HYBRID FORECASTING SYSTEM BASED ON CASE-BASED REASONING AND ANALYTIC HIERARCHY PROCESS FOR COST ESTIMATION

Sangyong Kim

School of Construction Management and Engineering, University of Reading,
Reading, Berkshire RG6 6AW, United Kingdom
E-mail: sykim@kyonggi.ac.kr

Abstract. Cost estimating of highway projects with high accuracy at the early stage of project development is crucial for planning and feasibility studies. Various research have been attempted to develop cost prediction models in the early stage of a construction life cycle. This study uses the hybrid estimating tool to provide an effective cost data management for highway projects and accordingly develops a realistic cost estimating system. This study focused on the development of a more accurate estimate technique for highway projects in South Korea at the early stage using hybrid analytic hierarchy process (AHP) and case-based reasoning (CBR). Real case studies are used to demonstrate and validate the benefits of the proposed approach. It is expected that the developed CBR system is to provide decision-makers with accurate cost information to asses and compare multiple alternatives for obtaining the optimal solution and controlling cost.

Keywords: case-based reasoning, analytic hierarchy process, highway, cost estimation.

1. Introduction

Successful management within the limited budget is an important concern in any construction project. Lack of information and reliable methods that support estimating process made it difficult to initiate estimating report during the project planning stage (Chou, O'Connor 2007). In order to control the cost within an acceptable level, it requires appropriate and accurate measurement of various project related determinants and the understanding of the magnitude of their effects. As such, the importance of early estimating cannot be over emphasized. Number of cost estimating models, however, has been limited in road and bridge construction.

Several studies have demonstrated focus on highway construction cost estimating in the past. Owing to the lack of detailed design information and drawings during the early stages, several technical methods have been developed to estimate construction costs based on limited information (Chou 2009). Although MRA (Multiple Regression Analysis) has been used to cost estimating based on statistics many times, it is not appropriate when describing non-linear relationships, which are multidimensional, consisting of a multiple input and output problem (Tam, Fang 1999). Chou *et al*. (2005) suggested heuristic simulation models to improve the accuracy and efficiency of highway budgeting estimates based on useful data from the TxDOT (Texas Department of Transportation). Parametric cost estimating models were developed using ANNs (Artificial Neural Networks) for reasons of its limitation, and the models were demonstrated that they were very useful at the early stages of a project life cycle

(Hegazy, Ayed 1998; Al-Tabtabai *et al*. 1999; Wilmot, Mei 2005). However, ANNs can lose their effectiveness when the patterns are very complicated or noisy, knowledge representation and problem structuring are ill-defined, and training is trapped in local minima (Hegazy *et al*. 1994).

CBR (Case-based reasoning) is a relatively recent problem solving technique that is attracting increasing attention because it seems to resemble more closely the psychological process humans follow when trying to apply their knowledge to the solution of problems. CBR is problem solving technique that reuses past cases and experiences to find solution to the problems. While other major AI (Artificial Intelligence) techniques rely on making associations along generalized relationships between problem descriptors and conclusions, CBR is able to benefit from utilizing specific knowledge of previously experienced, evaluate the proposed solution and update the system by learning from this experience (Kolodner 1993; Shin, Han 1999; Kim, K. J., Kim, K. 2010).

Especially CBR systems have been proposed as effective alternatives to the support of decision-making. Several studies have demonstrated potential applications of CBR in construction areas. K. J. Kim and K. Kim (2010) proposed a preliminary cost estimation model using CBR and GAs (Genetic Algorithms) for determining the important weights of attributes. Kang *et al*. (2010) developed quantity-based construction cost estimating system using CBR and GAs. Ji *et al*. (2010) suggested CBR model for improving cost prediction accuracy in multifamily housing projects. Yau and Yang (1998) confirmed that CBR is a quite effective for selec-

ting a retaining wall system at the project planning stage. Dzeng and Tommelein (2004) developed a generic CBR system to facilitate schedule reuse and the new CBR system to develop a decision-support system to aid a project manager in seeking subcontractor registration. Luu *et al.* (2005) approached to procurement criteria selection and modeling bridge deterioration (2002), and suggested the ways to reduce the problem of hazard identification. CBR has also been used for cost-estimating of construction projects (An *et al.* 2007; Doğan *et al.* 2008), bid mark-up estimation (Dikmen *et al.* 2007), and cost budgeting for pavement maintenance (Chou 2009). Learning from previous researches and applications CBR can make very reasonable estimating without using specific experts and rules. For example, the cost of a construction project is influenced by a number of factors including the duration, the location, the year of construction, and the size of a project. The problem to be investigated is whether using the values for these factors, collected from completed previous projects, realisation cost for future projects can be reliably estimated. Therefore, the new innovative CBR approach was used to express the concept of the system developed in this study.

Problem of estimating future highway construction cost with regard to both 4 main divisions and total construction cost is discussed in the paper. The study is organized as follows. The next section describes the objectives and methodology of this study. The following section shows how this data was analyzed to verify its consistency and completeness and to obtain the knowledge required for the highway application. Then, 48 actual cases of highway project data constructed in South Korea, from 1996 to 2008, have been used as the source of cost data and in developing a CBR application for systematic highway project cost estimation. The next section briefly presents the CBR system that was developed specifically to generate CBR applications for modelling cost estimates and the steps followed in developing an application. Finally, the testing procedures and the validation results are discussed.

2. Objectives and methodology

The major aim of this study is to develop the hybrid CBR decision support system for estimating of highway project costs. The study goals included: (1) estimation of highway project costs at the early stage by 4 main divisions and total cost as well; (2) extracting significant CFs (Cost Factors) based on previous studies and interview with experts; (3) developing weighs values for CF using AHP (Analytic Hierarchy Process). As a result, the developed system provides a useful benchmark which is capable of assisting in identifying the CFs which demonstrated a strong relationship with highway project costs.

CFs are very complicated which requires intelligent processing to get a precise view of the effects of the cost attributes on project cost (Boussabaine, Elhag 1999). The data is required which corresponds to all the CFs which are known from previous studies. First of all, this study summarized literature review and identified significant CFs which affect a highway project costs. Furthermore,

industrial interviews were conducted to assist with selecting these factors. When potential CFs were identified, the weights of data were calculated by AHP. In addition, appropriate CBR system was developed and examined, and preliminary testing of developed system was carried out, using a relatively small number of data sets. The system is developed by means of an MS Excel-Based Visual Basic Application.

