Classification of Engineering Consultancy Firms Using Self-Organizing Maps: A Scientific Approach

Mansour N. Jadid
Department of Building Science and Technology, University of Dammam
P.O.Box 30973, Alkhobar 31952, Saudi Arabia
mnjadid@yahoo.com

Abstract—The present study analyzed consultants in Saudi Arabia. Engineering consultancy firms work on projects worth billions of US dollars annually in the fields of design and supervision. The quality of the services provided in this field has a major impact on the national economy through various aspects of these services, such as the lifetime of the projects, their environmental effects, their efficiency, the esthetics of cities, and the societal impacts of the projects. Thus, the author investigated the knowledge, experience, human resources and financial situations of these firms, and the scientific qualifications of their professional staff. A survey was conducted using several consultancy firms and a database was constructed, while some of these firms were prequalified. The results of this exercise were revised and checked thoroughly, then used to construct a questionnaire to evaluate the firms. This study focused on applying a scientific approach to the classification of engineering consultancy firms using self-organizing maps (SOMs). An electronic version of the questionnaire was made available for all firms to complete to build the required database. A template was created using Microsoft Word to summarize the information collected, which included the prequalification data for firms and the levels of projects they handled. The summaries were converted into Excel format and used to feed an artificial neural network program. Using this program, the required information was extracted with a type of Kohonen network known as an SOM. The results were presented as tables and figures.

Index Term—algorithm; classification; clustering; consultant; Kohonen; learning; neural network; self-organizing map; supervisor.

I. LITERATURE REVIEW REGARDING CONSULTANCY FIRMS IN SAUDI ARABIA

The development and growth of engineering consultancy firms in Saudi Arabia needs to keep pace with the growing need for the services provided by these firms, given the growing need for engineering projects to be completed in Saudi Arabia and several massive development booms in Saudi Arabia in recent years. The number of consultancy firms in Saudi Arabia was less than 100 in 1976 but it increased rapidly to 1350 in 1998. These firms are distributed throughout Saudi Arabia: Riyadh province has 42% of them, Jeddah 28%, Dammam 7%, Alkhobar 6%, and Makkah and Medina 4%. The remaining 13% are distributed throughout the other provinces according to the Saudi Council of Engineers [1].

The capabilities and qualifications of these firms vary. Some have high standards, with high potential capabilities in urban planning, architectural creativity, excellence in engineering innovation, and superior management performance. Another class belongs in the middle, which are strong on the creative side of architecture but weak in the field of engineering, or vice versa. In general, these firms cooperate with other firms that have higher architectural and engineering capabilities to ensure good performance. Finally, there are firms with low capabilities where much effort and the use of specialized technical expertise are required to upgrade their performance so they can provide architectural services and engineering consultancy at an adequate level. In 2003, the value of the projects completed by engineering consultancy firms in Saudi Arabia in the field of design and supervision was estimated at between 3 and 4 billion US dollars. In 2000, 90% of the engineering work conducted was performed by 10% of the firms, which means that 90% of the firms are competing for a 10% market share, Abbas [2]. These indicators show that the market for engineering consultancy services is large, but there needs to be a reorganization so there is a balanced distribution of work among the consultants in the market. Consultancy firms need to increase their levels of performance to compete seriously in the market and they need to comply with the latest qualification requirements specified by the Saudi Council of Engineers, as well as following the rules and standards set by the Ministry of Commerce and Industry. This paper is based on a study of engineering consultancy firms in the Saudi Arabia, which includes more than 1440 firms distributed throughout the country. The author analyzed a sample of 500 firms from various provinces.
II. INTRODUCTION TO SELF-ORGANIZING MAPS

Self-organizing maps (SOMs), also known as Kohonen neural networks, Kohonen [3,4] use an unsupervised learning process to modify the internal state of a network to model the features found in a training dataset. This type of network has two layers: an input layer to obtain information from the outside, and an output layer to send information to the outside. When information is provided to an SOM, an output neuron is selected as a winner. This neuron is the output of the network and it corresponds to one of the classified groups, as shown in Fig. 1. An SOM is a type of neural network that has the ability to learn by detecting regularities and correlations in its inputs to predict future responses. This type of neural network model is to analyze and visualize high-dimensional data while preserving topological relationships. It projects a high-dimensional signal space onto a two-dimensional grid of nodes, so this type of network belongs to the class of competitive learning networks.

