УДК 004.02 DOI: 10.31673/2412-4338.2019.024452

Dibrivniy O. A. *State University of Telecommunications, Kyiv*

APPLICATION-ORIENTED USAGE OF NEURAL NETWORKS

Article dwells upon most recent trends emerging in the field of artificial neural network, such as Capsule Networks, Convolutional Neural Networks (CNN), Deep Reinforcement Learning (DRL), Lean and Augmented Learning, Supervised Model, Networks With Memory Model, Hybrid Learning Models. Comparative analysis of usage different types of neural networks architecture is provided.

Keywords: *neural networks, trends in neural networks, artificial neural networks, artificial neurons, deep learning for neural networks.*

Дібрівний О. А. *Державний університет телекомунікацій, Київ*

ПРИКЛАДНЕ ЗАСТОСУВАННЯ НЕЙРОННИХ МЕРЕЖ

В статті розглянуто класичні приклади побудови та роботи нейронних мереж, в залежності від їх класифікації. Розглянуто основні структурні елементи будь-якої нейронної мережі, та описано їхнє функціональне призначення. Розглянуто два типи нейронних мереж в залежності від типу зв'язку: мережі прямого зв'язку (FNNs) та рекурентні нейронні мережі (RNNs). *Для обраних типів мереж наведено Гаусівські функції активації прихованих ланок. Також було розглянуто основні типи навчання нейронних мереж такі як: контрольоване, неконтрольоване та підтримуване навчання. В статті основну увагу приділяється розвитку можливостей нейронних мереж в залежності від обраної архітектури та потреб які виникають перед цифровою спільнотою. Описуються основні тенденції розвитку моделей нейронних мереж, методик їх навчання та роботи з додатковими ресурсами для підтримки ефективної роботи різних видів нейронних мереж. Так наприклад в статі розглядаються основні особливості капсульних та згорткових нейронних мереж. Проводиться огляд специфік покращеного машинного навчання такого як: глибоке підтримуване навчання та аргументовано направлене навчання нейронних мереж. Розглядаються потенційні можливості керованого навчання та створення нейронних мереж з пам'яттю, концентруючись на мережах з довготривалою пам'яттю з використанням алгоритму еластичних вагових коефіцієнтів та прогресуючих нейронних мереж. Також описуються можливості гібридного навчання для збільшення продуктивності та точності роботи нейронних мереж. Метою даної статті було сформулювати список трендів в розвитку нейронних мереж на 2019 рік, засновуючись на аналізі можливостей та перспектив різних конструкцій та методів навчання нейронних мереж. Також проведено аналіз можливостей та продуктивності використання різних типів мереж для однакових цілей.*

Ключові слова: *нейронні мережі, тенденції в нейронних мережах, штучні нейронні мережі, штучні нейрони, глибоке навчання для нейронних мереж.*

Дибривный О. А. *Государственный университет телекоммуникаций, Киев*

ПРИКЛАДНОЕ ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ

Статья посвящена последним тенденциям, возникающим в области искусственных нейронных сетей, таким как капсульные сети, сверточные нейронные сети (CNN), глубокое усиленное обучение (DRL), бережливое расширенное обучение, модели под наблюдением, сети с моделью памяти, гибридные модели обучения. В статье приведены примеры сфер использования разных типов архитектур сетей, а также примеры использования разных моделей для одинаковых целей.

