PERGAMON

Computers & Operations Research 26 (1999) 1243–1265

computers & operations research

www.elsevier.com/locate/orms

Predicting the success of nations at the Summer Olympics using neural networks

Edward M. Condon[a], Bruce L. Golden[b,*], Edward A. Wasil[c]

[a]*College of Engineering, University of Maryland, College Park, MD 20742, USA*
[b]*Robert H. Smith School of Business, University of Maryland, College Park, MD 20742, USA*
[c]*Kogod College of Business Administration, American University, Washington, DC 20016, USA*

Received January 1998; received in revised form November 1998

Abstract

In this paper, we construct several models that try to predict a country's success at the Summer Olympic Games. Our data set consists of total scores for over 271 sporting events for 195 countries that were represented at the 1996 Summer Games and information we gathered on 17 independent variables. We build linear regression models and neural network models and compare the predictions of both types of models. Overall, the best neural network model outperformed the best regression model.

Scope and purpose

Every four years, an enormous amount of attention is focused on the Summer Olympic Games. Sports fans and analysts do their best to predict the outcomes with respect to many sporting events and the overall performance of nations competing at the Olympics. In this paper, we develop and compare socio-economic-based models for predicting the success of nations at the Summer Olympics.

Keywords: Artificial neural networks (ANNs); Ordinary least-squares (OLS) regression; Weighted least-squares (WLS) regression; Summer Olympics; Data analysis

1. Introduction

The 1996 Summer Olympic Games in Atlanta were the largest Games held since the inception of the modern Olympic Games in 1896. Almost 11,000 athletes from 197 nations participated in 271 events from 26 sports. It is estimated that over five million spectators and visitors came to Atlanta

* Corresponding author. P.O. Box 149, Simpsonville, MD 21150, USA. Tel.: + 1-301-405-2232; fax: + 1-301-314-8787
E-mail address: bgolden@rhsmith.umd.edu (B.L. Golden)

0305-0548/99/$ - see front matter
PII: S0305-0548(99)00003-9

[1] and that 3.5 billion people watched the Olympic Games on television [2]. Total ticket sales of 8.6 million surpassed the previous record for Olympic Games ticket sales by three million [1].

While more countries than ever were represented in the final medal count at the 1996 Summer Olympic Games (78 countries received at least one medal), some countries regularly win more medals than others. Many attempts have been made in the literature to explain these differences. Population is cited most frequently as a factor affecting Olympic success. However, we only need to examine the recent performances of nations such as Cuba and India to see that other factors are involved (in 1996, Cuba won 25 medals, while India won one medal). Past research efforts to account for differences in success between nations have relied on linear modeling techniques such as ordinary least-squares (OLS) regression.

In this paper, we construct several models to predict a country's success at the Summer Olympic Games. The first models are traditional OLS linear regression models. The second group of models are artificial neural network models (ANNs). The artificial neural network models are constructed using a computer program developed by faculty and students in the Robert H. Smith School of Business at the University of Maryland. Our data set consists of total scores (defined by the weighted sum of the top eight places in each event) for 195 countries that were represented at the 1996 Summer Games and information we gathered on 17 independent variables. (We point out that, of the 197 nations that participated in the 1996 Summer Olympic Games, socio-economic data were unavailable for two of the countries.) We compare the results of the artificial neural network models to results of the regression models using mean absolute error (MAE) of predictions as our measure of performance. We conjecture that the ability of the neural networks to model complex nonlinear relationships among variables should result in better predictions.

In Section 2 of this paper, we discuss Olympic success and provide an overview of past research on factors relating to Olympic success. A summary of the relevant literature is presented.

In Section 3, we discuss neural networks, network architectures, and training algorithms. The standard backpropagation training algorithm is mentioned along with enhancements to the algorithm that are used in this paper.

In Section 4, we discuss how the data are obtained and compiled. We provide a detailed description of the dependent variable and present the process we used to select the independent variables.

In Section 5, we discuss the models that we construct to predict the total score for each country that participated in the 1996 Summer Olympic Games. The regression models are built using all of the data. For the neural network models, the complete data set is split into training and testing sets. The training set is used to train the neural network models and the testing set is used to select the best set of parameter values for the neural network models. We use network pruning during training to improve performance. MAE values are computed using the testing set data for both the regression and neural network models.

In Section 6, we summarize our modeling efforts and provide suggestions for future research.

2. Olympic success

In 1896, the first modern Olympic Games were held in Athens, Greece. Athletes representing 14 nations competed in 43 events. Since then, the size and interest in the Olympic Games have grown enormously. In the XXVI Summer Games held in Atlanta in 1996, athletes from 197 nations competed in 271 events. The Summer Olympic Games are held every four years and provide nations an opportunity to showcase their best athletes in international competition.

Officially, the Summer Games are intended to be a competition between athletes and not countries. Until 1908, athletes entered as individuals rather than as members of a national team. The awarding of gold, silver, and bronze medals for first, second, and third places began in 1904. Although the International Olympic Committee (IOC) does not officially recognize national medal totals, these totals are often presented in the Organizing Committee of the Olympic Games (OCOG) reports and in the popular media. Typically, national medal totals are updated daily by the media during the Olympic Games. Some sports publications even try to predict national medal totals prior to the start of the Olympic Games [3]. Table 1 illustrates one such prediction.

Despite official claims by the IOC that "the Olympic Games are competitions between athletes in individual or team events and not between nations" [4], a ranking of nations according to total medals won is reported frequently. Table 2 shows the top three countries ranked by total number of medals won for all of the modern Summer Olympic Games.

Seven studies that investigate the differences in performance among countries have appeared in the open literature. A summary of data sets that have been used in the literature is given in Table 3. For a more detailed discussion of the data and analysis, see [5].

Our review of the literature reveals that computing Pearson correlation coefficients and using regression analysis are the most popular methods for examining relationships between variables and Olympic success. However, one must be aware of the assumptions being made about the data set when choosing a method of analysis or a specific model. In particular, linear regression assumes normality of errors for all observations. This assumption is often violated.

Unlike Pearson correlation coefficients and regression models, artificial neural networks are nonparametric procedures that do not require assumptions to be made about the data set being modeled. Another advantage of using artificial neural networks is that nonlinear relationships may be captured in the model. We are not aware of research in the open literature that investigates nonlinear relationships between influential factors and Olympic success.

