

Statistical learning problem of artificial neural network to control roofing process

Azariy Lapidus¹, Aleksandr Makarov^{1*}

¹Moscow state university of civil engineering, Yaroslavskoye shosse, 26, Moscow, Russia, 129337

Abstract. Now software developed on the basis of artificial neural networks (ANN) has been actively implemented in construction companies to support decision-making in organization and management of construction processes. ANN learning is the main stage of its development. A key question for supervised learning is how many number of training examples we need to approximate the true relationship between network inputs and output with the desired accuracy. Also designing of ANN architecture is related to learning problem known as “curse of dimensionality”. This problem is important for the study of construction process management because of the difficulty to get training data from construction sites. In previous studies the authors have designed a 4-layer feedforward ANN with a unit model of 12-5-4-1 to approximate estimation and prediction of roofing process. This paper presented the statistical learning side of created ANN with simple-error-minimization algorithm. The sample size to efficient training and the confidence interval of network outputs defined. In conclusion the authors predicted successful ANN learning in a large construction business company within a short space of time.

1 Introduction

In recent years, neural networks developed and tested for construction project organization and management can support decision making for construction process safety [1] and project management [2], they are also able to predict and evaluate specific construction processes such as: building underground structures [3], road building [4], earthwork [5], and other. The authors conduct a research in construction process organization and management to build roof structures at the planning and operation stage, assisted by the ANN methodology. Previously, the network architecture was designed (fig. 1) and functioning of each layer was determined. The main parameters of the system (Q_w – qualification of workers, C_s – construction supervision, It – information technology, T_s – team structure, W_c – weather condition) are located on the second layer. These parameters are divided into basic organizational and technological factors (for example, g_1 - working experience and competencies, g_4 – foreman). Third layer includes four evaluation criteria of construction process (T – time, S – safety, Q – quality, C – cost) and last layer merges all information that passed through the network into an integrated indicator P (construction process potential). To measure ANN inputs, the authors used three linguistic variables represented by fuzzy sets [6] with arguments inside the interval [0,1]. The operator – an

* Corresponding author: anmakarof@yandex.ru

expert in the field – estimates each factors by determining a linguistic term and selecting a point of the membership function graph. The abscissa vector

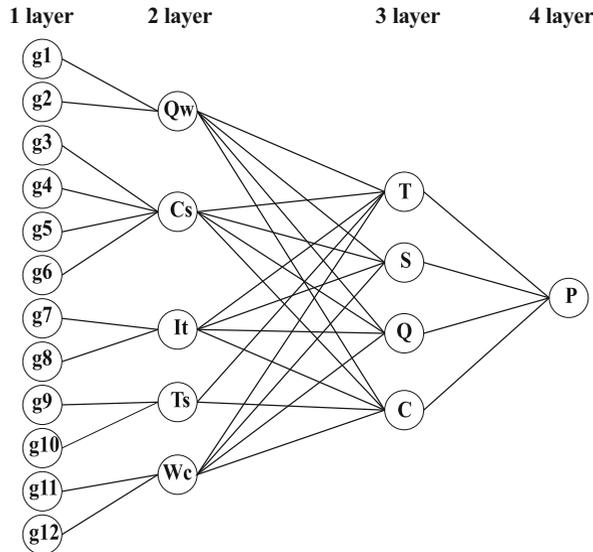


Fig. 1. ANN architecture.

of these points for all factors is input data of the network. ANN learning is a process whereby weight coefficients of synapses are tuned in a certain way so that the network deduces reliable results. A number of approaches to ANN learning exist. For a multilayer perceptron, supervised learning by mean the backward propagation of errors is one tried and trusted approach. This algorithm needs a training sample comprised of an input signal vector and a corresponding output signal vector. Formation this sample in real construction conditions is a laborious process that requires open access of researchers to all stages and participants of construction process. The organization of that access is impossible without the direct interest of the decision maker of construction project. Therefore, it is very important to estimate in advance a sufficient training sample size that must be obtained from construction site for the successful learning of the created ANN. Also providing enough training examples to make the learning algorithm statistically reliable is a key objective in the machine learning theory [7], and we are going to discuss a solution to it for the developed ANN.

