KNOWLEDGE-BASED MULTIPLE REGRESSION MODEL FOR COMPLEX ENGINEERING DESIGNS

Bi Zhou*, Masaaki Hoshino**
周碧、星埜正明

*M. Eng., Graduate student, Dept. of Transportation Eng., Nihon University
(24-1, Narashinodai 7, Funabashi, Chiba 274-8501, Japan)
**Member of JSCE, Dr. Eng., Professor, Dept. of Transportation Eng., Nihon University
(24-1, Narashinodai 7, Funabashi, Chiba 274-8501, Japan)

This paper explores the application of the Knowledge-Based Multiple Regression Model (KBMRM) for the complex engineering designs, gives a comprehensive presentation of MRM, develops the measures of model adequacy and puts them use to software. The representation and utilization of different types of complex engineering design models and knowledge acquisition are discussed. In addition to the representation of the knowledge, the introduction purpose of KB and MRM is to prevent the misuse of MRM and to solve the model-building problem. Future direction of the system is suggested for coming development. A sample design of a cable-stayed bridge is presented to further explain the model and the results show the potential application of this model to the complex engineering designs. The model can provide us with a penetrating insight into the situation being analyzed and view the situation as a whole.

Key Words: *multiple regression model, knowledge-based method, engineering designs, structural designs*

1. INTRODUCTION

The engineering design process is a complex and difficult problem to model, involving complicated and sophisticated developing environments, advanced numerical analysis and optimization techniques, engineering knowledge based on experience and heuristics, conformance to governing standards, which themselves are complex. In analyzing complex engineering problems, many mathematical models have been advanced considerably in the past centuries. However, a common feature of some of these models, is the fact that they solve the engineering problems by nature biased toward numerical, rather than descriptive the situation of being analyzed. Admittedly, if a suitable program is available, a solution may be obtained on a digital computer, but it is felt that an engineering model that implicitly brings knowledge from several different domains in the field of application into the search space gives a direct and exact solution of the problem. This is mainly because of the very nature of the engineering design process, which requires engineering judgment, intuition, experience and creative abilities. At the same time, the model with knowledge provides us with a penetrating analysis of the problem being analyzed, it enables us to get a feeling regarding the related factors (*candidate knowledge*) of the different components of the design, and to view the situation as a whole.

(1) Modern Complex Engineering Design Environments

Modern engineering design is known to be the process of complicated and sophisticated developments[1] for which effective solutions depend upon cooperative participation by a number of related factors:

(1) Multi-agent, the engineering design project represents a class of complex synthesis problems for which effective solutions should depend upon cooperative participation by a number of autonomic specialists.

(2) Multi-purpose, hundreds of thousands of extremely complex individual components for particular purpose are assembled together.

(3) Multi-million and multi-year, that requiring knowledgeable interaction among many project control specialists in order to make economical expenses and rational schedules.

(2) Complex Structural Designs

The engineering design is a complex process, involving advanced numerical analysis and optimization techniques, engineering knowledge based on experience and heuristics, and the complexity of the structure itself:

(1) Varied and colorful structural types.

(2) Nonlinear effects, e.g., of cable-stayed bridges, are caused by large displacements, bending moments, interaction of axial forces and shortening of members due to bowing. The principle of the superposition may be applied only with certain limitations.

(3) Differential erection methods, have an important influence on the development of the structural design. To fulfill the mechanical requirements, or for convenience during the erection, segmental methods and temporary works (e.g., temporary false works, hinges, cables) are usually used. Structural system conversion (e.g., hinged to fixed, determinate to indeterminate) or structural secondary effects often occur during the construction.

(4) Time-dependent effects, of creep and shrinkage of concrete and relaxation of prestressing tendons (esp. in segmental erection), or of large displacements in girder and pylons, sag effects in cable stays and anchorage slip loss (e.g., cable-stayed bridges, suspension bridges), are usually complex analyses.

The success of the overall design depends both upon the quality of individual components, and upon the effectiveness with which their efforts can be coordinated and directed toward global design objectives. Before these multi-factor situations, we should neglect the secondary contradiction while dealing with the main one. Therefore, downstream relationships are sometimes previously predicted out according to the knowledge without the benefit of properly jumbled communicated upstream solutions or their justifications.

2 Multiple Regression Models (MRM)

Regression analysis is an effective technique that unifies various practical methods for modeling and investigating the relationship between variables[2) 3) 4) 5) 6) 7) 8) 9) 10) 11)]. Regression models that employ more than one independent variable are called *Multiple Regression models* (MRM). Suppose that there are reasons for if a random variable Y has given probability distribution at a fixed value x of another variable, so that the mathematical expectation is

$$E(Y \mid x) = g(x, \beta), \tag{1}$$

where β is a set of *regression coefficients* determining the function $g(x)$, and that it is required to determine the values of these coefficients from results of observations; x acts as an "independent" variable, is called a *regression* (or *regression function*).

Depending on the nature of the engineering design problem and the aim of analysis, we use a model based on such assumptions: x is a controllable design variable, and the observed design value Y can be written in the form

$$y_i = g(x_i, \beta) + \varepsilon_i, \quad (i=1,2,...,n) \tag{2}$$

where the variables ε_i characterize the *errors*, which are independent and identically distributed with mean zero and constant variance.

Equation (2) may be written in matrix notation as

$$\{Y\} = [X]\{\beta\} + \{\varepsilon\}, \tag{3}$$

or

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \tag{4}$$

where $\{Y\}$ is an $(n\times 1)$ vector, $[X]$ is an $(n\times(k+1))$ matrix (includes multi-degree polynomial), $\{\beta\}$ is a $((k+1)\times 1)$ vector, and $\{\varepsilon\}$ is an $(n\times 1)$ vector. Suppose that $n>k+1$.