Table 1
Top five countries at the Atlanta Games as projected in *Sports Illustrated* [3]

Projected final medal standings				
Country	Gold	Silver	Bronze	Total
United States	48	42	48	138
Germany	23	25	35	83
Russia	24	20	23	67
China	21	16	9	46
Australia	8	20	13	41

Table 2
Top finishing countries ranked by total medals won

Year	Location	First	Second	Third
1896	Athens, Greece	Greece	United States	Germany
1900	Paris, France	France	United States	Great Britain
1904	St. Louis, USA	United States	Germany	Canada
1908	London, Great Britain	Great Britain	United States	Sweden
1912	Stockholm, Sweden	Sweden	United States	Great Britain
1916			Canceled because of war	
1920	Antwerp, Belgium	United States	Sweden	Finland
1924	Paris, France	United States	France	Finland
1928	Amsterdam, Holland	United States	Germany	Finland
1932	Los Angeles, USA	United States	Italy	France
1936	Berlin, Germany	Germany	United States	Hungary
1940			Canceled because of war	
1944			Canceled because of war	
1948	London, Great Britain	United States	Sweden	France
1952	Helsinki, Finland	United States	Soviet Union	Hungary
1956	Melbourne, Australia	Soviet Union	United States	Australia
1960	Rome, Italy	Soviet Union	United States	Italy
1964	Tokyo, Japan	Soviet Union	United States	Japan
1968	Mexico City, Mexico	United States	Soviet Union	Japan
1972	Munich, Germany	Soviet Union	United States	East Germany
1976	Montreal, Canada	Soviet Union	United States	East Germany
1980	Moscow, USSR	Soviet Union	East Germany	Bulgaria
1984	Los Angeles, USA	United States	Romania	Germany
1988	Seoul, South Korea	Soviet Union	East Germany	United States
1992	Barcelona, Spain	Soviet Union	United States	Germany
1996	Atlanta, USA	United States	Russia	Germany

Table 3
Summary of data sets examined in the open literature

Authors	Year published	Year of Games	Location of Games	Number of competing countries	Number of countries included in study
Jokl et al. [6]	1956	1952	Helsinki, Finland	69	33–66
Jokl [7]	1964	1960	Rome, Italy	83	Not specified
Ball [8]	1972	1964	Tokyo, Japan	93	36
Novikov and Maximenko [9]	1972	1964	Tokyo, Japan	93	53–84
		1968	Mexico City, Mexico	112	40–52
Levine [10]	1974	1972	Munich, Germany	129	98–129
Grimes et al. [11]	1974	1972	Munich, Germany	129	95
Kiviaho and Makela [12]	1978	1964	Tokyo, Japan	93	30

3. Artificial neural networks

We assume that the reader is familiar with the basics of artificial neural networks. In particular, we assume familiarity with multilayer, feedforward networks, training, testing, supervised learning, memorization, generalization, and standard backpropagation. (For details, we recommend [13] as a gentle introduction to ANNs.) For the purposes of this paper, we use the terms neural network and artificial neural network interchangeably.

An enhancement to the standard backpropagation algorithm that is used in this paper is pruning. One of the problems that can be encountered using the backpropagation algorithm is overfitting. This occurs when a neural network memorizes the input training patterns and becomes unable to generalize successfully to new input patterns. Pruning seeks to reduce overfitting by eliminating links that may not store important information (see [14]). These links, while reducing errors during training, may contribute meaningless information to the final output of a network when new patterns are presented. Fig. 1 shows an example of a pruned network.

Two parameters are used in our implementation of pruning. The first parameter is the critical weight value that is used to determine which links to prune. After a network has been trained, the weight on each link is examined. Links with weights whose absolute values are less than the critical weight value are pruned from the network. After removing links from the network, the remaining weights are reinitialized and the network is retrained using the backpropagation algorithm. The second parameter is the number of removal iterations. This determines how many times the pruning and retraining process is repeated.

4. Data sets and variable selection

Our goal is to build regression models and neural network models that predict the total score of a country participating in the 1996 Summer Olympic Games. The total score for each country is

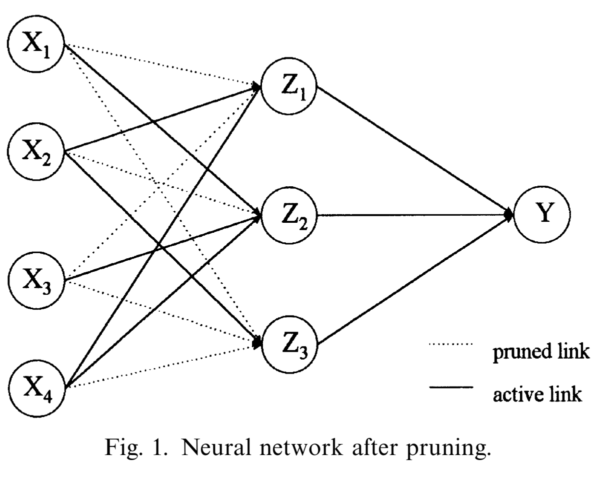

Fig. 1. Neural network after pruning.

calculated by assigning points to the top eight placing countries for each of the 271 Olympic events. Socio-economic information (e.g., population and national product) is used to predict total score. All of the information used to construct our models is available prior to the start of the Summer Olympic Games. Thus, our models could be used to predict the total score for a country prior to the start of the Summer Olympic Games.

4.1. Dependent variable

Other than the tabulation of medal counts in the popular media, there is no official method for scoring a country's performance in the Olympic Games. We use the simple method for scoring that is shown in Table 4. We award points to the countries of the top eight places in the final competition of each event. Although our scoring method favors the top three places (recall that, in Olympic competition, first place earns a gold medal, second place earns a silver medal, and third place earns a bronze medal), we also award points when an athlete or team performs well in the finals but is not awarded a medal. We define the *total score* for a country as the sum of the points awarded to a country across all Olympic events. Total score is our dependent variable.

