Construction Conceptual Cost Estimates Using Support Vector Machine

Min-Yuan Cheng[1], Yu-Wei Wu[2]

Abstract—**Conceptual cost estimate plays an essential role in project feasibility study. In practice, it is performed based on estimator's experience. However, due to the inaccuracy of cost estimate, budgeting and cost control are planned and executed inefficiently. Support Vector Machines (SVMs), an Artificial Intelligent technique, is used to conduct the construction cost estimate. The algorithms of SVMs solve a convex optimization problem in a relative short time with satisfied accurate solution.**

Applying SVMs, the construction conceptual cost estimate model is developed for owners and planners to predict the construction cost of a project. The impact factors of cost estimate are identified through literature review and interview with experts. The cost data of 29 construction projects are used as training cases. Based on the training results, the average prediction error is less than 10% and the computation time is less than 5 minutes. The error is satisfied for the conceptual cost estimate of a project during the planning and conceptual design phase. Case studies show SVMs can efficiently and accurately assist planners to predict the construction cost.

Index Terms—**Construction Cost, Conceptual Cost Estimate, Support Vector Machines.**

I. Introduction

Construction conceptual cost estimate provides a basis of planners to evaluate the project feasibility in the conceptual planning phase. The impacts of inaccurate cost estimating on project feasibility as well as profitability are significant. Overestimated costs result a low feasibility divesting client to own new projects. On the other hand, an underestimated cost could mislead planners to a high feasibility, which cause client additional costs in the construction phase. Thus, overestimated or underestimated costs affect clients' profits requiring a method to measure as accurate as possible.

The conceptual cost estimate is experience oriented. In conceptual planning phase, cost estimators can only estimate building cost according to preliminary design and project concepts. Under inadequate information circumstance, cost estimators refer to historical cases, and then judge conceptual cost based on their experiences. Nevertheless, building cost is effected by numerous factors. Some of these factors are full of uncertainty such as geological property and decorative class. Due to such complex and uncertain evaluation process, estimators evaluate building cost using a simple linear manner cannot accurately evaluate the costs. As a result, present building cost estimates are rough.

Hsieh(2002) employs the Evolutionary Fuzzy Neural Inference Model (EFNIM) to develop an evolutionary construction conceptual cost estimate model. In the model, Genetic Algorithms are primarily used for optimization; Fuzzy Logic for representing uncertainty and approximate reasoning; and Neural Networks for fuzzy input-output mapping. However the computation run time to search optimal solution takes very long. In order to reduce run time, this study using Support Vector Machine (SVM) to estimate construction cost.

The remainder of the paper is organized as follows: In section 2, we introduce Neural Networks (NNs) and Evolutionary Fuzzy Neural Inference Model (EFNIM). In section 3, We define the regression problem and present our approach using SVMs. In section 4 this study compares prediction accuracy and required effort of the SVMs with EFNIM and NNs. Finally in section 5, we conclude and discuss avenues for future work.

II. Neural Networks and Evolutionary Fuzzy Neural Inference Model

Most problems in construction management are complex, full of uncertainty, and vary with environment. Genetic Algorithms (GAs), Fuzzy Logic (FL), and Neural Networks (NNs) have been successfully applied in construction management to solve various kinds of problems. These three computing methods offset the demerits of one paradigm by the merits of another. Considering the characteristics and merits of each method,(Ko 2002) combines the above three techniques to develop an Evolutionary Fuzzy Neural Inference Model (EFNIM). In the model, GAs is primarily concerned with optimization; FL with imprecision and approximate reasoning; and NNs with learning and curve fitting; Thus, the best adaptation mode is automatically identified.

The architecture of the EFNIM is shown in Fig. 1. The proposed EFNIM is a fusion of FL, NN, and GA paradigms. The combination of FL, NNs, and GAs offset the demerits of one paradigm by the merits of another. In the formulated model, FL is primarily concerned with imprecision and approximate

This work was supported in part by the R.O.C Department of Construction Engineering, National Taiwan University Of Science And Technoloy.

1.Min-Yuan Cheng is with the Department of Construction Engineering, National Taiwan University Of Science And Technoloy, ROC(e-mail: myc@mail.ntust.edu.tw).

2.Yu-Wei Wu is with the Department of Construction Engineering, National Taiwan University Of Science And Technoloy, ROC (e-mail: D9305503@mail.ntust.edu.tw).

reasoning; NN with learning and curve fitting; and GAs with optimization.

