Real-Time Decision-Making with Partial Information for Construction Management

Wen-der Yu, Shao-Shung Lo, and Gang-wei Fan

***Abstract*—Real-time decision-making with partial information is commonplace in the daily tasks of a construction manager/engineer. However, traditional decision support systems (DSSs) do not support partial information inference. Data pre-processing was adopted to solve the data-incompleteness problem. Unfortunately, such approach may be biased. This paper presents a newly developed neuro-fuzzy system, named Variable-attribute Fuzzy Adaptive Logic Control Network (VaFALCON), for decision-making under partial information environment with the untreated original incomplete data. Two case studies with 91.7% and 83.3% attribute information were conducted to test the proposed VaFALCON system. It is found that the average accuracy recovery ratio (AARR) is between 90.57% and 95.68% for testing data with 91.7% partial information, and is between 86.03% and 93.67% for testing data with 83.3% partial information.**

***Index Terms*—neuro-fuzzy system, real-time decision making, partial information, AI.**

I. Introduction

MANY construction management activities rely on real-time decision-making with tools such as decision support system (DSS). Such activities include conceptual cost estimation for emergent construction works, duration estimation for selection of alternative methods, risk assessment of site conditions, etc. However, complete information required for a DSS is hard to acquire due to time constraint in real-time processes. Most traditional DSSs were unable to handle incomplete information in their reasoning processes and thus failed to provide useful assistance for the decision maker before complete information is available. This has severely reduced the usability of the DSS. In this paper, a newly developed VaFALCON [1] approach is proposed to handle partial information for a DSS. The proposed VaFALCON adopts the VANS (variable attribute network structure) [1] method to construct the neuro-fuzzy rule-based system, which is able to handle any combination of available attribute values. Thus, partial attribute information can be fully utilized during the decision-making process. Two real world cases are studied to test the applicability of VaFALCON for handling various degrees of partial attribute information.

The reset o this paper is organized in the following manner: In the second section, previous researches related to partial information decision-making are reviewed; in the third section, the VaFALCON approach with VANS method is described for solving the partial information decision-making problems; in the forth section, two real world cases are tested by VaFALCON with various degrees of partial attribute information; in the fifth section, conclusions are drawn and future directions of research are recommended.

II. Previous Researches Related to Partial Information Decision-Making

In recent years many researchers have tackled the problems of missing attributes values [8], [2], [3], [4] [5]. Among those works, *Rough Set* enables the expression of imprecise data in a precise way by a set of conditional attributes in a data table [4]. The *Rough Set* provides indiscernibility relation for handling missing attributes [6]. In the other approaches, LEM1 and LEM2 methods are proposed to search for a set of decision rules for classification of incomplete attribute values [7]. Moreover, C4.5 method was proposed by Quinlan to generate induction rules from a set of incomplete data [5]. The decomposition method was proposed by another researcher to handle incomplete attribute values [8]. The decomposition method consists of four steps: (1) Decomposition—greedily generating filling patterns that meet certain properties; (2) Splitting data into subtables according to filling patterns; (3) Inducing classifier from subtables; and (4) Integration—Inducing classifier from answers of classifiers based on subtables.

Even though some of the above approaches provided promising results for constructing precise classifiers, the procedure of most above mentioned methods are quite tedious and involved. In some approaches, *a priori* knowledge is required to construct the classifiers. It is desirable to develop a method that takes incomplete attribute values as complete data in the learning (or training) and inference processes of a DSS.

This work was financially supported by the National Science Council of Taiwan Government under Grant NSC 93-2211-E-216-012. Sincere appreciations are given from the authors to the sponsor

W. D. Yu, Associate Professor, Institute of Construction Management, Chung Hua University. Hsinhu, 30012, Taiwan (phone:+886-3-5186748; fax: +886-3-5370517;(e-mail: wenderyu@ chu.edu.tw).

S. S. Lo, Ph.D. Student, Department of Technology Management, Chung Hua University. Hsinhu, 30012, Taiwan (e-mail: Kareh075@ms17.hinet.net).

