Available online at http://jnrm.srbiau.ac.ir
Vol.1, No.4, Winter 2016
ISSN:1682-0169

Journal of New Researches in Mathematics

Science and Research Branch (IAU)

Using Neural Network to Determine Input Excesses, Output Shortfalls and Efficiency of Dmus in Russell Model

D. Modhej[a], M. Sanei[b*], N. Shoja[c]

[(a,b)] Departments of Applied Mathematics, Islamic Azad University, Central Tehran Branch, Tehran, Iran
[(c)] Department of Mathematics, Firoozkooh branch, Islamic Azad University, Firoozkooh, Iran

Received Autumn 2015, Accepted Winter 2016

Abstract

Data Envelopment Analysis (DEA) has two fundamental approaches for assessing the efficiency with different characteristics; radial and non-radial models. This paper is concerned the non-radial model of Russell which is a non linear model. Conventional DEA for a large dataset with many inputs/outputs would require huge computer resources in terms of memory and CPU time. Artificial Neural Network (ANN) is one of the most popular techniques for non linear models and for measuring the relative efficiency of a large dataset with many inputs/ outputs. Also in the last decade researches focused on efficiency evaluation via DEA as well as using ANN. In this paper we will estimate the input excesses and the output shortfalls in addition to efficiency of Decision Making Units (DMUs) in Russell model through ANN. The proposed integrated approach is applied to an actual Iranian bank set; the result indicates that it yields a satisfactory solution.
works.

Keywords: *Artificial Neural Network, Data Envelopment Analysis, Russell model, Efficiency, Input excesses, Output shortfalls*

*. Corresponding author: masoudsanei49@yahoo.com

1. Introduction

Data Envelopment Analysis (DEA) is a mathematical programming to evaluate the Efficiency of homogeneous Decision Making Units (DMUs) with multiple inputs and multiple outputs. Over the last decade DEA has been utilized in a wide range to assess the performance of organizations in public and private sectors such as hospitals, banks, schools, etc .Different DEA models are extended and used in practice because of different empirical axioms and corresponds to different characteristics of the production frontiers and production possibility set. As a matter of fact, the DEA models are commonly categorized into two types with distinguishing qualities, namely, the radial and non radial models. The radial models based on the assumption that inputs or outputs change proportionally, However, this assumption in reality needs care because data set may be substitutional and don't change proportionally. The CCR [4] and the BCC [2] models as the first works in DEA are radial models. In the return, non radial models are applied when inputs and outputs may change non-proportionally. A series of non-radial DEA models such as the additive [5], the Russell [7, 11] and the SBM [12] models have been developed in the traditional DEA framework. Artificial Neural Network (ANN), modeled after human thinking paradigm, acquires knowledge through iterative learning process and weight adjustment between interconnected neurons. In doing so, ANN achieves generalization and abstract learning from limited set of information and provides nonlinear mapping and predictive power. In the last decade, researchers focused on efficiency estimation via ANN, as well as applying DEA. Applications of ANN and DEA do not need strict assumptions. This advantage amplifies their implication areas. In literature it is straight forward to find proposed models which apply DEA and ANN together. As a recent instance of DEA-ANN application, Karamali et al. [9] examined the capability of ANN in sensitivity analysis of the parameters of efficiency analysis in DEA. Wu et al. [14] integrated DEA and ANN to examine the relative branch efficiency of a big Canadian bank. The results in the whole were comparable with the normal DEA. Emrouznejad et al. [6] combined ANN with DEA to introduce an approach to estimate efficiency of DMUs in large data sets. The results indicated that the ANN-DEA prediction for efficiency score appears to be a good estimate for the majority of DMUs. Azadeh et al. [1]

proposed algorithm to assess the impact of personnel efficiency attributes on total efficiency through DEA, ANN and Rough Set Theory (RTS). Wang [13] proposed a non-parametric efficiency analysis method based on the adaptive ANN technique. The proposed computational method is able to find a stochastic frontier based on a set of input –output observational data. Simulation experiments demonstrated the neural-network-based method would be effective as adaptive non-parametric efficiency analysis. Celebi et al. [3] explored a novel integration of neural network and DEA for evaluation of suppliers under incomplete information of evaluation criteria. The above literate, exhibits the applicability of DEA and ANN together are reliably. In fact, DEA and ANN can be applied together to support each other or be possibility for each other.

