DEVELOPMENT OF A NEURAL NETWORK MODEL OF AN INTELLIGENT MONITORING AGENT BASED ON A RECURRENT NEURAL NETWORK WITH A LONG CHAIN OF SHORT-TERM MEMORY ELEMENTS

*Osamah Adil Raheem (1), I. A. Aksenov (2), Yu. V. Redkin (3), A. V. Gorshkov (4), S. A. Sorokin (4), I. V. Atlasov (5), O. Ja. Kravets (6)**

(1) University of Wasit, Iraq; (2) Vladimir State University named after Alexander and Nikolay Stoletovs, Vladimir; (3) Admiral Ushakov Maritime State University, Novorossiysk; (4) Research Institute of Computing Complexes named after M.A. Kartsev, Moscow; (5) Moscow University of Ministry of Internal Affairs of Russian Federation named by V. Ja. Kikot, Moscow; (6) Voronezh State Technical University
(1) Iraq; (2-6) Russian Federation

* Corresponding Author, e-mail: csit@bk.ru

Abstract: The article continues to review the approach to designing the architecture of a distributed information monitoring system and quality management of communication services provided by the infrastructures of the Internet of Things and the industrial Internet of Things, based on solutions that support machine-to-machine and human-machine interaction. The development of a neural network model of an intelligent monitoring agent based on a recurrent neural network with a long chain of short-term memory elements is proposed. The matrix structure of the LSTM network memory cell is proposed, which takes into account the spatio-temporal correlation of load parameters associated with the time lag of its propagation and is a matrix of connectivity of LSTM network hyperparameters and accumulated values of load parameters of monitoring nodes in the vicinity of a controlled monitoring node, taking into account the characteristics of the time series of propagation of load in stationarity moments..

Key words: distributed information monitoring system, quality management of communication services, network traffic forecasting, neural networks, decentralized control.

1. INTRODUCTION

The article continues [1] the approach development to designing the architecture of a distributed information monitoring system and quality management of communication services provided by the infrastructures of the Internet of Things and the industrial Internet of Things, based on solutions that support machine-to-machine and human-machine interaction.

A review and analysis of previous studies are also provided in [1].

2. MODIFICATION OF A DECENTRALIZED MONITORING SYSTEM SCHEME BASED ON A VARIETY OF INTELLIGENT AGENTS

The choice of a decentralized monitoring system (MS) scheme based on a set of intelligent monitoring agents (IMA) requires modification of the generalized MS scheme with centralized management [1] by converting it into a peer-to-peer interaction scheme of a subset of IMA. The result of this modification is shown in Figure 1.

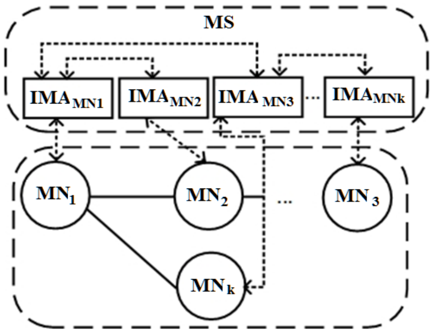

Figure 1. Generalized structure of a decentralized type monitoring system based on a subset of intelligent monitoring agents; MN – monitoring node

It follows from Figure 1 that a set of IMA's is placed on a set of monitoring nodes (MN), forming an interacting pair $MN_k \leftrightarrow IMA_{MN_k}$. In this case, MN_k is part of the environment IMA_{MN_k} with which it interacts, obtaining a vector of load parameters $\bar{X}_k(t^j_{sb})$ of the k-th MN at a time t^j_{sb}. Another part of the environment with which IMA_{MN_k} interacts is a subset of IMA_{MN}, formed in accordance with the weighted digraph of the propagation of the load parameter vector $G(MO,E), \langle \{MN_i, MN_k\}, \prec \rangle \in E, MN \in MO \cdot$

