Basic concepts of dynamic recurrent neural networks development

N. Boyko[1], P. Pobereyko[2]

[1]*Lviv Polytechnic National University; e-mail: nataliya.i.boyko@lpnu.ua*
[2] *Lviv Polytechnic National University; e-mail: sdgnatkovich@ukr.net*

Abstract. In this work formulated relevance, set out an analytical review of existing approaches to the research recurrent neural networks (RNN) and defined precondition appearance a new direction in the field neuroinformatics – reservoir computing. Shows generalized classification neural network (NN) and briefly described main types dynamics and modes RNN. Described topology, structure and features of the model NN with different nonlinear functions and with possible areas of progress. Characterized and systematized well-known learning methods RNN and conducted their classification by categories. Determined the place RNN with unsteady dynamics of other classes RNN. Deals with the main parameters and terminology, which used to describe models RNN. Briefly described practical implementation recurrent neural networks in different areas natural sciences and humanities, and outlines and systematized main deficiencies and the advantages of using different RNN.

The systematization of known recurrent neural networks and methods of their study is performed and on this basis the generalized classification of neural networks was proposed.

Key words: recurrent neural network, dynamic system, learning algorithms, reservoir computing, unsteady dynamics.

INTRODUCTION

Recurrent Neural Network (RNN) is a neural network (NN) where connections between units form a directed cycle. It is a dynamic system, its current state is

determined not only by input messages, but is dependent on the previous state of the network and due to this RNN has unlimited memory. As contrasted with neural networks they keep in memory input information about the delay for an indefinite period of time. Therefore, it is necessary to analyze the possibility of the research results of dynamic systems invoke in such related fields of science as physics, the theory of nonlinear dynamical systems, chaos theory and so on for the RNN synthesis and dealing with the issues of training.

SETTING

The typical structure of the neural network is shown on Fig. 1 below by authors. In our study the attention is focused on recurrent neural systems or, in other words, reaction-coupled networks. But the study will not be full in case there will be no short summary about the known issues of neural network.

Most researchers [2, 13] emphasize two ways of NN learning. They are divided into a network supervised learning and online unsupervised learning. NN supervised learning is hold by using such pairs as input and target vectors. The output vector is set, output of network is calculated and the result is compared with the corresponding target vector. Then in order to minimize errors weight coefficients of the input vector are changed in a certain algorithm. The vectors of learning set are

formed consequently: errors calculation and further coefficient changes for each input vector is performed until the error reaches the needed value throughout the training structure.

NN unsupervised learning does not require a target vector for its outputs. Learning set consists only of input vectors. Learning algorithm modifies network weight coefficients so that only outgoing vectors are in the output.

Also scientists [1, 5] allocate the so-called separation by way of presenting examples: providing single examples and "page" examples. NN state change occurs after each input presentation in the first method. In the second method based on preliminary analysis submission set examples are carried out simultaneously.