In this study we try to measure input excesses, output shortfalls and efficiency score of each DMU in Russell model through ANN. This method proposes for a large dataset which needs a long time and huge system computer and offers considerable computational savings.

The rest of this paper is organized as follows. In section 2 Russell model is explained. Section 3 describes Artificial Neural Network by back propagation algorithm. BP algorithm for solving Russell model is introduced in section 4. This is followed by numerical example in section 5. Conclusion remark appears in section 6.

2. About Russell model

Development of the axiomatic foundations of efficiency measurement began with Fa¨re et al. [7], who suggested three desirable axioms for efficiency indices: homogeneity, monotonicity, and indication of efficient input/output vectors. To meet the axiomatic foundations, [7] introduced an input-oriented non-radial model which was later extended in [8] into a jointly aggregate measure of output and input efficiency. The DEA model extended by Fa¨re [8] is not reduced to the LP formulation. The efficiency measure of the model is referred to as Russell Measure (RM). Then the model is revisited by Pastor et al. [11] (referring to it as the Enhanced Russell Measure (ERM)). Specifically, suppose that there are n DMUs indexed by j=1, 2… n. Each DMU consumes m different inputs and produces s different outputs, the observed input and output vectors of DMU_j denoted by $Xj=(x_{1j},x_{2j},\dots,x_{mj})$ and $Y_j=(y_{1j},y_{2j},\dots,y_{sj})$,all component of vectors X_j and Y_j are non-negative and each DMU at least has one positive input and output. The following non-linear programming is the ERM:

$$\min\ r=\frac{\frac{\sum_{i=1}^{m}\theta_i}{m}}{\frac{\sum_{r=1}^{s}\varphi_r}{s}}$$

$$s.to$$

$$\begin{aligned}
&\sum_{j=1}^{n}\lambda_j x_{ij}\le\theta_i x_{io} && i=1,2,\ldots,m \qquad (1)\\
&\sum_{j=1}^{n}\lambda_j y_{rj}\le\varphi_r y_{ro} && r=1,2,\ldots,s\\
&\lambda_j\ge 0 && j=1,2,\ldots,n\\
&0\le\theta_i\le 1;1\le\varphi_r && \forall i,r.
\end{aligned}$$

θ_i,φ_r indicate the input excesses and the output shortfalls, respectively. The following properties are considered as important in designing this model.

(P1) The measure should be invariant with respect to the unit of measurement of input and output data. (Units invariant)

(P2) The measure should be monotone decreasing in each input and output slack. (Monotone)

(P3) The measure should be invariant under parallel translation of the coordinate system applied. ([10])

Specially, the numerator and denominator values of ERM model, which are also available, can be separately interpreted as the average input efficiency and the average output inefficiency, respectively, with $r^*=1$ if and only if full efficiency is attained.

Definition 1 (Full ERM-efficiency): DMU_o is full ERM-efficient if and only if at an optimum:

1) $\theta_i^*=1\ \ i=1,2,...,m$

2) $\varphi_r^*=1\ \ r=1,2,...,s$

The $\sum_{j=1}^{n}\lambda_j=1$ can be added in above model, if the constant return to scale does not exist.

3. Artificial Neural Network

ANN is an information-processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information.

Although ANNs arose to model the brain, they have been applied when there is no theoretical evidence about the functional form. In this way, ANNs are data-based, not model-based. The key element of this paradigm is the novel structure of the information-processing system. It is composed of a large number of highly interconnected processing elements working in unison to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition, function approximation, data classification and so on in different areas of science.