3. Case-based reasoning

CBR is a not a kind of computerized tool that imitate the analogical reasoning of human brains in problem solving (Rivard *et al.* 1998). The principle of CBR is based on the assumption that similar problems have similar solutions. According to Riesbeck and Schank (1989), CBR solves problems by capturing previous experiences and matching the important features of new problem to those of the old cases that have been successfully solved. The main source of knowledge in CBR is the case that can be reused even if it is partially matching the problem in hand (Yang, Yau 1996). Especially, CBR can deal efficiently with both numerical and nominal data, and can handle effectively cases that have incomplete data or variable data structures (Arditi, Tokdemir 1999). Furthermore, CBR has powerful learning capabilities that do not require time-consuming training and testing operations (Yang, Yau 1996). Table 1 lists CBR applications in various domains.

Aamodt and Plaza (1994) call the top level task of CBR problem solving and learning from experience which directly matches two phases, maintenance and application, as shown in Fig. 1. In the six-*Re* processes, changes initiated from outside of the CBR can be modelled easily:

- *Retrieve* the most similar cases from stored previous cases;
- *Reuse* the retrieved cases to attempt to solve the problem;
- *Revise* the proposed solution if necessary;
- *Retain* the new solution as a part of a new case;
- *Review* the results from applying the solution;
- *Restore* the case into case base library.

4. Selection of cost factors

The factors affecting the project cost were selected as the attributes that would be used as the input data for prediction CBR system. The data came from application of the selecting procedure presented in Fig. 2. At first, the literature review was conducted to identify which CFs were used in order to accomplish the cost estimating of highway projects. While there are only a few studies available on CFs which are suspected to influence construction cost, numerous studies have taken place in the highway construction projects. They are listed in Table 2.

A questionnaire survey was designed to obtain the primary data for this study. A pilot survey was first carried out to test the relevance and comprehensiveness of the questionnaires before a full scale survey was conducted. The respondents were given a choice of being

Table 1. Summary of CBR applications in various domains

Domain	Authors (Year)	Contents	System
Design	Schmitt (1993)	Design and creativity	CBD
	Fenves *et al.* (1995)	Conceptual structural design	SEED
	Demirkan (1998)	Interior design applications	N/A
Fabrication	Roddis and Bocox (1997)	Bridge fabrication error solution expert system	CB-BFX
Method Selection	Yau and Yang (1998)	Retaining wall selection system	CASTLE using ESTEEM
Bidding/ Prequalification	Chua *et al.* (2001)	A CBR bidding system	CASEBID
Management System	Tah *et al.* (1999)	Large-scale data repository	CBRidge Planner
	Arditi and Tokdemir (1999)	Predict the outcome of construction litigation	ESTEEM
	Morcous *et al.* (2002)	Modeling infrastructure deterioration	CBRMID
Cost	Mukhopadhyay *et al.* (1992)	Software development effort estimates	Estor
	Marir and Watson (1995)	Estimates the cost of refurbishing houses	ELSIE
Planning and Scheduling	Lee *et al.* (1998)	Scheduling of apartment construction	FASTRAK-APT
	Dzeng and Tommelein (1997, 2004)	Scheduling of power plant boilers	CasePlan
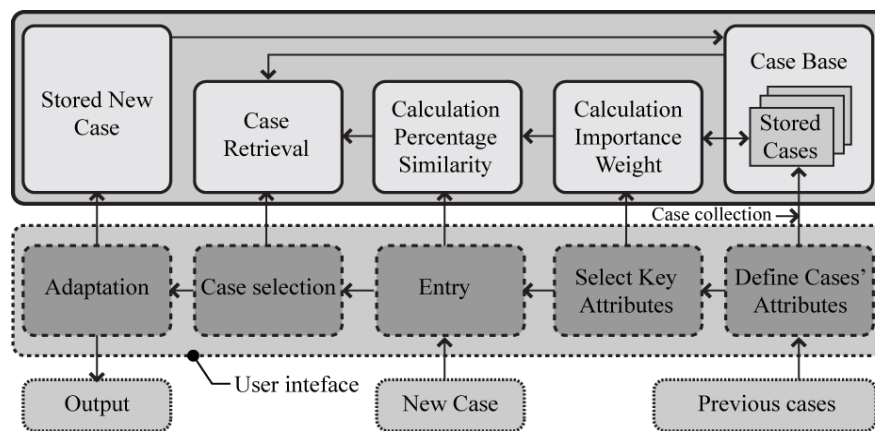
Fig. 1. Case-based reasoning cycle

