World Applied Sciences Journal 6 (9): 1177-1189, 2009 ISSN 1818-4952 © IDOSI Publications, 2009

# The Use of Artificial Neural Networks for Technology Selection in the Presence of Both Continuous and Categorical Data

## Reza Farzipoor Saen

Department of Industrial Management, Faculty of Management and Accounting, Islamic Azad University-Karaj Branch, P.O. Box: 31485-313, Karaj, Iran

**Abstract:** The objective of this paper is to propose a new use of Artificial Neural Networks (ANNs) for technology selection in the presence of both continuous and categorical data. To indicate the relative importance of the ANN inputs to the result of the network, a methodology for sensitivity analysis is presented. A numerical example demonstrates the application of the proposed method.

Key words: Technology selection % Artificial neural networks % Backpropagation algorithm

## **INTRODUCTION**

Selecting the right technology is always a difficult task for decision makers. Technologies have varied strengths and weaknesses which require careful assessment by the purchasers. Technology selection models help decision maker choose between evolving technologies. The reason for a special focus on technology selection is due to the complexity of their evaluation which includes strategic and operational characteristics.

A relatively new technique in the field of continuous and categorical modeling is the use of Artificial Neural Networks (ANNs). ANNs were successfully applied to a variety of problems, varying from image processing to bankruptcy prediction. Now that ANNs are becoming more and more widely accepted and outperform some classical continuous and categorical techniques in certain cases, it is clear that a technology selection model on the basis of ANN is an option which has to be considered.

Neural networks are technologies that acquire the ability to learn for the computers. They teach the relationships between the inputs and outputs of the events to the computers by using patterns. By the help of taught data various generalizations are made, similar events are interpreted, required decisions are made and related problems are solved. ANNs are generally the software systems that imitate the neural networks of the human brain. It is also possible to accept the ANNs as a parallel distributed data process system. ANNs can be applied successfully in learning, relating, classification, generalization, characterization and optimization.

In this paper, the ANN is used for technology selection as it has following strengths:

- <sup>C</sup> In general, there are a large number of computation activities to be performed at the same time during technology selection process. An ANN is inherently parallel and naturally amenable to expression in a parallel notation. Therefore, it is a superior method in technology selection process.
- C ANN handles both continuous and categorical data.
- C ANN produces good results even in complicated domains.
- C ANN is available in many off-the-shelf packages.

The objective of this paper is to propose a new use of ANNs for technology selection in the presence of both continuous and categorical data. In addition, to indicate the relative importance of the ANN inputs to the result of the network, a methodology for sensitivity analysis is introduced.

This paper proceeds as follows. In Section 2, literature review is presented. Section 3 discusses the proposed steps for technology selection. Numerical example and concluding remarks are discussed in Sections 4 and 5, respectively.

Corresponding Author: Reza Farzipoor Saen, Department of Industrial Management, Faculty of Management and Accounting, Islamic Azad University-Karaj Branch, P.O. Box: 31485-313, Karaj, Iran

Literature Review: Some mathematical programming approaches have been used for technology selection in the past. Shehabuddeen et al. [1] focused on the experience of operationalizing of a framework for technology selection. This is achieved through the application of a software tool, which is based on the structure provided by the framework. They illustrated how theoretical concepts presented in the framework relate to "real-life" technology selection considerations. Al-Ahmari [2] developed a Decision Support System (DSS) to help Decision Makers (DMs) to evaluate and select the proper Computer Integrated Manufacturing (CIM) technologies, based on several quantitative and qualitative factors. Kengpol and O'Brien [3] developed a decision support tool for the selection of advanced technology to achieve rapid product development.

Utturwar *et al.* [4] reduced the computational cost of technology selection by decomposing the process into two smaller sub-problems. They attempted to exploit the structure of the technology compatibility matrix to improve the efficiency of the technology selection process in aircraft design.

Lee and Kim [5] presented a methodology using Analytic Network Process (ANP) and Zero One Goal Programming (ZOGP) for information system projects selection problems that have multiple criteria and interdependence property. Lee and Kim [6] described an integrated approach of interdependent information system project selection using Delphi method, ANP and Goal Programming (GP).

Chan *et al.* [7] presented a technology selection algorithm to quantify both tangible and intangible benefits in fuzzy environment. They described an application of the theory of fuzzy sets to hierarchical structural analysis and economic evaluations. To justify information technology systems, Kulak *et al.* [8] proposed crisp and fuzzy multi-attribute axiomatic design approach.

Malladi and Min [9] showed how an Analytic Hierarchy Process (AHP) model could be utilized to select the optimal access technology for a rural community under a multiple number of criteria. Then, they formulated a mixed integer programming problem that would provide the optimal access technologies for a multiple number of homogeneous communities that were pooling resources such as budgets for fixed and variable costs. Finally, they showed how the problem could be extended to the case of heterogeneous communities where the fixed and variable costs vary among communities. Hajeeh and Al-Othman [10] used AHP to select the most appropriate technology for seawater desalination. Jaganathan *et al.* [11] proposed an integrated fuzzy AHP based approach to facilitate the selection and evaluation of new manufacturing technologies in the presence of intangible attributes and uncertainty.

Parkan and Wu [12] demonstrated the use of and compare some of the current Multiple Attribute Decision (MADM) and performance measurement Making methods through a robot selection problem borrowed from Khouja [13]. Particular emphasis were placed on a performance measurement procedure called Operational Competitiveness Rating (OCRA) and a MADM tool called Technique for Order Preference by Similarity to Ideal Solution (TOPSIS). But, Wang [14] offered comments on Parkan and Wu [12] based on an examination of their proposed OCRA method. Since the premise of the OCRA method is that the cost/revenue ratios must be known, costs and revenues cannot be measured in any units other than dollar value in any practical cases. This property makes the OCRA method faulty. Further, it is shown that the invalid weighting approach used in the OCRA method provides an illusion to management that a cost category with large cost/revenue ratio is more important than a cost category with small ratio. The conclusion is that a performance analysis using the OCRA method can be invalid.

