Some approaches to improving the quality of artificial neural network training

Yefim Rozenberg
JSC Railway Signalling Institute (JSC NIIAS)
Moscow, Russia
greatfime@gmail.com

Alexey Olshansky
JSC Railway Signalling Institute (JSC NIIAS)
Moscow, Russia
lexolshans@gmail.com

Ignat Dovgerd
JSC Railway Signalling Institute (JSC NIIAS)
Moscow, Russia
ignatikus@bk.ru

Gleb Dovgerd
JSC Railway Signalling Institute (JSC NIIAS)
Moscow, Russia
christmas1409@yandex.ru

Alexander Ignatenkov
JSC VTB Capital AM
Moscow, Russia
a.ignatenkov@gmail.com

Paul Ignatenkov
Smolensk State University
Smolensk, Russia
beat.pi@gmail.com

***Abstract*—The paper deals with improving the quality of artificial neural network (ANN) training. The research covers a complex neural network consisting of 2-dimensional Kohonen network and Wilshaw and von der Malsburg network capable of solving scheduling problems in transport. Existing results of using optimal control theory for ANN training are analysed; the authors suggest a new technique based on the direct neural control. Comparative error values during the training process using both the traditional methods and a new approach are presented. The new technique proves to be better than the traditional one for considered neural networks.**

Keywords—artificial neural network, Kohonen network, multilayered neural network, control

I. Introduction

Issues related to scheduling have always been of great significance for railway industry. Among the most common scheduling tasks one can mention routing, timetabling, volume planning, timetabling and volume planning, etc. Solving these tasks with strict methods we face certain problems such as combinatory complexities, exhaustive searches, computer memory deficiency, and time-consuming computations. In this case a number of heuristic algorithms are used (the Monte Carlo algorithm, evolutional algorithms, neural networks etc.). The present paper is aimed at illustrating how neural networks (a special category) solve timetabling tasks and create methods to control the quality of ANN training. The ANN under investigation [1] looks like 2-dimensional modification of the Kohonen and Wilshaw and von der Malsburg network.

When seeking a neural network solution of every task we should answer the following questions:

1. How to translate the task into the language “understandable” for the neural networks; how to find the correspondence between the states of neurons and the values of optimized parameters?
2. How to construct a network energy function with given constraints and given target function?

Immediately we run into two difficulties:

1. How to establish a correspondence between the members of a network energy function and the members of the general form of network energy?
2. How to calculate weighting factors for penalty functions?

One of the first attempts to overcome these shortcomings with regard to railway transport dates back to 2015 [1] and is connected with the development of a multilayer artificial neural network with variable signal conductivity (abbreviated as MANN VSC) to be applied for scheduling. Currently, this subject is considered to be the main source for research in the field of improving the quality of education.

ANN VSC is a hybrid neural network combining the characteristic features of a multilayer perceptron, the Wilshaw – von der Malsburg network with the Hopfield network.

II. About optimal control in neural networks tasks

Recently, the scope of application of neural networks has expanded considerably. The most popular tasks are synthesis of control systems, identification tasks, data processing, information recovery tasks, scheduling problems and other original activities (e.g. creating new pictures and arts).

Despite routine modifications of the structure and topologies of ANN and training methods, ANN is a system controllable only by using sets of recommendations based on heuristic approaches [2], numerical experiments, etc. Most authors emphasize that the quality of ANN training and the development and creation of neural network solutions is a complicated scientific problem. Sometimes we may see certain attempts of combined application of ANN and optimal control theory as a rigor mathematical method applicable for any task.

Paper [10] contains an attempt to create an algorithm for the development of the deep convolutional neural networks using manifold compactification. This approach is suitable for computer vision ANN but it is inconvenient for MANN with variable signal conductivity due to dissimilarity of their structures.

The theses [9] are more relevant for the ANN under consideration but it is impossible to apply the general idea of [9] because MANN follows its own rules of output calculation. Traditionally an artificial neural network implements an epoch as a full sequence of pairs “input-output” but MANN under consideration does not work with the set of different examples [8].

We should focus on paper [6] where the author suggests a genetic algorithm to optimize the vector of hyperparameters for convolutional neural networks. The closest result is in [7] where an asynchrony mover is a control object and two neural networks are suggested. The first network creates a control signal; the second one catches the difference between the desired output and the measurable output.

Paper [3] deals with constructing the optimal time sequences which consist of weights between neurons of a dynamic ANN. In [3] the two-point boundary value nonlinear problem is solved. It yields optimal rules of the ANN training. The weight matrix of the ANN in every time step (epoch) is set as an optimal time sequence. The authors note that at best the weight matrix at the final time step relates to the symmetric matrix constructed by J.J. Hopfield for associated memory [4].

Initial conditions are set as an input vector concatenated several training samples.

The functional (the criterion) of quality minimizes the value which is an opposite value of correlation between the output of the neuron and the desired output of the neuron at the final time step of controlling. During the time interval between the first step and the last step of controlling the functional penalizes miscorrelation level between the desired output and the answer of activation function of each neuron.

