CONCEPTUAL QUANTITY MODELING OF SINGLE SPAN HIGHWAY BRIDGES BY REGRESSION, NEURAL NETWORKS AND CASE BASED REASONING METHODS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MERT AŞIKGİL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
CIVIL ENGINEERING

JUNE 2012

Approval of the thesis:

CONCEPTUAL QUANTITY MODELING OF SINGLE SPAN HIGHWAY BRIDGES BY REGRESSION, NEURAL NETWORKS AND CASE BASED REASONING METHODS

submitted by **MERT AŞIKGİL** in partial fulfillment of the requirements for the degree of **Master of Science in Civil Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen ___________
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Güney Özcebe ___________
Head of Department, **Civil Engineering**

Assoc. Prof. Dr. Rıfat Sönmez ___________
Supervisor, **Civil Engineering Dept., METU**

Examining Committee Members:

Prof. Dr. M. Talat Birgönül ___________
Civil Engineering Dept., METU

Assoc. Prof. Dr. Rıfat Sönmez ___________
Civil Engineering Dept., METU

Prof. Dr. Murat Gündüz ___________
Civil Engineering Dept., METU

Asst. Prof. Dr. Aslı Akçamete ___________
Civil Engineering Dept., METU

Hüseyin Karancı, MSCE ___________
Planning Specialist, TeknoKA Consultancy

Date: June 07, 2012

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name : Mert Aşıkgil

Signature :

ABSTRACT

CONCEPTUAL QUANTITY MODELING OF SINGLE SPAN HIGHWAY BRIDGES BY REGRESSION, NEURAL NETWORKS AND CASE BASED REASONING METHODS

Aşıkgil, Mert
M.Sc., Department of Civil Engineering
Supervisor: Assoc. Prof. Dr. Rıfat Sönmez

June 2012, 73 Pages

Conceptual estimation techniques play an important role in determining the approximate costs of construction projects especially during feasibility stages. Moreover, pre-design estimates are also crucial for the contractors. With the help of the conceptual predictions companies can determine approximate project costs and can gain several advantages before tendering phase.

The main objective of this thesis is to focus on modeling of quantities instead of costs and to develop quantity take-off models for pre-design cost estimation of bridge projects. Majority of the existing studies focus on modeling of costs for conceptual cost estimation. This study includes modeling of the quantity take off items in a specific single span highway bridge using three different techniques namely, linear regression, neural network and case based reasoning. During this study 40 single span highway bridge projects whose owner is Republic of Turkey

General Directorate of Railways, Ports and Airports Constructions were investigated and models for each work item were developed. Then by integrating the quantity take off estimations with unit costs, total project costs were calculated. As a result by evaluating the prediction performance of the models, comparison of the methods was achieved. Results are discussed along with the advantages of the proposed method for conceptual cost estimation of bridge projects.

Keywords: Regression, Neural Network, Case Based Reasoning, Quantity Take off, Conceptual Estimation

ÖZ

TEK AÇIKLIKLI KARAYOLU KÖPRÜLERİNDE REGRESYON ANALİZİ, YAPAY SİNİR AĞI VE VAKA BAZLI ÇÖZÜMLEME METODLARI KULLANARAK KAVRAMSAL MİKTAR MODELLEMESİ

Aşıkgil, Mert
Yüksek Lisans, İnşaat Mühendisliği Bölümü
Tez Yöneticisi: Doç. Dr. Rıfat Sönmez

Haziran 2012, 73 Sayfa

Kavramsal tahmin teknikleri inşaat projeleri tahmini maliyetlerini belirlemede, özellikle fizibilite aşamasında büyük bir öneme sahiptir. Bunun dışında tasarım öncesi tahminlerin taahhüt firmalarına da katkısı büyüktür. Bu tarz çalışmalarla firmalar ihale öncesi fiyat mertebelerini tahmin edebilecekleri için ihale öncesi kararlarda çeşitli avantajlar elde etmiş olurlar.

Bu tezin ana amacı fiyatlar yerine miktar modellemesine odaklanarak köprüler için tasarım öncesi fiyat tahmini elde etmek için metraj modelleri oluşturmaktır. Mevcut çalışmalar kavramsal maliyet tahmini için maliyetleri modellemeye odaklanmıştır. Bu çalışma tek açıklıklı karayolu köprüleri metraj kalemlerinin, regresyon, yapay sinir ağı, ve vaka bazlı çözümleme metotları kullanılarak oluşturulan modelleri içermektedir. Bu çalışma esnasında işvereni Türkiye

Cumhuriyeti Ulaştırma Bakanlığı Demiryolları Limanlar ve Hava Meydanları İnşaatı Genel Müdürlüğü olan 40 tek açıklıklı karayolu köprüsü incelenmiş, her bir metraj kalemi için model kurulmuş ve tahmin edilen miktarların, birim fiyatlarlar ile birlikte kullanılması sonucu proje toplam fiyatları da tahmin edilmiştir. Ayrıca modellerin tahmin performanslarına göre uygulanan metotların karşılaştırılması yapılmıştır. Sonuçlar kıyaslanmış önerilen metodun kavramsal maliyet tahmini için avantajları belirtilmiştir.

Anahtar Kelimeler: Regresyon, Yapay Sinir Ağı, Vaka Bazlı Çözümleme, Metraj, Kavramsal Tahmin

To My Family

ACKNOWLEDGEMENTS

I would like to express my gratitude to my Supervisor, Assoc. Prof. Dr. Rıfat Sönmez, without whom I would not be able to complete my thesis, for his patience, guidance and encouragement at each step of this study. His unlimited assistance, comments and constructive critiques that shaped my study throughout the working process should never be forgotten.

I would like to express my gratitude to executives of Yüksel Domaniç Ltd. I would also like to express my gratitude to my supervisors and colleagues at METAG İnş. Tic. Aş. for their support through this study. I would mostly like to thank to Mr. Okan Gözen, Mr. Ertan Tüz, Mr. Ahmet Turhan and Mr. Altuğ Öztekin for their understanding and facilities that they provided for the completion of this study.

I would also like to express my grateful thanks to my father Ahmet Aşıkgil, my mother Nevin Aşıkgil, and my brother Barış Aşıkgil who never left me alone and provide endless support throughout my life, deserves special emphasis. I would like to thank to them for their endless love.

I would like to especially thank to my friend and colleague Hüseyin Karancı who gave invaluable supports throughout this study also deserves special emphasis.

I would also like to especially thank to my friend Alper Aldemir who never left me alone and provides endless support throughout my life, also deserves special emphasis.

Finally, I would like to express my sincere thanks to my friends, Bengi Kayışoğlu, and Gözde Çelik whose support I felt with me all through this study.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

CBR	: Case Based Reasoning
CBRM	: Case Based Reasoning Models
ID	: Iterative Dichotomiser
MAER	: Mean Absolute Error Rate
MAPE	: Mean Average Percent Error
MSE	: Mean Squared Error
NN	: Neural Network
R^2	: Coefficient of Determination
RPA	: Republic of Turkey General Directorate of Railways, Ports and Airports Constructions
SSE	: Sum of Squared Error

CHAPTER 1

INTRODUCTION

Construction is one of the most important industries in the world. From prehistoric age people tried to construct structures including bridges, amphitheatres, dams, electricity pylons, roads and canals. Initially structures were built for only human want, but with the passing of time this situation turned into project concept.

Project is defined in Project Management Body of Knowledge (PMBOK) as a temporary endeavor because it has a defined beginning and finishing date, and unique since it is not a routine operation (Project Management Institute, 2008). Alternatively, Oberlender (2000) defined project as set of activities which has certain scope, budget and schedule. As can be seen Oberlender's (2000) description one of the project constituent is the budget.

Cost of project is not only crucial for the client but also crucial for the contractor. During the decision phase client's feasibility decision is mostly based on cost estimation, and on the other hand because the lowest price is awarded in tendering stage, cost estimation is also crucial for the contractors.

Estimation is an attempt to predict the actual cost and can be divided into two groups as conceptual (early) estimates and detailed estimates. Reference point for the division is the detailed design. Before the detailed design because there are no detailed drawings etc. estimates are classified as conceptual. After the detailed design stage with the help of detailed specifications and drawings, detailed estimates are done for bidding.

Early estimates have important roles on projects and these roles can be summarized as follows:

Early estimates

- Serve as the owner's feasibility estimate
- Aid engineering company to a specific budget
- Assist owner for developing funding
- Serve as basis for evaluating contractor bids

Because the poor estimates can lead to losing of business opportunities, wasted development effort, lower than expected returns, and unrealistic budget both clients and contractors should give due importance to the conceptual estimation (Oberlender, 2000).

Conceptual estimation methods can be specified as;

- Unit Cost Method
- Factor Method
- Probabilistic Modeling and Simulation
- Parametric Estimation

Methods listed above are generally used for direct conceptual cost estimation. For example, in unit cost method, predetermined cost value per unit is taken into account for estimating the total cost of project. Moreover, in factor method, cost estimation can be made with the help of the historical data as a factor of the equipment cost for industrial projects. If it is not possible to analyze the relationship between variables, probabilistic modeling or simulation should be preferred. Furthermore, estimation can be done by using parametric estimation method when there are relationship between independent and dependent variables by means of statistical analysis.

For detailed estimation, in order to be accurate, quantity take off should be performed by the estimator after the preparation of detailed drawings. In the past quantity take offs were performed only manually and it was tedious and time consuming. With the help of developing technologies, time dedicated for quantity take off were decreased, however it is still not possible to make calculations for feasibility evaluation before the detailed drawings are developed. Therefore in the literature most of the conceptual cost estimation studies were done by estimating directly cost of the project without quantifying project works.

Apart from most of the studies related with the conceptual estimation in the literature, this study aims to present a methodology to predict the quantity of the works and to make cost estimation by using the predicted quantities. In order to make prediction for the works, three methods were used namely linear regression, neural network and case based reasoning method. The conceptual quantity take-off methodology is applied to 40 single span pre-stressed concrete beam highway bridges in Turkey. The owner of the projects is Republic of Turkey General Directorate of Railways, Ports and Airports Constructions (RPA).

Rest of the study can be listed as follows:
In chapter 2, the current literature is reviewed.
In chapter 3, the details of the modeling techniques are given.
In chapter 4, information about data used are given
In chapter 5, preparation of the linear regression model and the results are presented.
In chapter 6, modeling of neural network and the results are described.
In chapter 7, the generation of the case based reasoning model and results are shown
In chapter 8, by comparing the models generated by each method the results of the each model are evaluated
In chapter 9, summary of the thesis, discussions and final statements are given

CHAPTER 2

LITERATURE REVIEW

Conceptual cost forecasts are done in order to determine the approximate budget of a construction project. Through this estimation study, client can set the budget, predict the bidding cost and manage the design in order to meet the budget (Ashworth, 1988). According to Marr (1977) estimation error at the pre-design stage may be as high as 20-40 percent of the actual cost. Moreover this error depends on level of expertise and familiarity of the estimator which can be said personal factors (Ashworth and Skitmore, 1983). Therefore, by generating database for similar projects and evaluating them, modeling can be done and with the help of these models errors depending on the personal factors can be decreased significantly.

In the literature various techniques have been used for predicting the predesigned cost values. Regression analysis, neural networks, case based reasoning methods are the main techniques which were used for conceptual cost estimation.

Kouskoulas and Koehn (1974) studied on predesigned cost estimation for buildings by evaluating the affects of the variables, which are building types, quality, building technology, number of floors, construction year and location. Karshenas (1984) did not arbitrarily select linear equation like Kouskoulas and Koehn (1974). By using least square method, Karshenas (1984) decided a function which described the cost variation best. The function was nonlinear and depended on two variables namely typical floor area and height of building. Karshenas

(1984) thought that type of construction had huge effect on the final cost so he made his model for only one type of structure which is multistory steel framed office buildings. In the analysis, typical buildings which don't have unusual floor heights or extraordinary extremely wide span are used and there were totally 24 building data built in United States (Karshenas, 1984). For evaluating the models both Kouskoulas and Koehn (1974) and Karshenas (1984) calculated residuals in their studies.

Regression modeling was used for cost estimation for bridge rehabilitation by Smith, Ballou and Pazer (1996). In their study total rehabilitation cost was divided into three parts. These were substructure, superstructure and deck. For calculating the related costs two approaches were generated. First one was estimating the total cost for parts and the second one was predicting the unit costs for parts. Moreover 72 bridge data were divided into two as 44 single spans and 28 multiple span bridges. Models were obtained by using non-linear relations. As a result estimation errors were observed much higher than the generally experienced errors in the literature and this situation described with the estimation complexity of the bridge rehabilitation rather than bridge construction.

In the literature, Yeh (1998) have a study about quantity estimating by using regression models and neural network analysis. In the study Yeh (1998) prepared neural network models, linear and nonlinear regression models in order to estimate the total weight of steel of a steel building frame and total weight of reinforced steel of reinforced concrete building frame. 8 independent variables were included in the quantity models for steel building frame and 10 independent variables were included in the quantity models for reinforced concrete building frame. At the end of the analyses it was seen that for both quantity models for steel building frame and reinforced concrete building frame neural network gave better results than the nonlinear regression models and nonlinear regression models gave better results than the linear regression models.

In order to clarify the accuracy of the early cost estimates Trost and Oberlender (2003) made a study by evaluating 45 potential factors affecting the prediction accuracy of the cost estimates. These potential factors grouped and condensed into 11 with factor analysis. Then with the help of multivariate regression analysis, by evaluating the p values insignificant factors were eliminated and "estimate score procedure" which is used for learning of prediction capacity of the estimate was generated.

Sonmez (2004) conducted a survey on building project's conceptual cost estimation by both using regression and neural network models. In his study Sonmez (2004) collected 30 continuing care retirement community projects which were constructed in the United States. Initially 7 variables were included in the regression analysis and with the backward propagation linear regression model which had 5 significant variables were formed. Moreover, two neural network models with different hidden units were generated by using same 5 significant parameters used in the regression model. For all models, closeness of fit and prediction performance of models was calculated. In this study even so MAPE which can be described as the accuracy value for regression model is 11.1 smallest MAPE value is 11.7 for neural network models.

Kim, An and Kang (2004) made a comparison between regression analysis, neural networks and case based reasoning techniques by evaluating the cost estimation performance on the 530 residential building projects constructed by general contractors between 1997 and 2000 in Korea. During this study it was stated that 1 multiple regression 75 neural network and 1 case based reasoning models were developed. Moreover results showed that the best prediction performance MAER value was obtained by the one of the neural network models as 2.97%. MAER value for linear regression and case based reasoning model were 6.95% and 4.81 % respectively. In the light of these results even though one of the neural network models gave the best result between each other, case based reasoning analysis was selected as the most efficient for the time and accuracy because the neural network result was obtained after generating 75 models that consumed much more time with respect to other modeling techniques.

In their research Ugwu and Kumaraswamy (2004) tried to compare the accuracy of the cost of highway bridges obtained and examined. For this comparison neural networks were used. First of all, they collected data of 804 Highway Bridge constructed between 1942 and 2001 in Hong Kong. Then neural network models were built. By using 25% of the project data as training data, results were calculated. Because of the high error, some of the noisy data were eliminated and remodeling was done. New outcome was satisfactory. For example mean absolute error decreased from 55.02 to 8.03 and therefore accuracy of the estimation increased. This situation was clarified by Ugwu and Kumaraswamy (2004) that the data noise could affect the results of the study significantly.