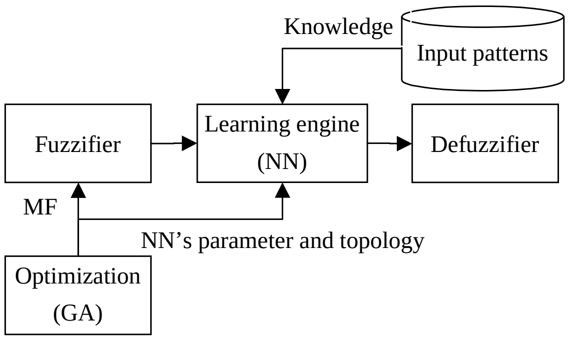

Fig. 1. EFNIM Architecture[1]

FL and NNs are complementary technologies. The combination of these two technologies into an integrated system appears a promising path towards the development of intelligent systems capable of capturing qualities characterizing the human brain. In Fig. 1, fuzzy inference engine and fuzzy rule base in the traditional fuzzy logic system are replaced by the NN. The NN is used to overcome the difficulties in acquisition of fuzzy rules and determination of composition operator and to offer a learning ability to the integrated system. The combination of the FL and NN is regarded as a "neuro with fuzzy input-output," which is also a neural network with both fuzzy inputs and fuzzy outputs. In this work, for convenience to describe the "neuro with fuzzy input-output," it is initialized by the FNN which is a general phrase to express fusion/union of FL and NN.

Although the FNN is more reasonable than traditional FL to simulate the characteristics and process of human inference, the FNN for learning different tasks has demonstrated the difficulty in selecting an appropriate topology as well as appropriate parameters for a network. In addition, the determination of suitable distribution for the MFs, for solving disparate problems is time consuming and the difficulty increases with problem complexity. GA is an effective approach to conquer the drawbacks of FNN. Therefore, the EFNIM employs GA to simultaneously search for the fittest shapes of MFs, optimum FNN topology, and optimum parameters of FNN.

III. SUPPORT VECTOR MACHINES

The theory of support vector machines (SVMs) is a new statistical technique and has drawn much attention on this topic in recent years. This learning theory can be seen as an alternative training technique for polynomial, radial basis function and multi-layer percept classifiers. SVMs are based on the idea of structural risk minimization (SRM) induction principle [2] that aims at minimizing a bound on the generalization error, rather than minimizing the mean square error. In many applications, SVMs have been shown to provide higher performance than traditional learning machines and has been introduced as powerful tools for solving classification and regression problems.

For the classification case, SVMs find a separating hyperplane that maximizes the margin between two classes. Maximizing the margin is a quadratic programming (QP) problem and can be solved from its dual problem by introducing Lagrangian multipliers [2]. In most cases, searching suitable hyperplane in input space is a too restrictive application to be of practical use. The solution to this situation is mapping the input space into a higher dimension feature space and searching the optimal hyperplane in this feature space. Without any knowledge of the mapping, the SVMs find the optimal hyperplane by using the dot product functions in feature space that are called kernels. The solution of the optimal hyperplane can be written as a combination of a few input points that are called support vectors. The equations of regression problems in SVMs are similar with the equations of classification problems except the target variables. By introducing the ε -insensitive loss function and doing some small modifications in the formations of equations, the theory of SVMs can be easily applied into regression problems.

Suppose we are given a set S of training points, (y1; x1),…., (yi; xi). Where $xi \in R^N$ and $yi \in R$ for i = 1,…,i. We wish to find the regression function,

$$f_R(x) = w \cdot x + b \qquad (1)$$

defined by the pair (w, b), where $w \in R^N$ and $b \in R$. SVMs approximate the function with the following characteristics: SVMs define the regression estimation with respect to the ε -insensitive loss function. SVMs minimize the risk based on the structural risk minimization (SRM) principle which is defined by the inequality $\|w\| \le const.$[2] Formally we can write this problem as a convex optimization problem by requiring:

$$\text{Minimize } \frac{1}{2} w \cdot w \qquad (2)$$

$$\text{subject to } \begin{cases} y_i - (w \cdot x_i + b) \le \varepsilon, \\ (w \cdot x_i + b) - y_i \le \varepsilon. \end{cases}$$

Analogously to the soft margin loss function, one can introduce slack variables ξ_i , ξ_i^* to cope with otherwise infeasible constraints of the optimization problem. Hence we arrive at the formulation:

$$\text{Minimize } \frac{1}{2} w \cdot w + C \sum_{i=1}^{l} (\xi_i + \xi_i^*) \qquad (3)$$

$$\text{subject to } \begin{cases} y_i - (w \cdot x_i + b) \le \varepsilon + \xi_i, \\ (w \cdot x_i + b) - y_i \le \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \ge 0. \end{cases}$$

