Hacettepe Journal of Mathematics and Statistics
Volume 39 (3) (2010), 439 – 447

A NEW ANN TRAINING APPROACH FOR EFFICIENCY EVALUATION

Sabri Erdem[*†] and Ipek Deveci Kocakoc[‡]

Received 21 : 05 : 2009 : Accepted 25 : 02 : 2010

Abstract

In this study, we propose a new Artificial Neural Networks (ANN) training approach that closes the gap between ANN and Data Envelopment Analysis (DEA), and has the advantage of giving similar results to DEA and being easier to compute. Our method is based on extreme point selection in a bandwidth while determining the training set, and it gives better results than the traditional ANN approach. The proposed approach is tested on simulated data sets with different functional forms, sizes, and efficiency distributions. Results show that the proposed ANN approach produces better results in a large number of cases when compared to DEA.

Keywords: Data envelopment analysis, Efficiency evaluation, Artificial neural networks, Training set selection.

2000 AMS Classification: 62 M 45, 90 C 08.

1. Introduction

All organizations aim essentially to survive whether they are profit or non-profit organizations. In today's global environment, one of the competitive strategies is low-cost leadership in the market. It is important to measure how resources are efficiently used between organizations. The more efficiently the resources are used, the higher competitive advantage in low cost leadership the organization has. That is why measuring the efficiency and making the efficiency comparisons properly is very important. Relevant measurement techniques are focused on measuring efficiency relatively because of the hardness of calculation of the real efficiency value that the organization can have. Consequently, there are numerous methods, both parametric and non-parametric, to compare the relative efficiency of organizations.

[*]Dokuz Eylul University, Faculty of Business, Business Administration Department, Quantitative Methods Division, Izmir, Turkey. E-mail: `sabri.erdem@deu.edu.tr`

[†]Corresponding Author.

[‡]Dokuz Eylul University, Economics and Administrative Sciences Faculty, Econometrics Department, Dokuzcesmeler Kampus, 35160 Buca, Izmir, Turkey. E-mail `ipek.deveci@deu.edu.tr`

Some problems may occur when using parametric methods like regression analysis, due to their underlying assumptions. Firstly, regression methods can estimate the production frontier by means of average value of the response variable. It causes the boundary efficiencies found in regression to be lower than the possible production frontier. Secondly a regression model generally considers one response variable, contrary to our multiple outputs. When compared to parametric methods, non-parametric methods like Data Envelopment Analysis (DEA) are more flexible, and their utilization is becoming widespread. DEA is based on the idea of a Frontier Production Function, which was first introduced by Farrell in 1957, as opposed to the Average Performance used in the Econometric Literature. This idea was then developed and formed as DEA by Charnes *et al.*[3]. A comprehensive literature study on DEA can be found in Emrouznejad *et al.* [6].

Data Envelopment Analysis (DEA) is a non-parametric extreme point method that compares each decision-making unit (DMU) with only the best ones. Most production functions are non-linear as a whole. DEA is based on estimating the frontier production function in a piecewise form, and uses the linear programming method as in Model (1.1). The frontier production function is estimated as an efficiency boundary formed by efficient production units.

$$\min_{\theta},_{\lambda} \theta \tag{1.1}$$

subject to

$$\begin{aligned} -y_i + Y\lambda &\geq 0, \\ \theta x_i + X\lambda &\geq 0, \\ \lambda &\geq 0. \end{aligned}$$

Model (1.1) shows an input oriented CRS model, where θ is a scalar and λ is a $Nx1$ vector of constants. The value of θ will be the efficiency score for the ith DMU. It will satisfy $\theta \leq 1$, with a value of 1 indicating a point on the frontier and hence a technically efficient DMU according to Farrel's definition[7]. Here the linear programming problem must be solved N times for each DMU in the sample. A value of the efficiency score (θ) is then obtained for each DMU. One of the basic problems when interpreting the DEA is that these efficiency estimations based on a single data set are very sensitive to noise and errors. In order to overcome this drawback, new approaches to assess the DEA efficiency values are also a current research interest.