In this case an optimal control strategy is founded as Lagrange problem for a task of an optimal program control of the multilayered perceptron with a sigmoid activation function.

Another way of control is applying PID-controllers as a control technique.

A few more papers concerning ANN application for scheduling tasks should be mentioned. These solutions refer to scheduling, too; nevertheless, they touch upon modification of ANN activation function or the ANN structure. Thus, paper [11] analyzes a pickup of empirical coefficients for multilayer perceptrons and describes transferring to stochastic methods of weight modification at Hopfield models, etc.

Papers [12-16] address NP-hard problems (timetabling tasks, path searching in graphs) and its neural network solutions with different types of ANN (MLP, LSTM, CNN, etc.) and with various key algorithms (genetic algorithms, adjusting ANN parameters, error back propagation, standard searching).

However, these papers, like other articles analyzed above, do not consider an artificial neural network as a controllable object using optimal control theory.

Paper [16] is a meta-study about various approaches to solving schedule problems with different recommendations – from project management techniques to neural expert systems but without any neurocontrol and adjustments.

Examination of articles [11-16] leads to conclusion that the problem of improving the quality of neural network solutions is being analyzed in many countries. However, mission statement with regard to neurocontrol as a control task with two ANN has not yet received attention it deserves. In the field of neurocontrol this problem is rightfully considered novel. It refers both to optimal control theory and hybrid neurocontrol.

III. About PID-control in neural networks

PID control of the ANN error signal is found with the following classical formula [5]:

$$G(s)=K_1+K_2/s+K_3*s, \tag{1}$$

where s is the argument of the transfer function, K_1 is the coefficient of proportional regulation, K_2 is the coefficient of integral regulation, K_3 is the coefficient of derivative regulation. The PID-controller is implemented in the programming language R in the RStudio environment and after that it is incorporated into the code of multilayered ANN. The novelty of this approach is in the controllable object (the MANN as a kind of ANNs) and in the universal algorithm to transform a concrete PID-control curve to a strict indicator which sets a direction of the MANN signal trajectory.

Fig. 1 shows dynamics of changing the error signal for the MANN consisting of 27 layers and 1920 neurons in each layer and with 185 schedules as a computational load without control.

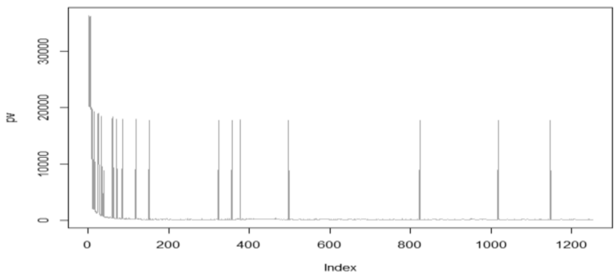

Fig. 1. The MANN error signal (a typical mode with a traditional algorithm [8], no control).

Fig. 2 shows the desired error change signal.

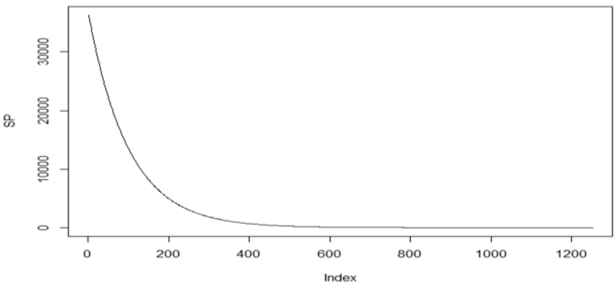

Fig. 2. Setup change (the principal view of the desired signal).

The authors organized and conducted about 1200 starts of the ANN with different parameters of the proportional (ranging from 0.1 to 1), the integral (from 10 to 40) and the differential (from 0.1 to 4.1) error components and the disturbations value from 5 till 60 points per every time step. It is not a not very efficient method of control because it provides only 10% stable trajectories. The stability is taken into consideration in a Lyapunov sense [5]. Computational experiments illustrate that the marginal critical value of the disturbations feed to the ANN is no more than 10-15% of the average error in the stable mode (Table 1.). This result cannot be evaluated as practical.

IV. Direct Neurocontrol for multilayers artificial neural networks and its advantages

Along with the traditional training algorithm the authors suggest a direct neurocontrol mode for training. The object to be controlled is a multilayered ANN with variable signal conductivity [1]; a three-layer perceptron with sigmoid activation functions is taken as a controller.

The main scheme of control is shown in Fig.3.

The ANN-controller is trained by the aggregation of triple sets “ The level of error per epoch” – “The level of error at the previous moment" – "The control signal from the previous time step to the present time step" or "The previous level of error" – "The current level of error" – "The control signal".

