Selection of the Project Delivery Systems for China’s Construction Projects with Artificial Neural Network and Data Envelopment Analysis

Jia Hu
Rinker School of Building Construction

Abstract

Purpose – The suitability of the project delivery system (PDS) selected for a project greatly influences the efficiency with which the project is executed. It is not an easy task to select an appropriate PDS as a large amount of ambiguous information exists. The aim of this paper is therefore to develop a PDS selection Model to help owner’s decision-making.

Design/methodology/approach – Similar projects are identified through the similarity metrics between the target project to be predicted and those in the database. Then some of the indicator values are examined and modified through DEA-BND model and then are trained by ANN model to predict an appropriate PDS for the target project. A survey was conducted by postal questionnaire to empirically validate the reliability of the model.

Findings – The indicator system of the PDS selection is established. Through the comparison of predicted results from different models, it is found out that the developed PDS selection model in this paper can predict PDS more precisely and shows higher reliability than the ANN model.

Originality/value – A new PDS selection model is developed by inputting project-specific data, which proves to be more accurate and less dependent on experts’ judgment. Its practical application will benefit the owner’s decision making in selecting the PDS.

Key Word: Construction Project, Project Delivery System, Data Envelopment Analysis, Artificial Neural Network, China

Introduction

Project Delivery system (PDS) describes how the project participants are organized to interact, transforming the owner’s goals and objectives into finished facilities (ASCE, 2000). There are several PDSs in China’s construction industry, such as Design-Bid-Build (DBB), Design-Build (DB), Construction Management (CM) and Engineering Procurement and Construction (EPC) (Chen, 2005).

PDS, as one of the critical factors of project success, affects project schedule, cost, quality and contract management to a great extent (Kumaraswamy, 2001; Chan, 2001; Khalil, 2002). Selecting an appropriate PDS can improve the project performance effectively (Oyetunji and Anderson 2006).

Different PDS selection models were once developed. Project performance under different PDSs can be predicted to serve as a basis of the PDS selection. Konchar (1998) empirically compared cost, schedule and quality performance of CM at risk, DB and DBB, using project-specific data collected from 351 U.S. building projects. However, the author did not definitely point out how to apply the research into PDS selection. Ling (2004b) developed models to predict performance of DBB and DB projects in Singapore through multivariate analysis.

Some other methods like AHP were also proposed. The priority of PDSs can be determined through the pairwise comparison matrix (AlKhalil, 2002; Mahdi, 2005). However, the accuracy of AHP is interfered by the experts’ uncertain and subjective judgments. Mafakheri (2005) utilized the interval AHP to determine the interval priorities for alternative PDSs and Rough Set theory to fully rank the
alternatives. But the full ranking depends on high degree of risks, which increases the inaccuracy. Besides, AHP tends to require a set of established indicators, including project participants, project characteristics and external environment (Alhazmi, 2000; Mahdi, 2005; Mafakheri, 2007). It is too complex if a large number of indicators are used. Careful selection of indicators is needed to reduce the number as well as their correlation.

Multi-attribute utility is a simple method for PDS selection (Love, 1998; Chan, 2001). The overall utility is calculated through multiplying the weights by utility of indicators. Speed, certainty, flexibility, quality, complexity, risk allocation, responsibility, arbitration and disputes, and price competition are often identified as its common indicators. The simplicity makes this model easy for practical use. However, the utility values of indicators always fail to reflect the actual status and the project may not achieve the specified objectives as initially expected.

Case-Based Reasoning (CBR) is a name given to a reasoning method that uses specific past experiences rather than a corpus of general knowledge (Riberiro, 2001). Luu (2003, 2005, 2006) has done much work in this field. CBR method is also introduced into this field. The cases (projects), similar with target one, are retrieved by calculating the similarity metrics between target and exemplar ones. Luu (2005) proposed indicators for CBR from the perspectives of owner’s characteristics and objectives, project characteristic and external environment. However, CBR oversimplified the selection process. The overall similarity metric is measured by linearly calculating similarity of respective indicators.

