Algorithm for neural network regeneration of labor costs based on the assessment of relevant construction data

Yousif Babakr Hussein Khoshnaw¹, Sergey Bolotin¹, Valeriya Bagulya², Haitham Bohan¹

¹Saint Petersburg State University of Architecture and Civil Engineering, Saint Petersburg, Russia
²Peter the Great St.Petersburg Polytechnic University, Saint Petersburg, Russia

E-mail: yousif.babakr@gmail.com, sbolotin@mail.ru, Valeriya.bagulya@mail.ru, haitham_kh9@yahoo.com

Abstract. In the scheduling of construction production, which is an integral function of the organization of construction, the task often arises of choosing adequate labor costs. There is an obvious way to solve this problem - this is the use of previously developed databases. However, its implementation requires the creation of such reference standards by which an adequate choice can be made. On the other hand, creating an entirely new labor base will require even more time and resources. In order to reduce all costs associated with the creation of the database, a neural network algorithm for the regeneration of labor costs is proposed based on the assessment of relevant data, which can be formed on the basis of executive documentation for completed construction projects and for organizations involved in their implementation. As a result, the regenerated norms of labor costs are based on the secondary use of statistics fixing actual labor costs.

1. Introduction
The effectiveness of the organization and management of construction is largely determined by the quality of scheduling, which, in turn, depends on the labor standards used. Such leading construction powers as the USA, Great Britain, France, the Russian Federation, Finland and others have their own base of labor cost standards, and their development is carried out by specialized organizations [1, 2]. For example, in the UK, BCIS company publishes Comprehensive Building Price Book and Davis Langdon an AECOM Company sponsor's Price Books. In the United States, Compass International company publishes Global Construction Costs Yearbook and Building Construction Cost Data Book (RSMeans). In France, Groupe Moniteur company publishes Le coût des travaux de bâtiment. In the USSR, labor costs norms were issued under the name “Unified Norms and Prices” (ENiR) [3] and they were created on the basis of data obtained by regulatory research stations. In the future, the developed norms were introduced by the state with a view to their practical use in the construction industry. In the Russian Federation, which is the successor of the USSR, similar rules are also introduced by the state, but in most cases they are advisory in nature. In Finland, the Construction Industry Association issues RATU standards made in the form of flow charts and they contain enlarged and operational information on labor costs, the recommended composition of the link, the necessary documents and plans, the required materials, machines and equipment, safety measures and quality assurance measures [2]. The companies and construction sites participating in the study are
located in various regions of Finland, and each rate is calculated based on data obtained from at least 10 sites, and the standard error does not exceed 10%.

Along with this, there are countries in which there is no own regulatory framework for labor costs, and for them the security problem can be solved by importing any existing database. However, for its adequate choice, it will still require the creation of some (limited in number) standards by which the procedure for finding the closest match can be carried out. The event of importing a suitable database can be replaced by the option of creating your own regulatory framework, but you should take into account the specific features of this approach [4].

The most used, when creating a database, is a timing method for measuring labor costs, which requires the creation of appropriate organizational structures and time-consuming. Another productive method is the so-called physiological approach, which ensures the objectivity of the measurement results.

Its essence lies in the fact that labor standards are determined on the basis of measurements of the energy costs of the employee, obtained from the following characteristics: oxygen consumed by the employee, heart rate, pulmonary ventilation, body temperature and lactic acid concentration in the blood. In the regulation of labor, the physiological method has established itself as the most accurate.

The analysis shows that to create a complete or even limited database of labor costs will require significant time and significant resource costs. Therefore, an alternative to the existing approaches may be the development of a procedure for the regeneration of labor costs based on the assessment of current data on construction conditions.

2. Discussion
The regeneration of labor cost norms is based on the secondary use of statistics that record actual labor costs. At the same time, relevant statistics can be generated on the basis of executive documentation related to completed construction projects. The process of translating the regenerated data into the form of labor cost standards can be defined as a regression process, supplemented by expert assessments of the conditions under which the actual values of labor costs are obtained.

Consider two possible alternatives associated with the formulation of the regression process. As the first alternative, we consider the use of classical regression analysis, and as the second alternative, we consider a more modern method based on the use of neural networks [5]. As you know, the method of multiple regression analysis depends on the degree of linearity of the models used, while the traditional regression analysis is focused on linear models. However, if the model differs from the linear one, either linearization of its parameters is applied, or a specific nonlinear regression algorithm is developed [6]. However, in any case, a priori information is required on the form of the nonlinear connection between the function and the arguments. For our task, the degree of non-linearity of the relationship between the conditions of labor processes, which include the assessment of labor and machine resources, supplied materials and weather conditions, and actual labor costs cannot be established a priori. Given this fact, the method based on modeling neural networks is of primary importance, since it does not require a priori information that determines the degree of non-linearity of the problem posed [7].