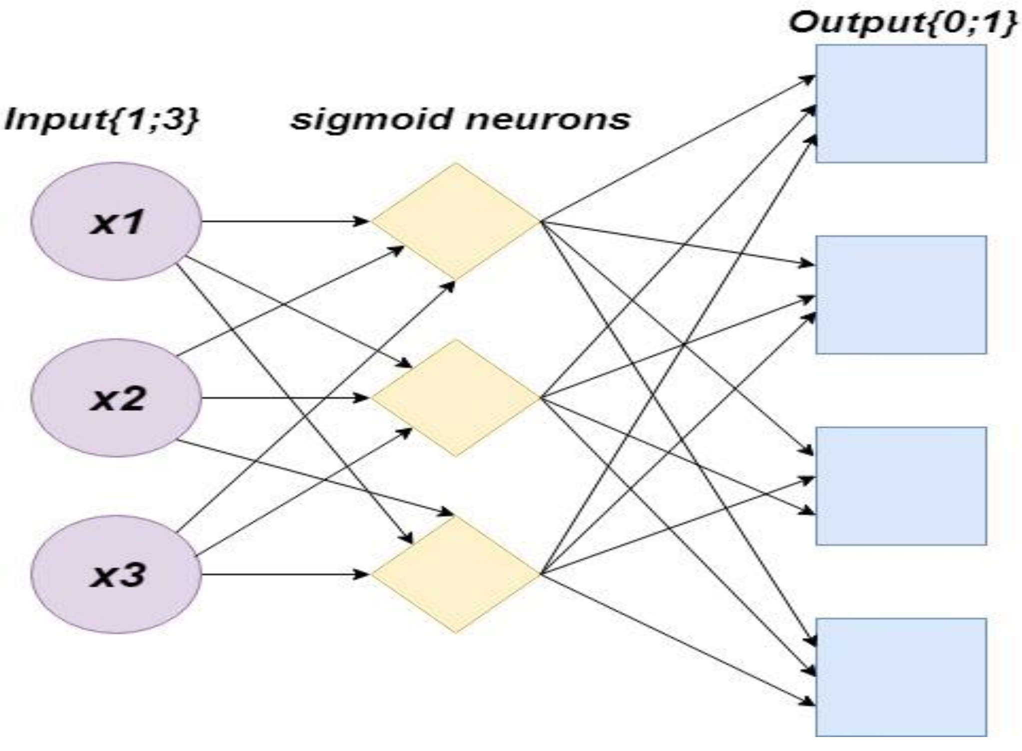
Ключевые слова: *нейронные сети, тенденции в нейронных сетях, искусственные нейронные сети, искусственные нейроны, глубокое обучение для нейронных сетей.*

1. Introduction

Neural Networks (NN) are computational models to mimic the human nervous system functionality. McCulloch and Pitts [1] suggested the first neuron model in 1943. The fundamental idea behind a neural network is to simulate multiple interconnected cells within a computer's "brain" so that it can learn from its environment, recognize different patterns and, in general, make decisions similar to a human being.

A basic neural network contains about millions of artificial neurons known as units. These units are arranged in layers with each layer connecting to the other side(Pic.1). The units are divided as follows:

1. Input Units — Designed to receive information from the outside environment
2. Hidden Units — These eventually feed into the Output Units. Each Hidden Unit is a squeezed linear function of its inputs.
3. Output Units — These signal how the Network should respond to the information it has recently acquired.



Pic.1. Architecture of classic neural network

Most of the neural networks are "Fully Connected." This implies that every hidden unit and every output unit is connected to every unit on the other side of the layer. The connections between each of the units are termed as "weight." The weight can be either positive or negative depending on the amount of influence it has on the other unit. Higher weights carry higher authority over interconnected units.

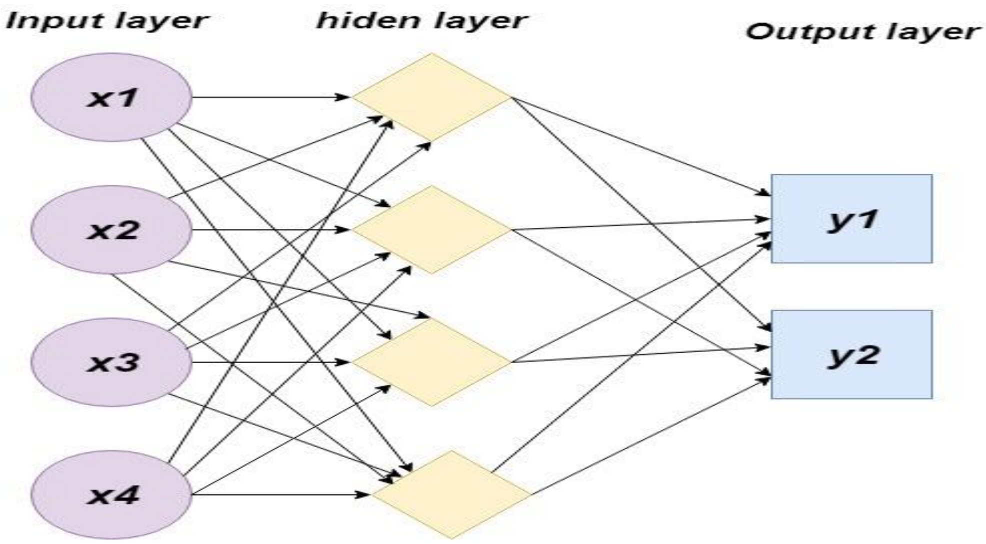
When a neural network gets trained, or just after the training when it starts operating, different patterns of information are fed into the network using different input units. These trigger the layers of the hidden groups, which then reach the output units. This is known as a Feedforward Network and is among the more commonly used designs.