In the 1996 Olympic Games, there were 197 participating countries and 271 events (see [5] for a listing of participating countries and Olympic events). Information on the top eight places in the finals of each event was obtained from the Atlanta Committee for the Olympic Games (ACOG) site on the Web (http://www.atlanta.olympic.org) [15]. The medal finishers were checked with *Sports Illustrated*'s listing of the final medal winners (see [16]). In Tables 5–7, we provide examples of the type of information that was available on the ACOG Web site. There were a few events for which information on all eight places was not available on the ACOG Web site or the ACOG results did not agree with the listings in *Sports Illustrated*. These events were women's road race (cycling), all judo events, all wrestling events, and open soling (yachting). Information on all eight place finishers for the cycling and wrestling events was obtained from the US Olympic organizations that govern cycling and wrestling. We could not obtain the fourth through eighth place finishers for the judo events from the US Olympic organization that governs judo. For the missing yachting places, we could not identify a US Olympic Sport Organization or other sporting body to contact. Thus, for

Table 4
Points assigned to top eight places

Place	Points assigned
First	5
Second	3
Third	2
Fourth	1
Fifth	1
Sixth	1
Seventh	1
Eighth	1

Table 5
Individual competition results from the ACOG Web site: Three levels of competition

Olympic results: men's archery individual			
	Archer	Country	Result
Finals			
Gold	Huish, Justin	USA	Winner
Silver	Petersson, Magnus	SWE	
Bronze	Oh, Kyo-Moon	KOR	Winner
	Vermeiren, Paul	BEL	
Semifinals			
	Huish, Justin	USA	Winner
	Vermeiren, Paul	BEL	
	Petersson, Magnus	SWE	Winner
	Oh, Kyo-Moon	KOR	
Quarterfinals			
	Huish, Justin	USA	Winner
	Frangilli, Michele	ITA	
	Vermeiren, Paul	BEL	Winner
	Torres, Lionel	FRA	
	Oh, Kyo-Moon	KOR	Winner
	Kim, Bo-Ram	KOR	
	Petersson, Magnus	SWE	Winner
	Jang, Yong-Ho	KOR	

Table 6
Individual competition results from the ACOG Web site: top eight places

Olympic results: swimming, women's 800 m freestyle				
	Rank	Athlete	Country	Time
Gold	1	Bennett, Brooke	USA	8:27.89
Silver	2	Hase, Dagmar	GER	8:29.91
Bronze	3	Vlieghuis, Kirsten	NED	8:30.84
	4	Kielgass, Kerstin	GER	8:31.06
	5	Dalby, Irene	NOR	8:38.34
	6	Evans, Janet	USA	8:38.91
	7	Guerts, Carla Louise	NED	8:40.43
	8	Hardcastle, Sarah	GBR	8:41.75

Table 7
Team competition results from the ACOG Web site

Olympic results: women's softball

	Country	Total	Win	Loss
Gold	USA	9	8	1
Silver	CHN	10	6	4
Bronze	AUS	9	6	3
	JPN	8	5	3
	CAN	7	3	4
	TPE	7	2	5
	NED	7	1	6
	PUR	7	1	6

judo and yachting, the fourth through eighth place finishers were not included in the computation of a country's total score.

An examination of previous research reveals that a number of different scoring methods have been used to measure Olympic success. Our scoring method tries to use as much of the available information on results as possible. By awarding one point to the fourth through eighth place finishers, we allow countries that received no medals in the 1996 Summer Games to have a positive total score.

Using our scoring method, 98 of 197 countries scored one point or more, as compared to the 79 countries awarded at least one medal. We point out that 19 countries received no medals and, with our scoring method, they now have a positive score. The total score for each country is given in [5].

Although the top eight finishers for most events were available, rankings below third (bronze medal) were not always clear. A bit of detective work was sometimes required.

4.2. Independent variables

In order to develop a set of independent variables, we examined the seven studies on Olympic success that have appeared in the open literature and took into account the factors (variables) that influenced Olympic success. In addition, we wanted to use variables for which data were easily available. Based on this search, we identified 29 candidate variables that could be used to construct our regression and neural network models. Data for these variables were available from *The World Factbook 1995* [17].

We point out that there are two difficulties with the data set of 29 candidate independent variables. First, data were not available for Palestine and Yugoslavia in *The World Factbook 1995* (perhaps due to unclear physical and political boundaries in each country in recent years), so these two countries are excluded from our analysis. Second, there are 12 variables for which data are unavailable for more than nine countries. We drop these 12 variables and use the remaining 17 independent variables shown in Table 8. In [5], the data for all 17 independent variables are given.

Table 8
Independent variables

Variable name	Short description	Number of observations available
AREA	Sum of all land and water areas delimited by international boundaries and coastlines	195
POPULATI	Population estimate based on population censuses	195
POP_GRW_	Percent population growth rate	195
BIRTH_RT	Birth rate per 1000	195
DEATH_RT	Death rate per 1000	195
INF_MOR_	Infant mortality rate per 1000 live births	195
LIFE_EXP	Life expectancy at birth for total population	195
AIRPORTS	Number of airports with usable runways	195
RAILROAD	Total length of railroad track	195
HIGHWAYS	Total length of paved and unpaved highways	193
NATL_PRO	National product (total output of goods and services in a country in a given year)	192
NP_PER_C	National product per capita (national product divided by population)	192
EL_CAPAC	Electric capacity	192
EXPORTS	Total value of exported goods in a given year	190
IMPORTS	Total value of imported goods in a given year	189
EL_PROD	Electric production	188
EL_PER_C	Electric consumption per capita	187

4.3. *Missing data*

In the data set for the 17 independent variables, there are a small number of missing observations. For example, with respect to highways, we have data for 193 countries. Data are missing for two countries: Aruba and Maldives. Overall, we need to fill in missing data for 37 variable entries. We propose to accomplish this in three ways: use the mean of existing values, use the minimum value of existing values, use a subjective estimate.

Using the mean of existing values to fill in missing data is the simplest of the three methods. The means can be computed easily and are reasonable guesses for missing entries. However, we note that many of the countries with missing data are either small countries (in area or population) or countries that have recently experienced major political and economic changes. The means are not likely to be good estimates of the missing entries in these instances. Therefore, we propose using the minimum value of existing values to fill in missing entries. We expect the actual values to be closer to the minimum values than the mean values. The minimum values can be computed easily.

Subjective estimates were also used to fill in the missing values. For each country with missing data, a country judged to be similar and having complete data was selected. Nearby countries that are similar in terms of area and population size were identified. Economic factors (such as national product, imports, and exports) were compared, when data were available. Variable values from the similar countries were used to fill in the values for the countries with missing data.

Table 9
Example of filled-in variable values

	Highways (km)		
	Mean	Minimum	Similar
Aruba	134,361	0	950

Table 9 shows an example of a missing variable value filled in by each of the three methods. The Netherlands Antilles was selected as a country similar to Aruba and its value for highways was used to fill in the value missing for Aruba.