G. W. Fan, Master Student, Institute of Construction Management, Chung Hua University. Hsinhu, 30012, Taiwan (e-mail: me080kimo@yahoo.com.tw).

III. VaFALCON AND VANS

A. Structure of VaFALCON

Variable-attribute Fuzzy Adaptive Logic Control Network (VaFALCON) [1] is a neuro fuzzy system that evolved from the Fuzzy Adaptive Logic Control Network (FALCON) originally developed by Lin and Lee [9]. Fig. 1 shows the generic structure of a FALCON model. A standard FALCON consists of five layers. Each layer consists of nodes with proper numbers of fan-in and fan-out connections represented by weights assigned to the nodes. The fan-in connections connect the nodes of the previous layer with the nodes of the current layer. The fan-out connections connect nodes of the current layer with nodes of the subsequent layer. An integration function is associated with the fan-in connections of a node. The integration function can be a summation, activation, or fuzzy operation. There are five layers in a FALCON: (1) Layer 1—input layer, which takes input attribute values from outside world; (2) Layer 2—input term nodes, which fuzzify the input attribute values into fuzzy variables; (3) Layer 3—rule nodes, which perform fuzzy AND operations; (4) Layer 4—output term nodes, which defuzzify the fuzzy functions concluded by the rule nodes; and Layer 5—output layer, which presents the system output to outside world.

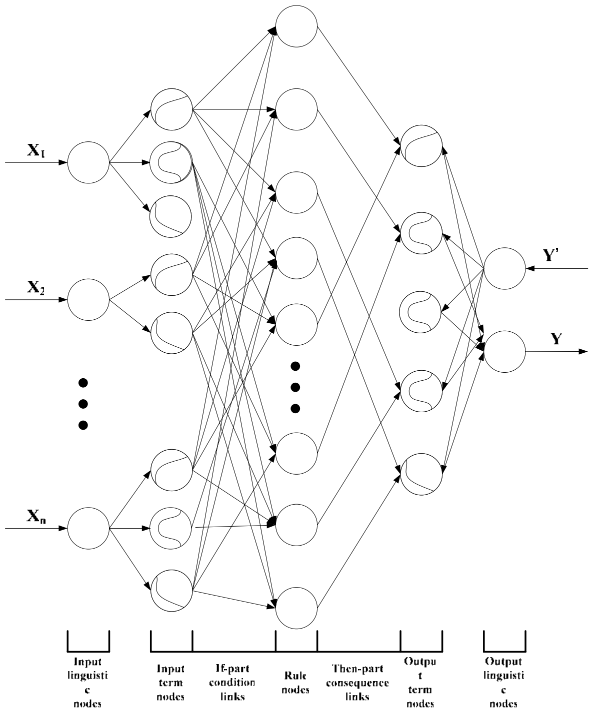

Fig. 1. A generic FALCON model (adapted from [9])

B. Variable-attribute network structure (VANS)

Similar to other neuro fuzzy systems, the traditional FALCON accepts only data with complete attribute values. Any missing attribute value will cause trouble in performing *Fuzzy AND* operation in Layer 3 of FALCON. Further propagations can not proceed consequently, and thus the system output can not be derived from the network at Layer 5.

In order to improve this problem, a Variable Attribute Network Structure (VANS) is proposed by Yu and Lin [1]. VANS adopts a flexible network structure that can adjust to the available attribute values in processing every single input dataset. Considered the FALCON in Fig. 2, where input attribute *a* is missing. The rule nodes connecting to fuzzy terms of attribute *a* are prohibited from further propagation.