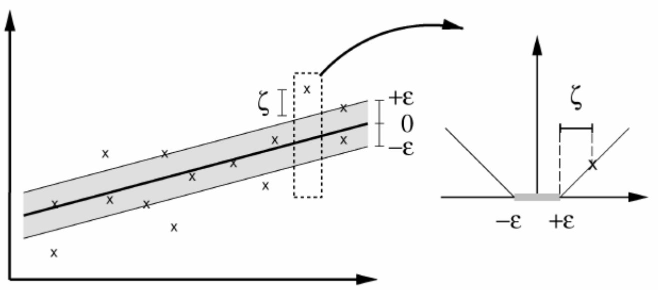

Fig. 2. The soft margin with ε -insensitive loss function.[3]

The constant C > 0 determines the trade off between the flatness of f_R and the amount up to which deviations larger than ε are tolerated. Figure 2 is a description of the situation graphically. The problem (3) can be solved by constructing a Lagrangian,

$$L=\frac{1}{2}\|w\|^2+C\sum_{i=1}^{l}(\xi_i+\xi_i^*)-\sum_{i=1}^{l}\alpha_i(\varepsilon+\xi_i-y_i+w\cdot\Phi(x_i)+b) \\ -\sum_{i=1}^{l}\alpha_i^*(\varepsilon+\xi_i^*+y_i-w\cdot\Phi(x_i)-b)-\sum_{i}^{l}(\eta_i\xi_i+\eta_i^*\xi_i^*) \quad (4)$$

Where $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0$. The solution to the quadratic programming problem is equivalent to determining the saddle point of the (4).At the saddle point, we obtain:

$$\frac{\partial L}{\partial b}=\sum_{i=1}^{l}(\alpha_i^*-\alpha_i)=0 \quad (5)$$

$$\frac{\partial L}{\partial w}=w-\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)x_i=0 \quad (6)$$

$$\frac{\partial L}{\partial \xi_i^*}=C-\alpha_i^*-\eta_i^*=0. \quad (7)$$

Substituting (5), (6) and (7) into the right hand side of (4), we see the problem reduces to

Maximize

$$-\frac{1}{2}\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)x_i\cdot x_j-\varepsilon\sum_{i=1}^{l}(\alpha_i+\alpha_i^*)+\sum_{i=1}^{l}y_i(\alpha_i-\alpha_i^*)$$

$$\text{subject to} \begin{cases} \sum_{i=1}^{l}(\alpha_i-\alpha_i^*)=0, \\ 0\leq\alpha_i^*\leq C, i=1,...,l \end{cases} \quad (8)$$

The Karush-Kuhn-Tucker conditions play an important role and are defined as:

$$\overline{\alpha_i}\left(\varepsilon+\overline{\xi_i}-y_i+\overline{w}\cdot x_i+\overline{b}\right)=0, i=1,...,l \quad (9)$$

$$\overline{\alpha_i^*}\left(\varepsilon+\overline{\xi_i^*}+y_i-\overline{w}\cdot x_i-\overline{b}\right)=0, i=1,...,l \quad (10)$$

$$\left(C-\overline{\alpha_i}\right)\overline{\xi_i}=0, i=1,...,l \quad (11)$$

$$\left(C-\overline{\alpha_i^*}\right)\overline{\xi_i^*}=0, i=1,...,l \quad (12)$$

The scalar $\overline{b}$ can be determined from the Karush-Kuhn-Tucker conditions. Substituting equation (6) into equation (1), the regression function can be rewritten as :

$$f_R(x)=\sum_{i=1}^{l}\left(\overline{\alpha_i}-\overline{\alpha_i^*}\right)x_i\cdot x+\overline{b}. \quad (13)$$

SVMs can process the training points in the feature space by a map $\varphi: R^N \to Z$, and then applying the standard regression algorithm of SVMs. Let $z=\varphi(x)$ denote the corresponding feature space vector with a mapping φ from R^N to a feature space Z. We just only provide a function K(.,.) called kernel which uses the points in input space only, can compute the dot product in feature space Z, that is

$$z_i\cdot z_j=\varphi(x_i)\cdot\varphi(x_j)=k(x_i,x_j) \quad (14)$$

Thus, the nonlinear regression problem can be found as the solution of

Maximize

$$-\frac{1}{2}\sum_{i=1}^{l}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)K(x_i,x_j)-\varepsilon\sum_{i=1}^{l}(\alpha_i+\alpha_i^*)+\sum_{i=1}^{l}y_i(\alpha_i-\alpha_i^*)$$

$$\text{subject to} \begin{cases} \sum_{i=1}^{l}(\alpha_i-\alpha_i^*)=0, \\ 0\leq\alpha_i^*\leq C, i=1,....,l \end{cases} \quad (15)$$

The decision function is then rewritten as

$$f_R(x)=\sum_{i=1}^{l}\left(\overline{\alpha_i}-\overline{\alpha_i^*}\right)K(x_i,x)+\overline{b}. \quad (16)$$

Then we can feed the data (xi) into the decision function and get output data (yi).