After you've sufficiently trained the neural networks with different learning examples, it then arrives at a stage where it can be presented with an entirely new set of inputs, not encountered in the training phase, and it can predict an output which is satisfactorily accurate.

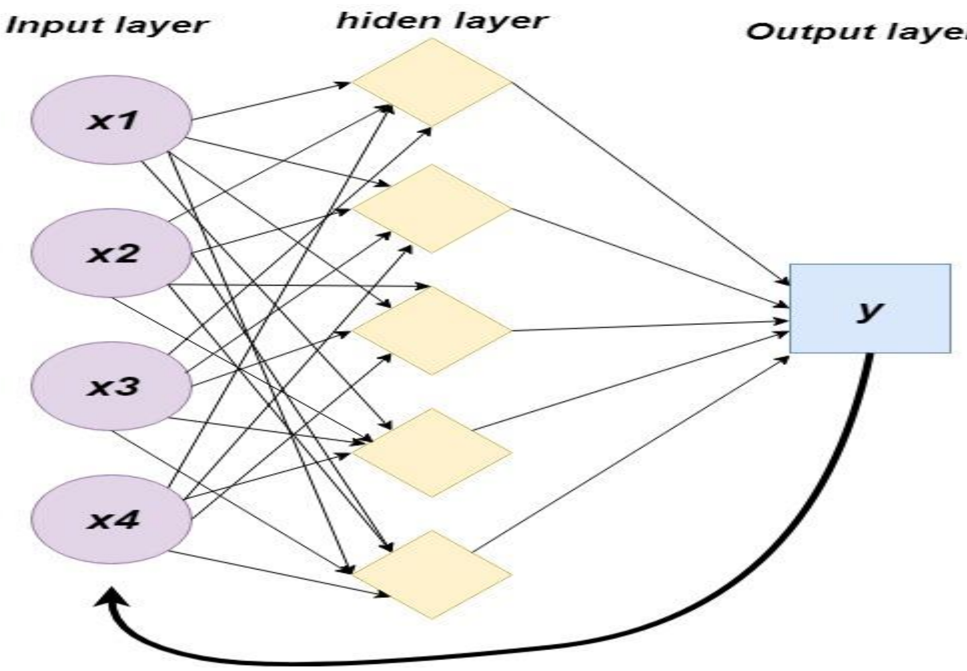
2. Analysis of research and publications

2.1 Types of Neural Networks

There are two types of neural networks; the Feed-Forward Neural Networks (FNNs) and the Recurrent Neural networks (RNNs) as shown in Pic.2 and Pic.3.