5. Modeling a country's total score in the 1996 Summer Games

In this section, we construct models that can be used to predict the total score for a country participating in the Summer Olympic Games. Two different modeling approaches are used: ordinary least-squares regression and neural networks. Two neural network models are trained and tested using a code developed at the University of Maryland (this is the Maryland Network Code, denoted by MNC). One model is built using standard backpropagation and a second model is built using standard backpropagation with pruning. To evaluate effectiveness, we calculate the mean absolute error of the predictions from each model.

Before constructing our models, we divided the countries into two sets. A random number between 0 and 1 was assigned to each of the 195 countries. The countries were then sorted from smallest to largest based on this random number. The first 130 countries, representing 2/3 of the data set, were placed in a training set. The remaining 65 countries, representing 1/3 of the data set, were placed in a testing set.

Splitting the data into training and testing sets allows us to identify the neural network parameter values that perform best. The regression models are built on all the data (that is, all 195 countries). The neural network models are trained on the 130 countries in the training set and tested on the 65 countries in the testing set. Testing is performed to find the best parameter settings for MNC. There are four parameters whose values we need to set: number of hidden nodes, learning rate, momentum, and pruning. After the entire data set was split into training and testing sets, the values for missing variable data were filled in using one of the three methods described earlier.

Whether or not the United States is included in a particular data set is also a concern. Since the total score of the United States (469 points) is much larger than the total score of other countries (Russia has the second-highest total – 308 points), it may behave as an outlier and affect the accuracy of our models. To investigate this effect, we first built a neural network model with the United States in the training set. We then built a neural network model with the United States in the testing set. Since the regression models are built on the entire set of data, there is no need to study this effect on the regression models.

5.1. Stepwise regression models

We use SPSS for Windows Release 6.1 to construct the OLS linear regression models for total score. The OLS models are built on the entire data set (training and testing data sets) using all 17 independent variables in a stepwise algorithm [18]. The first model was built using the data set with missing values filled in with variable means. The SPSS output is shown in Table 10.

The F-statistic, which measures overall model fit, is statistically significant at the 1% level. The six independent variables present in the model are significant at the 5% level.

The variance inflation factor (VIF) is an important multicollinearity diagnostic. A VIF value larger than 10 implies serious problems with multicollinearity for a data sample [19]. Since all of the VIFs in the OLS model are less than 10, multicollinearity does not appear to be a problem.

A plot of the residuals against the predicted values for the OLS model is shown in Fig. 2. The outward sloping funnel pattern in Fig. 2 indicates that the variance of the residuals is an increasing function of the prediction. This is a violation of the OLS assumption that prediction errors have a constant variance. Using a transformation that compensates for the nonconstant variance may improve predictions. We consider this later. In Fig. 3, we see that the histogram of standardized residuals indicates a slight departure from normality.

A second model was built with the missing data filled in with the existing minimum values of the variables. The SPSS output, a plot of residuals, and a histogram of standardized residuals are shown in [5]. The results are very similar to those of the first model.

Table 10
OLS linear regression model output: missing values filled in with means

R^2	0.833			
Adjusted R^2	0.828			
Standard error	23.206			
Observations	195			
Analysis of variance				
	DF	SSE	MSE	F-statistic
Regression	6	505 168.992	84 194.832	156.343
Residual	188	101 242.823	538.526	
	Coefficient	t-statistic	p-value	VIF
Intercept	− 146.321	− 5.042	0.000	
AIRPORTS	− 0.009	− 3.237	0.001	3.632
DEATH_RT	3.341	4.569	0.000	3.921
EXPORTS	0.308	8.142	0.000	2.220
LIFE_EXP	1.917	5.226	0.000	5.726
NP_PER_C	− 0.001	− 3.471	0.001	2.326
RAILROAD	0.002	12.646	0.000	4.077

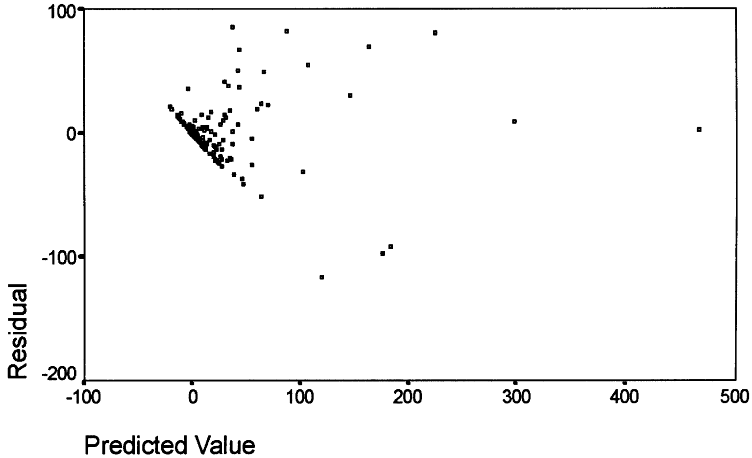

Fig. 2. Plot of residuals for OLS regression model: missing values filled in with means.

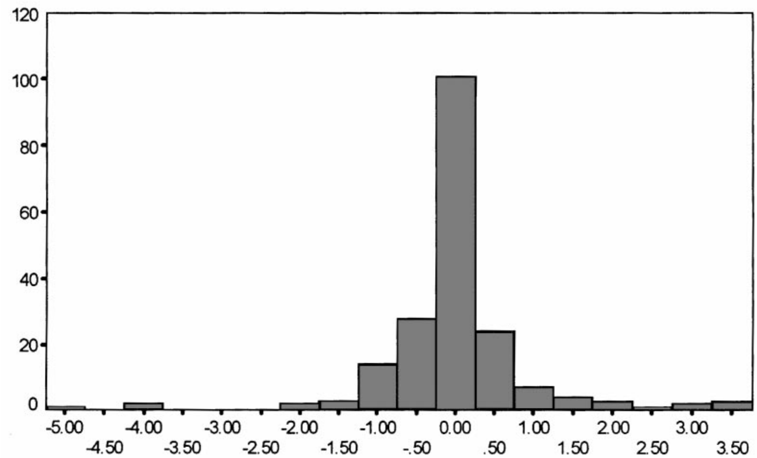

Fig. 3. Histogram of standardized residuals for OLS regression model: missing values filled in with means.

A third model was built with the missing data filled in with the subjective estimates. The estimates were obtained by examining variable values of countries that were similar in terms of geographic location, size, and economic factors to the countries with missing data. The SPSS output, a plot of residuals, and a histogram of standardized residuals are shown in [5]. The results are very similar to those of the first model.

