Proceedings of the Federated Conference on Computer Science and Information Systems pp. 181–186
DOI: 10.15439/2016F195
ACSIS, Vol. 8. ISSN 2300-5963

Evaluating Model of Traffic Accident Rate on Urban Data

Jianshi Wang
Department of System Innovation
Graduate School of Engineering
University of Tokyo
Email: jianshiwang0329@gmail.com

Yukio Ohsawa
Department of System Innovation
Graduate School of Engineering
University of Tokyo
Email:ohsawa@sys.t.u-tokyo.ac.jp

Abstract*—Public safety, especially the daily traffic accident is concerned by the public. Previous studies have already discussed accident reasons associated with accidents statistically. There is a method called Innovators Marketplace on Data Jackets created by Professor Ohsawa. This method is used to externalize the value of data via stakeholders' requirement communication. This paper applied the solution from an IMDJ workshop to research this topic creatively. This novel solution suggested to do analysis on the combination of urban data and traffic accident rate to find the impact factors to the traffic accident rate in the urban system. This paper used factor analysis, structure equation modeling and data mining to construct a theoretical frame for traffic accident rate analysis for urban data. Different accident indexes, such as total number of accident, fatality rate, injury rate, and casualty rate are combined to construct a traffic accident risk evaluation model. This paper chosen the urban data as the solution from IMDJ workshop, such as population structure information, vehicle information, road characters, public traffic system information, and the other kinds of data to explore factor meaning, and to identify relationships between different factors. It segmented these urban data based on their categories, and determined accident risk for each section. By doing analysis on not only the original data but also the changing rate of these data each year, the result analytical results showed that traffic accident rate on urban data could be described by the combination of population structure, road characters, public traffic system and public facilities. These four sections affects traffic accident rate significantly during the development of urban; however, the vehicle factor does not have influence on traffic accident rate. And it proves the solution from IMDJ workshop is not only novel but also practical strongly. Making some solution from IMDJ into reality, we will find another new way to affect the world.

I. INTRODUCTION

In Beijing, the fatal accident rate was 39.56 per million populations in 2014, it decreased roughly 2.8% in comparison with 2013, even though the population of 2014 has increased 1.7%. Different losses caused by traffic accident in urban are considerable. Many previous researches on traffic safety tried to find methods for preventing traffic accidents. Among these researches, the analytical methods are used frequently, including Clustering Analysis (CA), Statement Statistic (SS), Regression Analysis (RA), evaluation of the correlation between independent variable sets and a dependent variable; but these models could not cover the causal relationship between these variables or combine different traffic accident indexes to identify accident risk. Even though all of the final purpose of them attempts to avoid traffic accident, they mostly used discriminate analysis or regression model in analyzing accident characteristics and identifying behaviors of drivers, weather condition or road sections prone to accidents, then constructed an accident prediction model by parametric approaches (i.e. on distribution model). However, these methods only analyzed accidents on drivers with different attributes and relevance, weather data or different road sections; they cannot define the causality between factors clearly enough. Some current studies have already utilized data mining, including non-parametric approaches where traditional models were not employed, to identify key accident related impact factors. However, data mining can just identify traffic accident risk indexes (i.e. number of traffic accident, of death, of injuries, or of casualties). But it is necessary to identify the latent correlations among the different factors which is suggested by solution from IMDJ, and find the correlations among the traffic accident rate and urban data-population structure, vehicles, road characters or some other cause-and-effect relationships. In this study, we focused on urban data, applied structural equation modeling to find impact factors associated with traffic accident. Important traffic accident factors were identified, population structure, vehicle information, road characters and other cause-and-effect relationship associated with traffic accident risk were discussed, and prediction model which describe the latent relationship between urban city and accident was constructed. This paper discussed six factor sets by constructing an accident risk causal framework based on urban data and the component factor sets of each feature and influence on traffic accident.