Günaydın and Doğan (2004) used neural network method for conceptual cost estimation of the 4-8 storey buildings in Turkey. In order to make this study 30 building data were collected and randomly selected 20% of the data were extracted to use in the testing phase. Generated model included 8 input variable, 1 hidden layer with 4 hidden units and 1 output namely, cost of project. During the preparation phase of the model activation function was selected as hyperbolic tangent function which gives out between -1 and 1. In this study Günaydın and Doğan (2004) made a sensitivity analysis for the 8 dependent variables and calculated the percentage errors for six testing samples. For these testing samples absolute minimum, maximum and average percentage errors were obtained as 2, 12 and 7 respectively.

Li, Shen and Love (2005) also used regression analysis in their study and dependent variable was construction cost of the office buildings. Data were composed of 23 steel and 14 Reinforced concrete buildings located in Hong Kong. For reinforced concrete models randomly selected 7 data were used for cross validation. On the other hand 11 data were included in cross validation calculation. Regression models were obtained as a linear by using computer software SPSS12.

Li, Shen and Love (2005) obtained the percentages error range for reinforced concrete buildings between -4.11 percent and +2.74 percent. Moreover it was stated that percentage error range was from -6.65 percent to +2.74 percent for steel office buildings.

Another cost estimation study including multiple regression method was conducted by Lowe, Emsley and Harding (2006). In this study, 286 building data were compiled in the United Kingdom were used for performing backward and forward analyses and 6 models were created. Lowe's, Emsley's and Harding's (2006) best model results can be summarized as 0.661 for coefficient of determination value and 19.30 % for mean average percent error (MAPE)

Doğan, Arditi and Günaydın (2006) studied on modeling of early cost estimation of structural parts of the residential building. Totally 29 projects were collected and randomly selected 5 of them were separated as a testing data. During the training and testing processes different optimization methods, namely genetic algorithm, gradient descent and feature counting were used in order to compare the effects of the techniques. Finally it was realized that for this study models which used genetic algorithm method gave better results with respect to other two methods.

In their study Wang, Chiou and Juan (2008) dealt with case based reasoning method by using 293 historical buildings. Their aim was to make model predicting the restoration cost of the diachronic buildings. Models were generated by using inductive indexing and nearest neighbor methods. For evaluating the prediction accuracy 10 randomly selected cases were used. In order to evaluate the results, when actual cost, original budget and case based reasoning solution values were compared, reasoning model was seen as satisfactory.

Chou (2009) used case based reasoning method for early cost estimating for maintenance of pavement projects. In this study nearly 300 projects data of pavement maintenance were collected. By dividing the data into random 10 parts, necessary testing sets were obtained. In each fold by extracting testing samples, with the remaining data case based reasoning models were trained. Average mean absolute estimation error obtained for different cost items were calculated as between 23% and 30%.

Sonmez and Ontepeli (2009) have studied on the early cost estimation of urban railway project by using linear regression and neural network modeling. By means of backward elimination 10 models were generated. Moreover 2 neural network models were used in order to see the nonlinear relation between the dependent and independent variables. Models were created by using 13 urban railway projects in Turkey. The length of the railway systems was between 1.7 to 18.5 km, and projects were constructed between 1986 – 2005. During this analysis independent parameters were divided into two parts. One part consisted of parameters which were thought to have a primary effect on the dependent variable. The other part included secondary parameters. First, by developing five regression models, significant primary variables were found and by adding the secondary parameters one by one to the final model, significance of the secondary parameters were evaluated. Moreover, two neural network models were tried by using all 6 primary independent variables and only 2 primary input parameters obtained from the regression analyses. In the light of these MAPE is calculated for regression model as 35.2, neural network 1 as 49.8 and neural network 2 as 33.3.

Karancı (2010) performed a detailed study of conceptual cost estimation by not only preparing regression, neural network and case based reasoning models but also comparing models with each other with regards to their predictive variability and accuracy. In the study 41 mass housing projects in Turkey were examined. In regression model Karancı (2010) divided total cost of projects into the 6 parts. 12 parameters were used as independent variables. Closeness of fit was used as an initial measure for model comparison. Moreover, with the help of 3 fold cross validation technique MAPE and MSE values were calculated. In the neural network analyses and regression models same dependent variables were used. However two more neural models were generated by changing the number of the dependent variables and number of the hidden units in the hidden layer. In the case based reasoning model, type of feature matching was selected as fuzzy range 95% and for same 9 dependent variables 9 models were developed. Similarly in addition to the closeness of fit calculation by using 3 fold cross validation method MAPE and MSE values were calculated for comparing prediction performance of the models. When the total cost of the projects' prediction performance values (MAPE) were taken into account, it was seen that the best prediction was done by case based reasoning model as 12.62% and the worst one was neural network. However, it should be noted that the MAPE difference for the total cost prediction between neural network and case based reasoning models was only 1.27%.

In the literature there is a study about cost estimation of pre stressed concrete beam bridges by Kim, Kim and Kang (2009). In this study cost estimation is based on the quantity of the standard works. Kim, Kim and Kang (2009) collected 72 PSC beam bridge design documents from 25 road construction projects in order to make an analysis on the cost data of bridges. Bridge structure was divided into elements like superstructure, substructure etc. and these elements were divided into standard work items such as manufacturing PSC beam, cost of material, rebar fabrication etc. Then cost portions of the elements were analyzed. Standard works' cost allocations were also identified. Moreover, standard unit quantities for standard work items were calculated. Furthermore, some ratios were identified between standard work items. By means of this study Kim, Kim and Kang (2009) made a model working with only three parameters namely, length of span, total length and width of bridge. Generated model validated by 39 projects and the result showed less than 4.04% estimation error.

Kim (2011) applied case based reasoning technique in order to predict the approximate cost of railway bridge projects. 123 bridges constituted the studies database. At the end of the study a mean absolute error 11.9% was reached. Moreover, Kim and Hong (2012) studied on construction costs of rail road bridges by investigating 134 rail road bridges. First case based reasoning method were used for estimating the total construction cost then by generating regression models and integrating the coefficients to the case based model, error was decreased from 31.4 % to 15.2.

There are also several studies in the literature other than cost estimation purposes, using linear regression, neural network and case based reasoning methods.

For example, Irish, Barrett, Malina and Charbeneau (1998) studied on prediction of the highway storm water loads by dividing the total loads into the 11 constituents. For all of the constituents, regression models were developed with the data collected from Austin. The size of the data varied in each model depending on the availability of the necessary independent variable. The maximum size of data that used in the model was 423. With the help of these models, variables which affected the constituents were identified.

Neural network and regression analyses were used in Wang's and Elhag's (2007) research for modeling risk of bridges. Independent variables were constituted of 4 risky criteria. Risk modeling was initiated with rating the 4 variables as "none, low-risk, medium-risk and high-risk" then transforming the qualitative ratings into numbers as 0, 1, 2 and 3 respectively. After quantitative data obtained, analyses were done on 66 bridge structure. In this study MAPE value was obtained as 10.21 for artificial neural network model and 17.03 for the multiple linear regression analysis.

Arditi and Tokdemir (1999) examined the court cases which were collected from Illinois circuit. Total data divided into two sections. First section composed of 102 cases was used in training and second one consisted of 12 data for testing. 43 input variables listed and then by eliminating the insignificant ones, 38 significant variables were reached. By performing case based reasoning and neural network techniques with both 43 and 38 features models were developed. At the end of this study prediction percentage was 83% for case based reasoning model and 67% was for artificial neural network model.

In the literature apart from the examples given above, case based reasoning method was strongly preferred in medicine. Some of the studies were summarized by Holt, Bichindaritz and Schmidt (2006).

As a summary it can be said that first regression analyses were strongly used for cost estimations in the literature, and then researchers started to include neural networks modeling in their studies in order to mimic the brain and to model the nonlinearity. Moreover, recently case based reasoning method was also preferred in order to make cost estimations. However, it can be stated that in detailed research studies regression was used as a benchmark for evaluation of the other methods.

The existing pre-design cost estimation models mainly focus on modeling of costs. There is very limited research on modeling of quantities. The main objective of this thesis is to fill this gap in the literature by focusing on modeling of the quantities. Within this context regression analysis neural networks and case based models will be explored to achieve accurate conceptual quantity take-off models for the bridge projects.

CHAPTER 3

REGRESSION ANALYSIS, NEURAL NETWORK AND CASE BASED REASONING MODELS

3.1. Multiple Linear Regression Method

Regression analyses are used for explaining the variability of the dependent variable by independent variables. While generating the regression models, output variable is defined by the input variables with factors. If there is more than one dependent variable and the model is linear, these are called multiple linear regression models. As an example multiple linear regression equation with "n" parameters is given below;

$$y_i = c + a_1x_{1i} + a_2x_{2i} + \ldots a_ix_{ni} + e_i \tag{1}$$

Where;
y is dependent variable
c is constant number
i is the case number
$x_{1, \ldots} x_n$ are independent varibles
$a_{1, \ldots} a_n$ are coefficients for related dependent variables
e is the error value

In this concept objective is minimizing the SSE which is defined as sum of square error. SSE can easily be found with the help of following formula.

$$SSE = \sum_{i=1}^{m}(\hat{y}_i - y_i)^2 = \sum_{i=1}^{m} e_i^2 \quad (2)$$

Where $\hat{y}_i$ is the predicted and y_i is the actual dependent parameter value for case i out of total m cases.

In regression analyses while developing a model two parameters have great importance. One of them is the level of significance (P value), which shows the significance of the variables' contribution in the model. The other parameter is coefficient of determination whose symbol is R^2. This parameter indicates how well the model fits to the data. Therefore in this study these two parameters were taken into consideration for developing regression models.

3.2. Neural Network Method

Artificial neural network is a modeling technique which simulates the brain structure in the human body. This system composed of input layer, hidden layers and output layer like shown below.

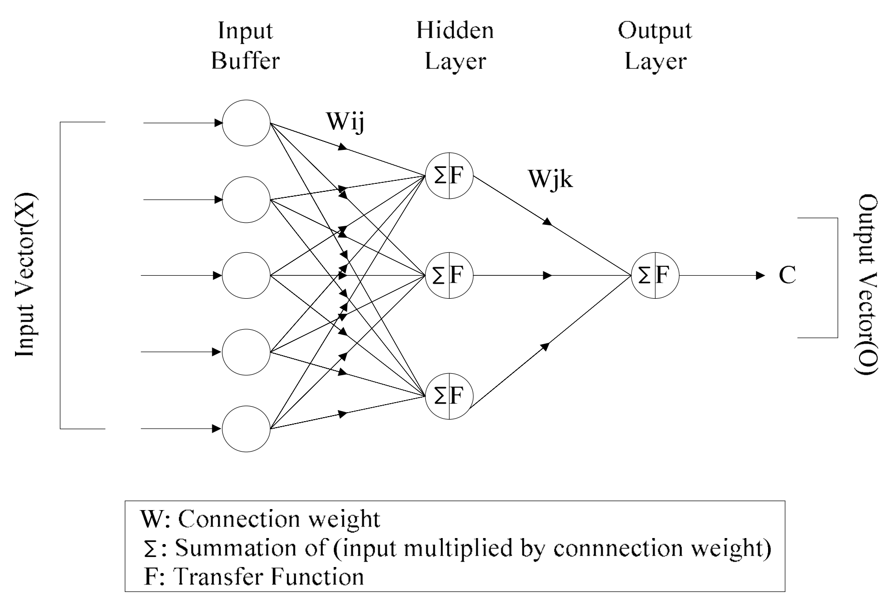

Figure 1 Neural Network Model

As can be seen from the Figure 1, first of all input parameters are given to the system and then model multiplies the input parameters by the related weights. Next, each hidden unit in the hidden layer receives the weighted sum of inputs from all sources. When the data is received, transfer function in the hidden unit is activated and this activation function gives output as an input parameter for the next step. This procedure continues until the output of the output layer is obtained. In this procedure there are several functions used in the hidden units. These can be linear, sigmoid or step function. In order to clarify the functions, functions' graphics are shown below in Figure 2.

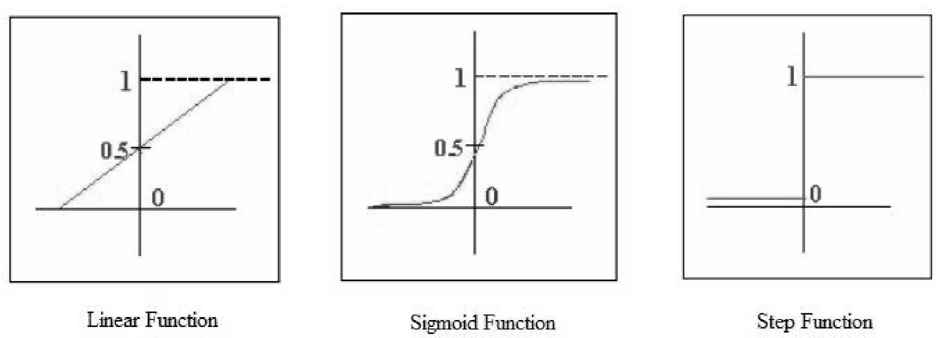

Figure 2 Transfer Functions

Sigmoid function which was used in this study is the most well-liked function type in neural network models in the literature. This function's mathematical expression is shown below.

$$f(x) = \frac{1}{(1+e^{-x})} \tag{3}$$

After the number of the hidden layer, numbers of the hidden units in each layer and functions to be used in the units are decided; weights in the model are determined by minimizing the estimation error of the neural network model.

3.3. Case Based Reasoning Method

CBR is another modeling method used for making predictions. In this method first of all a database called case base is constituted with the actual cases. Then retrieval criteria are defined. After that by means of the case based reasoning technique, most similar cases to the target case are retrieved. Briefly, CBR method checks the data stored in the predefined case base with respect to pre-defined similarity concept. Moreover with this method, after the retrieval as a solution to current case, adaptation can be done by the user. This procedure which is shown in Figure 3 was expressed by Yau and Yang (1998).

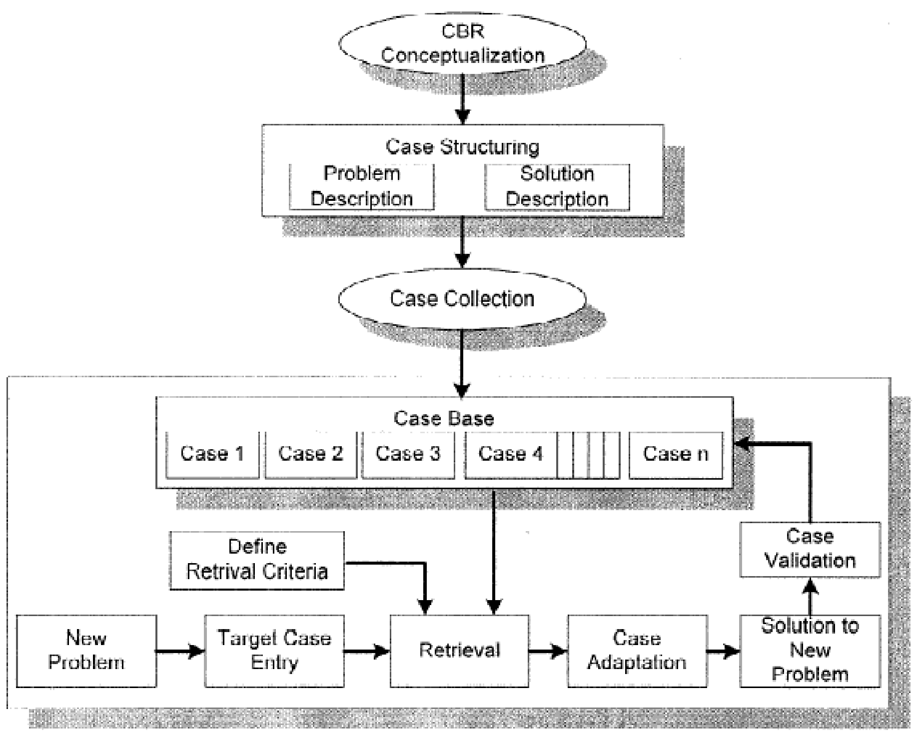

Figure 3 General CBR Development Procedure (Yau and Yang, 1998)

This system was also explained with four steps by Özorhon (2004) as follows.