Also, researchers [8, 11] distinguishe NN by neuron model features, for example, neurons with different nonlinear functions.
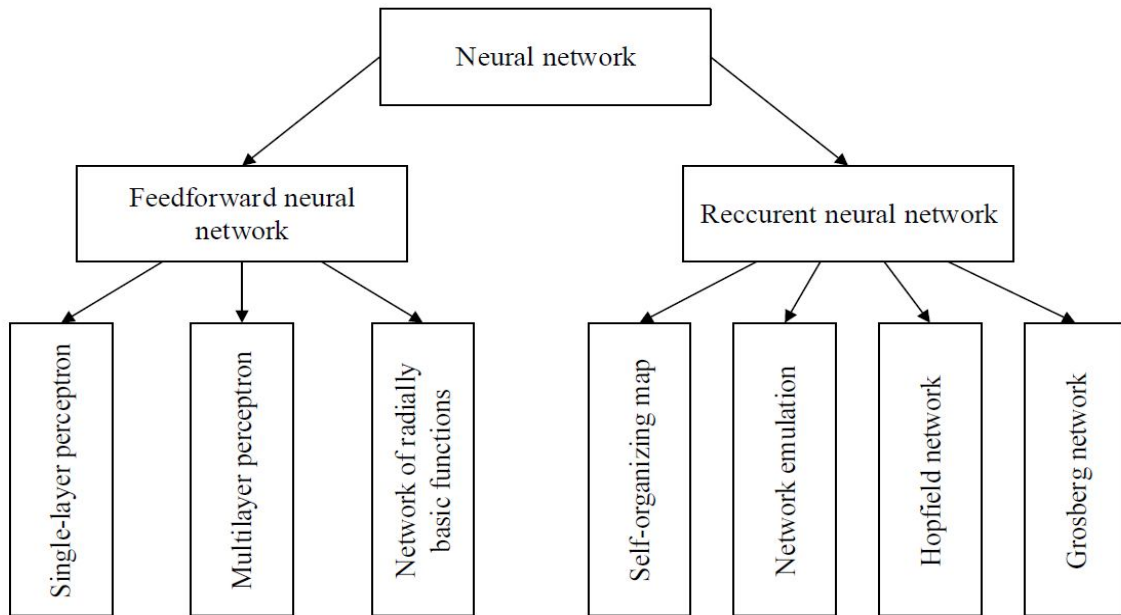
RECURRENT NEURAL NETWORK STRUCTURE

Depending on the structure RNN are divided into fully connected and local connected. Considering a fully connected RNN each neuron is linked to any other neuron network. As for RNN local connections each neuron is adjacent just to other neurons. These relationships are formed deterministically or randomly. Neurons of RNN local connections are more independent.

RNN with a small number of items has structure with one-, two- and three-dimensional latitude in a discrete state space. The state of the network is characterized by the values of the neurons outputs. Remote control systems (RCS) are use for their mathematical description. If the number of items is large, the discrete space state is switched to a continuous one and RNN structure is simulated by uninterrupted environment with distributed parameters. The dynamics of these infinite-dimensional space systems are described with the help of the following

special wave equations: partial differential equations or integral-differential equations.

Various methods are used in order to analyze the RNN structure, including method which is based on graph theory. It involves calculating various indicators, such as the reachability matrix, routes, cycles, routes, clustered index etc. It is possible to estimate the interaction of elements approximately in RNN using these indicators.

Fig.1. Recurrent neural network classification

In the space of discrete outputs values of RNN changing network conditions can be described by a system of difference or differential equations (1-2):

$$X[t+1] = g_1(X[t],U[t]); \qquad (1)$$

$$\frac{dX}{dt} = g_2(X,U)); \qquad (2)$$

where: $X = (x1, x2, ..., xn)$ is a system state; n - the dimension of the system; x1 - the value of RNN i exit; U is input effects; g_1 and g_2 - function and image accordingly that describe the system state dependence to the previous state and the input (determine the evolution of the system over time).

Depending on the type of g_1 and g_2RNN will belong to a different class of dynamical systems, which classification overlaps with the dynamic classification of RNN described above. One of the major classification feature of RNN belonging to the class of linear or nonlinear dynamic systems is linearity or non-linearity of the functions g_1 and g_2.

The analysis of RCS can be made depending on the purpose and characteristics of dynamic systems (linear / non-linear, continuous / discrete). The authors propose to use following for analysis:

1) physical principles of dynamic systems (analysis of energy conversion);

2) variational principles for remote control system;

3) formal groups and invariant theory;

4) geometric approach;

5) analysis of the existence and periodic motions stability.

THE ANALYSIS OF RNN DYNAMIC. RNN WITH DISCRETE STATE SPACE

Stability. RNN belong to dynamic systems that can be stable (unstable) in small and large. Stability in the large guarantee stability in the whole state space and stability in the small guarantee stability only at certain points.

Normally, stability in large can be identified only for a small class of RNN (e.g.: Hopfield NN). A stability in small demonstrates only the behavior of the system in equilibrium points. Therefore numerous special characteristics are develped that help to judge the stability of the system in its phase trajectory.