The ANN is composed of processing elements (nodes or neurons) and connections. The nodes are interconnected layer-wise of interconnection among themselves. Each node in the successive layer receives the inner product of synaptic weights, with the outputs of the nodes in the previous Layer. The operation of a neuron is shown in Fig. 1

Each node (X_1, X_2,…, X_n) has an output signal connected to each of the other nodes. Each connection is assigned a relative weight. A node's output depends on the specified threshold and transfer function F(X). The ANN has been shown to be effective for addressing the complex nonlinear problem.

Two types of learning networks are, respectively, supervised and unsupervised. For a supervised learning network, a set of training input vectors with a corresponding set of target vectors is trained to adjust the weights in an ANN. For an unsupervised learning network, a set of input vectors is proposed; however, no target vectors are specified. In this study, a supervised learning network is more suitable for the multi-response with censored data estimated problem. Among the several well-known supervised learning ANNs are back-propagation (BP), learning vector quantization, and counter propagation network. The BP model is the most extensively used and can provide better solutions for many applications. Therefore, the BP model is selected in this study. A BP neural network consists of three or more layers, including an input layer, one or more hidden layers, and an output layer. Fig. 2 illustrates a basic BP neural network with three layers.

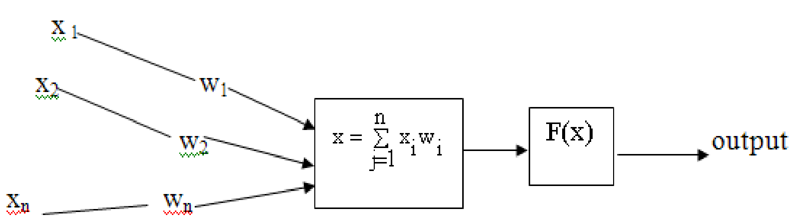

Fig.1

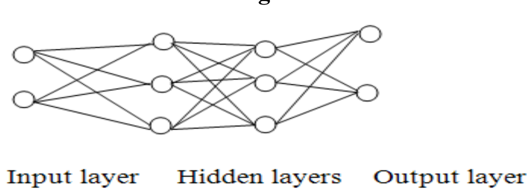

Fig.2

Input layer Hidden layers Output layer

Selection of the number of neurons in hidden layer is important because the neural network' mapping accuracy and ability to generalize from the training data significantly relies on the number of units in this layer. A network with small hidden neurons cannot completely distinguish the structure present in the training sample. Meantime large number of neurons in hidden layer makes system memorize the data, memorization of the system causes high error in test results and prevents to generalize the results for the other sample data. BP neural network learning works on a gradient-descent algorithm. The BP neural network initially receives the input vector and directly passes it into the hidden layer. Each element of the hidden layer is used to calculate an activation value by summing up the weighted input, and the sum of the weighted input will be transformed into an activity level by using a transfer function. Each element of the output layer is then used to calculate an activation value by summing up the weighted inputs attributed to the hidden layer. Next, a transfer function is used to calculate the network output.

In case there is a difference between the desired output and the output produced by the network, the connection weights should be altered and adjusted so as to minimize the Mean Squared Error (MSE) as follows: The actual network output is then compared with the target value. The error evaluation method that is used in this study is mean square error (MSE).

$$MSE = \frac{\sum_{i=1}^{n}(y_i - e(i))^2}{n} \tag{2}$$

n: The number of training sample

$e(i)$: The expected output

y_i: Real output related to sample training

BP neural network algorithm refers to the propagation of errors of the nodes from the output to the nodes in the hidden layers. These errors are used to update the weights of the network. After the knowledge representation is determined, the BP neural network will be trained to attempt the prediction behavior.