To estimate the regression coefficients $\{\beta\}$, we may minimize the *least squares function* L,

$$L = \sum_{i=1}^{n} (y_i - g(x_i))^2 = \sum_{i=1}^{n} \varepsilon_i^2 = \{\varepsilon\}'\{\varepsilon\} = (\{Y\} - [X]\{\beta\})'(\{Y\} - [X]\{\beta\}). \tag{5}$$

The regression coefficients $\{\beta\}$, say $\{\hat{\beta}\}$, must satisfy

$$\left.\frac{\partial L}{\partial \beta}\right|_{\hat{\beta}} = -2[X]'\{Y\} + 2[X]'[X]\{\hat{\beta}\} = 0. \tag{6}$$

Equations (6) are the *least squares normal equations*, and we can obtain the regression coefficients

$$\{\hat{\beta}\} = \left([X]'[X]\right)^{-1}[X]'\{Y\}. \tag{7}$$

When $n < k+1$, the regression coefficients is underdetermined. However, a direct solution can be obtained by using the concept of *pseudoinverse*[12)]. Assume that the rank of matrix $[X]$ is n and define the pseudoinverse of matrix $[X]$, $[X]^*$ thus

$$[X]^* = [X]'\left([X]\,[X]'\right)^{-1} \tag{8}$$

Then the regression coefficients are

$$\{\hat{\beta}\} = [X]^*\{Y\} + [Q]\,[w], \tag{9}$$

where $\{w\}$ is an $((k+1)-n)$ column matrix of arbitrary coefficients and $[Q]$ is an $((k+1)\times((k+1)-n))$ matrix formed from any $((k+1)-n)$ independent columns of the matrix $[R]$ thus

$$[R] = [I] - [X]^*[X]. \tag{10}$$

The predicting function using (9) exactly matches the function at the design pairs for any values of w_i. Thus, non-unique predictions are obtained when predictions are underdetermined.

Finally, the estimated regression model can be written as

$$\{\hat{Y}\} = [X]\{\hat{\beta}\}. \quad (11)$$

In scalar notation, the estimated model is

$$\hat{y}_i = \hat{\beta}_0 + \sum_{j=1}^{k} \hat{\beta}_j x_{ij}. \quad (i=1,2,...,n) \quad (12)$$

(1) Measures of MRM Adequacy

A number of techniques can be used to measure the adequacy of a MRM[4) 11) 13) 14)]. Here, we briefly discuss some of the more effective MRM adequacy measures (detailed explain also refers to examples below).

The *total sum of squares of deviation Syy* is partitioned into a *sum of squares due to regression SSR* and a *sum of squares due to error SSE*, say

$$S_{yy} = SS_R + SS_E = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 + \sum_{i=1}^{n}(\hat{y}_i - y_i)^2, \quad (13)$$

where $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

Therefore, the *multiple correlation coefficient*, which can be used as a measure of the amount of reduction in the variability of y obtained by using the related variables, is

$$M_{CC} = \sqrt{\frac{SS_R}{S_{yy}}}, \quad (14)$$

or the reduction by using *one* related variable with *others* held constant, the *partial correlation coefficient* is

$$P_{CC_i} = \frac{\beta_i}{\sqrt{C_{ii}SS_E + \beta_i^2}}, \quad (i=1,2,...,n) \quad (15)$$

where C_{ii} is the diagonal element of $\left([X]'[X]\right)^{-1}$ corresponding to β_i, and the *standard residual* is

$$S_R = \sqrt{\frac{SS_E}{n-k-1}}. \quad (16)$$

In the current study, *normal distribution*, *F-distribution*, *T-distribution*, and *Beta-distribution* are in the development of MRM.

(2) Prediction of New Observations

MRM can be used to predict future observations on y corresponding to particular values of the independent variables[11) 14)], say $x_{01}, x_{02},..., x_{0k}$. If $[x_0]' = [1, x_{01}, x_{02},..., x_{0k}]$, then a point estimate of the future observation y_0 at the point $x_{01}, x_{02},..., x_{0k}$ is

$$\hat{y}_0 = [x_0]\{\hat{\beta}\}. \quad (17)$$

A $100(1-\alpha)$ percent prediction interval for this future observation is

$$\hat{y}_0 - \psi \le y_0 \le \hat{y}_0 + \psi, \quad (18)$$

where

$$\psi = t_{\alpha/2,n-k-1}\sqrt{\frac{SS_E}{n-k-1}\left(1 + [x_0]'\left([X]'[X]\right)^{-1}[x_0]\right)}, \quad (19)$$

and $t_{\alpha/2,n-k-1}$ is the value of *T-distribution* with n-k-1 degrees of freedom.

(3) Stepwise Multiple Regression models (sMRM)

The *Stepwise Multiple Regression models* (sMRM), can be used as an effective variable selection technique[11)]. The procedure iteratively constructs a sequence of regression models by adding or removing variables at each step. The criterion for adding and removing a variable at any step is usually expressed in terms of a partial F-test. Let F_{in} be the value of the F statistic for adding a variable to the model, and let F_{out} be the value of the F statistic for removing a variable from the model. We must have $F_{in} \geqq F_{out}$, and usually $F_{in}=F_{out}$.

3. KNOWLEDGE-BASED APPROACH

The last few years have witnessed rapid development of knowledge-based systems to address a broad range of tasks in the engineering design domain[1) 15) 16) 17) 18) 19) 20) 21) 22)]. The knowledge-based method is the method that performs its task based on knowledge. The performance of the system increases with more knowledge and knowledge is supposed to be more easily added or modified than a program can be changed to implement a new feature of a system.

As illustrated before, the engineering design is not solely a mathematical process; MRM allows for the representation of aggregation and association relationships between different levels of abstraction of knowledge as variables. Knowledge, that is gained through experience or learned from someone with more experience are often used by us when making approximations for initial designs, optimizing and checking for detailed designs.

(1) KB and MRM

In addition to representing the knowledge for the complex engineering designs, the purpose we introduce KB and MRM is to prevent the misuse of regression analysis and to solve the model-building problems.

Without underlying theoretical considerations and further analysis, regression analysis is frequently misused[11)]. Care should be taken in selecting variables with which to construct models and in determining the form of the approximating function. It is quite possible to develop relationships among variables that are completely unrelated in a practical sense. E.g., one might attempt to relate the internal forces of a determinate structure and the temperature applying on it with the length of the structure. A perfect line may even appear to provide a good fit to the data, but the relationship is an unreasonable one on which to rely. A strong observed association between variables does not necessarily imply that a causal relationship exists between those variables.