The three regression models were used to predict total scores for countries in the testing data sets. Table 11 summarizes the MAEs. We see that filling in missing values with minimum values improves predictions over the other two fill-in methods. In addition, we see that including the United States in the testing set increases the MAE of the testing set predictions.

Table 11
Summary of testing set MAEs for OLS regression models

Fill-in method	US in testing set	US in training set
Mean	12.024	11.997
Minimum	11.851	11.823
Similar	11.915	11.885

Table 12
Neural network options

Options	Selections
Fill-in method	Mean, minimum, similar
Location of United States	Training set, testing set
Number of input nodes	6, 17

Table 13
Parameter settings searched

Parameter	Number of input nodes	
	6	17
Number of hidden nodes	2, 4, 6, 8, 10, 12	9, 13, 17, 21, 25, 29
Learning rate	0.3, 0.5, 0.7	0.3, 0.5, 0.7
Momentum	0.3, 0.5, 0.7	0.3, 0.5, 0.7
Number of initial seeds	3	3
Pruning	Off	Off

5.2. MNC models

The Maryland Network Code is a personal computer program developed over the last six years by faculty and students in the Robert H. Smith School of Business at the University of Maryland, College Park (see [14] for more details about the program). MNC is a powerful research program that lets a user set the values for a wide variety of parameters. MNC uses the standard backpropagation algorithm for training and can also perform network pruning. MNC's strength lies in its wide variety of features that provide extensive modeling flexibility.

When constructing our neural network models, there are three questions that we need to answer. Which data set of independent variables do we use? (Recall that there are three such data sets with missing values filled in with mean, minimum, or similar values.) Do we include the United States in the training set or in the testing set? Do we use the six independent variables (from the regression models) as inputs to the neural network model or all 17 independent variables? Based on the answers to these questions, there are 12 neural networks that we can construct. We summarize the neural network options in Table 12.

We conduct an experiment to find the best settings for five parameters in each model: number of hidden nodes, learning rate, momentum, number of initial seeds, and pruning. The values of the first four parameters can be increased or decreased while the pruning parameter can be turned on or off. Neural network training is performed on the training data set and testing is performed on the testing data set. MAE is used to evaluate the different parameter settings. In Table 13, we show the range of parameter values that we searched over during training (note that pruning is turned off). In

Table 14
Best parameter settings: United States included in the testing set

	Fill-in method	Number of hidden nodes	Learning rate	Momentum	Test set MAE
6 inputs					
	Mean	10	0.3	0.7	12.275
	Minimum	8	0.5	0.3	11.923
	Similar	10	0.3	0.7	12.138
17 inputs					
	Mean	21	0.7	0.3	12.432
	Minimum	13	0.3	0.5	11.871
	Similar	21	0.5	0.3	11.966

Table 15
Best parameter settings: United States included in the training set

	Fill-in method	Number of hidden nodes	Learning rate	Momentum	Test set MAE
6 inputs					
	Mean	8	0.5	0.7	10.185
	Minimum	10	0.5	0.7	9.745
	Similar	10	0.5	0.7	9.992
17 inputs					
	Mean	9	0.7	0.7	9.872
	Minimum	29	0.7	0.7	10.578
	Similar	13	0.7	0.7	9.758

Tables 14 and 15, we show the best parameter settings when the United States is included in the testing set and when the United States is included in the training set.

5.3. MNC models with pruning

We also build neural network models with pruning enabled (that is, pruning is turned on). To use pruning, we need to set the critical weight value and the number of removal iterations. We conduct an experiment to determine which values work well for our network (the structure of our neural network model is given in Table 16). We select three values for the critical weight (0.1, 0.5, 0.9) and two values for the number of removal iterations (1, 3). Overall, we experiment with 24 neural networks and compute each network's MAE on the testing set. The results of our experiment are given in Tables 17–20.

Using the results in Tables 17–20, we make the following observations.

- Setting the number of removal iterations to a value of three works well in practice. With the exception of four cases, the MAE values were lower when we set the number of removal

Table 16
Neural network options used to test pruning settings

Options	Selections
Missing data fill-in method	Mean
Location of United States	Training set, testing set
Number of input nodes	6, 17

Table 17
MAEs for 6 inputs, filling in missing values with column means, US in training set

Critical weight	Number of removal iterations	
	1	3
0.1	9.684	9.684
0.5	10.185	10.026
0.9	9.959	9.959

Table 18
MAEs for 17 inputs, filling in missing values with column means, US in training set

Critical weight	Number of removal iterations	
	1	3
0.1	9.872	9.798
0.5	8.870	8.870
0.9	9.872	9.583

Table 19
MAEs for 6 inputs, filling in missing values with column means, US in testing set

Critical weight	Number of removal iterations	
	1	3
0.1	12.275	12.215
0.5	12.141	11.912
0.9	12.275	12.042

Table 20
MAEs for 17 inputs, filling in missing values with column means, US in testing set

Critical weight	Number of removal iterations	
	1	3
0.1	11.136	10.811
0.5	10.235	10.202
0.9	11.047	11.047

iterations to a value of three than they were for a value of one. In four cases, the MAE values were the same.

- Once we have set the number of removal iterations to a value of three, we see that, over the four experiments, a critical weight of 0.1 gives the lowest MAE once and a critical value of 0.5 gives the lowest MAE three times.

Based on these two observations, we set number of removal iterations to a value of three and the critical weight to a value of 0.5.

Once the pruning settings have been determined, we conduct an experiment to find the best settings for the remaining four parameters in each model: number of hidden nodes, learning rate, momentum, and number of initial seeds. Neural network training is performed on the training data set and testing is performed on the testing data set. MAE is used to evaluate the different parameter settings. In Table 21, we show the range of parameter values that we searched over during training. In Tables 22 and 23, we show the best parameter settings when the United States is included in the testing set and when the United States is included in the training set.

The MAE values in Tables 22 and 23 are lower than the corresponding MAE values given in Tables 14 and 15 (recall that pruning is turned off in these two tables). In these experiments, pruning seems to improve a network's performance. In Fig. 4, we can see which arcs are pruned from a network for one of the cases of six input nodes. Table 24 gives a summary of the unpruned and pruned architectures and network performance associated with Fig. 4.

