Proceedings of ICAD2004
The Third International Conference on Axiomatic Design
Seoul – June 21-24, 2004
ICAD-2004-21

A STUDY ON SOLVING MULTI-GROUP CLASSIFICATION PROBLEMS

Kyung Soo Lee
exeter@yonsei.ac.kr
Graduate School
Department of Mechanical Engineering
Yonsei University
134 Shinchon-Dong, Seodaemun-Gu
Seoul, Korea 120-749

Sung Woon Cha
swcha@yonsei.ac.kr
School of Mechanical Engineering
Yonsei University
134 Shinchon-Dong, Seodaemun-Gu
Seoul, Korea 120-749

Dong Wook Lee
dongwook@yonsei.ac.kr
Graduate School
Department of Mechanical Engineering
Yonsei University
134 Shinchon-Dong, Seodaemun-Gu
Seoul, Korea 120-749

ABSTRACT

Solving multi-group classification problems has been improved by overcoming the limit of conventional statistical methods supported by development of artificial intelligence methods. But a number of studies based on various methods are still going on in many academic fields. This paper presented a new method by applying Set approach based on Axiomatic design to Pairwise Comparison Method which accelerates discriminations at multi-group classification problems. For verification, evaluating bond rating is applied to Neuro-Genetic model, and a new object function is retrieved to overcome the difference of the number of data which may occur at the Pairwise Comparison. At last, validity of this method is verified by comparing the result of new method with the result of preceding studies.

Keywords: Pairwise Comparison, classification, Neuro-Genetic model, set approach design, Credit rating

1 INTRODUCTION

Multi-group classification requires highly sophisticated expert knowledge compared to Pairwise Comparison Method. Consequently, Multivariate Discriminant Analysis, a statistical method, has been long used. Since late 90's, with active research into artificial intelligence, applying those results into solution method has led to current status. Multi-group classification Analysis reveal that compared to the conventional statistical method, models utilizing artificial intelligence such as Neuro-Genetic Model and CBR have resulted in superior results not only in muti-group classification problem, but also in pairwise comparison method. However, artificial intelligence method has its own setbacks including selection of input variables and understanding of the model[8].

This study aims to present the modified pairwise comparison method to enhance discriminations at multi-group classification problems. By applying set approach concept based on Axiomatic design to pairwise comparison method, design matrix is derived. We set out to solve the problem by further applying Neuro-Genetic method. As multi-group classification problem is not confined to specific studies but rather is faced by overall studies, verifying its general applicability is important. Therefore, to verify its universality, we first apply axiomatic design which has universality in various design problems to corporate bond rating issues. We then set out to test applicability to the multi-objective function problems.

2 MULTI-GROUP CLASSIFICATION MODEL

2.1 BREAKDOWN OF MULTI-GROUP CLASSIFICATION PROBLEM

Multi-group classification problem can be further broken down into twofold. The criterion for the breakdown is the group's order. In determining credit rating or credit analysis, each group has its own orders. On the other hand, more general problems including the function classification, the type/pattern classification and the product classification are independent of the orders. Other branches of studies also include both types of multi-group classification problems. However, overall order-independent problems are more prevalent.

2.2 CONVENTIONAL MULTI-GROUP CLASSIFICATION MODEL

In business management studies, multi-group classification model is most often called for in problems related to credit rating and credit analysis. In determining credit rating, methods utilizing the artificial intelligence have been most frequently studied and there are active researches are being proceeded in the methods using CBR, neuro-genetics, and hybrids[6,7]. Particularly, to enhance discriminative power, studies use Ordinal Pairwise Partioning to improve on the existing model[1] as well as combining neuro-genetic algorithm and case-based inference[6]. In engineering studies, various types of researches are being proceeded including pattern classification based on the artificial intelligence. However, more focus has been given to optimization studies rather than classification problems. Therefore, researches are being pursued in the artificial intelligence model and statistical method including CCD(Central Composite Design) based on design of experiments.

3 THEORETICAL BACKGROUND AND APPLICATION

3.1 OVERVIEW OF AXIOMATIC DESIGN AND SET APPROACH

Defining basic rule of design as independent axiom and information axiom, the axiomatic design is an approach in assisting creative process in design by providing scientific ground[2,3]. Particularly, by showing inter-relationship between the factors in designing through design matrix, it assists in achieving more accurate design process. Set approach is essentially based on the basic concepts of axiomatic design. By using grouping of variables, it aims to solve over-design problem, by which appropriate design matrix is derived. In addition grouping provides various design ideas and problem solving method in rational and systematic ways[4]. Another words, set design by set approach is objective methodology from the axiomatic design's perspective in trying to solve problems rising in optimum design.

