EFFECTS OF SIMPLE AND COMBINATION VARIABLES ON TRAFFIC ACCIDENTS USING HOMOGENEOUS ROAD SEGMENTS

by Terrance M. Rengarasu **, Toru Hagiwara*** and Masayuki Hirasawa****

(1) Introduction

Most of the previous studies with respect to traffic accident modeling have used simple variables. Using the simple variables implies a hypothesis that the variables affect the traffic accidents individually. However a previous study by Shanker et al.[1] has shown that combination variables have a significant effect on the traffic accidents. The study by Shanker et al. created combination variables by combining two simple variables. Threshold values for the two variables were set. A dummy variable was used to indicate whether the two simple variables met the threshold. The mean or the 75th percentile of the simple variable was used as threshold points. As simple this method is however; the setting of the threshold point is challenging. When there are more than two simple variables to combine the difficulty in deciding the threshold point increases. There is a need to develop a clear methodology to create combination variables.

The first aim of this study is to develop combination variables using data mining techniques. The second aim of the study is to find the effects of simple and combined road way variables on the traffic accidents. The final aim of the study is to compare accident predicting abilities of two negative binomial models. One model using fixed length road segments and the second model using homogeneous road segments.

(2) Data Handling

The selection of a type of road segment length is the first step in developing an accident prediction model. Past studies have used fixed length type[1)2)] or homogeneous type road segments[3)]. Using fixed length type segments in accident models simplifies the data collection procedures immensely. However using fixed length type road segments may divide one road way factor (example: a curve or a bridge) into two or more segments. Thus the developed model may not be able to capture the effect of that particular roadway factor. Homogeneous road segments are divided taking the roadway factors to into account. According to the aims of this study two databases had to be created. The first database with homogenous road segments and the second with fixed length road segments. For the first database with homogeneous road segments the roads were divided into smaller road segments based on horizontal alignment of the road. Road segments were either straight or curved. Figure 1 shows how the homogeneous road segments are divided. For the second database with fixed length road segments the roads were divided into 1km long segments.

Accident data of the national roads 5, 38, 39, 40, 44 and 274 of Hokkaido, Japan for 16 years from year 1989 to year 2004 were used in this study. All types of accidents except pedestrian accidents, Rear end accidents and accidents with trains were used in the study. Accident data was collected from the Traffic Accident Analysis System (TAAS) developed by the Civil Engineering Research Institute for Cold region, Hokkaido, Japan[4)].Accident data was collected for the homogeneous and 1 km road segments separately. Accident data was divided into two periods; non winter and winter. To represent the winter and non winter a dummy variable called ‘Winter’ was introduced.

Data for the simple roadway variables were obtained from the detailed and up-to-date roadway database of Japanese highways (MICHI Database). Using MICHI database roadway data for the homogeneous road segments and 1 km road segments were collected separately. Data for hilliness, maximum slope, horizontal radius, bendiness, tunnels, bridges, snow sheds, maximum shoulder width, average lane width, number of lanes truck lane, AADT, densely inhabited district (DID) and Railway stations were collected.

Accident data and simple roadway variable data were combined to make two databases; one with homogeneous road segments and other with 1 km long road segments. Unlike the database with 1 km segments database with homogeneous segments had segments with different length. Segment lengths varied from 1m to 4200m. Small segments are not use full when come to predicting potential accident prone areas. A minimum value of 200 m was selected as the least road segments length to be used in the model.

In the database with homogeneous segments there were 3,888 segments with a mean length of 5,621 m. In total data for 2415.338 km long road was collected. 29,589 accidents were recorded. Database with 1 km segments had 2790 segments covering data for 2790km long road. 33,726 accidents were recorded for the database with 1 km long segments. The accident frequency distributions for the database with homogeneous road segments and 1 km road segments are shown in Figure 1 and Figure 2 respectively. Descriptive statistics of the variables in the two databases are shown in Table 1.