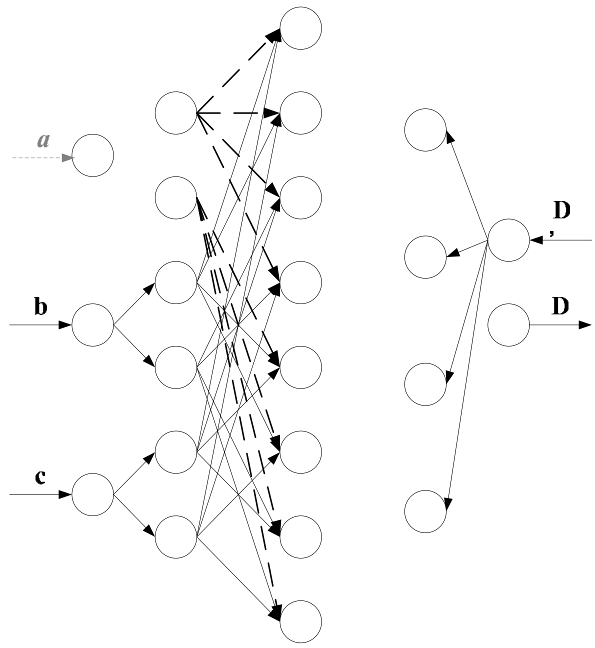

Fig. 2. Connections of FALCON for incomplete attribute vlues l ([1])

The main idea of VANS is to ignore the attributes with missing values. Thus the network of Fig. 2 degrades to the one depicted in Fig 3, where the attribute *a* (with missing attribute value) and its associated fuzzy term nodes are deleted from the network. The resulted FALCON contains only two (*b* and *c*) input attributes. The signal propagation process of the degraded network follows the rules of original FALCON.

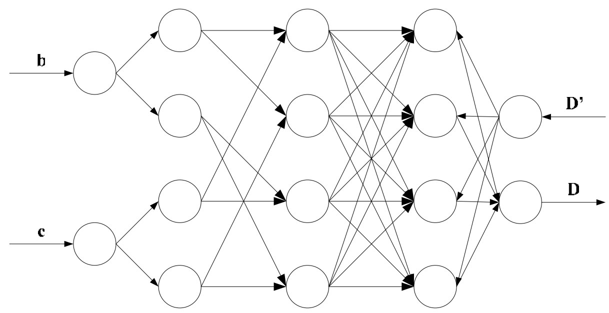

Fig. 3. Degraded FALCON for incomplete attribute vlues l ([1])

VANS is achieved with a special treatment of the missing attribute value by substituting them with values higher than 1.0 after fuzzification operation of Layer 2 in FALCON. The special treatment avoids missing attribute values to influence the results of *fuzzy AND* operation. The missing attributes are “invisible” to the network, and thus they are ignored during the propagation process. With VANS, a VaFALCON is constructed, which is able to process any combination of available attribute values in the network learning and fuzzy inference algorithms of the traditional FALCON. The function of handling variable attribute values provides VaFALCON capability for processing missing value problem.

IV. CASE STUDIES

This section presents two real world cases of partial information decision-making problems in construction management. There are four input attributes for the data of both cases. Totally, 4 (attributes)×3 (testing datasets) = 12 attribute fields are considered in each testing experiment. Two scenarios are designed to test the three cases: (1) decision-making with 91.7% information (one missing attribute value); (2) decision-making with 88.3% information (two missing attribute values). In Scenario I, one value is taken away from the 12 attribute fields every time. The residual attribute information equals to $(1-\frac{1}{12})\times 100\% = 91.7\%$ In Scenario II, two values are taken away from the 12 attribute fields every time. The residual attribute information equals to $(1-\frac{2}{12})\times 100\% = 83.37\%$. The system estimation accuracy is calculated by Equation (1).

$$Acc(\%) = \left\{1 - Abs\left(1 - \frac{Estimated}{Actual}\right)\right\} \times 100\% \qquad (1)$$

In Equation (1), *Estimated* is the output generated by the system, *Actual* is the actual result observed from real world, and *Acc.* is the percentage accuracy of the estimation. The absolute value is taken within the parenthesis to avoid minus values.