In the last decade, researchers focused on efficiency estimation via Artificial Neural Networks (ANN), as well as applying DEA. Although ANN models perform the efficiency measurement with respect to an average, as for regression methods, and do not try to estimate the frontier production function, they have the advantage of building the non-linear models well. This means that ANN cannot estimate the frontier production function like DEA does, since it uses the average approach, but is efficient in non-linear models. Different modifications have been suggested to overcome this drawback; however studies have been unable to find any common solution accepted by everyone [9].

The ANN approach used by Athanassopoulos and Curram [1] to analyze efficiency can be regarded as one of the pioneer applications. It was similar to regression analysis between the input and output variables, and the model obtained had a regression curve that passed through the input-output set. This model produced worse results than DEA because of the estimation problem of the efficiency frontier. Later, various ANN approaches have been suggested by many researchers to get closer to the efficiency frontier. Costa and Markellos [4], Shale *et al.* [11], Santin *et al.* [9], and Delgado [5] are some researchers who have compared DEA with different ANN approaches using different methods. Santin *et al.* [9] presents an experimental study which compares various approaches such as Stochastic Frontier Analysis (SFA), Ordinary and Corrected Ordinary Least Squares (OLS and COLS), Constant and Variable Return to Scale (CRS and VRS respectively). A recent paper published by Santin [10] suggests using a thick frontier approach, by adding the maximum residual term to the average output predicted by the MLP for each DMU, but to different segments of the distribution of the dependent variables. However, none of the studies used the same simulation schema or model, so it is not possible to compare the results of all these approaches.

In this study, we offer a new ANN training approach to efficiency estimation that is easy to apply and produces similar outputs to DEA. Our approach is based on the estimation of the frontier production function by including the extreme points within a certain bandwidth in the training set. In fact, the method of selecting the extreme points in this study is inspired by the block maxima method that was initiated by Gumble [8] in extreme value theory studies. The block maxima method is based on the limit behavior of normalized maximum of a random sample [2], and considers the maximum value generally within blocks of time intervals. This paper uses a different method for determining the bandwidths of the blocks.

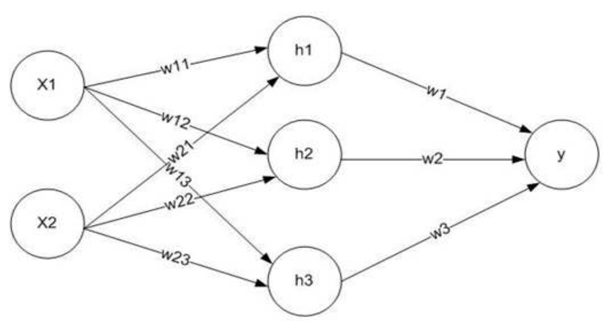
We propose to train the ANN by a selected sample of decision units. In this study we use an ANN that is multi-layered, feed forward, back propagated and multi-perception. Computations and simulations are executed using computer programs coded in MATLAB.

The rest of the paper is organized as follows: Section 2 describes the ANN, Section 3 introduces the proposed method, Section 4 shows the generating data and compares alternatives, and finally Section 5 concludes the paper by presenting conclusions based on our work.

2. Artificial neural networks

ANN's are inspired by the biological nerve system, and the idea was launched in 1943 by logiest Walter Pitts and neurophysiologist Warren McCulloch to get a new viewpoint for real life problems [13]. ANN, with its powered structure, has become one of the most popular non-parametric techniques, especially for non-linear models. ANN's are differentiated by considering the purpose of usage and structure of the problem such as optimization, forecasting, pattern recognition, clustering and classifying. The multi-layered multi perception, feed forward and back propagated type of ANN [14] shown in Figure 1 has a single hidden layer and models the relation between y as output variable, and x_1 and x_2 as input variables. In this model h_1, h_2, and h_3 show the neurons in the hidden layer.