*Key words: Negative Binomial, Homogeneous segments, Combination variables
**Terrance M. RENGARASU, Graduate School of Engineering, Hokkaido University
(N13, W8, Kita-Ku, Sapporo, 060-8628, Japan Tel: +81117066212 Fax: +81117066211)
***Toru HAGIWARA Ph.D.,Graduate School of Engineering, Hokkaido University
(N13, W8, Kita-Ku, Sapporo, 060-8628, JapanTel/Fax: +81117066214)
****(Masayuki HIRASAWA Ph.D., Civil Engineering Research Institute for Cold Region, Japan
(Hiragishi 1-3-1-34, Toyohira-Ku, Sapporo, 062-8602, Japan, Tel: +81118411738 Fax: +81118419747)

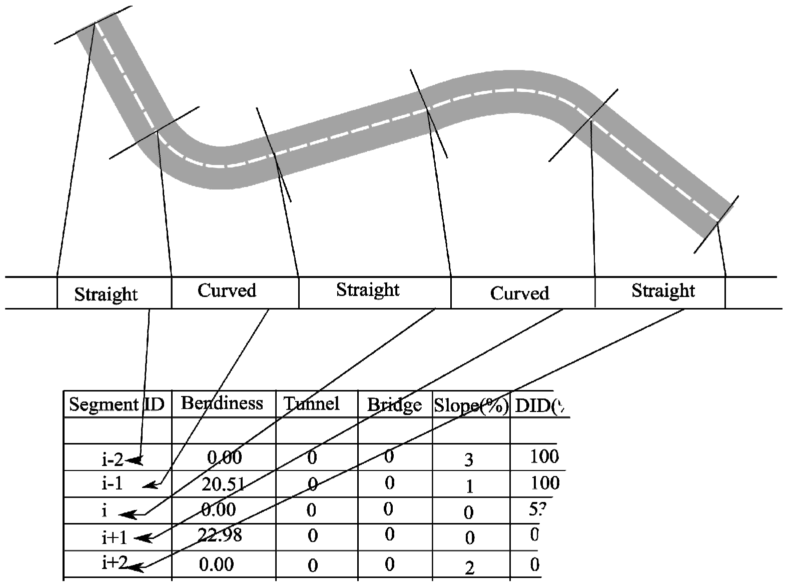

Figure 1 Illustration of how the homogeneous road segments are divided.

Table 1 Descriptive statistics of the variables in the two databases

Variable	Unit	Homogeneous segments				1km segments			
		Mean	Standard Error	Min	Max	Mean	Standard Error	Mini	Max
Horizontal radius	m	550956.1	8673.85	70	1,000,000	53823.48	2542.57	0	1,000,000
Bendiness	Degree	21	0.56	0	225.91	8.73	0.59	0	295.22
Maximum slope	%	2.21	0.03	0	8.7	2.33	0.04	0	8.7
Tunnel	Dummy	0.03	>0.01	0	1	0.04	>0.01	0	1
DID	%	5.85	0.39	0	100	10.01	0.6	0	100
Bridge	Dummy	0.24	>0.01	0	1	0.37	>0.01	0	1
Snow Shed	Dummy	0.01	>0.01	0	1	0.02	0.002	0	1
Average Lane width	m	3.3	>0.01	1.75	5.71	3.29	>0.01	1.75	5.01
Number of Lanes	Number	1.11	>0.01	1	3	1.148	>0.01	1	3
Maximum Shoulder width	m	1.61	>0.01	0.5	4	1.68	>0.01	0.5	5.25
AADT	Vehicles /direction/day	10212	129.04	1461	52473	10991.78	178.05	1461	53625.6
Winter	Dummy	0.5	>0.01	0	1	0.5	>0.01	0	1
Railway Stations	Stations /km	0.06	>0.01	0	0.8	0.06	>0.01	0	1.37
Uphill lane	Dummy	0.04	>0.01	0	1	0.036	>0.01	0	1
Segment Length	m	621.23	11.18	200	4098	1000	0	1000	1000

(3) Combination variables

Three combination variables were computed; combination of the cross sectional class variables, combination of the road geometry class variables and the combination of the adjacent segments. The variable 'Combination of cross sectional class variables' and 'Combination of the road geometry class variables' were computed using CHAID decision tree (Chi-squared Automatic Interaction Detection). Accident rate per kilometer was used as the target variable. Figure 3 and Figure 4 shows the decision trees developed in Combination of cross sectional class variables and Combination of the road geometry class variables. Combination of the adjacent segments was calculated as the average of the accident rates of the two adjacent segments.

