Abstract: Human resource allocation holds the key to success in the construction of labor-intensive public infrastructures. However, due to the difficulty in obtaining relevant onsite data, there has been a lack of quantitative numerical analysis and discussion in published literature. In this study, a database for assessing human resource allocation in pavement engineering was established by collecting detailed information from various construction projects. Fourteen influence factors were summarized through literature review and consultation with experts in the field. Thirty two road-smoothing projects were then randomly selected. Using the rough set approach and an artificial neural network model, a model for assessing human resource allocation in pavement engineering was developed. The model validity is verified by an average accuracy of 88.63%. Therefore, this proposed model can be viewed as a useful tool for estimating human resource demand in pavement engineering. It can also effectively alert the authority to avoid a shortage in manpower, preventing the construction project from falling behind schedule or even early termination as a result of inappropriate resource allocation.

DO:10.6135/ijprt.org.tw/2013.6(5).643

Key words: Allocation; ANN; Project human resources; Rough sets.

Introduction

Background and Motivation

Construction project engineering involves a special kind of project planning. Viewing the entire life cycle of a project, it includes planning, design, construction, and the following operation management. The construction stage, which is a labour intensive phase, often requires a considerable amount of labour input. Therefore, human resource planning becomes an important factor that needs consideration. At present, Taiwan is faced with a declining birth rate which, in combination with the massification of higher education, has resulted in the younger generation being reluctant to work in the construction industry.

In an effort to ensure economic development, expand domestic demand, and maintain the momentum of domestic economic growth, the government has been focusing on the development of several public infrastructure projects. However, according to the recent statement from the Taiwan Construction Engineering Association, in recent years as the economy recovers, construction projects in the private sector are thriving, while, those in the public sector are faced with an increasing shortage of skilled labor. This has resulted in projects falling behind their schedules. The completion date is often affected by factors that constructor are not liable for, which leads to many contractual disputes.

Studies conducted on projects that fall behind their schedule show that a shortage of manpower is often responsible for delays in construction. Currently, domestic construction companies allocate their on-site resources based primarily on past experiences.

Discussion in literature of the factors influencing human resource planning in project engineering is mostly limited to a case-by-case qualitative description. Some studies use questionnaires to gather information, but only focus on quantifiable factors, while unquantifiable conditional factors are neglected or not specifically analysed due to a lack of data. As a result, there is no numerical analysis or quantitative discussion of human resource allocations and the existing assessment approaches lack completeness and reliability. In view of this, it is the purpose of this study to construct a quantitative model for assessing human resource allocation in project engineering. The proposed model could assist the authority and construction firms with on site resource demands and help allocate workforce accordingly, therefore, ensuring the delivery of quality construction projects.

Research Aim

Based on the discussion above regarding human resource allocation in the construction of public infrastructures, the following objective is proposed: Construct a model for assessing human resource allocation in project engineering and estimating the total manpower demand. The model can serve as a tool for the authority to examine the human resource allocations made by contractors. It can also assist contractors in making effective resource allocations. This can prevent a construction project from falling behind schedule, and reduce the number of related contractual disputes.

Research Limitations

Constrained by data availability, the adopted estimation method requires many hypothetical assumptions. Therefore, before the results can be analyzed, the content and nature of the estimated data must be understood in order to avoid a misunderstanding of the research results.

1 Quantitative data of human resources in project engineering is difficult to obtain, and therefore discussion of the factors affecting human resource allocation in such settings has
generally been qualitative and on a case-to-case basis. Questionnaires have been used to gather relevant information, however, until this study, there has not been any quantitative numerical discussion. Data used in the analysis and discussion of influence factors in the proposed model are collected from the bidding and tender documents that are required in public project procurement. The data should contain the basic characteristics of the projects, and are therefore sufficient for the data exploration technique used in the model construction.

(2) The obtained data pertains to human resource allocation for road-smoothing projects in pavement engineering. Among human resources is referred to blue-collar workers. The proposed assessment model is still applicable in other types of engineering projects, however, the important core attributes from the rough set have to be re-analyzed and selected accordingly.

(3) The current study only focuses on demand estimation of the “quantity” of onsite workers in public engineering projects. There is no data for assessing the “quality” of the workers, which is not part of the scope of this work.

Literature Review

Human Resource Management in Construction Projects

It is believed that the personnel in a company are the key to its competitiveness. Discussions of Human Resource Management (HRM) in a company or for a project began appearing in literature decades ago [1, 2]. At the beginning of the development of HRM certain industries were not specified. A basic survey was conducted to investigate human resource (HR) practices in the U.S. construction industry. The conclusion of this study contained common practices in HRM such as planning and a written statement on HR philosophy, making it a step towards helping practitioners appreciate the potential contribution of HR practices in the construction industry [3]. Yet a study published in 1996 pointed out that the promise shown by HRM has not been matched by its performance in the construction industry [4]. Little information was accessible on its applications to the construction industry in the late 20th century.