3.2 PROCESS FOR MODIFIED PAIRWISE COMPARISON METHOD

Figure 1 shows the methodological process for this study.

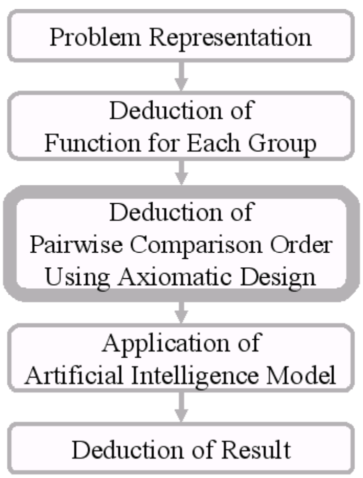


Figure 1. Process of new pairwise comparison method

Step 1 : Defining the problem
Define input variables, grouping and restraints in multi-group classification problem.

Step 2 : Deduction of function for each group
Using various methodologies, select meaningful input variables and deduct functions for each groupings.

Step 3 : Deduction of pairwise comparison order
Using axiomatic design and set approach, deduct pairwise comparison order.

Step 4 : Application of artificial intelligence model
Select the most appropriate artificial intelligence model for the given problem.

Step 5 : Deduction of result
Through the proposed methodology, solve the multi-group classification problem.

4 MODEL DEVELOPMENT AND APPLICATION

4.1 PROBLEM AND DEFINITION: DATA AND VARIABLES

Multi-group classification problem used in the study is a problem of credit rating classification. Credit rating data from National Information and Credit Evaluation, Inc for the period from 1991 to 1995 are used. By eliminating variables with data missing, total of 3832 data are applied to the suggested methodology. For the verification, randomly selected data, 30% from each rating, are utilized. For input variables, 12 input variables suggested by Shin(1999) are used to compare with results from other studies[6]. In addition, to apply to the neuro-genetic model[5], deduction of objective function with number of data for each group in mind.

Table 1. Number of companies in each rating

Ratings	Company #	%
A1	253 (177 / 76)	6.6
A2	819 (573 / 246)	21.4
A3	1296 (907 / 389)	33.8
B	1391 (974 / 417)	36.3
C	73 (51 / 22)	1.9
Sum	3832	100.0

Table 2. Name of variables (Kyung-shik Shin, 1998)

Variables	Name
x1	Firm classification by group types
x2	Firm types
x3	Total assets
x4	Stockholders' equity
x5	Sales
x6	Year after founded
x7	Gross profit to sales
x8	Net cash flow trends for 3 years
x9	Financial expenses to sales
x10	Dependence for liability
x11	Depreciation to total expenses
x12	Working capital turnover

4.2 DEDUCTION OF FUNCTION BY EACH GROUPING

To select meaningful input variables, data are classified in 5 classes for pairwise comparison. Class 1 is classified by A1 and A2&A3&B&C; Class2 by A2 and A1&A3 &B&C; Class3 by A3 and A1&A2&B&C; Class4 by B and A1&A2&A3&C; Class5 by C and A1 & A2 & A3 & B. Then, weights for each class are sought by using back propagation neural network model with 12 input nodes and 13 hidden nodes and 1 output node. Based on the calculated weights, meaningful factors are deducted by comparing each class's weights. Following deduction of functions for each groups, equation (1) is deducted

A1 = F(x1, x3, x4, x5, x6, x9, x10, x11, x12)
A2 = F(x1, x4, x5, x7, x8, x10, x12)
A3 = F(x1,x2,,x3,x4,x5,x6,x7,x8,x9,x10,x11,x12)
B = F(x3, x9, x12)
C = F(*) (1)

Group C's data consist only 1.9% of the total data, which made it impossible to find relevant variables. Therefore, group C are not functioned. In deducting pairwise comparison, it is deemed appropriate to apply it to the last classification process considering the small numbers.

4.3 DEDUCTION OF FUNCTION BY EACH GROUPING

Figure 2 depicts Venn diagram of equation (1) for application to set approach design.