Figure 1. A multi-layered feed forward back propagated ANN model



It is possible for an ANN to learn by changing the initial weights; w_{ij}, ($i = 1, 2, \ldots,$ number of input variables, $j = 1, 2, \ldots,$ number of neurons in the hidden layer) and w_j in an iterative way, using the back propagation of error that occurs between the observed output data and its estimated value. ANN is assumed to have "learned" when the overall error between observed and estimated output decreases to a certain tolerance level. Learning performance is decided by the performance of the ANN on the test dataset that are not used during the learning step.

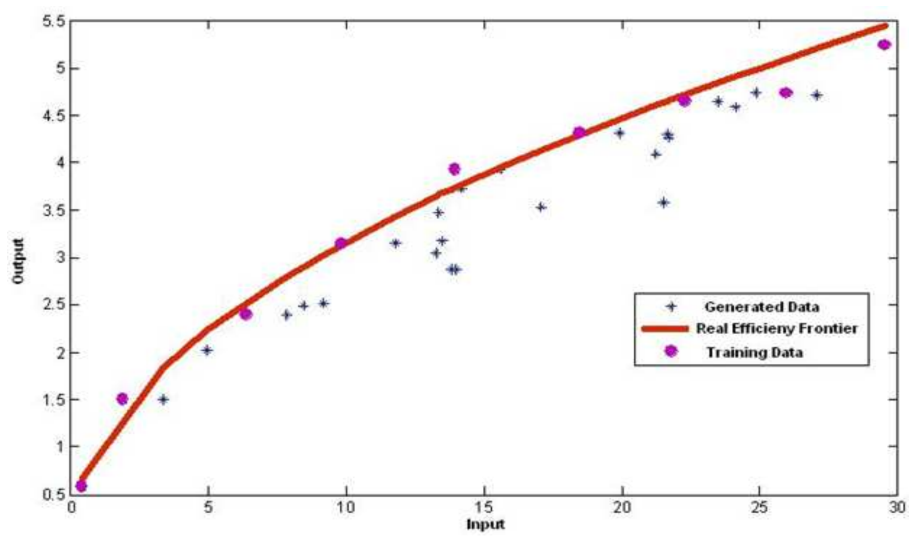
Another important point about the multi layered ANN is the method of activation in the hidden layer. It receives weight data from inputs and transforms them into weight values between hidden layers and output layers. These transformations are achieved by processing the received data by means of a certain activation function and transmitting it to the output layer. In this transformation step the sigmoid activation function is mostly preferred as in Equation 2.1.

$$f(x) = \frac{1}{1 + e^{-\sum x_i w_{ij}}}. \tag{2.1}$$

3. The proposed ANN training approach for DEA

In this study our proposed approach is different from the traditional ANN-for-DEA approaches in that we aim to estimate the real frontier product function. Our proposed ANN (PANN) approach takes the mean value of the input variable and corresponding maximum output value in a certain bandwidth into the training set, and also the extreme points are added to the set. Here the prepared dataset is used to train the ANN. Thus it is possible to estimate an efficiency frontier above the real dataset. The real efficiency frontier with a generated dataset of 25 observations and prepared training set can be seen in Figure 2. Our training set is very near to the real efficiency frontier. Thus, PANN results are expected to give good estimates of efficiency. The ANN Model constructed in this study has single input and output variables, 5 hidden layer neurons, and sigmoid as the activation function.

Figure 2. Appearance of the learning set



4. Generating data sets and estimation of efficiencies

Table 1 shows the structure of cases used for the simulation. Data are generated for 16 different cases, and results are calculated and presented in Table 2.