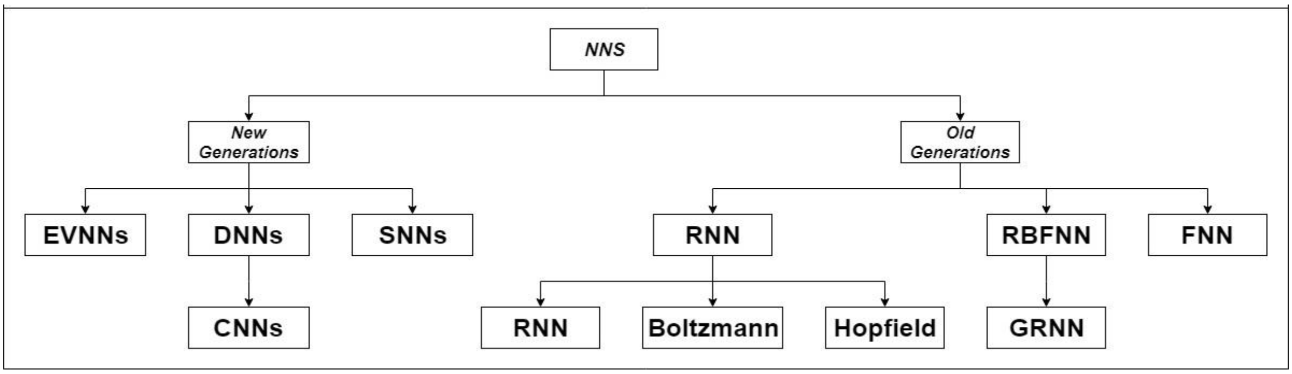
Pic.2. Feed forward neural network



Pic.3. Recurrent neural network

In FNN the data processing occurs only in the forward direction while in RNN the data processing can take place in the forward and the backward directions which enable RNN to have a memory of the output. FNNs can be categorized into the typical FNNs, RBFNNs and Generalized

Regression Neural Networks (GRNN) as shown in Pic. 4.



Pic.4. Neural networks types development

FNNs use dot product between the inputs and the weights and sigmoid activation functions in the hidden layer as given by equations (1) and (2). RBFNNs use the Euclidean distance between the inputs and the weights and Gaussian activation functions in the hidden layer as explained in equations (3) and (4).

$$Y_{FNN_S}=(\sum_{k=1}^{M}(\phi(\sum_{i=1}^{N}(x_i*\omega_i+b_i)))*\omega_0), \quad (1)$$

$$\phi(x) = \frac{1}{e^{-x}+1}, \quad (2)$$

$$Y_{\text{RBFNNs}}=(\sum_{k=1}^{M}(\theta(\sum_{i=1}^{N}(x_i*\omega_i+b_i)))*\omega_0), \quad (3)$$

$$\theta(x) = e^{\beta\|x-\mu\|^2}, \quad (4)$$

where x_i are the inputs, ω_i are the input weights, b_i is the input biases, w o are the output weights, β is the width of the radial basis function, and μ is the mean of the natural distribution.

The RNNs can be classified further into Hopfield, Boltzmann Machine, Self-organization Maps (SOM), Linear Vector Quantization (LVQ), and many other types. Fig. 5 demonstrates the different variations of RNNs. There is no fixed equation to calculate the output of the RNNs since the recursive connection can be placed in the hidden or the output layer and sometimes only part of the network is recursive. RNNs can be used in the continuous and the discrete time. The recursive links in the RNNs are memories of the former states or outputs of a given system.

2.2. Activation functions

Every neuron in the hidden layer or the output layer processes its input when it is fired with the so-called ’activation function’. These activation functions can be linear, binary, logistic, radial basis, or competitive functions. Each type of activation function can perform better in different tasks. For example, radial basis functions are better than linear functions in the approximation or regression problems while competitive functions perform better for classification tasks.

2.3. Neural Networks’ Learning Types

Neural networks can learn by supervised, unsupervised and Reinforcement Learning (RL). For supervised learning, the network is given labelled inputs and target examples while in unsupervised and reinforcement learning the network will be given inputs and it will have to find out the correct output.

1) Supervised Learning: One of the most typical supervised learning algorithms is the back-propagation. The backpropagation aims to change the parameters of the neural networks to follow a particular target or criteria. Back-propagation is based on gradient descent local optimisation which makes the derivative of the error function equals to zero.

2) Unsupervised Learning: In the Unsupervised learning, there is no error feedback to the network and the input data in unlabelled, and there is no supervisor. In this kind of learning, the neural network will have to cluster or classify the input data based on their statistical features, such as the mean, variance, and standard deviation. The most common types of unsupervised learnings are the cluster analysis. Self-organizing maps are one of the types of unsupervised learning which are based on competitive learning. In competitive learning, the output neurons compete between each other. For a neuron to be the winner, its weight vector should be the most similar to the input pattern. Usually, the similarity is measured based on the Euclidean distance between the weight of a neuron and the input pattern.