Table 2. Highway research and relevant CFs

Authors	Year	Objectives	Cost Factors	
Hegazy and Ayed	1998	Budget cost	·Project type	·Year
			·Project scope	·Season
			·Soil condition	·Duration
			·Water bodies	·Size
			·Location	·Capacity
Al-Tabtabai *et al.*	1999	Mark-up estimation	·Preservation of utilities	·Location
			·Type of road	·Soil Nature
			·Type of consultant	·Hauling distance
			·Construction of detours	
Wilmot and Mei	2005	Total construction cost	·Price of labour	·Duration
			·Price of material	·Location
			·Price of equipment	·Bid volume
			·No. of plan changes	·Bid variable
			·Change in specification	·Contract type
Chou *et al.*	2007	Internet-based preliminary cost estimation	·Proposed main lane no.	·Project length
			·Shoulder width	·Location
			·Lane width	
Williams *et al.*	2009	Construction data collection	·Geo. design standard	·Project length
			·Length of loops/ramps	·Bridge type
			·Length of curb/gutter	·Bridge length
			·Median length/type	·Bridge width
			·Lane length	
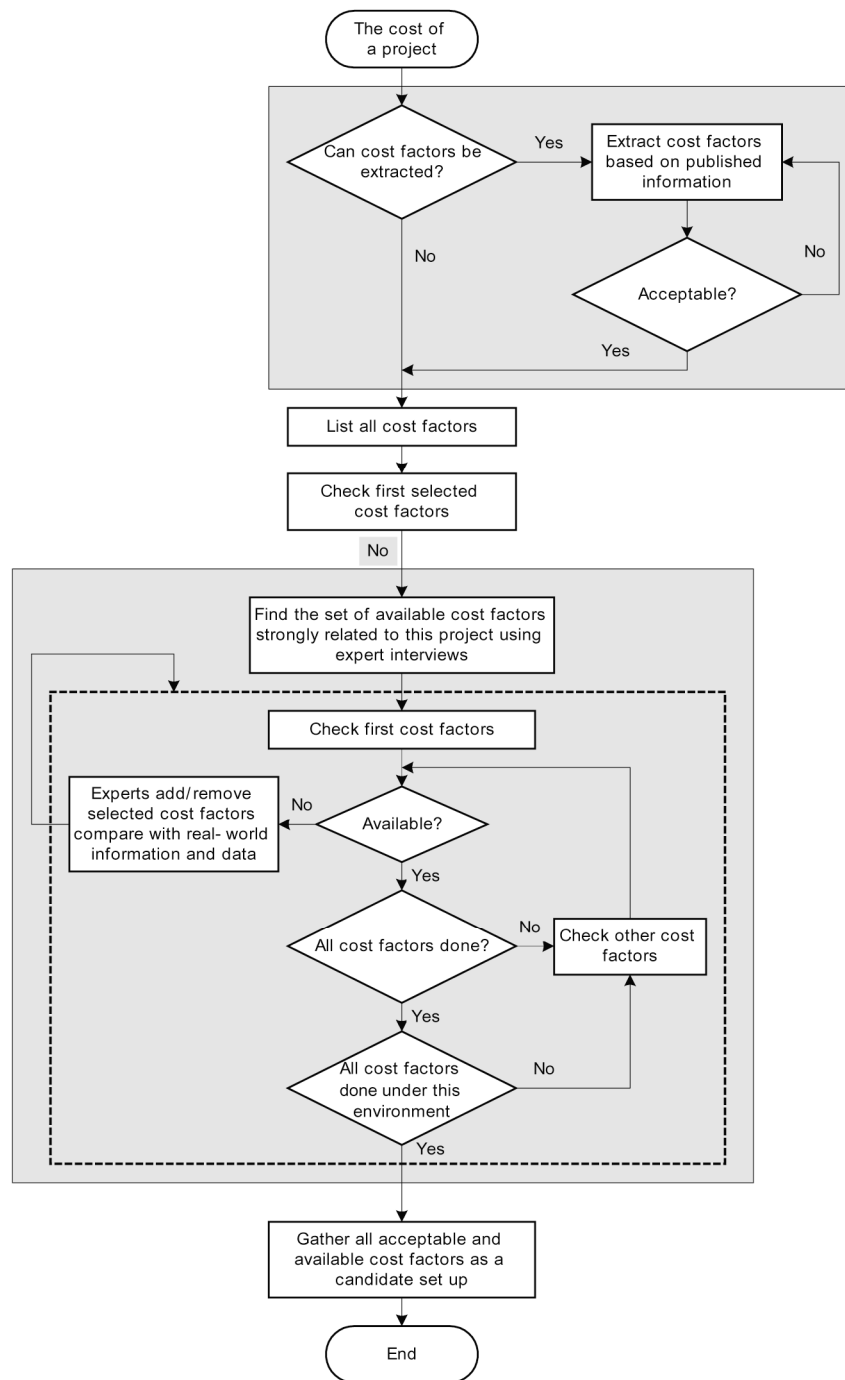
Fig. 2. Procedure of selection of relevant CFs

interviewed by telephone or to self-administer the questionnaires, and to send them back to the researchers. Three construction firms and 18 practitioners have been contacted to get feedback and comments about CFs of the data. The data included estimated material costs, actual costs, and general information based on 24 CFs, which was deemed potentially important to the accuracy of early estimates. The CFs were categorized as shown in Table 3.

Time Standardization

A cost index represents the relative scale of cost for a fixed quantity of goods or services between different periods, and provides a good means for forecasting future construction costs that change over time in response to changing demand, economic conditions, and prices (Ostwald 2001). The data collected for developing a CBR system have diverse characteristics and differences, such as when and where the projects were constructed. Such differences may cause incorrect prediction results (Kim, Kang 2003; Kim *et al.* 2005). A cost index ought to be a reliable tool for estimating future costs of construction activities, where construction activities are conducted months or years after costs were estimated (Huang 2007). The developed CBR cost estimation system includes therefore appropriate means which allow to reflect the change in overall highway construction costs over time.

Table 3. Determinants of project cost

No	Input variables	Unit	Values			Project significant factors
			Min.	Max.	Average	
1	Completed year	–	From 1996 to 2008			Time
2	Actual duration	months	24	85	58	
3	Contract's duration	months	24	68	50	
4	Time extension	months	0	36	8	
5	Design expenses	won	239,137,621	6,371,670,030	2,161,041,843	Cost
6	Contingency	won	310,459,368	8,566,800,000	2,815,219,944	
7	Type of site	–	1. Narrow 2. Medium 3. Large			General Information
8	Project scope	–	1. New 2. Rehabilitation			
9	Frame type of bridge	–	1. Concrete 2. Steel 3. Concrete+Steel			
10	Length of highway	km	3.34	49.00	8.39	
11	Ratio of bridge	%	0.30	9.80	1.68	
12	Wide of highway	m	23.4	37.8	26.7	
13	Wide of bridge	m	2.6	28.4	15.4	
14	No. of lanes	–	4	8	5	
15	Pavement type	–	1. Concrete 2. Ascon			
16	Asphalt	won/kg	19,000	35,200	24,647	Material
17	Cement	won/kg	1,450	3,300	2,070	
18	Bar steel	won/kg	210,000	363,000	263,053	
19	Sheet steel	won/kg	298,000	497,090	387,308	
20	Shape steel	won/kg	280,000	451,000	351,632	
21	earthwork	won	2,753,587,524	86,929,808,126	27,845,691,355	Division
22	pavement	won	178,560,000	62,218,194,183	19,920,876,974	
23	drainage/structure	won	991,358,249	29,316,852,276	9,381,476,650	
24	appurtenant/safety facilities	won	54,560,000	58,918,564,655	18,909,602,940	
25	total cost	won	8,390,793,722	237,383,419,240	76,057,647,919	

At first, the data used to establish the CBR system were collected from projects completed in 1996 and 2008. The data had to be converted to the identical time reference point defined by the Korea Institute of Construction Technology (KICT). The cost data of all the reference cases were converted to May 2005 cost level using the road cost index provided by the KICT. The detailed source of the road cost index of South Korea, which is announced monthly, was used for the construction cost adjustments. For example, May 2008 data were converted into May 2005 data by multiplying May 2008 cost data by the value $(100.0/122.8 = 0.81433)$ calculated by dividing 100.0, the index value for May 2005, by 122.8, i.e. the index value for May 2008. The road cost index applied to the conversion was official statistical data prepared to estimate the price fluctuation of input resources by 100.0 time scale as the price of a direct road construction cost input in a project at a certain point in time.