IV. MODEL VALIDATION

Hsieh (2002) identified 10 factors influencing building cost in the conceptual phase. These factors can be categorized into two groups: owner's preliminary requirements and site investigations. Table I reveals 26 input patterns and three validation cases collected from Hsieh (2002). These patterns are real collective housing projects located in northern area of Taiwan with reinforced concrete structures from 1997 to 2001.

The performance of SVMs, NNs, and the EFNIM in building cost estimating is compared in Table III. The accuracy of each method is evaluated using the RMSE. As shown in Table 3, Based on the training results, the average prediction error of SVMs is less than 18% and the computation time is less than 5 minutes, the EFNIM excels the rest of the methods and computation time is more than 300 minutes.

According to the literature review, the deviation of present conceptual cost estimating is about 25% (Zhong 1992; Huang 1993). Comparing the estimating model between SVMs and EFNIM, SVMs dramatically promotes the computation speed. The SVMs improves time requirement for developing the solution as well as prediction accuracy for the problem. Thus, planners can apply the SVMs in conceptual planning phase to estimate building cost. According to the inferred building cost, clients can make proper decisions to assess the project feasibility. In addition, the proposed method may assist clients to make various decisions, such as budgeting, tendering and awarding, and financial planning.

TABLE I PATTERNS FOR CONCEPTUAL ESTIMATING OF BUILDING COST

Input patterns												
	Buildin	Input										Output
no.	g cost	1	2	3	4	5	6	7	8	9	10	11
1	19519	872.97	1	67	1	50	6423.93	10	3	2	2	0.4325
2	21370	1646.83	1	0	1	24	6163.21	11	2	2	3	0.5203
3	29499	2168.94	1	0	1	13	6240.83	14	2	2	3	0.9056
4	18631	1777.74	2	150	1	79	8314.86	7	1	1	2	0.3904
5	27844	2869.94	1	46	1	146	17348.25	9	2	1	3	0.8272
6	29731	3756.89	1	25	1	87	5966.23	10	2	3	3	0.9167
7	17041	3018.93	1	0	1	201	21766.78	16	2	2	2	0.3150
8	14129	890.62	1	127	1	12	9545.90	12	3	1	1	0.1770
9	12154	5778.86	1	150	1	227	39390.72	14	3	1	1	0.0833
10	22070	2862.80	1	0	1	78	15240.96	12	2	2	3	0.5535
11	16906	1851.19	1	74	1	182	11910.44	14	3	2	2	0.3086
12	14864	2148.71	2	43	1	46	15659.10	14	2	2	1	0.2118
13	12287	3225.90	1	90	1	128	15900.32	9	2	1	1	0.0896
14	18894	2786.71	2	0	1	133	32888.41	24	4	2	2	0.4029
15	19629	833.04	1	152	1	126	7775.01	14	3	2	2	0.4377
16	19151	1315.67	3	243	1	10	6899.00	6	2	3	2	0.4150
17	25601	2302.02	1	87	1	58	18465.64	14	3	3	3	0.7208
18	23977	806.99	1	33	1	72	7854.34	14	3	3	3	0.6438
19	27083	4375.36	1	0	1	49	14029.39	11	2	3	3	0.7911
20	14713	790.06	1	97	1	32	5571.66	12	2	2	1	0.2046
21	24599	1554.47	1	104	1	194	11872.69	12	2	3	3	0.6733
22	21430	2781.05	1	25	1	57	5966.23	10	1	2	2	0.5231
23	20702	7607.80	1	96	1	236	25861.58	17	3	1	2	0.4886
24	23141	1889.97	1	84	1	196	10888.98	13	2	3	3	0.6042
25	12294	2920.92	1	137	1	134	15500.46	8	2	1	1	0.0900
26	12182	1358.97	2	29	1	80	8146.24	9	2	1	1	0.0847
27	25285	1748.88	1	36	1	86	13736.74	12	2	3	3	0.7059
28	22656	2315.94	1	147	1	83	12034.57	11	2	1	3	0.5812
29	16016	3146.93	2	52	1	144	18691.88	9	2	1	2	0.2664

Note: The captions of above numbered columns are: (1): Site area (in square meters). (2): Geology property. (3): Influencing householder number. (4): Earthquake impact. (5): Planning householder number. (6): Total floor area (in square meters). (7): Floor over ground (in stories). (8): Floor under ground (in stories). (9): Decoration class. (10): Facility class. (11): Normalized building cost. Building cost is in NTD (New Taiwan Dollars) per square meters.