In recent years more and more scholars started to adopt HRM theories into the construction industry. This was seen as the start of a relationship between strategic HR configurations and organizational performance [5]. A later study found that most models explaining project success were based on theory rather than on empirical proof and that, numerous papers concentrated on impact factors influencing project success [6]. For project-oriented companies such as companies involved in the construction industry, HRM is critical. An overview study provided findings supporting the above-mentioned viewpoint [7]. Findings associated with most project HR allocation decisions showed reactive [8]. A study to understand employee resourcing in construction was carried out, findings included the discovery of weak relationships among deployment, HR planning, team deployment, performance management, employee involvement, and training development [9]. A further study proposed that strategic management of HR in construction increased productivity, reduced absenteeism, and reduced turnover. These findings were not revolutionary in themselves but they may be useful to construction firms [10]. Gareis (2005) suggested that construction firms were concerned with five points in terms of how to perform HRM [11]. Lately there have been more and more studies discussing HRM in construction projects, including studies on the allocation of manpower using advanced tools [12-15].

Applications Using Artificial Neural Networks and Rough Set in Construction

Applications using artificial neural networks (ANN) to solve problems have been widely discussed among industries for years. Studies using ANN can be found in numerous construction fields such as cost estimate, contractor selection, scheduling, budgeting, performance, quality controlling, knowledge sharing, and even risk hedging [16-18]. Outcomes yielded from ANN-enhanced approaches have solved problems in practice [19-23], thus, ANN is a typical and effective technique, which, provides an adaptive platform for developing hybrid approaches. Such studies are also common and successful in construction or construction related fields [24-26].

Rough set theory was introduced [27] and has been used to apply intelligent applications for recent years [28]. Various applications using rough-set feature selection depend on lower approximation to minimize data, and require no human input. A practical case that applied ANN and rough-set theory which, based feature section to ICP package products, explained the feasibility of using this method [29]. In recent years, scholars have become aware of the advantages of using a rough set approach but this has been limited to topics such as identification of measures for construction safety and productivity analysis [30].

Research Design

Research Structure and Procedure

In this study, the commonly adopted causal model in the research of human resource estimation is used as the foundation for prediction and objective mathematical quantification as the means. The research structure is shown in Fig. 1. There are three main stages. The first stage is data gathering, pre-processing, expert consultation, and derivation of the variables. In addition, specialists are needed to survey data in order to raise its accuracy. The second stage is model construction. Model 1 uses the back-propagation artificial neural network method, and model 2 uses both the rough set method and the back-propagation artificial neural network method. The third stage adopts a K-fold cross validation method to calculate the accuracy of data for both models, analyse the results and make suggestions.

The research procedure is as follows:

1. Determine 8 variables to be analyzed through literature review and consultation with experts in the field.
2. Log in the bidding and tender documents in government procurement, build files and a table of records.
3. Filter the data by eliminating extreme values and treating missing values. Obtain 32 sets of data.
Data pre-processing

Research variables

The first stage

Model 1

Rough set theory to determine core attributes

Back propagation artificial network algorithm

The second stage

Model 2

Back propagation artificial network algorithm

K-fold cross validation method

Compare the accuracy of Model 1 and 2

Results analysis and discussion

The third stage

Fig. 1. Workflow.

4. Select important attribute factors using the rough set theory.
5. Use back-propagation artificial neural network to perform training and testing. Calculate the number of hidden layers and the learning factor $\alpha$. Construct two models, with and without using the rough set approach, to select the important attribute factors.
6. Use the K-fold cross validation method to verify the reliability and accuracy of the data. Perform training and testing using $K = 10$.
7. Analyse and discuss the results from the two models.

Data Sources

This study collects data related to public infrastructures from bidding and tender documents in government procurement. Data from completed road-smoothing projects in pavement engineering in the past three years are analysed. According to statistics, the Taipei City Government and the Taoyuan County Government have carried out the most road-smoothing projects domestically. Data from 32 cases from these two organs are collected and analysed in this study.