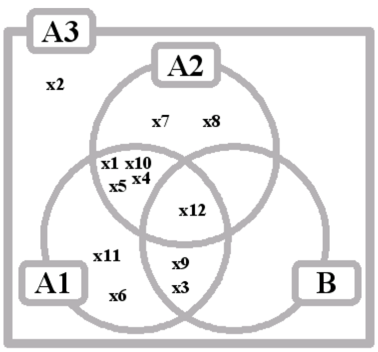
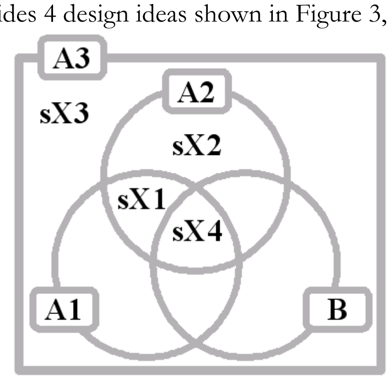


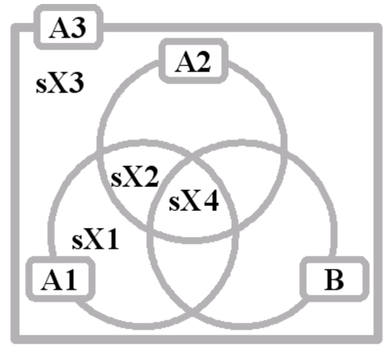
Figure 2. Classification of variables to each group

Grouping of variables based on the completed Venn diagram provides 4 design ideas shown in Figure 3, 4, 5 and 6.



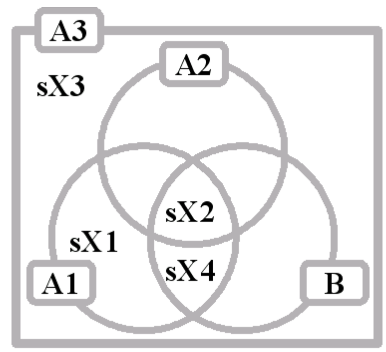
where, sX1 = {x1, x4, x5, x6, x10, x11}
sX2 = {x7, x8 }
sX3 = {x2}
sX4 = {x3, x9, x12}

Figure 3. The first design solution



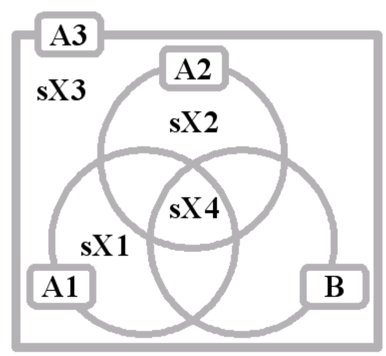
where, sX1 = {x6, x11}
sX2 = {x1, x4, x5, x7, x8, x10 }
sX3 = {x2}
sX4 = {x3, x9, x12}

Figure 4. The second design solution



where, sX1 = {x6, x11}
sX2 = {x1, x4, x5, x7, x8, x10, x12}
sX3 = {x2}
sX4 = {x3, x9}

Figure 5. The third design solution



where, sX1 = {x6, x11}
sX2 = { x7, x8}
sX3 = {x2}
sX4 = { x1, x3, x4, x5, x9, x10, x12}

Figure 6. The forth design solution

This design idea are expressed in axiomatic design equation as equation (2) ~ (5).

$$\begin{Bmatrix} A3 \\ A2 \\ A1 \\ B \end{Bmatrix} = \begin{bmatrix} \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & 0 & \times & \times \\ 0 & 0 & 0 & \times \end{bmatrix} \begin{Bmatrix} sX3 \\ sX2 \\ sX1 \\ sX4 \end{Bmatrix} \tag{2}$$

$$\begin{Bmatrix} A3 \\ A1 \\ A2 \\ B \end{Bmatrix} = \begin{bmatrix} \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & 0 & \times & \times \\ 0 & 0 & 0 & \times \end{bmatrix} \begin{Bmatrix} sX3 \\ sX1 \\ sX2 \\ sX4 \end{Bmatrix} \tag{3}$$

$$\begin{Bmatrix} A3 \\ A1 \\ B \\ A2 \end{Bmatrix} = \begin{bmatrix} \times & \times & \times & \times \\ 0 & \times & \times & \times \\ 0 & 0 & \times & \times \\ 0 & 0 & 0 & \times \end{bmatrix} \begin{Bmatrix} sX3 \\ sX1 \\ sX4 \\ sX2 \end{Bmatrix} \tag{4}$$

$$\begin{Bmatrix} A3 \\ A1 \\ A2 \\ B \end{Bmatrix} = \begin{bmatrix} \times & \times & \times & \times \\ 0 & \times & 0 & \times \\ 0 & 0 & \times & \times \\ 0 & 0 & 0 & \times \end{bmatrix} \begin{Bmatrix} sX3 \\ sX1 \\ sX2 \\ sX4 \end{Bmatrix} \tag{5}$$

