Canonical Correlation Analysis between Road Environmental Conditions and Traffic Accident Severity

So Young Sohn
Professor
Computer Science & Industrial Systems Engineering
Yonsei University
Seodaemoon-Gu, Shinchon-Dong134
Seoul, Korea
Fax: +82-02-364-7807
E-mail: sohns@yonsei.ac.kr

Hyung Won Shin
Graduate Student
Computer Science & Industrial Systems Engineering
Yonsei University
Seodaemoon-Gu, Shinchon-Dong 134
Seoul, Korea
Fax: +82-02-364-7807
E-mail: won3@yonsei.ac.kr

Abstract: We analyze the traffic accident records (TAR) data in an attempt to model the relationship between the road environmental conditions and the severity of traffic accidents. First we examine the quality of TAR with respect to variables related to road facility, which can cause traffic accidents. As a result, it is recommended that the data collection system has to be revised in a way to better reflect the condition of road facility such as quality of road surface, signpost, and light, etc. Next, canonical correlation analysis is used to identify the relationship between a set of variables which describe the road environmental conditions (road type, width, curve, and traffic signal), and the severity of accidents (death, major injury, minor injury, report, no injury). It is found that road curve is highly related to fatal accidents. Additionally, the condition of a traffic signal appears to play an important role in decreasing the fatal accidents.
Key Words: Traffic Accidents, Data Quality, Canonical Correlation, Quantification Theory II, TASF

1. INTRODUCTION

The rapidly increased amount of traffic in Korea in 1990s has caused many safety problems. In order to reduce the number of such traffic accidents, it is necessary to characterize the causes of accidents. For that purpose, controlled experiments can be performed under several scenarios representing various road traffic conditions. However, such experiments can be expensive. Another approach to identify accident severity related factors is to utilize retrospective data available such as TAR (Traffic Accident Records). TAR typically contains information related not only to the accident severity, but also to potentially influential factors in the accidents such as road type, traffic mode, weather condition, driver characteristics, and vehicle conditions. The main features of TAR in Korea can be summarized as follows: (1) TAR is a collection of each road traffic accident reported annually. Therefore the amount of data is large. (2) Numerous factors in TAR are recorded in the form of categorical variables with various levels. Therefore there are ample chances of multivariate data exploration. However, what has been done by the authorizing organization is limited to few publications of simple analysis of contingency tables of sets of two or three categorical variables. Recent attempt to utilize TAR includes application of data mining and fusion algorithms by Clarke, Forsyth, & Wright (1998), Sohn & Shin (1999), Sohn & Lee (2000), and. Sohn & Shin (2001) tried categorical dimension reduction technique to select crucial information on the pattern extraction for the accident severity and developed both neural network and decision tree models. In an effort to improve the accuracy of individual classification models, Sohn & Lee (2000) proposed fusion algorithms to combine the individual results for road traffic accident data. Classification is not the only tool to find the pattern from the large amount of data. Other forms of multivariate analyses including clustering and association are the alternatives to data mining. Factor analysis is one of the most frequently applied multivariate accident data analysis tool. In view of the categorical nature of TAR, Oppe (1992) showed how SAS programs such as PRINCALS and CANALS could be applied to categorical multivariate road accident data analyses in place of principal component analysis and canonical correlation analysis, respectively. However, quantification theory II that corresponds to categorical canonical analysis has scarcely applied to identify the related factors to the severity of road traffic accident.

In this paper, categorical canonical correlation analysis is employed to understand the association of road traffic accident severity with groups of other categorical accident related variables. Canonical correlation analysis is used to two groups of variables, which we believe to have some underlying correlation. When these two groups of variable are categorical, one can apply Quantification theory II.

Organization of this paper is as follows. In section 2, contents and accuracy of road condition in TAR data are explained. In section 3, canonical correlation analysis is briefly reviewed and applied to TAR data. Section 4 contains conclusion and some remarks on the results of our canonical correlation analysis.

2. TAR in KOREA

In Korea, TAR is filled in for each road traffic accident reported to the police station that has jurisdiction over the place where the accident occurred. TAR consists not only of the multiple-choice categorical columns to represent the characteristics of accident, but also of blank spaces to describe the situation concerning an accident. In order to manage traffic accidents information in a database form, TAR is then summarized into a Traffic Accident Statistics Form (TASF). The TASF consists of two (Main and Supplementary) forms. The main form is for the first and second persons involved in the accident. Supplementary form is necessary for the third party or an additional person involved in the accident. The main form of the TASF is completed for each accident where the first 50 columns are basic information concerned with the characteristics of accident. Additional columns 51-54 are needed for an injury accident, and columns 55-69 are needed to collect information regarding a fatal accident. The remaining columns (77-79) are filled in for a highway traffic accident. A brief sketch of the TASF is given in Table 1 along with the names of factors and their levels[6][8].

