different approaches to solving the problem. One of them within the intensive direction framework is based on heuristic not fully explored structu-

rally redundant increasing methods of various NN architectures functioning reliability and which are combined with the extensive direction.

2. Methodology

In this section, different artificial neural network (NN) architectures approaches are using to solving the generalization and developing prototypes problem of artificial neural networks input reactions as features multidimensional nonlinear interpolation result of recognized samples using a network architecture with artificially introduced structural redundancy.

2.1. Prototypes Generalization and Development as Multidimensional Nonlinear Interpolation in a Neural Network Computing Basis

In recent years, in addition to classical neural networks architectures that solve classification problems, optimization and search, new architectures have appeared such as autoencoders, generative adversarial neural networks but their design and operation principle is basically similar to networks designed for pattern recognition (Figure 1).

A common feature characteristic to all connectionist paradigms is the inability to construct a probabilistic or analytical distribution dependence of neural network weighting coefficients values on the training vectors values components at the input. Parametrically, a neural network is a “black box” and remains so for any even the most modern, accurate and reliable deep learning architectures. There is no way to trace in detail and step by step how the network output values were calculated.

However, the experimentally established networks structural features, their design principles and training are known to ensure the stability and such effects reproducibility as:

- generalization in multilayer feed-forward networks [14];
- associated prototypes development in dynamic networks with feedback [15].

Such features have one thing in common to initialize them, it is necessary to operate with the structural network redundancy, i.e. neurons number and location (Figure 2).

The neural network training process is represented as an approximation problem (and functioning as an interpolation problem in the narrow sense) [4]. In this case, the network itself acts as one nonlinear operator (Figure 2). This point of view allows us to consider generalization as the input data nonlinear interpolation result [16]. The network carries out correct interpolation mainly due to continuity (differentiability), neuron activation functions nonlinearity as well as through certain mathematical procedures which for multilayer feed-forward networks ensures overall output function continuity.

However, the overall output function continuity and the trained network functioning reliability are different concepts. Continuity only provides validity when it eliminates network overfitting. Those, when the training samples are similar

