Expert Systems with Applications 38 (2011) 9913–9917

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Range estimation of construction costs using neural networks with bootstrap prediction intervals

Rifat Sonmez *

Department of Civil Engineering, Middle East Technical University, Ankara 06531, Turkey

ARTICLE INFO

Keywords:
Neural networks
Cost estimation
Bayesian regularization
Bootstrap method
Construction projects

ABSTRACT

Modeling of construction costs is a challenging task, as it requires representation of complex relations between factors and project costs with sparse and noisy data. In this paper, neural networks with bootstrap prediction intervals are presented for range estimation of construction costs. In the integrated approach, neural networks are used for modeling the mapping function between the factors and costs, and bootstrap method is used to quantify the level of variability included in the estimated costs. The integrated method is applied to range estimation of building projects. Two techniques; elimination of the input variables, and Bayesian regularization were implemented to improve generalization capabilities of the neural network models. The proposed modeling approach enables identification of parsimonious mapping function between the factors and cost and, provides a tool to quantify the prediction variability of the neural network models. Hence, the integrated approach presents a robust and pragmatic alternative for conceptual estimation of costs.

1. Introduction

In early stages of construction projects, detailed design drawings are not usually available and conceptual estimation of costs is required for making budgeting and feasibility decisions. Cost models provide a powerful alternative for conceptual estimation of construction costs. However, development of cost models can be challenging as there are several factors impacting costs, and there is usually sparse and noisy data available for modeling.

Regression models have been used commonly to quantify the impact of factors on project costs (Kaiser, 2006; Karshenas, 1984; Kouskoulas & Koehn, 1974; Lowe, Emsley, & Harding, 2006; Sonmez, 2008). Regression analysis requires the user to decide a priori on the class of relations (linear, quadratic, etc.) to be used in modeling. Determination of the class of relations between the factors and project costs may become complicated, especially when multiple cost components are considered as the dependent variables.

Neural networks (Adeli & Wu, 1998; Cheng, Tsai, & Sudjono, 2009; Duran, Rodriguez, & Consalter, 2009; Hegazy & Ayed, 1998; Kim, Seo, & Kang, 2005; Sonmez, 2004) and case-based reasoning (Chou, 2009; Dogan, Arditi, & Gunaydin, 2006; Wang, Chiou, & Juan, 2008) models have been proposed in recent years for modeling of costs as an alternative to regression analysis. Neural network and case-based reasoning cost models usually provide a point estimate for estimating costs. However, a single point prediction does not include any information regarding the level of variability included in the estimated costs. Inclusion of estimation variability is very crucial for management decisions as conceptual cost estimates usually include a high amount of uncertainty.

The level of uncertainties included in the cost estimates can be quantified by developing range estimates using simulation techniques (Touran & Wiser, 1992; Wang, 2002). However the impacts of parameters on project costs are not generally included in simulation techniques. Parametric range estimation can be performed by using prediction intervals in regression models but, in this method a priori decision on the class of relations is required (Sonmez, 2004, 2008). Within this content, the main purpose of this study is to develop a method for range estimation of costs, which can identify the impact of parameters on costs easily, and can also quantify the level of uncertainties included in the estimated costs. The remainder of the paper is organized as follows: Section 2 is devoted to description of the project data and formulation of the cost modeling problem. In Section 3, neural network models are described. Bootstrap prediction intervals are presented in Section 4. Finally, concluding remarks are made in Section 5.

2. Modeling of building costs

Construction cost models in general reflect experiences that are unique to a construction organization for a certain project type. In this study cost models are developed for continuous care retirement community (CCRC) projects. CCRCs are living units for

* Tel.: +90 3122102422.
E-mail address: rsonmez@metu.edu.tr

0957-4174/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2011.02.042

Table 1
Factors impacting cost.

No	Description
$X1$	Total gross residential, commons, nursing facilities, and structured parking area in m^2
$X2$	Construction cost index
$X3$	City cost index
$X4$	Number of stories
$X5$	Percent area of commons and nursing facilities in the total building area
$X6$	Percent structured parking area in total area
$X7$	Total gross building area per residential unit
$X8$	Site area in m^2
$X9$	Major demolition on site
$X10$	Site waste treatment
$X11$	Wood frame
$X12$	Steel frame
$X13$	Concrete frame
$X14$	Steel and concrete frame
$X15$	Masonry structure
$X16$	Wood exterior finish
$X17$	Vinyl exterior finish
$X18$	Masonry exterior finish
$X19$	Plaster exterior finish
$X20$	Number of elevator stops
$X21$	Project duration in months

Table 2
Cost components.