Procedures involved in the accident data transfer (TAR->TASF->DB), however, are not simple. First of all, high quality of TAR is essential, because it is used as a basis of any further related database and comprehensive analysis. However, often, much important information is missing from TAR, which makes further analysis less valuable.

Due to many other responsibilities of policeman, typically part-time personnel who may not have sufficient training would do data coding for TASF from TAR, which is manually done first before going through computer input. One of the difficulties that unskilled part-time personnel may have is the fact that many items in TAR and TASF do not have one-to-one corresponding relationship. Additionally many factors have multi-level categories and all of these categories are not displayed in the limited form of TASF. The arrangement of TAR and TASF is not necessarily the same that when one works on coding job for TASF based on TAR, one has to look around the TAR back and forth several times to finish coding. As described, there are many problems to be corrected in order to achieve better quality of accident data related information.
A careful examination of TASF reveals many other data quality related problems. For instance, we analyze the sample data obtained from three police stations in Seoul, Korea (Cheongryangri, Seudaemun, and Songpa) which include a total of 11564 accidents reported in 1996.
As displayed in Table 2 and Figure. 1, most frequently found road condition at the time of accident is observed to be 'Not related to road condition' (code 600) and (code 400) which is not even included in the pre-defined category, comprising a total of 98.2%. In view of the degree of importance of Road Condition factor in analyzing the accident, TASF requires this factor have 27 categories. However, TAR does not have these many levels of category and coding person has to imagine which category the corresponding road condition of accident would fall by reading descriptions of accident in TAR.

Table 1. Traffic Accident Statistics Form in Korea

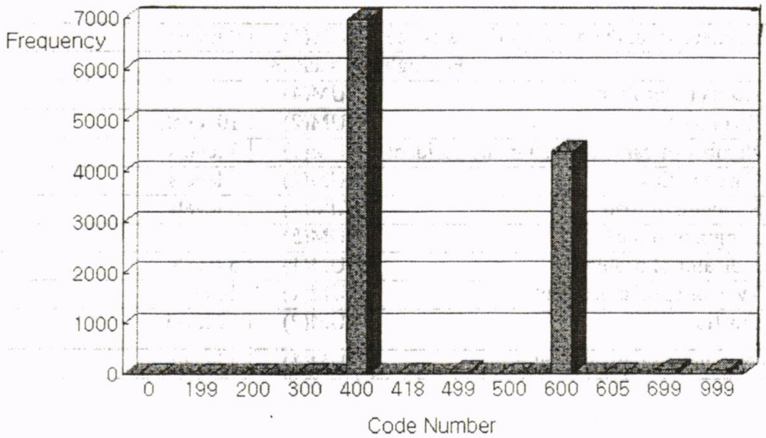
No	Factor	Type	Level
1	From classification number	CHAR(1)	
2	Police station code	CHAR(4)	
3	Report ID	CHAR(5)	
4	Accident type	NUM(1)	5 levels
5	Total number of fatality	NUM(2)	
6	Total number of injury	NUM(2)	
7	Total number of minor injury	NUM(2)	
8	Total number of minor injury report	NUM(2)	
9	Amount of property damage	NUM(2)	(first party=1) (second party=2)
10	Time of accident occurrence	CHAR(8)	
11	Date of week	CHAR(1)	7 levels
12	Day / night	CHAR(1)	2 levels
13	Code for 10	NUM(4)	
14	Residence code	NUM(4)	(1) (2)
15	Accident location x y coordinate	NUM(5)	(longitude) (latitude)
16	Road type	NUM(2)	
17	Road number	NUM(4)	
18	Occupation number	NUM(2)	(1) (2)
19	Age	NUM(2)	(1) (2)
20	Gender	NUM(1)	(1) 3 levels (2) 4 levels
21	Education	NUM(1)	6 levels
22	Weather	NUM(1)	5 levels
23	Special Accident	NUM(2)	18 levels
24	Road type	NUM(2)	12 levels