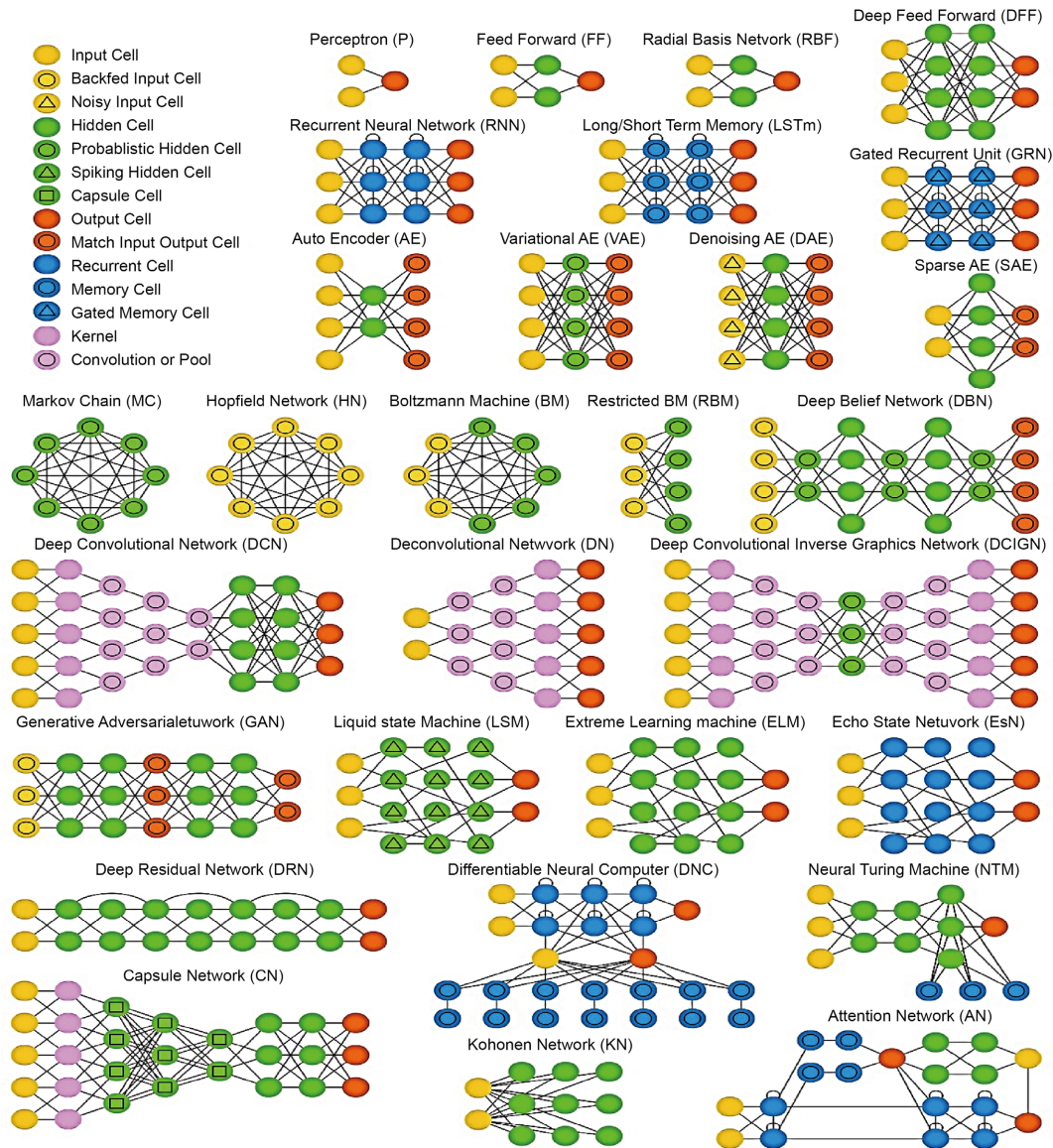


Figure 1. All NN architectures variety are “black boxes” varieties [13].

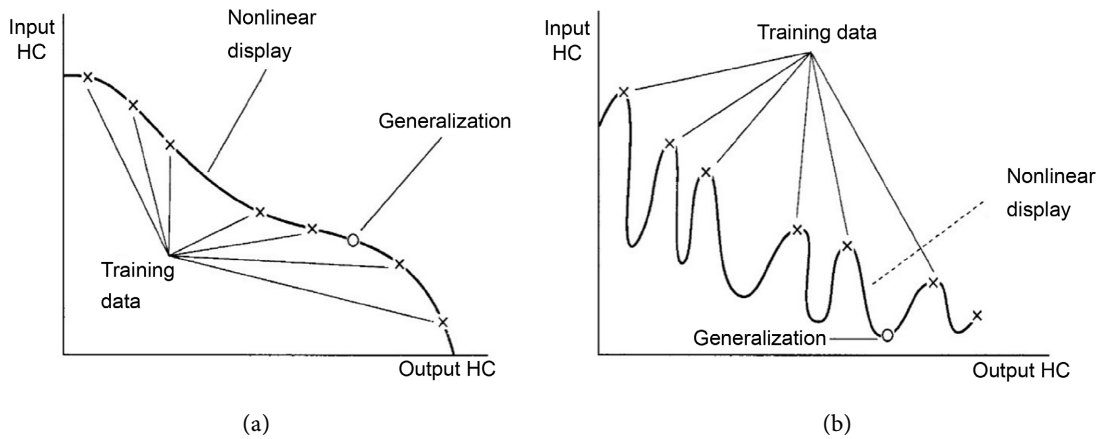


Figure 2. Reliable (a) and bad (b) generalization in a neural network.

and to distinguish them, it is necessary to increase the approximation accuracy or when there are too many examples and it is necessary to increase layers number, neurons in each layer, then according to the training results the network (with certain training algorithms) may be overtrained (the training sample examples are restored accurately but practically does not recognize similar samples). This effect is observed when training in additive or (even worse) multiplicative noise which is present in training examples but is not the simulated mapping function characteristic itself.

The interpolation problem is one of the main numerical methods problems. With its help approximate analytical representation problems, differentiation, the table-specified functions integration or functions with complex analytical representation are solved.

A broader concept the approximation uses methods for calculating approximate function values and its derivatives in the case when the certain fixed points function values are known. These points set is sometimes given to us by external circumstances, in machine learning case—by the training sample parameterization examples. The approximation concepts hierarchy is illustrated in **Figure 3**.

