Defining Contractor Performance Levels

Khaled Nassar, Ph.D., AIA
Associate Professor, American University in Cairo

In many countries, state and private owners often categorize contractors into various groups based on their capacities as well as their previous performance. These categories are then used for several purposes including contract award and prequalification. A number of different statistical algorithms can be used to categorize the contractors. No one algorithm can be used globally in all situations since the performance of these algorithms depend partially on the actual data set being used. Therefore this paper presents a new model for grouping contractors based on historic data. The model developed here utilizes qualitative as well as quantitative measures about the contractors’ performance. The model addresses the issue of classifying contractors into various performance categories using different crisp and fuzzy clustering algorithms and assesses the performance of these algorithms with appropriate validity measures. The model was validated using actual data from previously recorded project information for 13 contractors. The analysis shows that the model can be used effectively to determine the performance level of the contractors and to perform clustering of contractors into different performance groups.

Keywords: Contractor Performance, Competency, Project Management, Building Construction

Introduction

Many owners are becoming more aware of the fact that the lowest bid does not always result in the lowest cost. Several owners have been awarding contracts based on a financial bid as well as a technical bid. Various evaluation methods have been used to evaluate the technical bids. Ranking the contractors for the purpose of awarding contracts is a very important process and has obvious severe implications for the owner and contractor alike. When evaluating the contractor performance, owners often classify the contractor into different performance groups based on their historic performance, which is the topic of this paper.


However, the research of (Singh and Tiong, 2005) is more relevant to this paper since the researchers developed a fuzzy decision framework for contractor selection. They presented a systematic procedure based on fuzzy set theory. The procedure was intended to evaluate the capability of a contractor to meet the owner’s requirements in terms of cost, time and quality. Shapley value was the main concept used to determine the global value or relative importance of each criterion in accomplishing the overall objective of the decision-making process. However, no algorithm or validation techniques were proposed. This is a major drawback since using different crisp and fuzzy techniques will usually result in different results. This is unacceptable practically and results in a situation where a contractor can be classified in one level using a specific algorithm and in a higher or lower level using another algorithm.

This paper presents a procedure carried out to rank contractors working in the Dubai. The work presented here was carried out on behalf of a large owner/developer for ranking the contractors bidding for the developer. The main goal was to categorize contractor into similar categories in terms of their performance defined by several attributes including their previous bidding performance. The developer, who collected historical data about the contractors it hired, was interested in categorizing the contractors into various levels of performance and not
Average
63.4445
73.1407
8.21776
69.9224
27.0016
76.3118
56.9419
23.4568
72.1218
60.3609
14.1917
5.403
72.4621
52.0563
19.2519
Table 1: Normalized Contractor Data

<table>
<thead>
<tr>
<th>Schedule Measures</th>
<th>Cost Measure</th>
<th>Safety Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Delay</td>
<td>Average</td>
<td>Total</td>
</tr>
<tr>
<td></td>
<td>number of</td>
<td>liquidated</td>
</tr>
<tr>
<td></td>
<td>late jobs</td>
<td>damages</td>
</tr>
<tr>
<td></td>
<td></td>
<td>charged</td>
</tr>
<tr>
<td>84.887</td>
<td>80.6194</td>
<td>42.3868</td>
</tr>
<tr>
<td>19.2519</td>
<td>54.2306</td>
<td>69.4147</td>
</tr>
<tr>
<td>52.0563</td>
<td>56.5077</td>
<td>49.9459</td>
</tr>
<tr>
<td>72.4621</td>
<td>66.7635</td>
<td>50.7969</td>
</tr>
<tr>
<td>5.40359</td>
<td>68.543</td>
<td>49.3667</td>
</tr>
<tr>
<td>14.1917</td>
<td>99.1761</td>
<td>82.3556</td>
</tr>
<tr>
<td>60.3609</td>
<td>32.7176</td>
<td>5.98519</td>
</tr>
<tr>
<td>72.1218</td>
<td>9.98966</td>
<td>68.5222</td>
</tr>
<tr>
<td>23.4568</td>
<td>98.4688</td>
<td>22.679</td>
</tr>
<tr>
<td>56.9419</td>
<td>13.0701</td>
<td>88.8126</td>
</tr>
<tr>
<td>76.3118</td>
<td>67.7162</td>
<td>27.9714</td>
</tr>
<tr>
<td>27.0016</td>
<td>65.1732</td>
<td>46.8571</td>
</tr>
<tr>
<td>69.9224</td>
<td>99.2644</td>
<td>68.6673</td>
</tr>
<tr>
<td>8.21776</td>
<td>59.3133</td>
<td>69.8207</td>
</tr>
<tr>
<td>73.1407</td>
<td>65.9579</td>
<td>96.9801</td>
</tr>
<tr>
<td>75.658</td>
<td>75.2405</td>
<td>3.64596</td>
</tr>
<tr>
<td>63.4445</td>
<td>55.1375</td>
<td>54.4996</td>
</tr>
</tbody>
</table>

Data Normalization for Clustering

Clustering techniques are among the unsupervised methods of data mining which aim at classification of objects based on similarities among them. The term “similarity” should be understood as mathematical representation of the similarity criteria under consideration. The performance of most clustering algorithms is influenced by the geometrical properties of the individual clusters but also by the spatial relations and distances among the clusters. Therefore various clustering algorithms were used in order to classify the contractors into different clusters. We used the toolbox for Matlab by Balaz et al 2008.

