KnowPrice2: Intelligent Cost Estimation for Construction Projects

Bernd Domer[1], Benny Raphael[2], and Sandro Saitta[3]

[1] Tekhne management SA, Avenue de la Gare 33, 1003 Lausanne, Switzerland
[2] National University of Singapore, 4 Architecture Drive, Singapore 117566, Singapore
[3] Ecole Polytechnique Fédérale de Lausanne, IMAC, 1015 Lausanne, Switzerland
domer@tekhne.ch, bdgbr@nus.edu.sg, sandro.saitta@epfl.ch

Abstract. Correct estimation of costs of construction projects is the key to project success. Although mostly established in early project phases with a rather limited set of project data, estimates have to be precise. In this paper, a methodology for improving the quality of estimates is presented in which data from past projects along with other knowledge sources are used. Advantages of this approach as well as challenges are discussed.

1 Introduction

Correct estimation of construction costs is one of the most important factors for project success. Risks vary with the contractual context in which estimations or price biddings are presented. Table 1 discusses three standard configurations.

Table 1. Possible contractual configurations for construction projects

	General planner	**General contractor**	**Total services contractor**
Design contract	One contract with general planner	One contract with general planner	One contract with total services contractor for design and construction
Contract for-construction works	Multiple contracts with each sub-contractor	One contract with general contractor	
Type of cost estimation	Cost estimation	Price bidding	Price bidding
Risk, when estimate is too high	Project will be stopped	Project might be executed by another contractor with a lower bid	Project might be executed by another contractor with a lower bid
Risk, when estimate is too low	Project will suffer quality problems due to tight budget	Bankruptcy of sub contractors	Bankruptcy of sub contractors or total services contractor

Decisions that have the biggest impact on project success (that is, related to cost) are taken in early project phases. Cost control over project duration can only prevent

I.F.C. Smith (Ed.): EG-ICE 2006, LNAI 4200, pp. 147 – 152, 2006.

an established budget from getting out of control. It cannot correct fundamental errors made in the beginning.

At the start of a project, many details are unknown and cost estimators have to make several assumptions based on their experience with similar projects in the past. Construction managers are not only interested in accurate estimates, but also in the level of risk associated with estimates. A systematic methodology for performing this task is desirable. This paper presents a methodology for cost estimation that has been developed and implemented in a software package called KnowPrice2. The primary objective of this paper is to compare the methodology with current practices in the industry, rather than to provide a detailed description of the approach, which is given in [10].

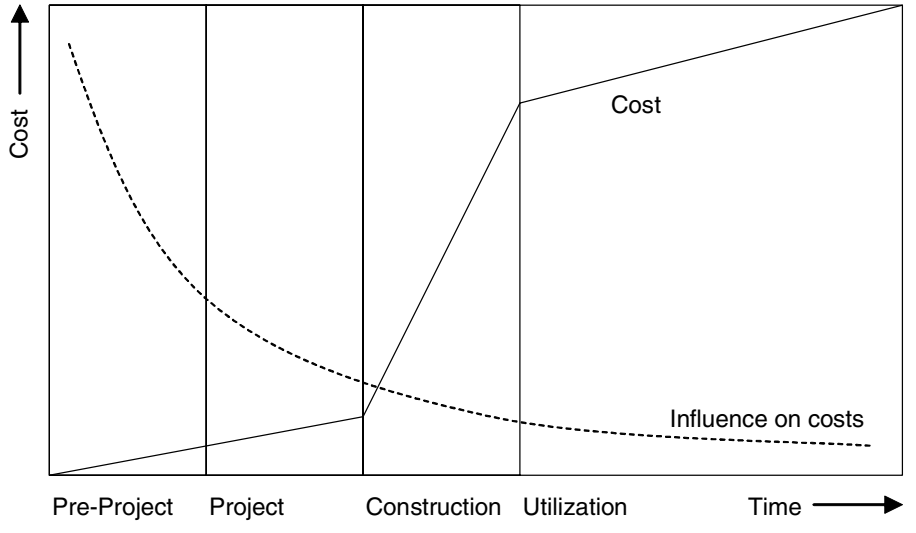

Fig. 1. Relationship between cost and the possibility to influence them in different project phases [1]

2 Estimating Construction Costs – Existing Methodologies and Data

Since the preparation of cost estimates is not a new challenge, several proposals on how to establish and structure construction budgets have been made. Databases with cost data of past projects have been prepared as well.