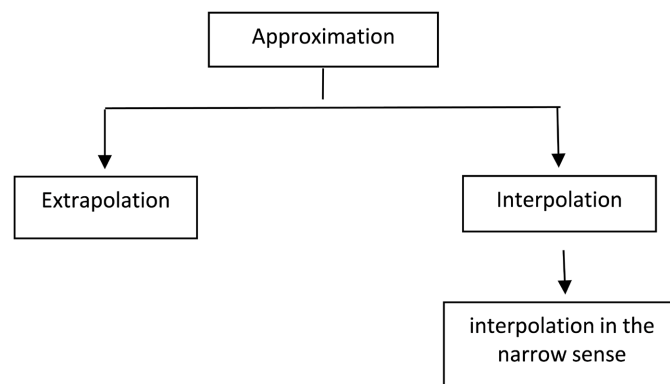


Figure 3. Interpolation in the narrow sense and extrapolation is the samples recognition not included in the training set.

Large number of approximation methods presence is caused by the historical theory and practice development of solving applied problems. Many methods arose as previous variants differing from them in the notation form, changing the calculations order which had the reducing goal of rounding errors influence in calculations.

One of the special approximation cases, the interpolation, implies that if the desired function $y(x)$ is specified by the training sample in the table form, i.e. in terms of experiment on a grid $\{x_n, n = 0, 1, \dots\}$ at the nodes of which the values $y_n = y(x_n)$ are known, then the task is to construct a function that restores the values of $y(x)$ at an arbitrary point x . At the same time, we must require a fairly simple behavior of $y(x)$: the function should not have “bursts” between neighboring nodes. Mathematically, it means that $y(x)$ should have a sufficient number of higher derivatives that are not too large in magnitude. Let us choose li-

nearly independent functions system $\{f_m(x), m = 0, 1, \dots\}$. Such linear functions combination is called a generalized polynomial $\Phi(x)$. Approximation of $y(x)$ by a generalized polynomial:

$$y(x) \approx \Phi_N(x) \equiv \sum_{m=0}^N c_m f_m(x) \tag{1}$$

where:

c_m —coefficients chosen with the condition that the generalized polynomial $\Phi_N(x)$, containing $(N + 1)$ coefficients accurately conveys the tabulated values function at the $(N+1)$ th node:

$$\sum_{m=0}^N c_m f_m(x) = y_n, \quad 0 \leq n \leq N \tag{2}$$

Such approximation method is called interpolation, and the coefficients c_m are found from the linear system solution (2). For it to be solvable it is necessary that:

$$\det[f_m(x_n)] \neq 0 \tag{3}$$

However, condition (3) is not necessary in neural network approximation case in a multilayer architecture.

Multidimensional interpolation means the constructing a function passing through points specified not on a plane but in space (three-, four-dimensional, etc.). Thus, instead of dependence (2) which we approximated with the function $f_m(x)$, we should find several coordinates function $f_m(x_1, x_2, x_3, \dots, x_z)$. If the nodes are in a regular grid form, for example, in the two-dimensional case, in a rectangular grid form then there are no fundamental problems with multidimensional interpolation construction. In the analysis under consideration, interpolation in the narrow sense of the word is of interest, which implies that the desired the function values $f_m(x_1, x_2, x_3, \dots, x_z)$ do not coincide with the nodes and lie in the intervals:

$$\begin{bmatrix} [x_1^1, x_1^2] & [x_1^2, x_1^3] & [x_1^3, x_1^4] & \dots & [x_1^{i-1}, x_1^i] & \dots & [x_1^{N-1}, x_1^N] \\ [x_2^1, x_2^2] & [x_2^2, x_2^3] & [x_2^3, x_2^4] & \dots & [x_2^{i-1}, x_2^i] & \dots & [x_2^{N-1}, x_2^N] \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ [x_j^1, x_j^2] & [x_j^2, x_j^3] & [x_j^3, x_j^4] & \dots & [x_j^{i-1}, x_j^i] & \dots & [x_j^{N-1}, x_j^N] \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ [x_M^1, x_M^2] & [x_M^2, x_M^3] & [x_M^3, x_M^4] & \dots & [x_M^{i-1}, x_M^i] & \dots & [x_M^{N-1}, x_M^N] \end{bmatrix} \tag{4}$$

where:

$(N-1)$ —is the intervals number between data points for each of the M variables.