Khouja [13] proposed a decision model for technology selection problems using a two-phase procedure. In phase 1, Data Envelopment Analysis (DEA) is used to identify technologies that provide the best combinations of vendor specifications on the performance parameters of the technology. In phase 2, a MADM model is used to select a technology from those identified in phase 1. Khouja [13] used MADM, to select a robot from the efficient robots. Baker and Talluri [15] proposed an alternate methodology for technology selection using DEA. They addressed some of the shortcomings in the methodology suggested by Khouja [13] and presented a more robust analysis based on cross-efficiencies in DEA. Farzipoor Saen [16] proposed a Minimax Regret-based Approach (MRA) that ranks the most appropriate technologies in the conditions that both ordinal and cardinal factors are present. Again, Farzipoor Saen [17], to select the best technology, introduced a model that is based on Imprecise Data Envelopment Analysis (IDEA). Talluri et al. [18] proposed a framework, which is based on the combined application of DEA and nonparametric statistical procedures, for the selection of Flexible Manufacturing Systems (FMSs). The strengths of this

methodology are that it incorporates variability measures in the performance of alternative systems, provides DM with effective alternative choices by identifying homogeneous groups of systems and presents graphic aids for better interpretation of results. To select the best advanced manufacturing technologies, Karsak and Ahiska [19] introduced a multi-criteria decision methodology that can integrate multiple outputs such as various technical characteristics and qualitative factors with a single input such as cost. Their model is derived from the cross-efficiency analysis, which is one of the branches of DEA model.

However, all the aforementioned references do not consider technology selection through ANNs. The advantages of ANNs over DEA are as follow:

- C No casualty commitment is required regarding the positive or negative influence that individual inputs have on the produced outputs. That is, increasing an input can actually have a negative effect on one or more outputs.
- C Performance targets are assessed for individual inputs and outputs of each Decision Making Unit (DMU) that do not have necessarily the same direction of improvement (i.e. expand or contract).
- C The neural network allows the use of both continuous value and classification input variables without the modeling enhancements that are necessary in the corresponding DEA models.
- C Neural networks employ validation procedures to test the adequacy of the proposed models for unseen DMUs.

**Proposed Steps for Technology Selection:** In the formation of an ANN system, a sustainable neural network model based on the nature of the problem should be selected and a neural network should be constructed according to the characteristics of the application domain. This application process of an ANN model design includes the steps below Choy *et al.* [20]:

**Design Stage:** In this stage, three kinds of variable are to be determined. They are the input and output attributes selection, the selection of the training method and the design of the hidden layer.

**1-Input and Output Attributes Selection:** Input and output selection is always a complex task for the ANN

model developer as there is no formal method for selecting variables for a model. Moreover, both under and over specification of input variables will most often generate suboptimal performance of the ANN model. The result is that on one side, if there are too many input variables, it can bring about poor generalization. On the other side, if there is insufficient amount of information representing critical decision criteria given to the model, then it is unable to develop a correct and accurate model.

ANNs work best when all the input and output values are between 0 and 1. This requires massaging all the values, both continuous and categorical to get new values between 0 and 1. By massaging the data set, the software used to produce and improve the ANN model performs better. Continuous values, such as sales price, range between two known values. Categorical values take one value from a list of values. Marital status, gender, account status, product code and so on are categorical values.

To massage continuous values, the lower bound of the range is subtracted from the value and the result is divided by the size of the range. This basic procedure can be applied to any continuous feature to get a value between 0 and 1. For categorical features, the fractions between 0 and 1 are assigned to each of the categories. For instance, if there are three categories, they would be assigned one to 0, another to 0.5 and the third to 1 (Berry and Linoff [21]).

2-Training Method Selection: After selecting the variables for the input and output, the next step is to determine the kind of training to be employed to best fit the problem. The learning algorithm can be divided into two distinct categories, namely, unsupervised learning and supervised learning. Both require a collection of training examples that enable the ANN to acquire the data set and produce accurate output values. The learning algorithm adjusts the connection weights after each iteration and this process continues until the network converges to a set of values in order to determine all of the inputs correctly. Among the available algorithms, Backpropagation algorithm designed by Werbos [22] and Rumelhart et al. [23] is the most suitable one as it has been extensively tested in many areas. The reasons of selecting Backpropagation type are because of its generality and the ease of implementation for most of the ANN systems. The basic idea of Backpropagation training is to use a gradient-descent approach to adjust and determine weights such that an overall error function such as sum of the squares of the errors can be minimized.

**3-Hidden Layer Design:** The design of hidden layer is dependent on the selected learning algorithm. For example, unsupervised learning methods normally require the quantity of neurons in the first hidden layer equal to the size of the input layer. Supervised learning systems are generally more flexible in the design of hidden layers. An increment of the number of hidden layers enables a trade-off between smoothness and closeness-of-fit. A greater quantity of hidden layers enables an ANN model to improve its closeness-of-fit, while a smaller quantity improves the smoothness or extrapolation capabilities of it. As a heuristic rule, the number of hidden layer neurons can be up to 2n+1 (where *n* is the number of neurons in the input layer).

**Training Stage:** The goal of the training stage is to obtain an accurate ANN model. In the training stage, the selection of the transfer function, learning rate, momentum, exit condition setting, Mean Square Error (MSE) and verification of the model are needed.

**1-Transfer Function:** A transfer function is needed to introduce the nonlinearity characteristics into the network. The nonlinear function will make the hidden neurons of multilayer network more powerful than just plain perception. The transfer function used is a standard function for Backpropagation, that is, the sigmoid transfer function. The sigmoid transfer function is chosen due to its ability to help the generalization of learning characteristics to yield models with improved accuracy. Given the nature of this function, data massaging in [0,1] range improves its performance.

**2-Parameters:** The Backpropagation training paradigm uses three controllable factors that affect the rate of learning of algorithm. They are the learning rate coefficient (**0**), momentum (") and the exit condition.