3) Reinforcement learning: Reinforcement learning is a type of unsupervised learning which relies on the learning from the interaction with the environment. The idea is that an agent (NN) will act, then it will observe the effect of the action and then based on the result of the action it will either be rewarded or punished. The reward or penalty can be used to adjust the NN parameters. The ultimate goal of the RL is to maximise the rewards of the actions.

3. Goals of research

The main goal of the study is to create a list of trends in the development of neural networks in 2019, building on previous studies of neural networks, in particular based on the advantages and disadvantages of individual designs and types of training neural networks. An important step in creating such a list is the selection and analysis of a wide range of specific neural networks, such as the architecture and methods of teaching the neural network.

4.Significant trends in in neural networks architecture

1) Capsule Networks are an emerging form of deep neural network. They process information in a way similar to the human brain. This essentially means that a Capsule Network can maintain hierarchical relationships.

This is in contrast to convolutional neural networks. Though convolutional neural networks are by far one of the most widely used neural networks, they fail to consider critical spatial hierarchies that exist between simple and complex objects. This leads to misclassification and a higher error rate.

When undertaking simple identification tasks, capsule networks provide a higher level of accuracy with a decrease in the number of errors. They also do not require a significant amount of data for training models.

2) Convolutional Neural Networks (CNN) have been around for ages and were inspired by biological processes - particularly the way how brain understands the signals it receives from the eyes. The state of the art visual recognition systems today use CNN algorithms to perform image classification, localization and object detection.

The interest in convolutional neural networks has been renewed because it's been heavily used for smart surveillance and monitoring, social network photo tagging and image classification, robotics, drones, and self-driving cars. The data scientists at Google, Amazon, Facebook etc. use this to do all sorts of image filtering and classification.

A closely related field is deep learning for computer vision which is what powers a barcode scanner's ability to "see" and "understand" the stripes in a barcode. That's also how Apple's Face ID recognizes you when it sees your face. To get started with deep learning for computer vision, there's tons of platforms offered that include Google's Vision API, Allegro.ai, Missinglink.ai etc.

3) Networks With Memory Model. One important aspect that distinguishes human beings from

machines is the ability to work and think discreetly. Computers can undoubtedly be pre-programmed to complete a specific task with extremely high accuracy. However, the problem occurs when you need them to work in diverse environments.

For machines to be suitable in real-world environments, neural networks have to be capable of learning sequential tasks without forgetting. It is essential for neural networks to be able to overcome catastrophic forgetting using the help of many different powerful architectures. These can include

1. Long-Term Memory Networks that can process as well as predict time series.
2. Elastic Weight Consolidation Algorithm that can slow down learning based on priority defined by previously completed tasks
3. Progressive Neural Networks that are immune from catastrophic forgetting is also capable of extracting useful features from already learned networks for use in a new task.

5. Significant trends in deep learning

1) Deep Reinforcement Learning is the form of a neural network which learns by communicating with its environment via observations, actions, and rewards. DRL has been successfully used to determine game strategies like those in Atari and Go. The famous AlphaGo program was used to defeat a human champion and has been also been successful.

DRL is essential as it is among the most general purpose learning techniques that you can use for developing business applications. It also requires significantly less data for training models. Another advantage is that you can train it by using simulation. This completely removes the need for labeled data.

2) Lean and Augmented Learning. By far, the biggest obstacle in Machine Learning in general and Deep Learning, in particular, is the availability of a significant amount of labeled data for training neural models. Two techniques can help address this – synthesizing new data and transferring a trained model for task A to task B.

Techniques like Transfer Learning (transfer the learning from one task to another) or One-Shot Learning (where learning occurs with only one or no relevant examples) make them Lean Data Learning techniques. Similarly, when new data is synthesized using interpolations or simulations, it helps to obtain more training data. ML experts usually refer to this as augmenting the existing data to improve learning.

Techniques such as these can be used to address a broader range of problems, especially where less historical data exists.

3) Supervised Model. A Supervised Model is a form of learning that infers a particular function from previously labeled training data. It uses a supervised learning algorithm that contains a set of inputs with the corresponding labeled correct outputs.