Solving the task problems of multidimensional interpolation in the narrow sense of the word arise when:

- 1) The original sample is not tied to an equidistant grid, i.e. represents disparate data series;

2) The desired multidimensional function arguments $f_m(x_1, x_2, x_3, \dots, x_z)$ are not independent.

2.2. Prototype Development in Dynamic Networks with Feedback

If in dynamic networks with “all to all” feedback type, the state of the i -th neuron in the system ($i = 1, 2, \dots, N$) is described by a continuous variable $\sigma_i(t)$, the pairwise neurons interaction is described by a matrix of weights with elements W_{ij} then the Lagrangian NS interaction which is the independent variables functional σ and W , is described by the expression [9]:

$$L = -E = -\frac{1}{2} \lambda \sum_{ij} W_{ij} \sigma_i \sigma_j \quad (5)$$

where:

$\lambda > 0$ is the interaction parameter.

Dynamic equations of NS for σ and W_{ij} :

$$\frac{1}{\gamma_1} \frac{d\sigma_i}{dt} - f_i(\sigma) = \frac{\partial L}{\partial \sigma_i} = \lambda \sum_j W_{ij} \sigma_j \quad (6)$$

$$\frac{1}{\gamma_2} \frac{dW_{ij}}{dt} - F_{ij}(W) = \frac{\partial L}{\partial W_{ij}} = \frac{1}{2} \lambda \sigma_i \sigma_j \quad (7)$$

where f and F are coefficients that prevent an unlimited increase in the modules σ and W .

The system dynamics (6), (7) significantly depends on its history. The training consists in the fact that in (6) it is applied an external influence acting during time t_s , as a result of which the vector σ takes on a stationary value φ^s corresponding to the “image” with components $\pm\sigma_0$. As a training result, the matrix W_{ij} in accordance with (7), changes by the amount $\Delta W_{ij} = \frac{1}{2} \gamma_2 \lambda t_s \varphi_i^s \varphi_j^s$. In this case, we assume that the training time t_s is significantly longer than the relaxation vector time σ to its stationary value φ^s . The training procedure can be repeated many times using images φ^s , $s = 1, 2, \dots, s_0$. Assuming for simplicity that before the training start $V = 0$, after the procedure end we obtain

$$W_{ij} = \sum_{s=1}^{s_0} \mu^s \varphi_i^s \varphi_j^s \quad (8)$$

where the coefficients μ^s depend on the duration of training.

In a discrete model the state $q = \varphi^k$ realizes a stable local system energy minimum if the number of training sample examples is small (9):

$$A \leq 0.14C, \quad (9)$$

where

A is the memorized images number, C is the neurons number in the network.

The situation in which $A > 0.14C$ is of interest, i.e. when there are an excess neurons number, it is in this case that the input reactions prototype is developed. If an images group $\Phi^k = (\varphi_1^k, \dots, \varphi_r^k)$, $k = 1, \dots, N$, of the training sample R is obtained with small random distortions $\Delta^k = (\delta_1^k, \dots, \delta_r^k)$ of some vector

$\Phi^o = (\varphi_1^o, \dots, \varphi_r^o)$, then the variation Δ of the vector Φ^o leads to a change in the network energy function corresponding to this vector by the amount

$$\Delta E = -\frac{1}{2} \left(\sum_{k=1}^N ((\Phi^o + \Delta^k)(\Phi^o + \Delta))^2 - \sum_{k=1}^N ((\Phi^o + \Delta^k)\Phi^o)^2 \right) \quad (10)$$

When $\{(-\Phi^o \Delta^k) \leq r, N \gg 1\} \Rightarrow \Delta E \geq 0$, therefore, the initial vector Φ^o corresponds to the minimum NS energy. This means that when the vector $P^o = (Z^o X^o K^o Y^o)$ is applied to the NS Hopfield input is stabilized in the state $(Z^o X^o K^o Y^o)$, in which Y^o —is the desired result in the previously unobserved situation of $Z^o X^o$ and K^o .