No	Description
$Y1$	Site development
$Y2$	Foundations and slab on grade
$Y3$	Structure
$Y4$	Enclosure
$Y5$	Interior finishes
$Y6$	Equipment and special construction
$Y7$	Conveying systems
$Y8$	Mechanical
$Y9$	Fire protection
$Y10$	Electrical
$Y11$	General requirements

seniors, and offer them access to coordinated social activities, dining and health care services. Models are developed using data of 20 CCRC projects compiled from a building contractor. The projects were built over a 13 year time frame, at 10 different locations in the United States. The data included information of 21 factors which are presented in Table 1. Factors used in this study are the variables related to building, site, and project conditions which might impact the project costs. The factors $X9, X10, \ldots, X19$ are binary variables, and are used to represent presence of a certain condition such as; major demolition on the site ($X9$). The cost components are the system costs, and are defined according to a cost breakdown structure. The contractor used 11 cost components to organize the CCRC costs, as shown in Table 2.

The task of cost modeling for CCRC projects is determination of the relations between the factors ($X1, X2, \ldots, X21$) and cost components ($Y1, Y2, \ldots, Y11$). Quantitative relations between the factors and cost components can be represented by a single overall model, or by an individual model for each cost component. In regression modeling the factors ($X1, X2, \ldots, X21$) are the independent variables and the cost components ($Y1, Y2, \ldots, Y11$) are the dependent variables. One of the main difficulties of cost modeling by regression analysis is determination of a proper model representing the relations between the factors and cost components adequately. Linear regression models without any interaction terms can be used to simplify the modeling process. However, linear models do not always guarantee adequate representation of the relations. An alternative approach, as implemented in this study, is to use neural networks to establish a mapping function between the factors and cost components.

3. Neural network models

Feed forward neural networks are used to develop an adequate cost model for the CCRC projects. The input buffer of the first neural network model consisted of 21 units, representing all of the factors ($X1, X2, \ldots, X21$) and, the output layer consisted of 11 units representing the cost components ($Y1, Y2, \ldots, Y11$), as shown in Fig. 1. Three neural networks with different number of hidden units were trained to determine the number of hidden units for the first neural network model. The neural networks had one hidden layer including 32 (Model-1a), 16 (Model-1b), and 8 (Model-1c) hidden units. Back propagation algorithm with an adaptive

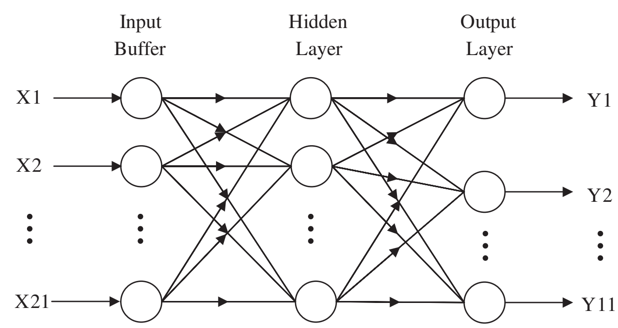

Fig. 1. Model-1.

Table 3
Prediction performance of Models 1a, 1b, and 1c.

Model	Nh*	MAPE
Model-1a	32	32.4
Model-1b	16	27.7
Model-1c	8	33.3

* Nh: Number of units in the hidden layer.

learning rate was used for training. In adaptive learning rate, the learning step size is kept as large as possible while maintaining a stable learning, by making the learning rate responsive to the complexity of the local error surface (Demuth & Beale, 2001). Leave-one-out cross validation was performed to evaluate the adequacy of the neural network models. One project data was not used during training, and the trained network was used to predict the total cost of that project. The procedure was repeated for all the projects, and predicted costs were compared with the actual estimated costs to assess the prediction performance. Mean absolute percent error (MAPE) was used as an error measure to evaluate the prediction performance. MAPE value for a cost model was the average of deviations between predicted total project cost and actual estimated total project cost in absolute values; expressed as proportion of the actual estimated cost. MAPE values for Model-1a, Model-1b and Model-1c are 32.4, 27.7 and 33.3 respectively as shown in Table 3. 16 hidden units are used for Model-1 based on the results.

3.1. Elimination of factors

The second cost model (Model-2) consisted of eleven neural networks ($N1, N2, \ldots, N11$) with each having one unit in the output layer representing the cost components ($Y1, Y2, \ldots, Y11$) respectively. For each neural network model the factors which may have a potential impact on the cost component was included