TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 19/67 pp114-120
Volume 13, Number S1, October 2008

Applying AHP-Based CBR to Estimate Pavement Maintenance Cost

Jui-Sheng Chou**

Department of Construction Engineering, Taiwan Tech (NTUST), 43 Sec.4, Keelung Rd., Taipei 106

Abstract: The cost-efficacy control of maintenance operations in developing countries has become critical to the infrastructure asset management after highway construction. To effectively manage numerous projects annually with limited resources, it is necessary to reasonably estimate costs during the process of making maintenance project selection decisions. This study outlines the modeling of case-based reasoning (CBR) estimation that compares and retrieves the most similar instance across the case library. Four CBR approaches were presented and assessed in terms of their mean absolute prediction error rates. The resulting model demonstrates the ability of estimating the pavement maintenance project costs with the satisfactory accuracy at the early stages.

Key words: case-based reasoning; analytical hierarchy process; cost estimation; project management

Introduction

During recent decades, numerous developing countries have undertaken large scale highway construction to boost urban economic development. As transportation network connectivity has been achieved, the focus of transportation agencies has gradually shifted from new construction to post infrastructure maintenance. Taiwan has a total roadway lane length over twenty thousand kilometers, of which approximately 350 million square meters annually needs to be maintained and repaired for road surface and subgrade base[1].

Roadway and bridge maintenance operations are important in maintaining service quality and user safety. For example, failure to ensure timely road surface repair or the poor pavement quality has led to incidents of heavy vehicles overturning and driver distraction, creating considerable public concerns.

Cost estimation can frequently be a time-consuming and cumbersome activity when there is no starting reference point of estimation for engineers to put a hand on. After creating a rapid and reasonable estimate, the agents gain more time to rank the reported maintenance projects according to consideration of the assessment criteria.

However, based on the survey findings and the in-depth interviews with experienced on-site engineers, most closeout information on past projects was reserved in paper-based formats and archived in the physical warehouses of the responsible district offices for possible auditing purposes only.

All the unprocessed data may contain valuable explicit or implicit knowledge generated by the experienced technicians and engineers. This growing body of knowledge should thus be made available for retrieval, reuse, and retaining. This study attempts to determine preliminary project cost with readily available information based on previous experience of pavement maintenance operations (PMO) to assist decision makers in project screening and budget initiation.

1 Literature Review

Retrieval of past knowledge and computer-assisted similarity comparison rather than individual experience-based comparison can accelerate the cost estimation process. Case-based reasoning (CBR) has

Received: 2008-05-30
** To whom correspondence should be addressed.
E-mail: jschou@mail.ntust.edu.tw

been identified as an appropriate solution for preparing early estimates for the following reasons: (1) Studies comparing CBR to other prediction techniques, such as neural networks[2] and parametric method[2,3], demonstrated that the CBR estimating model outperformed other models in the length of time it could be used for, quality which it maintained over extended periods, and its ability to provide a solution despite a lack of precision in certain information, (2) CBR uses past problem-solving experience like cognitive learning and applied adaptations of intuitive solutions to new cases[4], (3) CBR can be applied to qualitative and quantitative data, reflecting the types of datasets existing in the real world, and (4) CBR provides a simple means of measuring construction cost given that most studies have identified the existence of non-linear relationships between cost and influential factors[5-8]. The following reviews current estimating techniques associated with construction projects and applications of CBR in the construction industry.

1.1 Cost estimation of engineering projects

Project cost is derived from the sum of direct construction cost, indirect cost, and engineering contingency. In a common PMO practice in Taiwan, direct costs comprise contributions from four main work items (average 19.3%) and 24 minor work items (average 0.95%) based on observation of the projects gathered in this investigation. The indirect cost generally includes construction management, supervision, consulting, pollution control fees and engineering insurance. Furthermore, contingency amount generally depends on project scale and uncertainties. The major work items are based on the high frequency of showing up and high percentage costs of historical maintenance projects.