4. Back-Propagation Russell model

In this section, ANN approach will be proposed to estimate the input excesses, the output shortfalls and the efficiency scores of DMUs in the Russell model. To do so, we design a new net work as follow: Back-propagation Russell model learns by iteratively processing a training sample. For each training sample, the weights are modified so as to minimize the MSE between the network's prediction and actual result as obtained in a conventional

Russell model. Generally before training can begin in any neural network, the user must decide on the network topology by specifying the number of variables in the input layer, the number of hidden layers (if more than one), the number of variables in each hidden layers, and the number of variables in the output layer. The inputs and outputs in the corresponding Russell model are used as variables in the input layer and the input excesses, the output shortfalls and the efficiency scores are defined as the variables in the output layer. In the present paper linear activation function is used in the output layer and sigmoid function is used in the hidden layer as follows:

$$F(\mathbf{x}) = \frac{1}{1+e^{-\mathbf{x}}} \tag{3}$$

The initial values of the weights may also affect the resulting accuracy. Once a network has been trained and its accuracy is not considered acceptable, it is recommended to repeat the training process with a different network topology or a different set of initial weights.

Neural network training could be made more efficient by performing certain preprocessing steps on the network inputs and targets. Network input processing functions transforms inputs into better form for the network use. The normalization process for the raw inputs has great effect on preparing the data to be suitable for the training. Without this normalization, training the neural networks would have been very slow. There are many types of data normalization. In this paper all values in the dataset has to be changed to contain values in the interval [0, 1]. The normalization is often accomplished by using a linear interpretation formula such as

$$x_n = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{4}$$

Where: x_i = original input or output value

x_n =normalized value of input or output value

x_{min} =minimum original input or output value

x_{max} =maximum original input or output value

Finally the analysis is conducted in three stages,

(1) Training ANN Russell model with a sample of DMUs,

(2) Testing ANN Russell with another sample of DMUs, and

(3) Estimating the input excesses, the output shortfalls and the efficiency scores of DMUs using the generated ANN Russell model.

Table1 : Inputs and outputs

Inputs		Outputs	
Input1	Personnel	Output1	Facilities
Input2	Payable interest	Output2	The total sum of four main deposits
Input3	Deferred receivables	Output3	Received interest

Table 2 : Summery statistic of input values

	Inputs		
	Personnel	**Payable interest**	**Deferred receivables**
Max	88.15	513160000000	1064400000000
Min	2.26	41603512	1482400
Average	14.28	8395600000	18453000000
Standard deviation	10.17	25090000000	76816000000
Median	11.59	3535700000	2500200000

Table 3 : Summery statistic of output values

	Outputs		
	Facilities	**Sum of deposits**	**Received interest**
Max	8818600000000	10857000000000	875880000000
Min	2087400000	9327600000	569445
Average	124830000000	124790000000	10444000000
Standard deviation	454580000000	464040000000	45925000000
Median	47275000000	62697000000	3197200000

5. Numerical Example

In this section, we consider 200 branches of an Iranian bank with three inputs and three outputs. Table 1 shows the labels of these inputs and outputs.

Tables 2 and 3 relate a summary of the statistical properties for inputs and outputs. Because of the large number of branches, it is endeavored to apply ANN of the previous section to measure input excesses, output shortfalls and the efficiency scores in Russell model. At first input data are normalized.

Then, branches are partitioned into two parts, training set and validation set. 75% of branches are used as a training set which determine the optimal network parameters, and the remaining 25% are used as a validation set to evaluate the network generalization capability.

After following pre-specified epochs or satisfying condition on MSE, the network learns non-linear mapping between inputs and outputs of the system. Subsequent to the network training, remnants of branches are considered to test the network. Parameters of the estimated neural network are presented in Table 4.

Statistical results of input excesses, output shortfalls and the efficiency scores in Russell model by ANN are given in Table5.

Table 4 : Estimated neural network parameters

Concept	Result
Number of neurons / input-hidden-output	6-15-7
Activation function / hidden/output	Sigmoid/linear
Epochs (max)	2000
Mean square error	0.01
Learning rate	0.6

Table 5 : Summery statistic of Russell model by ANN

	efficiency	input excesses			output shortfalls		
	Γ	θ_1	θ_2	θ_3	φ_1	φ_2	φ_3
Max	1.0110	0.9631	1.0021	1.0130	8.3800	5.7900	14.8200
Min	0.1000	0.2600	0.2400	0.0600	0.9741	1.1033	0.9835
Average	0.4338	0.9095	0.6387	0.5705	1.8897	2.0397	2.5845
Standard deviation	0.2638	0.1787	0.2564	0.3066	1.1592	1.1956	1.9837
Median	0.3350	1.00	0.5700	0.5400	1.5150	1.6050	1.9800