However, it should be noted that number of data for each class differ as many as 20 times. When using pairwise comparison in particular, too much discrepancy in number of data present a lot of problems. Therefore, it is imperative to select a design with number of data in mind. Practically, tests A3 (1296) and C (73) using artificial intelligence reveals under 50% of hit rate, which shows degrading discriminating power with too much difference in the number of data. Consequently, it is appropriate to pairwise compare A1 as the last process as it contains relatively fewer data. Based on the comparison, new design equation is deducted excluding A1 and C class.

$$\begin{Bmatrix} A3 \\ A2 \\ B \end{Bmatrix} = \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \end{bmatrix} \begin{Bmatrix} sX3 \\ sX2 \\ sX4 \end{Bmatrix} \tag{6}$$

$$\begin{Bmatrix} A3 \\ B \\ A2 \end{Bmatrix} = \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \end{bmatrix} \begin{Bmatrix} sX3 \\ sX4 \\ sX2 \end{Bmatrix} \tag{7}$$

Equation (6) confirms the order of B → A2 → A3→ A1 & C whereas equation (7) confirms A2 → B → A3→ A1 & C. By applying simple Neuro-Genetic model reveals that pairwise comparison by equation (6) has 4% higher hit rate than when equation (7) is used. As a result, model application only follows equation (6).

4.4 APPLICATION OF ARTIFICIAL INTELLIGENCE : NEURO-GENETIC METHOD

4.4.1 SUMMARY OF MODEL APPLICATION

By the deducted pairwise comparison order, data are reclassified into 4 classes which are shown in Table 3.

Table 3 Classification of data by pairwise comparison order

Class	Classification	Target
Class 1	B & A1, A2, A3, C	B
Class 2	A2 & A1, A3, C	A2
Class 3	A3 & A1, C	A3
Class 4	A1 & C	A1, C

The table shows that it reclassifies investment grades groups (A1, A2, A3) and speculative grades groups(B,C). This is very similar to grouping by expert groups.

Reclassified data shown in table 3 are applied to Neuro-Genetic model and target groups are classified. The reason for using Neuro-Genetic model is due to the fact that there are big differences in the number of data for each group, thus requiring classification model with number of data consideration. To do so, it is appropriate to use a method which enables the model designer to determine appropriate objective function. Therefore, artificial intelligence model with proven record of credit rating and neuro-genetic model which incorporates designer's ideas are combined.

4.4.2 APPLICATION OF MODEL

First of all, the artificial intelligence model uses aforementioned 3-layer perceptron and number of nodes in the 3-layer are the same. The most important process in the neuro-genetic model is for the neuro-genetic algorithm to determine the objective function which is able to find the most appropriate. The objective function used in the study is shown in the equation (8).

Minimize
ABS(Hiting Rate of Class - Hiting Rate of Target)
And
Maximize
(Hiting Rate of Class + Hiting Rate of Target) (8)

The rationale for the equation (8) is that even with differing number of data in each class, it is able to maintain its universality by making both classes' hit ratios similar. Table 4 shows the result of test using neuro-genetic method.

Table 4 Test of object function

Class		Data #	General (%)	Specific (%)
A1 & A3	A1	253(16.3%)	74.7	84.2
	A3	1296(83.7%)	96.5	88.8
	Total	1549	93.0	88.1
A2 & A3	A2	819(38.7%)	57.6	71.6
	A3	1296(61.3%)	86.9	72.7
	Total	2115	75.6	72.3
B & A3	B	1391(51.8%)	76.7	76.7
	A3	1296(48.2%)	76.5	76.5
	Total	2687	76.6	76.6

General objective function has generally higher hit rate. However, with higher discrepancies in the number of data resulted in differing degree of hit rates. On the contrary, objective function by equation (8) minimizes discrepancies among groups. The hit rate is not far behind general objective function. Therefore, peculiarity due to the differences in the group can be overcome with universality by using this equation.

4.4.3 RESULT OF MODEL APPLICATION AND ANALYSIS

Table 5 shows results by applying new pairwise comparison method using neuro-genetic model to classes in the table 3.