Many endeavors have been made to overcome the existing defects. Artificial neural network (ANN) is introduced as an attempt to overcome the existing defects. In the ASCE Journal of Computing, for example, over 12% of papers published from 1995 to 2005 (54 out of 445) have used the term “neural” as part of their title (Flood, 2008). Ling et al. (2004) developed a nonlinear prediction method using ANN to predict the performance of DB projects, which could be used as a basis of PDS selection. But it should say that it is not easy to apply ANN models to the PDS selection because of the difficulty in acquiring the real-world project data. Data Envelopment Analysis (DEA) is used to measure the efficiency of DMU (Decision Making Unit) (Charnes and Cooper 1978). Simos and Maroulis (2007) applied DEA to calculate the efficiency of DBB, DB, CM and DBM in road projects. But this research is limited to road projects.

It is necessary to improve the current existing methods to enhance their prediction accuracy and effectiveness due to the importance of the role of the PDS played in China’s rapidly developing construction industry. Therefore, for solving the current existing problems, the aim of this paper is to develop an accurate and reliable PDS selection model to solve the current existing problems.

Research Methods

Artificial Neural Network

Artificial Neural Network (ANN) does not fully rely on existing experiences and knowledge and can extract the internal rules from the training samples using specified learning rules, through which the internal relationship between inputs and outputs can be found out. Compared to the multiple regression method, weights are not needed to consider and ANN has stronger self-adaptation and fault-tolerant abilities. In this paper, the PDS is defined as the output of ANN, while indicators of the PDS selection are defined as its inputs.

Data Envelopment Analysis
Data Envelopment Analysis (DEA) doesn’t need pre-estimated parameters. It has superiority in avoiding subjective influence, simplifying calculation and reducing deviation. In this paper, BND (Bound Variable) model, one of the DEA models, is selected to examine the validity of project data and then indicator values will be correspondingly modified. Its most significant characteristic is to allow the input and output indicators to fluctuate between the specified bounded ranges. Those data with significant deviation shall be corrected by BND model for two reasons: first, the average Production Possibility Frontier (PPF) can be generated and the significant deviation of individual DMU will be reduced by adjusting input indicators. Second, the efficiency value of a project will be changed into “1” (maximal efficiency) by adjusting input indicators into projection values, which will replace its original values to achieve the modification of indicator values.

**Indicators of the PDS Selection**

The indicator system is an integral part of the prediction model. Establishment of a universal set of indicators for PDS selection seems to be an endless issue which has become a black hole to attract researchers on PDS selection. It is not necessary to hold up all the technical development to wait for establishing such indicators (Chan, 2007). Based on the literature review on previous studies, the indicators can be categorized into four groups: project objectives, project characteristics, characteristics of owner and contractor and external environment.

Indicators of the PDS selection, identified from 15 papers, fall into the above-mentioned four groups mentioned above. As to indicators with the same or similar meanings but with different expressions, only one expression form is selected here for purpose of clarification and unification. For example, “flexibility” is selected to represent the indicator “Potential for design changes during construction”.

**Table 1  List of the indicator of the PDS selection**

<table>
<thead>
<tr>
<th>Indicators</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Project objectives</strong></td>
<td></td>
</tr>
<tr>
<td>Delivery speed</td>
<td>✓</td>
</tr>
<tr>
<td>Schedule delay</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>Cost growth</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>Cost certainty</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>Quality performance</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td><strong>project characteristics</strong></td>
<td></td>
</tr>
<tr>
<td>Project type</td>
<td>✓</td>
</tr>
<tr>
<td>Project Scale</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>Complexity</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>Ability to define the project scope</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>Flexibility</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>Disputes</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td><strong>Characteristics of owner and contractor</strong></td>
<td></td>
</tr>
<tr>
<td>Owner’s willingness to be involved</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>Owner’s willingness to take risks</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>Owner’s available human resources</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td>Contractor’s capability</td>
<td>✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓</td>
</tr>
<tr>
<td><strong>External environment</strong></td>
<td></td>
</tr>
</tbody>
</table>
Project schedule, cost and quality are commonly used as project objectives and they are also identified for the PDS selection. Schedule delay is the most frequently used indicator for project schedule. Delivery speed is eliminated because it is mainly used for building projects and not comparable among different project types. Therefore, only Schedule delay is selected to represent the project schedule.

As to project cost, Cost growth and Cost certainty are both frequently used. Cost certainty means the certainty over the cost for completion of the project. As seen from Table 1, most researchers selected only one of them, because a high correlation exists between them. Only Cost growth is selected in this paper.