Thus, in contrast to the traditional multi-agent systems (MAC) Grid scheme of agent interaction (interaction of the type "each with each"), the study proposes to reduce the number of IMA interaction links. That is, as a source of analysis, IMA_{MN_k} uses only the values of the parameter vector $\bar{X}_k(t^j_{sb})$ and the values of the parameter vectors formed by MN_{k-1} and MN_{k+1}, which are associated with the MN_k time lag of their propagation according to the type of digraph G(MO,E). It can be assumed that the subset $\{MN_{k-1}, MN_{k+1}\}$ is the source of retrospective values of the parameter vectors $\bar{X}_{k-1}(t^j_{sb})$ and $\bar{X}_{k+1}(t^j_{sb})$, and, in turn, MN_k is included in the subset of sources of retrospective parameters for other neighboring MN.

In graph theory, such a generated subgraph of the digraph G(MO,E), consisting of all vertices conjugate to MN_k and all edges connecting two such vertices, is called the Neighborhood of the vertex MN_k and is denoted $N_{G(MO,E)}(MN_k)$ [2].

3. SUBSTANTIATION OF A VARIANT OF THE NEURAL NETWORK MODEL OF AN INTELLIGENT MONITORING AGENT

The scheme of MS organization proposed in Figure 1 with a decentralized structure based on a set $\left\{IMA_{MN_1},..,IMA_{MN_k}\right\}$ suggests that the most important component of IMA is a learning block that provides intelligent analysis of the values of vectors of load parameters of parameters $\bar{X}_{k-1}(t_{sb}^{j})$, $\bar{X}_{k}(t_{sb}^{j})$, $\bar{X}_{k+1}(t_{sb}^{j})$. Since it is required to perform a retrospective analysis of a time series, therefore, according to the general classification of neural network models [3], a class of recurrent neural networks (RNN) [4].

The RNN structure takes into account the results of converting the output vector of parameters at the previous stage to process the input vector of parameters at the next stage of network operation (Figure 2). RNN should be considered as several copies of the L_t layer, activation of which occurs at a time that transmits the parameters of the subsequent copy to the L_{t+1} layer.

For each time point t, the values $y^{<t>}$ of the L_t layer output and its characteristic variable $a^{<t>}$ are determined by expressions (4)-(5):

$$a^{<t>}=g_1(W_{aa}a^{<t-1>}+W_{ax}x^{<t>}+b_a) \tag{4}$$

$$y^{<t>}=g_2(W_{ya}a^{<t>}+b_y), \tag{5}$$

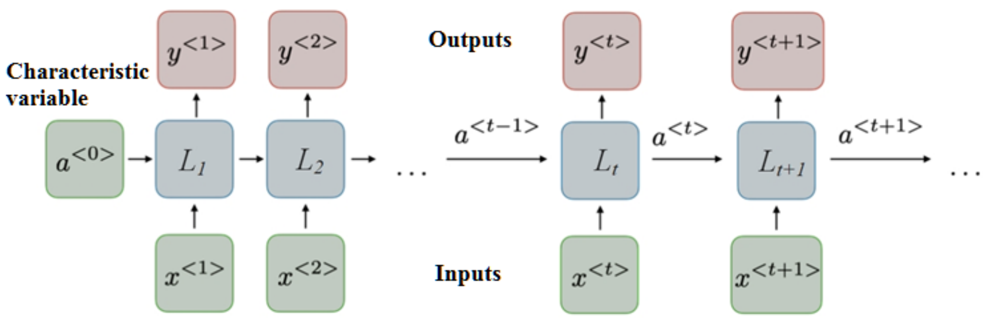

Figure 2. Generalized structure of a recurrent neural network

where g_1 and g_2 are activation functions – usually a logistic sigmoid σ or hyperbolic tangent - tanh), and W_{aa}, W_{ax} are the weighting coefficients of the corresponding activation and input parameters, b_a is the displacement vector of these parameters.

Taking into account (4)-(5), the layer structure is shown in Figure 3.