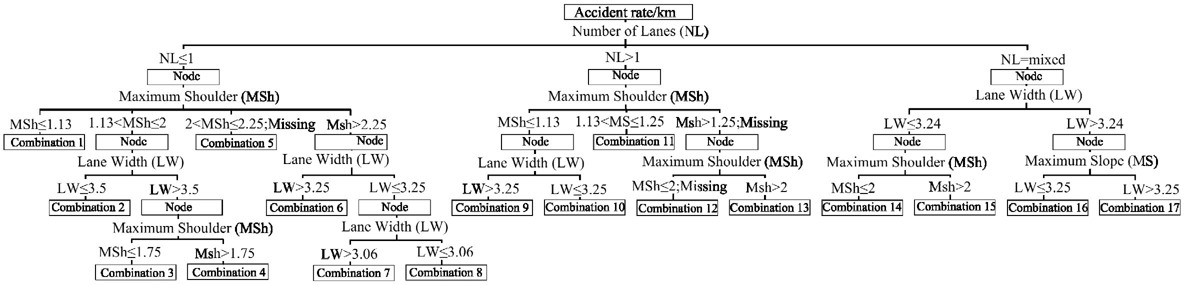

Figure 3 Decision tree for the Combination of cross sectional class variables

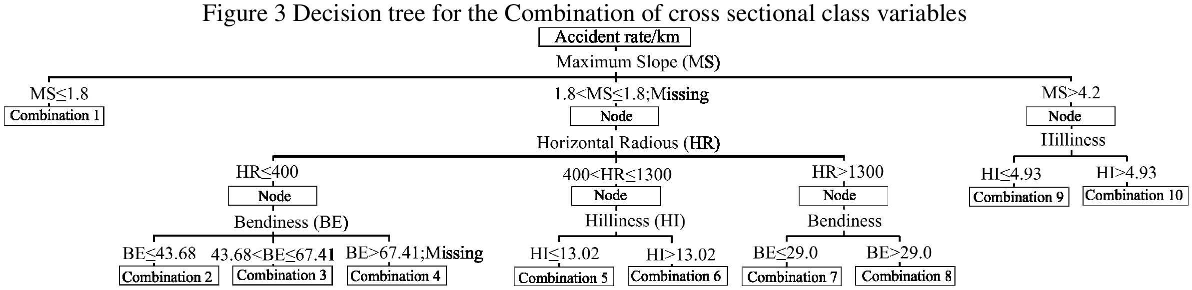

Figure 4 Decision tree for the Combination of the road geometry class variables.

(4) Results

Results of the Negative-Binomial regression models are shown in Table 2. Only the significant variables are shown in the table.

(a) Results of the model using homogeneous road segments

The overdispersion factor was found to be significant in the Negative Binomial model developed with the homogeneous road segments. The model achieved a log likelihood ratio of 0.40. A total of 15 variables were found to have a significant effect on the number of accidents at a critical P value of 10%. The bendiness increased the number of accidents while the hilliness decreased the number of accidents. Roadway structures tunnel and bridge has significant increasing effect on the number of accidents. In the cross sectional class variables Maximum shoulder width and average lane width have a significant increasing effect while number of lanes has a significant decreasing effect on the number of accidents. In-built are indicator class DID percentage and Railway stations showed a significant increase in the number of accidents. AADT showed a significant increasing effect. Winter showed a decreasing effect on number of accidents. Segment length showed a increasing effect on accidents. Three variables used for the interaction; (Interaction of Adjacent Segments, Interaction of cross sectional, Interaction of Longitudinal variables) showed significant increasing effect on number of accidents.

(b) Results of the model using 1km road segments

The developed model showed a significant overdispersion factor. The model achieved a log likelihood ratio of 0.28. Total of 7 variables showed a significant influence on the number of accidents. In the road geometry class bendiness and Maximum grade showed a significant increasing effect while hilliness showed a significant decreasing effect on the number of accidents. In the built are indicator class DID percentage showed a significant increasing effect on number of accidents. In The road cross section class number of lanes showed an increasing effect on the number of accidents. Winter showed a significant increasing effect on the number of accidents. AADT showed increasing effect on the number of accidents.

(5) Discussions and Conclusions

(a) Comparison of the two models

In this study same data was used to create two kinds of databases one with a homogeneous road segment and other with 1 km road segments. By comparing the log likelihood ratio of the two models it is clear that the model with the homogeneous segments have a higher level of accuracy. Further the model with the homogeneous segments has more significant variables in the model. Therefore it can be concluded that the usage of homogeneous road segments and the usage of interaction variables increase the accident predicting ability of a model.