Another important problem in model-building is how to select the independent variables which quite likely include all of the important variables, but we are not sure that all of these variables are necessary to adequately model the dependent variable *y*. Usually, We would like the final model to contain enough independent variables so that in the intended use of the model it will perform satisfactorily. On the other hand, to keep the model maintenance costs to a minimum, we would like the model to use as few independent variables as possible. In most problems, there is no single regression model that is "best" in terms of the various evaluation criteria that have been proposed. A great deal of experience and judgment with the system being modeled is usually necessary to select the appropriate independent variables by KB and MRM.

(2) Complex Engineering Design Models

In the engineering design, there are many design variables that are inherently related, and it is necessary to explore the nature of these relationships. FIG. 1 presents the engineering designs focusing on the structural design; the notation is based on Julia[23]. The figure identifies, at a high level of abstraction, the structure and its relationships involved in the engineering design: a successful *structure* must satisfy (be subject to) the *function* specified (do what it was designed to do), the *style* designed and the *cost* given. The *structure* can be considered as an aggregation of structural *members*, in a 1 : many relationship. The *Members*' spatial coordinates are represented by the *style* through the *structure*, and the *members* are assembled by different *erection methods and phases*.

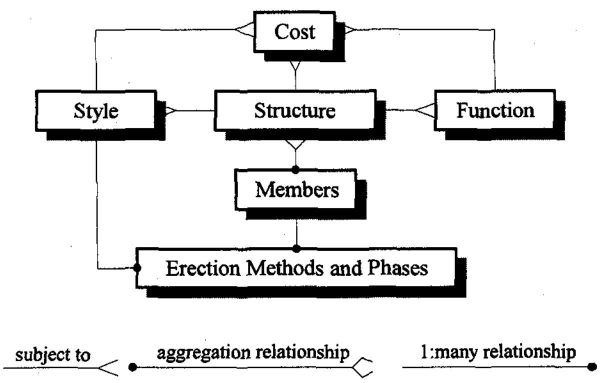

FIG. 1. Structure and Its Relationships

Therefore, the engineering design can be considered as the participants involved from several disciplines that carry out a variety of tasks. Each task (*simple type*, in a broad sense) uses its own relationship between its own sub-tasks to fulfill the task. Certainly, there is no definitive functional relationship between them, but there does seem to be some kind of relation. E.g., the ultimate strength of a prestressed concrete beam may be determined by the properties of all the prestressed bars in the beam and the concrete itself; The mechanical characteristic of a rigid frame bridge is strongly under the influence of the ratio of the side span to the middle span and the ratio of the second moment inertia of beams to columns. A synthetic task (*multiple type*) is a task that reflects shared knowledge between different types of tasks. E.g., in estimating the cost of a structure, we should take into consideration of the cost of every structural component, design and erection cost, erection method, local price index etc. Some of these relationships are discussed as follows:

a) Structure and Structural Members (Simple type)

According to equations (1) and (3), the relationship between the design of the structure D^{str} and the design of the structural members D_j^{meb}, $(j=1,2,...nMembers)$ can be expressed as the expected design of D^{str} for the aggregation of the independent structural members D_j^{meb}

$$E[D^{str}|D_j^{meb}] = [D^{meb}]\{\beta\} . \quad (j=1,2,...,nMembers) \qquad (20)$$

Therefore, the equation (12) becomes

$$\hat{D}_i^{str} = \hat{\beta}_0 + \sum_{j=1}^{nMembers} \hat{\beta}_j D_{ij}^{meb} , \quad (i=1,2,...,nPairs, j=1,2,...,nMembers) \qquad (21)$$

This model describes a hyperplane in the *nMembers*-dimensional space of the independent member designs D_{ij}^{meb}. The coefficient $\hat{\beta}_j$ represents the expected change in response $\hat{D}_i^{str}$ per unit change in D_{ij}^{meb} when all the remaining independent member designs D_{ik}^{meb} $(k \neq j)$ are held constant. We also call $\hat{\beta}_j$ *as partial regression coefficients*, because they describe the partial effect of one independent variable when the other independent variables in the models are held constant.

b) Structural Members and Complex Design Phases (Simple type)

The words "design phases" in this paper includes design phases (e.g. calculation steps) and erection phases. Here, we use the cable-stayed systems below to illustrate the model; they are general enough to apply to any types of structure in any design phase.

During the erection of the cable-stayed systems having radial or parallel cables, there are a few methods of cable adjustment to obtain the most convenient distribution of internal forces and girder displacements under dead load. In the case of cantilever methods of erection, to minimize repeated adjustment of separate cables, it is useful to calculate the required tension and the *sensitivity* of all cables attachments to the stiffening girder during all phases of the erection. Traditionally, the *substructure method*[24] is used to obtain the *sensitivity*; however, it should "remove" all the cables every time to apply unit loads or unit displacements. Recent study of the *influence matrix method*[25] has improved this method to form the *sensitivity matrix* by the concept of the *influence matrix*. However, similar to the *substructure method*, it is difficult to reflect the actual situation during the erection phases systematically.

The expected design of the ith member D_i^{meb} for the aggregation of the independent design phases of the ith member at the jth phase D_{ij}^{phase} is

$$E[D_i^{meb} | D_{ij}^{phase}] = [D_i^{phase}]\{\beta\}\,. \quad (i=1,2,...,nMembers,\ j=1,2,...,nPhases) \qquad (22)$$

So, the equation (12) becomes

$$\hat{D}_{ij}^{meb} = \hat{\beta}_0 + \sum_{k=1}^{nPhases} \hat{\beta}_k D_{ijk}^{phase}\,, \quad (i=1,2,...,nMembers, j=1,2,...,nPairs,\ k=1,2,...,nPhases) \qquad (23)$$

c) Design Points and Related Factors (Multiple Type)

As mentioned before, engineering designs (design points) usually involve participants from several disciplines (related factors) that carry out a variety of tasks. Similar to the *simple type*, the expected value of the ith design point D_i^{point} for the aggregation of the independent related factors of the ith point at the jth factor D_{ij}^{factor} is

$$E[D_i^{point} | D_{ij}^{factor}] = [D_i^{factor}]\{\beta\}\,, \quad (i=1,2,...,nPoints,\ j=1,2,...,nFactors) \qquad (24)$$

and the equation (12) becomes

$$\hat{D}_{ij}^{point} = \hat{\beta}_0 + \sum_{k=1}^{nFactors} \hat{\beta}_k D_{ijk}^{factor}\,, \quad (i=1,2,...,nPoints,\ j=1,2,...,nPairs,\ k=1,2,...,nFactors) \qquad (25)$$

(3) Knowledge Absorption

The complete process of the knowledge engineering that is necessary to the development of MRM requires both knowledge absorption (knowledge collection), knowledge evolution (knowledge evaluation, knowledge modification) and make the knowledge available to software developments and for actual design usage (knowledge representation).