The labeled inputs are compared with the labeled outputs. Given the variation between the two, you can calculate an error value, and an algorithm is then used to learn the mapping between the input and the output.

The end goal here is to approximate the mapping function to the extent that if a new input data is received an accurate output data can be predicted. Similar to a situation where a teacher supervises a learning process, the learning process halts when the algorithm has achieved a satisfactory level of performance or accuracy.

4) Hybrid Learning Models.

Various types of Deep Neural Networks, including GANs and DRL, have shown a lot of promise when it comes to their performance and widespread application concerning different sorts of data. That said, Deep Learning Models cannot model uncertainty in a way that Bayesian or Probabilistic approaches can.

Hybrid Learning Models can combine the two approaches and utilize the strength of each. Some examples of such Hybrid Models include Bayesian GANs and Bayesian Conditional GANs.

Hybrid Learning Models allow for the possibility to expand the range of business problems that can be addressed to include Deep Learning with Uncertainty. This will allow for higher performance as well as explainability of models which can encourage more widespread adoption.

6.Comparative analysis of usage different types of neural networks architecture.

Most recent usage of all neural network models is image classification/ According to [9] even with a small number of hyperparameters, Capsule Networks can efficiently classify faces (on a given datasets). Additionally, Capsule Networks, like classical neural networks, requires more samples per class to achieve a lower test error. Good results on baseline and Capsule Networks architectures for face datasets are obtained through small variation within classes (relatively small degree of horizontal and vertical face rotation, relatively fixed face position with respect to image). Training Capsule Networks requires more computational resources than CNN because the outputs of primary capsules are activity vectors (with instantiation parameters) rather than scalars, leading to higher memory requirements because of the increased dimensionality. Moreover, when increasing the size (height px., width px.) of input images, the number of cells used in GPU RAM increases near exponentially. The authors argue that Capsule Networks has the potential to achieve a higher performance with some modifications in the hyperparameters. Having an 8D vector for routing capsules seems not enough for a complex dataset such as CIFAR-100, and 16D, 32D, 64D, may give a boost in performance, but also cause usage of more powerful GPU(s) to fit such complex model to memory. Another alternative could be decreasing the batch size to 1 sample, which may be harmful to the training process because the gradient might go in a wrong direction. Having good preprocessed dataset may solve this problem because in that case, every new sample pushes the gradient on the right track, but it may take longer to converge [10]

As for other field of usage of different neural network models we can highlight next:

The benefit of using CNNs is their ability to develop an internal representation of a two-dimensional image. This allows the model to learn position and scale in variant structures in the data, which is important when working with images.

CNNs are used for:

–image data;

–classification prediction problems;

–regression prediction problems.

More generally, CNNs work well with data that has a spatial relationship.

The CNN input is traditionally two-dimensional, a field or matrix, but can also be changed to be one-dimensional, allowing it to develop an internal representation of a one-dimensional sequence.

This allows the CNN to be used more generally on other types of data that has a spatial relationship. For example, there is an order relationship between words in a document of text. There is an ordered relationship in the time steps of a time series.

Although not specifically developed for non-image data, CNNs achieve state-of-the-art results on problems such as document classification used in sentiment analysis and related problems.

CNNs are tried to use for:

–text data;

–time series data;

–sequence input data.

And using RNNc includes the following set.

The Long Short-Term Memory, or LSTM, network is perhaps the most successful RNN because it overcomes the problems of training a recurrent network and in turn has been used on a wide range of applications. RNNs in general and LSTMs in particular have received the most success when working with sequences of words and paragraphs, generally called natural language processing.

This includes both sequences of text and sequences of spoken language represented as a time series. They are also used as generative models that require a sequence output, not only with text, but on applications such as generating handwriting.

It is not rational to use RNNs for:

– text data;
– speech data;
– classification prediction problems;
– regression prediction problems;
– generative models.

Recurrent neural networks are not appropriate for tabular datasets as you would see in a CSV file or spreadsheet. They are also not appropriate for image data input.

Usually RNNs are not used for:

– tabular data;
– image data.

RNNs and LSTMs have been tested on time series forecasting problems, but the results have been poor, to say the least. Autoregression methods, even linear methods often perform much better. LSTMs are often outperformed by simple MLPs applied on the same data.