Table 6 : Regress analysis for bank branches

Parameter	
Slope	0.975
Intercept	0.015
R^2 coefficient	0.958

From Table 5, it can be seen that efficiency scores for certain bank branches are more than 1. This is not voidable in the DEA context. It is not, however, a surprising result for DEA-ANN, since ANN with statistical properties generate a stochastic frontier according to efficient DMUs. The adaptive learning capability of Russell model can be observed by strong correlation between the predicted and actual Russell results as summarized in Table 6.

5. Conclusion

DEA is a non-parametric technique which is widely used for evaluation the efficiency. In Russell model from input excesses and the output shortfalls we can measure the efficiency of each DMU. DEA for a large dataset with many input/output variables and/or many DMUs would require huge computer resources in terms of memory and CPU time. This paper compounded a neural network with DEA to set up an algorithm to evaluate the Input excesses, the output shortfalls and the efficiency of DMUs in large data set. The proposed algorithm has been tested on 200 Iranian banks. The results turn out that the prediction of ANN-Russell model shows to be a fruitful estimate for the majority of DMUs. Also, analysis of error appears that the larger dataset have the smaller error.

References

[1] Azadeh, A., Saberi, M., Tavakkoli Moghaddam, R., Javanmardi, L., 2011. An integrated Data Envelopment Analysis–Artificial Neural Network–Rough Set Algorithm for assessment of personnel efficiency. Expert Systems with Applications, 38 (3), 1364-1373.

[2] Banker,R.D., Charnes,A., Cooper, W.W.,1984.Models for the estimation of technical and scale inefficiencies in data envelopment analysis. Management Science,30,1078-1092.

[3] Celebi, d., Bayraktar, d., 2008 .An integrated neural network and data envelopment analysis for supplier evaluation under incomplete information. Expert Systems with Applications, 35 (4), 1698-1710

[4] Charnes, A., Cooper, W.W., Rhodes, E., 1978. Measuring the efficiency oh decision making units. European Journal of operational Research, 2(6), 429-444.

[5] Charnes, A., Cooper, W.W., Golany, B., Seiford, L.M., Stutz, J., 1985. Foundations of data envelopment analysis and Pareto–Koopmans empirical production functions. Journal of Econometrics 30, 91– 107.

[6] Emrouznejad, A., Shale, E.A., 2009. A combined neural network and DEA for measuring efficiency of large scale data sets. Computer and industrial Engineering , 56, 249-254.

[7] F¨are, R.S., Lovell, C.A.K., 1978. Measuring the technical efficiency of production. Journal of Economic Theory, 19, 150-162.

[8] Fa¨re, R.S., Grosskopf, S., Lovell, C.A.K, 1985. The Measurement of Efficiency of Production. Boston: Kluwer Academic Publishers

[9] Karamali, L., Memariani, A., Jahanshahloo G.R.,2014. ANN-DEA Integrated Approach for Sensitivity Analysis in Efficiency Models. Iranian Journal of Operations Research, 4, 14-24.

[10] Pastor, J .T., 1996. Translation invariance DEA: A generalization. Annals of Operations Research, 66, 93-102.

[11] Pastor, J.T., Ruiz, J.L., Sirvent, I., 1999. An enhanced DEA Russell graph efficiency measure. European Journal of Operational Research, 115, 596–607.

[12] Tone K., Tsutsui M., 2009. Network DEA: A slacks-based measure approach. European Journal of Operational Research, 197,243–52

[13] Wang, S., 2003. Adaptive non-parametric efficiency frontier analysis: a neural-network-based model. Computer Operation Research, 30,279–295.

[14] Wu, D., Yang, z., Liang, L . 2006. Using DEA-neural network approach of evaluate branch efficiency of a large Canadian bank. Expert Systems with Applications, 31, 108-115.