Knowledge absorption is the act or process by putting in action causes or agents over which we have control, and purposely extracting knowledge from a domain expertise:

(1) "*Individual*", we can obtain knowledge from public guides, standards, recorded designs, hypothetical designs or elicited from experts in the field of application, individually.

(2) "*Cooperation*", the design engineering is regarded as in the area of architecture, structure, and construction, where it is common for architectural designs, structural designs, construction designs, and facility maintenance designs to be handled by four unrelated departments. Interaction and feedback among disparate participants often occurs too late in the design process to effect improvements. We should meet the architects for architectural problems, meet the residence engineers or experts who are familiar with the site for information about the construction specification.

(3) "*Public*" is a good way of improving our designs. We could make our designs known to the public, engage ourselves into arguments, reply raised queries, and gain valuable knowledge.

(4) "*Discussion*" is another method, that, our fruits could be approved and improved by the face-to-face panel meetings.

(4) Knowledge Evolution and Representation

The knowledge base (KB) is the brain trust of the engineering design. FIG. 2 depicts the formation of KB. We can consider the knowledge evolution as any modification or any reconstruction in KB, which allows it to perform better the second time on repetition of the same task or another

TABLE 1. Sample Rules

Rule	Condition	Actions	Source
Absorb_influence [6]	Structure IS DETERMINATE	Secondary_generalized_force IS NONE	common sense
Absorb_recencey [1]	Phase[current] < Phase[new]	Knowledge[current] = Knowledge[new]	phase
MRM [1]	Multiple_correlation_coefficient <= Expectation AND		MRM
	Standard_residual >= Expectation	GOTO COLLECTION	MRM
Analysis_cable_Stress [1]	Cable_stress[phase] > Allowable_stress AND		analysis
	Temporary_measure_effect IS POOR	GOTO COLLECTION	site situation
Erection_time_CableBr [14]	6m <= Block_length <= 12m AND		design
	2m <= Block_depth <= 3m AND		design
	68 <= Cable_number <= 124 AND	Erection_time=-0.2550+0.1620	MRM*
	Block IS PRECAST	*Cable_space+0.0008*Cable_number	site situation
Cost_CableBr_[8]	Relation IS Cable_space AND	1. Influence_Cable_space (0.9923) >	MRM*
	Cable_number AND	Influence_Cable_number (0.9920) >	MRM*
	Girder_depth	Influence_Girder_depth (0.6480) AND	MRM*
		2. $\hat{y}_0 - \psi \le \mathrm{Cost}[[x_0],\alpha] \le \hat{y}_0 + \psi$	MRM*
Moment_P_girder_CableBr [5]	Relation IS Cable_space AND	1. Influence_Cable_number (0.9709) AND	MRM*
	Cable_number AND	Influence_Girder_depth (0.7808) >	MRM*
	Girder_depth	Influence_Cable_space (0.0964) >	MRM*
		2. $\hat{y}_0 - \psi \le \mathrm{Cost}[[x_0],\alpha] \le \hat{y}_0 + \psi$	MRM*

MRM* refers to TABLE 2.

task drawn from the same situation. Sometimes, we may not find out until too later that, our absorbed knowledge is not the best, so knowledge absorption must be made based on abroad and deep understanding of the practical situation. In addition, the selection between vast amount of absorbed knowledge at different stages of different levels of abstraction is often different. To examine them along with the changing of the situation, and produce a description of the situation, we should pick out the knowledge that is regarded as the most relevant and the best explanation of the situation. We can consider four criteria:

(1) Influence, has the absorbed knowledge the most directly influence available?

(2) Reliability, how was the knowledge come from? how trustworthy is it?

(3) Recency, is it the latest knowledge available? will new knowledge be available promptly?

(5) Practicality, is the knowledge practicable for the actual situation?

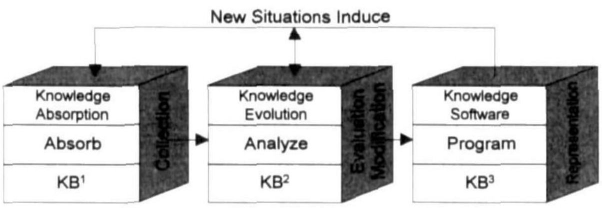

FIG. 2. Formation of KB

It is important that the representation of KB should support inference, decision and learning. We use a number of *temporary views* to look at the design situation from different angles; each angle bringing out different features of the description of the situation.

Firstly, the related factors that are considered relative to the design situation are collected into the system to form the different *temporary views*. This collection may be done for the different cases it is subject to in order to identify possible problems with the design situation, for which the system may choose them to fit the design situation accordingly.