(b) Effects of variables

The model with homogeneous road segments showed that tunnels and bridges have a significant increasing effect on number of accidents. Because of the high construction cost and difficult terrain road segments immediate to tunnels and bridges are generally of poor alignment. The model suggests that the number of accidents is increasing in the densely inhabited segments of the road and in the road segments with a railway station. Flow rate will be higher in the densely inhabited area in addition high number of cross roads would tend to increase the number of accidents. Increase in accidents in the segments with a railway station is a reasonable finding, since railway stations are located in the central part of the town and main commercial activities are located around it. According to the model the number of accidents in winter decrease significantly. This may seem like an unexpected result. However thorough study of the accident data revealed that the number of winter accidents is less than that of summer accidents. It can be argued that the number of accidents is decreasing due to the decrease in traffic flow, slow driving speeds and extra vigilant driving in the winter. Difficult driving conditions might increase some types of accidents during the winter. Study by Rengarasu et. al[2)] showed that number of head on accident increase in the winter. However when different types of accident are grouped together, this study finds that number of accidents decrease in the winter.

Bendiness, Hilliness, Maximum Shoulder width, Lane width and Number of lanes are used in the model as a variable directly and as two interaction variables. Therefore the results of Bendiness, Hilliness, Maximum Shoulder width, Lane width and Number of lanes have to be interpreted using the variable and the interaction variable. According to the model increase in the bendiness increases the number of accidents. This can be interpreted in two ways. Firstly accidents are more frequent in curved segments than in the straight segments of the road. Secondly among the curved segments high bendiness has high number of accidents. Increase in the bendiness means road more curved need drives constant attention and high chance of making a mistake. Increase in bendiness result in short sight distance decreasing the driver's ability to see the road ahead. Increase in hilliness means a poor alignment of the road. Generally the number of accidents is expected to rise with hilliness of the road. However the model indicated that increase in hilliness decreases the number of accidents. This result is unexpected. Investigation into the interaction variable might help.

Combination variables used for the cross section and the road geometry showed a significant increasing effect. This indicates that there is a combination effect of variables on the number of accidents. Having established that there is a combination effect on accidents the corresponding tree can be used to identify the 'safe' and 'unsafe' combination of variables. Combination variable for the adjacent segments showed a significant increasing effect on the number of accidents. This suggests that if a segment is accident prone then the effect is carried on to the next segment too.

Table 2 Results of Negative Binomial Regression Model with Homogeneous road segments and 1 km segments

Parameter	Homogeneous segments			1km segments		
	Estimate	t-statistic	P-value	Estimate	t-statistic	P-value
Constant	-17.245	-35.432	0	-8.087	-27.248	0
Bendiness	0.003	2.925	0.003	0.003	5.851	0
Hilliness	-0.006	-2.273	0.023	-0.003	-1.853	0.064
Maximum Grade	Not significant			0.003	1.769	0.077
Tunnel Dummy	0.299	3.16	0.002	Not significant		
Bridge Dummy	0.107	2.587	0.01	Not significant		
DID Percentage	0.007	9.012	0	1.188	21.944	0
Railway Stations	0.515	3.113	0.002	Not significant		
Maximum Shoulder	0.075	2.21	0.027	Not significant		
Lane Width	0.177	2.089	0.037	Not significant		
Number of Lanes	-0.174	-1.921	0.055	0.367	5.856	0
Log AADT	1.157	27.419	0	1.022	28.512	0
Winter Dummy	-0.105	-3.019	0.003	0.065	1.949	0.051
Log Segment length	0.997	28.005	0	Not used		
Combination of Adjacent Segments	0.009	9.946	0	Not used		
Combination of cross section variables	0.032	6.012	0	Not used		
Combination of Longitudinal variables	0.03	3.723	0	Not used		
Over dispersion	0.451	19.866	0	0.43	23.059	0
Log likelihood function	-6213.67			-6805.95		
Restricted log likelihood	-10412.1			-9472.64		
ρ2	0.4			0.28		

(6) References

1. Shanker, V *et al.* Effects of roadway geometrics and environmental factors on rural freeway accident frequencies, accident analysis and prevention, Vol.27, No. 3, 371-389, 1995.
2. Rengarasu T. M. *et al.* Effects of road geometry and season on head-on and single-vehicle collisions on rural two lane roads in Hokkaido, Japan, Journal of the Eastern Asia Society for Transportation Studies, Vol. 7 pp.2860-2872, 2007
3. Young J.K. and Kara M. K., Safety effects of speed limit changes: use of panel models, including speed, use, and design variables, Transportation Research Record 1908, TRB, National Research Council, Washington, D.C, pp. 148-158,2005,
4. Hirasawa, M. and Asano, M. Development of traffic accident analysis system using GIS, Proceedings of the Eastern Asia Society for Transportation Studies, Vol.4, 1193-1199, 2003.