NEURAL NETWORK MODEL FOR ESTIMATING CONSTRUCTION PRODUCTIVITY[a]

Discussion by H. Yucel Ersoz,[3] Associate Member, ASCE

Labor costs make up a large portion of the total cost of a construction project. Therefore, success of a construction company in today's competitive market largely depends on accurate estimation of productivity, which is no easy task. Productivity in construction is greatly affected by work conditions that change from project to project. A good estimate of productivity requires a careful analysis of work conditions and their impact on productivity.

It is difficult or impossible to quantify the impact of work conditions on productivity. How much does the performance of a formwork crew improve under favorable weather conditions? How much faster can a very experienced crew erect structural steel? It is quite hard to answer such questions quantitatively. The authors have made an important contribution in the area of cost estimating by presenting an excellent method for analysis of impact of work conditions on productivity.

Although it is a good solution to the problem at hand, the method presented is not without problems. The discusser feels that various aspects of the procedure need to be discussed and clarified.

Let us begin with a brief examination of how impact of work conditions on productivity can be analyzed. Assume that there are n factors that affect productivity of a construction crew, i.e., a formwork crew. The work conditions encountered in each past project can be represented by a vector. Let $\mathbf{A}_k$ be the work conditions vector for the kth project.

Therefore, $\mathbf{A}_k = (x_{k1}, x_{k2}, \ldots, x_{kn})$.

In this case, each x_{ki} is a factor that affects productivity. For example, x_{k1} might be weather conditions encountered in the kth project, x_{k2} might be repetitiveness of work in the same project, and so on. Let $\mathbf{B}_k$ denote the productivity in the same project. $\mathbf{B}_k$ is a function of $\mathbf{A}_k$.

Therefore, $\mathbf{B}_k = f(\mathbf{A}_k)$.

What we want to know is what the productivity will be for any given set of working conditions. We are not necessarily interested in the parametric expression of the function f. A neural network is just the tool we need to achieve this goal. The authors have made a good choice.

The neural network selected is perfectly trained if it can estimate $\mathbf{B}_k$ correctly for all past projects by the end of the training. This can happen if the following conditions are fulfilled:

1. All factors that affect productivity are considered.
2. All training pairs are consistent. In other words, if input vectors of two training pairs are the same, expected output should also be the same.
3. It is mathematically possible to find the function f with the selected network type.

Let us discuss these conditions on the following examples. Consider the following training pairs

$$\text{Training pair 1: } \mathbf{A}_1 = (2, 4, 3, 1, 7, 9) \geqq \mathbf{B}_1 = 15.5$$

$$\text{Training pair 2: } \mathbf{A}_2 = (2, 4, 3, 1, 7, 2) \geqq \mathbf{B}_2 = 12.3$$

$$\text{Training pair 3: } \mathbf{A}_3 = (2, 4, 3, 1, 7, 9) \geqq \mathbf{B}_3 = 13.4$$

If we omit the sixth factor

$$\mathbf{A}_1 = (2, 4, 3, 1, 7) \geqq \mathbf{B}_1 = 15.5$$

$$\mathbf{A}_2 = (2, 4, 3, 1, 7) \geqq \mathbf{B}_2 = 12.3$$

Although $\mathbf{A}_1$ and $\mathbf{A}_2$ appear to be the same, $\mathbf{B}_1$ and $\mathbf{B}_2$ are different. The training pairs are inconsistent and we have ''noise'' in the training data. It will be more difficult for the network to estimate productivity accurately by the end of the training. This is a violation of Condition 1. Consider training pairs 1 and 3. They have the same input, but different outputs. Condition 2 has been violated. Again, training data is noisy.