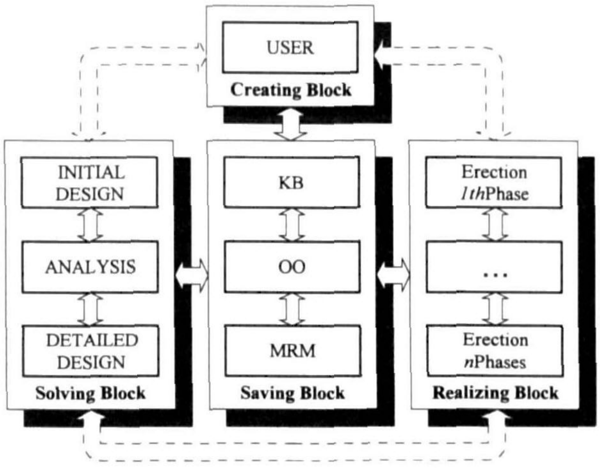

FIG. 3. Engineering Design Focusing on KBMRM

Secondly, MRM is used to analysis the *temporary views* and judges the adequacy of them. Only those related factors that are of relevance or have higher *sensitivity* to the design situation are tend to be reflected in the model. In the current system, we use the *partial correlation coefficient* (*Pcc*) to choose the related factors and the *multiple correlation coefficient* (*Mcc*) to choose the model. For saving the model converging time, we found it is accurate enough if Mcc^2 is greater than 80%.

Thirdly, according to the model chosen by the second stage, if the situation induces, the MRM coefficients are used for overriding (modifying and removing) the existing KB, which aim to represent exceptions or defaults.

The representation could be capable of handing the qualitative knowledge as well as the numerical one. The production rule is of form: IF <CONDITION> THEN <ACTIONS>. Sample rules in the present study are represented at TABLE 1.

The KBMRM representation of both numerical and qualitative knowledge based on rules is clearer and more convenient for designers to make decisions than the representation of only numerical, such as pure neural

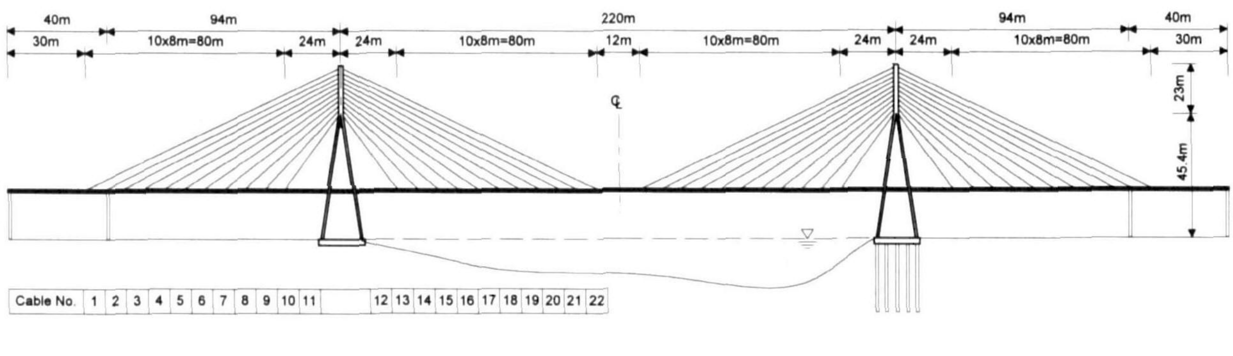

FIG. 4. HuangHe Cable-Stayed Bridge

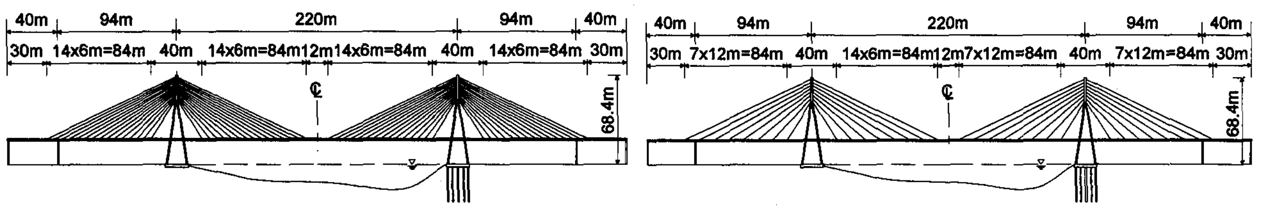

FIG. 5. Plans of Cable Space is 6m and 12m

networks or only qualitative knowledge[22]. Especially, when the knowledge in the form of rules is vague, designers are usually difficult to make any accurate and convincing decision.

4. SOFTWARE PROTOTYPE

FIG. 3 shows an outline of the proposed refined architecture of the engineering design focusing on KBMRM. KBMRM has been used as a *saving block* between the *creating block*, the *solving block* and the *realizing block*; where OO, which is not mentioned in this paper, standards for the object-oriented method[36]. In practice, MRM can be called as a subroutine in FORTRAN or a function in C within any OS circumstances such as Dos or Windows.

As there is no need to use any special software development tool, MRM can become a part of any program written in any procedural language. After the regressing is over, the file containing the relationships between design points and their related factors can be read into any program to suit the machine requirements. In addition to these advantages, the KB system, as we know, has an extremely user-friendly nature, explanation facility, line-of-reasoning and other usual features.

5. DESIGN OF A CABLE-STAYED BRIDGE

The HuangHe Cable-Stayed Bridge[26] (FIG. 4), is a five-span prestressed concrete double-plane cabled-stayed bridge built in China. The total width is 19.5m with two triangle boxes of 2.75m in height, and the cable space is 8m. The erection of the superstructure was performed by the balanced cantilever method, with tensioning cables in pairs simultaneously. FIG. 5 shows other two plans that the cable space is 6m and 12m, respectively. TABLE 2 shows some of the results of the mechanical and the economical comparison of the three plans whose cable space is 6m, 8m and 12m with the variations of cable numbers and girder depth, respectively. Only typical data have been listed in the table. Obviously, KBMRM provides us for establishing the relationships between important control design variables and their related factors with the shared knowledge. E.g. from the table we can build a model that expresses the cost as a function of the three independent variables of the cable space, the cable number (it can also be considered as the stiffness of the cable), and the girder depth. On the other

TABLE 2. Mechanical and Economical Comparison (multiple type)