II. LITERATURE REVIEW

This paper builds a theoretical framework for traffic accident risk by the analysis on the latent relationship between different aspects of urban data and traffic accident to identify causal relationships. In the past, Etienn(2006) used logit regression analysis to discuss reasons that led to different degree of injury condition in the traffic accident. Gender, age, and protective gear were related to the condition of injury. By using bivariate and simple analysis, Wong and Chung(2007) applied a rough set approach for accident chains to analyzing the influence of traffic accident for different factors. They found that a single factor accident chain had poor quality and

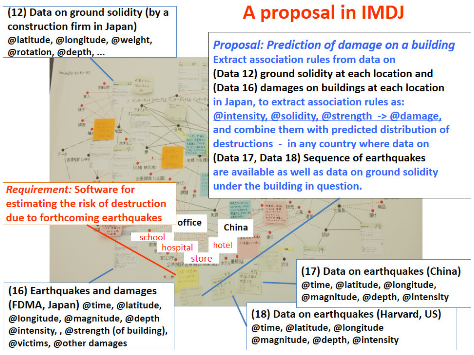

Fig. 1. A proposal in IMDJ

a multi-factor accident chain should be used when analyzing accidents. John(2007) has used mixed logit model to construct a prediction model for serious traffic accident. Variables related to daily traffic flow, such as average daily traffic flow, percentage of big vehicles, and number of access roads, and variables related to road features, such as curvature and road friction were significant variables impacting road accident risk. Niloofar(2015) utilized highly disaggregate spatial units for an analysis of area specific variables associated with urban traffic accident. These researches suggested that factors affecting traffic accident risk include drivers, road structure of city, public traffic system, which may affect each other. On the other hand, this paper applies the solution from IMDJ workshop to do factor analysis and structure equation modeling to analyze these features as well as the cause-and-effect relationships between urban factors and traffic accident rate. The main goal of this paper is necessary in order to have a clear enough structure of urban and traffic accident for improving safety which will be done with urban planning in the future.

III. MOTIVATION OF RESEARCH

The reason of doing this research comes from one solutions of IMDJ workshop. The IMDJ(Innovators Marketplace on Data Jackets) has been metioned about many times aboved. And this research is the application of a solution from IMDJ workshop, so I think it is necessary to introduce this amazing method. The purpose of IMDJ is making a social environment where analysits and decision makers in active businesses and science could be provided with data they need. During the workshop, each members could give requirements and provide solutions by checking or combining DataJackets on the Keygraph. And finally members can buy/sell their solutions in a reasonable condition. The Fig.1 shows a proposal in IMDJ workshop. DataJackets are small pieces of information containing the abstracts of data that exist but cannot be disclosed. And Keygraph created by Professor Ohsawa is the high-effective visualization method to show the relationship between different Datajackets by analyzing contents in Data jackets. Before an IMDJ workshop, the topic of the workshop should be determined by the organizer. During an IMDJ workshop which topic is ”Improving the quality of urban life”, a requirement about public safety has been asked. That requirement asked for a method to check the safety condition of objective people in an area. And another member supplied a solution of checking the safety index of the area by using combination of two Datajackets. One is the structure data of the city, and the other is the accident rate of the city. The research angle of this solution is totally different from the previous researches We selected the urban data just like the solution suggested to do this research. The society has paid lots of cost on traffic accidents significantly. And it is time to do the research to reveal the complex relationship between urban data and public saftery. The solution from IMDJ workshop gives a very novel and meaningful angle to do research on this problem.