A. Case I—Building Construction Cost Estimation

Building construction cost estimation is a difficult task during the early stage of a construction project as most design information is not available at that stage. Traditional approaches rely on domain experts (experienced cost estimators) in performing the conceptual cost estimation. However, the domain experts are difficult to find, expensive to educate, and likely to leave. An example from Yu [10] is selected for case study. In the selected example, 4 attributes were identified as attributes among the nearly 30 parameters originally collected, including (1) retaining wall type (RWT); (2) No. of floors above ground (F); (3) No. of floors under ground (SF); (4) total floor area (A). One single output, construction cost estimation (TWD), is recorded in the database. Totally 25 data are collected from historical building construction project by surveying the final project reports provided by public owners. 22 data sets are used for learning and the rest 3 data are used for testing. The Training sets are shown in Table I. The complete testing sets are shown in Table

TABLE I
TRAINING DATASETS OF CASE I

ID	Retaining wall type (RWT)	Floor (F)	Sub-floor (SF)	Area (A) (m^2)	Total cost (TC) (TWD)
1	3	7	2	2959	43459663
2	3	12	2	7449	164000000
3	3	13	2	15178	255260938
4	2	6	1	918	15769572
5	2	7	1	1502	22722088
6	2	7	1	1721	32171609
7	2	12	1	4518	77150600
8	2	16	2	27866	407150000
9	2	20	3	38255	525000000
10	1	4	1	2630	28500000
11	1	6	1	2958	36154899
12	1	8	2	3855	47714078
13	1	8	1	7316	89872402
14	1	12	2	8331	114884225
15	1	12	2	8351	122814797
16	1	12	1	9396	122923137
17	1	12	2	10810	173313000
18	1	12	1	20993	329966802
19	1	14	2	31513	533289382
20	1	14	2	32955	557683814
21	1	12	1	13989	185164666
22	2	12	1	5560	95530000

RWT: 1= steel rail pile, 2= replace aggregate method, 3= curtain wall method.

TABLE II
COMPLETE TESTING DATASETS OF CASE I

ID	Retaining wall type (RWT)	Floor (F)	Sub-floor (SF)	Area (A) (m^2)	Total cost (TC) (TWD)
T.1	3	7	2	3318	46696021
T.2	1	12	1	28059	434390623
T.3	3	6	2	3223	50394716

RWT: 1= steel rail pile, 2= replace aggregate method, 3= curtain wall method..

II.

Before testing the partial information scenarios, the complete testing datasets (see Table II) are tested. The testing result recorded is 96.63%. Then, two testing scenarios are applied to the three testing datasets shown in Table II. The testing result for Scenario I is shown in Table III, where the values in the row of “Rank” represent the influential ranking of each attribute on the testing accuracy. It is found from Table III that the information of “No. of floors under ground”(SF) influences the testing accuracy most significantly. The overall average testing accuracy for Scenario I is 87.52% in Case I.

TABLE III
TESTING RESULTS FOR ONE MISSING ATTRIBUTE VALUE IN CASE I

ID	Missing attribute			
	RWT	F	SF	A
T.1	0.9301	0.9047	0.7420	0.9663
T.2	0.7977	0.7967	0.9871	0.8563
T.3	0.9321	0.9067	0.7165	0.9663
Average	0.8866	0.8694	0.8152	0.9296
Rank	3	2	1	4

For Scenario II, combinations of missing values in any two attribute fields are tested. The cross testing results of missing attribute values in Scenario II are shown in Fig. 4 at the end of this paper. The testing results are shown in Table IV. It is found that missing values in both "No. of floors above ground (F)" and "No. of floors under ground (SF)" influence the testing accuracy most significantly. These two attributes are also the top two most significant attributes found in Table III. The overall testing accuracy for Scenario II is 80.33%.

By testing with scenario I and II, the most significant influential attributes that affecting the accuracy of estimation for building construction cost are identified. Moreover, it was found that the overall testing accuracy for Scenario I is 87.52% and for Scenario II is 80.33%, vs. the 96.63% testing accuracy for complete attribute information.