**Learning Rate (0):** The learning coefficient governs the speed that the weights can be changed over time, reducing the possibility of any weight oscillation during the training cycle.

**Momentum ("):** The momentum parameter controls over how much iteration an error adjustment persists. There is no definitive rule regarding the momentum, ". In general, it is set to 0.5, which is half of the maximum limit for training to reduce the damping effect. **Exit Condition:** ANNs use a number of different stopping rules to control the termination of the training process. One of the stopping rules is "stop it after a specified number of epochs".

**3-Verification:** The residual entropy of the trained network is a measure of its generalization. When the residual entropy increases, the performance of the generalization decreases, meaning that the model still needs modification. The residual entropy is monitored during training by means of MSE. It is the squared error between the output response of network and the training target.

**Generalization stage:** After the training stage, the ANN learning performance is evaluated by running it using the validation data set. There are two steps to be performed in order to accept the model. One is recall and the other is validation.

**Recall:** A validation data set is applied to check the degree of the generalization of the trained model. By doing so, the size of the generalization error is determined and minimized. This step is called Recall.

**Validation:** A network is said to be generalized well when the output is correct or close enough for an input. In such cases, the model is ready for use.

Numerical Example: For illustration purposes, the technology selection approach proposed in this paper is used for robot selection. The data set for this example contains historical data on 100 industrial robots. Table 1 depicts the robot attributes. These data were used to train and test the ANN model. The inputs utilized were cost, vendor reputation and load capacity. Vendor reputation is included as a categorical input. For this categorical input, the fractions between 0 and 1 are assigned to each of the categories. Assume that there are three categories including low reputation, moderate reputation and high reputation. The value 0 would be assigned to low reputation. The value 0.5 would be assigned to moderate reputation. The value 1 would be assigned to high reputation. The output, indicated in the last column of Table 1, is efficiency scores of the robots obtained by CCR model. The CCR model is one of the techniques of DEA<sup>1</sup>. DEA proposed by Charnes et al. [24] (Charnes, Cooper, Rhodes (CCR) model) and developed by Banker et al. [25] (Banker, Charnes, Cooper (BCC) model) is an approach for evaluating the efficiencies of decision making units (DMUs).

	Network in	puts	Network output		
Robot No.	Cost (10000\$)	Vendor	Load	Efficiency (using CCR model)	
1	7.2	0.5	60	0.077	
2	4.8	0.5	6	0.003	
3	5	1	45	0.2	
4	7.2	0.5	1.5	0.069	
5	9.6	1	50	0.104	
6	1.07	0	1	0.002	
7	1.76	0	5	0.006	
8	3.2	0.5	15	0.156	
9	6.72	0	10	0.003	
10	2.4	0	6	0.005	
11	2.88	0.5	30	0.174	
12	6.9	0.5	13.6	0.072	
13	3.2	1	10	0.313	
14	4	1	30	0.25	
15	5.00	0.5	47	0.145	
10	8	0	15	0.004	
18	6.3	1	10	0.159	
19	.94	0.5	10	0.532	
20	.16	0	1.5	0.02	
21	2.81	1	27	0.356	
22	3.8	0.5	.9	0.132	
23	1.25	1	2.5	0.8	
24	1.37	0.5	2.5	0.365	
25	3.63	0	10	0.006	
26	5.3	1	70	0.189	
27	4	0	205	0.109	
28	6	0.5	105	0.108	
29	3	0.5	10	0.1	
31	5	0	23 64	0.027	
32	3	0.5	11	0.167	
33	6	1	27	0.167	
34	9	0	84	0.02	
35	2.5	1	545	0.804	
36	4	1	65	0.25	
37	8	0	45	0.012	
38	2.6	0.5	51	0.205	
39	1	1	58	0.143	
40	4	0.5	54	0.135	
41	0	0.5	0.5	0.07	
42	19	0	8	0.009	
44	3	0.5	45	0.174	
45	5	0.5	245	0.189	
46	4	1	2	0.25	
47	7	0.5	54	0.077	
48	2	0.5	54	0.27	
49	5	1	6	0.2	
50	8	1	38	0.125	
51	4	0	38	0.02	
52	4	0	158	0.084	
53	7	0.5	148	0.106	
54	2	0.5	14	0.071	
55 56	2	0	9	0.01	
50 57	9	0.5	415	0.289	
58	2	1	44	0.5	
59	3	0	55	0.039	
60	5	0.5	65	0.113	
61	4	1	74	0.252	
62	8	0	65	0.017	
63	1	0.5	24	0.5	
64	5	0.5	547	0.318	

Table 1. Dalated attributes for 100 ashets

World Appl	. Sci. J.,	6 (9):	1177-1189,	2009
------------	------------	--------	------------	------

Table 1: Continued				
65	4	1	558	0.51
66	8	0	21	0.006
67	1	0	470	0.1
68	2	0.5	50	0.266
69	4	0.5	587	0.419
70	5	1	60	0.2
71	1	1	70	0.1
72	3	0	890	0.631
73	6	0	5	0.002
74	4	0.5	54	0.135
75	2	1	8	0.5
76	6	1	66	0.167
77	6	0.5	55	0.09
78	4	0	4	0.002
79	5	0	54	0.023
80	9	0.5	21	0.056
81	5	1	321	0.307
82	1	1	12	0.1
83	4	0.5	54	0.135
84	5	0.5	44	0.104
85	2	1	24	0.5
86	3	0	245	0.174
87	4	0	279	0.148
88	5	0.5	654	0.363
89	6	0.5	445	0.229
90	8	1	21	0.125
91	7	1	48	0.143
92	5	0.5	15	0.1
93	2	0.5	35	0.25
94	4	1	45	0.25
95	6	0	12	0.004
96	6	0	55	0.02
97	3	0	56	0.04
98	2	1	54	0.5
99	4	1	45	0.25
100	4	1	14	0.25

Since the efficiency scores of the robots 67, 71 and 82 equal to one, they have been selected as efficient robots that should be bought. However, the other robots have not been selected, because their efficiency scores are less than one. Note that these measures are not exhaustive by any means, but frequently used in robot's performance evaluation. In an application of this methodology, decision makers must carefully identify appropriate inputs and outputs to be used in the decision making process.