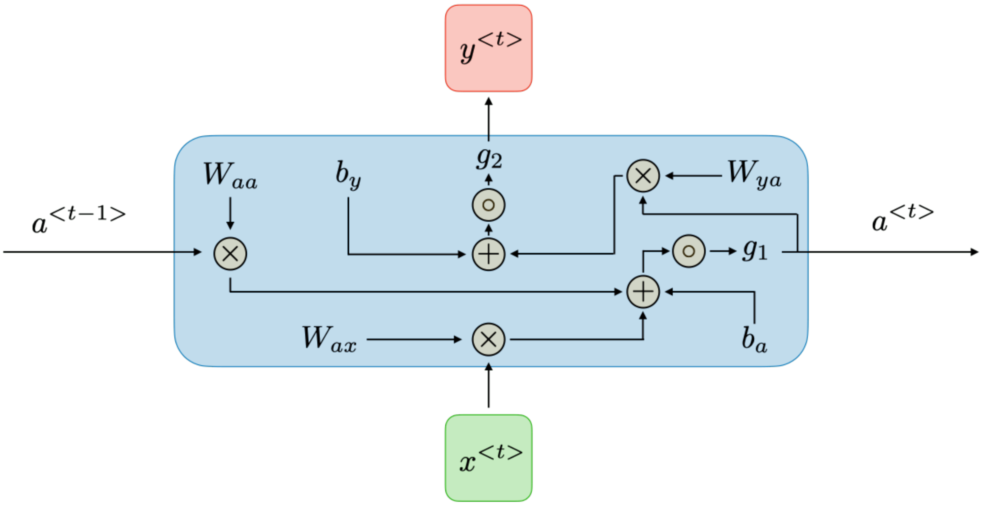

Figure 3. Generalized structure of the recurrent neural network L_t layer

The main problem of the impossibility of using RNN classical architecture in long-term forecasting tasks is the problem of vanishing or exploding gradients during the execution of the error back propagation (BPE) task by gradient descent methods, which are most applicable for RNN networks. Since the activation functions in the BPE problem are linear and at the same time represent the ratio of derivatives of the error value over the period T, even small changes in the values of the weighting coefficients W_{aa} and W_{ax} will lead to an exponential increase or decrease in the gradient values, which is explained by the presence of bifurcation points [5]. Accordingly, this leads to the difficulty of obtaining dependencies over long time intervals – obtaining reliable context from past network states.

4. THE PROBLEM OF GETTING CONTENT

In part, the problem of obtaining context in long-term T intervals is solved by special architectures of classical RNNs, for example, such as Jordan and Elman RNNs [6]. A feature of these RNN architectures is the introduction of feedbacks from network outputs (Jordan RNN) or from outputs of hidden layer neurons (Elman RNN).

Since it is necessary to minimize the delay time of the control action t^j_{delay}, and the network learning time is one of the components t^j_{proc} in its composition, to eliminate this shortcoming, the study suggests using the LSTM network – a long chain of short-term memory elements, which is a type of RNN architecture. The main difference between LSTM and RNN with feedbacks is the possibility of their use in predicting time series, in particular, in cases where the values of the required parameters are separated by time lags with indeterminate duration and boundaries. This is achieved by the fact that the LSTM architecture, along with the L_t layer elements discussed in Figures 3 and 4, contains a memory element $Cell_{LSTM}$ – a recurrent module with a special structure

(Figure 5) [7], which has the ability to memorize the values of the parameters of temporal dependencies $x^{<t>}$ and $a^{<t>}$. By combining layers of $Cell_{LSTM}$-cells that store the values of temporal dependencies, both on short (t±1) and long (t±n) time intervals, it is possible to increase both the depth of the forecast due to the accumulation of retrospective data, and its accuracy due to a retrospective analysis of the influence of these parameters on the output value. Figure 4 shows that the $Cell_{LSTM}$-cell structure, in addition to the characteristic variable $a^{<t>}$, contains a special variable $c^{<t>}$ called the $Cell_{LSTM}$-state. This variable displays the temporal state $Cell_{LSTM}$ – information about the data stored, accumulated or deleted at time t-1 in the data cell. The variable $c^{<t>}$ is controlled by special filter functions, called "gate" in the LSTM architecture. The classical approach to development $Cell_{LSTM}$ defines a "gate" of four types: input gate, output gate, forget gate, candidate cell state gate. Expressions (6)-(9) define the basic functions for gates:

$$i_t=\sigma(W_{xi}x_t+W_{ai}a_{t-1}+b_i) \quad (6)$$

$$o_t=\sigma(W_{xo}x_t+W_{ao}a_{t-1}+b_o) \quad (7)$$

$$f_t=\sigma(W_{xf}x_t+W_{af}a_{t-1}+b_f) \quad (8)$$

$$c'_t=\tanh(W_{xc}x_t+W_{ac}a_{t-1}+b_{c'}) \quad (9)$$

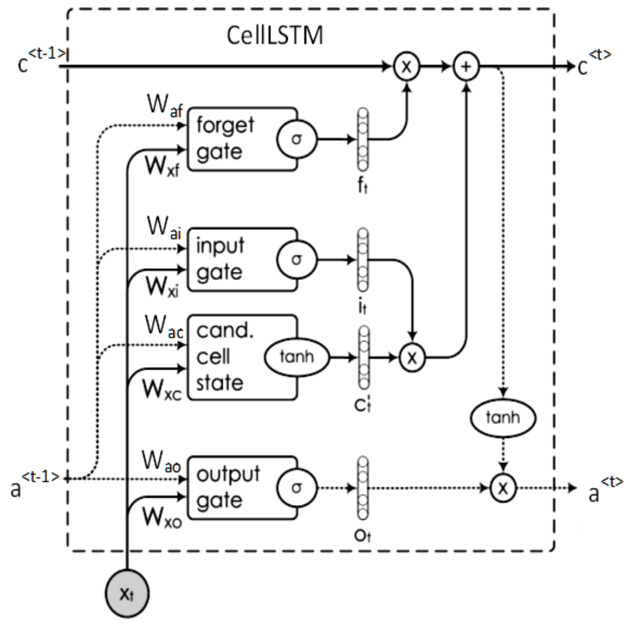

Figure 4. Generalized structure of the LSTM network memory $Cell_{LSTM}$-cell

It also follows from Figure 4 that the current state $c^{<t>}$ depends on whether the previous state $c^{<t-1>}$ will be replaced by a new "candidate" state (10):

$$c^{<t>}=f_t \otimes c^{<t-1>}+i_t \otimes c'^{<t>} \quad (10)$$

In addition, the value of the characteristic variable $a^{<t>}$ also depends on the current state $c^{<t>}$ (11):

$$a^{<t>}=o_t \otimes \tanh(c^{<t>}) \quad (11)$$

The considered structure of $Cell_{LSTM}$ is basic and in the process of improving the LSTM architecture has been modified for applicability in specific conditions. For example, in conditions when the value of the output "gateway" is zero (the "gateway" is closed), the behavior of the LSTM network will not depend on the $Cell_{LSTM}$-state (degeneration into an RNN structure). To eliminate this, the cell state coefficient, implemented by matrix multiplication, is introduced into the function of the basic gateways.

Interesting applications of similar control mechanisms for components of distributed systems are presented in [8, 9].

5. CONCLUSION

The process of designing the structure of the neural network model, which is the basis of the training unit of the intelligent monitoring agent, has been completed. A choice is made of the structure of a variant of a recurrent neural network, which provides the possibility of solving the problem of predicting the values of a time series of load parameters on monitoring nodes located in the vicinity of a controlled monitoring node and associated with a time lag of the propagation of moments of unsteadiness of this load.

A class of recurrent networks with long short-term memory (LSTM networks) has been reasonably chosen as the basis of the neural network model. The matrix structure of the LSTM network memory cell is proposed, which takes into account the spatial-temporal correlation of load parameters associated with the time lag of its propagation. The matrix is a matrix of connectivity of the hyperparameters of the LSTM network and the accumulated values of the load parameters of monitoring nodes in the vicinity of the monitored monitoring node, taking into account the characteristics of the time series of propagation of the moments of unsteadiness of the load.

