http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 5 (6): 1143-1145, 2006
ISSN 1812-5638

Multivariate Interpolation Model to Estimate the Effort Component of Software Projects

Mitat Uysal
Department of Computer Engineering,
Dogus University-Kadiköy-Acibadem, Istanbul, Turkey

Abstract: In this study, a multivariate interpolation model was developed to estimate the effort component in the software projects. The data set that was used consists of two independent variables, first is Developed Lines (DL) and second is Methodology (ME) and one dependent Variable Effort (E). The data set is taken from (Shin and Goel, 2000) and the results that are obtained in my work are compared with the results of Shin and Goel (2001) that are produced using a different model based on RBF.

Key words: Software engineering, software cost estimation, multivariate interpolation, data analysis, curve fitting

INTRODUCTION

Estimation of resources, cost and schedule for a software development effort requires experience, access to good historical information and the courage to commit to quantitative measures when qualitative data are all that exist (Pressman, 1992).

The importance of software cost estimation is well documented. Good estimation techniques serve as a basis for communication between software personnel and non-software personnel such as managers, sales people or even customers (Knafe, 1995).

As estimation model for computer software uses empirically derived formulas to predict data that are a required part of the software project planning step (Pressman, 1992).

Resource models consist of one or more empirically derived equations that predict effort (in person-months), project duration (in chronological months), or the other pertinent project data.

Basili (1980) described four classes of resources models:

- Static single-variable models
- Static multi-variable models
- Dynamic multi-variable models
- Theoretical models

The static single-variable model takes the form:

$$\text{Resources} = C_1 * (\text{Estimated characteristics})^{c_2}$$

where the resources could be effort, project duration, staff size or requisite lines of software documentation. The constants C_1 and C_2 are derived from data collected from past projects. The basic version of the Constructive Cost Model or COCOMO is an example of a static single-variable model.

Static multi-variable model has the following form:

$$\text{Resources} = C_{11}\, e_1 + C_{21}\, e_2 + \ldots$$

where e_i is the ith software characteristics and C_{21}, C_{22} are empirically derived constants for the ith characteristics (Pressmann, 1992)

A dynamic multivariate model projects resource requirements as a function of time.

A theoretical approach to dynamic multivariable modeling hypothesizes a continuous resource expenditure curve and from it, derives equations that model the behavior of the resource. The Putnam Estimation Model is a theoretical dynamic multi-variable model. Some new models are proposed for software cost estimation. One of them is Peters and Ramanna Model based on an application of the Choquet integral (Peters and Rammana, 1996)

This is a form of multi-criteria decision-making where the numeric value computed by the Choquet integral is an expression of the degree of preference of one technology over another in developing a software system.

Neural Networks are another tool to develop software cost estimation models. Idri *et al.* (2002) proposed a new model for this. They have used the full purpose COCOMO '81 dataset to train and to test the network. The obtained accuracy of the network was acceptable.

SOFTWARE COST MODEL THAT IS USED IN THIS WORK

The following software cost model is used in present study:

$$E = f\,(DL, ME)$$

Where E is effort, DL is Developed Lines and ME is methodology used in the software project. f is a nonlinear function in terms of DL and ME.

Multivariate interpolation method was used to find the interpolated values of DL and ME using a data set that contains E, DL and ME values obtained from past projects. This data set is taken from (Shin and Goel, 2000).

TWO DIMENSIONAL INTERPOLATION

The two dimensional function table is an array of functional values $f_{i,j} = f(DL_i, ME_j)$ on a rectangular grid, (DL_i, ME_j) as shown in Fig. 1.

Double lagrange interpolation is to apply the lagrange interpolation method twice in two dimensions (Nakamura, 2002). Therefore, the interpolation uses all the data points in the table. Suppose the function table has M columns and N rows. The coordinates of the points are denoted by (DL_M, ME_N) and the functional values by $f_{m,n}$. Then, double Lagrange interpolation is given by

$$g(DL,ME) = \sum_{m=1}^{M}\sum_{n=1}^{N}\Phi_m(DL)\Psi_n(ME)f_{m,n}$$

Where Φ_m and Ψ_n are shape functions given by

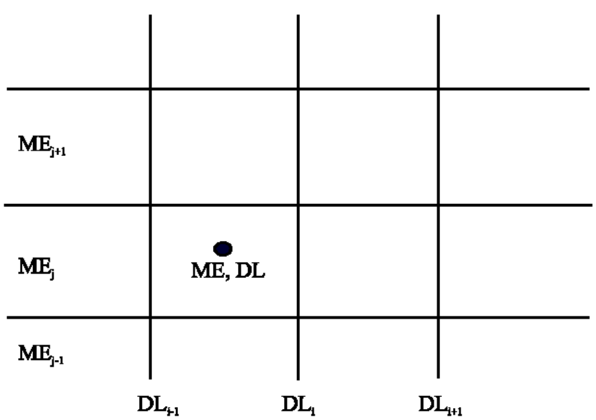

Fig. 1: Rectangular grid

$$\Phi_m(DL) = \prod_{k=1,k\neq m}^{M}\frac{DL - DL_k}{DL_m - DL_k}$$

$$\Psi_n(ME) = \prod_{k=1,k\neq n}^{N}\frac{ME - ME_k}{ME_n - ME_k}$$

Recognize that $\Phi_m(DL)\ \Psi_n(ME)$ is a two dimensional shape function that becomes zero at all the data points except at (DL_m, ME_n).