To get the values of cost and load capacity between 0 and 1, the lower bound of the range in each column vector is subtracted from the value and the result is divided by the size of the range. Table 2 depicts the massaged values.

The dataset is divided into a training set, test set and validation set. A set of 68 data rows was used in the training of the ANN model. These data were used to build the ANN model. Sixteen data rows were used to test it and sixteen data rows were used to validate it. The test set and validation set were created by randomly selecting data from the Table 2. Table 3 is the summarized testing results of the best network topology with respect to the

Table 2:	The massaged Network int	l values		Network output
Robot				
No. (DMII)	Cost	Vendor reputation	Load	Efficiency (using CCR model)
<u>(Dine)</u> 1	0.745763	0.5	0.066472	0.077
2	0.491525	0	0.005736	0.003
3	0.512712	1	0.049601	0.2
4	0.745763	0.5	0.000675	0.069
5	1	1	0.055224	0.104
6	0.096398	0	0.000112	0.002
0	0.169492	0	0.004611	0.006
8	0.322034	0.5	0.015859	0.156
9 10	0.094913	0	0.005736	0.005
11	0.288136	0.5	0.03273	0.174
12	0.713983	0.5	0.014284	0.072
13	0.322034	1	0.010235	00.313
14	0.40678	1	0.03273	0.25
15	0.372881	0.5	0.05185	0.143
16	0.711864	1	0.088966	0.148
17	0.830508	0	0.015859	0.004
18	0.650424	1	0.010235	0.159
19	0.082627	0.5	0.010235	0.532
20	0 28072	0	0.000675	0.02
21	0.28072	0.5	0.029350	0.132
23	0.115466	1	0.0018	0.8
24	0.128178	0.5	0.0018	0.365
25	0.367585	0	0.010235	0.006
26	0.544492	1	0.077719	0.189
27	0.40678	0	0.229558	0.109
28	0.618644	0.5	0.117085	0.108
29	0.512712	0.5	0.010235	0.1
30	0.300847	1	0.027106	0.333
31	0.512712	0	0.070971	0.027
32 33	0.500847	0.5	0.01130	0.167
33	0.018044	0	0.029350	0.02
35	0.247881	1	0.611967	0.804
36	0.40678	1	0.072095	0.25
37	0.830508	0	0.049601	0.012
38	0.258475	0.5	0.056349	0.205
39	0.724576	1	0.064222	0.143
40	0.40678	0.5	0.059723	0.135
41	0.830508	0.5	0.072095	0.07
42	0.194915	1	0.001237	0.5
43	0.184322	0	0.00/986	0.009
44	0.300847	0.5	0.049601	0.174
45 46	0.312712	0.5	0.274347	0.189
40 47	0.40078	0.5	0.059723	0.25
48	0.194915	0.5	0.059723	0.27
49	0.512712	1	0.005736	0.2
50	0.830508	1	0.041728	0.125
51	0.40678	0	0.041728	0.02
52	0.40678	0	0.176696	0.084
53	0.724576	0.5	0.165448	0.106
54	0.724576	0.5	0.014734	0.071
55 56	0.194915	0	0.00911	0.01
50 57	0.026441	1	0.465/52	0.289
51 58	0.930441	0.5	0.004611	0.056
50 59	0.174913	0	0.040470	0.039
60	0.512712	0.5	0.072095	0.113
61	0.40678	1	0.082218	0.252
62	0.830508	0	0.072095	0.017
63	0.088983	0.5	0.025981	0.5
64	0.512712	0.5	0.614217	0.318

## World Appl. Sci. J., 6 (9): 1177-1189, 2009

Table 2: Continued

65	0.40678	1	0.626589	0.51
66	0.830508	0	0.022607	0.006
67	0.088983	0	0.527612	0.1
68	0.194915	0.5	0.055224	0.266
69	0.40678	0.5	0.659206	0.419
70	0.512712	1	0.066472	0.2
71	0.088983	1	0.077719	0.1
72	0.300847	0	1	0.631
73	0.618644	0	0.004611	0.002
74	0.40678	0.5	0.059723	0.135
75	0.194915	1	0.007986	0.5
76	0.618644	1	0.07322	0.167
77	0.618644	0.5	0.060848	0.09
78	0.40678	0	0.003487	0.002
79	0.512712	0	0.059723	0.023
80	0.936441	0.5	0.022607	0.056
81	0.512712	1	0.360027	0.307
82	0.088983	1	0.012485	0.1
83	0.40678	0.5	0.059723	0.135
84	0.512712	0.5	0.048476	0.104
85	0.194915	1	0.025981	0.5
86	0.300847	0	0.274547	0.174
87	0.40678	0	0.312788	0.148
88	0.512712	0.5	0.734563	0.363
89	0.618644	0.5	0.499494	0.229
90	0.830508	1	0.022607	0.125
91	0.724576	1	0.052975	0.143
92	0.512712	0.5	0.015859	0.1
93	0.194915	0.5	0.038353	0.25
94	0.40678	1	0.049601	0.25
95	0.618644	0	0.012485	0.004
96	0.618644	0	0.060848	0.02
97	0.300847	0	0.061973	0.04
98	0.194915	1	0.059723	0.5
99	0.40678	1	0.049601	0.25
100	0.40678	1	0.014734	0.25