5.4. Summary of results

Table 25 shows average MAE comparisons for the different fill-in methods. We see that, on average, filling in missing values with minimum values or similar values produces better results

Table 21
Parameter settings searched

Parameter	Number of input nodes	
	6	17
Number of hidden nodes	2, 4, 6, 8, 10, 12	9, 13, 17, 21, 25, 29
Learning rate	0.3, 0.5, 0.7	0.3, 0.5, 0.7
Momentum	0.3, 0.5, 0.7	0.3, 0.5, 0.7
Number of initial seeds	3	3
Pruning	On	On
Number of removal iterations	3	3
Critical weight	0.5	0.5

Table 22
Best parameter settings: United States included in the testing set

	Fill-in method	Number of hidden nodes	Learning rate	Momentum	Test set MAE
6 inputs					
	Mean	10	0.3	0.7	11.912
	Minimum	12	0.5	0.5	11.404
	Similar	6	0.5	0.5	11.381
17 inputs					
	Mean	17	0.7	0.7	10.202
	Minimum	17	0.7	0.7	10.274
	Similar	13	0.7	0.7	10.200

Table 23
Best parameter settings: United States included in the training set

	Fill-in method	Number of hidden nodes	Learning rate	Momentum	Test set MAE
6 inputs					
	Mean	10	0.7	0.3	10.026
	Minimum	12	0.7	0.7	9.741
	Similar	12	0.5	0.7	9.468
17 inputs					
	Mean	9	0.7	0.7	8.870
	Minimum	29	0.5	0.5	9.056
	Similar	29	0.7	0.7	9.560

than filling in with mean values. We also see that, on average, the neural network models perform better than the OLS linear regression models.

Table 26 shows average MAE comparisons for models using the pruning algorithm. We see that models trained with the pruning algorithm turned on perform better for both networks with 6 input nodes and 17 input nodes. We also see that, on average, networks with 17 input nodes perform better than networks with 6 input nodes.

Table 27 shows average MAE comparisons for the United States in the training set and the United States in the testing set. The small difference in the OLS linear regression models is expected. These models were built using all of the data and then changes were made in the testing set for MAE computations. However, the neural network models were trained only on the data in the training set. Table 27 shows that MNC models trained with the United States in the training set perform better on average than MNC models trained with the United States in the testing set.

5.5. *Some remaining issues*

The plots of the residuals against the predicted values for the OLS models (e.g., see Fig. 1) show an outward sloping funnel pattern. This indicates nonconstant variance and suggests the need for a transformation or the use of weighted least-squares regression [18]. For the sake of completeness, we explored weighted least squares (WLS) using SPSS for Windows Release 6.1. Weights were assigned to each observation based on the reciprocal of the variance of that observation's error term (or a suitable approximation). This gave more weight to observations with smaller error variances [20]. We plotted the OLS residuals against each independent variable included in the final OLS model to determine which variable contributed most to the nonconstant variance. Although the plots indicated the variables AIRPORTS and RAILROAD to be the most likely candidates, we created WLS models using each independent variable as an estimator for the weights. Table 28 summarizes the testing set MAEs of the predicted scores for the various models.

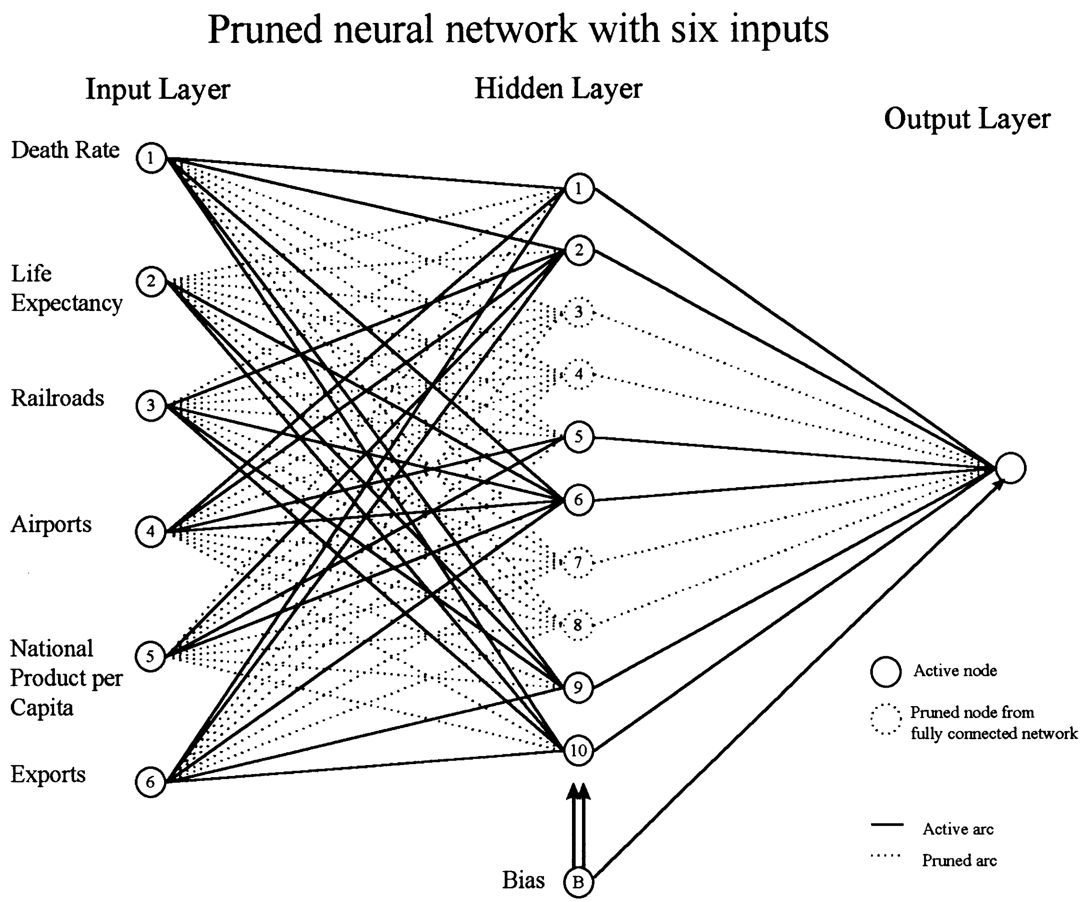

Fig. 4. Pruned neural network trained with US in testing set, missing values filled in with mean values.