1) Acquisition of Cases
2) Indexing
3) Retrieval
4) Adaptation

3.3.1. Acquisition of Cases

Acquisition of cases step is the procedure of collecting old cases. In this procedure it is important that the case base structure is prepared in manageable structure because it is the direct resource of the retrieval procedure (Gupta, 1994).

3.3.2. Indexing

Indexing is the one of the most important step in the case based reasoning method which provides the collection of the suitable cases at the right times. In fact it is a coding problem which should give the necessary information to the case and with the help of this information related cases can be referenced during the retrieval step (Chua et al., 2001). Indicators assigned during this process should include the idea where the case can be used. Therefore, indices should be both abstract and concrete in order to widen the future use and be recognizable in future (Özorhon, 2004). In this respect, Kolodner (1993) thinks that human beings are more capable of deciding indices than algorithms, so that Kolodner (1993) purposed that the indicators should be selected by hand for practical applications.

Case based reasoning method works with a group of indices to seek and select the cases which are close to the existing situation. Indexing of the cases can be made with the three main approaches which are nearest neighbor, inductive reasoning and knowledge guided indexing (Barletta 1991, cited in Gupta, 1994). Most of the time systems use all of these methods in the same modeling process (Özorhon, 2004).

* When the nearest neighbor technique is preferred, in the retrieval step cases which have most closely attributes to the existing problem are selected. Moreover if all of the weights of the features defined equally, cases which have more matching feature are selected by the CBR model (Gupta, 1994).

* The scope of the inductive reasoning is to make a generalization for decision rules by using old cases. In this approach unlike nearest neighbor technique, cases are selected from database in accordance with the feature importance in the model and the retrieval decision is done by heuristically by an algorithm (Gupta 1994).

* Knowledge guided indexing is used when domain knowledge is preferred to determine the similarities between the old and the current case. In order to use this method, “if – then” rules are activated. However, because it is difficult to apply this method and get the expositive knowledge, knowledge guided indexing is used to improve the other two indexing techniques rather than the independent usage (Gupta 1994).

3.3.3. Retrieval

Retrieval is the procedure that most similar cases were obtained with the help of the case similarity indexes. Obtaining of the relevant cases depends on good indexing of the cases which choose a suitable set of indices. According to the pre-defined similarity functions, CBR model examines the relevancy of all old case in the case base with respect to current case and retrieves the most similar cases to the user (Yau and Yang, 1998).

Case similarity indexes are specific for each case and these indexes depend also on the current case. When the current case changed by the user new case similarity indexes are developed by the model. High value of the similarity index shows the high relevancy between the case worked on and the target case (Yau and Yang, 1998).

3.3.4. Adaptation

When the most relevant case or cases are retrieved, case based reasoning model should adapt the result for the requirements of the problem case. Adaptation rules were categorized by (Kolodner, 1993) in two groups. First one was defined as structural adaptation where rules defined for adaptation are directly applied to the solution and the second one was defined as derivational adaptation where algorithms, methods and rules are reused for redeveloping a solution.

Yau and Yang (1998) described the adaptation procedure as a user defined system which directly calculate, capture or manipulate the own values of the retrieved cases to produce a remedy to the new problem.

After the case goes through the adaptation process, next and the final step is the inclusion of the adapted case into the database in order to use in the future. With the help of this cycle CBR model's database is increased and following problems take the advantage of the expanding case base. According to the solution evaluation if the results are satisfactory system decides to include the solution in database otherwise failure explanation is provided to the system library (Gupta, 1994).

CHAPTER 4

DESCRIPTION OF DATA

For this study, first of all, design documents of several highway bridge projects in Turkey, and quantity take off and cost estimation results prepared by an engineering firm in Turkey were compiled. Owner of the projects is Republic of Turkey General Directorate of Railways, Ports and Airports Constructions (RPA). After the general investigation on the data set, noisy data which can affect the research results were searched. By noisy data description it was intended to mean the data which is not trusted. When a set of data is examined, sometimes it can be encountered with some information which can not be possible in real life. In this study no data was eliminated because of the irrationality. However, in order to be consistent in the research, data which is not related with this study were eliminated. For example one bridge was eliminated because it is not single span bridge. Shortly, by means of the elimination of the unrelated data, final database was obtained. Number of the bridges examined for this study and project names for which the bridges were designed are summarized in Table 1 below. The data includes the both cost and quantity take off items value for each bridge.

Table 1 Composition of the Database

Project Name	Number of Bridges
Tokat-Turhal Railway	2
Kırşehir-Yerköy Railway	5
Ankara-İzmir Railway, Polatlı-Afyon Part	24
Turkey-Georgia (Kars- Tbilisi)	9
TOTAL	**40**

Moreover, general properties of the bridges used in this research are listed below.

1) Highway bridges constructed as an overcrossing in a railway project.
2) Single span, length with 20, 25 or 30m
3) Width of bridge between 9.5m and 14.1m
4) Foundation above the ground water level (This item affects the unit cost of the earth works. It may also affect the design of foundation)
5) Soil profile consists of stiff clay or deep cohesionless conditions
6) Foundation without pile
7) Pre-stressed concrete beam (I Section – BS40)
8) No skewness: Abutment Axes of the bridge are perpendicular to the axis of the highway road.
9) Bridges were designed according to AASHTO 2002

Bridges used in this study are structurally composed of pre-stressed concrete beams and abutments. Moreover in order to provide the rigidity, floor beams were also included in the design phase. General longitudinal and cross section views of the bridges are shown Figure 4 and Figure 5 respectively.

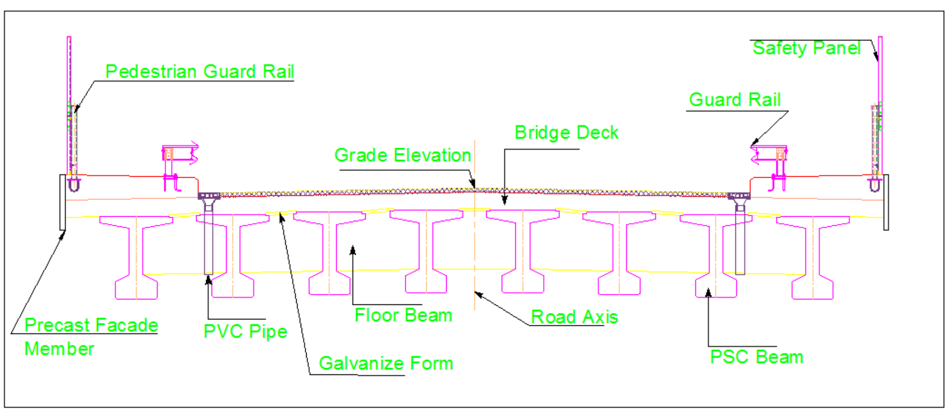

Figure 4 Typical Cross Section View

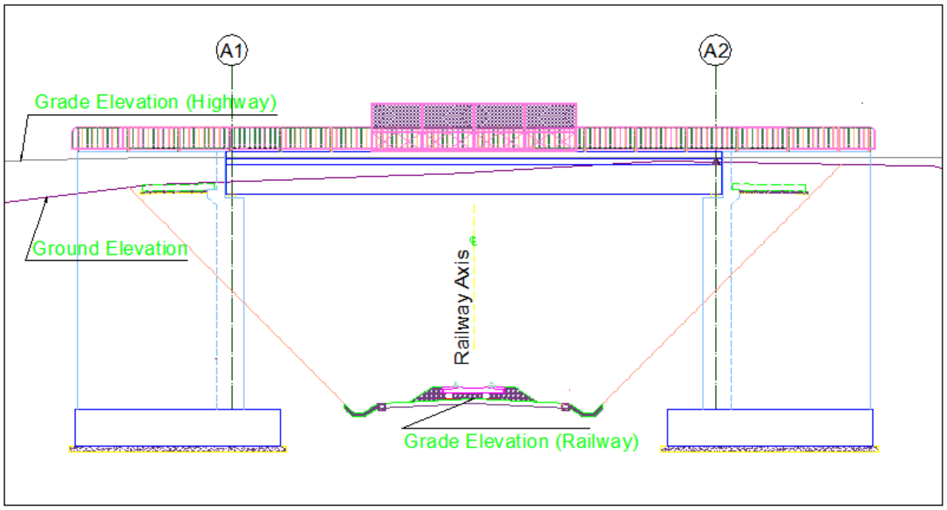

Figure 5 Typical Longitudinal Section View

As can be seen from Figure 4 and Figure 5, typical bridges consist of structural part and supplementary items such as PVC rainwater drop-pipe, guard rail, pedestrian guard rail etc. When the design, quantity take off and cost estimation documents prepared by the design office were investigated it was seen that the highway bridge construction had been divided into 19 items for estimating the final cost of the bridges after design phase. It was also seen that by the design office, a system had been generated by giving item numbers to the quantity take of items for both railway and highway bridges. Inherently, although some of the items were common, there were several items that are used in railway bridges but not in highway ones or vice versa. In this system unit costs had been calculated with respect to this system items by the design office engineers. It was understood that during this calculations engineers in the design office had integrated some of the items to another item. For example, in the S-102.3 (Concrete BS25) item shown in Table 2, formwork which is necessary for execution of the item had been included. Unit cost for the items also included the labor costs to accomplish the items and the bridge construction. The 19 items are shown with their item number, description and their unit in Table 2.

Table 2 Quantity Take off Items

Quantity Take off Item No:	Quantity Take off Item Description	Unit
S-101	Earthwork for Foundations	m^3
S-102.1	Concrete BS14	m^3
S-102.3	Concrete BS25	m^3
S-103.1-a	Reinforcing Bar St III (≤ Ø12)	Ton
S-103.1-b	Reinforcing Bar St III (> Ø12)	Ton
S-103.3	Prestressing Steel	Ton
S-104	Placed Stone Filling	m^3
S-109	Waterproofing for Bridge Deck	m^2
S-111(Not included in this study)	Pedestrian Guard Rail	ton
S-112	Pre-stressed Concrete Beam Execution and Installation	m^3
S-113	Elastomeric Bearing	dm^3
S-114 (Not included in this study)	Expansion Joint Execution (80 mm)	m
S-115 (Not included in this study)	Grating Execution	kg
S-116 (Not included in this study)	PVC Rainwater Down-Pipe	m
S-117	PVC Service Pipe	m
S-118	Galvanize Form	m^2
S-120 (Not included in this study)	Safety Panel	m
S-121 (Not included in this study)	Precast Facade Member	m^2
S-123 (Not included in this study)	Guard Rail	m

In order not to be confused in the analyses, quantity take off item numbers were used as indicated in the cost estimation prepared by the design office. There are some item numbers between S-101 and S-123 but not listed the Table 2 above. This situation occurs because the missing items are not related with the selected bridge data. Missing items can be related to only railway bridges as described earlier or to elements which the highway bridges in this study do not have. For example, because the bridges that analyzed in this research do not have any piling the list does not also include related entry. Moreover, seven items in the Table 2 namely, S-111 (Pedestrian Guard Rail), S-114 (Expansion Joint Execution 80 mm), S115 (Grating Execution), S116 (PVC Rainwater Down-Pipe), S-120 (Safety Panel), S-121 (Precast Facade Member) and S-123 (Guard Rail) were not included in this study. The reason behind this is that quantity of the S-120 (Safety Panel) is 17m in all the cases. Moreover, quantity of the S-121 (Precast Facade Member) item is only multiplication of height and length of the member, but height of this component is variable with respect to the design engineer's choice. In other words, if the height of the member is known it can be calculated with 100% accuracy. Furthermore, S-111, S-114, S115, S116, and S-123 can be directly calculated by using related parameters. Therefore for the dependent variables defined above no model was generated. After decision of the quantity take off items which would be estimated, 12 dependent variables were obtained. These are the items listed in Table 2 except S-111, S-114, S115, S116, S-120, S-121 and S-123.

CHAPTER 5

LINEAR REGRESSION

5.1. Determination of Parameters

In this stage, first of all by brainstorming, the question that which parameters should be used in this analysis was searched. Several properties of the bridges were selected as an input parameter, and then with these parameters, trial models were developed for each of the quantity take off items. During this phase it was considered that although models gave good estimation results with respect to the after design estimations, several models did not produced reasonable coefficients. In other words, expected signs for the coefficients were not obtained for some of the parameters. When the situation was researched, it was understood that some of the quantity take off data were miscalculated by the design engineers. These data were corrected by means of the detailed design documents. Moreover, after these corrections, revisions on the parameters were done. Because the first decided parameters were too much and some of them were interconnected, by considering the general fact affecting the quantity take off items, interrelated parameters were combined and the number of the independent variables were decreased as much as possible in order to have a parsimonious model. A parsimonious model can be defined as a model fitting the data adequately without using any unneeded variables. Parsimony concept is very crucial during modeling because in operation parsimonious models generally give better results (Pankratz, 1983). Furthermore, it was seen that if the independent variables are excessive or complicated, interpretation of the variables can be hard and confusion can occur. In order to obey the principle of the parsimony, parameters to be used in the models were updated and finally by means of the investigations on bridge structure and cost relations, 7 parameters were decided to be included in this study. These 7 Parameters are:

1) Length of Span (Var1)
2) Width of Bridge (Var2)
3) A_0 Value (Var3)
4) Distance Between Grade Elevation of Highway and Railway (Var4)
5) Average Excavation Height (Var5)
6) Maximum Abutment Height (Var6)
7) Average Abutment Height (Var7)

Except A_0 value, all of the independent variables were measured or calculated in meter. In order to be precise in the analyses all these parameters listed above were measured on the detailed design documents including detailed drawings and detailed structural analysis reports. To accomplish the consistency on measured value of the dependent parameters, standardization for each value measurement was defined. In order to clarify the details of measurements, standardizations used during the measurements are shown below with the figures and explanations

5.1.1. Length of Span (Var1)

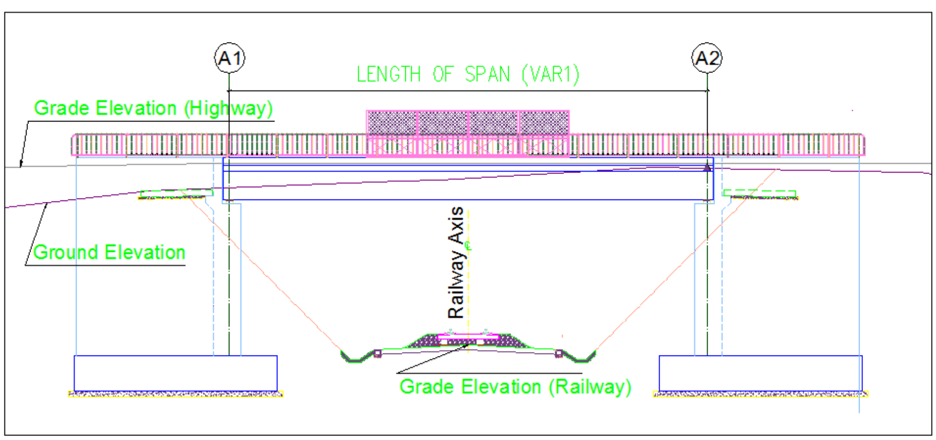

Figure 6 Illustration of Var1

As can be seen from the Figure 6 length of span (Var1) was defined as the distance between two abutment axes. In other words, variable mentioned above is the length of the pre-stressed concrete beams design calculation length.