IV. RESEARCH EXTENT AND METHODOLOGY

This study focused on urban data in Beijing City. Beijing is still in construction after Beijing Olympics 2008, and it is becoming bigger and advanced every year. Public facility and mass rapid public transit lines have been under construction in Beijing City. And also with some policy changed, the other social structures have been keeping changing every year, such as the different structure of population, the GDP of Beijing, the sales of different products and etc. These different structures have affected all of aspects in Beijing City. In respect of methodology, this paper used structure equation modeling with population, driver, vehicle, road characters, public transportation system, other urban data as independent variables, and accident rate as the dependent variable. All these data related to this urban were compiled into six categories. In addition to applying structure equation modeling, this paper used structure equation modeling method to explain the relationship between factors. To choose whether observation variables in each aspect are appropriate before model construction, each variable was subjected to factor analysis. It is necessary to exclude the unrelated factors in order to reduce the interference terms. After determining the variables, this paper analyzed the quantized the effect between these variable.

A. Data content

Urban data: We covered 30 factors, and clustered these factors into 6 sections according their categories. These 30 factors involving population, population in different age, number of truck, number of bus, gasoline consumption each year, number of hospital, number of police stations, road area, length of bus lane, green area in urban and some other aspects of urban. We assume, these data actually have the relationship among themselves and affect the traffic situation, it also includes the traffic accident naturally. Accident data: Accident rate was expressed by the sum of Fatality rate, Injury rate and Casualty rate multiplied by different weight.

$$Fatality\ rate = number\ of\ deaths/number\ of\ accident \quad (1)$$

$$Injury\ rate = number\ of\ injuries/number\ of\ accident \quad (2)$$

$$Casualty\ rate = number\ of\ casualties/number\ of\ accident \quad (3)$$

$$Accident\ rate = (1) * 3 + (2) * 2 + (3) * 1 \quad (4)$$

Here, the above accident rate means the personal safety loss of the traffic accident.Theses different weights mean the degree of damage after accident based on the influence of traffic accident. The Fatal consequence is the most serious one, so I give the weight 3 to it; to the injury consequence, I give the weight 2 to evaluate it; to the casualty, because it prefers to refer to a range of affected people, I just use 1 to depict its weight. And the number of accident refers to the number of traffic accidents in that year. These traffic fatality, injury and casualty data were collected for the period 2010 to 2014 from the following sources: publications in the same period such as Beijing Statistics yearbook, official statistic site likes National Bureau of Statistics of China and official city traffic and accident reports.

B. Variables

We chose the data by which can be used to describe the urban objective situation in order to find the relationship between structure of urban, involving the structure of urban planning and social structure in urban, and accident rate. Then identified causality with the accident rate on the result. With the development of the urban and technology, the structure of urban is keeping changing every year. Doing analysis on the cause of the traffic accident to find the main reason by using statistical method or clustering method is common method. But we could not ignore the relationship between urban data and traffic accident rate. The changes in urban can affect every small aspects in our life, absolutely involving the accident. For example: alcohol drinking sales in urban may reflect the change of drunk driving rate in urban, and drunk driving is one of main reason of traffic accident, this chain shows the alcohol drinks sales has a tacit relationship with accident. And there must be other causes of the change in traffic accident from this kind of chain based on the urban system.

C. Model Framework

To build a traffic accident risk model on urban construction, different urban data were compiled. In this paper, we collected and compiled different data in an equal structure. This structure means that these urban data are on the same important position at the beginning of analysis. In this structure, the relationship does not only exist between the traffic accident rate and different urban factors, but also among these factors. It is more like a net of relationship. The model of analysis should be much more comprehensive, only in this way the result has persuasiveness, and intuitive. Fig.2 shows the framework in this research.

D. Analysis process

1) step1: Preprocessing the urban data: Using the raw data to do analysis is easily to ignore the influence of small change automatically during the analysis. It is necessary to preprocess the urban data. The purpose is to enlarge the influence of these factors and increase the efficiency of analysis.Also the final visualization will be clear enough to observe. Transforming the raw data into different ratio value, such as gender ratio, potential support ratio, vehicle type ratio, road light per meter, length ratio between road and total public traffic system, the ratio of green area, living space, and etc. Fig.3 shows the preprocess of urban data.