5. Determining the weight

The "weight" indicates how much attention should be paid to the factor during the matching process in CBR cycle (Kolodner 1993). It reflects the importance of that factor relative to other factors. It was found that considered values of weights influence the project cost prediction at most (Arditi, Tokdemir 1999; Chua *et al.* 2001; Luu *et al.* 2005; An *et al.* 2007; Kim, K. J., Kim, K. 2010). The determination of an appropriate CF weighting method is a major issue for effective case retrieval and indexing in CBR cycle (Park, Han 2002; An *et al.* 2007). The major issue in CBR is to retrieve not just

a similar past case but a usefully similar case to the problem. Previous approaches used GAs, gradient search, and feature counting. The problem with GA applications comes from difficulty in identification of appropriate fitness function which would successfully incorporate problem specific information (GA 2008). Gradient search may stagnate at local optima and fail to find the optimal global solution for certain starting solutions (Albright, Windston 2007; Kim, K. J., Kim, K. 2010). It is very difficult for feature counting to reliably state that one feature is more or less important than another based solely on human intuition (Arditi, Tokemir 1999).

For this reason, the integration of domain knowledge into the case retrieving and indexing process is highly recommended in developing a CBR system. This section utilizes a hybrid approach using AHP to case base retrieval process in an attempt to increase overall cost accuracy. If this hybrid approach is carried out well, the CBR system can deliver better estimation of costs (Shin, Han 1999). It can operate more accurately or at a lower cost level, it will be able to provide a better understanding of the effects of CFs interaction and variation.

Analytic Hierarchy Process

AHP is a multi-factor decision-making method that uses hierarchical structures to represent a decision problem and then delivers priorities for the decision-maker throughout judgments (Saaty 1986, 1987, 1990; Dyer 1990). Many previous studies (Dyer, Forman 1992; Al-Harbi 2001; Chwolka, Raith 2001; An *et al.* 2007; Podvezko 2009; Medineckiene *et al.* 2010) consider the

Table 4. Scale of relative importance (Saaty 1990)

Intensity of Relative Importance	Definition	Explanation
1	Equal importance	Two activities contribute equally to the objective
3	Moderate importance of one over another	Experience and judgment slightly favour one activity over another
5	Essential or strong	Experience and judgment strongly favour one activity over another
7	Very strong importance	An activity is strongly favoured and its dominance is demonstrated in practice
9	Extreme importance	The evidence favouring one activity over another is of the highest possible order of affirmation
2, 4, 6, 8	Intermediate values between the two adjacent judgments	When compromise is needed
Reciprocals of above nonzero number	If activity *i* has one of the above nonzero numbers assigned to it when compared with activity *j*, then *j* has the reciprocal value when compared to *i*	–

AHP methodology to be well suited for decision-making due to its role as a synthesizing mechanism in decisions. For example, An *et al*. (2007) compared three different weighting methods and concluded that the AHP was more accurate, reliable, and explanatory than decent gradient methods for determining the relative important weights for making preliminary estimates of new construction costs. Once the hierarchy is built, the decision-maker systematically evaluates its components, which represent considered factors, by comparing their importance in a pair-wise manner. This study applies the AHP to calculate the weights of the aspects and the attributes within each aspect. Pair-wise comparisons of importance of the factors at each level of an AHP are made in terms of importance when comparing factors with respect to their relative importance (Zahedi 1986; Harker, Vargas 1987; Podvezko 2009; Medineckiene *et al*. 2010) (see Table 4).

The last step is devoted to the measurement of the overall consistency of provided AHP judgments by means of the CR (Consistency Ratio) proposed by Saaty. The CR provides a way of measuring errors introduced during elicitation of expert opinions. The value of consistency index is applied with this regard Eq. (1) (Chen *et al*. 2010):

$$CI = \frac{(\lambda_{\max} - n)}{n - 1}, \quad (1)$$

where *n* is the number of compared factors, and $\lambda_{\max}$ is the maximum eigenvalue of a judgment matrix which corresponds to the group of compared factors. The CR value is given by division of the CI value by the Random consistency index value. The RI value depends on number of compared factors. RI values for different numbers of factors are presented in Table 5.

Appropriate CR value justifies extracting expert knowledge that can guide effective retrievals of useful weights. The weight values expressing importance of each CF are presented in Fig. 3. They will be assigned to the considered attributes for case based retrieval of the most similar process plans due to the effective similarity function in the proposed application area.

Table 5. Random consistency index (Saaty 1990)

n	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.9	1.12	1.24	1.32	1.41	1.45	1.49
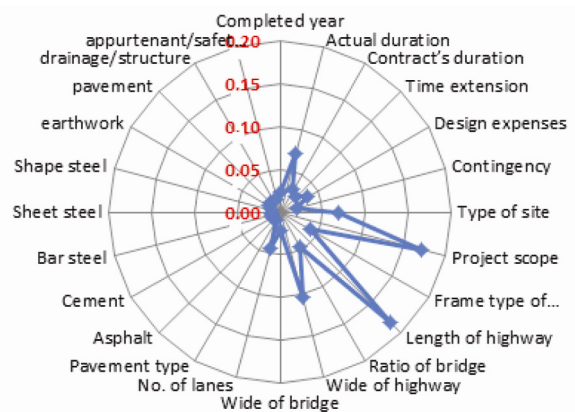
Fig. 3. Weights of CFs obtained by means of AHP application

6. Case study

Expert knowledge can be applied to assess importance weights. The expert is expected to have the required knowledge and experience to decide which model or system makes good predictions. CBR applications can be created using the hybrid AHP-CBR application development tool. The CBR system searches for matched cases contained in the case base and summarises them into a set of acceptable solutions. Decision-makers select then one of the recommended solutions. The system's interface is organized following the basic process used to construct the AHP-CBR application. The reasoning structure of proposed system is presented in Fig. 4. The following six steps are involved in CBR application:

Step 1. *Case base definition*: the first step is used to define the initial components of the system. The names and value types for CFs are defined. The selected CFs should provide the best description of relevant construction cost influencing attributes which result from prior experience. Table 2 presents an illustrative example of a case based library contents.
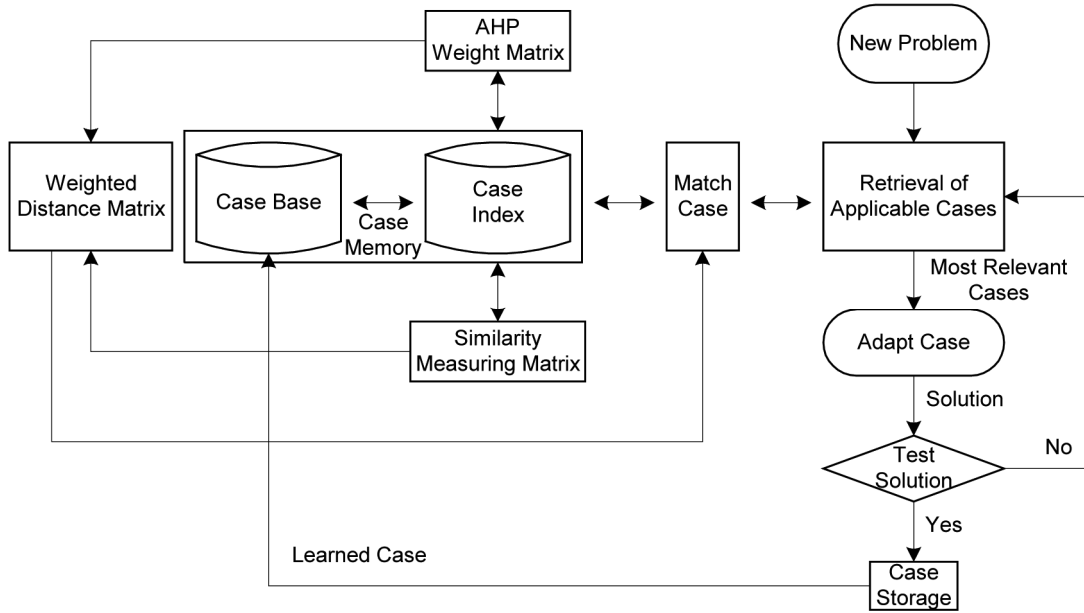
Fig. 4. Hybrid AHP-CBR system

Step 2. *Similarity definition*: the step deals with a way the similarity between a new problem description and the case based library items is assessed. The methodology and various metrics for determining similarity during case base retrieval are defined. SI (Similarity Index) is assessed both at the case level (comparing cases against each CF) as well as at the CF level (comparing the value of each CF value to the new entered CF values). Weighted case similarities between the new problem and cases included in the case base library are estimated according to the following formula:

$$SI = \frac{\sum\limits_{i=1}^{n}(W_i \times SS_i)}{\sum\limits_{i=1}^{n}(W_i)} \times 100. \qquad (2)$$

In Eq. (2), *SI* is a calculated numerical value which demonstrates the degree of similarity between a case in the case based library and the investigated problem case (Yau, Yang 1998). *SI* is normalized into a scale from 0 to 1 for easy comparison. *Weights* (*W*) of each CF can be either assigned by the decision-makers or AHP. *SS* (*Similarity score*) is calculated on the basis of values of the CFs: numerical and nominal. For the nominal factor, the *SS* equals 1 when the two values are identical and 0 otherwise.

For the numerical factor, *SS* is calculated by Eq. (3). In Eq. (3), $V_{\text{case based}}$ represents value of a factor for the cases stored in a case based library. V_{problem} value corresponds to the target case for predicting highway costs. A more detailed classification method is applied to improve the accuracy in this study when decision-maker selects one of retrieved cases. It is possible to select the best matching case from the case based library. Consequently, the new *SS* formula has been developed and proposed here which not only expresses the difference of compared

cases but also makes verification of the minimum and maximum relationship of the cases possible. The similarity score in the developed formula is referred to as $SS_{revised}$ to distinguish it from the *SS* used to retrieve similar cases:

$$SS_{revised} = \frac{1}{\left|V_{\text{case based}} - V_{\text{problem}}\right| + 1}. \qquad (3)$$

Finally, *SI* is calculated due to Eq. (2).

Step 3. *Case definition*: this step is used to fill in the case information for each case to be stored in the case based library. A case collection interface is then applied for introducing data for the real highway project cases into the library. CF values which describe the cases must conform to the defined types. 48 highway cases are included in the case based library in the prepared CBR system.

Step 4. *Rule definition*: rules are used to compute *SI* and to adapt a retrieved similar case to better meet the needs of the new problem. Rules are used to address the differences that exist between a new problem case (target case) and the retrieved similar case. The rules are applied to account for the differences and advise on what the plausible outcomes of a comparison might be. Rules can be used to change CF values based on comparison.

Step 5. *Application interface*: after case retrieval is complete the system returns a list of cases with SI values indicating their similarity to the target case. Their scores indicate their relevance to the problem at hand. The decision-makers can apply the selected case to help decide how to solve the current problem. The selected case can be then adapted to better assist in making a decision.

Step 6. *System validation*: to determine whether the predicted project cost provided by AHP-CBR is a good estimate of the problem case, three methods that have been reported by Yau and Yang (1998), Arditi and Tokemir (1999) and Koo *et al.* (2010) are used. Each of these methods makes use of the overall case *SI* for each retrieved case. These methods are as follows:

1. The problem case is compared to the characteristics of the retrieved case that has the highest overall *SS*;
2. The problem case is compared to the most frequent characteristics in the top ten retrieved cases, or fewer if ten are not available, that have an overall *SS* greater than or equal to 0.75 (75%);
3. *The problem case is compared to the average characteristics of the top five retrieved cases, or fewer if five are not available that have an overall SS greater than or equal to 0.75 (75%). The average of the predicted condition is weighted using the overall SS to magnify the importance of the retrieved cases which have higher SS.*

According to the CBR concept, the case with the highest *SI* in the case base library may be considered to have the most similar project characteristics compared to the test case in this study. Also, each of the 4 division costs (see Table 7) is identifiable from the selected case based on retrieved *SI*. These results may be used as references in the decision-making process as well.

7. Result for the sample CBR system application

As mentioned above, the research was carried out by employing AHP method to assign importance weights to each CF. Different error calculation formulae have been used by previous studies. The *Mean Absolute Estimation Error* (*MAEE*) calculated due to Eq. (4) is applied for expressing the system performance:

$$MAEE = \frac{\sum_{i=1}^{10} \frac{|\text{Cost}_{CBR} - \text{Cost}_{ACT}|}{\text{Cost}_{ACT}}}{n} \times 100\%, \quad (4)$$

where: Cost_{CBR} represents output for CBR application; Cost_{ACT} expresses actual cost, and n denotes the number of testing cases.