5.1.2. Width of Bridge (Var2)

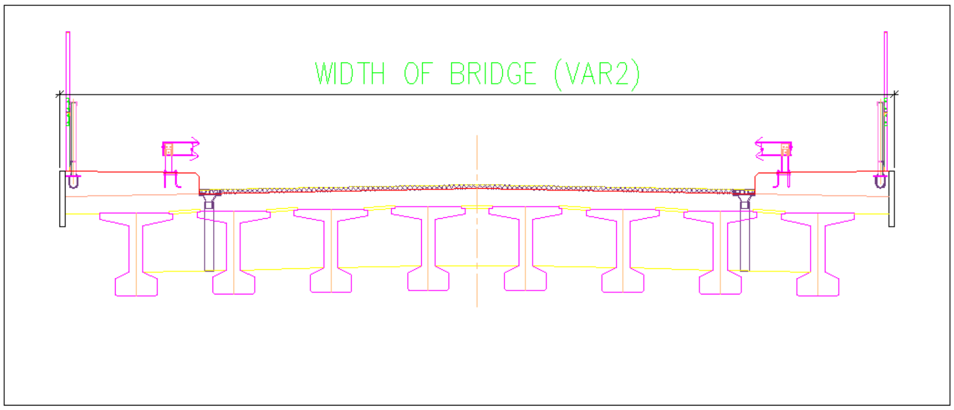

Figure 7 Illustration of Var2

Width of bridge was defined as the distance between two edge points in the cross section including the precast façade member thickness as shown in Figure 7.

5.1.3. A_0 Value (Var3)

A_0 Value is the seismic coefficient used for earthquake analyses which can be determined by Table 3 and Figure 8, if there is no detailed study on this

Table 3 Seismic Coefficient Values (TEC 2007)

Earthquake Region	A_0
1	0.4
2	0.3
3	0.2
4	0.1

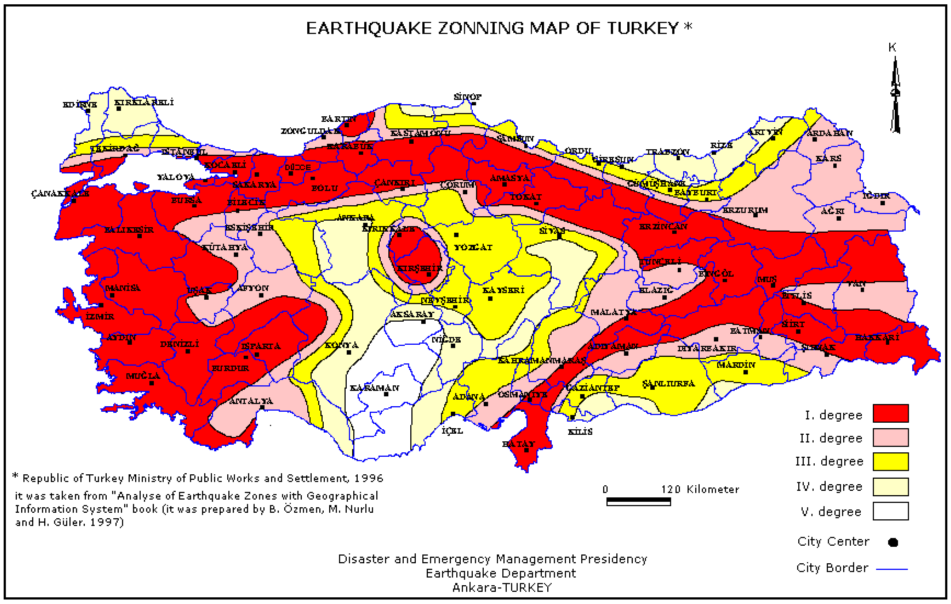

Figure 8 Earthquake Zonning Map of Turkey(Özmen, Nurlu and Güler 1997)

5.1.4. Distance Between Grade Elevation of Highway and Railway (Var4)

Var4 was measured as the elevation difference of grade elevation of highway and grade elevation of railway on railway axis on longitudinal section view of bridge. Measured distance as a Var4 can be realized from the Figure 6.

5.1.5. Average Excavation Height (Var5)

In order to calculate the average height of excavation variable, abutments' foundations and the orientation of the abutment axes were standardized. For example, even though the foundation length may change from bridge to bridge it was considered as 8.5m after examination of the whole data. Part behind the axis was decided as 6.25m and part in front of the axis was taken as 2.25m. With the help of this assumption and examination of the ground profile and cutting profile, average height of excavation were determined.

Steps used for deciding the Var5 can be listed as follows:

1) Offsetting the axis of abutments forward 2.25m on longitudinal section

2) Offsetting the axis of abutments backward 6.25m on longitudinal section

3) Calculating the average elevation difference between ground profile and grade elevation of the railway between newly defined lines for each abutments. If there is a cutting between these predefined points, cutting edge between the specified points is used as ground profile.

4) Finally, Var5 is calculated by taking average of the value computed for each abutment, explained in step 3

The procedure can be understood by means of Figure 9 below.

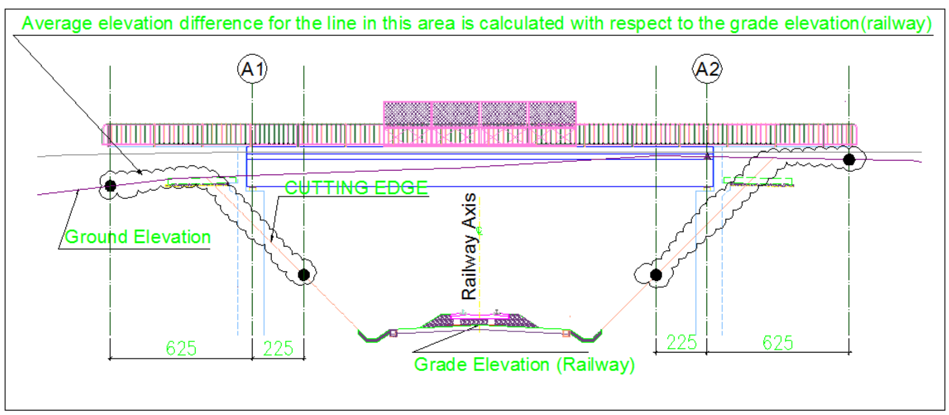

Figure 9 Illustration of Var5

5.1.6. Maximum Abutment Height (Var6)

Actually this variable includes two properties of the bridge design. One of them is distance between grade elevation on the axis of abutment and ground elevation on the axis of the abutment. Second one is the angle of the ground profile or cutting edge. In order to include the angle of ground profile or cutting edge in the analyses, standardization used in definition of the Var5 is also preferred in determination of Var6. Investigations on the detailed showed that ground profile or cutting edge affects the depth of foundations, by this means quantities and cost of project. Therefore, abutment height is calculated by both including the distance between the grade elevation on the abutment axis and ground elevation (or cutting edge if it is critical), and distance between the ground or cutting edge elevation on the axis of abutment and on the axis offset forward 2.25m. Shortly Var6 can be measured by performing the following phases.

1) Offsetting the axis of abutments forward 2.25m on longitudinal section
2) Investigating which one of ground profile or cutting edge elevation on the new axis defined in the first phase is critical
3) Measuring distance between the grade elevation on the axis of abutment and point which is the junction of the newly defined axis and critical ground profile.
4) Steps up to phase 4 are done for both two abutments and value higher is selected as maximum abutment height.

Height of abutment which is defined for this study is show in Figure 10 below.

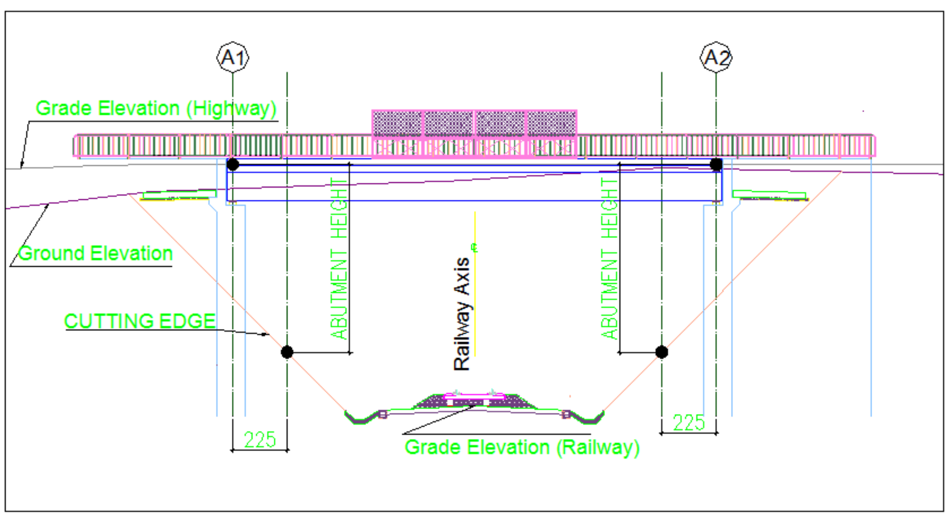

Figure 10 Illustration of Var6

5.1.7. Average Abutment Height (Var7)

Average abutment height was also defined as an independent variable in order to realize that whether dependent variable is related with maximum or average height of abutment in project. Var7 is calculated by implementing first three steps described in section 5.1.7 for both two abutments and taking average of them.

5.2. Preparation of Regression Models

In this stage, for regression model generation SPSS 18 software was used. During the modeling phase first, all of the 7 independent variables defined in section 5.1 and 12 dependent variables described in chapter 4 entered in the software as database. After that from the "Analyze" menu "Linear Regression" was selected. On the linear regression window, whose sample display is shown in Figure 11, for each quantity take off model, related dependent and independent variables were defined. Moreover, drop dawn menu which is called "method" was selected as backward and from the options menu, removal criteria was set to 0.1. This means that linear regression models were developed by using backward elimination method by removing the independent variables whose significance (p value) was greater than 0.1.

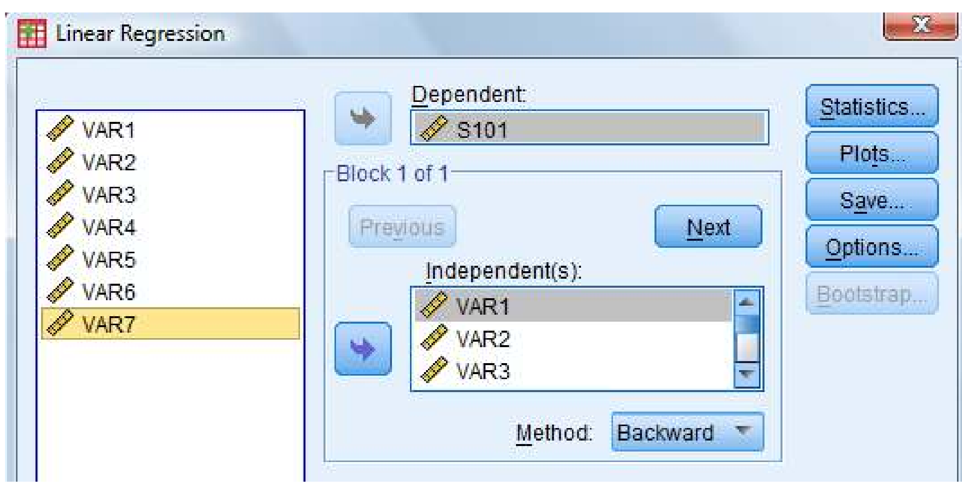

Figure 11 Linear Regression Window

Subsequent to the analyses, output documents including coefficient of determination (R^2) and p values of the coefficients for each model generated during the backward elimination and final models were obtained with the help of SPSS. Moreover, after the investigations of these important values of regression analysis, main models were selected.

During the preparation phase of the models, because the logics behind the work items are different from each other, in other words because the reasons affecting the each quantity take off value may vary from item to item, independent variables only considered as related with the independent variable for each work item were selected in the SPSS software. Starting parameters used in each model are shown in the Table 4

Table 4 Starting Parameters

Quantity Take off Item No:	Starting Parameters for Models
S-101	Var1, Var2, Var3, Var4, Var5, Var6, Var7
S-102.1	Var1, Var2, Var3, Var4, Var6, Var7
S-102.3	Var1, Var2, Var3, Var4, Var6, Var7
S-103.1-a	Var1, Var2, Var3, Var4, Var6, Var7
S-103.1-b	Var1, Var2, Var3, Var4, Var6, Var7
S-103.3	Var1, Var2
S-104	Var1, Var2, Var3, Var4, Var6, Var7
S-109	Var1, Var2
S-112	Var1, Var2
S-113	Var1, Var2, Var3
S-117	Var1, Var2
S-118	Var1, Var2

With the help of the regression analyses, to achieve parsimony concept, insignificant parameters which did not have enough impact on the dependent variables were eliminated by backward elimination. In Table 5 variables which constituted final models for each quantity take off item, in other words significant parameters used as final independent variables in the models are shown.

Table 5 Parameters Constituting Models

Quantity Take off Item No:	Parameters Constituting Models
S-101	Var2, Var3, Var5
S-102.1	Var2, Var3, Var6
S-102.3	Var1, Var2, Var4, Var6
S-103.1-a	Var1, Var2, Var3
S-103.1-b	Var2, Var3, Var4, Var6
S-103.3	Var1, Var2
S-104	Var2, Var4, Var6
S-109	Var1, Var2
S-112	Var1, Var2
S-113	Var1, Var2
S-117	Var2
S-118	Var1, Var2

As can be seen from Table 5, var7 (average abutment height) is not present in the parameters which constitute the models. It means that var7 did not have significant contribution in any of the models. Therefore it was eliminated during the backward elimination process. On the other hand from the Table 5 it is seen that var6 (maximum abutment height) is one of the significant parameter for dependent variables S-102.1(Concrete BS14), S-102.3(Concrete BS25), S-103.1-b (Reinforcing

Bar St III (> Ø12)), and S-104 (Placed Stone Filling). When this situation is interpreted, it can be understood that maximum abutment height variable which was defined in section 5.1.6 is enough to link the relationship between work items mentioned above and height of abutments. Moreover, interpretations on the var6 and var7 can be clarified that in the design phase of the project, abutments were designed by considering the critical one. The described case was examined on the detailed design documents and it was considered that the designers generally prefer examining the higher distance between grade elevation (highway) and ground elevation on the abutment axis, because of the both structural and efficiency reasons; design of the abutments mostly depends on only the critical abutment.

To show the significance levels of the parameters selected in final model for each quantity take off part and how well the models fits to the data , P values of the coefficients and R^2 values are given in Table 6.