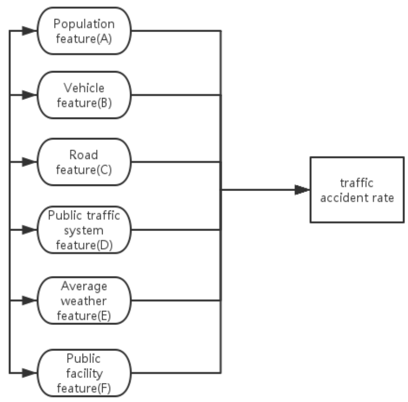

Fig. 2. model framework

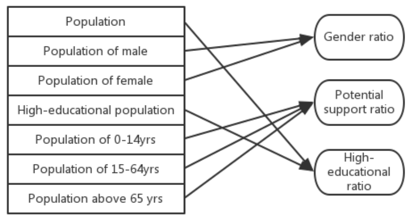

Fig. 3. Example of preprocessing

2) step2: Clustering latent factor sets: In order to distinguish the effects of different aspects in urban, before the analysis, it is necessary to define the preprocessed data into different sets. After the main components were analyzed, variables have been clustered in six features-population, vehicle,road characters, public traffic system, average weather data, public facility data. In this way, the result can give the structure and relationship between accident and variables, also among variables themselves. The TABLE I shows the result of clustering.

$Population\ feature \longrightarrow A\ factor\ sets(A)$
$Vehicle\ feature \longrightarrow B\ factor\ sets(B)$
$Road\ feature \longrightarrow C\ factor\ sets(C)$
$Public\ traffic\ system\ feature \longrightarrow D\ actor\ sets(D)$
$Weather\ average\ feature \longrightarrow E\ factor\ sets(E)$
$Public\ facility\ feature \longrightarrow F\ factor\ sets(F)$
$\alpha.Urban\ data\{A, B, C, D, E, F\} \Rightarrow Accident\ Rate$
$\beta.\Delta Urban\ data\{A, B, C, D, E, F\} \Rightarrow \Delta Accident\ Rate$

Population feature(A)	Vehicle Feature(B)
A1.All population A2.Gender ratio A3.High educational ratio A4.Low educational ratio A5.Potential support ratio A6.Driver ratio	B1.Number of vehicle B2.Type ratio(Big/small) B3.Driver/number of vehicle B4.Average gasoline consumption
Road characters(C)	Public traffic system(D)
C1.Length of road C2.Area of road C3.Road light ratio C4.Number of bridge C5.Wide road/narrow road	D1.Length of bus line D2.Length of subway line D3.Subway station/urban area D4.Total passengers per year D5.Bus passenger/subway passenger
Average weather feature(E)	Public facility(F)
E1.Good weather/Bad weather E2.Strong wind day E3.Days of low visibility	F1.Area of parks/areas F2.Green area F3.Hospitals F4.Schools F5.Police station

TABLE I

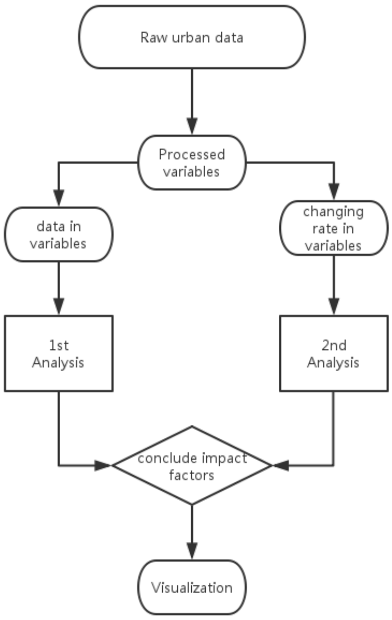

Fig. 4. Impact factors from result of analysis

3) step3: Calculating the changing rate. The whole process of analysis has been designed to do two times in order to confirm the effect of impact factors. The first used data in TABLE I directly, and the second used changing rate of factors in each year. The reason of second analysis on changing rate is to confirm the relationship between factors and the accident rate. I selected the factors which have strong relationship with accident rate by checking the P-value of analysis results. And two results can show overlap part of significant impact factors clearly, also it can confirm the influence of impact factors and quantize the influence of these impact factors.