Expert Consultation

Through consultation with 10 experts and scholars, who have served over 15 years in the field of public engineering, a common opinion is that in the proposed research, analysis of dimensions and factors affecting human resource allocation is complicated by the unique case-by-case nature of projects in the construction industry. Using information extracted from bidding and tender documents in government procurement, we hope to discuss systematically the dimensions and factors affecting human resource allocation in the construction industry. The idea of developing an assessment model that estimates on-site manpower allocation meets the approval of most experts and scholars. Bidding and tender documents are required of all organizations in charge of projects; therefore, they should be the most complete system for information on domestic public engineering projects. The documents contain the basic characteristics of such projects, and should be sufficient to derive the variables used in evaluating human resources. Experts and scholars approved the fact that published literature on onsite human resource management in construction projects often chose road area, construction duration, contract value, and site location as major influence factors. In addition, they also proposed other possible factors that could affect human resource demand. These include bid-to-budget ratio, administrative category and level of competency of the host organization, whether there is appropriate project management, and whether the planning, design and supervision are commissioned to professional engineering firms.

Variables Description

In the proposed model, the cumulative number of labours is a dependent variable, and independent variables are those derived from the bidding and tender documents in government procurement. The selected variables are introduced below:

1. Project duration
   The project duration is the number of days between the start date and the completion date of the project.
2. Contract value
   The contract value is the actual expense of the project, taking into consideration the effect of changes in the design.
3. Road area
   The road area of the project is a variable factor that measures the project scale.
4. Bid-to-budget ratio
   The bid-to-budget ratio is the ratio of bid price over project budget. It is often used as an indicator of underbidding.
5. Duration Calculation
   The duration calculation is concerned with how the project...
duration is calculated. There are three ways to calculate this variable: contract valid days, calendar days, and work days.

6. Site location
The site location is the location of the construction site.

7. Administrative category of the host organization
The host organization can be classified into two categories: central and local.

8. Administrative category of the competent body
The competent body can be classified into two categories: central and local.

9. Administrative category of the executive organ
The executive organ can be classified into two categories: central and local.

10. Professional category of the executive organ
The executive organ can be classified into two categories: professional and non-professional engineering entities.

11. Project Management Model
Whether or not a project management model is introduced to assist management.

12. Planning approach
Whether the planning is carried out by a hired professional consulting firm or by the host organization itself.

13. Design approach
Whether the design is carried out by a professional consulting firm or by the host organization itself.

14. Supervising approach
Whether the construction supervision is carried out by a professional consulting firm or by the host organization itself.

The subjects of the present study are completed road-smoothing projects in pavement engineering. Only data from projects by Taipei City Government and Taoyuan County Government have been obtained so far. Therefore, some of the 14 independent variable factors do not have discriminability. There are only 8 factors that can be analysed: duration, contract value, road area, bid-to-budget ratio, site location, planning approach, design approach, and supervising approach.

**Model 1: Back-propagation Artificial Neural Network**

When using the artificial neural network, in order to achieve good learning precision, learning speed, and recall speed, the following parameters must be considered. A trial and error approach should be employed to find the most optimal network structure. The parameter settings are tabulated in Table 1.

(1) Type of artificial neural network: back-propagation artificial neurons

(2) Number of hidden layers: one or two hidden layers often produce the best convergence. Without a hidden layer, the nonlinear relationship between output and input cannot be established. Too many layers overcomplicate the network and result in slower convergence. In general, one layer is sufficient for most problems. In this study, one layer is adopted in training.

(3) Processing elements in the hidden layers (Number of neurons): This parameter can also be called the number of nodes. In general, a large number of processing elements produces more accurate results, but leads to slower convergence. After a threshold however, more elements do not reduce the error further. In this study, 20 processing elements are used in testing.

(4) Learning rate: If the learning rate is too large, the algorithm becomes unstable. If it is too small, convergence is time consuming.

(5) Momentum coefficient (Alpha): Used to update the weights in training. The momentum tends to keep the weights moving in the same direction. The momentum coefficient must be in the interval of 0 to 1. A larger momentum coefficient helps to avoid local minima in the network, and therefore, a value of 0.9 is adopted in this study.

<table>
<thead>
<tr>
<th>Parameter settings</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network Model</td>
<td>Back Propagation</td>
</tr>
<tr>
<td>Input Data</td>
<td>32 Sets</td>
</tr>
<tr>
<td>Number of Training and Testing Groups</td>
<td>10</td>
</tr>
<tr>
<td>Hidden Layers</td>
<td>1</td>
</tr>
<tr>
<td>Processing Elements</td>
<td>20</td>
</tr>
<tr>
<td>Learning Rate</td>
<td>0.4</td>
</tr>
<tr>
<td>Momentum Coefficient</td>
<td>0.9</td>
</tr>
</tbody>
</table>

**Model 2: Rough Set and Back Propagation Artificial Neural Network**

Model 2 first resorts to the rough set theory to identify important attribute variables from influence factors, and then uses the back propagation artificial neural network as explained in Model 1. This is expected to significantly improve the accuracy of the model, thus, developing an assessment prediction model with better performance. The rough set theory mainly applies to nominal attribute data. Therefore, when using the rough set technique, if the attribute data are continuous real numbers, they must be discretized. Among the attributes in this study, contract value, cumulative number of labours, duration, bid-to-budget ratio, and road area are continuous data. A two-step clustering method is used to perform the discretization. The developed database is as shown in Fig. 2.