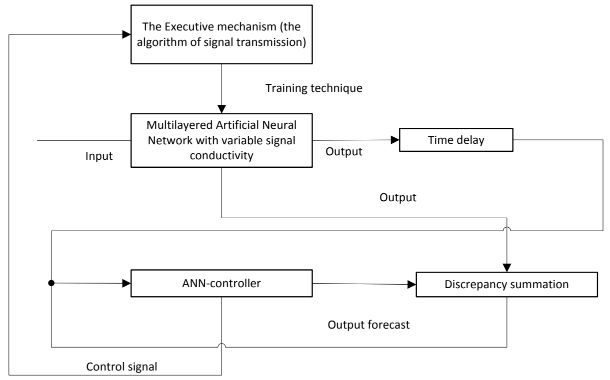

Fig. 3. The scheme of a direct neurocontrol mode.

The current error signal of the ANN and the previous one are gathered and entered the trained and ready multilayer perceptron. An answer signal of the ANN-controller entered the discrepancy summation and actuating mechanism (an algorithm). Hereinafter the value of summated discrepancy is also fed by the ANN-controller.

The control scheme described above was tested for the concrete scheduling problem (the railway branch Arkhara – Volochaevka, 27 railway stations). The task included 185 trains per 24 hours.

The results of testing are given in the table 1.

TABLE I. A COMPARISON OF DIFFERENT TRAINING METHODS

Training error (points)	*Traditional algorithms*	*PID-controller (the best configuration with K1/K2/K3 = 0.1/40/2.1)*	*Direct neurocontrol*
Min	75	*362*	*193*
Max	134795	*211585*	*57895*
Median	5469	*471*	*210*
Average	16548	*1830*	*384*
SD	6687	*4485*	*1180*
Rate of error overshoot	50	*15*	*0.4*

V. CONCLUSIONS

Thus, this work shows the principal possibility to control the multilayered artificial neural network with variable signal conductivity. The three layered perceptron with the sigmoidal activation function is used as a controller. The solutions achieved using multilayered artificial neural networks with direct neurocontrol are of much better quality (as compared with those obtained with traditional algorithms.

ACKNOWLEDGMENT

This research was provided by Russian Foundation for Basic Research (the project #17-20-01065 "A theory of railway transport system neural network control").

REFERENCES

[1] A. Olshansky and A. Ignatenkov, "One approach to control of a neural network with variable signal conductivity," Information tecnologiesn and Nanotechnologies (ITNT), pp. 984-987, 2017.

[2] A.V. Nazarov and A.I. Loskutov, "Neural networks algorithms of forecast and system optimization," Saint Petersburg: Science and technic, 2003, 384 p.

[3] O. Fahotimi, A. Dembo and T. Kailath, "Neural network weight matrix synthesis using optimal control techniques," Advances in Neural Information Processing Systems-2 (NIPS-2), USA, Denver, Colorado, USA, Stanford,1989.

[4] J.J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," Proc. Natl. Acad. Sci. Biophysics, USA, vol. 79 pp. 2554-2558, 1982.

[5] R.C. Dorf and R.H. Bishop, "Modern control systems," Pearson, 2011.

[6] Y.R. Tsoy, "Neuroevolution algorithm and software for image processing: The dissertation on competition of a scientific degree of Candidate of technical sciences," Tomsk Polythechnic University, 2007, 209 p.

[7] A.M. Sagdatullin, "A neural network controller for the velocity value of an asynchronic motor," Theses of Russian Congress for Control Problems, Russia, Moscow, Institute of control sciences of Russian Academy of Sciences, pp. 4485-4498, 2014.

[8] A.M. Olshansky and A.V. Ignatenkov, "Development of an artificial neural network for constructing a train schedule," Bulletin of the Ryazan State Radio Engineering University, vol. 55, pp. 73-80, 2016.

[9] I.M. Kulikovskikh, "Reducing computational costs in deep learning on almost linearly separable training data," Computer Optics, vol. 44, no. 2, pp. 282-289, 2020. DOI: 10.18287/2412-6179-CO-645.

[10] Yu.V. Vizilter, V.S. Gorbatsevich and S.Y. Zheltov, "Structure-functional analysis and synthesis of deep convolutional neural networks," Computer Optics, vol. 43, no. 5, pp. 886-900, 2019. DOI: 10.18287/2412-6179-2019-43-5-886-900.

[11] A.S. Jain, "Meeran S. Job-shop scheduling using neural networks," International Journal of production research, vol. 36, no. 5, pp. 1249-1272, 1998.

[12] Z. Li, Q. Chen and V. Koltun, "Combinatorial optimization with graph convolutional networks and guided tree search," Advances in Neural Information Processing Systems, pp. 539-548, 2018.

[13] A. Milan, "Data-driven approximations to NP-hard problems," Thirty-First AAAI Conference on Artificial Intelligence, 2017.

[14] J. Bruck and J.W. Goodman, "On the power of neural networks for solving hard problems," Neural Information Processing Systems, pp. 137-143, 1988.

[15] A. Chaudhuri and K. De, "Job Scheduling Problem Using Rough Fuzzy Multilayer Perception Neural Networks," Journal of Artificial Intelligence, vol. 1, no. 1, 2010.

[16] S.J. Noronha and V.V.S. Sarma, "Knowledge-based approaches for scheduling problems: A survey," IEEE Transactions on Knowledge and Data Engineering, vol. 3, no. 2, pp. 160-171, 1991.