Table 6 P Values of Coefficients and R^2 Values for Models

Model	R^2 Value	P Value					
		Var1	Var2	Var3	Var4	Var5	Var6
S-101	0.869	-	0.000	0.013	-	0.000	-
S-102.1	0.873	-	0.000	0.001	-	-	0.000
S-102.3	0.643	0.011	0.000	-	0.012	-	0.013
S-103.1-a	0.968	0.000	0.000	0.000	-	-	-
S-103.1-b	0.734	-	0.062	0.000	0.003	-	0.098
S-103.3	0.925	0.000	0.000	-	-	-	-
S-104	0.792	-	0.000	-	0.010	-	0.000
S-109	0.870	0.000	0.000	-	-	-	-
S-112	0.836	0.000	0.000	-	-	-	-
S-113	0.675	0.000	0.000	-	-	-	-
S-117	0.456	-	0.000	-	-	-	-
S-118	0.855	0.000	0.000	-	-	-	-

As can be seen from Table 6 coefficient of determination values of the models were between 0.643 and 0.968 except that S-117. R^2 value for S-117 is 0.456. Only for this model it seems fitting of the model to the data is not good but still it can be interpreted as acceptable. On the other hand other 11 models' R^2 value results revealed that the fitting of these models to related data are in adequate level.

When the P values were examined it can be seen that in order to predict the quantity of earthwork for foundations item there are three parameters needed namely, width of bridge, A_0 value and average excavation height. According to these analyses it seems that concrete BS14 item depends on width of bridge, A_0 value and maximum abutment height but concrete BS25 item depends on length of span, width of bridge, distance between grade elevation of highway and railway and maximum abutment height. By means of this information it can be said that because the place of uses for BS14 and BS25 are different from each other, parameters which affect the quantity results are also different. Same logic is valid for reinforcing steel items, S-103.1-a and S103.1-b. For example, because the deck of bridge includes reinforcement whose diameter is equal to 12mm but does not include reinforcement whose diameter is greater than 12 mm, although length of span is significant parameter for S-103.1-a, it does not have significant contribution in model for S-103.1-b. According to the analyses results, width of bridge, distance between grade elevation of highway and railway and maximum abutment height are the independent variables which constitute the final model for placed stone filling item. Moreover with regards to pre-stressing steel item, only two parameters namely length of span and width of bridge are the final model parameters. Other than these, var1 (length of span) is also present as a significant parameter for models whose dependent parameters' descriptions are waterproofing for bridge deck, pedestrian guard rail, pre-stressed concrete beam execution and installation, elastomeric bearing, grating execution, pvc rainwater down-pipe, galvanize form and guard rail. Apart from the description above Var2 (Width of bridge) was saved by the regression analysis in the models whose dependent parameters' descriptions are pre-stressed concrete beam execution and installation, elastomeric bearing, expansion joint execution, pvc service pipe and galvanize form. Regression coefficients of the final models for each quantity take off item are shown in Table 7.

Table 7 Final Models for Each Quantity Take off Item

Model	Regression Coefficients
S-101	-572.30 + 83.56 x Var2 + 1008.82 x Var3 + 140.36 x Var5
S-102.1	-3.43 +3.83 x Var2 + 21.35 x Var3 + 1.37 x Var6
S-102.3	-1534.58 + 28.21 x Var1 + 61.30 Var2 + 75.51 x Var4 + 36.47 x Var6
S-103.1-a	-50.89 + 1.61 x Var1 + 3.03 x Var2 + 12.82 x Var3
S-103.1-b	-104.99 + 2.01 x Var2 + 146.24 x Var3 + 9.12 x Var4 + 2.32 x Var6
S-103.3	-16.49 + 0.66 x Var1 + 0.42 x Var2
S-104	-955.31 + 62.14 x Var2 + 47.05 x Var4 + 50.27 x Var6
S-109	-304.85 + 12.82 x Var1 + 17.09 x Var2
S-112	-189.64 + 8.02 x Var1 + 7.74 x Var2
S-113	-99.78 + 3.83 x Var1 + 5.31 x Var2
S-117	1.55 + 10.69 Var2
S-118	-1118.95 + 33.80 x Var1 + 76.68 x Var2

5.3. Validation of Regression Models

By evaluating the R^2 value results it can be understood that how well the model fits to the data however, good fit of model does not always mean that the model give accurate predictions (Sonmez, 2008). Therefore, in this study in order to check the prediction performance of the final regression models, cross validation method was used.

In the validation stage three folds cross validation was used. It means that randomly selected one third of the data is not used during modeling but after models are developed, predictions are obtained for the selected one third of the data. Then same procedure is applied for another one third of the whole data which are not included the previously selected data. After that finally same method is implemented to the last one third of the data.

In this stage to achieve random selection principle first of all random number from 1 to 40 were generated in the Microsoft Office Excel (2007). Then projects were put in order according to the randomly generated numbers. Because the number of the total projects is 40, equal division could not be achieved. Total database was divided in to three groups that consist of 13, 13 and 14 projects respectively. That is to say, after putting order with respect to the randomly generated numbers, initially first 13 projects, then second 13 and finally last 14 projects were used for evaluating the prediction performance of the models which were generated by including the data not used in the prediction stage.

Mean average percent error (MAPE) technique was used for assessing the prediction performance of each model. Formulation implemented to calculate the MAPE value is shown in equation (4)

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|actual_i - predicted_i|}{actual_i} \ x \ 100 \tag{4}$$

Validation results achieved by applying the 3 – fold cross validation technique for the prediction performance of all linear conceptual quantity take off models are shown in Table 8.

Table 8 Validation Results for Linear Regression Models

Model	Quantity Take off Item Description	MAPE
S-101	Earthwork for Foundations	9.23
S-102.1	Concrete BS14	3.02
S-102.3	Concrete BS25	9.15
S-103.1-a	Reinforcing Bar St III (≤ Ø12)	3.96
S-103.1-b	Reinforcing Bar St III (> Ø12)	11.04
S-103.3	Pre-stressing Steel	7.82
S-104	Placed Stone Filling	7.39
S-109	Waterproofing for Bridge Deck	3.87
S-112	Pre-stressed Concrete Beam Execution and Installation	4.95
S-113	Elastomeric Bearing	6.70
S-117	PVC Service Pipe	8.29
S-118	Galvanize Form	6.73

As can be seen from Table 8, MAPE values are between 3.02 and 11.04. Highest MAPE value which is 11.04 was obtained from the model whose dependent variables description is reinforcing bar St III (> Ø12). Furthermore, earthwork for foundations, concrete BS25 are the other two works items whose model prediction performance results are around 9 %. In the light of these it can be said that because before detailed design stage, height of the abutments can not be determined exactly, works items including the var5 and var6 have relatively higher MAPE value than remaining models. However as a result from Table 8 it can be interpreted that the all models have sufficient prediction performance for this study.

CHAPTER 6

NEURAL NETWORK MODELS

After the linear regression analyses, the study continued with generation of neural network models to search for non-linear relations between variables and quantity take-off items. Therefore in this chapter preparation of neural networks models and validation of them will be discussed.

6.1. Preparation of Neural Network Models

In preparation stage of the neural network modeling SPSS 18 software was used. The data described in chapter 4 constituted the database of neural network models. Moreover, variables which were determined by backward elimination process in regression analyses and used in the final regression models were included in the neural network modeling process as input variables. General procedure followed in the SPSS 18 to generate neural networks models are as follows.

After the data was imported to the software, from the "Analyze" menu "Multilayer Perceptron" was selected. In "Variables" window for each model, one quantity take off item was entered in "Dependent Variables" part. Then related independent variables were included in "Covariates" section. "Factors" section was not preferred because it is specified for categorical variables which are also called discrete variables. Although Var3 (A_0) variable seems to be categorical, because by means of detailed calculations A_0 can get value different from the ones specified Table 3. Moreover in the same window "Rescaling of Covariates" part was defined as "Normalized" (See Figure 12). Normalization can be described as scaling of the values in the training part of the database between 0 and 1. In other words minimum value of the data set gets the value 0 and the maximum one become 1.

Normalization procedure is applied using the formula (5) below:

$$X_{Normalized} = (X - Min) / (Max - Min) \quad (5)$$

Where;

X : Case value of the parameter to be normalized

Min : Minimum value for related parameter in the training data

Max: Maximum value for related parameter in the training data

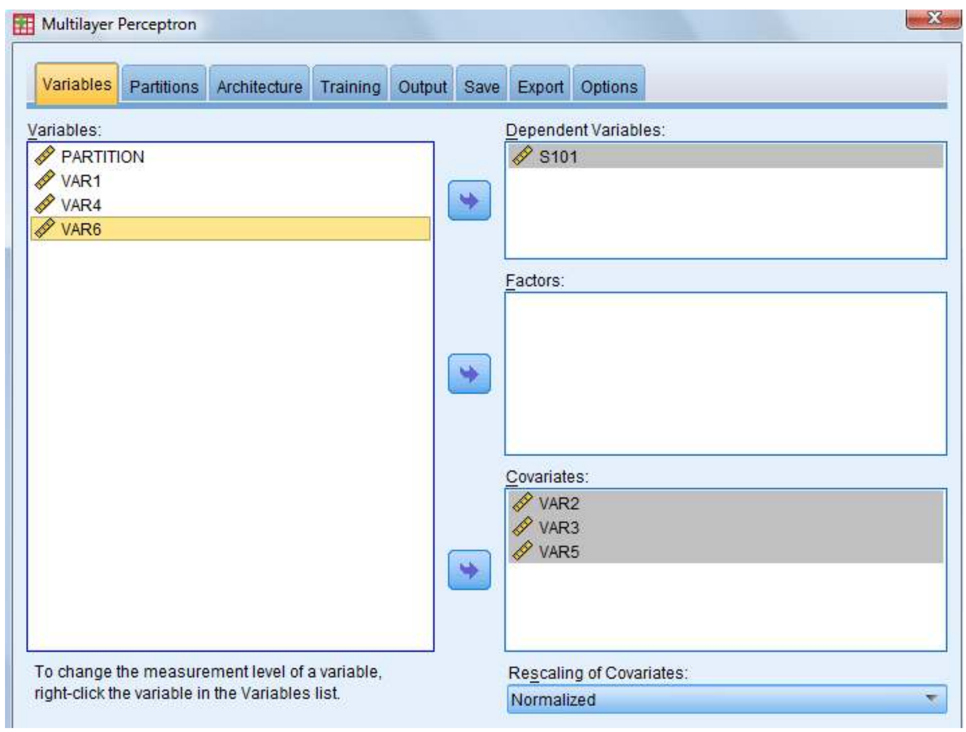

Figure 12 NN Multilayer Perceptron Variables Window

After “Variables” window was completed, “Partition” window as shown in Figure 13 was filled.

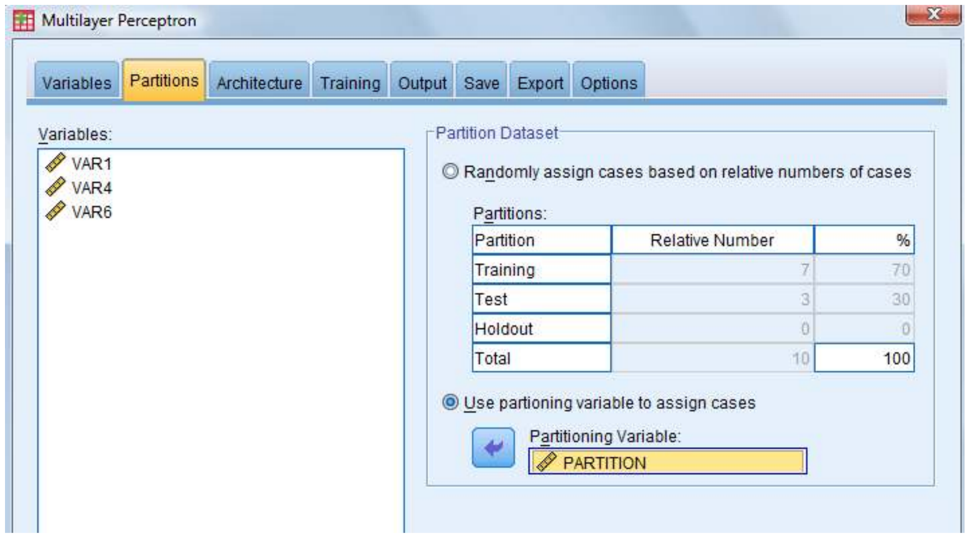

Figure 13 NN Multilayer Perceptron Partitions Window

In "partitions" window "variables" part shows the remaining variables which were not included in that model. On the right hand side there are two options that can be selected by the user. Upper option allow user to make random selection by only giving the relative numbers or percentages while determining the training, testing and hold out part of data set. However with the help of second option user can directly define which cases are used as training, testing or hold out. In order to use second option extra variable which composed of 1, 0, or -1 should be defined. "1" is implemented to show the case included in training part. 0 is used for testing and -1 is for holdout partitions. Training sample is used for obtaining the neural network model. Testing sample is data set apart from the training sample which tracks the error on the model. Holdout part is independent set of data records used to assess the final neural network (PASW Neural Networks 18). In this stage while preparing the main models, whole data set were assigned as training sample. However, in the validation phase data set were divided in to three parts as implemented in the regression models validations process, then models were developed with related cases assigned as training sample and the evaluation were done by holdout sample.

Structures of the models were developed by using “Architecture” window shown in Figure 14.

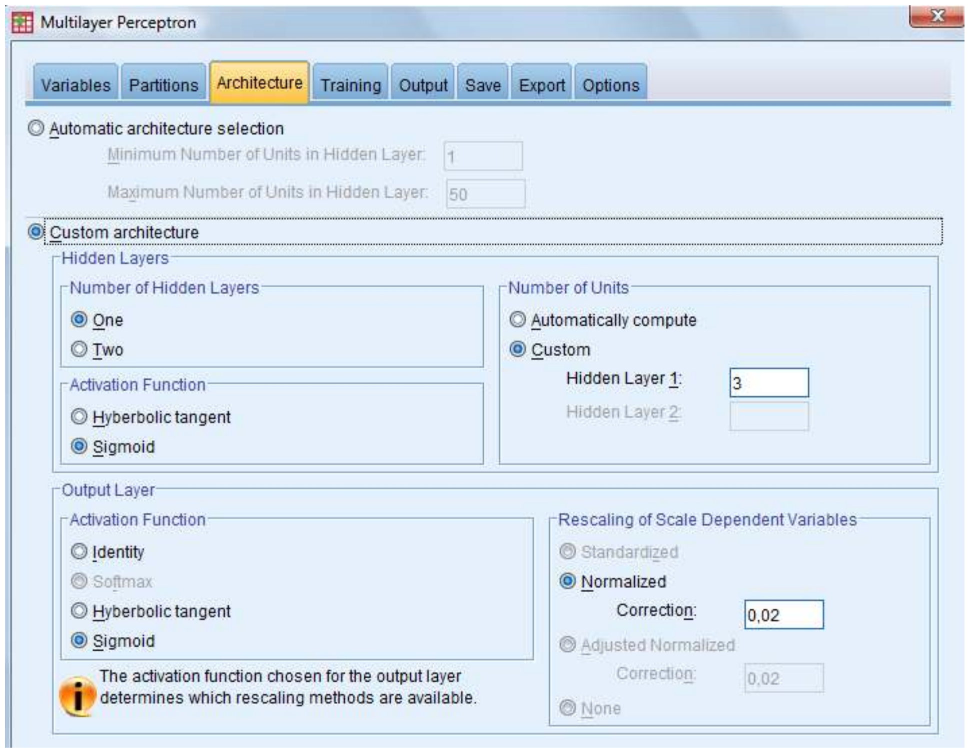

Figure 14 NN Multilayer Perceptron Architecture Window

While generating the models, one hidden layer was defined and generally average of the number of input and output variables was used as number of units in the hidden layer. Although there is no strict rule on formation of neural network model architecture, one hidden layer selection is thought as sufficient for most of the implementations. Sigmoid is the most popular transfer function and with the help of sigmoid function neural network models can learn the relationship between independent and dependent parameters. Therefore in this study sigmoid function was selected as activation function both in hidden and output layers. Normalization is the necessary rescaling method for scale-dependent variables when the output layer uses the sigmoid activation function. So that, "Rescaling of Scale Dependent Variable" was defined as "normalized" with a correction. In normalization process without correction minimum value for dependent variable gets 0 and maximum one gets 1. However, 0 and 1 are the limits for range of sigmoid function and they are not within the range. Therefore correction is applied to the normalization formula which was given in Formula (5).Corrected one used for the dependent variables is shown in the Formula (6) below.

$$X_{Normalized} = [X-(Min-\varepsilon)] / [(Max+\varepsilon)-(Min-\varepsilon)] \qquad (6)$$

After the constitution of the architecture for models training options were arranged. "Training" window shown in Figure 15 is composed of three types of training namely, batch, online and mini-batch, and two type of optimization algorithm namely scaled conjugate gradient and gradient descent.