Design pairs	Cable space (m) (*i=1*)	Cable number (*i=2*)	Girder depth (m) (*i=3*)	Girder moment, axial force area* Moment+ (10^5 t-m²)	Moment- (10^5 t-m²)	Axial force (10^4 t-m)	Erection time** (m/day)	Cost*** kilo-Yuan/m² Actual	MRM	Difference %
1	6	124	2.00	2.89	-3.00	9.65	0.81	7.30	7.17	1.81
2	6	124	2.75	5.79	-5.73	10.96	0.81	7.39	7.32	0.91
3	6	124	3.50	10.25	-10.17	12.44	0.81	7.28	7.48	2.63
4	8	92	2.00	3.24	-3.16	8.05	1.11	6.56	6.51	0.82
5	8	92	2.75	6.45	-6.24	9.11	1.11	6.66	6.66	0.00
6	8	92	3.50	11.60	-10.92	10.31	1.11	6.76	6.81	0.78
7	12	68	2.00	3.80	-4.52	5.30	1.74	6.90	7.19	4.08
8	12	68	2.75	7.60	-8.94	5.90	1.74	7.50	7.35	2.09
9	12	68	3.50	15.06	-15.86	6.53	1.74	7.64	7.50	1.87
MRM (first-degree)										
					{*Pcc*}			SS_R	S_R	*Mcc*
	(*i=1*)	(*i=2*)	(*i=3*)		(*i=1*)	(*i=2*)	(*i=3*)			
Moment+	0.4533	0.0037	5.9955	-13.3628	0.7808	0.0946	0.9709	131.2089	1.2140	0.9731
Moment-	-1.0580	-0.0513	-5.8377	22.4672	-0.9481	-0.7994	-0.9707	136.3069	1.1873	0.9751
Axial force	-0.6939	0.0168	1.3944	9.2583	-0.9884	0.8226	0.9548	46.6331	0.3568	0.9932
Erection time	0.1620	0.0008	0.0000	-0.2550	1.0000	1.0000	0.0000	1.3518	0.0000	1.0000
Cost	0.4737	0.0503	0.2044	-2.3222	0.9923	0.9920	0.6480	1.0531	0.1974	0.9186
MRM prediction										
Moment+	$\psi=1.214t_{\alpha/2,5}(1+[x_0]'([X]'[X])^{-1}[x_0])^{1/2}$					Moment-	$\psi=1.1873t_{\alpha/2,5}(1+[x_0]'([X]'[X])^{-1}[x_0])^{1/2}$			
Axial force	$\psi=0.3568t_{\alpha/2,5}(1+[x_0]'([X]'[X])^{-1}[x_0])^{1/2}$					Erection time	$\psi=0$			
Cost	$\psi=0.1974t_{\alpha/2,5}(1+[x_0]'([X]'[X])^{-1}[x_0])^{1/2}$					Cost([8,92,3]',0.05)	6.14 <= Cost <= 7.27 (α=0.05)			

Note: * include dead load; ** blocks are precast; ***according to the price index in 1991.

hand, build a model that can be used to predict the cost at given values of the three independent variables. Similarly, we can also build a model that expresses the girder internal forces as a function of the three independent variables, which can help us to estimate the quality of necessary tension in the girder with the variations of the three independent variables. The quantity of *Mcc* is also called the *coefficient of determination*, and it is often used to judge the adequacy of a regression model. If our model fitted perfectly, the residuals would all be zero so *Mcc* would be one. Clearly, $0 \leqq Mcc \leqq 1$. We can refer loosely to Mcc^2 as the amount of variability in the data explained or accounted for the model. E.g., in the data of cost, we have $Mcc^2 = 0.9186^2 = 0.8438$, that is, 84.38 percent of the variability in the cost is accounted for by the model when the three independent variables are used. One may, however, be more interested in the degree of relationship between design variables and *one* of the variables with *all* the other variables removed (held constant). Thus, the partial correlation coefficient {*Pcc*} can be observed. E.g., from the table, we can see the girder's depth has a greater influence on the girder's moment area than the cable number does. In the latter design decision, we can omit the influence of the cable number.

In addition to the elicited knowledge listed in TABLE 1, we could also conclude several unconditional rules about this design situation.

The positive girder moment:

(1) the positive girder moment area is in direct ratio to the cable space;

(2) the positive girder moment area is in direct ratio to the cable number;

(3) the positive girder moment area is in direct ratio to the girder depth;

The girder axial force:

(1) the girder axial force area is in inverse ratio to the cable space;

(2) the girder axial force area is in direct ratio to the cable number;

(3) the girder axial force area is in direct ratio to the girder depth;

The erection time:

(1) the erection time is in direct ratio to the cable space;

(2) the erection time is in direct ratio to the cable number;

(3) the erection time has no relation to the girder depth; (due to the site situation)

...

TABLE 3 shows the results of the design points and their cable sensitivity according to the erection phases. The erection phases may be any combination when necessary. In addition, we can establish relationships between the internal forces at the design points and the cable forces to find out the influence of the cable forces to the design points including the information of the erection phases (it can be at any phase). The relationships can then be used to express or predict the internal forces of the design points for given cable forces (such as measured cable forces). This method would also be used for design optimization or design control purpose (such as displacements, stresses, etc.). The values of

TABLE 3. Design Point and Its Cable Sensitivity According to Erection Phases (simple type)