Table 24
Summary of network architectures (US in testing set, missing values filled in with mean values)

	Number of input nodes	Number of hidden nodes	Number of arcs from input to hidden nodes	Number of arcs from hidden to output nodes	Total number of arcs	MAE of test set
Pruning turned off	6	10	60	10	70	12.275
Pruning turned on						
initial configuration	6	10	60	10	70	—
final configuration	6	6	24	6	30	11.912

The results show that the WLS models created using AIRPORTS and RAILROAD to estimate the weights perform well when the United States is not included in the testing set. In these cases, the WLS models outperform the best ANN models (with 17 input nodes and pruning). However, the WLS models perform poorly when the United States is included in the testing set. In these

Table 25
Average MAE comparisons for fill-in methods

Model	Fill-in method			All
	Mean	Minimum	Similar	
OLS linear regression	12.011	11.837	11.900	11.916
MNC	10.697	10.574	10.055	10.609
Both	10.960	10.826	10.826	

Table 26
Average MAE comparisons for pruning algorithm

Pruning	6 inputs	17 inputs	Both
Off	11.043	11.080	11.061
On	10.655	9.660	10.158
Combined	10.849	10.370	

Table 27
Average MAE comparisons

Model	US in training set	US in testing set
OLS linear regression	11.902	11.930
MNC	9.721	11.498
Both	10.157	11.584

Table 28
Summary of testing set MAEs for WLS, OLS, and ANN models

Weighting variable	Correlation with score	US in testing set			US in training set		
		Mean	Minimum	Similar	Mean	Minimum	Similar
AIRPORTS	0.649	19.178	19.071	19.111	8.139	7.951	7.999
DEATH_RT	– 0.018	13.624	13.156	13.204	13.511	13.036	13.091
EXPORTS	0.750	14.001	10.505	12.991	10.770	9.707	10.369
LIFE_EXP	0.249	12.666	12.517	12.546	12.614	12.466	12.501
NP_PER_C	0.313	12.921	11.729	12.569	12.123	11.532	12.314
RAILROAD	0.845	12.792	12.792	12.792	8.656	8.656	8.656
OLS		12.024	11.851	11.915	11.997	11.823	11.885
ANN		10.202	10.274	10.200	8.870	9.056	9.560

cases, both the OLS and ANN models noticeably outperform the WLS models. Furthermore, an examination of the residual plots for the WLS models reveals that much of the dispersion is still present, indicating that the models do not adequately correct for the nonconstant variance.

Our scoring method, shown in Table 4, is quite reasonable. Since, in each event, the top three finishers receive medals, the points assigned are greater than for finishers numbered 4–8, who each receive a single point. On the other hand, it is also somewhat subjective. Would the rankings change significantly if the scoring method changed slightly?

We explored this issue, at least partially, in the following way. Suppose we assign $(9 - i)^{\delta}$ points to place i for $i = 1, \dots, 8$. If $\delta = 0$, then the first eight finishers receive a score of 1. If $\delta = 1$, then the scores from high to low become 8, 7, 6, ..., 1. If $\delta = 0.5$, then the scores from high to low become $\sqrt{8}$, $\sqrt{7}$, $\sqrt{6}$, ..., $\sqrt{1}$. Using this mechanism, we compared the original rankings (based on Table 4) with rankings obtained for $\delta = 0, 0.5$, and 1. In each case, we applied the ANN model with 17 input nodes and pruning. Minimum values were used to fill in missing entries.

The results are displayed in Figs. 5 and 6. The 65 countries in the testing set are ranked (based on Table 4) and the associated delta rankings are also presented. In each plot, the delta rankings are indicated by small squares while the original rankings are represented by the 45° line. Deviations

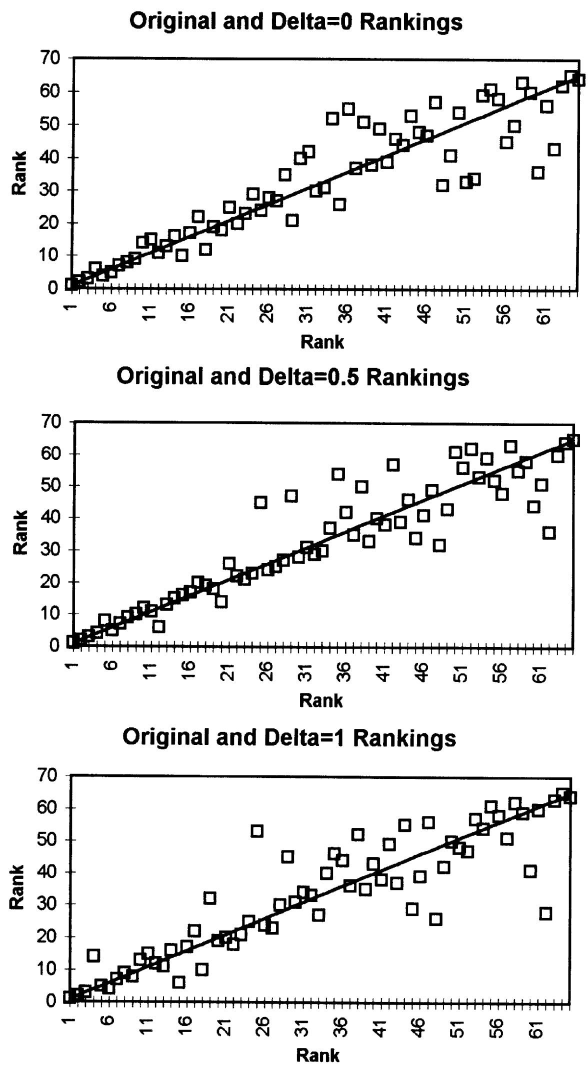

Fig. 5. Sensitivity of rankings when US is in the testing set.

from the 45° line indicate differences in rankings. It is clear that there are differences, but not substantial differences, especially if we focus on the top 10 countries in each plot.

5.6. Practical applications of findings

Overall, neural network models outperform regression models for predicting a country's total score at the Summer Olympic Games. In addition, the three methods for filling in missing values produce different results and the use of the pruning algorithm improves the performance of the neural network models.