## Table 3: The best topology selection results

				Average
Network		Learning		MSE of
topology	Learning algorithm	rate	Momentum	training
3-4-1	Batch Backpropagation	0.1	0.5	0.00683
3-5-1	Batch Backpropagation	0.1	0.5	0.00943
3-3-1	Batch Backpropagation	0.1	0.5	0.00288
3-4-1	Batch Backpropagation	0.08	0.5	0.00744
3-3-1	Batch Backpropagation	0.08	0.5	0.00843
3-4-1	Batch Backpropagation	0.11	0.5	0.00558
3-4-1	Batch Backpropagation	0.09	0.5	0.00879
3-6-1	Batch Backpropagation	0.1	0.5	0.01057
3-3-1	Batch Backpropagation	0.1	0.7	0.01121
3-3-1	Batch Backpropagation	0.1	0.6	0.00735
3-3-1	Batch Backpropagation	0.1	0.4	0.01481
3-4-1	Incremental Backpropagation	0.1	0.4	0.00958
3-4-1	Incremental Backpropagation	0.1	0.5	$0.00276^{*}$
3-5-1	Incremental Backpropagation	0.1	0.5	0.00553
3-3-1	Incremental Backpropagation	0.1	0.5	0.00769
3-3-2-1	Incremental Backpropagation	0.1	0.5	0.00857
3-2-3-1	Batch Backpropagation	0.1	0.5	0.00317
3-2-3-1	Batch Backpropagation	0.2	0.3	0.00846
3-2-3-1	Batch Backpropagation	0.1	0.6	0.00796
3-2-3-1	Batch Backpropagation	0.2	0.5	0.00364

20 possible different sets of learning algorithm, learning rate, momentum, number of hidden layers and number of neurons in each hidden layer. From Table 3, with the input nodes of cost, vendor reputation and load capacity in the

Table 4: The results of ANN		Table 4: Continued
Robot No. (DMU)	Efficiency (using ANN)	65
1	0.069	66
2	0.054	67
3	0.156	68
4	0.067	69
5	0.063	70
6	0.057	/1
	0.056	12
8	0.165	75
10	0.055	74 75
10	0.055	76
12	00.07	77
13	0.315	78
14	0.227	79
15	0.145	80
16	0.094	81
17	0.053	82
18	0.101	83
19	0.449	84
20	0.059	85
21	0.393	86
22	0.132	87
23	0.763	88
24	0.359	89
25	0.054	90
26	0.145	91
27	0.055	93
28	0.096	94
30	0.050	95
31	0.054	96
32	0.177	97
33	0.111	98
34	0.053	99
35	0.968	100
36	0.24	
37	0.053	Table 5: Average v
38	0.222	Inputs
39	0.09	Cost
40	0.132	Vendor reputation
41	0.064	Load capacity
42	0.555	
43	0.056	input lover one
44	0.180	input layer and
45	0.122	of 0.1 gives the
40	0.07	training). Here
48	0.292	input laver 4
49	0.148	the and the sector of 1
50	0.075	the output laye
51	0.054	Increment
52	0.054	a learning als
53	0.074	sigmoid trans
54	0.069	signold trans.
55	0.056	0.1. Momentun
56	0.183	80000 epochs"
57	0.059	During th
58	0.598	During un
59	0.055	set are repeate
6U	0.101	compares its p
01	0.244	and adjusts a
02 62	0.053	ta alma 1
64	0.181	technology se
<u></u>	0.101	on the training

World Appl. Sci. J., 6 (9):	11	77	-11	89,	2009	
_			~			

90		0.041
82		0.841
83		0.132
84		0.099
85		0.577
86		0.056
87		0.055
88		0.214
89		0.114
90		0.074
91		0.09
92		0.096
93		0.282
94		0.232
95		0.054
96		0.054
97		0.055
98		0.608
99		0.232
100		0.221
Table 5: Average value	s of the inputs and network	output
Inputs	Average value	Network output
Cost	0.45768	0.119609
Vendor reputation	0.5	
Load capacity	0.106282	

0.592 0.053 0.063 0.29 0.299 0.159 0.897 0.0610.054 0.132 0.561 0.116 0.081 0.054 0.054 0.059 0.24

input layer and the  $\{3-4-1\}$  topology with a learning rate of 0.1 gives the best result (minimum average MSE of training). Here 3-4-1 represents the three nodes in the input layer, 4 nodes in the hidden layer and one node in the output layer.

Incremental Backpropagation algorithm is selected as a learning algorithm<sup>2</sup>. The transfer function used is sigmoid transfer function. The learning rate is set to 0.1. Momentum is set to 0.5. Stopping rule is "stop it after 80000 epochs".

During the training phase, the values in the training set are repeatedly feed through the ANN. The network compares its predicted output value to the actual output and adjusts all its weights to improve the model for technology selection. Performance of the ANN model on the training set was then evaluated by MSE. After

<sup>2</sup>Alyuda Forecaster Version 1.6 was used in the training and testing stages of the ANN model.

running for 80000 epochs, ANN model obtained an average MSE of 0.00276 in training. Therefore, it is possible to say that the ANN model can estimate the values of the output with very small mistakes. At this point, the network is ready for technology selection. The results of ANN for the inputs of neural network have been depicted in Table 4.

Sensitivity Analysis: ANNs are opaque. Even knowing all the weights on all the neurons throughout the network does not give much insight into why the network works. Research into rule extraction from networks may bring unequivocally good results. Until then, the trained network itself is the rule and other methods are needed to peer inside to understand what is going on. Fortunately, a technique called "Sensitivity Analysis (SA)" can be used to understand how opaque models work (Berry and Linoff [21]). SA does not provide explicit rules, but it does indicate the relative importance of the inputs to the result of the network. SA uses the test set to determine how sensitive the output of the network is to each input. The following are the basic steps:

**Step 1:** Find the average value for each input.

**Step 2:** Measure the output of the network when all inputs are at their average value.

**Step 3:** Measure the output of the network when each input is modified, one at a time, to be at its minimum and maximum values (usually 0 and 1 respectively).

For some inputs, the output of the network changes very little for the three values (minimum, average and maximum). The network is not sensitive to these inputs. Other inputs have a large effect on the output of the network. The network is sensitive to these inputs.

Table 5 presents the average value for each input and indicates the output of the network when all inputs are at their average value.