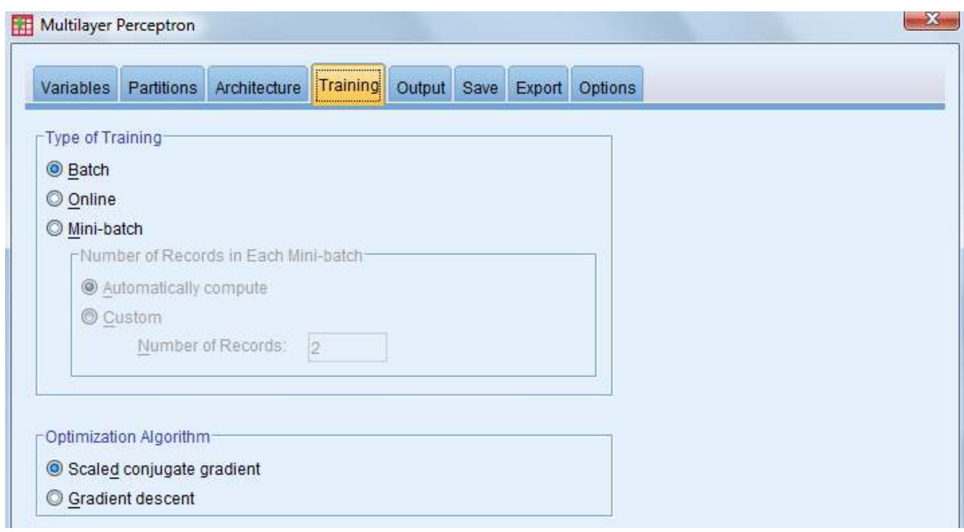

Figure 15 NN Multilayer Perceptron Training Window

Batch type updates the weights only after using all training data records. In other words batch training evaluates all information from the all cases in the database. This option is preferable because it minimizes the total error directly, but, if the database is large, it updates the weights many times until the stopping criterion is met. Therefore it is most useful for "smaller" datasets (PASW Neural Networks 18).

Online type updates the weights after every single training data record. It means that, this type of training evaluates the information from one case at a time. Procedure is started with one case and continues with other cases. After the evaluation of each case weights are updated. If the stopping criterion is not met after the evaluation of all cases, training continues by recycling. Online type of training is better than batch and mini-batch types when the database is composed of many cases and the model includes many input parameters. Under these circumstances online type of training reaches quicker and more reasonable results than the other two types (PASW Neural Networks 18).

Mini-batch type of training is suitable for the datasets which are medium-size. In mini-batch type training data are divided into groups of nearly equal size and weights are updated after evaluating each group. Mini-batch can be stated as a type between batch and online training (PASW Neural Networks 18).

In this study because the total number of cases included in the database is 40, it was classified as small dataset. Therefore, as a training type “batch” was preferred. After that, “scaled conjugate gradient” was selected as an optimization algorithm. The reason behind the selection is that Møller (1990) stated, gradient descent algorithm has bad convergence rate and it depends on the user specified variables whose selection does not have any theoretical basis. Moreover, scaled conjugate gradient algorithm gives faster results and if the demand for reduction in error increases, higher the speed up values is reached (Møller, 1990).

By implementing the procedures described above models were generated for each quantity take off items. Synaptic weights and the constant values which are called "Bias" in SPSS neural network output table are shown in Table 9, Table 10 and Table 11

Table 9 Synaptic Weights and Bias Values for Model Including 1 Hidden Unit

Predictor		Predicted	
S117		Hidden Layer 1	Output Layer
		H(1:1)	S117
Input Layer	(Bias)	2.587	
	VAR2	-22.725	
Hidden Layer 1	(Bias)		0.400
	H(1:1)		-5.110

In Table 9, Table 10 and Table11 it can be seen that in the neural network models generated in this study, input layer are composed of independent variables, hidden layer includes different numbers of hidden units and output layer contains output unit. Figures shown in these tables specify constant values which are indicated as "bias" term and synaptic weights used while evaluating the related unit in prediction stage. For example, in order to calculate the value to pass through hidden unit 1 (H(1:1)) bias value and weights for each variables which were obtained with the neural network modeling are used. After total value of the multiplications of each variables and related weights linked to the H(1:1) is added to the bias value, input used for the sigmoid function in the hidden unit is reached. After that, multiplication of the output of the sigmoid function and related weight is used as a component of next layer's units. This procedure continues by adding the multiplication of the units' outputs with related weights to the bias value and using these results as input for transfer function in related hidden units.

Table 10 Synaptic Weights and Bias Values for Models Including 2 Hidden Units

Predictor		Predicted		
S101		Hidden Layer 1		Output Layer
		H(1:1)	H(1:2)	S101
Input Layer	(Bias)	1.007	-2.715	
	VAR2	-0.272	0.976	
	VAR3	-0.449	0.380	
	VAR5	-2.515	3.287	
Hidden Layer 1	(Bias)			-2.104
	H(1:1)			-2.867
	H(1:2)			5.541
S102.1		Hidden Layer 1		Output Layer
		H(1:1)	H(1:2)	S102.1
Input Layer	(Bias)	1.715	-2.430	
	VAR2	-2.167	2.547	
	VAR3	-0.818	1.729	
	VAR6	-3.131	-3.288	
Hidden Layer 1	(Bias)			0.464
	H(1:1)			-5.432
	H(1:2)			4.278
S103.1-a		Hidden Layer 1		Output Layer
		H(1:1)	H(1:2)	S103.1-a
Input Layer	(Bias)	3.288	-2.657	
	VAR1	-4.077	2.694	
	VAR2	-2.811	2.804	
	VAR3	-2.441	-3.190	
Hidden Layer 1	(Bias)			-0.011
	H(1:1)			-8.153
	H(1:2)			5.380
S103.3		Hidden Layer 1		Output Layer
		H(1:1)	H(1:2)	S103.3
Input Layer	(Bias)	-0.374	3.292	
	VAR1	-2.107	-5.671	
	VAR2	-0.949	-0.291	
Hidden Layer 1	(Bias)			4.314
	H(1:1)			-2.459
	H(1:2)			-8.517
S104		Hidden Layer 1		Output Layer
		H(1:1)	H(1:2)	S104
Input Layer	(Bias)	-1.128	0.496	
	VAR2	2.997	1.350	
	VAR4	-1.305	-4.807	
	VAR6	2.037	-1.004	
Hidden Layer 1	(Bias)			-1.857
	H(1:1)			5.055
	H(1:2)			-3.530

Table 10 Synaptic Weights and Bias Values for Models Including 2 Hidden Units (Cont.)

Predictor		Predicted		
S109		Hidden Layer 1		Output Layer
		H(1:1)	H(1:2)	S109
Input Layer	(Bias)	2.295	3.187	
	VAR1	-3.608	-4.436	
	VAR2	-1.415	-1.134	
Hidden Layer 1	(Bias)			4.502
	H(1:1)			-4.053
	H(1:2)			-6.030
S112		Hidden Layer 1		Output Layer
		H(1:1)	H(1:2)	S112
Input Layer	(Bias)	5.119	2.776	
	VAR1	-4.940	-4.627	
	VAR2	-3.745	-9.606	
Hidden Layer 1	(Bias)			4.138
	H(1:1)			-8.824
	H(1:2)			2.260
S113		Hidden Layer 1		Output Layer
		H(1:1)	H(1:2)	S113
Input Layer	(Bias)	3.772	4.309	
	VAR1	-0.537	0.592	
	VAR2	-5.156	-7.377	
Hidden Layer 1	(Bias)			5.060
	H(1:1)			-6.177
	H(1:2)			-4.890
S118		Hidden Layer 1		Output Layer
		H(1:1)	H(1:2)	S118
Input Layer	(Bias)	-0.328	3.415	
	VAR1	-0.891	-0.795	
	VAR2	-1.382	-5.336	
Hidden Layer 1	(Bias)			3.422
	H(1:1)			-1.538
	H(1:2)			-7.978

Table 11 Synaptic Weights and Bias Values for Models Including 3 Hidden Units

Predictor		Predicted			
S102.3		Hidden Layer 1			Output Layer
		H(1:1)	H(1:2)	H(1:3)	S102.3
Input Layer	(Bias)	-0.228	1.376	6.302	
	VAR1	-0.313	1.966	-2.530	
	VAR2	-1.009	-0.410	-6.080	
	VAR4	-0.354	0.366	-1.468	
	VAR6	-1.287	-3.734	2.168	
Hidden Layer 1	(Bias)				6.968
	H(1:1)				-0.245
	H(1:2)				-3.544
	H(1:3)				-6.402
S103.1-b		Hidden Layer 1			Output Layer
		H(1:1)	H(1:2)	H(1:3)	S103.1-b
Input Layer	(Bias)	0.040	5.006	1.323	
	VAR2	-0.449	-2.709	0.360	
	VAR3	0.059	-5.044	-1.378	
	VAR4	2.740	-0.113	-2.632	
	VAR6	2.187	1.402	-2.322	
Hidden Layer 1	(Bias)				0.495
	H(1:1)				5.069
	H(1:2)				-5.455
	H(1:3)				-2.104

6.2. Validation of Neural Network Models

In this phase in order to make the neural network models comparable with linear regression models database were divided into three parts like in regression modeling procedure and by assigning same cases used in validation of regression, neural network models' validations were achieved. In the regression and neural network models as well as case based models, same validation sets were used. Results are shown in Table 12.

Table 12 Validation Results for Neural Network Models

Model	Quantity Take off Item Description	MAPE
S-101	Earthwork for Foundations	9.30
S-102.1	Concrete BS14	3.66
S-102.3	Concrete BS25	6.61
S-103.1-a	Reinforcing Bar St III (≤ Ø12)	4.73
S-103.1-b	Reinforcing Bar St III (> Ø12)	7.46
S-103.3	Pre-stressing Steel	5.95
S-104	Placed Stone Filling	7.54
S-109	Waterproofing for Bridge Deck	4.26
S-112	Pre-stressed Concrete Beam Execution and Installation	1.86
S-113	Elastomeric Bearing	2.32
S-117	PVC Service Pipe	7.76
S-118	Galvanize Form	5.62

As can be seen from the Table 12, MAPE values are between 1.86 and 9.30. These results can be interpreted that all neural network models have sufficient prediction performance for this study.

CHAPTER 7

CASE BASED REASONING MODELS

In this stage of the study, software called ESTEEM Version 1.4 was used to generate case based reasoning models for quantity take off items. Following sections in this chapter, preparation of the case based reasoning models and validation of them will be discussed.

7.1. Preparation of Case Based Reasoning Models

Case based reasoning model preparation was composed of 4 steps as follows.

1) Assigning Case Base Definitions
2) Assigning Cases
3) Assigning Similarity Definitions
4) Arranging End User Interface Editor

7.1.1. Assigning Case Base Definitions

In this step, feature names and value types were assigned to the software. Feature names of case based reasoning models were composed of both dependent and independent variables' names used in the regression and neural network models of related quantity take of items. For example, model generated with regression and neural network methods for work item S101 included Var2, Var3 and Var5 as independent variables, so that feature names of the case based reasoning model for S101 were assigned as Var2, Var3, Var5, and S101. During the feature value type selection phase, "numeric" was defined for all dependent and independent variables because the whole database used in the analyses were suitable for this choice. Table 13 shows, case base definitions assigned for each model.

Table 13 Case Base Definitions for Each Model

Quantity Take off Item No:	Feature	Feature Description	Feature Type
S-101	Var2	Width of Bridge	Numeric
	Var3	A_0 Value	Numeric
	Var5	Average Excavation Height	Numeric
S-102.1	Var2	Width of Bridge	Numeric
	Var3	A_0 Value	Numeric
	Var6	Maximum Abutment Height	Numeric
S-102.3	Var1	Length of Span	Numeric
	Var2	Width of Bridge	Numeric
	Var4	Distance Between Grade Elevation of Highway and Railway	Numeric
	Var6	Maximum Abutment Height	Numeric
S-103.1-a	Var1	Length of Span	Numeric
	Var2	Width of Bridge	Numeric
	Var3	A_0 Value	Numeric
S-103.1-b	Var2	Width of Bridge	Numeric
	Var3	A_0 Value	Numeric
	Var4	Distance Between Grade Elevation of Highway and Railway	Numeric
	Var6	Maximum Abutment Height	Numeric
S-103.3	Var1	Length of Span	Numeric
	Var2	Width of Bridge	Numeric
S-104	Var2	Width of Bridge	Numeric
	Var4	Distance Between Grade Elevation of Highway and Railway	Numeric
	Var6	Maximum Abutment Height	Numeric
S-109	Var1	Length of Span	Numeric
	Var2	Width of Bridge	Numeric
S-112	Var1	Length of Span	Numeric
	Var2	Width of Bridge	Numeric
S-113	Var1	Length of Span	Numeric
	Var2	Width of Bridge	Numeric
S-117	Var2	Width of Bridge	Numeric
S-118	Var1	Length of Span	Numeric
	Var2	Width of Bridge	Numeric

7.1.2. Assigning Cases

40 projects were utilized in order to generate the main models of the work items. In this stage all cases were stored in the database of the software by assigning their feature names and values. Figure 16 gives the example screen of ESTEEM software used during the case assigning stage of the model preparation.