Since the project defined in this stage is purely conceptual, it is not necessary to precisely estimate all work activities and it is reasonable to set an upper bound for the number of cost components that should be retained based on observation of the sudden drop and subsequent flattening. The selection criteria is for work items with cost percentage above 1%, and thus the selected items included asphalt concrete surface removal, asphalt surface rehabilitation, pavement markings (2 mm), and tack coat with bituminous materials. Based on estimates of the major costs, minor item costs can be obtained as relative cost percentages of the major items, and are totaled along with indirect cost and contingency to yield the total project costs.

Many studies have explored the non-linear relationships between cost and project characteristics[2,5,8,9]. Neural network (NN) is another feasible alternative for early cost prediction owing to its ability to model complex systems given a minimal amount of data input[7,10,11]. However, the NN approach is a black box technique and has a steep learning curve for estimators[2]. Some researchers recognize that cost estimates should be represented as a range rather than as a single value[12-14]. Probabilistic cost estimating model with Monte Carlo simulation can be used to design project conditions. A range of cost estimates can then be developed for various scenarios, and through this requires abundant statistical data.

1.2 Applications of case-based reasoning

Case-based reasoning (CBR) is based on the rationale of characterizing projects using a number of essential attributes. These attributes are then used with weighting values to match similar cases. CBR is an idea of cognitive model that incrementally and sustainably learns from particular past experiences. CBR requires a series of systematic procedures to extract relevant knowledge from the experience, integrate a case into an established knowledge structure, index the case for later matching, adapt a new case with similar cases, and save to knowledge base.

Numerous studies have identified CBR as particularly appropriate for application to the conceptual stage of domain knowledge. Particularly, CBR has been used alone or with other data mining techniques in construction management[15,16], construction duration and cost models of civil engineering projects[2,17-19], bidding and international market selection[20-21], and construction method selection system[22].

2 Similarity Measure Using Eigenvector Method

An (2007) compared three different weighting methods and concluded that the analytical hierarchy process (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[18]. This paper determined attribute weights using the groundwork of AHP of Saaty , also known as the eigenvector method (EM), which includes a series of algebra procedures designed to bring out experience from previous cases, enabling pair-wise comparison and consistency checking of engineering judgments.

For each alternative considered it is necessary to elicit the relative weights of considering features through a ratio sale. The first step involves conducting a questionnaire survey and then establishing the pair-wise comparison matrices based on the evaluation attributes. This investigation used a pair-wise comparison matrix, which requires $n(n-1)/2$ comparisons with n attributes listed in the distributed survey forms.

To examine the validity and reliability of the information drawn out from the experienced knowledge, the consistency index (C.I.) can be calculated based on the largest eigenvalue and associated eigenvector of an n by n positive reciprocal matrix. Judgmental consistency concerns the preferred cardinal transitivity in the pair-wise comparison matrices[23].

A questionnaire survey was administered to gather data for considering the attributes and evaluating the weights. Subsequently based on the expert opinions and judgments, seven attributes found to affect preliminary pavement repair costs were included as input attributes, namely project location, construction area, project length, project width, traffic volume and road capacity, project duration, and environmental condition.

Pair-wise comparisons were performed through which one attribute dominated the other. Thirty-eight completed survey forms were returned out of the 68 forms delivered, for a response rate of 55.9%. All subjects have over five years of engineering experience in pavement maintenance projects and have worked in the construction industry for an average of over 12 years. The consistency ratio of the returned forms was 0.0899, which was within the acceptable level (0.1)[23].