2 Methods

ANN learning can be described as search for function $F(x, w)$ in the space of all possible function Ω , which approximates work of a real-life system with the permissible departure/error and sufficient reliability. According to the statistical learning theory [8], the objective is to minimize risk functional by parameter w :

$$R(w) = \int Q(x, y, w)P(x, y)dydy = \int [y - F(x, w)]^2 P(x, y)dx dy, \quad (1)$$

where $Q(x, y, w)$ - loss-function, $P(x, y)$ - unknown joint probability distribution of the network's ins and outs.

Because recovering density of probability distribution over the sampling is a labor-intensive task, all ANN learning algorithms based on precedents, particularly backpropagation, seek to minimize the empirical risk

$$R_{emp}(w) = \frac{1}{N} \sum [y - F_{emp}(x, w)]^2 \quad (2)$$

Since a training sample size is always finite, the main issue about reaching the stopping criterion is about how function $F_{emp}(x, w)$ successfully obtained from experimental data can minimize the risk functional (1) – in other words, how the error after learning will differ from the network's error in real-life operation.

Solution to this objective is related to deciding the condition of convergence of empirical averages to expected values [9]; it appears as

$$P\{\sup |R_{emp}(w) - R(w)| > \varepsilon\} < p, \quad (3)$$

and it means that with probability of $1 - p$, disparity $R_{emp}(w) - \varepsilon \leq R(w) \leq R_{emp}(w) + \varepsilon$ is true simultaneously for all input values; where ε - bound of confidence interval or error boundary.

As they analyzed the conditions of convergence (3) for training samples of finite length $n = N$ and as they generalized the Glivenko theorem for classification of images, Soviet researchers V. Vapnik and A. Chervonenkis introduced the notion of capacity h of function set Ω , known internationally as the Vapnik-Chervonenkis dimension or *VC-dimension* [10]. We used this parameter to find permissible conditions for regular convergence over the sample size of $N \succ h$, from which one can isolate the expression for error boundary:

$$|R_{emp}(w) - R(w)| \leq \frac{\varepsilon_0}{2} \left(1 + \sqrt{1 + \frac{4R_{emp}(w)}{\varepsilon_0}}\right), \text{ where} \quad (4)$$

$$\varepsilon_0 = 4 \frac{h(\ln \frac{2N}{h} + 1) - \ln p}{N}. \quad (5)$$

Since $R_{emp}(w) \square 1$, this expression means that the sample size depends on h alone. For neural network with sigmoid activation function, the established range of *VC-dimension* [11, 12] has the following orders: $\inf h = O(W^2)$, $\sup h = O(W^4)$, where W – number of free network parameters, i.e. synaptic weights.

Training of a created network stops at Layer 3, and the synaptic weights of neuron P are beyond the training array; then $W = 32$ and *VC-dimension* will have the following orders:

$$\inf h = O(10^3), \sup h = O(10^6). \quad (6)$$

Considering (4, 5) and the condition $N \succ h$ we find that training sample size is bounded by the values of (6).

The resulting theoretical training sample size sufficient to ensure statistical convergence of the ANN has a very large gaps between its boundaries; one of the reasons is that this is a conservative estimate for a scenario with random initial values of weights, zero *a priori* information about the system except the number of links, and no consideration for the specifics of different training algorithms. Subsequently, known results from practical runs of different neural networks demonstrated that to implement statistical convergence in real-life missions, a probably approximately correct learning (PAC-learning) [13, 14] is more preferable; it was also found that for the classification problem, the complexity of the training sample depends more on weights than on their number [15]. This is why, to evaluate the capacity of the function set Ω we proposed that the *VC-dimension* be

substituted with its upscale version: *fat-shattering dimension* [16, 17], provided that the ANN learning algorithm ensures maximized share of accurate output values, i.e. with minimum error in learning data subjected to the restriction