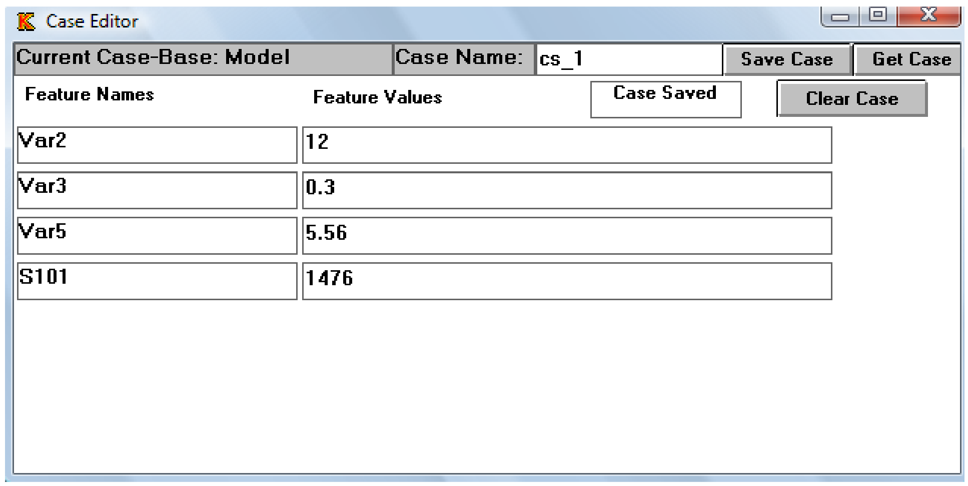

Figure 16 Case Editor Window

7.1.3. Assigning Similarity Definitions

At this step of model preparation, similarity definition is needed. ESTEEM has 3 similarity definitions namely, feature counting method, weighted feature and inferred feature computation.

7.1.3.1. Feature Counting Method

Feature counting method uses the concept of nearest neighbor technique which was described in section 3.3.2. In this method while ESTEEM is examining the similarity between the case in database and current case to be predicted, all variables' weights are automatically set as 1. Because all of the weights are taken as 1, in this method all of the features have equal contribution in the model. In other word this type of similarity definition skips the dominance of one variable to another one during the prediction stage. Therefore it was not used in this study.

7.1.3.2. Weighted Feature Computation

In ESTEEM, there are 3 alternatives presented for generating the weights of the features namely, ID3, gradient descent and manual weight assigning method.

- *ID3 Method*

ID3 is the Quinlan's(1986) algorithm that produce a decision tree for the cases in database. After that by means of the tree weights for each variable are computed. In ESTEEM if ID3 method is selected, then the variable being dependent and independent variables are defined by the user to generate the weights of features. After running this procedure features which have zero weights are eliminated. The disadvantages of ID3 method in ESTEEM is that user must select the feature match type as "equal". Although this method takes the dominance of the features into consideration, because there is a must for feature match type preference it was also not used in this study

- *Gradient Descent Method*

Gradient descent method is superior to ID3 method because when the user prefers to utilize gradient descent method all feature match types can be included in the analysis. Therefore this method was used while generating the case based reasoning models in this study. As it is described in the ID3 method part, during the gradient descent method application stage user should specify the source and target features to be used in the model. Moreover, after specification of the source and target features ESTEEM introduce new window shown in Figure 17, consisting of "step size updating method", "starting step size", "ending step size", "step size update parameter", and "number of cases tested per step" parts. Step size updating method can be selected as geometric or arithmetic. In geometric method, starting step size called delta is decreased to the result of the multiplication step size update parameter and current delta. On the other hand, if arithmetic method is chosen delta update is done by subtracting the step size update parameter from the current delta. Ending step size is specified as the stopping criteria for process Karancı (2010).

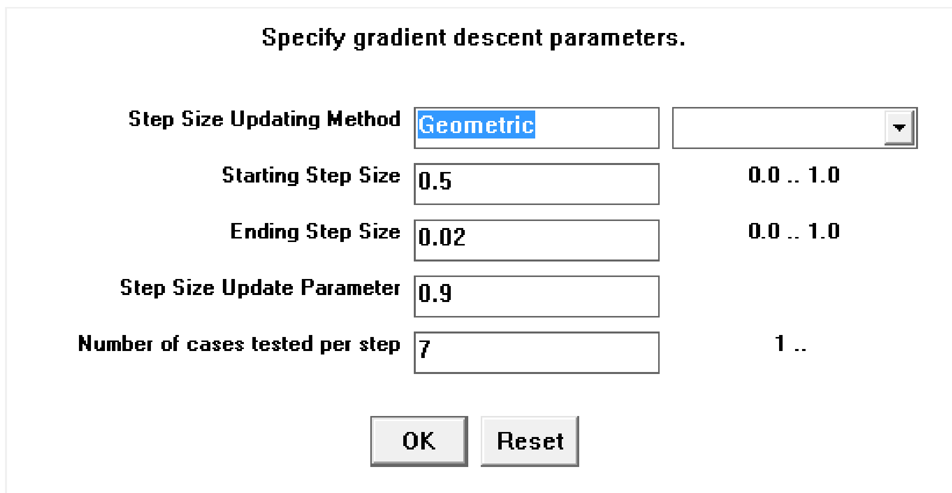

Figure 17 Gradient Descent Parameters Window

Algorithm of gradient descent method can be clarified as follows. First of all, from the database several random cases are selected. Then the most similar cases to the randomly pre-selected ones determined. This determination of the similar cases is based on the source features' current weights. In other words, according to match types of the features, similarities are calculated for each feature and by using weights and feature similarities case similarities are calculated. After the similar cases are found, considering consistency between target features of the matching cases, source features' current "weight updates" are calculated. After analyzing various random cases the updates vector are normalized, multiplied by Delta and added to the current weight vector. After that delta is decreased with step size update parameter and until the result of delta is equal or less than ending step size this process which is summarized in Figure 18 (Doğan, Arditi and Günaydın 2006).

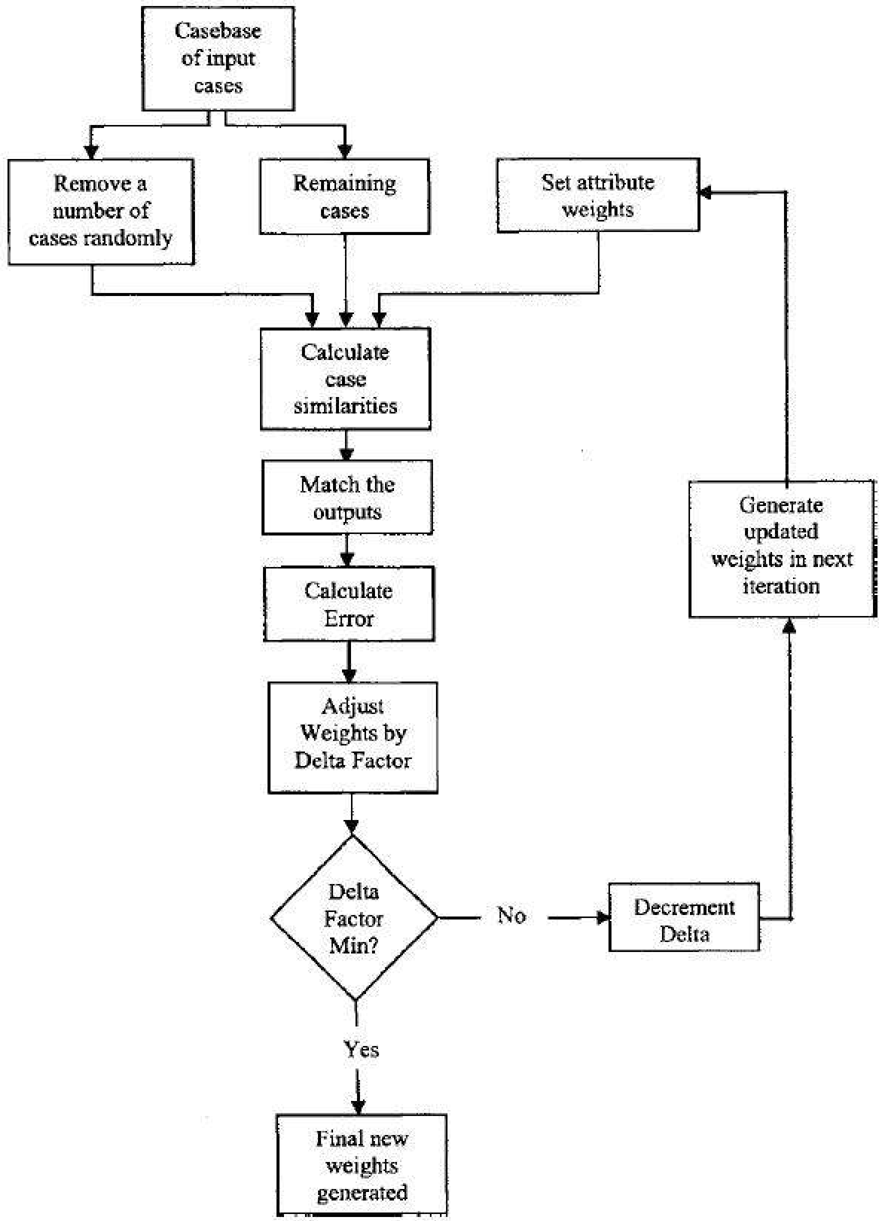

Figure 18 Gradient Descent Process (Doğan, Arditi and Günaydın 2006)

- *Manual Weight Assigning Method*

Manual weight assigning method provides opportunity to the user to assign feature weights manually. If the user does not have experience on the cases and is not sure about the weights of features ESTEEM allows the user to generate the weights by using its algorithms.

7.1.3.3. Inferred Feature Computation

In ESTEEM the last option for the similarity definition is inferred feature computation which can be thought as knowledge guided indexing described in chapter 3. This weight generation method is based on the rules defined by the user. By using these pre-defined rules, system calculates the weights for each source features.

In this study, as specified earlier gradient descent method was preferred for determining the weights of source features. In order to assess the similarity of source features gradient descent method in ESTEEM allow user to define the match type of the feature. For numeric type of data software gives 6 alternatives namely, “Equal”, “Range”, “Fuzzy Range”, “Absolute Range”, “Absolute Fuzzy Range” and “Inferred” to choose. Equal type is used for exact match and if the exact matching is satisfied software return the result as 1 otherwise 0. Range type checks whether the smaller value is in the specified percent range of the bigger value for assigning 1 or 0. When the fuzzy range is selected, returned value depends on specified percent range of the bigger value. In this type if the smaller one is in the range, result is calculated by specifying smaller feature’s location in the range by proportioning. Absolute range type runs as the description for range type but other than that in this type distance is specified instead of percentage. Similarly, absolute fuzzy range type is works like fuzzy range type but unlike fuzzy range, absolute fuzzy range type takes the range as distance rather than percent of bigger feature. Inferred feature match determines the similarity score depending o predefined rule in the rule base (ESTEEM, 1996). Preferred feature matching types in case based reasoning models are shown in Table 14.

Table 14 Types of Feature Matching Used in Models

Quantity Take off Item No:	Feature	Feature Description	Feature Matching Type
S-101	Var2	Width of Bridge	Fuzzy Range: Tol.: 90%
	Var3	A_0 Value	Fuzzy Range: Tol.: 90%
	Var5	Average Excavation Height	Fuzzy Range: Tol.: 90%
S-102.1	Var2	Width of Bridge	Absolute Fuzzy Range: Tol.: 4.6
	Var3	A_0 Value	Absolute Fuzzy Range: Tol.: 0.2
	Var6	Maximum Abutment Height	Absolute Fuzzy Range: Tol.: 5.41
S-102.3	Var1	Length of Span	Fuzzy Range: Tol.: 90%
	Var2	Width of Bridge	Fuzzy Range: Tol.: 90%
	Var4	Distance Between Grade Elevation of Highway and Railway	Fuzzy Range: Tol.: 90%
	Var6	Maximum Abutment Height	Fuzzy Range: Tol.: 90%
S-103.1-a	Var1	Length of Span	Fuzzy Range: Tol.: 90%
	Var2	Width of Bridge	Fuzzy Range: Tol.: 90%
	Var3	A_0 Value	Fuzzy Range: Tol.: 90%
S-103.1-b	Var2	Width of Bridge	Absolute Fuzzy Range: Tol.: 4.6
	Var3	A_0 Value	Absolute Fuzzy Range: Tol.: 0.2
	Var4	Distance Between Grade Elevation of Highway and Railway	Absolute Fuzzy Range: Tol.: 2.63
	Var6	Maximum Abutment Height	Absolute Fuzzy Range: Tol.: 5.41
S-103.3	Var1	Length of Span	Absolute Fuzzy Range: Tol.: 10
	Var2	Width of Bridge	Absolute Fuzzy Range: Tol.: 4.6
S-104	Var2	Width of Bridge	Absolute Fuzzy Range: Tol.: 4.6
	Var4	Distance Between Grade Elevation of Highway and Railway	Absolute Fuzzy Range: Tol.: 2.63
	Var6	Maximum Abutment Height	Absolute Fuzzy Range: Tol.: 5.41
S-109	Var1	Length of Span	Absolute Fuzzy Range: Tol.: 10
	Var2	Width of Bridge	Absolute Fuzzy Range: Tol.: 4.6
S-112	Var1	Length of Span	Absolute Fuzzy Range: Tol.: 10
	Var2	Width of Bridge	Absolute Fuzzy Range: Tol.: 4.6
S-113	Var1	Length of Span	Absolute Fuzzy Range: Tol.: 10
	Var2	Width of Bridge	Absolute Fuzzy Range: Tol.: 4.6
S-117	Var2	Width of Bridge	Absolute Fuzzy Range: Tol.: 4.6
S-118	Var1	Length of Span	Absolute Fuzzy Range: Tol.: 10
	Var2	Width of Bridge	Absolute Fuzzy Range: Tol.: 4.6

7.1.4. Arranging End User Interface Editor

In this step, in end user interface window, user arranges the retrieve stage of the case based reasoning method. There are three columns namely target case entry features, retrieved cases features and selected case features. Source features are entered in target case entry features column. In retrieved case features column variable which is needed to see in retrieval case list should be assigned. Selected case features column arranges that which of the variables are shown when a specific case is selected.

To sum up, during the preparation of case based reasoning models first case base definitions which are composed of the variable names and types were defined. As can be seen from the Table 13 all the variables are numeric. After that, by assigning the all cases, database was prepared. Then by selecting the feature matching type and using weighted feature computation, weights for the features were produced. Decision of the feature matching type and the weight generation method was based on the trial models. In other words several models were generated with different types of features matching and weight generation methods and then by the help of investigations on the results absolute fuzzy range or fuzzy range with gradient descent method was preferred. As can be seen from the Table 14 most of the models were prepared by using absolute fuzzy range type whose tolerance values were determined by taking difference between the maximum and minimum value of the related features. For example, if a features max value in case base is 11 and min value in database is 1, and tolerance is defined as 11-1=10, this means that 1 unit distance decreases the 10% of similarity. In this example 5's similarity to 6 is 90% or 3's similarity to 8 is 50%. Moreover for some models it was seen that fuzzy range type gave better prediction performance so that these models were modified by using fuzzy range feature matching type. In the light of this information by using all cases main models' feature weights were generated. In Table 15 feature weights of the all models are shown.