SOM has been applied to many applications in engineering such as modeling, analysis, prediction and estimation, as well as other disciplines. SOMs were used by ChandraShekar and Shoba [5] to retrieve textual information from the World Wide Web, while Liu and Weisberg [6] demonstrated the use of SOMs in the fields of meteorology and oceanography to produce a powerful pattern recognition and feature extraction scheme. Gonçalves et al. [7] applied an unsupervised method to the classification of remotely sensed images, which was based on clustering using SOM. The main concept of their proposed method achieved cluster analysis using images based on a set of SOM prototypes, instead of working directly with the original patterns of the scene. In the financial market, Khan et al., [8] proposed a method for improving stock picking using an SOM instead of the traditional method. An unsupervised neural network approach based on SOM was implemented using the NeuroSolution software package by Fekihal and Yousif [9]. In this method, the work was concentrated on the design and implementation to classify transcribed speech samples and determine mental disorders.

In the area of geotechnical engineering, the classification of soils was addressed using two methods based on SOM to construct a classification model as mentioned by Mutalib et al., [10]. A new face-based biometric system using SOMs was developed by Raja and JosephRaj [11] where a supervised learning technique improved the performance and robustness of recognition to exploit the physiological or biological characteristics of humans for recognition. An SOM was also implemented to address a problem related to the level of service for urban streets, which involved a classification problem with four urban classes in an Indian context, as described by Mohapatra and Bhuyan [12].

![Fig. 1. Simple illustration of the processing in an SOM network](image_url)

A. Neural network-based clustering

The process of clustering aims to identify similarity relationships between data objects in a high-dimensional signal space. An SOM was used in the present study to classify data into classes by unsupervised learning. During unsupervised learning, the training dataset only contains input variables and the SOM network attempts to learn the structure of the data. An iterative algorithm is used to train the SOM network, which starts with an initial random set of radial centers and the algorithm adjusts them gradually to reflect the clustering of the training data.
B. Unsupervised learning using neural software

Using a neural software program, the author employed an SOM to perform the unsupervised classification of consultants according to their design and supervision experience. The network used for the SOM in the software was placed in a folder “/consultant project/final/employee-education” with the training and testing data set, which contained 80 training and 30 test records.

C. SOM learning algorithm

SOM is an unsupervised learning algorithm. It is a visualization technique used to represent higher dimensional data such as 1-D, 2-D, or 3-D behavior. The functions of the SOMs involve two stages: a learning stage where the map is built using a network for organization based on a competitive process using the training set; and a prediction stage where new vectors are moved rapidly to locations on the converged map so they can classify or categorize the new data easily. The simple steps required by the SOM algorithm are as follows:

- Initialize
- Sample
- Match
- Update

D. The SOM process

The SOM uses a square grid of $n^2$ neurons, are arranged in $n$ rows and $n$ columns, where the number of cycles is specified during each cycle and all of the observations are presented to the map once. The order of presentation may be random or user-defined, although the default order is stored in a data sheet in an updatable format. The number of variables entered in the sheet matches the total number of columns exactly, and the number of observations entered in the input sheet is less than or equal to the number of rows in the data sheet.

E. Scope and limitations

The scope of work relies on engineering consultancy firms in Saudi Arabia and only the basic iterative Kohonen algorithm was applied. It runs through a number of epochs, where it executes each training case during every epoch. To classify the consultants based on the cost of their previously completed projects, the data were exported to an Excel sheet by Microsoft Office [13] with eight columns and 80 rows, where six columns were used for the inputs and the first two columns were omitted. The sample data for 80 consultants are shown in Table I. The cluster sizes and the configurations are shown in Tables II and III, respectively (the cluster information plus the mean and variance of each cluster are also shown).