TABLE IV
TESTING RESULTS FOR TWO MISSING ATTRIBUTE VALUES OF CASE I

Combination of missing attributes	Testing accuracy	Rank
RWT+F	0.7987	4
RWT+SF	0.7725	2
RWT+A	0.8607	6
F+SF	0.7442	1
F+A	0.8526	5
SF+A	0.7911	3
Overall average accuracy	0.8033	

B. *Case II—Building Construction Cost Estimation*

Curtain wall method has been widely adopted in urban construction projects in Taiwan. Social costs can be very high under inappropriate management practice. Therefore, the accurate duration estimation of such works is important for effective project planning and management in the crowed and congested urban construction sites. An example from Yang [11] is selected for case study. In the selected example, totally 27 historical datasets were collected from major consultant firms of Taiwan. Among which 24 are used for training and 3 are used for testing. The input attributes identified by Yang are: (1) excavation depth (m); (2) quantity of walls; (3) construction method; and (4) soil type. Two qualitative attributes are transformed into numeric values: (1) construction methods— 1 means ML method, 2 represents MHL; (2) the soil type—Clayey as 1, Sandy-clayey as 2, Sandy as 3, Sandy-gravel as 4, Gravel as 5, and Clayey-gravel as 6. The Training sets are shown in Table V. The complete testing sets are shown in Table VI.

TABLE V
TRAINING DATASETS OF CASE II

ID	Dig depth (DD) (m)	No. of units (NU)	Method (M)	Soil type (S)	Duration (D) (day)
1	3	7	2	2959	43459663
2	3	12	2	7449	164000000
3	3	13	2	15178	255260938
4	2	6	1	918	15769572
5	2	7	1	1502	22722088
6	2	7	1	1721	32171609
7	2	12	1	4518	77150600
8	2	16	2	27866	407150000
9	2	20	3	38255	525000000
10	1	4	1	2630	28500000
11	1	6	1	2958	36154899
12	1	8	2	3855	47714078
13	1	8	1	7316	89872402
14	1	12	2	8331	114884225
15	1	12	2	8351	122814797
16	1	12	1	9396	122923137
17	1	12	2	10810	173313000
18	1	12	1	20993	329966802
19	1	14	2	31513	533289382
20	1	14	2	32955	557683814
21	1	12	1	13989	185164666
22	2	12	1	5560	95530000

M: 1= ML method, 2= 2 represents MHL.
S: 1= Clayey, 2= Sandy-clayey, 3= Sandy, 4= Sandy-gravel, 5= Gravel, Clayey-gravel.

TABLE VI
COMPLETE TESTING DATASETS OF CASE I

ID	Dig depth (DD) (m)	No. of units (NU)	Method (M)	Soil type (S)	Duration (D) (day)
T.1	3	7	2	3318	46696021
T.2	1	12	1	28059	434390623
T.3	3	6	2	3223	50394716

M: 1= ML method, 2= 2 represents MHL.
S: 1= Clayey, 2= Sandy-clayey, 3= Sandy, 4= Sandy-gravel, 5= Gravel, Clayey-gravel.

At first, the complete datasets (see Table II) are tested. The testing accuracy is 91.84%. Then, two testing scenarios are applied to the three datasets shown in Table VI. The testing result for Scenario I is shown in Table VII. From the values in the row of "Rank", it is found from that the attribute information of "Soil type"(S) influences the testing accuracy most significantly. The overall average testing accuracy for Scenario I is 87.88% in Case II.

TABLE VII
TESTING RESULTS FOR ONE MISSING ATTRIBUTE VALUE IN CASE II

ID	Missing attribute			
	DD	NU	M	S
T.1	0.9552	0.9151	0.9124	0.8715
T.2	0.9011	0.8968	0.9176	0.8449
T.3	0.8591	0.8451	0.7888	0.8376
Average	0.9051	0.8857	0.8729	0.8513
Rank	4	3	2	1

The cross testing results of missing attribute values in Scenario II are shown in Fig. 5. For Scenario II, the testing results are shown in Table VIII. It is found that missing values in both "soil type (S)" and "construction method (M)" influence the testing accuracy most significantly. These two attributes are also the top two ranked most significant attribute found in Table III. The overall testing accuracy for Scenario II is 86.02%.