The n-fold cross-validation was adapted in the next phase to evaluate the performance of the AHP-CBR system and to reinforce the reliability of results. The 6-fold approach can be considered the effective form of reliability analysis of the measurement system. For example, MAEE of 9.17% with 62.5% of the estimates within 10% of the MAEE correspond to the results of the AHP-CBR system application, while 87.5% of the estimates within 15% are obtained for a 1-fold approach application. The results as listed in Table 6. And, the mean error (difference) rate of 1-fold compared to 6-fold is equal to 9.09%. The corresponding output accuracy of the established AHP-CBR system meets the fifth class requirements with regard to carrying out project screening and feasibility study due to the definition of American Association of Cost Engineers (AACE). Test cases allowed not only to predicted total construction cost estimation error rate but also to predict estimation error rate for each of the four division cost. Obtained results are shown in Table 7. Retrieved similarity index values obtained for the selected case based problems comprise therefore valid reference points for the decision-making process.

Table 6. Cost reasoning errors for 6-fold cross-validation of the AHP-CBR system

Error rate (%)	1-Fold		2-Fold		3-Fold		4-Fold		5-Fold		6-Fold	
	Fre.(%)	Cum.(%)	Fre.(%)	Cum.(%)	Fre.(%)	Cum.(%)	Fre.(%)	Cum.(%)	Fre.(%)	Cum.(%)	Fre.(%)	Cum.(%)
0–2.5	0	0	1(12.5)	1(12.5)	0	0	2(25.0)	2(25.0)	1(12.5)	1(12.5)	1(12.5)	1(12.5)
2.5–5.0	1(12.5)	1(12.5)	2(25.0)	3(37.5)	2(25.0)	2(25.0)	0	2(25.0)	2(25.0)	3(37.5)	2(25.0)	3(37.5)
5.0–7.5	2(25.0)	3(37.5)	0	3(37.5)	2(25.0)	4(50.0)	2(25.0)	4(50.0)	1(12.5)	4(50.0)	1(12.5)	4(50.0)
7.5–10.0	2(25.0)	5(62.5)	2(25.0)	5(62.5)	1(12.5)	5(62.5)	0	4(50.0)	0	4(50.0)	1(12.5)	5(62.5)
10.0–12.5	1(12.5)	6(75.0)	0	5(62.5)	1(12.5)	6(75.0)	2(25.0)	6(75.0)	2(25.0)	6(75.0)	1(12.5)	6(75.0)
12.5–15.0	1(12.5)	7(87.5)	1(12.5)	6(75.0)	0	6(75.0)	1(12.5)	7(87.5)	1(12.5)	7(87.5)	0	6(75.0)
15.0–17.5	1(12.5)	8(100.0)	1(12.5)	7(87.5)	2(25.0)	8(100.0)	0	7(87.5)	1(12.5)	8(100.0)	0	6(75.0)
17.5–20.0	0	8(100.0)	1(12.5)	8(100.0)	0	8(100.0)	1(12.5)	8(100.0)	0	8(100.0)	1(12.5)	7(87.5)
20.0–22.5	0	8(100.0)	0	8(100.0)	0	8(100.0)	0	8(100.0)	0	8(100.0)	1(12.5)	8(100.0)
Min.	2.76	–	1.80	–	2.76	–	0.79	–	2.21	–	1.27	–
Max.	15.36	–	19.25	–	16.81	–	18.98	–	17.43	–	20.07	–
MAEE	9.17	–	9.47	–	9.25	–	8.76	–	8.49	–	9.42	–

Table 7. The error rate of each cost division (1-fold)

Division	Error rate (%)							
	Case1	Case 2	Case 3	Case 4	Case 6	Case 7	Case 8	Case 9
Earthwork	9.52	7.07	10.53	11.72	6.53	9.57	15.90	2.69
Pavement	9.34	7.49	10.01	12.50	6.31	9.93	14.82	2.83
Drainage/Structure	9.57	7.14	10.66	11.85	6.49	9.49	15.72	2.65
Appurtenant/Safety facilities	9.29	7.42	9.88	12.37	6.35	10.01	15.00	2.87
MAEE	9.43	7.28	10.27	12.11	6.42	9.75	15.36	2.76

8. Conclusion

Cost estimating system has become an integral part of any advanced cost management modelling. Such systems make estimation of the accurate project cost and improvement in cost prediction rate possible. Presented research therefore focused on developing the hybrid AHP-CBR system which provides accurate predictions of the future cost of different highway projects.

The contribution of this research pertains to four areas. At first, obtaining the higher predictive accuracy of cost estimate and guide to decision-maker at the early planning stage is addressed. Developed AHP-CBR system reduces the time required to build a cost list for project activities and makes reduction of processing time and cost possible. At second, the extracted CFs for highway projects significantly improve system performance with regard to the cost estimation. This finding contributes to the current body of knowledge on approximate cost estimating, and may serve as a useful guide for future data collection efforts and cost estimation system development. At third, this research proposes the identification of an alternative similarity score measuring formula. The introduced similarity measure makes investigation of contrast between the developed similarity measure and the classical SS measures possible when CFs are used to describe a case. And finally, the weights of CFs are calculated using AHP.

In order to enhance the capabilities of the CBR approach in cost estimating, numerous problems should be explored in the future research. The problems include: development of proprietary indices for adjusting the cost due to difference in a project location, development of more justified weights using different weight estimation methods, collection of more project cases into the case based library for improving accuracy, and identification of important CFs in accordance with different phases of the project planning and realisation.

References

Aamodt, A.; Plaza, E. 1994. Case-based reasoning: foundational issues, methodological variations, and system approaches, *AI Communications* 7(1): 39–59.

Albright, S. C.; Winston, W. L. 2007. *Management science modelling*. Mason, Ohio: Thomson South-Western. 936 p.

Al-Harbi, K. M. A. 2001. Application of the AHP in project management, *International Journal of Project Management* 19(1): 19–27. http://dx.doi.org/10.1016/S0263-7863(99)00038-1

Al-Tabtabai, H.; Alex, A. P.; Tantash, M. 1999. Preliminary cost estimation of highway construction using neural networks, *Cost Engineering* 41(3): 19–24.