<table>
<thead>
<tr>
<th>Consultant ID</th>
<th>Previous Design Projects Cost (US dollars)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Less than 15,000</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>92</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>393</td>
</tr>
</tbody>
</table>
### Table II
**CLUSTERING INFORMATION USED IN THE NETWORK**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of variables used for clustering</td>
<td>6</td>
</tr>
<tr>
<td>Number of observations used for clustering</td>
<td>80</td>
</tr>
<tr>
<td>Number of Clusters</td>
<td>4</td>
</tr>
</tbody>
</table>

### Table III
**MEAN AND VARIANCE OF CLUSTERS**

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>Cluster 1</th>
<th>Cluster 2</th>
<th>Cluster 3</th>
<th>Cluster 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cost (US Dollars)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15,000 to 30,000</td>
<td>69.8</td>
<td>43.0</td>
<td>26.7</td>
<td>27.6</td>
<td>61.0</td>
</tr>
<tr>
<td>30,000 to 150,000</td>
<td>138.9</td>
<td>62.3</td>
<td>53.8</td>
<td>20.4</td>
<td>40.3</td>
</tr>
<tr>
<td>150,000 to 300,000</td>
<td>84.8</td>
<td>28.4</td>
<td>14.7</td>
<td>1.1</td>
<td>5.1</td>
</tr>
<tr>
<td>Greater than 500,000</td>
<td>51.8</td>
<td>17.0</td>
<td>2.5</td>
<td>0.9</td>
<td>8.0</td>
</tr>
<tr>
<td>Greater than 1,000,000</td>
<td>49.8</td>
<td>15.4</td>
<td>0.0</td>
<td>0.3</td>
<td>4.3</td>
</tr>
</tbody>
</table>

### Table IV
**CRITERIA PARAMETERS SPECIFIED FOR THE KOHONEN NETWORK ANALYSIS**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of observations</td>
<td>80</td>
</tr>
<tr>
<td>Number of variables</td>
<td>5</td>
</tr>
<tr>
<td>$n^2 = \text{number of neurons in the map}$</td>
<td>4</td>
</tr>
<tr>
<td>$n = 4$, to obtain a number of clusters less than or equal to</td>
<td>16</td>
</tr>
<tr>
<td>Number of training cycles</td>
<td>80</td>
</tr>
</tbody>
</table>

### Table V
**LEARNING PARAMETERS SPECIFIED FOR THE KOHONEN NETWORK ANALYSIS**

<table>
<thead>
<tr>
<th>Learning Parameters</th>
<th>Start value</th>
<th>End value</th>
<th>Decay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sigma for the Gaussian neighborhood as percentage of map width</td>
<td>0.9</td>
<td>0.1</td>
<td>exponential</td>
</tr>
</tbody>
</table>

### III. CLUSTER ASSIGNMENT BASED ON PROJECT COST

During the map training process, the variables in the dataset were scaled so the values of each variable lay between $(-1)$ and $(1)$. The outputs were generated to the output sheet and the classified data, cluster labels, cluster means, and variances were saved in a new workbook, as shown in Table III. Finally, a plot was generated in the weight sheet to visualize how many data points were captured in each portion of the map (Fig. 2).
IV. CLASSIFICATION RESULTS

A total of 80 observations were clustered into four groups using unsupervised learning, i.e., cluster 1 = 11, cluster 2 = 38, cluster 3 = 7 and cluster 4 = 24 which were classified from the dataset. Table VI shows the output of the cluster assignment process and Table VII shows the cluster assignments for various consultants.

**TABLE VI**
UNSUPERVISED OUTPUT OF THE CLUSTER ASSIGNMENT PROCESS

<table>
<thead>
<tr>
<th>Consultant ID</th>
<th>Cluster ID</th>
<th>Consultant ID</th>
<th>Cluster ID</th>
<th>Consultant ID</th>
<th>Cluster ID</th>
<th>Consultant ID</th>
<th>Cluster ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>21</td>
<td>1</td>
<td>41</td>
<td>2</td>
<td>61</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>22</td>
<td>1</td>
<td>42</td>
<td>2</td>
<td>62</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>23</td>
<td>2</td>
<td>43</td>
<td>2</td>
<td>63</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>24</td>
<td>2</td>
<td>44</td>
<td>2</td>
<td>64</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>25</td>
<td>4</td>
<td>45</td>
<td>2</td>
<td>65</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>26</td>
<td>4</td>
<td>46</td>
<td>4</td>
<td>66</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>27</td>
<td>4</td>
<td>47</td>
<td>2</td>
<td>67</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>28</td>
<td>3</td>
<td>48</td>
<td>2</td>
<td>68</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>29</td>
<td>2</td>
<td>49</td>
<td>2</td>
<td>69</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>30</td>
<td>3</td>
<td>50</td>
<td>2</td>
<td>70</td>
<td>1</td>
</tr>
</tbody>
</table>