REFERENCES

[1] Kravets O.Ja., Atlasov D. I., Gorshkov A.V., Sidorenko E.V., Mutina E.I., Aksenov I.A., Redkin Yu.V. Designing the architecture of a distributed system for information monitoring of IoT and IIoT infrastructures traffic. *International Journal on Information Technologies and Security*, vol. 16, no. 1, 2024, pp. 49-56.

[2] Anu J., Charles D. Vertex neighborhood restricted edge achromatic sums of graphs. *Discrete Mathematics Algorithms and Applications*, vol. 15, no. 08, 2023, art. no. 2250169. DOI: 10.1142/S1793830922501695.

[3] Wang H., Lu H., Sun J., Safo S. Interpretable deep learning methods for multiview learning. *BMC Bioinformatics,* vol. 25, no. 1, 2024. DOI: 10.1186/s12859-024-05679-9.

[4] Song X., Wang X., Song Y., Zuo X., Wang Y. Hierarchical Recurrent Neural Networks for Graph Generation. *Information Sciences*, vol. 589, no. 9, 2021. DOI: 10.1016/j.ins.2021.12.073.

[5] Liu S., Ni'mah I., Menkovski V., Mocanu D., Pechenizkiy M. Efficient and effective training of sparse recurrent neural networks. *Neural Computing and Applications*, vol. 33, no. 6, 2021. DOI: 10.1007/s00521-021-05727-y.

[6] Khan A. Training Learning Weights of Elman Neural Network Using Salp Swarm Optimization Algorithm. *Procedia Computer Science*, vol. 225, no. 1, (2023). DOI: 10.1016/j.procs.2023.10.188.

[7] Bengio Y., Simard P., Frasconi P. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, vol. 5, no. 2, 1994, pp. 157–166.

[7] Bengio Y., Simard P., Frasconi P. Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, vol. 5, no. 2, 1994, pp. 157–166.

[8] Shahapurkar A., Rodd S.F. Efficient feature aware machine learning model for detecting fraudulent transaction in streaming environment. *International Journal on Information Technologies and Security*, vol. 14, no. 3, 2022, pp. 3-14.

[9] Yildirim M., Sokullu R. Deep Learning-Based Decoding for Phase Shift Keying. *International Journal on Information Technologies and Security*, vol. 13, no. 1, 2021, pp. 39-51.

Information about the authors:

Osamah Adil Raheem - PhD in Engineering Sciences, university teacher, University of Wasit, El-Cut, Muhfaza Wasit, Iraq, areas of scientific research - information technologies, computer networks, system analysis, optimization

Ilia Antonovich Aksenov – PhD in economics, associate professor of Vladimir State University named after Alexander and Nikolay Stoletovs, areas of scientific research -economic information systems

Yuriy Viktorovich Redkin – PhD, associate professor of Admiral Ushakov Maritime State University, Novorossiysk, areas of scientific research - wireless networks, digital signal processing, control and data processing systems

Alexey Vladislavovich Gorshkov – PhD in Engineering Sciences, General Director of the Scientific Research Institute of Computing Complexes named after M.A. Kartsev, Moscow, areas of scientific research – system analysis, optimization, simulation of complex objects

Sergey Aleksandrovich Sorokin – Dr. Sc. (IT), Chief Engineer of Research Institute of Computing Complexes named after M.A. Kartsev, Moscow, areas of scientific research – system analysis, optimization, simulation of complex objects

Igor Victorovich Atlasov - Dr. Sci (Phys.-Math.), professor of Moscow University of Ministry of Internal Affairs of Russian Federation named by V. Ja. Kikot, Moscow, areas of scientific research – system analysis, optimization, simulation of complex objects

Oleg Jakovlevich Kravets – Dr. Sci (IT), professor of Voronezh State Technical University, areas of scientific research - system analysis, optimization, simulation of complex objects

Manuscript received on 22 March 2024