An, S.-H.; Kim, G.-H.; Kang, K.-I. 2007. A case-based reasoning cost estimating model using experience by analytic hierarchy process, *Building Environment* 42(7): 2573–2579. http://dx.doi.org/10.1016/j.buildenv.2006.06.007

Arditi, D.; Tokdemir, O. 1999. Comparison of case-based reasoning and artificial neural networks, *Journal of Computing in Civil Engineering* ASCE 13(3): 162–169. http://dx.doi.org/10.1061/(ASCE)0887-3801(1999)13: 3(162)

Boussabaine, A. H.; Elhag, T. M. S. 1999. Knowledge discovery in residential construction project cost data, in *Proc. of the 15th Annual ARCOM Conference*, 15–17 September, 1999, Liverpool, UK, 489–498.

Chen, W. T.; Chang, P.-Y.; Chou, K.; Mortis, L. E. 2010. Developing a CBR-based adjudication system for fatal construction industry occupational accidents, *Expert Systems with Applications* 37(7): 4867–4880. http://dx.doi.org/10.1016/j.eswa.2009.12.028

Chou, J.-S. 2009. Web-based CBR system applied to early cost budgeting for pavement maintenance project, *Expert Systems with Applications* 36(2): 2947–2960. http://dx.doi.org/10.1016/j.eswa.2008.01.025

Chou, J.-S.; O'Connor, J. T. 2007. Internet-based preliminary highway construction cost estimating database, *Automation in Construction* 17(1): 65–74. http://dx.doi.org/10.1016/j.autcon.2007.03.001

Chou, J.-S.; Wang, L.; Chong, W. K.; O'Connor, J. T. 2005. Preliminary cost estimates using probabilistic simulation for highway bridge replacement projects, in *Construction Research Congress 2005,* April, 2005, American Society of Civil Engineers, San Diego, California, USA, 467–471. http://dx.doi.org/10.1061/40754(183)92

Chua, D. K. H.; Li, D. Z.; Chan, W. T. 2001. Case-based reasoning approach in bid decision making, *Journal of Construction Engineering and Management* ASCE 127(1): 35–45. http://dx.doi.org/10.1061/(ASCE)0733-9364(2001) 127:1(35)

Chwolka, A.; Raith, M. G. 2001. Group preference aggregation with the AHP – implications for multiple-Issues agendas, *European Journal of Operational Research* 132(1): 176–186. http://dx.doi.org/10.1016/S0377-2217(00)00121-1

Dikmen, I.; Birgonul, M. T.; Gur, A. K. 2007. A case-based decision support tool for bid mark-up estimation of international construction projects, *Automation in Construction* 17(1): 30–44. http://dx.doi.org/10.1016/j.autcon.2007.02.009

Demirkan, H. 1998. Integration of reasoning systems in architectural modelling activities, *Automation in Construction* 7(2–3): 229–236. http://dx.doi.org/10.1016/S0926-5805 (97)00056-3

Doğan, S. Z.; Arditi, D.; Günaydin, H. M. 2008. Using decision trees for determining attribute weights in a case-based model of early cost prediction, *Journal of Construction Engineering and Management* ASCE 134(2): 146–152. http://dx.doi.org/10.1061/(ASCE)0733-9364(2008)134: 2(146)

Dyer, J. S. 1990. Remarks on the analytic hierarchy process, *Management Science* 36(3): 249–258. http://dx.doi.org/10.1287/mnsc.36.3.249

Dyer, R. F.; Forman, E. H. 1992. Group decision support with the Analytic Hierarchy Process, *Decision Support Systems* 8(2): 99–124. http://dx.doi.org/10.1016/0167-9236(92) 90003-8

Dzeng, R.-J.; Tommelein, I. D. 1997. Boiler erection scheduling using product models and case-based reasoning, *Journal of Construction Engineering and Management* ASCE 123(3): 338–347. http://dx.doi.org/10.1061/(ASCE)0733-9364(1997)123:3(338)

Dzeng, R.-J.; Tommelein, I. D. 2004. Product modelling to support case-based construction planning and scheduling, *Automation in Construction* 13(3): 341–360. http://dx.doi.org/10.1016/j.autcon.2003.10.002

Fenves, S. J.; Rivard, H.; Gomez, N.; Chiou, S. C. 1995. Conceptual structural design in SEED, *Journal of Architectural Engineering* ASCE 1(4): 179–186. http://dx.doi.org/10.1061/(ASCE)1076-0431(1995)1:4(179)

GA. 2008. *Introduction to genetic algorithm* [online], [cited 5 August 2011]. Available from Internet: http://lancet.mit.edu/~mbwall/presentations/IntroToGAs/

Harker, P. T.; Vargas, L. G. 1987. The theory of ratio scale estimation: Saaty's analytic hierarchy process, *Management Science* 33(11): 1383–1403. http://dx.doi.org/10.1287/mnsc.33.11.1383

Hegazy, T.; Ayed, A. 1998. Neural network model for parametric cost estimation of highway projects, *Journal of Construction Engineering and Management* ASCE 124(3): 210–218. http://dx.doi.org/10.1061/(ASCE)0733-9364 (1998)124:3(210)

Hegazy, T.; Fazio, P.; Moselhi, O. 1994. Developing practical neural network applications using back-propagation, *Computer-Aided Civil and Infrastructure Engineering* 9(2): 145–159. http://dx.doi.org/10.1111/j.1467-8667.1994.tb00369.x

Huang, G. 2007. *Cost modeling based on support vector regression for complex products during the early design phases.* PhD thesis. Department of Industrial and Systems Engineering, Virginia Polytechnic Institute and State University, Virginia, USA.

Ji, C. Y.; Hong, T. H.; Hyun, C. T. 2010. CBR revision model for improving cost prediction accuracy in multifamily housing projects, *Journal of Management in Engineering* ASCE 26(4): 229–236. http://dx.doi.org/10.1061/(ASCE)ME.1943-5479.0000018

Kang, B. J.; Cho, J. H.; Lee, J. S.; Lee, J. S.; Chun, J. Y. 2010. Development of quantity-based reinforced concrete structural construction cost estimating system using CBR+GA model, *Architectural Institute of Korea* 26(3): 139–146.

Kim, G. H.; Kang, K. I. 2003. A study on a model for neural networks training using genetic algorithms for predicting cost estimates of apartment projects at the early project stage, *Journal of Architectural Institute of Korea* 19(10): 133–142.

Kim, G. H.; Seo, D. S.; Kang, K. I. 2005. Hybrid models of neural networks and genetic algorithms for predicting preliminary cost estimates, *Journal of Computing in Civil Engineering* ASCE 19(2): 208–211. http://dx.doi.org/10.1061/(ASCE)0887-3801(2005)19:2(208)

Kim, K. J.; Kim, K. 2010. Preliminary cost estimation model using case-based reasoning and genetic algorithms, *Journal of Computing in Civil Engineering* ASCE 24(6): 499–505. http://dx.doi.org/10.1061/(ASCE)CP.1943-5487.0000054

Kolodner, J. 1993. *Case-based reasoning.* 1st ed. San Mateo, CA: Morgan Kaufmann. 668 p.