Table 15 Weight of Features in CBRM Models

Quantity Take off Item No:	Feature	Feature Description	Weight of Feature
S-101	Var2	Width of Bridge	0.149
	Var3	A_0 Value	0.211
	Var5	Average Excavation Height	0.640
S-102.1	Var2	Width of Bridge	0.474
	Var3	A_0 Value	0.376
	Var6	Maximum Abutment Height	0.149
S-102.3	Var1	Length of Span	0.012
	Var2	Width of Bridge	0.451
	Var4	Distance Between Grade Elevation of Highway and Railway	0.022
	Var6	Maximum Abutment Height	0.515
S-103.1-a	Var1	Length of Span	0.016
	Var2	Width of Bridge	0.336
	Var3	A_0 Value	0.648
S-103.1-b	Var2	Width of Bridge	0.226
	Var3	A_0 Value	0.519
	Var4	Distance Between Grade Elevation of Highway and Railway	0.154
	Var6	Maximum Abutment Height	0.101
S-103.3	Var1	Length of Span	0.260
	Var2	Width of Bridge	0.740
S-104	Var2	Width of Bridge	0.469
	Var4	Distance Between Grade Elevation of Highway and Railway	0.336
	Var6	Maximum Abutment Height	0.195
S-109	Var1	Length of Span	0.094
	Var2	Width of Bridge	0.906
S-112	Var1	Length of Span	0.190
	Var2	Width of Bridge	0.810
S-113	Var1	Length of Span	0.223
	Var2	Width of Bridge	0.777
S-117	Var2	Width of Bridge	1.000
S-118	Var1	Length of Span	0.120
	Var2	Width of Bridge	0.880

7.2. Validation of Case Based Reasoning Models

While determining the final output of the models after the retrieval process, when the similarity scores of the cases were obtained from the software, average of the results of cases which have highest similarity score were taken. Validation process described previously in regression and neural network chapters were applied in the same manner and the prediction performance which can be seen in Table 16 for each model were calculated.

Table 16 Validation Results for CBRM Models

Model	Quantity Take off Item Description	MAPE
S-101	Earthwork for Foundations	11.30
S-102.1	Concrete BS14	3.56
S-102.3	Concrete BS25	4.98
S-103.1-a	Reinforcing Bar St III (≤ Ø12)	3.18
S-103.1-b	Reinforcing Bar St III (> Ø12)	7.31
S-103.3	Pre-stressing Steel	7.61
S-104	Placed Stone Filling	7.23
S-109	Waterproofing for Bridge Deck	5.51
S-112	Pre-stressed Concrete Beam Execution and Installation	2.50
S-113	Elastomeric Bearing	1.38
S-117	PVC Service Pipe	7.65
S-118	Galvanize Form	5.58

As can be seen from the Table 16, MAPE values are between 1.38 and 11.30. These results can be interpreted that all case based reasoning models have sufficient prediction performance for this study.

CHAPTER 8

COMPARISON OF METHODS

The prediction performances of regression, neural network and case based models are summarized in Table 17.

Table 17 Prediction Performance Comparison of Methods

Model	Quantity Take off Item Description	MAPE	MAPE	MAPE
		Regression	Neural Network	Case Based Reasoning
S-101	Earthwork for Foundations	9.23	9.30	11.30
S-102.1	Concrete BS14	3.02	3.66	3.56
S-102.3	Concrete BS25	9.15	6.61	4.98
S-103.1-a	Reinforcing Bar St III (≤ Ø12)	3.96	4.73	3.18
S-103.1-b	Reinforcing Bar St III (> Ø12)	11.04	7.46	7.31
S-103.3	Pre-stressing Steel	7.82	5.95	7.61
S-104	Placed Stone Filling	7.39	7.54	7.23
S-109	Waterproofing for Bridge Deck	3.87	4.26	5.51
S-112	Pre-stressed Concrete Beam Execution and Installation	4.95	1.86	2.50
S-113	Elastomeric Bearing	6.70	2.32	1.38
S-117	PVC Service Pipe	8.29	7.76	7.65
S-118	Galvanize Form	6.73	5.62	5.58

As can be seen from the Table 17 different methods gave better results in different quantity take off items. For example regression models for work part S-101, S-102.1 and S-109 have the best prediction performance among the three methods. On the other hand neural network models for S-103.3 and S-112 have better average accuracy on prediction than both models obtained by using regression and case based reasoning methods. Moreover when the models for S-102.3, S-103.1a, S-103.1b, S-104, S-113, S-117, and S118 were taken into consideration it was seen that case based reasoning method gave better results than the other two methods. Although regression models prediction performance were calculated between 3.02% and 11.04% neural networks and case based reasoning models' prediction performance were between 1.86% - 9.3%, and 1.38% - 11.3% respectively.

Other than these comparisons, in order to find the effect of conceptual quantity take off study on the final cost estimation of the projects first of all by multiplying the unit price of the works items determined by the design firm for year 2009 with the related quantities obtained from each model, costs for each item were determined and added to each other to find the total costs. With the help of these calculations, predicted costs for each project were achieved. Then actual quantities calculated after the finishing of the detailed design stage were multiplied with the year 2009 unit cost and by summation of actual work item costs, actual projects costs for year 2009 were reached. Finally by implementing the MAPE evaluation procedure described previously, prediction performances of the cost models which were determined by using the conceptual quantity take off models' estimations were attained. The results are shown in Table 18 below.

Table 18 Prediction Performance of Total Cost Model

Model	**MAPE**	**MAPE**	**MAPE**
	Regression	**Neural Network**	**Case Based Reasoning**
Total Cost	6.77	4.70	3.87

In the light of Table 18 it can be said that with the 6.77% MAPE value for prediction performance, regression method is the worst method for this study in determining the total project cost. Moreover, as can be seen from Table 18 total cost models generated by using conceptual quantity take off models have MAPE value of 4.70% and 3.87% when the neural network and case based reasoning methods were used respectively.

CHAPTER 9

SUMMARY AND CONCLUSIONS

As a conclusion, quantity take off and cost estimation studies for the construction projects are crucial for the companies in the construction industry. When companies overestimate the total cost of a project, firms may miss the chance of winning the bid and when the final cost is underestimated it may result as a loss for the company. Because in either case, outcome influences enterprises negatively, this topic is one of the main concern for all establishments in the construction industry. Moreover, not only contractors but also clients are affected by the results of the estimations. At the decision phase while determining whether the implementation of the projects are feasible or not, main issue is estimating the final cost of the work with the highest level of accuracy with very limited knowledge about the project.

In order to overcome the problems caused by the predictions, contractors and clients take a precaution by forming estimation teams consisting of experienced engineers. However because of the lack of detailed data in pre-design phase of the project and the time limitations on engineers for predicting the total cost of works, new procedures were investigated to determine the cost of projects before the detailed drawings are prepared.

Linear regression, neural network and case based reasoning methods that were most commonly used in the literature for cost estimation of the projects which usually consist of buildings. Moreover, these studies included the direct cost prediction of the projects without estimating the quantity take off items. In this study by using linear regression, neural network, and case based reasoning techniques cost of the projects were not directly predicted. First with the help of the methods described previously, quantity take off items were modeled and then by using the unit costs of each item, cost of each part were predicted and finally total cost of projects were achieved by means of summation process of work items.

Study started with the data collection of highway projects. When the raw data were collected, some of which were defined as noisy or realized out of predefined area were eliminated. In other words, in order to get realistic and satisfactory results by defining the limitations of the database, data which were out of the scope or which were seen as unreliable were eliminated. Limitations of the data which were clarified previously can be summarized as follows:

- Highway bridges
- Single Span, length between 20m and 30m including the edge values
- Width of Bridge between 9.5m and 14.1m
- Foundation above the ground water level
- Soil profile consists of stiff clay or deep cohesionless conditions
- Foundation without pile
- Pre-stressed concrete beam (I Section – BS40)
- No skewness
- Bridges were designed according to AASHTO 2002

After the preparation of the database by using linear regression method, investigating the significance values of coefficients, and coefficient of determination values, final linear regression models' parameters were determined. During this stage the parsimony concept was considered to build models with good prediction performance. When the linear regression models were completed, by utilizing the same parameters which were included in the final regression models, neural networks models were generated. Similarly, next to the carrying out the neural network quantity take off models, case based reasoning method was performed in order to make models for each work items.

After generation of the models by applying each method, to compare the prediction ability of methods for every model 3-fold cross validation process was implemented. Moreover, by integrating the unit costs with the estimated quantities of work parts final costs of the projects were predicted. 3-fold cross validation was also applied to the total costs of the projects. In the light of this study as a summary several advantages and disadvantages of methods encountered are listed below

Linear Regression:

- Because of the P-value analysis, linear regression method gave great chance to identifying the parameters to be used in the final model achieving parsimony concept.
- When prediction performances of the total project costs were taken into consideration this method gave the worst result among the 3 methods

Neural Network:

- During the modeling process, because the starting weights were different as default in software, neural network method generated different synaptic weights in other words different models, even if the database, dependent and independent parameters were not changed.
- In order to reach final model, trial models were generated.
- When prediction performances of the total project costs were taken into consideration this method gave better result than linear regression method.

Case Based Reasoning:

- During the modeling process, because of the random case selection process in the gradient descent algorithm, case based reasoning method generated different feature weights in other words different models, even if the database, dependent and independent parameters were same.
- Because the cases in the database were similar to each other best prediction performance for the total project costs were achieved.

To sum up, the main purpose of this study was predicting the quantity take off items constituting of the single span highway projects and the total cost of projects by using estimated quantities. Therefore 144 models were developed including main models and validation process. Moreover, in order to determine the final models extra trial models were prepared. With the help of this study several advantages listed below can be taken.

- Estimated quantities can be used for checking the design of the bridge.
- By integrating the unit costs for each item of a specific year, total project cost can be predicted.
- By using these estimations in feasibility study, more feasible projects can be implemented.
- Clients can check the contractors' tender prices.
- Contractors can check the quantity take off measurements and total cost calculations before the tender phase.
- With the help of these models, inexperienced persons can realize the effects of the each work item.
- Planning of the project can be done safer and in advance by means of prediction of quantities.
- Predicting the quantities increases the prediction ability of the indirect cost estimations (For example warehouse sizes can be determined more accurately).
- By using the method giving the best result for each item total cost can be estimated more accurately.

In the further research by enhancing the database in other words by collecting and investigating more cases, models can be improved. Moreover conceptual quantity take off process can be implemented to construction projects other than single span highway projects.

REFERENCES

Arditi, D. and Tokdemir, O. B. (1999). Comparison of case-based reasoning and artificial neural networks. Journal of Computing in Civil Engineering, 13(3), 162-169

Ashworth, A. (1988). Cost studies of buildings, Longman Scientific and Technical, Harlow, Essex, U.K.

Ashworth, A. and Skitmore, R.M. (1983). Accuracy in estimating. Chartered Institute of Building, Ascot.

Barletta, R. (1991). An introduction to case -based reasoning. AI Expert, 6(8), 42 – 49.

Chou, J. S. (2009). Web – based CBR system applied to early cost budgeting for pavement maintenance project. Expert Systems with Applications, 36, 2947 – 2960.

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

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

Esteem Software. (1996). Esteem 1.4: *Case based reasoning development tool*, San Mateo, California.

Gunaydin, H. M., and Dogan, S. Z. (2004). A neural network approach for early cost estimation of structural systems of buildings. International Journal of Project Management, 22, 595-602.

Gupta, U. G. (1994). How case – based reasoning solves new problems. Interfaces, 24(6), 110 – 119.

Holt, A., Bichindaritz, I. and Schmidt, R. (2006). Medical applications in case-based reasoning. The Knowledge Engineering Review, 20(3), 289-292.

Irish, L. B. Jr., Barrett, M. E. Malina J. F. Jr. and Charbeneau, R. J. (1998). Use of regression models for analyzing highway storm-water loads. Journal of Environmental Engineering, 124(10), 987-993.

Karancı, H. (2010). A comparative study of regression analysis, neural networks, and case – based reasoning for early range cost estimation of mass housing projects. MSc thesis, Middle East Technical University., Graduate School of Natural and Applied Sciences, Ankara, Turkey.

Karshenas, S. (1984). Predesign cost estimating method for multistory buildings. Journal of Construction Engineering and Management, 110(1), 79-86.

Kim, B. S. (2011). The approximate cost estimating model for railway bridge project in the planning phase using cbr method. KSCE Journal of Civil Engineering, 15(7), 1149-1159.

Kim, B. S. and Hong, T. (2012). Revised case-based reasoning model development based on multiple regression analysis for railroad bridge construction. Journal of Construction Engineering and Management, 138(1), 154-162

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

Kim, K. J., Kim, K. and Kang, C. S. (2009). Approximate cost estimating model for psc beam bridge based on quantity of standard work. KSCE Journal of Civil Engineering, 13(6), 377-388.

Kolodner, J. (1993). Case-based reasoning. San Mateo, CA: Morgan Kaufmann Publishers, Inc.

Kouskoulas, V. and Koehn, E. (1974). Predesign cost-estimation function for buildings. Journal of the Construction Division, ASCE, 100(CO4), 589-604.

Li, H., Shen, Q.P. and Love, P.E.D. (2005). Cost modeling of office buildings in Hong Kong: an exploratory study. Facilities, 23(9/10), 438-452.

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

Marr, K.F. (1977). Standards for construction cost estimates. Transactions of the American Association of Cost Engineers, Paper B-6, 77-80.

MØller M. F. (1990). A scaled conjugate gradient algorithm for fast supervised learning.

Oberlender, G.D. (2000), Project Management for Engineering and Construction, The McGraw-Hill Companies, Inc., Boston, MA

Özmen, B., Nurlu, M. and Güler, H. (1997). Analyze of earthquake zones with geographical information system.

Özorhon, B. (2004). Organizational memory in construction companies: A case – based reasoning model as an organizational learning tool. MSc thesis, Middle East Technical University., Graduate School of Natural and Applied Sciences, Ankara, Turkey.

Pankratz, A. (1983). Forecasting with univariate Box - Jenkins models. John Wiley & Sons, New York, NY. pp 81 – 82.

PASW Neural Networks 18, http://www.uky.edu/ComputingCenter/SSTARS/SPSS/18_Manuals/PASW_Neural Network_18.pdf

Project Management Institute (2008). A Guide to the Project Management Body of Knowledge (PMBOK Guide) – Fourth Edition. Project Management Institute, Newton Square, Pennsylvania.

Quinlan, J. R. (1986). Induction of decision trees. *Mach. Learn.*, 1(1), 81 – 106.

Smith I.N.C., Ballou, D.P. and Pazer, H.L. (1996). Modeling the costs of bridge rehabilitation. Transpn Res.-A, 31(4), 281-293.

Sonmez, R. (2004). Conceptual cost estimation of building projects with regression analysis and neural networks. Canadian Journal of Civil Engineering, 31(4), 677-683.

Sonmez, R. (2008). Parametric range estimating of building costs using regression models and bootstrap. Journal of Construction Engineering and Management, 134(12), 1011 – 1016.

Sonmez, R. and Ontepeli, B. (2009). Predesign cost estimation of urban railway projects with parametric modeling. Journal of Civil Engineering and Management, 15(4), 405-409.

Trost, S. M., and Oberlender, G. D. (2003). Predicting accuracy of early cost estimates using factor analysis and multivariate regression. Journal of Construction Engineering and Management, 129(2), 198-204.

Turkish Ministry of Public Works and Settlement (2007). Specification for structures to be built in disaster area, TEC, Ankara, Turkey.

Ugwu, O. O. and Kumaraswamy, M. M. (2004). Neural Network based decision support for estimating cost of highway bridges – A Hong Kong study. COBRA 2004.

Wang, H. J., Chiou, C., and Juan, Y. K. (2008). Decision support model based on case – based reasoning approach for estimating the restoration budget of historical buildings. Expert Systems with Applications, 35, 1601-1610.

Wang, Y.M. and Elhag, T. M. S. (2007). A comparison of neural network evidential reasoning and multiple regression analysis in modelling bridge risks. Expert Systems with Applications 32, 336-348.

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

Yeh, I. C. (1998). Quantity estimating of building with logarithm-neural networks. Journal of construction engineering and management, 124(5), 374-380.