Fabrication or design phases	Cable No.											Design points*	
	1 22 (*i=1*)	2 21 (*i=2*)	3 20 (*i=3*)	4 19 (*i=4*)	5 18 (*i=5*)	6 17 (*i=6*)	7 16 (*i=7*)	8 15 (*i=8*)	9 14 (*i=9*)	10 13 (*i=10*)	11 12 (*i=11*)	Moment (t-m)	Axial force (t)
	Cable forces** (t)												
1											359.67	-80.88	203.40
2										243.57	360.32	227.23	363.10
3									190.41	276.06	394.10	-200.84	539.79
4								176.75	234.61	313.63	422.65	-781.92	746.56
5							185.98	223.93	273.72	342.01	438.69	-1271.41	985.59
6						205.97	230.95	261.85	302.34	359.87	444.93	-1618.35	1254.28
7					229.77	246.59	265.95	289.45	321.43	369.79	445.35	-1844.90	1549.05
8				253.98	265.70	278.14	291.69	308.53	333.40	374.52	442.81	-1986.01	1866.67
9			277.10	285.63	293.95	301.79	310.03	321.24	340.48	376.09	439.02	-2070.85	2204.59
10		298.58	305.13	311.00	315.61	319.13	322.79	329.43	344.26	375.69	435.12	-2120.02	2560.78
11	318.32	323.62	328.06	330.90	331.92	331.59	331.43	334.45	345.97	374.42	431.19	-2147.33	2933.58
MRM (first-degree polynomial)													
	$\{\hat{\beta}\}$											$\hat{\beta}_0$	*SR*
	(*i=1*)	(*i=2*)	(*i=3*)	(*i=4*)	(*i=5*)	(*i=6*)	(*i=7*)	(*i=8*)	(*i=9*)	(*i=10*)	(*i=11*)		
Moment	0.1932	0.2477	0.3240	0.3585	0.2606	-0.2597	-1.8532	-4.7340	-5.1835	1.2240	15.3689	-5608.6261	0.0000
Axial force	0.5151	0.4006	0.1950	-0.0074	-0.1540	0.0592	1.5090	5.1527	7.3754	0.7546	-37.0687	13535.8841	0.0000
sMRM (first-degree polynomial)													
	$\{\hat{\beta}\}$											$\hat{\beta}_0$	*SR*
	(*i=1*)	(*i=2*)	(*i=3*)	(*i=4*)	(*i=5*)	(*i=6*)	(*i=7*)	(*i=8*)	(*i=9*)	(*i=10*)	(*i=11*)		
Moment	0.0000	0.1077	0.0000	0.0000	-0.1237	-0.5772	-1.5757	-2.9606	-2.2730	1.1312	0.0000	-80.6602	8.2822
Axial force	1.2789	0.0000	2.0361	0.0000	0.0000	2.6012	0.0000	0.0000	2.1333	0.0000	0.0000	257.9257	156.0918

Note: * from the left support of 40m+94m=134m at the girder; ** no prestressed forces in cables.

SR show that the sMRM method gives a rougher relationship than MRM does, but it gives a simpler and direct explain of the situation.

6. PROBLEMS

As we know, in most engineering problem, the design variables are relevant, reciprocal and interactive on each other, and have time-dependent effects. Therefore, the problems are followings:

(1) The independent design variable x_i are usually intercorrelated. In situations where this intercorrelation is very large, we find there are serious effects on the general applicability of the model. To detect the presence of this intercorrelation, the *variance inflation factor* for $\hat{\beta}_j$

$$VIF\left(\hat{\beta}_j\right)=\frac{1}{\left(1-Mcc_j^2\right)},\ (j=1,2,...,k) \tag{26}$$

is introduced, where Mcc_j^2 is the *coefficient of multiple determination* resulting from regressing x_j on the other k-1 regressor variables. If any factors exceed 10, we should consider the intercorrelation. Hoerl et al.[27) 28)] have proposed *ridge regression* as an alternative to ordinary least squares. In the *ridge regression*, the regression coefficients are obtained by solving

$$\{\beta^*\}(l)=\left([X]'[X]+(l)[I]\right)^{-1}[X]'\{Y\}, \tag{27}$$

where $l \geqq 0$ is a constant. Generally, values of l in the interval $0 \leqq l \leqq 1$ are appropriate. A good discussion of the practical use of ridge regression has been made by Marquardt et al.[29)].

(2) If the dependent and independent design variables are time-dependent on time series, and if their correlated factors are not included in the model, the assumption of uncorrelated errors is often untenable. We refer this situation as *autocorrelation.* The common test statistic for *autocorrelation* is the *Durbin-Watson statistic*[3)30)]. The only effective remedial measure when *autocorrelation* is present is to build a model that accounts explicitly for the autocorrelative structure of the errors.

7.CONCLUSIONS AND FUTURE DIRECTION

There are three ways to represent the complex engineering designs by using KBMRM:

(1) Functions, which are primarily intended for procedural knowledge

(2) Rules, which are primarily intended for heuristic knowledge based on experience.

(3) Object-oriented programming, which are primarily intended for heuristic knowledge and procedural knowledge.

The present study has taken an approach to the multiple regression generic model for the complex engineering design process by approaching the problem area with the view of knowledge base. Such a generic model has developed that provides for the representation of the engineering knowledge during the design process and has the ability of explanation and problem-solving features. The development of this generic model resulted in the absorption of the relationships between the design points and their related tasks involved in the engineering design. The model is applicable to a wide range of engineering design types by developing the model for specified task(s). From above examples, we can conclude that the KBMRM method shows the potential application for the knowledge-based complex engineering designs, and has a considerable practical value to the process of our structural designs and of optimizing our designs.

The development of the object-oriented programming (OOP)[1) 23) 31) 32) 33)34)35)] approach was beyond the scope of the present paper but would do much improve the representation of knowledge with an *adding-MRM-overriding* method to the new candidate knowledge[36)]. Pure OOP implementations lack the overall control features necessary to guide design progress and maintain solution consistency in conformance with global objectives. MRM is used as a knowledge acquisition tool with provisions of intelligent global data to explain and predict the situation which is analyzed, and the object-oriented paradigm, which allows for the representation of knowledge in a number of different formats, is chosen for the design and implementation of the generic model for engineering design. By the use of the MRM's nature and of the OOP's inheritance as a mechanism for information elision, it leads to a great efficiency of representation, avoiding redundancy of explicit information and predicting for future designs.

Future attention will turn to the integration of modular systems to provide a unified approach to the complex engineering problems which should reflect the design phases and the erection phases, characterized by the need for interactive contributions by the multiple design environments. This area includes:

(1) Developing integrated systems that address the different activities involved in the design, focusing on knowledge sharing between engineering and construction.

(2) Representing and evolving the complex engineering design knowledge that is not necessarily numerical or algorithmic in nature[36)].

(3) Managing and integrating the complexity of the software which used in the different domain and allowing them for maintenance and redevelopment.

ACKNOWLEDGEMENTS

The first writer should greatly appreciate the financial support for this research from Kameda Gumi Co., Ltd., Japan.