The main types of dynamic. There are three main dynamic types of RNN depending on stability:

• stable dynamics: RNN for a limited time goes to the stable state of equilibrium (in this mode when the input signal is changed the transition from one state to another is possible),

• oscillating dynamics: RNN condition that describes a closed circular trajectory - limiting cycle; on the one hand we can assume that the limit cycle encodes some information, on the other hand there are variations, while information is transferred to the phase relationships,

• unstable dynamics: RNN trajectories with close initial conditions diverge over time; if the growth of the system is not limited, it tends to infinity, otherwise chaotic dynamics is set when the system changes spontaneously.

It should be noted that there is chaotic dynamics where the state changes of RNN are unpredictable, and

the process is stochastic by the casual observer. In terms of physics the system "is in itself" and response weakly to input signals. In a chaotic mode the piece of information contained in RNN is quite large, but it is very difficult to get it. Researchers [15-18] study the behavior of dynamic systems in chaotic conditions in many laboratories. They offer options for this dynamic interpreting. Such type of dynamics is unstable in small but steady in large and in some cases the dynamics can be seen as vibrational.

Different dynamic RNN regimes can be considered as separate stages of solving a particular problem. For example, the author distinguish the following regimes: a chaotic regime, which corresponds to the initial solutions finding, i.e.: the choice between various hypotheses; oscillatory regime reflects the processes of switching attention from one solution to another; stable regime fulfills requirements of found solution.

RNN dynamics depends both on the structure and the properties of the components. There are cases when RNN is not stable though its elements are robust and vice versa – RNN is stable, and its elements are not stable. However, studies of NN stable regimes usually form NN with persistent elements to study the vibrational regimes with oscillating elements and, finally, to study the chaotic regimes with chaotic oscillators.

The authors propose an analytical study of simple systems (of second- and third-order) dynamics using the theory of bifurcations and catastrophes. It is the level of neuron (the synapse). These theories do not work at RNN level. Therefore, in general, for arbitrary RNN structures the only possible way of its dynamics analyze is computing simulation, which implementation is to calculate the mode performance of the network. There are the following theories on the basis of determining these parameters: the theory of probability and mathematical statistics, theory of stochastic processes, information theory, theory of deterministic chaos, fractal and sync theory.

Transient and steady regime. The authors provide RNN dynamic as the sum of two components: sustainable and transition ones. The transient component is asymptotically reduced to zero, and after that there is only steady component. Both transient and steady components (regimes) are used for image recognition:

- in transient regime after the input image is added its dynamics is superimposed on RNN dynamics – it is possible to recognize the input image on "snapshot" of the resulting dynamics later; in this case, a short-term memory property is used accounting background of input signal change (stimulus),

- in steady regime after the input image is added the system enters the attractor after a while (fixed state, limit cycle or chaotic attractor) that enable recognition of input image.

RNN with steady regimes. Researchers [3-4] offered steady in large RNN. After state change or input signal they move into a state of equilibrium to the points of fixed attractor after a while. One of the most famous members of RNN is Hopfield network (HN). It is fully connected RNN with symmetric matrix of communications.

At the macro level HN is described by the special function of energy (Lyapunov function). In the process of HN function its energy decreases until HN go into the state of corresponding attractor. HN got two main implications: associative memory and optimization. HN induced a new trend in RNN research – attraktor NN [15, 19]. Scientists [12] have been proposed various modifications, that eliminate its main disadvantages, i.e: increase memory, reduce the number of false attractors – chimeras and increase speed. In the simplest case learning algorithm varied and, as a consequence, connectivity between neurons is changed as well. In some approaches the status of each neuron is not described by a single number but vector. Also more complex nonlinear models of neurons and connections are used (terminal attractor, synergetic computer of Haken). The principle of NN remained the same as in Haken NN in all cases.

Overall, despite the large number of RNN models with stable regimes, they are not widely spread but became a convenient model for the study of HN. Finally, some researchers [16] critisized attractor idea and predictable dynamics:

- key calculations are performed by chaotic dynamics in human mind and convergence means permanent state of peace,

- RNN with point attractor can not handle dynamic image.