The eight influence factors discussed above are set as conditional attributes. The cumulative number of labours is set as a decision attribute to distinguish the results. Following the construction of the database for human resource demand estimation model, attributes are reduced in accordance with the rough set theory. Analysis results are tabulated in Table 2. Two reduced sets are obtained. In Table 2, “Size” indicates the number of attributes in each set. The fourth column “Reducts” represents the reduced set after redundant attributes are deleted. "1" in the second column (Positive Region, Pos. Reg.) means that prediction of decision attribute made from a reduced data set is consistent with that made from the original data set. In this case, “1” in the positive region in Table 2 shows that the prediction of the cumulative number of labors made from the reduced and original data sets are consistent. The stability coefficient (SC) represents the stability of dynamic reduction. 1 represents the most stable condition. Table 2 shows that there is no concern with stability and data consistency after the attribute reduction.
Table 2. Reduct Results.

<table>
<thead>
<tr>
<th>Size</th>
<th>Pos. Reg.</th>
<th>SC</th>
<th>Reducts</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>attr9, attr3, attr4,</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>attr9, attr5,</td>
</tr>
</tbody>
</table>

Table 3. Summary of Important Attributes.

<table>
<thead>
<tr>
<th>No.</th>
<th>Attribute</th>
<th>Occurrence</th>
<th>Percent</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>attr 1</td>
<td>Contract Value</td>
<td>0</td>
<td>0.00%</td>
<td>×</td>
</tr>
<tr>
<td>attr 2</td>
<td>Cumulative Number of Labor Workers</td>
<td>-</td>
<td>-</td>
<td>Dependent</td>
</tr>
<tr>
<td>attr 3</td>
<td>Construction Duration</td>
<td>1</td>
<td>50.00%</td>
<td>⊙</td>
</tr>
<tr>
<td>attr 4</td>
<td>Planning Approach</td>
<td>1</td>
<td>50.00%</td>
<td>⊙</td>
</tr>
<tr>
<td>attr 5</td>
<td>Design Approach</td>
<td>1</td>
<td>50.00%</td>
<td>⊙</td>
</tr>
<tr>
<td>attr 6</td>
<td>Supervising Approach</td>
<td>0</td>
<td>0.00%</td>
<td>×</td>
</tr>
<tr>
<td>attr 7</td>
<td>Passing Rate of the Bid</td>
<td>0</td>
<td>0.00%</td>
<td>×</td>
</tr>
<tr>
<td>attr 8</td>
<td>Site Location</td>
<td>0</td>
<td>0.00%</td>
<td>×</td>
</tr>
<tr>
<td>attr 9</td>
<td>Road Area</td>
<td>2</td>
<td>100.00%</td>
<td>⊙</td>
</tr>
</tbody>
</table>

Note: ⊙: Important Attributes, ×: Unimportant Factors

The important attributes are further analysed and summarized in Table 3. The third column “Occurrence” indicates the number of appearances of the attribute variable in the two reduced sets. The fourth column “Percent” is the ratio of the number of appearances of the attribute variable in each reduced set. It is worth noting that in the original rough set theory, attributes that appear in all reduced sets become important attributes while the rest are non-important attributes. The attributes with zero occurrences are unimportant factors. This can however be adjusted. Huang et al. (2010) decided to adopt those attributes with an occurrence rate higher than 50% as important attributes and the rest are non-influential factors. As a result, among factors affecting the cumulative number of labours in public engineering projects, duration, planning approach, design approach, and road area are important factors, and the others are non-influential factors.

Research Results

Avoiding samples with discrepancy and large variation in the data sets can effectively lower the error associated with random sampling. The K-fold cross validation method can ensure that the samples used in constructing the model and those used in testing are independent. Some scholars prefer using this method to assess the accuracy of results. In this study, this method is adopted to analyse the accuracy of results yielded from the prediction model.