TABLE VIII
TESTING RESULTS FOR TWO MISSING ATTRIBUTE VALUES FOR CASE II

Combination of missing attributes	Testing accuracy	Rank
DD+NU	0.8867	6
DD+M	0.8797	5
DD+S	0.8568	4
NU+M	0.8543	2
NU+S	0.8545	2
M+S	0.8294	1
Overall average accuracy	0.8602	

C. *Findings from Case Studies*

It can be seen from the results of the testing cases that VaFALCON is able to recover knowledge from partial information. In order to evaluate the capability of knowledge recovery for VaFALCON, consider the knowledge recovery ratio (*KRR*) defined in Equation (1).

$$KRR_{\alpha\%}(\%) = \frac{Acc_{\alpha\%}}{Acc_{100\%}} \qquad (2)$$

In Equation (2), $KRR_{\alpha\%}$ (%) means the knowledge recovery ratio (in percentage) for α % attribute information; $Acc_{\alpha\%}$ is system testing accuracy of α % attribute information defined in Equation (1); $Acc_{100\%}$ is system testing accuracy of complete attribute information.

The testing result of the two case studies show that VaFALCON is capable of providing decision-making support under partial information environment. It is found that $KRR_{91.7\%}$ is 90.57% in Case I and 95.68% n Case II; $KRR_{83.3\%}$ is 86.03% in Case I and 93.67% in Case II. Even though the performance of VaFALCON in knowledge recovery of partial attribute information is case sensitive (performance is much better in Case II than in Case I), the knowledge recovery of the two cases in either scenario is no less than the available information.

The second finding from the two case studies is that the most influential attributes for estimation accuracy can be identified by system testing with incomplete datasets. In case I, "No. of floors above ground (F)" and "No. of floors under ground (SF)" are the top two most influential attributes. The combination of these two missing attribute values also influences the estimation accuracy most significantly in Scenario II. Similarly, In case II, "soil type (S)" and "construction method (M)" are the top two most influential attributes. The combination of these two missing attribute values also influences the estimation accuracy most significantly in Scenario II. By identifying the most significant attributes for a DSS, the user can focus his/her efforts on collecting the most important information for decision-making. As a result, the efficiency of decision-making can be improved.

V. 5. CONCLUSIONS AND FUTURE WORK

Decision-making in construction management is usually time-constrained. Partial information decision-making is commonplace for construction managers/engineers. However, the traditional decision support systems (DSSs) require the user to provide complete attribute values. This is not realistic in real world applications. This research proposes a VaFALCON neuro fuzzy system for handling partial attribute information in decision-making for construction management. The VaFALCON adopts VANS flexible network structure that is able to learn and inference with any combination of input attributes. As a result, datasets with missing attribute values can be utilized as the complete datasets to make the most use of the available information.

From two case studies, it is found that the proposed VaFALCON is able to recover 90.57% to 95.68% system accuracy from 91.7% attribute information, and recover 86.03% to 93.67% system accuracy from 83.3% attribute information. Moreover, the most influential attributes for estimation accuracy can be identified by testing of VaFALCON. Therefore, the user can focus his/her efforts on collecting the most important information for decision-making. As a result, the efficiency of decision-making can be improved.

Due to time constraint, only two scenarios (91.7% and 83.3% attribute information) are tested. More experiments on other degrees of partial attribute information should be performed in future work. Effect of partial information in both training datasets and testing datasets should also be researched in the future to see the cross impact of missing values.

Acknowledgment

The authors sincerely appreciate the sponsor, National Science Council of Taiwan Government, for providing financial support for this research.

References

[1] W. D. Yu and H. W. Lin, "A VaFACON neuro fuzzy system for mining of incomplete construction databases," *Automation in Construction*, 2005, 13 pp. (in press)

[2] N. H. Bshouty and D. K. Wilson, "On learning in the presence of unspecified attribute values," *Proceedings of the 12th Annual Conference on Computational Learning Theory, COLT'99*, pp. 81-87, ACM, 1999.

[3] S. A. Goldman, S. Kwek, and S. D. Scott, "Learning from examples with unspecified attribute values," *Proceedings of the 10th Annual Conference on Computational Learning Theory,* pp. 231-242, ACM, 1997.

[4] M. Kryszkiewicz, "Properties of incomplete information systems in the framework of rough sets," *In L. Polkowski and A. Skowron, editors, Rough Sets in Data Mining and Knowledge Discovery*, pp. 422-450, Physica-Verlag, 1998.