Various performance measures were considered for the analysis. Quantitative measures where calculated from a historic database. These quantitative measures can be broken down into 3 main categories; schedule, cost and safety. Qualitative measures on the other hand were assessed subjectively using the Analytical Hierarchical Process. Therefore, a data set containing all the performance measures versus the 21 contractor was compiled. Eight performance measures were used in all: the average delay, the average number of late jobs, ratio to average bid, Disabling Injury Severity Rate, the Average Days Charged, managerial and customer service, environment and sustainability (the last two being the qualitative measures). More information on the calculation and collection of these measures can be found in Nassar 2008, however in this paper we will focus on the algorithms used as described in the next section.
Before clustering can be carried out, the data had to be normalized as the various performance measure recorded was in a different scale (i.e. the delay may be in days, where as the weighted delay is in dollar.days). Normalization therefore entails setting a fixed scale for all the data. This can be done using by scaling with relation to the minimum and maximum value of each criterion, or alternatively through normalization through the variance which was the technique used in the analysis presented here due the relatively small data sample. The following equation was used for normalization:

\[ X = \frac{X_{\text{old}} - \bar{X}}{\sigma_X} \]

Figure shows 1 three sets of charts; first the un-normalized raw data for the average delay performance measure, and second the same data after normalization according to the min-max and finally the data normalized by variance which was used in this research.

Once the data has been normalized, the main goal becomes trying to fit contractors in various performance categories. Here we must first try to determine the number of performance categories to use and secondly determine which of the different clustering algorithm will produce the best results (by best we mean most consistent). This means that if one were to use a certain clustering algorithm a specific contractor may be assigned to one performance category while the use of another algorithm may result in the same contractor being assigned to lower or higher performance category. This is obviously unacceptable to the owners or the contractors who need a reliable way to classify contractors. Therefore a number of different algorithms were used as described below.
Consider the various projects in the data set, as an $n$-dimensional row vector $x_k = [x_{k1}, x_{k2}, ..., x_{kn}]^T$. A set of $N$ observations is denoted by $X = \{x_k | k = 1, 2, ..., N\}$, where $n$ is the number of contractors and $N$ are the various performance measures considered. The goal is to find a partition matrix $U = [\mu_{ik}]$. The first algorithms considered were two typical hard clustering algorithms, namely K-means and K-medoid. These are simple and popular, though the results are not always reliable. For an $N \times n$ dimensional data (where $n$ is the number of contractors and $N$ are the various performance measures considered) one of $c$ clusters is allocated by minimizing the sum of squares, i.e.

$$
\sum_{i=1}^{c} \sum_{k \in A_i} \|x_k - v_i\|^2
$$

where $A_i$ is the set of data points in the $i$-th cluster and $v_i$ is the mean of those points. In K-medoid clustering the cluster centers are the nearest objects to the mean of data in one cluster $V \| x_i^c \in X \| i < c$. The results of the C-means clustering are shown in figure 2 for two of the performance measures; ration to average bid and average delay.

Next we considered the Fuzzy C-means algorithm which is based on minimizing an objective function defined as

$$
\sum_{i=1}^{c} \sum_{k \in A_i} \|x_k - v_i\|^2
$$
The objective function is actually a measure of the total variance of $x_k$ from $v_i$. The minimization of the c-means function represents a nonlinear optimization problem that can be solved by using a variety of available methods, ranging from grouped coordinate minimization, over simulated annealing to genetic algorithms. The most popular method, however, is a simple Picard iteration which what was used in our research. Another Fuzzy algorithm considered was the Gustafson and Kessel extension of the fuzzy c-means algorithm by employing an adaptive distance norm, in order to detect clusters of different geometrical shapes in one data set. Each cluster has its own norm-inducing matrix $A_i$, which yields the following inner-product norm:

$$ J \left( \mathbf{U}, V \right) = \sum_{i=1}^{c} \sum_{k=1}^{N} w_{ik} \sum_{j=1}^{m} \| x_{ik} - v_j \|_A^2 $$

where,

$$ V = v_1, v_2, \ldots, v_c, v_j \in \mathbb{R}^n $$

is a vector of cluster centers, which have to determined and

$$ D_{ik}^2 = \| x_{ik} - x_{ik} \|_A^2 = \mathbf{w}_k - x_{ik} A \mathbf{w}_k - x_{ik} $$

is a squared inner-product distance norm.