The conditions mentioned above pose some problems for our analysis. Construction workers are not machines, always behaving the same way under the same conditions. Even under apparently identical work conditions, different productivity values might be obtained. It is unrealistic to expect to be able to input into the neural network all factors that affect a worker's productivity. Furthermore, experts' opinions about the work conditions are subjective. Different experts will have different opinions about the work conditions encountered. In fact, it may or may not be possible for each expert to be perfectly consistent in his own evaluations. It is almost impossible to fulfill either of the first two conditions. Therefore, the authors' probabilistic approach seems to be the best choice. We do have more problems, however:

To analyze the variation of productivity, the authors propose to divide the likely range of productivity values based on historical data into segments, which they call ''productivity zones.'' For example, a productivity range between 0.1 m.hrs/sq.ft and 0.9 m.hrs/sq.ft can be broken into four productivity zones, each one being 0.2 m.hrs/sq.ft wide. How large should the productivity zones be? The larger the zones, the easier it will be to train the network. This convenience will, of course, come with a price: if we use larger zones, our estimates for future projects will be more uncertain. When more data becomes available in the future, the zones might be narrowed and the neural network weights can be ''fine tuned.'' What are the authors recommendations on selecting the zone width?

With large productivity zones, the likelihood values assigned to neighboring zones during training will most probably have to be reduced, because productivity will be less likely to fall into a neighboring zone. Is trial and error the only way of finding the best likelihood values for neighboring zones during training? If zones are narrowed as more data becomes available in the future, finding the right values might require quite a bit of effort.

Using Kohonen's approach, we must pay attention to more than one output node. When is the network trained enough? What should be the error function? Sum of squares of errors at output nodes might be a good indicator. Could the authors clarify this?

Although they do not make any difference in the final result, the negative values in Table 5 are somewhat disturbing. For example, the neural network is telling the user that the likelihood of having a productivity value in zone 7 is -0.1. Semantically, this does not seem to be a valid statement. Could this have been avoided?

The results of the estimate produced by the prototype model are summarized in Table 8. (Note: Table 7 appears to contain data from zones 1, 2, 3, and 4. In the example used to demonstrate the prototype model, the authors selected only Zones 2, 3, and 4 for the productivity estimate. It seems that Column

[a]December 1997, Vol. 123, No. 4, by Jason Portas and Simaan AbouRizk (Paper 14120).

[3]7600 Wood Hollow Dr., Apt. 1018, Austin, TX 78731.

J. Constr. Eng. Manage., 1999, 125(3): 212-213

1 should be labeled ''Zone 2,'' Columns 2 and 3 should be labeled ''Zone 3,'' and Column 4 should be labeled ''Zone 4'' in this table. This is consistent with the zone ranges defined in Tables 4 and 6.) It appears that the authors treated all three zones equally while calculating the results. That is, the mean and standard deviation values presented in Table 8 were calculated using all the data in Table 7. This creates a problem, because zone 4 has a smaller effect on the estimate than does zone 1 although it was assigned a larger likelihood value. It makes more sense to pick data points from each zone in proportion to the likelihood value assigned to that zone. In this example, we could get a better estimate by picking three data points from zone 2 for every five data points we pick from zone 4 and for every eight data points we pick from zone 3. Then, the estimate would reflect the different likelihood values of each zone. The data points could be picked from each zone randomly. Of course, the more random data points we pick, the more representative the estimate will be. As long as the number of data points picked from a zone is proportionate to its likihood value, picking the same data point more than once, if needed, should not be a problem.

The biggest problem that will be encountered once a system like this is implemented by a construction company is the consistency of experts' subjective opinions. The values that should be assigned to the neighboring nodes during training depend on the consistency of experts' input. If there is a high employee turnover in a company, then experts using the system are likely to input more inconsistent data to the system. Would that be a problem in the future? What are the authors' opinions on that? A set of descriptive guidelines that explains how work conditions should be evaluated by estimators and field personnel might be helpful. If the company develops such guidelines, inconsistencies can be minimized and the system can produce good estimates. For example, if job site access is a factor that affects the productivity in an activity, pictures taken at different project job sites with various site access conditions might be included in the guidelines so that ''limited site access'' or ''somewhat limited site access'' can be interpreted the same way by everyone.

As indicated by the authors, the company's information management system should be rearranged to provide the input to the network more readily. Then the input would be more reliable than experts' recollections of a project that was completed 10 years ago by a project manager who does not work for the company anymore. On the other hand, an estimator who has just been hired by the company can look at the evaluation guidelines and the work conditions noted for the project, and make use of that information in his/her next estimate.

The authors should be commended for their attempt to solve the difficult problem of analyzing how productivity of a construction crew is affected by work conditions. The method presented is sound and its implementation is feasible. However, it should be kept in mind that the system will produce good results only if the construction company using it makes the necessary changes discussed here in its information management system.