The relative importance weighting attributes, W_j, for the selected work items, as listed in Table 1, were subsequently employed in Eq. (1) to measure the similarity between the new case and previously stored cases[24]. Similarity between the cases increased as this value approached one.

$$\text{Similarity}(f^{\mathrm{I}}, f_i^{\mathrm{R}}) = \frac{\sum_{j=1}^{n} W_j \times \text{Sim}(f_j^{\mathrm{I}}, f_{ij}^{\mathrm{R}})}{\sum_{j=1}^{n} W_j} \tag{1}$$

where $\text{Similarity}(f^{\mathrm{I}}, f_i^{\mathrm{R}})$ represents the case similarity measure between the input case, f^{I}, and the retrieved i^{th} case of the case-base, f_i^{R}. Furthermore, $\sum_{j=1}^{n} W_j$ represents the summation weights of n attributes in each case. $\text{Sim}(f_j^{\mathrm{I}}, f_{ij}^{\mathrm{R}})$ is calculated as a ratio representing the i^{th} attribute similarity measure of the input case, f_j^{I}, and the retrieved i^{th} case from the case library. Equation (2) provides the measurement formula for the above attribute similarity. The closer the value is to one, the greater the similarity of the two attributes.

$$\text{Sim}(f_j^{\mathrm{I}}, f_{ij}^{\mathrm{R}}) = \frac{\min(|f_j^{\mathrm{I}}|, |f_{ij}^{\mathrm{R}}|)}{\max(|f_j^{\mathrm{I}}|, |f_{ij}^{\mathrm{R}}|)} \tag{2}$$

where min(·) represents the minimum value of the new input case, f_j^{I}, and the retrieved i^{th} case for the i^{th} attribute and max(·) represents the higher of the above two values.

Notably, the weights of tack coat using bituminous materials were identical to those of asphalt concrete surface removal. This phenomenon occurs because this work item cost is calculated proportionally to the other one based on the observation of the collected project cost data. The relationship is thus deemed to possess the same characteristics for the attribute weights of both work items.

Besides the weights determined using EM, the assumption of equal importance for each attribute was adopted as an alternative weighting method for comparing the modeling results. Figure 1 presents the flow chart of similarity measure calculation using the weights (W_i) provided in Table 1. The retrieved item costs are then summed and divided with a fixed cost percentage to derive the total project cost.

In addition to reasoning out item-based cost from the attributes comparison, a project-based reasoning was performed to obtain the most similar total project cost to the new input case. By doing so, it is possible to identify whether summing component costs using equal weight (EW) or EM, or calculating a direct reasoning aggregated cost is the best approach.