The results of unsupervised classification for various activities are shown in Fig. 3. Fig. 3a shows the results for the personnel of the consultancy firms, including civil, architectural, structural, electrical, and mechanical engineers. Fig. 3b shows the cost of previously completed projects. Fig. 3c shows the firms’ specialties during project supervision, while Fig. 3d shows the employees’ qualifications.
Consultancy firms’ manpower (Civil, Architect, Structure, Electrical, Mechanical) (a)

Previous completed project cost (15,000-30,000, 30,000-150,000, 150,000-300,000, > 300K, > 500K, > 1,000K) (b)

Consultancy firms’ specialties (Design supervision) (c)

Employees’ qualifications (PhD, Masters, Bachelors, Diploma) (d)

Fig. 3. Results of unsupervised classification

V. CLASSIFIED DATABASE OF CONSULTANTS

After processing these steps, a classified database was generated with 20 columns that contained data on 80 consultancy firms. The first two columns represented the ID and name of the firm, while the remaining columns represented the classification data on a scale from 1 to 5. In this classification, a firm with ID=1 was classified in class B based on its employees’ education. The same firm was classified as: class C based on its experience in civil and electrical engineering; class E based on mechanical and architectural engineering; and class A based on structural engineering. The “Design” and “Supervision” columns contained the class values for these criteria. The classified database for this sample is shown in Table VIII with a list of the 80 consultants included. The classification scale is shown in Table IX.
TABLE VIII
CLASSIFIED DATABASE FOR THE SAMPLE OF CONSULTANTS OBTAINED USING A NEURAL NETWORK

| ID Consultant’s | Employees’ Education | Civil | Electrical | Mechanical | Architecture | Structural | Commercial | Educational | Residential | Hospitals | Sports | Hotels | Commercial | Educational | Residential | Hospitals | Sports | Hotels |
|-----------------|----------------------|-------|------------|------------|--------------|------------|------------|------------|------------|------------|---------|--------|--------|------------|------------|------------|---------|--------|--------|
| 1               | 2 3 5 5 1 5 5 5 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 |        |            |            |              |            |            |            |            |            |         |        |        |            |            |            |         |        |        |
| 2               | 5 5 4 1 5 5 4 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 |        |            |            |              |            |            |            |            |            |         |        |        |            |            |            |         |        |        |
| 3               | 1 4 1 4 2 5 5 5 0 0 0 2 3 0 0 0 0 0 0 0 0 0 |        |            |            |              |            |            |            |            |            |         |        |        |            |            |            |         |        |        |
| 4               | 1 1 2 3 5 5 5 5 0 5 5 3 2 4 0 2            |        |            |            |              |            |            |            |            |            |         |        |        |            |            |            |         |        |        |
| 5               | 5 5 1 4 2 4 4 4 0 4 5 6 5 3 0 6            |        |            |            |              |            |            |            |            |            |         |        |        |            |            |            |         |        |        |
| 6               | 3 5 4 4 2 4 4 5 5 6 5 5 0 6 5 5 5          |        |            |            |              |            |            |            |            |            |         |        |        |            |            |            |         |        |        |
| 7               | 2 3 2 1 1 1 1 1 1 5 1 1 0 1 4 1 1          |        |            |            |              |            |            |            |            |            |         |        |        |            |            |            |         |        |        |
| 8               | 1 3 1 3 1 1 1 4 4 1 0 3 1 1 1 1 0 0        |        |            |            |              |            |            |            |            |            |         |        |        |            |            |            |         |        |        |
| 9               | 5 5 5 5 1 5 5 5 4 0 0 5 5 5 5 0 0          |        |            |            |              |            |            |            |            |            |         |        |        |            |            |            |         |        |        |
| 10              | 1 2 5 4 1 5 4 5 5 2 0 1 5 3 0 1          |        |            |            |              |            |            |            |            |            |         |        |        |            |            |            |         |        |        |

TABLE IX
CLASSIFICATION SCALE

<table>
<thead>
<tr>
<th>Classification</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1/Class A</td>
<td>Excellent</td>
</tr>
<tr>
<td>Class 2/Class B</td>
<td>Very Good</td>
</tr>
<tr>
<td>Class 3/Class C</td>
<td>Good</td>
</tr>
<tr>
<td>Class 4/Class D</td>
<td>Fair</td>
</tr>
<tr>
<td>Class 5/Class E</td>
<td>Average</td>
</tr>
<tr>
<td>Class 6/Class F</td>
<td>Below Average</td>
</tr>
<tr>
<td>Class 7/Class G</td>
<td>Fail</td>
</tr>
</tbody>
</table>