SIMULATION RESULTS

Applying the two variables (DL and ME) model mentioned above, the effort model surface as a function of DL and ME was obtained as shown Fig. 2.

In Fig. 3, the contours of the E = f(DL, ME) surface are shown.

From these figures, we can note that the effort values are increasing as DL increases. Also, there is a slight decrease in estimated effort with increasing values of methodology. Both of these trends are similar to the results of Shin and Goel's who obtained the same figures

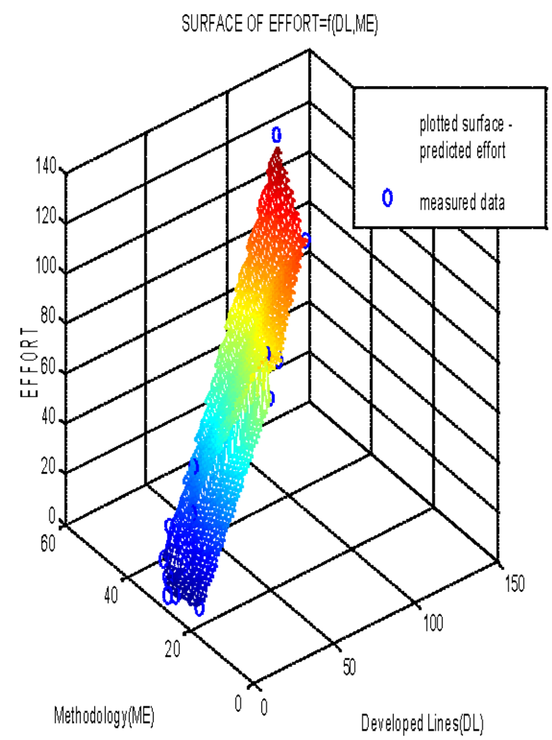

Fig. 2: Surface of effort = f(DL, ME)

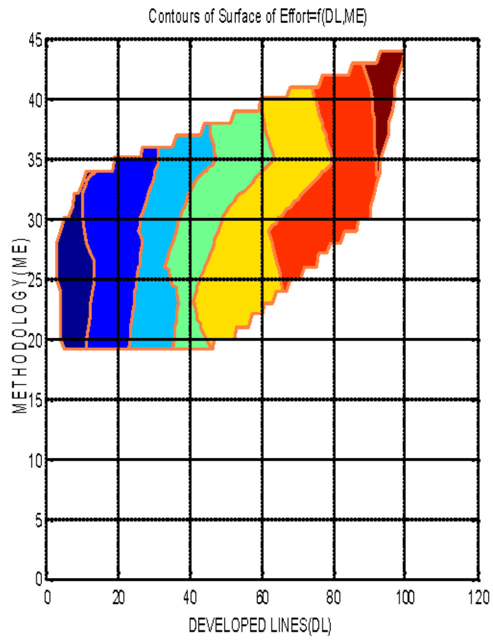

Fig. 3: Contours of surface of effort = f(DL, ME)

using radial basis function networks. The present results are not exactly same as results of Shin and Goel's but similarities can be seen.

CONCLUSIONS AND FUTURE WORK

In this study, I have developed a two variables interpolation model for modeling empirical data in software engineering applications. If the required data is given, then model obtains the E = f (DL, ME) surface and computes the values of this surface in the desireds points.

In future work, a GUI will be developed for the two variables interpolation model.

REFERENCES

Basili, V., 1980. Models and Metrics for Software Management and Engineering. IEEE Computer Society, Press.

Idri, A., T.M. Khoshgoftaar and A. Abran, 2002. Can neural networks be easily interpreted in software cost estimation. World Congress on Cop. Intelligence, Honolulu, May 12-17.

Knafe, G.J, J.A. Morgan, R.L. Follenweider and R.M. Karcich, 1995. Software failure data analysis using the least squares approach and the time per failure concept. Intl. J. Reliability, Quality, Safety Eng., 2: 161-175.

Nakamura, S., 2002. Numerical analysis and graphic visualization with MATLAB, Prentice Hall.

Peters, J.F. and S. Ramanna, 1996. Application of the choquet integral in software cost estimation. IEEE, 2: 862-866.

Pressman, R.S., 1992. Software Engineering, A Practitioner's Approach, Mc Graw Hill.

Shin, M. and A.L. Goel, 2000. Empirical data modeling in software engineering using radial basis functions. IEEE Trans. Software Eng., 26: 567-576