Figure 1-3 show the output of the network when each input is modified, one at a time, to be at its minimum and maximum values (usually 0 and 1 respectively). For vendor reputation, the output of the network does not change for the three values (minimum, average and maximum); i.e. the network is not sensitive to the vendor reputation. Other inputs including cost and load capacity have a large effect on the output of the network. The network is sensitive to these inputs. For the cost, the average is threshold that beyond which has not a significant effect on the network output; i.e. if the cost is being increased, the network output will be decreased slightly. However, load capacity has a positive effect on the network output.

There are variations on this procedure. It is possible to modify the values of two or three inputs at the same time to see if combinations of inputs have a particular importance. In this case, the analysis should be repeated for the minimum and maximum values of the inputs to see how sensitive the network is at the extremes. If SA produces significantly different results, then there are higher order effects in the network that are taking advantage of combinations of inputs.

Figure 4-6 show the output of the network when two inputs are changed simultaneously. For two pairs of inputs, the output of the network changes significantly; i.e. the network is sensitive to the combinations of inputs. The pair of vendor reputation-load capacity has a positive effect on the network output; i.e. if this combination is being increased, the network output will be increased. However, the pair of cost-vendor reputation has not a significant effect on the network output; i.e. if this combination is being increased, the network output will be fixed. The pair of cost-load capacity has a negative effect on the network output; i.e. if this combination is being changed, the network output will be decreased. Since SA produces significantly different results, then there are higher order effects in the network that are taking advantage of combinations of inputs.

**Rank-Sum Test:** It is often necessary to test statistically the difference between two groups, i.e., the results of DEA and the results of ANN, in terms of efficiency. Do differences occur by chance or are they statistically significant? This subsection deals with such statistical issues. Since DEA and ANN are nonparametric techniques, so the nonparametric statistics is used. For this purpose, the rank-sum test developed by Wilcoxon-Mann-Whitney [26] may be used to identify whether the differences between two groups are significant.

Rank-sum test is one of the nonparametric statistical tests based on the ranking of data. Given statistically independent data belonging to two groups, this test serves to test whether the hypothesis that the two groups belong to the same population or whether they differ significantly.





Fig. 1: The output of the network when cost is modified to be at its minimum (0) and maximum (1) values



Fig. 2: The output of the network when vendor reputation is modified to be at its minimum (0) and maximum (1) values



Fig. 3: The output of the network when load capacity is modified to be at its minimum (0) and maximum (1) values

World Appl. Sci. J., 6 (9): 1177-1189, 2009



Fig. 4: The output of the network when cost and repeatability are changed simultaneously



Fig. 5: The output of the network when cost and load capacity are changed simultaneously



Fig. 6: The output of the network when cost and velocity are changed simultaneously

Let the data in two groups be represented by A (*efficiency scores obtained by DEA*) and B (*efficiency scores obtained by ANN*). Then A and B are merged to arrive at a sequence C in which the data are arranged in descending order. There is

$$\begin{split} B=& [0.968, 0.897, 0.841, 0.763, 0.608, 0.598, 0.592, 0.577, \\ 0.561, 0.555, 0.449, 0.448, 0.393, 0.359, 0.357, 0.315, 0.299, \\ 0.292, 0.29, 0.282, 0.244, 0.24, 0.24, 0.232, 0.232, 0.227, 0.222, \\ 0.221, 0.216, 0.214, 0.191, 0.186, 0.183, 0.181, 0.177, 0.165, \\ 0.159, 0.156, 0.148, 0.145, 0.145, 0.132, 0.132, 0.132, 0.132, \\ 0.122, 0.116, 0.114, 0.111, 0.101, 0.101, 0.099, 0.096, 0.096, \\ 0.094, 0.09, 0.09, 0.084, 0.081, 0.075, 0.074, 0.074, 0.07, 0.07, \\ 0.069, 0.069, 0.067, 0.064, 0.063, 0.063, 0.061, 0.059, 0.059, \\ 0.055, 0.055, 0.054, 0.056, 0.056, 0.056, 0.055, 0.055, 0.055, \\ 0.055, 0.054, 0.054, 0.053, 0.053, 0.053, 0.053, 0.053, 0.053] \\ (n=100) \end{split}$$

Then these sequences into a new sequence are merged, C, with length m + n = 100+100 = 200:

C = [1, 1, 1, 0.968, 0.897, 0.841, 0.804, 0.8, 0.763, 0.631]0.608, 0.598, 0.592, 0.577, 0.561, 0.555, 0.532, 0.51, 0.5, <u>0.5, 0.5, 0.5, 0.5, 0.5, 0.449, 0.448, 0.419, 0.393, 0.365,</u> 0.363, 0.359, 0.357, 0.356, 0.333, 0.318, 0.315, 0.313, <u>0.307</u>, 0.299, 0.292, 0.29, <u>0.289</u>, 0.282, <u>0.27</u>, <u>0.266</u>, <u>0.252</u>, <u>0.25</u>, <u>0.25</u>, <u>0.25</u>, <u>0.25</u>, <u>0.25</u>, <u>0.25</u>, <u>0.25</u>, 0.244, 0.24, 0.24, 0.232, 0.232, 0.229, 0.227, 0.222, 0.221, 0.216, 0.214, 0.205, 0.2, 0.2, 0.2, 0.191, 0.189, 0.189, 0.186, 0.183, 0.181, 0.177, <u>0.174</u>, <u>0.174</u>, <u>0.174</u>, <u>0.167</u>, <u>0.167</u>, <u>0.167</u>, 0.165, 0.159, <u>0.159</u>, 0.156, <u>0.156</u>, 0.148, 0.148, 0.148, 0.145, 0.145, 0.143, 0.143, 0.143, 0.143, 0.135, 0.135, 0.135, 0.132, 0.132, 0.132, 0.132, 0.132, 0.132, 0.125, 0.125, 0.122, 0.116, 0.114, 0.113, 0.111, 0.109, 0.108, <u>0.106</u>, <u>0.104</u>, <u>0.104</u>, 0.101, 0.101, <u>0.1</u>, <u>0.1</u>, 0.099, 0.096, 0.096, 0.094, 0.09, 0.09, <u>0.09</u>, 0.084, <u>0.084</u>, 0.081, <u>0.077</u>, <u>0.077</u>, 0.075, 0.074, 0.074, <u>0.072</u>, <u>0.071</u>, 0.07, 0.07, <u>0.07</u>, 0.069, 0.069, <u>0.069</u>, 0.067, 0.064, 0.063, 0.063, 0.061, 0.059, 0.059, 0.059, 0.057, 0.056, 0.056, 0.056, 0.056, <u>0.056</u>, <u>0.056</u>, 0.055, 0.055, 0.055, 0.055, 0.054, 0.054, 0.054, 0.054, 0.054, 0.054, 0.054, 0.054, 0.054, 0.054, 0.053, 0.053, 0.053, 0.053, 0.053, <u>0.04</u>, 0.039, <u>0.027</u>, <u>0.023</u>, <u>0.02</u>, <u>0.02</u>, <u>0.02</u>, <u>0.017</u>, <u>0.012</u>, <u>0.01</u>, <u>0.009</u>, <u>0.006</u>, <u>0.006</u>, <u>0.006</u>, <u>0.005</u>, <u>0.004</u>, <u>0.004</u>, <u>0.003</u>, <u>0.003</u>, <u>0.002</u>, <u>0.002</u>, <u>0.002</u>]