In most European countries, building codes define how to measure building surfaces and volumes [eg. 2, 3] and to structure building costs in cost groups [eg. 4, 5]

Although building codes for structures have been unified at European level, this has yet to be done for the above mentioned guidelines. A proposal has been made [6].

This means that even today the way building surfaces are measured as well as the structure of cost groups differ from country to country. As a result, the comparison of cost data from different countries is difficult.

The Swiss Research centre for Rationalization in Building and Civil Engineering (CRB) proposes methodologies such as the costs by element method [7]. Cost data of construction elements is provided and updated each year, but it does not account for regional variation of prices or building types. This approach needs a very detailed breakdown of all building elements and is therefore not the method of choice for quick estimates.

Data of existing buildings have been collected as well. Germany's BKI, the community of architectural chambers publishes each year a well structured and exhaustive catalogue of project costs [8]. The structure divides buildings into different types, gives examples for each type and provides standard deviation for price data. Regional data is included as well, since prices vary locally.

Although it is very tempting to use this data collection for swiss projects, construction is not yet as "globalized" as other industries. As discussed previously, data collections of other countries cannot be re-used without serious cleaning and adaptation.

3 KnowPrice2 Strategy

3.1 Background

KnowPrice2 is a software package for cost estimation that was developed by the Swiss Federal Institute of Technology (EPFL) in collaboration with an industrial partner, Tekhne management SA. It links case data with a unique approach for establishing construction project budgets. The employed methodology differs significantly from other database approaches such as [9], since case-based reasoning strategies are combined with relationships between variables that are discovered by data mining.

3.2 Methodology

The total project cost is estimated using knowledge from different sources. Knowlege sources include generic domain knowledge in the form of rules, cases consisting of data related to past projects and relationships that are discovered by mining past project data. Depending on what data is known about the current project, appropriate type of knowledge is used. For example, if all the variables that are needed for applying the rules are available, generic domain knowledge is used. Otherwise, relationships that are discovered by data mining are used. Only when relevant relationships are not available, case data is used.

Generic domain knowledge contains equations for computing costs of building elements by summing up costs of components or using unit costs. Each case contains characteristics of a building which includes information such as types, quantities and costs of elements. Data mining aims to discover relationships between building characteristics and costs. Rules of the following form are discovered through this process:

Where q_1, q_2, etc. are quantities and c_1, c_2, .. b are coefficients that are determined by regression using relevant case data. The data mining process is guided by the knowledge of dependencies which is provided by domain experts. A dependency relationship indicates that certain symbolic and quantitative variables might influence a cost variable. The exact relationship between variables is determined by the data

mining module through analysing case data. More details of the data mining technique are presented elsewhere [10].

The CBR module selects relevant cases to a new project using a similarity metric. By default, a similarity metric that gives equal importance to all the input variables is used. In addition, users can define their own similarity metrics by specifying relevant variables and weights. Selected cases are used to estimate the variations in the values of independent variables.

$$\begin{aligned} &\text{IF symbolic\_variable EQUALS value THEN} \\ &\text{cost\_variable} = c_1 * q_1 + c_2 * q_2 + c_3 * q_3 + c_n * q_n + b. \end{aligned} \tag{1}$$

Steps involved in the application of the cost estimation methodology are the following:

1. Users input known data related to a new project
2. The system creates a method for computing the total cost using generic rules and relationships that are discovered by data mining
3. Variations in the values of independent variables are determined from similar past cases and are represented as probability density functions (PDF).
4. Monte-carlo simulation is carried out for obtaining the probability distribution of total cost using PDFs of independent variables.

Since the methodology computes the PDF of the total cost instead of a deterministic value, information such as the likelihood of an estimate exceeding a certain value is also available. This permits choosing the bid price at an acceptable level of risk. This is not possible using conventional deterministic cost estimation.

3.3 Program Structure

The knowledge base is structured into five modules, namely, Generic domain knowledge, Dependencies, Cases, Ontology and Discovered knowledge.