25	Width of road	NUM(1)	7 levels
26	Population density	NUM(1)	4 levels
27	Traffic control device	NUM(1)	7 levels
28	Road line type	NUM(1)	10 levels
29	Road surface condition	NUM(1)	10 levels
30	Freight equipped	NUM(2)	(1) (2) 7levels
31	Type of vehicle	NUM(2)	(1) 26 levels (2) 29 levels
32	Car shape	NUM(1)	5 levels
33	Usage of vehicle	NUM(2)	35 levels
34	Engine displacement	NUM(2)	
35	Accident mode	NUM(2)	41 levels
36	Speed of car before the accident	NUM(2)	19 levels
37	Purpose of passing	NUM(2)	20 levels
38	Presence of special operator	NUM(1)	7 levels
39	Violent driving	NUM(1)	9 levels
40	Drunk driving	NUM(2)	13 levels
41	Violations/behavior	NUM(5)	120 levels
42	Cause of accident	NUM(3)	(person cause) (vehicle cause) (road environmental cause)
43	Rule violation	NUM(2)	12 levels
44	Type of license	CHAR(10)	10 levels
45	Status of license	NUM(2)	12 levels
46	Handicapped license status for vehicle	NUM(1)	5 levels
47	Status of handicapped driver	NUM(1)	6 levels
48	Number of years with license	NUM(2)	11 levels
49	Protective device	NUM(2)	13 levels
50	Degree of damaged vehicle	NUM(1)	8 levels
For injury accident			
51	Degree of injury body	NUM(1)	(1) 6 levels (2) 7 levels
52	Injured body part	NUM(2)	(1) 11 levels (2) 12 levels
53	Condition of injured body part	NUM(2)	(1) 11 levels (2) 12 levels
54	Vehicle part/road caused injury	NUM(2)	(1) 17 levels (2) 18 levels
For fatal accident			
55	Barrier median segregation facility	NUM(1)	7 levels
56	Separation of road	NUM(1)	4 levels
57	Speed limit	NUM(2)	11 levels
58	Automatic transmission	NUM(1)	5 levels
59	Vehicle make	NUM(2)	
60	Body type	NUM(2)	
61	Insurance	NUM(1)	8 levels
62	Vehicle inspection status	NUM(1)	6 levels
63	Maximum load	NUM(2)	13 levels
64	Loading condition	NUM(1)	10 levels
65	Activity of vehicle/passenger right before the accident	NUM(2)	(1) 19 levels (2) 20 levels
66	Passing type	NUM(1)	6 levels
67	Driver occupation	NUM(1)	7 levels
68	Driver frequency	NUM(1)	9 levels

69	Distance from the passenger's residence	NUM(1)	9 levels
For highway accident			
70	Accident location	NUM(4)	
71	Road type	NUM(2)	10 levels
72	Number of cars involved in the accident	NUM(1)	7 levels
73	Curve radius	NUM(1)	7 levels
74	Structure of road	NUM(1)	5 levels
75	Length of tunnel	NUM(5)	
76	Indicator of problem	NUM(1)	5 levels
77	Type of special accident	NUM(1)	10 levels
78	Traffic	NUM(2)	12 levels
79	Presence of replacement driver	NUM(1)	4 levels

Table 2. Road Condition Classification in TASF

Level		code
Poor Road Shape		601
Poor Shape of Intersection		602
Poor Sight		603
Road Status	Poor Road Surface	604
	Snowfall or freezing	605
Traffic Obstruction	Under Construction	606
	Telephone Pole or Signboard	607
	Permitted Parking	608
	Traffic Congestion	609
	Accident	610
	Etc.	611
Traffic Safety	Signal Lamp	612
	Side Walk	613
	Median Strip	614
	Road Reflection Mirror	615
	Lighting System	616
	Road Crossing	617
	Etc.	618
Sign Imperfection	Regulation Sign	619
	Direction Sign	620
	Route Sign	621
	Caution Sign	622
	Supplementary Sign	623
No Sign		624
Other causes of road condition		625
Survey Incapable		699
Not related to road condition		600
Person Unidentified		0
Train Accident or Sufferer of Independent Accident		999



0~500: Not included in pre-defined codes

Figure.1.Frequency of Levels of Road Condition at the Time of Accident

As presented in Figure.1, road condition contains much useless information. Therefore, we use four other variables (road type, road width, signal lamp, and road shape) which can represent environmental condition of road in order to understand the relationship between accident severity and environmental road condition. Contingency tables between each variable and accident severity are displayed in Tables 3-6.