REFERENCES

1) David, V. M. and Chris, H., (1991). "Model for Communication in Automated Interactive Engineering Design". *J. Comput. Civ. Engrg., ASCE*, 5(1), pp. 4-23.
2) Balakrishnan, A. George, V., D. and Lotfi, Z., (1965). *Probability, Random Variables, and Stochastic Processes*. Mcgraw-Hill Book Comp.

3) Eric, H. A., and John, E. J., (1977). *Statistical Methods for Social Scientists*. Academic Press Inc., New York.
4) Hald, A., (1952). *Statistical Theory with Engineering Application*. John Wiley & Sons, New York.
5) Harald, C., (1966). *Mathematical Methods of Statistics*. Overseas Publications, Ltd. Tokyo.
6) Hwang, J. N., Lay, S. R., Martin, R. D., Schimert, J., (1994). "Regression Modeling in Back-Propagation and Projection Pursuit Learning". *Translations on Neural Networks, IEEE*, 5(3), pp. 342-353.
7) John, R. W., (1967). *Prediction Analysis*. D. Van Nostrand Company Inc., Toronto.
8) Robb, J. M., (1980). *Aspects of Multivariate Statistical Theory*. John Wiley & Sons, New York.
9) Samuel, S. W., (1962). *Mathematical Statistics*. John Wiley & Sons, New York.
10) Warren G., (1976). *Statistical Forecasting*. John Wiley & Sons, New York.
11) William, W. H. and Douglas, C. M., (1980). *Probability and Statistics in Engineering and Management Science*. John Wiley & Sons, New York.
12) Carpenter, W. C. and Barthelemy, J. F., (1994). "Common Misconceptions About Neural Networks As Approximators". *J. Comput. Civ. Engrg., ASCE*, 8(3), pp. 345-357.
13) Snee, R. D., (1977). "Validation of Regression Models: Methods and Examples". *Technometrics*, 19(4), pp. 415-428.
14) Yi, S., (1993). *Mathematical Regression Methods in Statistical Forecasting*. PhD thesis, TongJi Univ.
15) Arciszewski, T., and Ziarko, W., (1992). "Machine Learning in Knowledge Acquisition". *Knowledge Acquisition in Civil Engineering*, T. Arciszewski and L. Rossman, Eds., ASCE, New York, N.Y.
16) Hohns, M., (1985). "Learning form Failures: Procedural Changes in the Design and Construction Process to Reduce Failures". *Reducing Failures of Engineered Fac.; Proc., NSF/ASCE Workshop, ASCE*, New York, N.Y., pp. 75-83.
17) Krice, G. J., (1996). "IFPATS: A Link Between Distributed AI Systems and Expert Users". *J. Comput. Civ. Engrg., ASCE*, 10(2), pp. 151-156.
18) Lu, S. and Chen, Y., (1987). "A Machine Learning Approach to the Automatic Synthesis of Mechanistic Knowledge for Engineering Decision Making". *Artificial Intelligence in Designing and Manufacturing*, 1(2), pp. 109-118.
19) Maher, M. L., (1992). "Automated Knowledge Acquisition of Preliminary Design Concepts". *Proc., ASCE Conf. for Comp. in Civ. Engrg.*, ASCE. New York, N.Y.
20) Reich, Y. and Fenves, S. J., (1989). "The Potential of machine Learning Techniques in Civil Engineering". *Artificial Intelligence in Engrg. Des. And Manufacturing*. 3(3).
21) Reich, Y. and Fenves, S. J., (1995). "System that Learns to Design Cable-Stayed Bridges". *J. Struct. Engrg.* 12(7), pp. 1090-1100.
22) Skibnieuski, M., Arciszewski, T. and Lueprasert, K., (1997). "Constructability Analysis: Machine Learning Approach". *J. Comput. Civ. Engrg., ASCE*, 11(1), pp. 8-16.
23) Julia, D. B., (1996). "Address Current Issues In Structural Design Software". *J. Comput. Civ. Engrg., ASCE*, 10(4), pp. 286-293.
24) Troitsky, M. S., (1977). *Cable-Stayed Bridge, Theory and Design*. Granada Publishing Ltd.
25) Xiao, R. C., (1991). "Influence Matrix Method for the Adjustment of Internal Forces and Displacements in Control Sectoins". *J. East China Highway*, No.5 pp. 55-60.
26) Shanghai Municipal Engineering Research Institute, (1992). *Evaluation for the HuangHe Cable-Stayed Bridge*. Shanghai, China.
27) Hoerl, A. E. and Kennard, R. W., (1970a). "Ridge Regression: Biased Estimation for Non-Orthogonal Problems". *Technometrics*, V12, pp. 55-67.
28) Hoerl, A. E. and Kennard, R. W., (1970b). "Ridge Regression: Application to Non-Orthogonal Problems". *Technometrics*, V12, pp. 68-82.
29) Marquardt, D. W. and Snee, R. D., (1975). "Ridge Regression in Practice". *The American Statistician*, V29, pp. 3-20.
30) Neter, J. and Wasserman, W., (1974). *Applied Linear Statistical Models*. Richard D. Irwin, Homewood, I11.
31) Bento, J., Feijo, B. and Smith, D. L., (1997). "Engineering Design knowledge Representation Based on Logic and Objects". *Comput & Struct.* 63(5), pp. 1015-1032.
32) Raphael, B. and Kumar, B., (1997). "Object Oriented Representation of Design Case". *J. Comput. & Struct.* 63(4), pp. 663-668.
33) Stumpf, A. L., Ganeshan, R., Chin, S. and Liu, L. Y., (1996). "Object-Oriented Method for Integration Construction Product and process Information". *J. Comput. Civ. Engrg., ASCE*, 10(3), pp. 204-212.
34) Booch, G., (1994). *Object-Oriented Analysis and Design with Applications.* 2nd ed. Addison Wesley Longman, Inc., California.
35) Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F. and Lorengen, W., (1991). *Object-Oriented Modeling and Design.* Prentice-Hall, Inc., New Jersey.
36) Zhou, B. and Hoshino, M., (1998). "OOMRM: Object-Oriented Approach for Multiple Regression Models in the Complex Engineering Representation". *Advances in Engineering Computational Technology. Civil-Comp Press. Edinburgh*, pp. 249-255.

(Received September 7, 1998)