Table 1. Case Structure

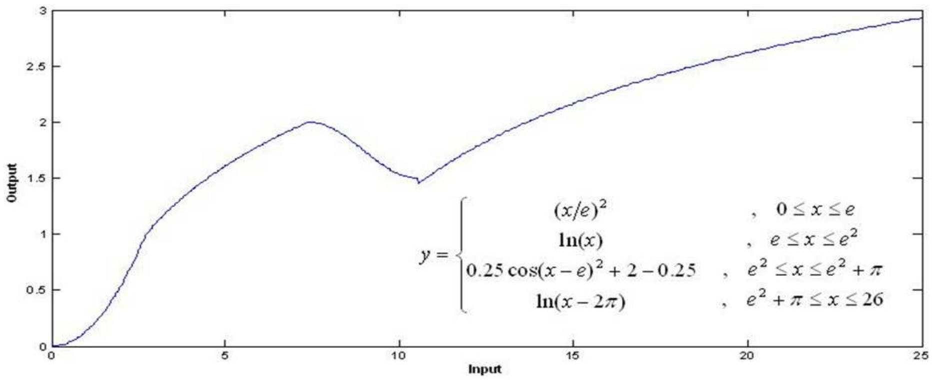
Factor	**Level**	**1**	**2**
Shape of the real efficiency frontier and returns to scale	2	Cobb-Douglas ($y = x^{0.5}$) (Decreasing returns to scale)	Non-linear $y = \begin{cases} (\frac{x}{e})^2 \\ \ln(x) \\ 0.25\cos(x-e)^2 \\ \qquad +2-0.25 \\ \ln(x-2\pi) \end{cases}$ (Partial-both increasing and decreasing-returns to scale)
Sample data size	2	25	100
Percentage of the efficient decision making units	2	0	25
Inefficiency distribution	2	Low inefficiency $U_i \sim$ Half Normal(0.10,0.1)	High inefficiency $U_i \sim$ Half Normal(0.25,0.2)
Total number of cases	16 (2^4)		

A Cobb-Douglas production function ($y = x^{0.5}$) and a non-linear piecewise continuous production function that has a certain shape and function (Figure 3) are considered for the shape of the real efficiency frontier.

This non-linear production function was used by Santin *et al.* [9] to show that their algorithm works well even for complicated production functions. It has two returns-to-scale and decreases and increases at different points. This is not typical of most DEA production functions, which like the Cobb-Douglas function usually only decrease.

On the other hand, it is expected that the ANN approach will produce better results for these kinds of function. By utilizing these two functions, we intended the study to cover different economies of scale in the underlying production or cost function being explored.

Figure 3. Piecewise non-linear production function



Data for the Cobb-Douglas production function is generated from the distribution $X \sim$ Uniform$(0, 30)$, and the non-linear production function from $X \sim$ Uniform$(0, 26)$. The variable X is regarded as the input variable, and is used to find the real production frontier values (Y^{eff}).

Vectors (U) that contain inefficiency are generated from a Half-Normal distribution. For the cases of low inefficiency, the mean and standard deviation are taken to be low. We take ($Y_i = Y_i^{\text{eff}} \times U_i$), which shows the output values are obtained from the variable Y by multiplication by inefficiency values. To reflect the "percentage of efficient decision making units" in the data set, output values that belong to a determined percentage of units are replaced by Y_{eff}. Therefore, the resulting data set involves the units that have perfect efficiency ($E_i = 1$) in some percentage.

In a traditional ANN approach all the data set is used for training purposes. The main contribution of our PANN approach is the utilization of a different data set selection method for ANN training. In the PANN approach, we use only pre-determined data with regard to a bandwidth. Training data is selected for each data set by determining a suitable bandwidth for that data set. In determining the bandwidths, $h = 1.06\sigma n^{-0.2}$ [12], where h is bandwidth, σ the standard deviation of the inputs and n the total number of inputs, is used. This formula is frequently used for density estimation as well. In order to cover the effect of extreme values, we included the minimum and maximum input values in the training data.