In [8], James McCaffrey uses the term neural network regression, and he believes that this type of regression is the most powerful form that ensures the correct implementation of the regression task. But its implementation requires the construction of an adequate neural network architecture, and this is the next debatable issue.

Basically, neural networks are based on a model of an artificial neuron, the inputs of which receive signals from the outputs of other neurons - xi, and in addition, each signal is characterized by a value that is usually called weight - wi. As a result, the overall output state of an artificial neuron is determined by the weighted sum of its inputs and the selected activation function. In artificial neural networks, various activation functions are used, among which the logistic function is the most popular [9]. Given the choice of the logistic function, the relationship between the inputs of an artificial neuron and its output – Y is determined by the following expression.
\[ Y = \frac{1}{1 + \exp \left( - \sum w_i x_i \right)}. \] (1)

It should be noted that the use of only one neuron makes the approximation task almost identical to the traditional method of regression analysis. Therefore, in the general case, to build an adequate neural network, many neurons are used that form such a property of the network as its multilayer. By a layer of an artificial neural network is meant a lot of neurons, to which signals from other neurons of a given network arrive in parallel. In the general case, the choice of the architecture of an artificial neural network is determined by the specifics of a particular task. For example, for the task of predicting the duration of construction, an artificial direct distribution network is used, which is called a multilayer perceptron [10]. The monograph [11] gives a recommendation on the sufficiency of using three hidden layers and the equality of the number of neurons in the input layer to the number of input parameters.

3. Description of the algorithm

At the inputs of a three-layer perceptron signals are given that simulate the conditions for ensuring the performance of work \( \lambda i \) on labor resources, machines mechanisms, and materials, as well as deviation from the norm of weather conditions, which increase labor costs. In this case, the values of the above conditions are determined on the basis of expert estimates, and the labor costs \( Y \) are determined on the basis of executive documentation, for example, according to form KC-2 [12]. Table 1 shows the presentation of the source data for the calculation and output of the results according to the norms of labor costs (HT3). This form, as well as the corresponding HT3 calculation program, is implemented in the environment of the Excel table processor.

The form of presentation of the source data and the conclusion of the calculation results.

| №  | Medium | The security of labor,\% | Labor costs (TK) |
|----|--------|-------------------------|-----------------|
|    | conditions | Workers (P) | Machines (M) | Transport (T) | Climate (k) | Fact (ФТ3) | Calculation (РТ3) |
| 1  | 66.5 | 60 | 65 | 70 | 80 | 0.81 | 0.81 |
| 2  | 76.5 | 70 | 75 | 80 | 90 | 0.74 | 0.74 |
| 3  | 87.0 | 80 | 90 | 90 | 100 | 0.72 | 0.73 |
| 4  | 85.5 | 75 | 75 | 100 | 80 | 0.71 | 0.71 |
| 5  | 87.5 | 90 | 95 | 80 | 100 | 0.66 | 0.67 |
| 6  | 83.5 | 90 | 95 | 70 | 100 | 0.70 | 0.70 |
| 7  | 78.5 | 65 | 75 | 90 | 90 | 0.77 | 0.78 |
| 8  | 78.5 | 80 | 65 | 75 | 100 | 0.75 | 0.75 |
| 9  | 78.5 | 75 | 75 | 80 | 90 | 0.78 | 0.77 |
| 10 | 84.0 | 80 | 90 | 85 | 90 | 0.73 | 0.73 |
| 11 | 82.5 | 84 | 65 | 86 | 80 | 0.78 | 0.78 |
| 12 | 81.5 | 65 | 89 | 96 | 82 | 0.71 | 0.71 |
| 13 | 81.4 | 87 | 92 | 70 | 94 | 0.70 | 0.70 |
| 14 | 84.3 | 66 | 94 | 97 | 97 | 0.69 | 0.69 |
| 15 | 84.4 | 71 | 83 | 98 | 85 | 0.71 | 0.71 |
| 16 | 72.8 | 64 | 90 | 72 | 94 | 0.67 | 0.67 |
Using data by input 4 variables and recommendations on the topology of the neural network, it can be calculated that the number of parameters of the three-layer perceptron is 53. Thus, the simulated non-linearity of the dependence of labor costs associated with 4 measured conditions of building production \(Y(X_1, X_2, X_3, X_4)\) is described by a mathematical model based on 53 parameters. However, neural networks work with numerical data lying in different ranges, normalization is performed for input and output variables, bringing all the data to a single range from 0 to 1. In the neural networks presented in Table 1, the initial data is called a training array. This data is limited by the volume of 10 options, which corresponds to the recommendations declared by the creators of the RATU base for the lower boundary of statistical representativeness [2]. Although it should be noted that in the review of neural networks [11], the following heuristic rule is given: the number of observations should be ten times greater than the number of connections in the network. For our purposes, in view of the informational complexity of the primary problem, solved on the basis of collecting a large number of statistical data, the training array is determined by 10 options.