RNN with vibrational regimes or oscillating NN. NN oscillator (NNO) are networks which state varies with time for periodic or close to the periodic law. Sometimes they are called neural attractors of "limit cycle" type. In terms of biology at the heart of NNO hypothesis about the relationship of the perception phenomenon and oscillatory neuronal activity is laid.

All NNO models differ in the following main parameters: type of oscillator elements; type of connections; structure.

The authors offer to use detailed models of biological neurons (Hodgkin-Huxley), simplified models of biological neurons (Hindmarsh-Rose, Fittshyu-Nagumo, Izhikevicha) model of "integration and excitement", models of neurons populations (Wilson-Forged) and models of physical oscillators that have no relation to the biological neuron (generator van der Pol, oscillator phase) as oscillators. The main requirement for oscillator is that it forms oscillating activity at the output. One of the key properties of NNO elements is their frequency. In NNO all elements can be of single frequency or be divided into groups with different frequencies.

The type of connections depends on the oscillator model. If it is the model of biological neuron, the connections are made through chemical synapses, if not, then through electrical ones. Depending on the oscillator model the electrical connection can be linear or nonlinear. Connections can have features of self-supervision (internal dynamics) and plasticity (self-learning).

The most important feature that distinguishes one oscillatory NN model from others is its structure. It includes the following components:

- the topology (connectivity) – fully connected, with local connectons, with weak connections, etc .,

- the balance of excitatory and inhibiting connections (oscillators), the presence of global inhibitor,

- distribution of power relations and delays (homogeneous and heterogeneous).

The research divide ONN into two groups: the mathematical study of ONN synchronization phenomena and orthonormal basis (OB) application for building the models of olfactory, visual cortex and motor systems, as well as memory and attention.

Within the first group the problem of the neural structures of complex spatial-temporal inhomogeneous structures (flat and spiral waves, solitons) formation is considered.

As for another group the application of OB is considered for modeling the various functions of the brain. Many scientific works are related to the research of OB [7]. Some of them [9] concerns the question of the complex dynamic modes origin, the chaotic one.

RNN with chaotic regimes. Chaos is a basic form of collective neuronal activity for all processes and functions of perception. It acts as a controlled noise sources to provide continuous access to previous memorized images and storing new ones. Chaos allows the system to be always active, it provides the system from awake or come to a steady state whenever the input influence changes.

Most researchers [10] agree that regime of orderly chaos is the best one in terms of storage and information handling. On the one hand, this regime has all the advantages of chaos and on the other hand, it can be controlled.

Much attention is paid to the orderly operation regime arrays to form such clusters as chaotic synchronization. This regime is between the two extreme regimes – full order and full chaos. It allows you to use the beneficial properties of chaotic systems to meet the clustering challenges.

RNN with continuous state space. RNN with a very large number of neurons is not considered as a discrete system in space, but as a continuous media. In this case, the wave equations are used to describe it. They determine the dynamics of neural fields u (r, t), that is formed by environment. The field u (r, t) characterizes the activity of RNN at t time at point with r coordinates. For connectivity is used a special communication function w (r), which determines the dependence of communication from a distance, for example, the function of "Mexican hat".

The dynamics of RNN environment is explored using a special method, i.e.: the Turing instability analysis. The wave equation of environment has one state of balance, a quiescent state. Then equation is lineared in the vicinity of this condition and parameters that characterize the type of dynamics are calculated. At a certain selection parameters (weight function, activation function) dynamics may be unstable. Such instability is called the Turing instability. This instability leads to the formation of inhomogeneous stationary structures called dissipative structures or Turing patterns. Dissipative one means that the structure is due to dissipative processes of energy dissipation in an open system.

The type and form of Turing patterns depend primarily on the number of measurement environment. The appearance of periodic global sustainable structures that are called bump or continuous attractors is specific for one-dimensional environment. Under certain conditions such complex in environments dynamic regimes as wave transmission can arise. They may alternate with state of rest, and may exist permanently in the form of solitons. In two-dimensional environments there are spiral waves.