After completing the learning step, input variables enter both traditional ANN and PANN, and estimations of the output variables $\hat{Y}^{\text{ANN}}$ and $\hat{Y}^{\text{PANN}}$ are calculated. Estimated output values are divided by output values to calculate the efficiency scores (E_i) as in Eq. (4.1) and (4.2).

$$(4.1) \qquad E_i^{\text{PANN}} = \frac{Y_i}{\hat{Y}_i^{PANN}},$$

$$(4.2) \qquad E_i^{\text{ANN}} = \frac{Y_i}{\hat{Y}_i^{ANN}},$$

$$(4.3) \qquad \text{MAD} = \frac{1}{n}\sum_{i=1}^{n} \left| E_i^{\text{real}} - E_i^{\text{computed}} \right|.$$

For each data set, the mean absolute deviations (MAD) of each approach are calculated by using Eq. (4.3) to compare the performance of the three approaches given in Table 1.

The approach that results in the lowest MAD values is emphasized in Table 2 by giving the values in bold.

Table 2. Results

Data Set No	Prod. Func. Shape	Data Size	Efficient	Inefficient Dist. No.	MADDEA	MADANN	MADPANN
1	C-D	25	0	Low	**0.0619**	0.1193	0.0692
2	C-D	25.	0	High	**0.0871**	0.2481	0.1421
3	C-D	25	25	Low	**0.0632**	0.0933	0.0685
4	C-D	25	25	High	**0.1115**	0.1942	0.1264
5	C-D	100	0	Low	0.0734	0.1171	**0.0526**
6	C-D	100	0	High	0.0988	0.2519	**0.0978**
7	C-D	100	25	Low	0.0724	0.0963	**0.0595**
8	C-D	100	25	High	0.1322	0.1982	**0.0842**
9	N-L	25	0	Low	**0.0635**	0.1195	0.0716
10	N-L	25	0	High	**0.0947**	0.2555	0.1533
11	N-L	25	25	Low	**0.0589**	0.0890	0.0687
12	N-L	25	25	High	**0.0630**	0.0947	0.0877
13	N-L	100	0	Low	0.0740	0.1177	**0.0614**
14	N-L	100	0	High	0.1000	0.2519	**0.0971**
15	N-L	100	25	Low	0.0721	0.0940	**0.0550**
16	N-L	100	25	High	0.0720	0.0946	**0.0563**
Average of MAD values					0.0824	0.1561	0.0856
Success rate for large data size					0	0	100%
Overall success rate					8/16=0.50 50%	0/16=0 0%	8/16=0.50 50%

C-D = Cobb-Douglas, N-L = Non-linear

When the results in Table 2 are evaluated, it can be seen that the proposed ANN approach produces very similar results to DEA. Both DEA and PANN give the best results in 8 out of 16 data sets. Thus, they have a success rate of 50% each.

On the other hand, the traditional ANN method gives the worst results for all of the data sets as expected. Lower MAD values also show that the efficiency frontier estimated by our ANN approach is close to the real efficiency frontier.

As we can see from the average MAD values, the traditional ANN approach produces the worst estimation (with an average of 0.1561), and the proposed ANN approach produces a very approximate estimation (with an average of 0.0856) to DEA.

We also conducted pairwise comparisons of MAD values by the Mann-Whitney U-test under the assumption of nonnormal distribution of MAD values (Table 3). It can be seen that the proposed method gives similar MAD values to DEA ($p = 0.616$), and is statistically better than ANN ($p = 0.001$).

Table 3. Results of pairwise comparisons of MAD values

$(i - j)$ comparison	Mean Rank of i	Mean Rank of j	Asymp. Sig. (two tailed)	Result (α=0.05)
DEA-ANN	10.53	22.47	0.000	Significant difference, DEA gives lower MAD
DEA-PANN	17.38	15.63	0.616	No difference
ANN-PANN	21.88	11.13	0.001	Significant difference, PANN gives lower MAD

All results show that the PANN approach can be used for efficiency estimation. For small (25 unit) data sets, DEA gives better results than PANN. The disadvantage of the bandwidth method we used is that it decreases the training sample size. When we train the network with a small sample, learning will not be sufficient. For bigger sets, however, the size of the training set is enough for learning and PANN gives better results. For small and scattered data sets like 2 and 10, the PANN results are not close to the DEA results. This means that if data set is small and has high variability, ANN learning gets harder.