[5] J. R. Quninlan, "Unknown attribute values in induction," *In A. M. Segre, editor, Proceedings of the Sixth International Machine Learning Workshop*, pp. 31-37, Morgan Kaufmann, 1980.

[6] J. Stefanowski and A. Tsoukiàs "Incomplete information tables and rough classification," *International Journal of Computational Intelligence*, Vol. 17, No. 3, pp. 545-566, 2001.

[7] J. W.Grzymala-Busse and M. Hu, "A comparison of several approaches to missing attribute values in data mining," *Rough Sets, and Current Trends in Computing, Second International Conference, RSCTC2000 Banff, Canada, October 16-19, 2000, Revised Papers* (W. Ziarko, Y. Y. Yao, Eds.), Springer, 2001.

[8] R. Latkowski, "On decomposition for incomplete data," Fundamenta Informaticae, Vol. 54, No. 1, pp. 1-16, 2003.

[9] C. T. Lin and C. S. G. Lee, "Neural-network-based fuzzy logic control and decision system," *IEEE Transactions on Computers*, Vol. 40, No. 12, pp. 1320-1336, 1991.

[10] J. S. Yu, "Developing building cost estimating system using case-based reasoning approach," *Master Thesis*, Department of Civil Engineering, National Central University, Chungli, Taiwan, R.O.C., 2001. (in Chinese)

[11] J. B. Yang, "An integrated knowledge acquisition and problem solving model for experience-oriented problems in construction management," *Dissertation in Partial Fulfillment of Requirements for Degree of Ph.D.*, National Central University, Chungli, Taiwan, R.O.C., 1997.

		T.1				**T.2**				**T.3**			
		RWT	F	SF	A	RWT	F	SF	A	RWT	F	SF	A
T.1	RWT												
	F	0.7706											
	SF	0.7365	0.7175										
	A	0.9301	0.8946	0.7144									
T.2	RWT	0.7615	0.7361	0.5734	0.7977								
	F	0.7605	0.7351	0.5724	0.7967	0.9069							
	SF	0.9509	0.9255	0.7628	0.9871	0.9871	0.9509						
	A	0.8201	0.7947	0.6320	0.8563	0.7841	0.8860	0.9709					
T.3	RWT	0.8959	0.8705	0.7079	0.9321	0.7635	0.7626	0.9529	0.8221				
	F	0.8705	0.8451	0.6824	0.9067	0.7381	0.7371	0.9275	0.7967	0.7726			
	SF	0.6803	0.6549	0.4922	0.7165	0.5479	0.5469	0.7373	0.6065	0.8153	0.7195		
	A	0.9301	0.9047	0.7420	0.9663	0.7977	0.7967	0.9871	0.8563	0.9321	0.8966	0.7630	

Fig.4. Cross testing of 83.3% (two missing attributes) information for Case I

		T.1				T.2				T.3			
		DD	NU	M	S	DD	NU	M	S	DD	NU	M	S
T.1	DD												
	NU	0.9380											
	M	0.9545	0.9071										
	S	0.9134	0.9071	0.8234									
T.2	DD	0.9379	0.8978	0.8763	0.8542								
	NU	0.9336	0.8935	0.8720	0.8500	0.9255							
	M	0.9544	0.9143	0.8928	0.8708	0.9027	0.8970						
	S	0.8817	0.8417	0.8201	0.7981	0.9255	0.9157	0.9225					
T.3	DD	0.8959	0.8558	0.9053	0.8122	0.8418	0.8375	0.8583	0.7856				
	NU	0.8819	0.8419	0.8392	0.7983	0.8278	0.8236	0.8444	0.7717	0.8823			
	M	0.8256	0.8464	0.7829	0.7420	0.7715	0.7673	0.7881	0.7154	0.8683	0.8008		
	S	0.8744	0.8804	0.8316	0.7907	0.8203	0.8160	0.8368	0.7641	0.8442	0.9099	0.9017	
	Ave	0.9083	0.8786	0.8493	0.8145	0.8593	0.8429	0.8500	0.7592	0.8649	0.8554	0.9017	

Fig.5. Cross testing of 83.3% (two missing attributes) information for Case II