Figure 3: The results based on Fuzzy C-means and the Gustafson and Kessel algorithms

However, the objective function cannot be directly minimized with respect to $A_i$, since it is linear in $A_i$. This means that $J$ can be made as small as desired by simply making $A_i$ less positive definite. To obtain a feasible solution, $A_i$ must be constrained in some way. The usual way of accomplishing this is to constrain the determinant of $A_i$.

The last algorithm considered is the fuzzy maximum likelihood estimates (FMLE) clustering algorithm, which employs a distance norm based on the fuzzy maximum likelihood estimates, proposed by Bezdek and Dunn as:

$$ D_{ik} \left( \mathbf{U}, V \right) = \frac{\sqrt{\det \mathbf{C}_{ii}}}{\alpha_i} \exp \left( \frac{1}{2} \mathbf{w}_k - v_{ij} \right) F_{w_i}^{-1} \mathbf{w}_k - v_{ij} $$
Note that, contrary to the GK (Gustafson and Kessel) algorithm, this distance norm involves an exponential term and thus decreases faster than the inner-product norm. The membership degrees are interpreted as the posterior probabilities of selecting the appropriate cluster for each contractor given the data point \( x_k \).

![Figure 4: The results based on FMLE algorithm](image)

The question now becomes which of the above algorithms to choose in order to classify the contractors appropriately. The answer is determined by evaluating each of the above algorithms according to various validity measures as described next.

### Validity Measures

Different validity measures have been proposed in the literature, none of them is perfect by oneself. Therefore we used several indices to compare the various algorithms. The first validity measure considered is the Partition Coefficient (PC): measures the amount of "overlapping" between the clusters and is defined as:

\[
PC(c) = \frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{N} \mu_{ij}^2,
\]

where \( \mu_{ij} \) is the membership of data point \( j \) in cluster \( i \). The most prominent disadvantage of PC is lack of direct connection to some property of the data itself. Another measure considered is the classification Entropy (CE) which measures the fuzziness of the cluster partition only as,

\[
CE(c) = \frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{N} \mu_{ij} \log \mu_{ij}
\]

The Partition Index (SC) on the other hand is the ratio of the sum of compactness and separation of the clusters. It is a sum of individual cluster validity measures normalized through division by the fuzzy cardinality of each cluster and is given by,

\[
SC(c) = \frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{N} (\mu_{ij})^m \left\| x_j - v_i \right\|^2 \\
\frac{N_i \sum_{j=1}^{N} \left\| x_j - v_i \right\|^2}{N_i \sum_{j=1}^{N} \left\| x_j - v_i \right\|^2}
\]

The Separation Index (S) uses a minimum-distance separation for partition validity.
\[
S(c) = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} (\mu_{ij})^2 \| x_j - v_i \|^2}{N \min_{i,j} \| x_j - v_i \|^2}
\] (4)

Other measures considered are: the Xie and Beni’s Index (XB) which aims to quantify the ratio of the total variation within clusters and the separation of clusters, Dunn’s Index (DI) which identifies compact and well separated clusters, and the Alternative Dunn Index (ADI) which aims at modifying the original Dunn’s index to become more simple, when the dissimilarity function between two clusters is rated in value differently. All these validity measures where calculated for the various clustering algorithms mentioned above and is displayed in Figure 5.

![Figure 5: Comparison of algorithms based on validity measures](image)

The results show that the best performing algorithm for our data set was the Fuzzy C-means algorithm. As such the contractors were classified according to that algorithm which showed clear clusters of contractors in 4 main groups based on the classification data shown in Table 1.

**Conclusions**

One of the main drawbacks of trying to group the various contractors into different performance categories is that the numbers of the data groups have to be decided a-priori. This may be a drawback since contractors may argue as to the validity of the categorization vis-a-vis the number of performance categories used and the rationale for deciding on a specific number of categories (i.e. a contractor may be grouped in the second performance category when contractors are grouped into 6 different categories, but may be grouped in the first if only 4 categories are selected). Therefore this paper presented a technique which can overcome these limitations and possibly open the way for the wider implementation of contractor classification in the construction industry, which in turn can be used for various managerial and contractual purposes.

**References**

Balazs Balasko, Janos Abonyi and Balazs Feil, (2008). Fuzzy Clustering and Data Analysis Toolbox For Use with Matlab


Luis Fernando Alarcón and Claudio Mourgues (2002), Performance Modeling for Contractor Selection, J. Mgmt. in Engrg., Volume 18, Issue 2, pp. 52-60