4) step4: After getting results of these two analysis, factors which have no relationships with traffic accident are deleted. We used data of these impact factors to calculate the influence by numeric on urban traffic accident. Making a graph to visualize the relationship and degree of influence between these impact factors and traffic accident rate is necessary. With this graph, impact factors and their influence between each other can be observed clearly. The numbers marked on lines mean the degree of correlation between them. The positive number means the positive correlation, and the negative number means the negative correlation. The dotted line means the relationship is bidirectional relationship, and the solid line means the direction of relationship between sections is just one way. The Fig.4 aboved shows the whole process of analysis.

V. RESULT

The TABLE II showed results from analysis. Two columns showed impact factors and their quantized influence calculated by the assumed model.

After the first analysis, A5 factor(potential support ratio), D2 Factor(length of subway line), D3 Factor(subway station/urban area), F3 Factor(number of school) and F4 Factor(number of school) has been selected, they are from three different sections. The first result showed the potential support ratio, length of subway line, subway station/area, and number of school are the impact factors, among them the length of subway line showed the strongest relationship (0.938) with traffic accident rate. It gives some reasonable assumptions: the Beijing attracts more working population (due to the A5(potential support ratio)), and with the construction of subway line system, it is much more convenient, more and more people choose the public traffic system, and the traffic accident rate has been controlled. And with the number of hospital, the number of fatality in traffic accident has been affected. The number of school means, firstly the increasing working population takes more students, secondly the number of school has the relationship with public traffic system. The second analysis has used the changing rate in the same factors. The first result could be used to combine with it, then the combination of results could solve the problem comprehensively. After the second analysis, A2 factor(gender ratio), A3 Factor(high educational ratio), C1 Factor(length of road), D2 Factor(length of subway line) and F4 Factor(num of school) has been selected, from four different sections. The result of second analysis showed that gender ratio, high educational ratio, length of road, length of subway line and number of school are the impact factors. It showed the number of school has the strongest relationship with traffic accident rate.And it is apparently to find that the influence of F4 factor(number of school) has been improved. With considering the factor-High educational ratio, it showed some reasonable possibility: with the increasing of educational level, more people come to recognize the importance of traffic safety.And they know how to protect their safety,also it means in the traffic accident, the student accident has affected it strongly. With choosing the impact factors from these two analysis, we can visualize

First analysis	Second analysis
A5.Potential support ratio(0.55)	A2.Gender ratio(0.056)
D2.Length of subway line(0.938)	A3.High educational ratio(0.129)
D3.Subway station/city area(0.036)	C1.Length of road(0.028)
F3.Hospital(0.558)	D2.Length of subway line(0.004)
F4.Number of school(0.07)	F4.Number of school(0.867)