Table 3. Frequency of each type of accident severity against road type

	Death	Major injury	Minor injury	Injury report	Property damage
Service region	0	26	34	0	131
(Inter) over 13m	5	324	304	3	414
(Inter)over 6m	3	97	104	2	163
(Inter)under 6m	0	32	60	0	82
Around intersection	2	47	56	0	48
Tunnel	1	8	7	0	6
Bridge	4	12	15	0	17
Etc	94	2552	3079	64	3768

Table 4. Frequency of each type of accident severity against road width

	Death	Major injury	Minor injury	Injury report	Property damage
Service region	0	26	34	0	131
Under 3m	0	5	8	0	17
Over 3m	4	189	226	5	367
Over 6m	15	515	664	12	902
Over 9m	22	619	753	9	758
Over 13m	34	770	886	14	1002
Over 20m	34	974	1088	29	1452

Table 5. Frequency of each type of accident severity against signal lamp

	Death	Major injury	Minor injury	Injury report	Property damage
Signal lamp light up	33	1296	1268	11	1796
Signal lamp blink	1	84	84	0	35
Signal lamp light out	1	104	104	0	57
Signal lamp break out	2	8	8	0	15
No signal lamp	72	1662	2195	58	2726

Table 6. Frequency of each type of accident severity against road shape

	Death	Major injury	Minor injury	Injury report	Property damage
Service region	0	26	34	0	131
left curve upward	0	3	3	0	7
Left curve downward	0	1	1	0	3
Left curve flat	2	10	12	0	18
Right curve upward	0	2	5	0	1
Right curve downward	0	0	4	0	2
Right curve flat	3	21	27	0	24
Straight upward	1	2	3	0	6
Straight downward	0	5	10	0	5
Straight Flat	103	3028	3560	69	4432

3. Canonical Correlation Analysis of Accident Severity

The main goal of accident data analysis is to reduce the number of accidents by extracting the patterns out of the data available. In this section, we use quantification theory II to analyze the patterns of accident severity. Quantification theory II is a categorical form of canonical analysis that tries to find the best linear relationship between two groups of variables [1][2]. In our analysis, one group of variables represents different levels of severity (Death, Major Injury, Minor Injury, Property Damage, Injury Report) while another group of variables includes TASF factors such as Road Type, Road Width, Road Shape and Signal Lamp. These factors are chosen in view of the prior idea of potential impact as well as the data quality. For instance, we do not include road condition due to its poor data quality even though it is considered as one of influential factors on the accident severity.

Data used for this canonical analysis is obtained from three police stations in Seoul, Korea (Cheongryangri, Seudaemun, and Songpa) and include a total of 11564 accidents reported in 1996.

In terms of notation, we let x_{1j} different levels of severity j=1,...,5 while x_{2k} four factors (k=1,...,4). Then in CCA, we attempt to find the linear combinations of the variables that give the maximum correlation between these two combinations. Let those linear combinations to be

$$y_1 = w_1 x_1 = \sum_j w_{1j} x_{1j}$$

$$y_2 = w_2 x_2 = \sum_k w_{2k} x_{2k}$$

Then we wish to find those values of w_1 and w_2 that maximize the correlation between y_1 and y_2 [4]. Note that x_{1j} and x_{2k} we used in the model are all categorical variables and we could apply quantification theory II due to this nature.

As a result of the preliminary analysis, we found four canonical models, which are significant at 5% level. The first two models which take into account 76.54% of the total variation is as follows for the severity:

Canonical Equation 1 for the accident severity (1)
= -1.66(death)-0.42(major injury)-0.80(minor injury)+0(injury report)+3.03(property damage)

Canonical Equation 2 for the accident severity (2)
= -3.13(death)+5.91(major injury)+4.88(minor injury)+0(injury report)+8.05(property damage)

Also, the first two canonical models for Road Type, Width, Signal Lamp and Road Shape are as follows:

Canonical Equation 1 for road condition (3)
= 4.45(Service Region)+0.05((Inter)over 13m)+1.00((Inter)Over 6m)+0.96((Inter)Under 6m)-0.44(Around Intersection)-1.82(Tunnel)-0.78(Bridge) +2.66(Under 3m)+1.04(Over 3m)+0.42(Over 6m)-0.76(Over 9m)-0.50(Over 15m)

+3.23(Signal Lamp lights up)-2.73(Signal Lamp lights out)+0.28(Signal Lamp break down)-0.42(No Signal Lamp)+1.92(Left Curve Upward)+3.67(Left Curve Downward)+0.20(Left Curve Flat)-4.30(Right Curve Upward)-1.72(Right Curve Downward)-1.44(Right Curve Flat)+1.06(Straight Upward)-2.12(Straight downward)