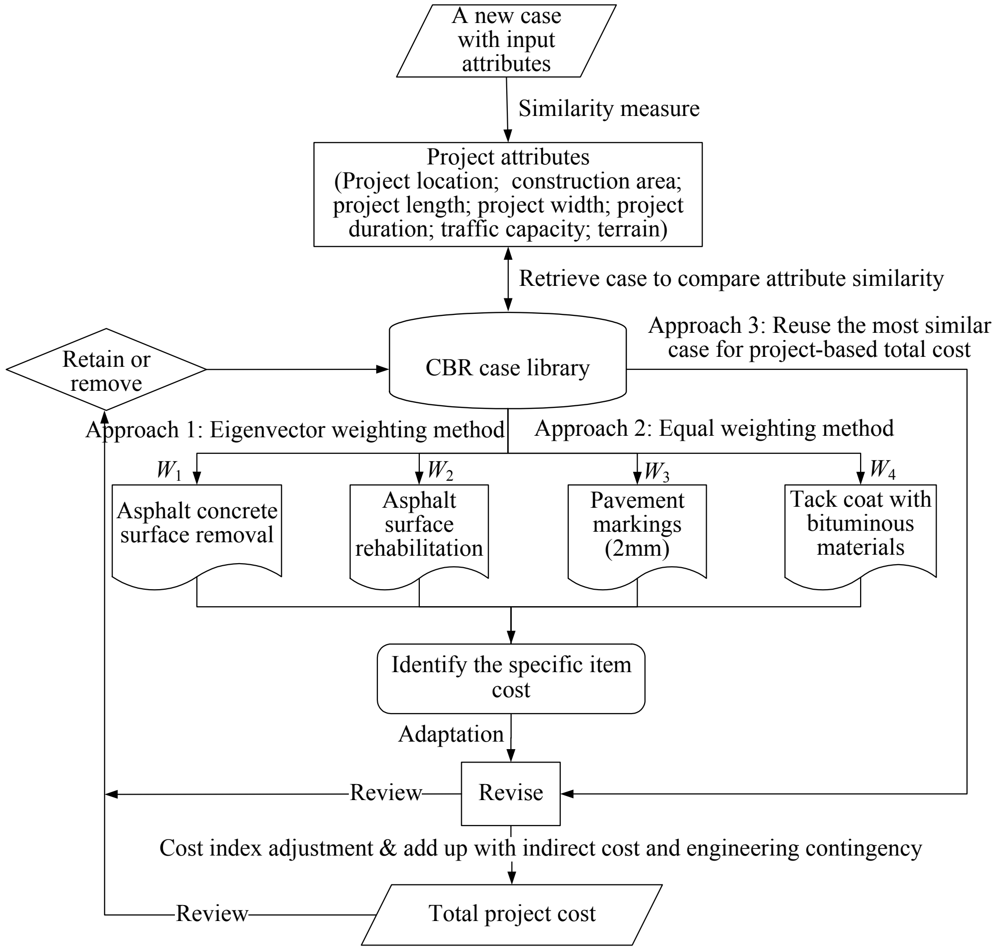

Fig. 1 Estimating procedure with various CBR approaches

Table 1 Relative importance weights using EM

Attributes	Asphalt concrete surface removal (W_1)	Asphalt surface rehabilitation (W_2)	Pavement markings (2 mm) (W_3)	Tack coat using bituminous materials (W_4)
	Weights based on the experienced engineers			
1. Project location	0.0494	0.0375	0.0388	0.0494
2. Construction area	0.2744	0.1966	0.1804	0.2744
3. Project length	0.1814	0.1767	0.1804	0.1814
4. Project width	0.1989	0.2166	0.2268	0.1989
5. Traffic volume and road capacity	0.0767	0.1005	0.1059	0.0767
6. Project duration	0.1504	0.1852	0.1922	0.1504
7. Terrain	0.0687	0.0869	0.0755	0.0687

3 Development of the CBR Models

The procedures of data handling, case base building, model configuration, and model validation are illustrated in the following subsections. To evaluate the performance of the CBR model, the holdout method and cross-validation were employed to assess the accuracy of model reasoning based on mean absolute estimation error rates.

3.1 Data collection and preprocessing

Approximately 300 historical pavement maintenance projects were obtained from 13 divisions of the Taiwan Directorate General of Highways (TDGH) that were geographically distributed throughout Taiwan. However, due to the noisy, incomplete, and inconsistent data or project information, thorough examination of the data yielded just 81 full project cases. In this case, the reusable ratio of reserved district records was somewhat low, at just 27%, owing to the complex nature of construction projects.

3.2 CBR model configuration

CBR is a computational paradigm for problem solving and has been already applied in various applications in the construction industry[2,15-17,19-22,25]. This study integrates CBR, EW, and EM for preliminary project cost estimation. During the retrieval process, the nearest neighboring method is the most widely used in CBR[18], and therefore was adopted in this study.

3.3 Model validation

CBR model performance was assessed using a *k*-fold cross validation, also known as rotation estimation. Kohavi (1995) indicated that a k value of ten seems an optimal number of folds[26]. Therefore the cases were randomly separated into ten mutually exclusive subsets of roughly equal size. Notably, the forecasting accuracy of the test cases was found to be the best by using a single analogy to generate the estimation. This study used equal weights (EW) and weights determined by the eigenvector method (EM) to reasoning (1) the total cost derived from the major component costs associated with a fixed cost percentage, and (2) aggregated cost of a project case, both of which were compared to actual cost for each testing case, so as to calculate the mean absolute estimation error using Eq. (3).

$$\mathrm{MAEE} = \frac{\sum \dfrac{\left|C_{\mathrm{CBR}} - C_{\mathrm{Act}}\right|}{C_{\mathrm{Act}}}}{n} \times 100\% \qquad (3)$$

where,

MAEE: Mean absolute estimation error (%);

C_{CBR} : Cost output via case-based reasoning;

C_{Act} : Actual cost;

n: No. of testing cases.