Project quality is measured with varied indicators by different researchers. In this paper, the formula developed by Ng (2002) is adopted to represent the level of Project quality, which is divided into Material quality, Workmanship quality and Design quality.
For *Project characteristics*, six indicators listed Table 1 are selected and can be seen from Table 3. The formulas of *Project complexity* and *Flexibility* are selected from the paper of Ng (2002).

Table 3 Description of indicators of project characteristics
Project complexity
Number of extra working packages resulting from the need for extra complexity
Number of total working packages of a project with basic complexity =

Ability to define the project scope

Variation caused by the indefinite project scope at contract signing
Total Variation =

Variation caused by the indefinite project scope at contract signing
Total Variation =

Flexibility

Anticipated adjustments
Original contracts sum =

Anticipated adjustments
Original contracts sum =
For characteristics of owner and contractor, the five indicators listed in Table 1 are selected. *Owner’s willingness to be involved* and *Owner’s willingness to take risks* are selected from the paper of Ng (2002).

Table 4  Description of indicators of characteristics of owner and contractor
For External environment, the indicators are measured by scale 1-9. As one of the widely used indicators, Market competitiveness means that “Level of competition in market with regards to this project” (Luu, 2005) and is subject to the number of available contractors. 1 = low competitiveness and limited available contractors; 9 = high competitiveness and extensively available contractors. Regulatory feasibility measures the degree to which the regulations influence the implantation of project. 1 = adverse influence; 9 = favorable influence. Technology availability measures the degree of availability of necessary technologies from market. 1 = quite hard to acquire necessary technologies; 9 = quite easy to acquire technologies.

The indicator value is expressed in the form of the triangular fuzzy number, which involves more information than the crisp number or expected value. When surveyed, respondents are required to give the triangular fuzzy number which includes three numbers, two are the boundary value and one is the most probable value.

**PDS Selection Model**

**Framework of PDS Selection Model**

By combined usage of ANN and DEA, the proposed model is composed of three parts: Selecting Similar Projects, Examining Indicator Values, and Training and Predicting. The framework of the model can be shown in Fig. 1.

![Fig. 1 Framework of the PDS selection Model](image)

**Stage 1 Selecting Similar Projects**

Similar projects are selected to increase the prediction accuracy of the ANN model. Then, according to the provided selection procedure from Fig. 2, choose projects which are similar with the
target project. The first step is to establish a project database including the **needed indicator** data. It is assumed that the project database has been established and have enough project data in this paper. Then, the work of selecting similar projects can be carried out according to the procedure shown in Fig.2.

As shown in Fig. 1, the first step is to retrieve projects from the project database. Based on the **indicator** of **project type**, projects of the same type are chosen as the target ones. Then, retrieval II is executed to select the most similar projects according to the **indicators** of **project scale**, **project complexity** and indicators of **external environment**. **Specify the critical value of the similarity metric to determine which projects are selected as similar projects.** The project whose similarity metric exceeds the critical value would be selected and those below will be eliminated, in order to ensure selecting enough similar projects with high degree of similarity.

Some indicator values are **hard** to define because the target project has not been started when the PDS selection is made. By inputting different values, alternative PDSs are predicted, from which the owner can choose the optimal one.

![Diagram](image)

**Fig. 2  Selection procedure of similar projects**

The similarity between the target project and projects in database should be calculated when retrieval II is implemented. The similarity between the target project T and project R from database can be calculated by the following formula:

\[
S(T, R) = \sum_{i=1}^{n} w_i \cdot \sin(f_i^T, f_i^R)
\]

Thereinto, \( n \) = the number of indicator; \( W_i \) means the weight of each indicator ( \( W_i \geq 0 \), \( \sum_{i=1}^{n} W_i = 1 \) ). The weights are directly given by owner. \( f_i^T, f_i^R \) respectively represent the value of indicator \( i \) of project \( T \) and project \( R \), and the value is either crisp or fuzzy. \( \sin(f_i^T, f_i^R) \) = the similarity between project \( T \) and project \( R \). Both the range of \( S(T, R) \) and \( \sin(f_i^T, f_i^R) \) are between 0 and 1. The similarity increases as the value of \( S(T, R) \) becoming larger. When it reaches 1, the two projects are completely matched.

\( \sin(f_i^T, f_i^R) \) can be calculated in two ways considering the following two situations:

1) When the indicator value is crisp number,
Thereinto, $f^T_i$, $f^R_i$ respectively represent the value of indicator $i$ of the target project and projects in the database; $\alpha$ and $\beta$ respectively represent the upper and lower bound of the value.