in which the underlined numbers belong to A.

Then *C* is ranked from 1 to N (=m + n). If there is a tie, the midrank is used for the tied observation.

 $R = [\underline{2}, \underline{2}, \underline{2}, 4, 5, 6, \underline{7}, \underline{8}, 9, \underline{10}, 11, 12, 13, 14, 15, 16, \underline{17},$ <u>18, 21.5, 21.5, 21.5, 21.5, 21.5, 21.5, 25, 26, 27, 28, 29,</u> <u>30</u>, 31, 32, <u>33</u>, <u>34</u>, <u>35</u>, 36, <u>37</u>, <u>38</u>, 39, 40, 41, <u>42</u>, 43, <u>44</u>, <u>45, 46, 50, 50, 50, 50, 50, 50, 50, 50, 54, 55.5, 55.5, 57.5,</u> 57.5, <u>59</u>, 60, 61, 62, 63, 64, <u>65</u>, <u>67</u>, <u>67</u>, <u>67</u>, 69, <u>70.5</u>, <u>70.5</u>, 72, 73, 74, 75, 77, 77, 77, 80, 80, 80, 82, 83.5, 83.5, 85.5, <u>85.5</u>, 88, <u>88</u>, <u>88</u>, 90.5, 90.5, <u>93</u>, <u>93</u>, <u>93</u>, <u>96</u>, <u>96</u>, <u>96</u>, 100, 100, 100, 100, <u>100</u>, <u>103.5</u>, <u>103.5</u>, 105, 106, 107, <u>108</u>, 109, <u>110</u>, <u>111</u>, <u>112</u>, <u>113.5</u>, <u>113.5</u>, 115.5, 115.5, <u>117.5</u>, <u>117.5</u>, 119, 120.5, 120.5, 122, 124, 124, <u>124</u>, 126.5, <u>126.5,</u> 128, <u>129.5,</u> <u>129.5,</u> 131, 132.5, 132.5, <u>134,</u> <u>135,</u> 137, 137, <u>137</u>, 140, 140, <u>140</u>, 142, 143, 144.5, 144.5, 146, 148, 148, 148, 150, 153.5, 153.5, 153.5, 153.5, 153.5, 153.5, 159, 159, 159, 159, 159, 166.5, 166.5, 166.5, 166.5, 166.5, 166.5, 166.5, 166.5, 166.5, 166.5, 174.5, 174.5, 174.5, 174.5, 174.5, 174.5, <u>178</u>, <u>179</u>, <u>180</u>, <u>181, 183.5, 183.5, 183.5, 183.5, 186, 187, 188, 189, 191,</u> <u>191, 191, 193, 194.5, 194.5, 196.5, 196.5, 199, 199,</u> <u>199</u>]

For example, the top three numbers in the sequence C have the same value 1 so (1+2+3)/3=2 which is their midrank in R.

Next, the ranking of the *A* data indicated by the underlined numbers is summed.

$$\begin{split} & S=2+2+2+7+8+10+17+18+21.5+21.5+21.5+21.5+21.5+21.5\\ &+27+29+30+33+34+35+37+38+42+44+45+46+50+50+50\\ &0+50+50+50+59+65+67+67+67+70.5+70.5+77+77+77+80+\\ &80+80+83.5+85.5+88+88+93+93+93+96+96+96+100+103.5\\ &+103.5+108+110+111+112+113.5+113.5+117.5+117.5+117.5+124\\ &+126.5+129.5+129.5+134+135+137+140+153.5+153.5+178\\ &+179+180+181+183.5+183.5+183.5+183.5+183.5+186+187+188+1\\ &89+191+191+191+193+194.5+194.5+196.5+196.5+199+199\\ &+199=9753.5\end{split}$$

Then statistic, *S*, follows an approximately normal distribution with mean m(m + n + 1)/2 and variance mn(m+n+1)/12 for m,n\$10. By normalizing *S*, there will be:

$$T = \frac{S - \frac{m(m+n+1)}{2}}{\sqrt{\frac{mn(m+n+1)}{12}}}$$

*T* has an approximately standard normal distribution. Using *T*, the null hypothesis that the two groups have the same population at a level of significance " can be checked. The hypothesis will be rejected if  $T \leq -T_{a_{\lambda}}$  or

 $T \ge T_{a_2}$ , where  $T_{a_2}$  corresponds to the upper  $a_2$ 

percentile of the standard normal distribution. In this example, there is T= -0.72446. If a = 0.05 (5%) is chosen, then it holds that  $T_{0.025}$  = 1.96. Since T = -0.72446>-1.96=  $-T_{0.025}$ , the null hypothesis at the significance level 5% is rejected. Consequently, the differences among the efficiency scores obtained by DEA and efficiency scores obtained by ANN are statistically significant.