Closure by Simaan AbouRizk,[4] Member, ASCE

The issues brought up by the discusser (Ersoz 1998) are not unique to this paper. In reality, most artificial neural networks development involves a significant amount of work in data collection and analysis. This process is iterative, and requires discipline to insure that the data sets are accurate, consistent, and representative of the system being modeled to the best possible degree.

The work presented in the paper was our second attempt at implementing neural networks for modeling labor productivity in a real industrial setting. The first relied exclusively on quantitative data and was not successful. In reality, companies do not (and perhaps should not) collect all the information found to have an impact on labor productivity. Given the nature of technical papers published in professional journals it is not possible to detail all aspects of the work conducted. The reader was, therefore, referred to the MSc thesis that was completed by Portas (1996) for the details. Many of the issues brought up by the discusser are discussed in Portas (1996).

Notwithstanding the above, the discussion brings about important points that will be addressed in this closure in a summarized manner.

1. Noise in training the data: It is true that in some cases the same input may lead to slightly varying productivity. The fact is that one cannot capture all factors that may impact productivity, hence this issue comes up but can be dealt with in two ways:
   - Review of data for consistency and integrity, and make necessary adjustments when problems are identified; and
   - Introduce fuzziness in the output to account for these fluctuations when no problems with the data can be further identified.

   Portas (1996) discusses how both approaches were utilized in our study, and the difficulties encountered in the process.
2. Selection of zone width can be a matter of trial and error as the writer indicates. This is no different, however, than the approach one would utilize to set up the neural network structure. For example, how many hidden layers, how many nodes in each layer, transfer function, etc., are all set based on loose guidelines, experience, and eventually trial and error. More recently we have implemented a Kohonen self-organizing network for that purpose with some success. The results are too premature to publish, however.

   As for ''paying attention to more than one output node,'' the weights can be distributed at the discretion of whoever is conducting the training. The approach we implemented is more like a supervised Kohonen Network, where the output classes are known. The approach is flexible enough to allow specifying the neighborhood zone as well as changing the gain term when justified.
3. The negative results can be avoided by using a transfer function that maps on the interval [0, 1] rather than the interval [−1, +1]. The latter interval was used in our implementation, as it produced better training results. The negative values can be simply neglected or masked away from the user in the recall phase since the program was developed in-house.
4. The weighted average can be calculated in a number of different ways, including the one suggested by the discusser (which, as stated, will give a more accurate prediction of the actual weighted average). In this implementation the weighted average was not encouraged as an indicator for the estimator to use, because it was desired that they will exert their judgment within the boundaries established by the zone prediction. One of the

[4]Prof. of Constr. Engrg. and Mgmt., Univ. of Alberta, 220 Civil Engineering Building, Edmonton, Alberta, Canada T6G-2G7.

J. Constr. Eng. Manage., 1999, 125(3): 212-213

main dangers of using the weighted average is obvious with a bimodal distribution of the output, for example. In that case the weighted average would be quite misleading.

5. The consistency of experts is a problem regardless of whether one is implementing a neural network or any other model. Although this is a concern, it can be easily managed, as we demonstrated by undertaking this study with a major contractor. In enhancing this approach, Knowles (1998) completed a study that involved a comprehensive sensitivity analysis of the network described in this paper. Among other issues addressed by Knowles was more detailed breakdown of the ''complexity'' factors in order to improve the consistency of answers regarding the issue of complexity, for example.

   Consistency of experts, however, should not deter one from undertaking such an approach, as it can be overcome by further refinement of the model (i.e., more factors, and more quantitative descriptors as opposed to qualitative ones), properly educating the person supplying the information as well as those collecting it, and finally, incorporating the person supplying the information as part of the training exercise when possible (i.e., the superintendent in our case).

In conclusion, the issues raised by the discusser are all important to consider when implementing the approach described in the paper.

It should be noted that the approach described in the paper was recently applied at a major industrial contracting company to predict productivity for pipe welding, fitting, handling. Both this implementation and the one described in the paper showed marked improvement over currently used methods. Prior to developing these neural networks, we benchmarked accuracy of predictions of productivity by estimators for the tasks we were to model. Once a marked improvement over currently employed methods is determined, a sensitivity analysis is conducted. The approach is then deployed to be used parallel to other existing tools while its performance is assessed. This process yields a number of issues that need to be addressed, including the ones indicated by the discusser. Only after all of the identified issues are resolved does one implement the tool in industry.