2) When the indicator value is fuzzy, and if the indicator $i$ of project $T$ and project $R$ satisfy:

$$f^T_i = (f^T_{i1}, f^T_{i2}, f^T_{i3}), f^T_{i1} \leq f^T_{i2} \leq f^T_{i3}$$ and

$$f^R_i = (f^R_{i1}, f^R_{i2}, f^R_{i3}), f^R_{i1} \leq f^R_{i2} \leq f^R_{i3},$$

then Graded Mean Integration-representation Distance (Hsieh, 1999) method is applicable.

Stage 2  Examining Indicator Values

Part of the indicator values of similar projects are examined and modified in order to be effectively applied into the ANN model. A DEA-BND (Bound Variable) model is selected to evaluate the efficiency of owner’s management during the construction stage (owner’s management during this stage is DMU here). In order to apply DEA-BND model, the value of each input indicator is expressed by triangular fuzzy number, which includes two boundary number and one probable number. Their boundary ranges can be obtained by questionnaire. By measuring the efficiency of projects using the DEA-BND model, the projection values of each input indicator can be calculated and substitute the original values as amendments to modify the input indicators.

1) Input indicators. Input indicators are chosen from three aspects: time, personnel and expense. The three indicators, Owner participation, owner’s available personnel and project scale, respectively represent time, personnel and expense inputs. Since the owner’s management is affected by contractor’s capability, the indicator contractor’s capability is also selected here.

2) Output indicators. As the owner’s management is to be evaluated by DEA-BND model, it is necessary to convert the project performance into indicators that can represent the output of owner’s management. This paper proposes the indicator of Project output satisfaction as the output indicator of DEA. It can measure the owner’s satisfaction to the outputs of management activities. Project output satisfaction can be generated from three aspects which are schedule, cost and quality. The formula is as follow:

$$S_{abc} = \alpha T + \beta C + \gamma Q$$

Where, $S_a$ = project outputs satisfaction, $T$ = owner’s schedule management satisfaction, $C$ = owner’s cost management satisfaction and $Q$ = owner’s quality management satisfaction. $a$, $b$ and $c$ respectively represent the importance weights of $T$, $Q$ and $C$, which can be specified by owner.

As satisfaction is quite difficult to be measured accurately, in this paper it is acquired from the value of schedule, cost, and quality management satisfaction, which can be respectively obtained by converting schedule delay, cost growth and project quality into values between from 1 and 100 using linear calculation, convex function or concave function.

Stage 3  Training and Predicting

Applying the ANN into training the modified similar projects, and after that, formulate the
nonlinear function between indicators and PDS to predicate which PDS is suitable for the target project. Back Propagation (BP) ANN is a multi-layer feed-forward neural network based on error back-propagation algorithm. In the application of ANN model, BP ANN is widely used. Therefore BP ANN is applied here to train and predict the similar projects.

The correlation of input indicators needs to be tested before the application of BP ANN model. Factor Analysis can be used to reduce or eliminate the correlation among the input indicators. A potential precondition of factor analysis is that there should be strong correlation among indicators. By KMO (Kaiser-Meyer-Olkin) and Bartlett’s test of sphericity, the correlation coefficient should exceed the commonly used value 0.3.

For BP ANN, set the random number between (-1, 1) as the original weights and design one hidden layer. The transfer function from the input layer to the hidden layer is single-stage Sigmoid function and the transfer function from the hidden layer to the output layer is linear function. The minimum error is set as 0.01. The PDS can be predicted by inputting the indicators of the target project into the trained BP ANN model.

**Questionnaire Design**

Questionnaire is designed according to the established indicator system. Values of the indicators can be obtained in two ways: (1) indicators whose variables in the formula can be obtained from respondents, values of them are calculated by formulas; (2) Values of them can be either defined by scale 1-9 or directly estimated by who. All the indicator values obtained by questionnaires should be normalized into the scale 1-9.

The respondents are required to provide the maximum and minimum of the indicator values. If they fail, then the fluctuation boundary will be 1, in other words, the indicator values will be changing between (value-1, value+1).

152 questionnaires in total were sent out to managers of owners and other participants of projects undertaken by the top 100 contractors of China. After 3-month questionnaire survey, 132 questionnaires were returned, 81 of which were valid. Among the projects referred in the questionnaires, up to 90% were invested more than 10 million RMB; over 70% were invested more than 100 million RMB; 43 of those were DB projects (DB projects in this paper contains EPC projects); 38 of those are DBB projects.