**Concluding Remarks:** This paper has introduced a new use of ANNs for technology selection. Such a technique seems to be able to support management in the critical and delicate task of selecting the best technology among a set of competing multi-attribute technologies.

Since, the ANN is built with sets of historical data, it is difficult to guarantee the network will provide satisfactory results, especially when the network is used in different situations where the input feed into the network is not from the same domain. Moreover, the ANN does not have the sensibility characteristic like the decision maker; it is not able to identify the environment changes which need to readjust the output to fit the environment, which leaves room for further improvement to the system.

To make sure that the predicted variable could be adjusted in response to change in the performance of technologies, updated data should be fed into the data set and passed to the ANN for training regularly.

The problem considered in this study is at initial stage of investigation and much further researches can be done based on the results of this paper. Some of them are as follow: Similar research can be repeated for dealing with ordinal data, fuzzy data and bounded data in the conditions that dual-role factors exist.

In this study, the proposed model has been applied to a problem related to robot selection. However, the same approach could be applied, to other problems related to selection of flexible manufacturing systems, computer integrated manufacturing systems, computer numerical control systems and many other technology selection decision cases.

### ACKNOWLEDGEMENT

The author wishes to thank an anonymous referee for the valuable suggestions and comments.

#### REFERENCES

- Shehabuddeen, N., D. Probert and R. Phaal, 2006. From theory to practice: challenges in operationalising a technology selection framework. Technovation, 26(3): 324-335.
- Al-Ahmari, A.M.A., 2007. Evaluation of CIM Technologies in Saudi Industries Using AHP. Intl. J. Adv. Manufacturing Technol., 34(7-8): 736-747.
- Kengpol, A. and C. O'Brien, 2001. The Development of a Decision Support Tool for the Selection of Advanced Technology to Achieve Rapid Product Development. Intl. J. Prod. Econom., 69(2): 177-191.
- Utturwar, A., S. Rallabhandi, D. Delaurentis and D. Marvis, 2006. A Two-Step Optimization Approach for Technology Selection. Engineering Optimization, 38(8): 889-908.
- Lee, J.W. and S.H. Kim, 2000. Using analytic network process and Goal Programming for Interdependent Information System Project Selection. Computers & Operations Res., 27(4): 367-382.
- Lee, J.W. and S.H. Kim, 2001. An integrated approach for interdependent information system project selection. Intl. J. Project Manage., 19(2): 111-118.
- Chan, F.T.S., M.H. Chan and N.K.H. Tang, 2000. Evaluation Methodologies for Technology Selection. J. Materials Processing Technol., 107(1-3): 330-337.
- Kulak, O., C. Kahraman, B. Oztaysi and M. Tanyas, 2005. Multi-Attribute Information Technology Project Selection Using Fuzzy Axiomatic Design. The J. Enterprise Inform. Manage., 18(3): 275-288.

- Malladi, S. and K.J. Min, 2005. Decision support models for the selection of Internet Access Technologies in Rural Communities. Telematics and Informatics, 22(3): 201-219.
- Hajeeh, M. and A. Al-Othman, 2005. Application of the analytical hierarchy process in the selection of desalination plants. Desalination, 174(1): 97-108.
- Jaganathan, S., J.J. Erinjeri and J. Ker, 2007. Fuzzy Analytic Hierarchy Process Based Group Decision Support System to Select and Evaluate New Manufacturing Technologies. Intl. J. Adv. Manufacturing Technol., 32(11-12): 1253-1262.
- Parkan, C. and M. Wu, 1999. Decision-Making and Performance Measurement Models with Applications to Robot Selection. Computers & Industrial Engineering, 36(3): 503-523.
- 13. Khouja, M., 1995. The Use of Data Envelopment Analysis for Technology Selection. Computers & Industrial Engineering, 28(1): 123-132.
- Wang, S., 2006. Comments on Operational Competitiveness Rating Analysis (OCRA). European J. Operational Res., 169(1): 329-331.
- Baker, R.C. and S. Talluri, 1997. A Closer Look at the Use of Data Envelopment Analysis for Technology Selection. Computers & Industrial Engineering, 32(1): 101-108.
- Farzipoor Saen, R., 2006. Technologies ranking in the presence of both cardinal and ordinal data. Appl. Mathematics and Computation, 176(2): 476-487.
- Farzipoor Saen, R., 2006. A decision model for technology selection in the existence of both cardinal and ordinal data. Applied Mathematics and Computation, 181(2): 1600-1608.

- Talluri, S., M.M. Whiteside and S.J. Seipel, 2000. A Nonparametric Stochastic Procedure for FMS Evaluation. European J. Operational Res., 124(3): 529-538.
- Karsak, E.E. and S.S. Ahiska, 2005. Practical Common Weight Multi-Criteria Decision-Making Approach with an Improved Discriminating Power for Technology Selection. Intl. J. Prod. Res., 43(8): 1537-1554.
- Choy, K.L., W.B. Lee and V. Lo, 2003. Design of an Intelligent Supplier Relationship Management System: A Hybrid Case Based Neural Network Approach. Expert Systems with Applications, 24(2): 225-237.
- Berry, M.J.A. and G. Linoff, 1997. Data Mining Techniques: for Marketing, Sales and Customer Support. John Wiley & Sons.
- 22. Werbos, P.J., 1974. Beyond Regressions: New Tools for Prediction and Analysis in the Behavioral Science, Ph.D. Dissertation, Committee on Applied Mathematics, Cambridge, MA: Harvard University.
- Rumelhart, D.E., D.E. Hinton and R.J. Williams, 1986. Learning Representation by Backpropagating Errors. Nature, 323(9): 533-536.
- Charnes, A., W.W. Cooper and E. Rhodes, 1978. Measuring the Efficiency of Decision Making Units. European J. Operational Res., 2(6): 429-444.
- Banker, R.D., A. Charnes and W.W. Cooper, 1984. Some Methods for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. Manage. Sci., 30(9): 1078-1092.
- Cooper, W.W., L.M. Seiford and K. Tone, 2007. Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA Solver Software. Springer, Second Edition.