The conditions of labor security, presented in table 1, are quantitatively determined in percent, which are obtained on the basis of the relevant expertise. Next, data on actual labor costs attributed to a unit of production are entered. For example, for the production of concrete structures, the unit of measurement of labor costs can be a person-shift, and the unit of output is a cubic meter of concrete structure.

A certain difficulty is the question associated with determining the length of the calculation step, which determines the calculation time. The monograph [11] shows the importance of this parameter, and adaptive algorithms for dynamic speed control are used to reduce the calculation time. The estimated values of labor costs presented in table 1 are obtained with a calculation step equal to 0.2 and the number of iterations equal to 40,000.

The data calculated on a three-layer perceptron are characterized by the following estimated characteristics. The pair correlation coefficient between working conditions and actual labor costs amounted to -0.865, a similar correlation for the calculated values was 0.863. Such coincidences show a high tightness of the relationship between the calculated and actual values, between which the pair correlation coefficient turned out to be 0.996. In practical terms, the high closeness of the connection allows using the appropriate statistical modeling to regenerate the source data in case of their insufficient representativeness. As an example, table 1 shows the data regeneration for options from 11th to 20th.

For the calculation of HT3, weights of the considered conditions have values that determine the weighted average labor conditions presented in the first column of table 1. To calculate the corresponding weights, Figure 1 shows the input of the corresponding expert estimates.

| \(B_p=\) | \(B_M=\) | \(B_T=\) | \(B_K=\) | answer option: |
|----------|----------|----------|----------|----------------|
| 0.400    | 0.100    | 0.400    | 0.100    |
| \(B_p\) \(B_M\) \(B_T\) \(B_K\) |
| \(=>\) a little bit greater |
| \(<=\) a little bit smaller |
| \(=>\) greater |
| \(<\) smaller |
| \(=>\) more greater |
| \(>=\) more smaller |
| \(=\) equal |
| \(0\) i dont know |

**Figure 1.** Examination of weights (significance) of individual working conditions
As an example, the following expert answers are presented, based on the method of pairwise estimation of weight ratios. The values of the provision of skilled workers are considered equal to the values of the transportation of materials. At the same time, the security of workers and materials is considered to be much more significant in comparison with the security of machines and the influence of weather conditions. This result can be represented as the following expression:

\[ B_P = B_T \quad \text{and} \quad B_P \gg B_M \quad \text{and} \quad B_P \gg B_K \]  

(2)

As a result of calculations implemented according to the method described in scientific articles [13, 14], the corresponding weights were obtained: \( B_P = 0.4; B_M = 0.1; B_T = 0.4 \) and \( B_K = 0.1 \).

For the final determination of the rate of labor costs, the linear trend extrapolation procedure was used, which was determined according to the calculation of the three-layer perceptron as a function of the growth of working conditions, as shown in Figure 2.

![Figure 2. A linear trend graph of the calculated labor costs versus the average value of fulfilling working conditions.](image)

The intersection point of the linear trend and the value of the full supply of working conditions, which corresponds to an abscissa of 100\%, shows the desired value of the norm of labor costs. As a final algorithm, we give a step-by-step procedure for the regression of the formation of labor costs.

A lot of construction objects are determined, on which there is executive documentation for the chosen type of work.

According to the executive documentation, the actual labor costs are determined, referred to the unit volume of work performed.

According to the adopted scale, which should be the same for the entire set of objects, assessed (in percent) the security of the actual conditions of work.

Taking into account the normalization of the entered evidence, a training array is formed.

The training array is processed in a neural network formed on the basis of a three-layer perceptron, as a result of which data are obtained that approximate the dependence of labor costs on the conditions for their receipt.

If there is insufficient representativeness of the obtained data, then data regeneration based on statistical modeling is used.

An extrapolation procedure is carried out, based on the determination of the linear trend of the calculated values of labor costs and the security of fulfilling the conditions for organizing the labor process. The intersection point of the trend line and the value of the full security of the conditions of the organization of the labor process shows the desired value of the norm of labor costs.