Table 5 Classification accuracy rates by new pairwise comparison method

Group	Training (%)	Validation (%)
A1	100	99
A2	67.4	67.1
A3	80.5	84.8
B	80.1	76.7
C	100	91
Total	79.2	79.1

Overall discriminative power is confirmed to be superior. However, A1 and C from Class 4 show 100% of accuracy. We view the result is due to the lack of data numbers, thus making neuro-genetic model inappropriate. Excluding Class 4, other classes show 77.3% of accuracy rate. To compare with conventional OPP method, OPP1 backward method[1] which has shown the most superior result, is used. The results show the new model brings better results (Table 6).

Table 7 compares results from other literature on the credit rating and the results from this study. The accuracy rate is higher when the new method is applied.

Table 6 Classification accuracy rates by OPP 1 backward method

Group	Accuracy (%)
A1	72.7
A2	76.0
A3	71.7
B	75.5
C	68.5
Total	74.0

Table 7 Performance comparison studies of compared with this study

	Year	Data #	Method	Accuracy
Kwon (1997)	1991 ~1993	3085	OPP 1 backward	73.6 %
Shin (1997)	1991 ~1995	2651	CBR	63 %
Shin (1999)	1991 ~1995	3886	GA-CBR	75.5 %
Park (2000)	1991 ~1995	3822	HMUR	68.15 %
This study	1991 ~1995	3832	New OPP 2	79.2 %

This confirms the new pairwise comparison model is superior to the OPP1-backward. The major determinants in the difference result from determining the pairwise comparison orders. Therefore, determining pairwise comparison orders based on axiomatic design provides more rational and systematic approach when compared to the conventional pairwise comparison method.

5 CONCLUSION

Multi-group classification problem has been improved along with the developments of artificial intelligence. However, there are further researches being done to improve the methods in various academic fields.

This study presents the new pairwise comparison method by combining axiomatic design and set approach. We applied the new method in determining pairwise comparison order and utilized neuro-genetic model in order to enhance accuracy. The model is verified by using credit rating problem and the results are compared to the previous studies to compare the accuracy. Results confirm that versatilities and universalities of axiomatic design. In particular, new objective function is derived and evaluated to test convergence and generalization in the neuro-genetic model.

Credit rating problem used in the study falls into multi-group classification problem with orders. In applying the model, pairwise comparison orders resulted in similar grouping with experts grouping. However, we view that the model is more appropriate for the multi-group classification problem with order-independence rather than problem with orders. Therefore, it requires further research into using axiomatic design to the order-independent multi-classification problems.

Furthermore further study can be pursued in the methodology to determine appropriate model and objective functions with number of data in consideration, which is deemed to be critical in pairwise comparison.

6 REFERENCES

[1] Kwon, Y. S., Han, I. G., & Lee, K. C., "Ordinal pairwise partitioning (OPP) approach to neural networks training in bond rating," *Intelligent Systems in Acounting Finance and Management*, Vol. 6, pp. 23–40, 1997.

[2] Suh N.P., *The Principles of Design*, New York: Oxford University Press, 1990. ISBN 0-19-504345-6

[3] Nam P. Suh, "Axiomatic Design : Advances and Applications", The Oxford University Press, 2001.

[4] Lee, K.S., Cha, S. W., "A Study on Design Method through Set Approach", Axiomatic Design Class Materials, 2002.

[5] Shin, K.S., Shin, T.S., Han, I, "Neuro-genetic Approach for Bankruptcy Prediction : A Comparison to Back-propagation Algorithms," *KORMS*, pp. 599-608, 1998.

[6] Kyung-shik, Ingoo Han, "Case-based Reasoning Supported by Genetic Algorithms for Corporate Bond Rating," *Int'l Journal of Expert Systems with Applications*, Vol.16, No.2, pp.85-95, 1999.

[7] Park, K.N., Lee, H.Y., Park, S.K., "A Hybrid Credit Rating System using Rough Set Theory", *Korean Management Science*, Vol. 25, No. 3, pp. 125~135, Sep. 2000.

[8] Hong, Seung-Hyun, Shin, Kyung-Shik, "Using GA based Input Selection Method for Artificial Neural Network Modeling : Application to Bankruptcy Prediction", *Journal of Korea Intelligent Information Systems Society*, Vol. 9, No. 1, pp. 227~249, June 2003.