$$\frac{1}{N} \sum [y - F_{emp}(x, w)]^2 \leq \frac{1}{\sqrt{N}}. \quad (7)$$

Papers [18, 19] demonstrated that using the *fat-shattering dimension* $fat_F(\varepsilon)$ one can successfully determine complexity of the required training set for implementation of neural network convergence in problems of complex classification, and in recovery of regression. If we use a probably-approximately correct learning model with an SEM-algorithm, where the convergence condition (3) is described with the function of losses

$$P\{Q[y - F_{emp}(x, w)]^2 - \inf Q[y - F(x, w)]^2 > \varepsilon\} < p, \quad (8)$$

which means that with probability of $1 - p$ the approximating function received from the learning sample will give a quadratic error not above ε , compared to the best function within the function set Ω .

Thus, for a neural network with inputs from the interval of $[0,1]$, and with outputs delimited by A , with l of computing layers, whose each neuron has an activation function that satisfies the Lipschitz condition $|\varphi(v_1) - \varphi(v_2)| \leq L |v_1 - v_2|$, and a limited sum of weights $\sum w \leq V$, we are going to have

$$fat_F(\varepsilon) \leq 16W(l \ln(LV) + 2 \ln(32W) + \ln(\frac{1}{\varepsilon(LV-1)})). \quad (9)$$

We used $fat_F(\varepsilon)$ to find a sufficient training sample size to meet the condition of convergence (8)

$$N \geq \frac{c}{\varepsilon^2} (fat_F(\varepsilon) \log^2(\frac{1}{\varepsilon}) + \log(\frac{1}{\delta})), \quad (10)$$

being of the following order

$$N = O(\frac{1}{\varepsilon^2} (\frac{A^2 V^{l(l+1)}}{\varepsilon^{2l}} \ln m + \ln(\frac{1}{\delta}))). \quad (11)$$

3 Results

The ANN built by these authors has three computing layers $l = 3$, its activation function does meet the Lipschitz condition at $L = 10$, $m = 12$; therefore, considering condition $\sum w = 1$ (2) and assuming that $\varepsilon = 0,8$ and $p = 0,05$ quoted from above (9-11) we get $fat_F(\varepsilon) = 11874$, while the necessary training sample will have order $O(1486)$ for the size of $N \geq 929$ that is sufficient to meet condition (8). Our result does not contradict the lower boundary of our training set (6) that we found using the *VC-dimension*.

Assume the best function of the set yields a squared error $\inf Q[y - F(x, w)]^2 = 0.01$. Then based on the results and considering (8), the output value for the ANN y_p from neuron P will produce this confidence interval,

$$P(|y - y_p| \leq 0,9) \geq 0,95. \quad (12)$$

and this means that following fuzzification the output from the ANN enters point B in the membership function graph, even as the actual value of the construction process potential is in section CD with probability of $p = 0,95$ (Fig. 2). Because neuron P unlike the rest is linear unit with activation function $y_p(v) = 10v$ and considering condition $\sum w = 1$ the argument interval of fuzzy sets will appear as $[0,10]$.

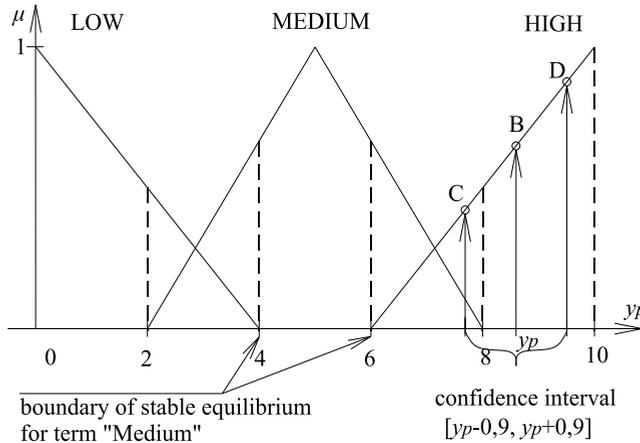


Fig. 2. Confidence interval of ANN outputs.