Canonical Equation 2 for road condition (4)
= 0.81(Service Region)+0.74(Intersection)-0.57((Inter) Over 6m)-0.19((Inter) Under 6m)-0.13(Around Intersection)-1.83(Tunnel)-6.48(bridge)+1.16(Under 3m)+1.11(Over 3m) +0.89(Over 6m)-0.91(Over 9m)-0.25(Over 15m)+0.18(Signal Lamp Light Up)-0.93(Signal Lamp Lights Out)-5.41(Signal Lamp Break Down)-1.41(No Signal Lamp) +0.71(Left Curve Upward)+0.24(Left Curve Downward)-3.23(Left Curve Flat)-1.31(Right Curve Upward)-2.85(Right Curve Downward)-2.31(Right Curve Flat)-7.44(Flat Upward)+1.34(Flat Downward)

Now we apply quantification theory II to utilize this result for finding the relationship between the two groups of variables. First of all, standardization is needed for the estimated coefficients from the mean values. In order to find the mean values, we multiply the number of accidents occurred for each severity level by the corresponding estimated coefficient and divide it by the total number of accidents as displayed in Table 7. As a result, we obtain the mean values for the first two canonical equations to be 1.832 and 6.320, respectively:

Table 7. Adjusted Coefficients for Accident Severity

Accident Type	Number of Cases	Canonical Equation 1		Canonical Equation 2	
		Coefficient	Quantifying Value	Coefficient	Quantifying Value
Death	109	-1.66	-2.492	-3.13	-9.450
Major injury	3098	-0.42	-1.252	5.91	-0.41
Minor injury	3659	-0.80	-1.632	4.88	-1.44
Injury report	69	0.00	-0.832	0	-6.320
Property damage	4629	3.03	2.198	8.05	1.73
Sum	11564				

We do the same analysis for another group of variables and the results are given in Table 8. One additional analysis done for this group of variables is finding the range of the coefficients for each categorical variable. This range can give us ideas about the relative impact of each factor on the variation of the severity.

Table 8. Adjusted Coefficients for Road Type, Road Width, Signal Lamp and Road Shape

Variable		Canonical Equation 1			Canonical Equation 2		
		Coefficient	Quantifying Value	Range	Coefficient	Quantifying Value	Range
Road Type	Service Region	4.45	4.338	6.270	0.89	0.859	1.460

	(inter)over 13m	0.05	-0.062		0.74	0.709	
	(inter)over6m	1.00	0.888		-0.57	-0.601	
	(inter)under6 m	0.96	0.848		-0.19	-0.221	
	Around Intersection	-0.44	-0.552		0.13	0.099	
	Tunnel	-1.82	-1.932		-1.83	-1.861	
	Bridge	-0.78	-0.892		-6.48	-6.511	
	Etc	0.00	-0.112		0.00	-0.031	
Road Width	Service Region	0.00	-0.104	3.420	0.00	-0.470	1.160
	Under 3m	2.66	2.764		1.16	0.690	
	Over 3m	1.04	1.114		1.15	0.680	
	Over 6m	0.42	0.524		0.89	0.420	
	Over 9m	-0.76	-0.656		0.91	0.440	
	Over13m	-0.50	-0.396		0.25	-0.22	
	Over 20m	0.00	-0.104		0.00	-0.470	
Signal Lamp	Light Up	3.23	2.408	5.960	0.18	0.920	6.33
	Switch On and Off	0.00	-0.882		0.00	0.740	
	Light Out	-2.73	-3.552		0.93	1.670	
	Breakdown	0.28	-0.542		-5.40	-4.660	
	No Signal Lamp	-0.42	-1.242		-1.41	-0.67	
Road Shape	Service Region	0.00	-0.01	7.970	0.00	-0.032	8.782
	Left Curve upward	1.92	1.930		0.71	0.678	
	Left Curve Downward	3.67	3.680		0.24	0.208	
	Left Curve Flat	0.20	-0.201		-3.23	-3.262	
	Right Curve Upward	-4.30	-4.290		1.31	1.278	
	Right Curve Downward	-1.72	-1.701		-2.85	-2.882	
	Right Curve Flat	-1.44	-1.430		-2.31	-2.342	
	Straight Upward	1.06	1.07		-7.44	-7.474	
	Straight Downward	-2.12	-2.11		1.34	1.308	
	Straight Flat	0.00	-0.01		0.00	-0.032	

(inter) means intersection road

Based on the standardized canonical equation 1, it appears that Road Shape is most influential factor by having the range of 7.97. The right hand sided curved road and straight downward road are associated with bodily injury related accident while property damage related accidents tend to occur on the left hand sided downward, upward or straight upward roads. Next influential variable is Road Type. Tunnel, bridge, and intersection are more prone to bodily injury while property damage occurs more often in service area. It turns out that when the Signal lamp does not exist or not working, bodily injury tends to occurs.