5. Conclusions and Discussions

In this study, we suggest a new approach for ANN estimation of efficiency. A classical Cobb-Douglas type production function and a mixed production function are utilized as functional forms containing increasing and decreasing economies of scale. Half Normal and Exponential distributions are used for inefficiency distributions. Small and large sized data sets are evaluated by simulation.

The model is developed for only the single input and single output case, but can easily be extended to multiple input and output cases. We utilized a single input-single output case since there is no available method in the literature to generate multiple input-multiple output data for efficiency analysis. The only comparison on multiple input-multiple output cases can be made by using real life data, but in this case MAD values cannot be calculated since the true frontier values would not be known. Therefore, in order to provide scientific fairness, this paper aims to confirm the proposed model by simulation studies before trying to use it on real life data. An extension of this paper might be a real life case study which has multiple inputs and outputs, but this is beyond the scope of the present paper.

Our approach produces better results than traditional ANN for estimating the efficiency frontier. The suggested approach also estimates better than DEA in large data sets. Since solving linear programming models of DEA for large data sets is complicated, ANN can be used for efficiency estimation. However, more improvements to the PANN approach should be made to accommodate all the information DEA gives about inefficient units. Although we used only one ANN structure, different ANN models, different learning rates and different data normalization techniques may be tried to improve the results for small data sets.

References

[1] Athanassopoulos, A. D. and Curram, S. P. *A comparison of data envelopment analysis and artificial neural networks as tools for assessing the efficiency of decision making units*, Journal of the Operational Research Society **47** 8, 1000–1016, 1996.

[2] Beirlant, J., Goegebeur, Y., Teugels, J., Segers, J., De Waal, D. and Ferro C. *Statistics of Extremes: Theory and Applications* (John Wiley and Sons, 2004).

[3] Charnes, A., Cooper, W. W. and Rhodes, E. *Measuring the efficiency of decision making units*, European Journal of Operational Research **2** (6), 429–444, 1978.

[4] Costa, A. and Markellos, R. N. *Evaluating public transport efficiency with neural network models*, Transportation Research C **5** (5), 301–312, 1997.

[5] Delgado, F. J. *Measuring efficiency with neural networks. An application to the public sector*, Economics Bulletin **3** (15), 1–10, 2005.

[6] Emrouznejad, A., Parker, B. R. and Tavares, G. *Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA*, Socio-Economic Planning Sciences **42**, 151–157, 2008.

[7] Farrell, M. J. *The Measurement of Productive Efficiency*, Journal of the Royal Statistical Society **120**, 253–281, 1957.

[8] Gumbel, E. *Statistics of Extremes* (Columbia University Press, New York, 1958).

[9] Santin D., Delgado, F. J. and Valino A. *The measurement of technical efficiency: A neural network approach*, Applied Economics **36**, 627–635, 2004.

[10] Santin, D. *On the approximation of production functions: A comparison of artificial neural networks frontiers and efficiency techniques*, Applied Economic Letters **15** (8), 597–600, 2008.

[11] Shale, E. A., Athanassopoulos, A. D and Curram, S. P. *A frontier-based neural network for assessing the efficiency of activity units*, Foundations of Computing and Decision Sciences **32**, Forthcoming.

[12] Silverman, B. W. *Density Estimation for Statistics and Data Analysis* (Chapman and Hall, London, 1986).

[13] Veelenturf, L. P. J. *Analysis and Applications of Neural Networks* (Prentice Hall, UK, 1995).

[14] Zhang, X. S. *Neural Networks in Optimization* (Springer, 2000).