According to the analysis on the project type, as to DB projects, the three project types respectively account for 60%, 28% and 12%; while as to DBB projects, the three project types respectively account for 26%, 26% and 48%.

**Data Analysis**

**PDS Prediction**

1) The Project #003 is randomly selected as the target project, and other projects are treated as projects in a project database. Considering the number of questionnaire is small, it is impossible to take account of all the project types. Thus the “Retrieval I” is omitted. When calculating the similarity metrics in the part of “Retrieval II”, this research assumes the weight of each indicator is equal. So, it is calculated that the maximum value is 0.89, and the minimum value is 0.35. Set critical value as 0.7, there are 31 projects exceeding 0.7, including 7 DB and 14 DBB projects.
2) Before applying the DEA-BND model, the boundary ranges of input indicators can be found from the received questionnaires. The output indicator is converted by indicators of project objectives using linear function. So, the input indicators are renewed by using projection values.

3) By calculating KMO (Kaiser-Meyer-Olkin) and Bartlett’s test of sphericity, the correlation coefficient turns out to be 0.51, which satisfy the precondition of factor analysis. As to determining the number of factors, it is required that either the accumulation contribution rate exceeds 85% or its eigenvalue exceeds the mean value.

Factor analysis is carried out through SPSS (Statistical Package for the Social Science). The 16 factors are extracted through factor rotation and among them the 10 factors with largest contribution rate are selected. Their accumulative contribution rate is 93.4%, and SPSS analysis shows there are very low correlation coefficients among them. The indicators processed by factor analysis are regarded as input indicators of the BP ANN model. The output indicator is PDS, among which the value of DBB is set as 1 and the value of DB is set as 2.

The indicator values of the target project are shown in Table 5. BP ANN is constructed to train and predict the modified similar projects. With the tool of MATLAB, the predicted result of Project #003 is 1.3115 (the actual value is 1).

Table 5   Indicator values of target project
Reliability of Model

The predicted result of target Project #003 is very close to the actual value. To validate the predicting reliability of the developed model, a comparative analysis is done in this paper.

Firstly, select a group of control projects, whose codes shown in questionnaires are in order and starting from 001#. Meanwhile, the amount should be the same as that of projects similar to Project #3, namely 31 projects with 14 DB projects and 17 DBB projects. By factor analysis, the 10 first factors with the highest contribution rate are selected as input indicators of BP ANN model.

Because the initial weights of ANN model are determined randomly, the model developed in this paper and the model developed by using control samples are run continuously for 10 times, and the result is shown as Table 6. The third column is the predicted result of BP ANN model which is trained
by using the 31 control projects. It can be seen that the mean value of the model developed in this paper is 1.11, and the sample variance is 0.018. Compared with the result of control projects, the developed model is closer to “1”, and the variance is smaller, which indicates that the model is reliable.

Table 6  Comparison of predicted values of different models (Project 003#)
Using the same analysis method, Project #018 and Project #035 are randomly selected to further validate the model reliability, the predicted mean values and variance values of the two projects are shown in Table 7. Similar conclusion can be reached as Project #003. If a large amount of projects are obtained and the “retrieval I ” is considered, the reliability of the developed model are likely to be higher.

Table 7  Predicted value comparison for different models (Project 003# and 035#)
Conclusion

Sixteen indicators from four categories (project objectives, project characteristics, characteristics of owner and contractor and external environment) are identified through literature review. The PDS selection model is developed with ANN and DEA methods. It is empirically found out that the developed model shows higher prediction reliability than ANN model. Project-specific data are inputted to improve the accuracy of prediction and reduce the deviation of experts’ subjective judgments. Project database is the basis of model prediction, so it is only by establishing the project database that the model can be more likely applied into practice, which is also the future work of this research. Besides, the model is compared with ANN model in this paper. The reliability of the model can be further proved through comparison with other PDS selection models.

Acknowledgement

The authors would like to acknowledge the support received from the National Natural Science Foundation of China funded project (No. 70772057), Charted Institute of Building (CIOB) and all the respondents of questionnaire.

Reference


[27] Mahdi I. M., Alreshaid K., Decision support system for selecting the proper project delivery method using analytical hierarchy process (AHP). International Journal of Project Management,