VI. CLASSIFICATION SUMMARY AND INTERFACE DESIGN

The results obtained after combining the data from the individual classification files are summarized in Table X. These results were grouped into two main categories: “Project Design” and “Project Supervision.” For commercial projects, 14/80 consultants were classified in class A, which comprised 28% of the total number of consultants that worked on commercial projects. For healthcare projects, 7/10 consultants were classified in class A while 10% were classified in class A for educational projects, and 4% were classified in class A for residential projects. These results are shown in Table X. The percentages for project design are shown in Fig. 4.
TABLE X
SUMMARIZED CLASSIFICATION RESULTS

<table>
<thead>
<tr>
<th>Class</th>
<th>Project design</th>
<th>Project supervision</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Commercial</td>
<td>Health</td>
</tr>
<tr>
<td>Class A</td>
<td>14 7 5 0 21 2 13 5 11 2 8 2</td>
<td></td>
</tr>
<tr>
<td>Class B</td>
<td>11 8 3 0 0 2 9 3 0 1 2 1</td>
<td></td>
</tr>
<tr>
<td>Class C</td>
<td>0 0 0 0 1 1 15 14 1 0 2 4</td>
<td></td>
</tr>
<tr>
<td>Class D</td>
<td>1 8 13 5 6 5 3 11 5 4 0 9</td>
<td></td>
</tr>
<tr>
<td>Class E</td>
<td>6 9 9 9 12 21 3 11 0 1 13 0</td>
<td></td>
</tr>
<tr>
<td>Class F</td>
<td>4 1 1 0 14 3 0 0 4 5 6 1</td>
<td></td>
</tr>
<tr>
<td>Class G</td>
<td>14 17 20 36 8 16 7 6 32 39 19 33</td>
<td></td>
</tr>
</tbody>
</table>

The consultant classification database was exported to Microsoft Access, Microsoft Office [13] and used as back-end storage. Active Server Pages - ASP [14] was used as a programming tool to link the database to a webpage to generate queries. Fig. 5 shows the relationships between the tables in the database. The interface page was designed in ASP using Macromedia Dreamweaver in both Arabic and English [15], and users had the option to explore and search the database or view the required information, as shown in Fig. 6.
VII. CONCLUDING REMARKS

In this study, an SOM was applied to aid the selection of appropriate consultant firms. There is wide variation in the requirements for the preparation of studies and designs for various types of buildings and facilities (commercial, educational, health, sports, infrastructure, etc.) and in the requirements for the supervision of special projects, so the present study was conducted based on a sample of engineering consultancy firms and design projects, where the author used a neural network to classify the firms. This study was conducted using a form of unsupervised learning known as Kohonen networks or SOMs.

Future scope of the research will include other unsupervised neural networks such as Bayesian networks which depend on density estimation techniques by building statistical models.

ACKNOWLEDGMENT

This research was supported by King Abdulaziz City for Science and Technology, Saudi Arabia, Grant #: AT-25-28. The author acknowledges and appreciates the contributions made by the team working on this project.

REFERENCES


Mansour N. Jadid is an associate professor of Civil Engineering in Department of Building Science and Technology and the General supervisor for the University of Dammam Projects in Dammam, University of Dammam, Saudi Arabia. Received the Ph.D in 1994 from the University of Edinburg, UK. MSCE from the Pennsylvania State University, USA in 1987 and BSCE from the University of Washington, Seattle, Washington, USA in 1981. His email mniadid@yahoo.com. He was awarded a Fulbright scholarship in the field of artificial intelligence in civil engineering at the University of Maryland, Maryland, USA. He has published more than 25 papers in referee scientific journals and conference proceedings. He presented his research in international conferences and seminars in Saudi Arabia, UAE, USA, Canada, Australia, Malaysia, and many other countries. He has supervised several funded research projects and has interest in artificial intelligence and neural network in the field of civil engineering, reinforced concrete, and construction and construction management. One of his published articles in ScienceDirect was selected as top 25 hottest articles in 2007.