2.3. Generalization in Fully Connected Multilayer Neural Networks

In the middle of the 20th century, F. Rosenblatt showed [14] that thanks to the redundant structure the multilayer neural network classifier allows one to move from a selective reaction to one image to a generalized reaction to a «similar» image, the characteristics of which may be completely different from the previous one. At the same time, the known basic transformation recognition methods classification included:

- a) Analytically descriptive method which consists in reducing the classes alphabet and the features dictionary to a simple canonical description which is invariant under the transformations under consideration;
- b) Image conversion method. It is mainly used to transform images based on their topology in a given n-dimensional space and has the serious disadvantage in requiring a huge neurons number in the network;
- c) The generalization method by dominance which consists in the fact that a change in the characteristics of X input image in which a certain characteristics part remains the same as the original one will not cause a change in the reaction Y of the neural network to the original image. In other words, many additional features introduction the values of which are common to “similar” input images X will facilitate the rapid formation of a more reliable class alphabet with a training sample fixed size. Actual it is a classic multidimensional nonlinear interpolation.

In the research process (on input images of similar ones, mainly according to one-to-one transformations group), the fourth development direction in generalizing multilayer neural networks characteristics was established, the generalization by similarity, the recognition rule for which states: «Any neural network fragment that reacts to cause X_1 will also react to all mappings $\Xi(X_1)$ of a given cause in a transformations group \mathfrak{R} with probability p ». Prototype, generalization, similarity are similar concepts that can be combined in multilayer neural networks architecture (including convolutional) to increase the recognition reliability of the desired vector characteristics values that previously extrapolated (interpolated) using the multidimensional linear extrapolation method (Figure 4).

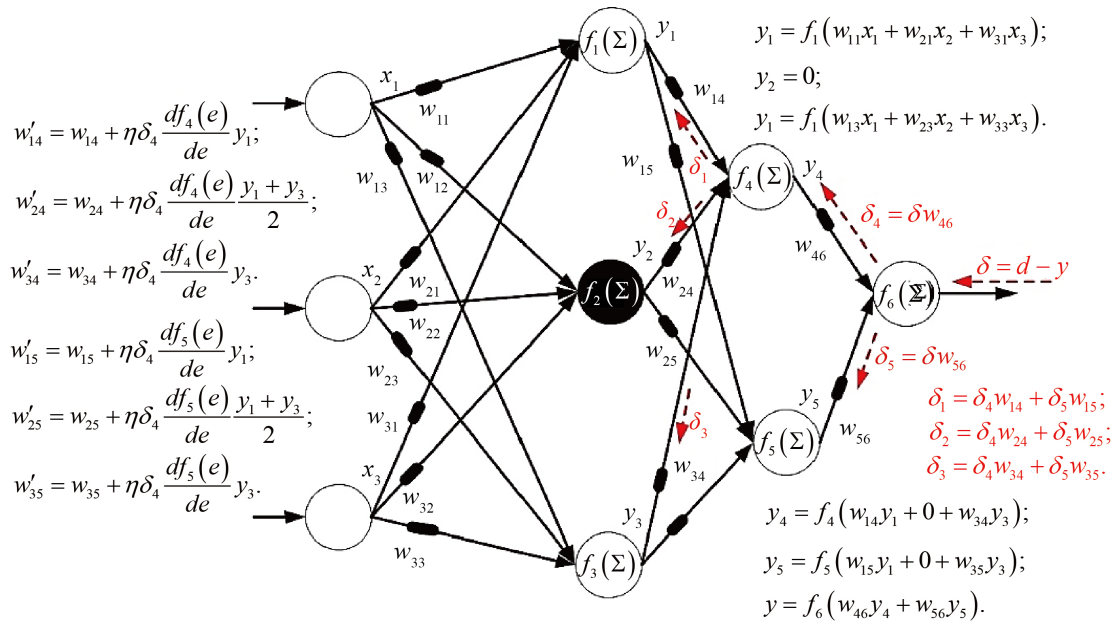


Figure 4. Forming principle explanation of the generalization effect by predominance due to redundancy in neurons number in a multilayer neural network hidden layers.

To do this, it is necessary to introduce artificial redundancy into multilayer neural networks architecture, i.e. introduce “false” neurons on which by analogy with dynamic networks the direct signals “potential” will accumulate during the learning process, but the neurons themselves do not participate in learning and minimizing the error functional at the neural network output.