4. Sensitivity analysis

We proceed to show the specific results of using the developed algorithm. The data presented in table 2 show the reproducibility of the calculation results, which determines the degree of closeness to each other of the independent results obtained in specific conditions.
10 repeated calculations of labor costs (HT3). The 1st line is the calculation when entering the expression (2), the 2nd is the calculation with additional data regeneration, the 3rd line is the calculation with \( B_T = B_M \) and \( B_P \ll B_K \).

| N | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|   | 0,630| 0,646| 0,672| 0,628| 0,66| 0,692| 0,623| 0,632| 0,646| 0,636|
| 2 | 0,676| 0,665| 0,606| 0,67| 0,68| 0,708| 0,688| 0,654| 0,672| 0,658|
| 3 | 0,67| 0,706| 0,672| 0,72| 0,67| 0,717| 0,652| 0,682| 0,724| 0,679|

The relative standard deviation, which is also defined as the coefficient of variation, is taken as a reproducibility criterion. The results of calculating labor costs for the 10 implementations presented turned out to be fairly close, and their relative error was less than 3%.

At the same time, the calculation time carried out in the macro program and implemented in the environment of the Excel table processor on an economy class laptop was less than three seconds. The first column shows the results of reproducibility without data regeneration, and the second with additional data regeneration. As can be seen from the results shown, the difference between the average values was less than 5%, with a coefficient of variation of less than 4%. As mentioned above, the error requirements established by the developers of the Finnish labor standards RATU are limited to 10% [2].

A much larger error in the result of the regression determination of labor costs is introduced by the error associated with the determination of percent compliance with the regulatory conditions for the implementation of the labor process. For example, if we reduce all the initial data for assessing percent compliance with regulatory conditions by 10%, then the calculated standard for labor costs will change by about the same amount.

From this it follows that in order to obtain more relevant results when using the proposed method, it is necessary to combine it with such expert assessment methods, the application of which provides the calculation of errors [13, 14].

5. Summary
To reduce the costs associated with the creation of a database of labor costs in construction, an algorithm for neural network regeneration of norms was developed based on the assessment of relevant data, which are formed according to the executive documentation of completed construction projects in organizations involved in their implementation, as well as expert assessments of the conditions of construction production.

References
[1] Bolotin, S.A., Kotovskaya, M.A. Analysis of the European and Russian regulatory bases of labor costs as applied to construction planning [Text] / S.A. Bolotin, M.A. Kotovskaya // Bulletin of civil engineers. - 2013. - No. 2 (37). - S.98-103.
[2] Porshneva, L.G. Finnish experience: cadres decide everything! / L.G. Porshneva // Bulletin of the National Association of Builders. - 2011. - No. 8 (15). - S.57-133; [Electronic resource]. – URL: http://сроамур.рф/doc/nostroy_bull/15.pdf. (Date of appeal: 30.10.2012).
[3] ENIR. Uniform Base Cost Indexes and Unit Prices for Construction, Installation, and Renovation - General Provisions. (USSR Gosstroy 43/512/29-50)
[4] Zotkina N.S. Organization, regulation and remuneration in construction. Uch. pos. St. Petersburg: Publishing House of the St. Petersburg State University of Economics and Finance, 2003, 202 p.
[5] Callan R. Basic concepts of neural networks: Tran. from English M : Publishing House "Williams", 2001.
[6] N.R.Draper, H.Smith. Applied Regression Analysis. Second Edition. John Willey & Sons, 1987.
[7] Boussabaine A.H. The use of artificial neural networks in construction management, Construction management and Economics, 1996.
[8] https://helpiks.org/3-10999.html
[9] Josh Patterson, Adam Gibson. Deep Learning. A Practitioner’s Approach. O’Reilly. 2017
[10] Bolotin S.A., Dadar A.X, Malsagov A.R. Prediction of construction time based on executive schedules, statistical and neural network modeling. Real estate: economics, management. 2017. No3, p. 59-63.
[11] Bengio, 2012. Practical Recommendation for Gradient-Based Training of Deep Architectures// https://arxiv.org/abs/1206.5533; Muller et al., 2012. Neural Networks: Tricks of the Trade, Second Edition.
[12] Rotachev A.G., Pavlov K.I. Permitting, executive and acceptance documentation in construction. Uch. pos. Novosibirsk: Publishing House of the Siberian State. University of Communication 2012.75 p.
[13] Bolotin S.A. Dadar A.X. The choice of the sensitivity of the scale for assessing the quality of the organization of construction // Izv. university. Construction, No. 9, 2010. SS. 34-38.
[14] Bolotin S.A. Dadar A.X. Determination of the error of the qualiometric assessment of the weights of additive quality indicators of construction schedules // Izv. university. Construction, No. 2, 2010. SS. 29-33.