Table 2 lists the complete set of evaluation results for the item cost reasoning model. Evidently, the item-level CBR model achieved a peak reasoning accuracy of 90% and a minimum reasoning accuracy of 43%. The 10-fold approach can be considered as an effective form of reliability analysis of the measurement technique.

Table 2 Cost reasoning errors for ten-fold cross-validation for major work items

Fold no.	Error rate (%)			
	Asphalt concrete surface removal (ASCR)	Asphalt surface rehabilitation (ASR)	Tack coat using bituminous materials (TCBM)	Pavement markings (2mm) (PM)
1	25	25	19	27
2	15	16	22	10
3	21	26	29	27
4	37	41	27	40
5	28	26	26	31
6	21	12	33	24
7	13	14	28	29
8	26	26	38	54
9	57	34	40	32
10	20	17	19	29
Min.	13	12	19	10
Max.	57	41	40	54
MAEE	26	23	28	30

3.4 Comparison of modeling approaches

Through the inspection of each major work item, the best fold for ACSR to holdout for testing was Fold no. 7 for which reasoning accuracy reached 87%, while ASR, TCBM, and PM were Folds no. 6, 1, and 2, respectively. This finding was then employed as an adaptation principle for selecting the best training sets for the distinct work items. Using Eq. (3), the reasoning error rates were calculated as listed in Table 3 using four different approaches.

Table 3 Summary of various CBR modeling approaches

Testing Case	Error rate (%)			
	EW-CBR		EM-	EM-CBR
	Model I[a]	Model II[b]	CBR[c]	(Adapted)[d]
1	23	26	18	2
2	7	12	7	5
3	66	68	16	2
4	37	40	37	8
5	36	40	36	5
6	15	0	19	10
7	40	42	43	7
8	24	32	24	7
Min.	7	0	7	2
Max.	66	68	43	10
MAEE	31	33	25	6

a. Total project cost is computed by adding the reasoning-out major work items with equal weights of attributes. b. Total project cost is computed by retrieving the most nearest case from the case library. c. Total project cost is computed by adding the reasoning-out major work items with eigenvector weighting method. d. Total project cost is computed by adding the reasoning-out major work items with eigenvector weighting method and usage of best training folds for distinct work items.

For the EW-CBR Model I, using major work item costs to back estimate the total project cost (component model), and Model II, reasoning the total project cost directly (aggregate model). Apparently, the Model I slightly outperformed than Model II. EM-CBR modeling enhanced the reasoning accuracy by 6~8 % compared to EW-CBR. EM-CBR (adapted) approach yielded an MAEE with 6% error rate which obtained the most accurate estimation results among the modeling approaches shown.

4 Conclusions and Future Work

This study presents a case-based reasoning approach to reuse the past experience of experts and determine the important weights of attributes. Reasoning procedure was exhibited for creating the most appropriate CBR model for rapidly estimating the PMO budgets with early information and screening the projects to be programmed during the year.

Two methods were proposed to estimate the PMO costs of a project with several CBR model configurations. The component CBR model appeared to estimate total project costs better than the aggregate model. The CBR model was found to be more effective in considering the experience-based weights of these attributes compared to the situation when it treated them as equally important.

Furthermore, the adaptation involving the use of the best training folds for distinct component models yielded more accurate and reliable total project cost estimation. Future research can be focused on the design and implementation of a web-based system embedded with the proposed CBR model to efficiently handle the similarity computation and case retrieval.

References

[1] MOTC. http://www.thb.gov.tw/. 2007 (cited Nov. 2006).

[2] Kim G H, An S H, Kang K I. Comparison of construction cost estimating models based on regression analysis, neural networks, and case-based reasoning. *Building and Environment*, 2004, **39**(10): 1235-1242.