Koo, C. W.; Hong, T. H.; Hyun, C. T.; Koo, K. J. 2010. A CBR-based hybrid model for predicting a construction duration and cost based on project characteristics in multifamily housing projects, *Canadian Journal of Civil Engineering* 37(5): 739–752. http://dx.doi.org/10.1139/L10-007

Lee, K. J.; Kim, H. W.; Lee, J. K.; Kim, T. H. 1998. Case and constraint-based project planning for apartment construction, *AI Magazine* 19(1): 13–34.

Luu, D. T.; Ng, S. T.; Chen, S. E. 2005. Formulating procurement selection criteria through case-based reasoning approach, *Journal of Computing in Civil Engineering* ASCE 19(3): 269–276. http://dx.doi.org/10.1061/(ASCE)0887-3801(2005)19:3(269)

Marir, F.; Watson, I. 1995. CBRefurb: case-based cost estimation, case based reasoning: prospects for application, *IEE Colloquium* 7: 1–5.

Medineckiene, M.; Turskis, Z.; Zavadskas, E. K. 2010. Sustainable construction taking into account the building impact on the environment, *Journal of Environmental Engineering and Landscape Management* 18(2): 118–127. http://dx.doi.org/10.3846/jeelm.2010.14

Morcous, G.; Rivard, H.; Hanna, A. M. 2002. Modeling bridge deterioration using case-based reasoning, *Journal of Infrastructure Systems* 8(3): 86–95. http://dx.doi.org/10.1061/(ASCE)1076-0342(2002)8:3(86)

Mukhopadhyay, T.; Vicinanza, S. S.; Prietula, M. J. 1992. Examining the feasibility of a case-based reasoning model for software effort estimation, *MIS Quarterly* 16(2): 155–171. http://dx.doi.org/10.2307/249573

Ostwald, P. F. 2001. *Construction cost analysis and estimating.* 1st ed. Prentice-Hall, Inc. 462 p.

Park, C.-S.; Han, I. 2002. A case-based reasoning with the feature weights derived by analytic hierarchy process for bankruptcy prediction, *Expert Systems with Applications* 23(3): 255–264. http://dx.doi.org/10.1016/S0957-4174 (02)00045-3

Podvezko, V. 2009. Application of AHP technique, *Journal of Business Economics and Management* 10(2): 181–189. http://dx.doi.org/10.3846/1611-1699.2009.10.181-189

Riesbeck, C. K.; Schank, R. S. 1989. *Inside case-based reasoning.* 1st ed. Lawrence Erlbaum Associates. 448 p.

Rivard, H.; Fenves, S. J.; Gomez, N. 1998. Case-based reasoning for conceptual building design, in *Proc. of the 1st International Conference on New Information Technologies for Decision Making in Civil Engineering*, October, 1998, University of Quebec, Montreal, Quebec, 355–366.

Roddis, W. M. K.; Bocox, J. 1997. Case-based approach for steel bridge fabrication errors, *Journal of Computing in Civil Engineering* 11(2): 84–91. http://dx.doi.org/10.1061/(ASCE)0887-3801(1997)11:2(84)

Saaty, T. L. 1986. Axiomatic foundation of the analytic hierarchy process, *Management Science* 32(7): 841–855. http://dx.doi.org/10.1287/mnsc.32.7.841

Saaty, T. L. 1987. Concept, theory, and techniques: rank generation, preservation, and reversal in the analytic hierarchy decision process, *Decision Sciences* 18(2): 157–177. http://dx.doi.org/10.1111/j.1540-5915.1987.tb01514.x

Saaty, T. L. 1990. How to make a decision: the analytic hierarchy process, *European Journal of Operational Research* 48(1): 9–26. http://dx.doi.org/10.1016/0377-2217(90)90057-I

Schmitt, G. 1993. Case-based design and creativity, *Automation in Construction* 2(1): 11–19. http://dx.doi.org/10.1016/0926-5805(93)90031-R

Shin, K.-S.; Han, I. 1999. Case-based reasoning supported by genetic algorithms for corporate bond rating, *Expert Systems with Applications* 16(2): 85–95. http://dx.doi.org/10.1016/S0957-4174(98)00063-3

Tah, J. H. M.; Carr, V.; Howes, R. 1999. Information modelling for case-based construction planning of highway bridge projects, *Advances in Engineering Software* 30(7): 495–509. http://dx.doi.org/10.1016/S0965-9978(98)00128-8

Tam, C. M.; Fang, C. F. 1999. Comparative cost analysis of using high performance concrete in tall building construction by artificial neural networks, *ACI Structural Journal* 96(6): 927–936.

Williams, R. C.; Hildreth, J. C.; Vorster, M. C. 2009. Highway construction data collection and treatment in preparation for statistical regression analysis, *Journal of Construction Engineering and Management* 135(12): 1299–1306. http://dx.doi.org/10.1061/(ASCE)CO.1943-7862.0000112

Wilmot, C. G.; Mei, B. 2005. Neural network modelling of highway construction costs, *Journal of Construction Engineering and Management* ASCE 131(7): 765–771. http://dx.doi.org/10.1061/(ASCE)0733-9364(2005)131: 7(765)

Yang, J.; Yau, U. 1996. Applications of case-based reasoning in construction engineering and management, in *Proc. of the 3rd Congress on Computing in Civil Engineering*, 17–19 June, 1996, American Society of Civil Engineers, Anaheim, California, 663–669.

Yau, N.-J.; Yang, J.-B. 1998. Applying case-based reasoning technique to retaining wall selection, *Automation in Construction* 7(4): 271–283. http://dx.doi.org/10.1016/S0926-5805(97)00072-1

Zahedi, F. 1986. The analytic hierarchy process – A survey of the method and its applications, *Interfaces* 16(4): 96–108. http://dx.doi.org/10.1287/inte.16.4.96

Sangyong KIM. PhD Candidate at the School of Construction Management and Engineering, University of Reading. MSc in construction management and material from Korea University. A member of CIOB, RICS, ASCE, JKIBC, and AIK. His research interests include cost estimation, information technology in construction, decision-making and analysis by artificial intelligence, and its applications in construction areas.