TABLE II

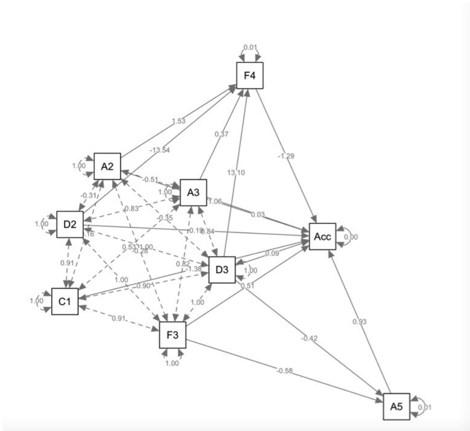

Fig. 5. Visualization of Impact Factors

the structure and the influence about them. Fig.5 showed the visualization of the result. The selected nine factors from these two analysis made this net diagram. As the interpretation of the diagram before, every impact factors showed the degree of influence, direction of the relationship and the combination of the relationship. For example, F3 (number of hospital) has affected the A5 (potential support ratio) by negative effect of 0.58, and the direction is just from F3 to A5. C1(Length of road) has the relationship with F3, they affected each other by the positive effect of 0.91. Generally, the traffic accident rate is affect by all of impact factors on this diagram. The A5(potential support ratio) has positive effect of 0.93 on traffic accident rate. The F4 (number of school) has negative effect of 1.29 on the traffic accident rate. The D2 (Length of subway line) has positive effect of 0.19 on the traffic accident. By gathering these coefficents, using these different factors as variables, we can define a model to describe the relationship between each factors and accident rate and the relationship among these different factors.

VI. DISCUSSION

"The more vehicles, the more accident.", it is a general consideration. But in this research, it showed there is no relationship with number of vehicles. Neither in the first analysis nor the second analysis, there is no result showing the accident rate has the relationship with vehicle factor. From the raw data, it showed even though the number of vehicle is increasing, the traffic accident rate has decreased, and by the result, it showed the vehicle data has no influence on the traffic accident data. On the other hand there is a strong relationship among the other urban data,such as-population structure, public traffic accident and traffic accident rate. Even though the traffic accident mainly about the vehicle caused accident on the road, the results here showed the other factors affected this rate obviously from the angle of urban structure data. And the other possibility exists, the changing data of vehicle variables has effects on the traffic accident rate, but the influence is too small, and the influence has been neutralized by other changed urban data. It still needs more and more data to make the model much more comprehensively, and make the correlation correctly.

VII. CONCLUSION

This paper constructed an assessment model for traffic accident risk which is novel in that we defined the relationship between urban data and traffic accident data. The four main sections-population structure, road characters, public traffic system and public facility had the significant impact on traffic accident rate, while the vehicle section was insignificant. A higher gender ratio (men/women) can reduce traffic accident risk, while a higher potential support increased the traffic accident risk. The longer length of road and public traffic lane can decrease the traffic accident rate.The results show that public traffic system and facilities are important factors to controlling the traffic accident rate. Traffic accident is mostly caused by people themselves, researchers can do many analysis on policy or reason analysis in order to reduce the traffic accident rate. However, from the new angle of IMDJ workshop, one should pay more attention to the significance from urban structure designing. This paper successfully made one traffic accident risk assessment model.Four sections-population structure, road characters, public traffic system and public facility are adequate for assessing traffic accident risk.It may require the deeper research in future, but this study was still restricted by its access to data. Although 30 variables were included,it was not enough. If the experiment add other kind of urban data, the model would be modified on causality and deduction. This paper was focus on the influence of urban data to traffic accident risk, it is suggestive to use this methodology to apply to other kind of public safe problem.

VIII. FUTURE WORK

Based on the conclusion of this paper, it showed that the relationship between urban data and accident could be quantized in number and described in a mathematical model. If there are other urban data, the model could be more perfect. So finding other urban data to construct the model more comprehensively. Trying to find what extent does this model match real sitiuation, and make it more comprehensive.

The requirement from that IMDJ workshop needs to check the safety condition of the objective people. So the model of traffic accident rate is not enough, because the public safety also covers the crime rate. From this solution of IMDJ, we prove the proposal which shows the traffic accident rate is related to the urban social data, so the crime rate of objective area could also be defined in that kind of mathematical model. And with combining these models, a synthesized prediction model which can describe danger index of area based on urban data could be made. It could be used in checking the danger status of objective area. Also it could give reference to urban construction in the future. Because this interesting research angle was from IMDJ workshop, and this useful result was from the applicaiton of solution in IMDJ. I think the more results from IMDJ would make more exciting and useful things. The IMDJ workshop let members from different fields to externalize use value of data by combining their familiar data and other datas noveltly. It could help us to dig the deep value of the hidden datasets. Also it could help us to share the experience from professional people in different fields. I could not wait to hold a workshop with topic of public safety to find more new points in solving these kinds of problems.