3. Conclusion

In many technical and physical problems, a strong mathematical solutions local dependence on changes in initial quantities is manifested. The solutions proximity is of great importance and can be effectively used when solving extrapolation (interpolation) problems of the characteristics of developing sample (or control actions) in conditions of insufficient information. The choosing the optimal model task and evaluating the results obtained becomes significantly more complicated with any interpolation methods generalization to multidimensional space and its solution is impossible without the modern computer technologies and software use. And interpolation nodes nonlinearity and non-equidistance, in relation to a neural network system under the finite training sample conditions causes high probability ensuring problem correct input samples recognition that were not previously observed, *i.e.* geometrically parameterized outside the interpolation nodes. A solution to the problem can be NN architectures with artificially introduced structural redundancy which makes it possible to increase the accuracy and reliability of recognition due to multidimensional nonlinear interpolation of recognized samples characteristics reflected in the generalization and prototypes development when feeding features that are not included in the training set as input.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Hendrycks, D. and Gimpel, K. (2017) A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks. *Conference Track Proceedings*. <https://openreview.net/forum?id=Hkg4T19xl>
- [2] Zhuravlev, D.V. and Smolin, V.S. (2023) The Neural Network Revolution of Artificial Intelligence and It's Development Options. *Proceedings of the 6th International Conference "Futurity Designing. Digital Reality Problems"*. <https://doi.org/10.20948/future-2023-16>
- [3] Vedyakhin, A., *et al.* (2021) Strong Artificial Intelligence: On the Approaches to Super-Intelligence. *Intellectual Literature*, 232.
- [4] Khaikin, S. (2006) *Neural Networks: A Complete Course*. 2nd Edition, Williams, 1104.
- [5] Zhang, C.Y., Bengio, S., Hardt, M., Recht, B. and Vinyals, O. (2017) Understanding Deep Learning Requires Rethinking Generalization. *5th International Conference on Learning Representations*, Toulon, 24-26 April 2017. <https://openreview.net/pdf?id=Sy8gdB9xx>
- [6] Vincent, J. (2022) ChatGPT Proves AI Is Finally Mainstream and Things Are Only Going to Get Weirder. <https://www.theverge.com/2022/12/8/23499728/aicapability-accessibility-chatgpt-stable-diffusion-commercialization>
- [7] Rose, J. (2022) Inside Midjourney, the Generative Art AI That Rivals DALL-E. <https://www.vice.com/en/article/wxn5wn/inside-midjourney-the-generativeart-ai-t-hat-rivals-dall-e>
- [8] Nazarov, A.V., Burlutsky, S.G. and Matasov, Yu.F. (2023) Neurometric Approach to Solving the Problem of Interpolation in Artificial Intelligence Systems Based on Machine Learning Technologies. *Collection of Articles. Reports of the IV International Scientific Conference "Aerospace Instrumentation and Operational Technologies"*, Part 2, St. Petersburg, 72-77.
- [9] Bose, A., Das, A., Dandi, Y. and Rai, P. (2021) NeurInt-Learning Interpolation by Neural ODE. Spotlight at the NeurIPS 2021 Workshop on the Symbiosis of Deep Learning and Differential Equations (DLDE).
- [10] Farago, A. and Lugosi, G. (n.d.) Strong Universal Consistency of Neural Network Classifiers. *Proceedings. IEEE International Symposium on Information Theory*, San Antonio, 17-22 January 1993, 431. <https://doi.org/10.1109/isit.1993.748747>
- [11] Rajput, D., Wang, W. and Chen, C. (2023) Evaluation of a Decided Sample Size in Machine Learning Applications. *BMC Bioinformatics*, **24**, Article No. 48. <https://doi.org/10.1186/s12859-023-05156-9>
- [12] Balki, I., Amirabadi, A., Levman, J., Martel, A.L., Emersic, Z., Meden, B., *et al.* (2019) Sample-Size Determination Methodologies for Machine Learning in Medical Imaging Research: A Systematic Review. *Canadian Association of Radiologists Journal*, **70**, 344-353. <https://doi.org/10.1016/j.carj.2019.06.002>
- [13] Leijnen, S. and Veen, F.V. (2020) The Neural Network Zoo. *Proceedings*, **47**, Article No. 9. <https://doi.org/10.3390/proceedings47010009>

- [14] Rosenblatt, F. (1965) Principles of Neurodynamics. In: *Perceptrons and the Theory of Brain Mechanisms*, Mir, 480.
- [15] Vedenov, A.A. (1988) Modeling of Thinking Elements. Nauka, 158.
- [16] Wieland, A. and Leighton, R. (1987) Geometric Analysis of Neural Network Capabilities. *1st IEEE International Conference on Neural Networks*, San Diego, Vol. 3, 85-392.