Researchers [18] give the hypothesis that waves can not propagate indefinitely and quickly fade in real biological systems because of the heterogeneity and symmetry break. In general, the processes that occur in the environment can be detected in discrete lattice, but this requires a corresponding dimension to a discrete version.

NN LEARNING. CLASSICAL APPROACHES TO RNN LEARNING

One of the most important properties of NN is the possibility of learning. Under learning is understood the process of setting parameters by simulating neural environment where it is placed. Type of training depends on the method of these parameters setting. To train a neural network signals that change the arbitrary parameters of NN from the environment must appear. They entail different neural response to incoming signals. RNN learning is generally classified as NP-complete problem, even for NN with one hidden layer. For a number of cases (RNN structure and simple learning case) study has polynomial complexity. The following statement is applied to RNN: "harder machine is studying – more complex is learning algorithm". As the RNN is much more complex than conventional NN direct distribution, the algorithms and their learning algorithms are more complicated than for the previous one .

Authors emphasize several different groups of approaches for RNN study:

- Attraktor RNN, where the needed attractors can be coded by means of weight coefficients setting on the basis of Heba rules.

- The usage of supervised learning algorithms based on optimization techniques of the algorithm type of inverse error propagation; the examples are following: Back Propagation Through Time, BPTT, Real-Time Recurrent Learning, RTRL, Recurrent Back Propagation, RBP, algorithms that use Kalman filtering. Since this algorithms optimization functionality is formed as the sum of errors on some time interval and same parameters are set on every step, there are problems with the convergence of methods, working time and computational cost.

- The usage of supervised learning algorithm considering RNN as part of recurrent one, feedback signals are treated as separate inputs, i.e.: contextual neurons. As a result, job training is simplified and reduced to conventional NN direct distribution learning algorithms. This type of training includes Elmana NN, Jordan NN and others.

- The unsupervised learning algorithms usage (Kohonen teaching rule, synaptic plasticity of NN impulse).

- The lack of training in the classical sense, changing system parameters (implicit learning) - setting weight coefficients of random values. In this case, the learning function is assigned to a special device - scanner that deals with the classification of RNN dynamics. The

principle is underlying the new NN paradigm, i.e.: reservoir computing.

The authors singled out another version of the training without explicit training by changing the weight coefficient of RNN. It is a new paradigm of learning where the way to communicate with the environment is represented as a change of phase portraits RNN behavior and the formation of dynamics in response to some effect (it can be compared with short-term memory in NN direct distribution in response to input impact) . The new paradigm of training is related to the new approach in the calculation - reservoir computing (RC).

RESERVOIR COMPUTING

The complexity and inefficiency of the known learning RNN algorithms make us look for new approaches and strategies for the use of their computing capabilities. The authors identify one of these approaches as reservoir computing (PB, Reservoir Computing, RC). The basic idea of these calculations is to use RNN reservoir dynamics with rich and powerful computing capabilities. The reservoir is formed randomly. His transfer to the appropriate dynamic regime of operation (status) is submitting the relevant continuous signal in input. The reservoir is formed in such way that for similar input signals this condition is similar but for different ones it is different as well. The output of reservoir is connected to such special devices as readers that solve the problem of classification, prediction, clustering, and so on for reservoir state. Reservoir integrates the dynamic of the input image in its condition and its readers recognize the input image.

RC necessity is justified by the fact that static and dynamic models of NN are limited. The use of these models makes it difficult to solve the problem of images recognition. Therefore, the problem of RNN training led to new approaches. Consequently RC has appeared combining the rich dynamics of RNN and the power of NN static training.