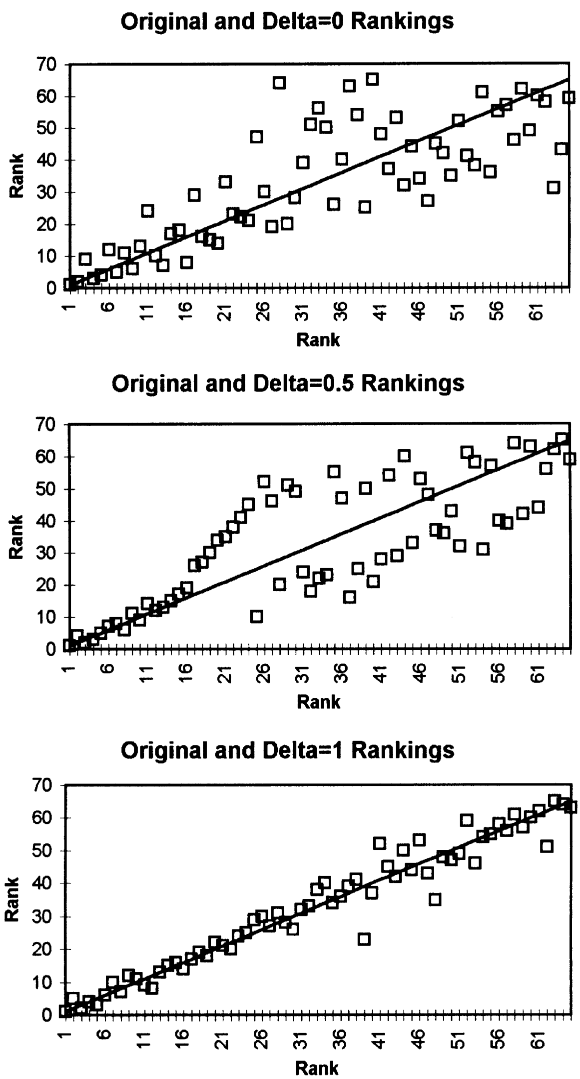

Fig. 6. Sensitivity of rankings when US is in the training set.

Although a country could use our models and new data to predict its total score in an upcoming Olympics, there is probably not much a country can do (in the short term) to make significant changes in the hopes of increasing its total score. In the long term, however, our models may provide clues as to how a country might improve its Olympic performance.

Our models could be used to account for performance differences among similar countries. This may also suggest other variables associated with Olympic success that are not included in our models. For example, all of our models (neural network and regression) predict a low total score for Kenya based on the 1995 data. However, Kenya produced a total score at the 1996 Summer Olympic Games that many larger countries (in terms of area, population, and economy) were not able to achieve (such as Mexico and Argentina). Examining more closely how Kenya prepares for Olympic competition or what other socio-economic variables impact performance may provide useful information to other small nations without substantial financial resources on how to improve their Olympic showings.

6. Conclusions

In this paper, we constructed models to predict the total score for a country participating in the 1996 Summer Olympic Games. We used data for 195 countries that participated in the 1996 Summer Games. We built 27 models: three models were OLS linear regression models and 24 models were neural network models. Our data sets had missing values for several independent variables and we filled in the missing values using means, minimums, and subjective estimates. The neural network models were built with the United States in the training set and with the United States in the testing set.

We compared the results of the neural network models to the results of the regression models using mean absolute error as our measure of performance. The best neural network model outperformed the best regression model. In every case, neural network models trained with the pruning algorithm performed better than the models trained without it when using the same data set. In terms of future work, several ideas come to mind.

We selected our variables, in part, because extensive information was readily available for them. However, other variables such as population density, economic system, and geographic distance from the Summer Games have been associated with Olympic success in previous studies. Adding new variables to the models may improve performance.

Another possibility for future work is including results and data from more than one Olympic competition. Building models based on more than one Olympic competition could provide valuable insight as to which variables remain stable over time.

References

[1] United States Olympic Committee (USOC) World Wide Web site. http://www.olympic-usa.org/games.
[2] Smith G. It's greek to U.S. Sports Illustrated 1996;85:32–8.
[3] Verschoth A. Who will win what? Sports Illustrated – Olympic Preview Issue 1996;231–48.
[4] International Olympic Committee (IOC). Olympic charter. Lausanne, Switzerland: International Olympic Committee, 1996.

[5] Condon E. Predicting the success of nations in the Summer Olympics using neural networks. Master's Thesis, University of Maryland at College Park, 1997.
[6] Jokl E, Karvonen M, Kihlberg J, Koskela A, Noro L. Sports in the cultural pattern of the world. Helsinki, Finland: Institute of Occupational Health, 1956.
[7] Jokl E. Health, wealth, and athletics. In: Jokl E, Simon E, editors. International research in sport and physical education. Springfield, IL: Thomas, 1964;218–22.
[8] Ball D. Olympic games competition: structural correlates of national success. International Journal of Comparative Sociology 1972;13:186–200.
[9] Novikov A, Maximenko A. The influence of selected socio-economic factors on the level of sports achievements in the various countries. International Review of Sport Sociology 1972;7:22–44.
[10] Levine N. Why do some countries win Olympic medals? Some structural correlates of Olympic games success: 1972. Sociology and Social Research 1974;58:353–60.
[11] Grimes A, Kelly W, Rubin P. A socioeconomic model of national Olympic performance. Social Science Quarterly 1974;55:777–83.
[12] Kiviaho P, Makela P. Olympic success: a sum of non-material and material factors. International Review of Sport Sociology 1978;2:5–17.
[13] Beale R, Jackson T. Neural computing: an introduction. London, UK: IOP Publishing Ltd., 1990.
[14] Sun X. Neural network models for the wire bonding process. Ph.D. Dissertation, College of Business and Management, University of Maryland, College Park, MD, 1994.
[15] Atlanta Committee for the Olympic Games World Wide Web site. http://www.atlanta.olympic.org. (Site is no longer active.)
[16] Sports Illustrated, August 1996 Olympic Commemorative Issue, 132–5.
[17] Central Intelligence Agency. The world factbook 1995. Washington, DC: CIA, 1995.
[18] Draper N, Smith H. Applied regression analysis, 2nd ed. New York: Wiley, 1981.
[19] Montgomery D, Peck E. Introduction to linear regression analysis. New York: Wiley, 1992.
[20] Carroll R, Ruppert D. Transformation and weighting in regression. New York: Chapman & Hall, 1988.

Edward M. Condon currently works in the Physics Department at the University of Maryland. He received a BS in Physics and Psychology and an MS in Systems Engineering, all from the University of Maryland.

Bruce L. Golden holds the France-Merrick Chair in Management Science in the Robert H. Smith School of Business at the University of Maryland. His research interests include neural networks, heuristic search, and applied operations research. He is currently Editor-in-Chief of the INFORMS Journal on Computing.

Edward A. Wasil is a Professor of Management Science in the Kogod College of Business Administration at American University in Washington, D.C. His research interests include the application of neural networks to business decision problems, network optimization, and the process of implementing management science models. Currently, Dr. Wasil serves as the Feature Article Editor of the INFORMS Journal on Computing.