[3] Duverlie P, Castelain J M. Cost estimation during design step: Parametric method versus case based reasoning method. *International Journal of Advanced Manufacturing Technology*, 1999, **15**(12): 895-906.

[4] Han S G, Lee S G, Jo G S. Case-based tutoring systems for procedural problem solving on the www. *Expert Systems with Applications*, 2005, **29**(3): 573-582.

[5] Chou J S, Peng M, Persad K R, et al. Quantity-based approach to preliminary cost estimates for highway projects. *Transportation Research Record*, 2006(1946): 22-30.

[6] Chengalur Smith I N, Ballou D P, Pazer H L. Modeling the costs of bridge rehabilitation. *Transportation Research Part A: Policy and Practice*, 1997, **31**(4): 281-293.

[7] Emsley M W, Lowe D J, Duff A R, et al. Data modelling and the application of a neural network approach to the prediction of total construction costs. *Construction Management and Economics*, 2002, **20(**6): 465-472.

[8] Lowe D J, Emsley M W, Harding A. Predicting construction cost using multiple regression techniques. *Journal of Construction Engineering and Management*, 2006, **132**(7): 750-758.

[9] Yu W D. PIREM: A new model for conceptual cost estimation. *Construction Management and Economics*, 2006. **24**(3): 259-270.

[10] Hegazy T, Ayed A. Neural network model for parametric cost estimation of highway projects. *Journal of Construction Engineering and Management*, 1998, **124**(3): 210-219.

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

[12] Diekmann J E. Probabilistic estimating: mathematics and applications. *J. Constr. Engrg. and Mgmt*., 1983, **109**(3): 297-308.

[13] Chou J S, Wang L, Chong W K, et al. Preliminary cost estimates using probabilistic simulation for highway bridge peplacement projects. In: American Society of Civil Engineers, Construction Research Congress. San Diego, California, USA, 2005: 467-471.

[14] Touran A, Suphot L. Rank correlation in simulating construction costs. *Journal of Construction Engineering and Management*, 1997, **123**: 297-301.

[15] Yau N J, Yang J B. Case-based reasoning in construction management. *Computer-Aided Civil and Infrastructure Engineering*, 1998, **13**(2): 143-150.

[16] Yu W D, Liu Y C. Hybridization of CBR and numeric soft computing techniques for mining of scarce construction databases. *Automation in Construction*, 2006, **15**(1): 33-46.

[17] Dogan S Z, Arditi D, Gunaydin H M. Determining attribute weights in a CBR model for early cost prediction of structural systems. *Journal of Construction Engineering and Management*, 2006, **132**(10): 1092-1098.

[18] An S H, Kim G H, Kang K I. A case-based reasoning cost estimating model using experience by analytic hierarchy process. *Building and Environment*, 2007, **42**(7): 2573-2579.

[19] Bailey S F, Smith I F C. Case-based preliminary building design. *Journal of Computing in Civil Engineering*, 1994, **8**(4): 454-467.

[20] Ozorhon B, Dikmen I, Birgonul M T. Case-based reasoning model for international market selection. *Journal of Construction Engineering and Management*, 2006, **132**(9): 940-948.

[21] Chua D K H, Li D Z, Chan W T. Case-based reasoning approach in bid decision making. *Journal of Construction Engineering and Management*, 2001, **127**(1): 35-45.

[22] Yau N J, Yang J B. Applying case-based reasoning technique to retaining wall selection. *Automation in Construction*, 1998, **7**(4): 271-283.

[23] Saaty T L. The Analytical Hierarchy Process. New York: McGraw-Hill, 1980.

[24] Kolodner J L. Case-based reasoning. San Mateo, California: Morgan Kaufmann Publishers Inc., 1993.

[25] Roddis W M K, Bocox J. Case-Based Approach for steel bridge fabrication errors. *Journal of Computing in Civil Engineering*, 1997, **11**(2): 84-91.

[26] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: The 14th International Joint Conference on Artificial Intelligence (IJCAI). San Francisco, CA, 1995: 1137-1145.