The authors offered several options for implementing reservoir computing:

• Backpropagation-Decorrelation, BPDC where BPDC learning algorithm is used that is simplifying of RTRL algorithm. If the whole RNN is studying in RTRL, the BPDC is constructed in such a way that only reading modules are training,

• Liquid State Machine, LSM, the reservoir is impulse RNN [17],

• Evolino – EVOlution of systems with Linear Output – reservoir is made up of special neurons that simulate continued short-term memory, the output layer is linear [20],

• Temporal RNN – biological reservoirs that are real cortical NN,

• Other reservoirs that reflect the work of purely analytical as well as really optical, biological, physical, and other quantum dynamical system.

• The authors distinguish three basic ways to form the reservoirs:

• to use general guidelines for creating 'good' reservoirs that do not depend on the type of problem and is solved by choosing topology, connectivity, coupling force, delays,

• adaptation reservoir – unsupervised learning is using examples of input: globally reservoir is formed in such way that it is in the right dynamic mode for a given input and has the necessary properties (for example, division); at the local level – self-organization of reservoir using rules of synaptic plasticity in the process of applying the input,

• reservoir supervised study, using examples of input and corresponding output data. In this case, a lot of reservoirs with different parameters is generated for a particular problem, the quality of recognition for each reservoir is assessed and the best of them is chosen. Basically reservoir computing has universal reservoir computing capabilities in terms of approximation of arbitrary nonlinear dynamical systems with fading memory. If you add back coupling from the reader to the reservoir, the possibility of system approximation with constant memory appears, particularly Turing machines. Readers are simple static training machines: weighted straight-line regression, adalina (with training in real time by the method of the smallest quadrats), perceptron, k nearest neighbors, reference vector machines, static NN.

We provide an example for understanding the idea of dynamical systems class on the basis of remote control (RC) and degree of its difficulty: in order to describe module neural networks that consists of 125 neurons in the presence of noise component and dynamic synapses, 250 correlate nonlinear stochastic RC for neurons and approximately 400 difference equations of the third order for synapses are needed.

The use of existing indicators of dynamic systems (for example, Kolmogorov entropy, fractal dimension, synchronization indices etc.) to determine the regimes of RNN even for simple cases of chaotic neural networks it is ineffective because of the specific use of dynamic regimes for solving data handling tasks. Therefore, methods of cybernetic physics and methods of new sections of nonlinear dynamics should be used for the analysis of complex RNN.

IMPLEMENTATION AREA OF RECURRENT NEURAL NETWORKS

Recurrent Neural Networks have practical use in the following areas:

1. Economics and Business: market prediction, loan default risk assessment, prediction of bankruptcy, evaluation of property value, dentifying of under- and overvalued companies, commodity and cash flow optimization, automatic reading and recognition of cheques and documents, security of plastic cards transactions.

2. Medicine: establishing diagnosis, medical images processing, patienthood monitoring, the factor analysis of treatment effectiveness, clarification of instrument readings from noise.

3. Avionics: radar signal detection, adaptive piloting of heavily damaged aircraft, unmanned drone.

4. Communication: compression of videos, fast coding-decoding, optimization of cellular networks and packets routing plan.

5. Internet: associative information search, information filtering, spam blocking, targeted advertise and marketing for e-commerce.

6. Automatic production: production process optimization, products quality control, monitoring and visualization of multidimensional data control, accidents prevention, robotics technology.

7. Politological and sociological technologies: forecasting of election results, analysis of opinion polls, predicting of ratings dynamics, identifying the important factors, objective electorate clustering, research and visualization of social population dynamics.

8. Safety and security systems: fingerprint, voice, signature identification, voice recognition in the crowd, car numbers recognition, analysis of airspace images, monitoring of information flows in computer networks and detection of intrusion and fraud.

9. Information input and processing: recognition and processing of hand-written payment, financial and accounting documents.

10. Geologic exploration: seismic data analysis, associative search techniques of mineral products, estimate of field resources.

CONCLUSION

The paper presents an analytical review of modern approaches to the study of recurrent neural networks. The systematization of known recurrent neural networks and methods of their study is performed and on this basis the generalized classification of neural networks was proposed.