Because any high-rise construction process, including one related to roofing, is a complex system with a lasting life cycle, considerable human factor, and certain unpredictable random disturbances, the construction developer's main objective then is to keep the key process variables that all combined constitute potential P , within some specific preset boundaries that determine the system's sustainable equilibrium. The authors identified such boundaries, within which only one linguistic term can be assigned to the measured component (Fig. 2). The above confidence interval (12) of output values for the ANN is narrower than the stable equilibrium interval, and therefore, we can use the resulting interval to successfully evaluate and monitor the dynamics of both the integrated construction process potential P , and its values by one of criteria (T , S , Q , C). As an example, to evaluate P based on the quality criterion, we assume value $w_{QP} = 1$ to the quality criterion $Q - P$, the remaining weights of neuron P $w_p = 0$. Also the resulting confidence interval is narrower than in (12), considering the statistical aspects of learning discussed above.

4 Discussion

The final stage of learning for the created ANN understands its test run on real-life construction projects. If the construction company wants the ANN to be efficiently trained and if its project managers get actively involved, up to 2-3 training examples can be collected in a week of construction activities. Seeing that the average life cycle of a roofing process is 2 to 3 months, a project can quite realistically yield a sample size $N = 20 \div 30$. A large construction business, particularly one focused on roofing, or a profiled construction auditor may handle up to ten construction projects at the same time – so the average training sample size can be approximately $N = 800 \div 1000$ in a year. The academic results of sufficient training sample presented in this paper with the use of the SEM algorithm suggest that after a year or two of training in a major construction business the created ANN will be able to produce results with accuracy of (12). Meanwhile, our

expressions (17, 22-23) lead us to believe that growth of sample size tends to make the confidence interval of ANN outputs narrower (12), and thus more accurate. Keeping the stability of learning algorithm under control, this will ensure continuous progress of the created ANN.

References

1. D.A. Patel, K.N. Jha, *SS*, **89**, 240-248 (2016)
2. F. Costantino, G.D. Gravio, F. Nonino, *IJPM*, **33**, 1744-1754 (2015)
3. E. Siami-Irdemoosa, S.R. Dindarloo, M. Sharifzadeh, *AiC*, **58**, 85-94 (2015)
4. M. Li, W. Chen, *PP*, **25**, 1212-1217 (2012)
5. B. Hola, K. Schabowicz, *AiC*, **19**, 570-579 (2010)
6. A. Lapidus, A. Makarov, *MATEC WoC*, **86**, 05003 (2016)
7. M. Anthony, *DAM*, **156**, 883-902 (2008)
8. V. Vapnik, *IEEE Trans. Neu. Net.*, **10**, 5, 988-999 (1999)
9. V. Vapnik, A. Chervonenkis, *TP&A*, **16**, 2, 264-280 (1971)
10. V. Vapnik, A. Chervonenkis, *TP&A*, **26**, 532-553 (1981)
11. P. Koiran, E.D. Sontag, *JC&SS*, **54**, 190-198 (1997)
12. M. Schmitt, *NIPS 1999*, 328-334, (1999)
13. D. Haussler, *I&C*, **100**, 78-150 (1992)
14. S. Floyd, M. Warmuth, *ML*, **21**, 269-304 (1995)
15. P.L. Bartlett, *NIPS 1996*, 134-140 (1996)
16. M.J. Kearns, R.E. Schapire, *JC&SS*, **48**, 3, 464-497 (1994)
17. N. Alon, S. Ben-David, N. Cesa-Bianchi, D. Haussler, *JACM*, **44**, 4, 615-631 (1997)
18. P.L. Bartlett, P.M. Long, R.C. Williamson, *JC&SS*, **52**, 434-452 (1996)
19. M. Anthony, P.L. Bartlett. *Neural network learning: theoretical foundations* (Cambridge university press, 1999)