The first module "Generic domain knowledge" implements rules to achieve the objective. Here, the objective is to compute the overall building cost, obtained by the summation of building cost classifications (BCC).

The second module "Dependencies" describes dependencies between a) the dependent variable and b) the influencing variable. This module is, from the industrial partner's point of view, one of the major achievements of KnowPrice2. Whereas in other programs the user has to describe relationships in a deterministic way, KnowPrice2 employs a different approach. In most cases, cost estimators cannot provide the deterministic expressions to relate building characteristics with BCCs directly. It is much easier to indicate that there is a relationship without giving precise values. The major achievement of Knowprice2 is that it evaluates the relationship between variables using existing data and data mining techniques [10].

Cases are input in a semi-automatic procedure using an Excel spreadsheet in which data is grouped into building characteristics such as surfaces, volumes, etc. and costs, structured according to the Swiss cost management system [5]. The same interface allows to management of cases.

The ontology module contains the decomposition hierarchy for organising variables, the data type of each variable, default values of variables and possible values of symbolic variables.

The module "discovered knowledge", contains a decision tree that organises all the relationships that are discovered by data mining.

3.4 Challenges

Collaboration between EPFL and Tekhne has been very close which means that the project did not suffer from major problems. One main challenge was to adapt the Dependencies module such that it can treat non-deterministic relations.

The second challenge was (and still is) to provide data for testing. In a first effort, a database with past projects of Tekhne has been created. This database has been used to test the basic functionality of KnowPrice2. It is not sufficient to do the fine tuning of the software.

4 Advantages of KnowPrice2

From the industrial partner's point of view, KnowPrice offers several advantages:

The approach used for the cost estimation depends no longer on personal experience only but links it with intelligent computational techniques to support the user. Employed methods have a sound scientific basis and increase the client's confidence in budgets thereby.

Costs can be estimated with incomplete project data (this means, when all project details are not known). It is up to the user to increase the degree of precision by increasing the number of case data and case variety.

Cases can be entered via Excel spreadsheets. This is very convenient for the user. Values for descriptive variables are proposed to guide user input. The amount of data entered depends on information present and not pre-fixed by the software.

The final cost is presented as a price range with associated probabilities. The client can choose the degree of risk he would like to take: either going for a low budget with a rather high risk of exceeding the estimate or to announce higher budgets with lower risks.

Costs can be related to similar previous projects and can thus be justified.

5 Future Work

The quality of results highly depends on the number and variety of cases entered. So far, only cases of the industrial partner (Tekhne) have been entered. KnowPrice2 works correctly so far, but needs definitely more data for proper testing.

Even though building cost classifications are defined in codes, they are not unambiguous. They leave space for interpretation and each user might apply them differently. The effects of this have not yet been examined.

KnowPrice2 has to be tested under real project conditions. This test might reveal necessities for changes in the program itself and necessary user interface adaptation.

Acknowledgements

This research is funded by the Swiss Commission for Technology and Innovation (CTI) and Tekhne management SA, Lausanne.

References

1. Büttner, Otto: Kostenplanung von Gebäuden. Phd Thesis, University of Stuttgart (1972)
2. DIN 277: Grundflächen und Rauminhalte von Bauwerken im Hochbau. Beuth Verlag (2005)
3. SIA 416: Flächen und Volumen von Gebäuden. SIA Verlag, Zürich (2003)
4. DIN 276: Kosten im Hochbau. Beuth Verlag (1993)
5. CRB (Ed.): Building cost classification. CRB, Zürich (2001)
6. CEEC (Ed.): Le code européen pour la planification des coûts, CEEC (2004)
7. CRB (Ed.): Die Elementmethode – Informationen für den Anwender. CRB, Zürich (1995)
8. BKI (Ed.): Baukosten Teil 1: Statistische Kostenkennwerte für Gebäude. BKI, Stuttgart (2005)
9. Schafer, M., Wicki, P.: Baukosten Datenbank "BK-tool 2.0", Diploma thesis, HTL Luzern (2003)
10. Raphael, B., et. Al.: Incremental development of CBR strategies for computing project cost probabilities, submitted to Advanced Engineering Informatics, 2006.
11. Campi, A., von Büren, C.: Bauen in der Schweiz – Handbuch für Architekten und Ingenieure, Birkhäuser Verlag (2005)