REFERENCES

1. **Benderskaia E.N., Zhukova S.V. 2011.** Neural network with chaotic dynamics in problems of cluster analysis, № 7, Pp. 74–86. (in Russian)
2. **Benderskaya E.N., Zhukova S.V. 2013.** Multidisciplinary Trends in Modern Artificial Intelligence: Turing's Way // AIECM – Turing, Book Chapters: Artificial Intelligence, Evolutionary Computation and Metaheuristics, Springer, Pp. 320–343.
3. **Coombes S. 2005.** Waves, bumps, and patterns in neural field theories // Biological Cybernetics, Vol. 93, № 2, Pp. 91–108.
4. **Dasgupta B., Siegelmann H., Sontag E.D. 1995.** On the Complexity of Training Neural Networks with Continuous Activation Functions // IEEE Transactions on Neural Networks, 1995, Vol. 6, № 6, Pp. 1490–1504.
5. **Dominey P.F. 1995.** Complex sensory-motor sequence learning based on recurrent state representation and reinforcement learning // Biological Cybernetics, Vol. 73, № 3, Pp. 265–274.
6. **Dunin-Barkowski W.L., Osovets N.B. 1995.** HebbHopfield neural networks based on one-dimensional sets of neuron states // Neural Processing Letters, Vol. 2, № 5, Pp. 28–31.
7. **Feng J., Brown D. 1998.** Fixed-point attractor analysis for a class of neurodynamics // Neural Computation, Vol. 10, Pp. 189–213.
8. **Kaneko K. 2006.** Life: an introduction to complex systems biology, Berlin: Springer-Verlag, p.369.
9. **Maass W., Natschлдger T., Markram H. 2002.** Real-time computing without stable states: a new framework for neural computations based on perturbations // Neural Computation, Vol. 11, Pp. 2531–2560.
10. **Schrauwen B., Verstraeten D., Campenhout J.V. 2007.** An overview of reservoir computing theory, applications and implementations // Proc. of the 15th European Symp. on Artificial Neural Networks, Pp. 471–482**.**
11. **Fedasyuk D., Yakovyna V., Serdyuk P., Nytrebych O. 2014.** Variables state-based software usage model // Econtechmod : an international quarterly journal on economics in technology, new technologies and modelling processes, Lublin ; Rzeszow, Volum 3, nomber 2, Pp. 15-20.
12. **Benderskaia E.N., Nikitin K.V. 2011.** Modeling of neural activity in the brain, № 6-2(138), Pp. 34–40. (in Russian)
13. **Rybytska O., Vovk M. 2014.** An application of the fuzzy set theory and fuzzy logic to the problem of predicting the value of goods rests // Econtechmod : an international quarterly journal on economics in technology, new technologies and modelling processes, Lublin ; Rzeszow, Volum 3, №2, Pp. 65-69.
14. **Benderskaia E.N. 2012.** An opportunity to demonstrate some characteristics of the synchronization to identify self-organizing clusters in neural networks with chaotic dynamics, Pp. 69–73 (in Russian)
15. **Magnitskii N.A., Sidorov S.V. 2004.** New methods of chaotic dynamics, Moscow, p. 320 (in Russian)
16. **Malinetskii G.G., Potapov A.B. 2006.** Nonlinear dynamics in problems of cluster analysis, Moscow: KomKniga, p.240 (in Russian)
17. **Piterson W., Weldon E. 1976.** Error-correcting codes. Moscow: Mir publ., p.593 (in Russian)
18. **Tiukin I.Iu., Terekhov V.A. 2008. Eshbi Vilyam. 2006.** Introduction to cybernetics. Moscow, p.432. (in Russian)
19. **Khaikin S. 2006.** Neural networks – Moscow–St.-Petersburg– Kiev: ID "Vil'iams", p.1103 (in Russian)
20. **Benderskaya E.N. 2013.** Nonlinear Trends in Modern Artificial Intelligence: A New Perspective // Beyond AI: Interdisciplinary Aspects of Artificial Intelligence. Topics in Intelligent Engineering and Informatics, Springer, Vol. 4., Pp. 113–124.