7. Conclusions

Neural networks are now an essential part of our modern life. They are part of almost every intelligent system. The evolution of the research in the neural networks made their application in complicated systems feasible. The current research trends in the neural networks include the evolutionary neural networks, rough and spiking neural networks and deep neural networks. There are many practical applications using the neural networks. AI lays the foundation for a new era and many of the breakthroughs in technology are purely based on one of this technology.

Modern applications and infrastructure are generating log data that is captured for indexing, searching, and analytics. The massive data sets obtained from the hardware, operating systems, server software and application software can be aggregated and correlated to find insights and patterns. When machine learning models are applied to these data sets, IT operations transform from being reactive to predictive.

When the power of AI is applied to operations, it will redefine the way infrastructure is managed. The application of ML and AI in IT operations and DevOps will deliver intelligence to organizations. It will help the ops teams perform precise and accurate root cause analysis.

AIOps will become mainstream in 2019. Public cloud vendors and enterprise are going to benefit from the convergence of AI and DevOps. Machine learning and artificial intelligence will become the key technology trends of 2019. From business applications to IT support, AI is going to impact the industry significantly.

References

1. T. Sun, H. Pei, Y. Pan, and C. Zhang, "Robust adaptive neural network control for environmental boundary tracking by mobile robots," *International Journal of Robust and Nonlinear Control* 2 (2013), Vol. 23, 123–136.

2. W. Sun, Z. Zhao and H. Gao, "Saturated adaptive robust control for active suspension systems," *IEEE Transactions on Industrial Electronics* 9 (2013), Vol. 60, 3889–3896.

3. F. F. El-Sousy, "Adaptive hybrid control system using a recurrent rbfn-based self-evolving fuzzy-neural-network for pmsm servo drives," *Applied Soft Computing* (2014), Vol. 21, 509–532.

4. M. M. Ferdaus, S. G. Anavatti, M. A. Garratt and M. Pratama, "Fuzzy clustering based nonlinear system identification and controller development of pixhawk based quadcopter," in *Advanced Computational Intelligence (ICACI), 2017 Ninth International Conference on. IEEE,*

(2017), 223–230.

5. F. Lv, G. Yang, W. Yang, X. Zhang and K. Li, "The convergence and termination criterion of quantum-inspired evolutionary neural networks", *Neurocomputing* (2017), Vol. 238, 157–167.

6. H. V. H. Ayala and L. dos Santos Coelho, "Cascaded evolutionary algorithm for nonlinear system identification based on correlation functions and radial basis functions neural networks," *Mechanical Systems and Signal Processing* (2016), Vol. 68, 378–393.

7. I. Vlachos, T. Deniz, A. Aertsen and A. Kumar, "Recovery of dynamics and function in spiking neural networks with closed-loop control," *PLoS computational biology* 2 (2016), Vol. 12, e1004720.

8. A.J. Al-Mahasneh, S.G.Anavatti and M.A.Garratt, "The Development of Neural Networks applications from Perceptron to Deep Learning", *Conference paper for School of Engineering and Information Technology,* The University of New South Wales at the Australian Defense Force Academey Canberra, ACT 2612 Australia (2017), 118-121.

9. R. Mukhometzianov, J. Carrillo, "*CapsNet comparative performance evaluation for image, Classification*", University of Waterloo, ON, Canada (2017).

10. P.M. Radiuk, "Impact of Training Set Batch Size on the Performance of Convolutional Neural Networks for Diverse Datasets", *Riga Technical University, Information Technology and Management Science*, December (2017), Vol. 20, 20–24.

Автор статті (Author of the article)

Дібрівний Олесь Андрійович – аспірант, асистент кафедри інженерії програмного забезпечення (Dibrivnyi Oles Andriiovych – postgraduate student, Assistant Professor at Software Engineering department). Phone.: +380 (93) 432 65 43. E-mail: jokeresdeu@gmail.com.

Дата надходження
в редакцію: 02.04.2019 р.

Рецензент:
д.т.н., професор Кузнецов О.П.
Білоруський Державний університет
інформатики та електроніки, Білорусь