MODEL FOR CONSTRUCTION BUDGET PERFORMANCE—NEURAL NETWORK APPROACH[a]

Discussion by Sandra L. Weber[5]

In their paper describing the use of an artificial neural network to model construction budget performance, the authors clearly present the background concepts and describe the data that were utilized to train the model and test its predictive performance. However, several conceptual and procedural questions arise from the paper.

[a]September 1997, Vol. 123, No. 3, by D. K. H. Chua, Y. C. Kog, P. K. Loh, and E. J. Jaselskis (Paper 14320).

[5]Assoc. Prof., Del E. Webb School of Construction, Arizona State Univ., Tempe, AZ 85287-0204.

In the study, RMS classification error (CE), a mathematical tool that eliminates human judgment, was used as a measure of how well the neural network model predicted actual results. However, it would seem that the relationship between classification error values and levels of actual performance should be stated. In the description of the model development, it is stated that with 27 unscaled input variables the resulting CE was 1.225 while, when the variables were scaled, the resulting CE was 0.905. What is the physical difference between a CE of 1.225 and a 0.905? How does a 1.198 CE for predicting previously unseen data compare with chance?

Also not addressed in the paper are why the number of input variables was reduced from 27 to 8 and the reasons for eliminating several of the variables. A variable reduction criterion based on sensitivity analysis is discussed. Table 3 presents the output change caused by a 5% change in the input variables. The table gives the output sensitivity for all 27 input variables. From the described process, those variables identified as having the greatest percentage change should be those chosen for use in the reduced eight-variable model. The authors state that the eight selected factors ''exhibited at least 10% changes in the network output when a +5% dithering of input values was made.'' This statement regarding the selection process for the eight-variable model seems unsupported by the data presented in Table 3. Three variables—*activities in execution plan*, with an 11.13% effect; *budget contingency*, with a 12.69% effect; and *safety inspection*, with an 11.22% effect—were not included in the eight-variable model. At the same time, three other variables identified as having effects below 10% were included in the eight-variable model. One of these, *control system budget*, had an effect of only 0.14%.

Fig. 4(a) presents the predictive performance of the eight-variable model when tested with the ''learning'' set using 20 of the original 22 input samples and forming 380 output measurements. The predictive performance appears to have a CE of 0.387, which is equivalent to 15% of the measurements being misclassified. When the trained network was presented with 20 ''test validation'' samples [Fig. 4(b)], a CE of 0.837 was achieved, with 40% of the samples misclassified. The authors make the statement that the error is smaller in the test validation sample than the scaled 27-input-variable model ''since some spurious factors that have led to inconsistent mapping have been eliminated.'' It appears that the word ''spurious'' has been used loosely. Considering Table 3, the logical analysis process described has not been followed in developing this supposedly smaller error model.

The discussion supporting the use of scaled values as inputs is very interesting. The authors argue that there should be little output effect for a 1% change in the input variable value for the *drawings completed before construction* example. This supporting evidence example is fairly clear because the limits are clearly defined as 0–100%. The other supporting evidence example is very different, however. The example is of the difference between *construction experience* of 15 or 16 years. In the case of a 30-year total career, the percentage difference is only 3.3%. This is a small change similar to that seen with completed drawings, which is likely the authors' point. However, in this example the magnitude of the effect is a function of the career duration to date, rather than the total career, as the reader is lead to believe. What would the output effect be in the case of managers with one year as opposed to two years of experience? The magnitude of the difference in experience is of only one year, similar to the authors' 15 and 16 year example. However, with a one to two year change in experience, there is a 50% increase in total experience. There seems to be a mistaken assumption that all integer differences, regardless of scale, have the same magnitude of effect.

J. Constr. Eng. Manage., 1999, 125(3): 212-213

Closure by D. K. H. Chua,[6] Member, ASCE, Y. C. Kog,[7] P. K. Loh,[8] and E. J. Jaselskis,[9] Associate Member, ASCE

The writers thank Prof. Sandra L. Weber for her interest and her remarks on their work.