K-fold Cross Validation Method

When constructing the prediction model, the collected data must be divided into a training set and a test set, and data samples in the two sets must be independent. The training set is used to train the prediction model, and then the test set is used to test the performance of the model and assess whether the model is general. The sample size affects the division between the training set and test set, thus influencing the performance of the developed model. The K-fold cross validation method is used in this study to divide the data samples. This method can effectively reduce discrepancy and variation. When the sample space is small, poor selection of training
data samples will not lead to less accurate prediction. The defect with this method is that one must repeatedly construct K number of classification models; thus, the modelling process is consequently time consuming. The procedure is as follows: the data set is first randomly divided into K independent sets (D1, D2, ..., Dk). Each set has the same size. The K-1 sub-sets are then used in turn as the training set, and the last sub-set is used as the test set. Repeat the training and testing for K times, until each independent set has been trained and tested. Finally, the average performance is calculated. A model developed in this way tends to be more objective. For example, when K=10, the above procedure is shown in Fig. 3.

In published literature, there is no clear standard for the number of folds in the training and testing sets. This is due to the difference in data attributes affecting the set division and model performance. Ron Kohavi (1995) discussed the determination of the optimal value of K in the K-fold cross validation method, discrepancy and variation is at the minimum when there are 10 folds (K=10). Therefore, in this study, the data set is randomly divided into 10 groups.

**Result Analysis**

In this study, 32 sets of data were extracted from the collected information in the pre-processing phase, during which missing data was attended to and normalization was performed. The 32 sets of data were then analysed using model 1 and 2, respectively. K-fold cross validation (K = 10) was applied as a final step to verify accuracy. Results show that the highest accuracy in model 1 is 98.97% and the lowest is 48.08%. The average accuracy is 84.58%. The highest accuracy in Model 2 is 99.80% and the lowest is 67.47%. Compared to Model 2, Model 1 seems not to perform well for some particular sets. This can be found in Table 4. In addition the average accuracy of Model 2 is 88.63%, representing that Model 2 not only has a higher average accuracy but also a greater highest and lowest accuracy. Therefore, selecting important attribute factors using the rough set theory before using the artificial neural network model produces better results than using the artificial neural network alone. It also proves that the process of selecting core attribute factors using the rough set theory can effectively improve the accuracy. Furthermore, it validates the application of the proposed model in assessing human resource demand in construction project engineering. The results are tabulated in Table 4.

**Research Findings**

**Comparison of Data Exploration Techniques**

This study uses the artificial neural network model and the rough set theory data exploration techniques to investigate the development of a human resource allocation model in pavement engineering. The results show better accuracy compared to those using the artificial neural network model only. It also proves that the process of selecting core attribute factors using the rough set theory can effectively improve the accuracy.

**Important Attributes in the Rough Sets**

The important attributes derived from rough sets prompt the following considerations: (1) Construction duration is a direct indication of project scale, and therefore it is not surprising that it is an important attribute for evaluating the decision attribute – cumulative number of labors. (2) There is a possible strong positive correlation between the road area and contract value. Road area is more precise in characterizing the decision attribute - cumulative number of labors, compared to the contract value. As a result, although contract value is a factor in evaluating the project construction scale, it is not one of the important attributes derived from the rough set analysis. (3) Both the planning and design approaches are attribute factors derived from the rough set analysis. The supervising approach, on the other hand, is not. The supervising approach factor should be highly positively correlated with the design approach factor. However, the design approach factor completely overrides the effect of the supervising approach factor.

**Conclusion and Suggestions**

Currently, domestic construction companies allocate their on-site...
resources based primarily on past experiences. Only some quantifiable factors are considered while unquantifiable conditional factors are neglected or not specifically analysed due to a lack of data. As a result, existing assessment approaches lack completeness and reliability. In this study, gathered quantitative data from actual cases are used to conduct a quantitative numerical discussion on various factors. Using the artificial neural network and the rough set data exploration technique a model for assessing human resource allocation in project engineering is developed. The results are compared with those obtained from the artificial neural network only model.

It is revealed that better accuracy is achieved by the proposed model. When compared to rough estimation based on past experiences, the proposed model provides a reliable tool for the authority and construction firms to manage the on-site manpower resources. Moreover, it offers the host organizations a good assessment model for manpower demand and an alert system that can prevent the occurrence of labour shortage and reduce related contractual disputes. The model developed in the present study is only applicable to road-smoothing projects in pavement engineering. In the future, data from other types of engineering projects can be collected. Using the rough set with artificial neural network combined data exploration model, the human resource allocation model in these fields can be constructed.

Other data exploration techniques could also be employed in the future to investigate other possibilities of human resource allocation models and discuss their differences. Human resource demand curve and estimation model for each construction phase could be developed by introducing analysis related to construction progress.

References