Canonical equation 2 may not be more powerful than canonical equation1 in terms of explanation of the phenomena. But it can do complementary role. That is it can explain what the first one could not do. According to the second equation, Road Shape is the best one to do such a role. Straight upward road can explain the death that cannot be explained by the curved right hand side upward driving while the straight downward driving complements curved left hand side downward driving for property damage..

4. Conclusion

In this paper, we first examine the quality of TAR in general. A careful examination of TASF reveals many data quality related problems. We suggest revamping data collection procedures along with redesign of TAR and TASF. Systematic training scheme is expected to complement the hardware related improvement.

Aside the poor quality factors such as road condition, canonical correlation analysis is used to identify the relationship between a set of variables that describe the road environmental conditions (road type, width, curve, and traffic signal), and different levels of severity of accidents (death, major injury, minor injury, report, no injury).

Based on our analysis, we observed the following. The most influential factor on the accident severity changes is Road Shape. It appears that right hand sided curved road and straight downward road are associated with bodily injury related accident while property damage related accidents tend to occur on the left hand sided downward, upward or straight upward roads. Next influential variable is Road Type. Tunnel, bridge, and Intersection are more related to bodily injury while property damage occurs more often in service area. It turns out that when the signal lamp does not exist or not working, bodily injury tends to occur. Also, according to canonical equation 2, Road shape can explain the part not identified by the canonical equation 1. Straight upward road can explain the death that cannot be done by the curved right hand side upward driving while the straight downward driving complements curved left hand side downward driving for property damage.

Reengineering of data collection system is left as further research topic to resolve many problems indicated in [11]. Also, other types of emerging algorithms such as multilevel association rule can be employed to understand the relationship between accident severity and road conditions. Comparison study to this kind of algorithm would of interest of future research.

References

[1] Afifi, A. A. and Clark, V. (1990). **Computer-Aided Multivariate Analysis**, New York: Chapman & Hall.

[2] Clarke, D.; Forsyth, R.; and Wright, R. (1998) Machine learning in road accident research: Decision trees describing road accidents during cross-flow turns, **Ergonomics, Vol. 41, No.7**,1060-1079

[3] Huh, M.H. (1992). **Quantification Methodology**, Free Academy (Seoul, Korea).

[4] Lai, P.L. and Fyfe, C. (1999). A neural implementation of canonical correlation analysis, **Neural Networks, Vol.12, No.10**, 1391-1397.

[5] NHTSA. (1995). GES National Accident Sampling System General Estimates System User' Manaual, DOT, USA.

[6] Research Institute for Police Science. (1998). Improvement of statistical information system for road traffic accident, Technical Report, (Seoul, Korea).

[7] Road Traffic Safety Society. (1992). Traffic accident investigation and management system, Technical Report, (Seoul, Korea).

[8] Road Traffic Safety Society. (1996). Improvement of traffic accident management system, Technical Report, (Seoul, Korea).

[9] Oppe, S. (1992). A Comparison of some statistical techniques for road accident analysis, **Accident Analysis & Prevention, Vol.24**, 397-423.

[10] Shin, H.W and Sohn, S.Y. (2001). Comparing Classification Accuracy of Ensemble and Clustering Algorithms Based on Taguchi Design, **Journal of the Korean Institute of Industrial Engineers, .Vol.27, No.1**, 47-53.

[11] Sohn (1997). Study on the road traffic accident data collection system and analysis, Technical report prepared for research institute for police science, (seoul, Korea).

[12] Sohn, S.Y. and Lee, S.H.(2000). Data fusion, ensemble and clustering to improve the classification accuracy for the severity of road traffic accident in Korea, **Journal of the Korean Institute of Industrial Engineers, Vol.26, No.4**, 354-362.

[13] Sohn, S.Y. and Shin, H.W.(1999). Comparison of data mining classification algorithms for categorical feature variables, **Korea IE Interface, Vol.12, No.4**, 551-556.

[14] Sohn, S.Y. and Shin, H.W. (2001), Pattern recognition for road traffic accident severity in Korea, Ergon**omics, Vol 44, No. 1**, 107-117.