As noted by Prof. Weber, the RMS classification error (CE) ''was used as a measure of how well the neural network model predicted actual results'' in the paper. Indeed it is the writers' intention to use the CE values to compare the prediction accuracy of the models. A greater CE value would indicate greater prediction error. Accordingly, a model with a CE value equal to 0.905 would be regarded as able to predict more accurately than one with a CE value of 1.225. On the average, the former is able to predict with less than one classification away from the actual classification, while the latter will predict with more than one classification away from the actual classification.

With regard to the CE value for predicting previously unseen data (including cases with incomplete information) by chance, additional investigation conducted has found that the CE value is 1.802 if the predicted classification is determined randomly using Monte Carlo simulation (with 100 cycles). Based on analytical computation the CE value is 1.895. The results have been obtained with the assumption that the chance of getting (predicting) any of the five classifications for a given outcome is 20%, or 1/5. With either approach, the CE value for classification by chance is much greater than 1.198, which is the CE value if classifications are predicted by the model.

The model for construction budget performance has been developed incrementally. As explained briefly in the paper, the idea is to identify the key determinants by first eliminating those factors that tend to be the ''effects'' of construction problems rather than the ''causes'' of budget performance, and also by removing those less significant factors. The average percent of output change caused by a 5% change in the inputs presented in Column 2 of Table 3 was obtained at the end of the first stage of the experiment. Altogether, five stages of experiments were performed. At the end of each stage of the experiment, the input values were increased by 5% through a built-in function in the software used, and a few factors showing negative influence and the less important factors were discarded. In between the first and last stages, the output changes caused by 5% input changes vary as the neural networks were to find a different optimal set of connection weights through the learning process for each model configuration. The values presented in Column 3 of Table 3 were obtained at the end of the fifth experimental stage. All key determinants identified at the end of the experiment show more than 10% on output change with 5% change in input value as described in the paper. To ensure that an optimal final model is obtained, the 19 factors removed earlier were added again to the final model, one at each time. It has been found that in the ''8 + 1'' models, the ''effect'' factors still show negative influence on construction budget performance while the less important factors show insignificant changes on output when their input values were changed by 5%. Therefore, the writers arrived at the conclusion that the final model comprising the eight key deterimants is an optimal one for the set of data used in the study. The findings are also similar to that in a separate investigation based on analytic hierarchy process (AHP) method (Chua et al. 1998). It follows that the writers considered the other 19 factors as ''spurious'' in the sense that they are not genuine in view of the final model.

The classification ranges for the conversion of objective measurements into scaled integer values have been established after informal discussions with a few experienced practitioners having up to 20 years' experience in the industry. In determining the classification ranges, the practitioners were asked ''if the factor can be divided into six scales, what is the likely range of measurements that corresponds to each of these scales?'' The writers agree with Prof. Weber's point that ''with a one to two year change in experience, there is a 50% increase in total experience.'' However, it must be realized that this 50% increase in total construction experience would still yield no significant difference in terms of the project manager's contribution toward construction budget performance. The writers realized from the onset that it is very difficult, if not impossible, to establish a set of perfect classification ranges. Nevertheless, for the purpose of classification there must be some thresholds set at certain points. In order to reduce the effect of subjectivity in setting the cutting points for the classification ranges, the concept of fuzzy membership degrees has been introduced into the classification process in a subsequent extension of the model (Chua et al. 1998).

[6]Sr. Lect., Dept. of Civ. Engrg., Nat. Univ. of Singapore, Singapore 119260.

[7]Adjunct Assoc. Prof., Dept. of Civ. Engrg., Nat. Univ. of Singapore, Singapore 119260.

[8]Res. Asst., Dept. of Civ. Engrg., Nat. Univ. of Singapore, Singapore 119260.

[9]Assoc. Prof., Civ. and Constr. Engrg. Dept., Town Engrg. Build., Room 450, Iowa State Univ., Ames, IA 50011.

APPENDIX. REFERENCE

Chua, D. K. H., Kog, Y. C., and Loh, P, K. (1998). ''A study on construction project success.'' *Res. Rep. No. RP950633*, Dept. of Civ. Engrg., National University of Singapore.

J. Constr. Eng. Manage., 1999, 125(3): 212-213