In Table I, (1), (3), (5), (6), (7), and (8) are quantitative factors, whereas columns (2), (4), (9), and (10) are qualitative factors. Qualitative factors are described in Table II.

TABLE II DESCRIPTION OF QUALITATIVE FACTORS FOR CONCEPTUAL COST ESTIMATING

Influencing factor	Qualitative option	Value
Geology property	Soft	1
	Medium	2
	Hard	3
Earthquake impact	Low	1
	High	2
Decoration class	Basic type	1
	Normal type	2
	Luxurious type	3
Facility class	Basic type	1
	Normal type	2
	Luxurious type	3

TABLE III GENERALIZATION COMPARISON FOR CONCEPTUAL ESTIMATING OF BUILDING COST

Pattern no.	Building cost	SVMs estimated cost	NNs estimated cost	EFNIM estimated cost
27	0.7059	0.6639	0.0009	0.6822
28	0.5812	0.7212	0.0009	0.7194
29	0.2664	0.3208	0.0009	0.2543
RMSE		0.09	0.5491	0.0813
Training time		Less than 5 minutes	More than 5 minutes	Above 300 minutes

Note:Real building cost is multiplied by 21093.02 (in NTD/m^2).

V. CONCLUSION

In this thesis, we will summarize the above-mentioned process of study and achievement of study as follows:

(1) This study has proposed a new method to predict the construction cost of a project. Such a model would solve construction conceptual cost estimate problem in a relative short time with satisfied accurate solution.

(2) Through the cost estimate method, the designing authority could calculate the engineering cost based on project contents proposed by customers. Customer could then assess the feasibility based on the proposed project contents and cost from related authority. A concrete project content and estimated cost could be established in time, as well as cost and requirement could fit the need.

(3) With the application of EFNIM, this study enhances the accuracy of price estimate effectively, and decreases the tolerance of rough cost estimate down to ±15%, while for sketchy cost estimate down to within ±10%.

(4) With the application of SVMs, this study predict the building cost efficiently, and training time is less than 5 minutes.

REFERENCES

[1] Chien-HO KO," Evolutionary Fuzzy Neural Inference Model (EFNIM) For Decision-Making In Construction Management," Ph.D. dissertation, Dept. of construction Engineering, National Taiwan University Of Science and Technology, Taipei, Taiwan, 2002.

[2] Chun-Fu Lin," Fuzzy Support Vector Machines," Ph.D. dissertation, Dept. of Electrical Engineering, National Taiwan University, Taipei, Taiwan, 2004..

[3] Wen-Shan Hsieh," Construction Conceptual Cost Estimates Using Evolutionary Fuzzy Neural Inference Model," M.S. thesis, Dept. of construction Engineering, National Taiwan University Of Science and Technology, Taiwan, 2002..

[4] C.-C. Chang, C.-W. Hsu and C.-J. Lin. The analysis of decomposition methods for support vector machines. *IEEE Transactionson Neural Networks*. 11(2000), 1003-1008.

[5] C.-C. Chang and C.-J. Lin. Training nu-Support Vector Classifiers: Theory and Algorithms . *Neural Computation* 13(2001), 2119-2147.

[6] Min-Yuan, Cheng., and Chien-Ho, Ko., "OO evolutionary fuzzy neural inference system for construction management" Journal of Construction Engineering and Management, 2003, ASCE (tentatively approved).

[7] George, Birrell., "Construction Cost Estimating in the Design Process", Journal of the Construction Division, 106,1980, pp.551-566.

[8] Robert, I. Carr., ＆ Fellow, ASCE., "Cost-Estimating Principles", Journal of Construction Engineering and Management, 115, 1989.

[9] Hojjat, Adeli., ＆ Mingyang, Wu., "Regularization Neural Network for Construction Cost Estimation" ,Journal of Construction Engineering and Management,124,1998,pp18-24.

[10] Tarek, Hegazy., Member, ASCE., ＆ Amr, Ayed., "Neural Network Model for Parametric Cost Estimation of Highway Projects" ,Journal of Construction Engineering and Management,124,1998, pp210-218.