Acknowledgment

This research was supported by Japan Science and Technology Agency (JST) and Core Research for Evolutionary Science and Technology (CREST).

References

[1] Al-Ghamdi, A. S., *Analysis of traffic accidents at urban intersections in Riyadh* Accident Analysis and Prevention, 35(5), pp. 717-724, 2003, http://dx.doi.org/10.1016/S0001-4575(02)00050-7

[2] K. Ivan, I. Haidu, J. Benedek, and S. M. Ciobanu*Identification of traffic accident risk-prone areas under low-light conditions* Nat. Hazards Earth Syst. Sci., 15, 2059-2068, 2015. http://dx.doi.org/10.5194/nhessd-3-1453-2015

[3] Ariana Vorko-Jovic, Josipa Kern, Zrinka Biloglav, *Risk factors in urban road traffic accidents* Journal of Safety Research, 37(1), pp. 93-98, 2006. http://dx.doi.org/10.1136/ip.2010.029215.429

[4] John C. Milton, Venky N. Shankar, Fred L. Mannering, *Highway accident severities and the mixed logit model: An exploratory empirical analysi* Analysis and Prevention, Volume: 40, Issue: 1, January, pp. 260-266, 2008. http://dx.doi.org/10.1016/j.aap.2007.06.006

[5] Kim, K., Nitz, L., Richardson, J., Li, L., *Personal and behavioral predictors of automobile crash and injury severity* Accident Analysis and Prevention, 27(4), pp.469-481,1995. http://dx.doi.org/10.1016/0001-4575(95)00001-G

[6] Y. Ohsawa, H.Kido,T.Hayashi,C.Liu,*Innovators Marketplace on Data Jackets for Externalizing the Value of Data via Stakeholders Requirement Communication* Procedia Computer Science,pp.709-716,2013. http://dx.doi.org/10.1007/978-3-319-13545-8_6

[7] Kuhnert, P. M., Do, K. A., McClure, R., *Combining non-parametric models with logistic regression: an application to motor vehicle injury data* Statistics and Data Analysis, 34(3), pp. 371-386,2000. http://dx.doi.org/10.1016/S0167-9473(99)00099-7

[8] ARCHER,J., VOGEL,K.,*The Traffic Safety Problem in urban areas* Royal Institute of Technology Publication,2000. http://dx.doi.org/10.1016/j.aap.2016.03.017

[9] Kathleen,L. Wolf and Nicholas Bratton,*Urban Trees and Traffic Safety: Considering U.S. Roadside Policy and Crash Data* International Society of Arboriculture,pp.170-179,2006. http://dx.doi.org/10.1061/(ASCE)0733-947X(1990)116:1(90)

[10] K. Ivan, I. Haidu, J. Benedek, and S.M.Ciobanu,*Identification of traffic accident risk-prone areas under low-light conditions* Nat. Hazards Earth Syst. Sci.,pp. 20592068, 2015. http://dx.doi.org/10.5194/nhess-15-2059-2015

[11] Zajac, S. S., Ivan, J. N., *Factors influencing injury severity of motor vehicle-crossing pedestrian crashes in rural Connecticut* Accident Analysis and Prevention, 35(3), pp. 369-379,2003. http://dx.doi.org/10.1016/S0001-4575(02)00013-1

[12] Milton, John C., Shankar, Venky N., Mannering, Fred L., *Highway accident severities and the mixed logit model: An exploratory empirical analysis* Accident Analysis and Prevention, Volume: 40, Issue: 1, pp. 260-266, 2008. http://dx.doi.org/10.1016/j.aap.2007.06.006

[13] Cass DT, Ross F, Lam L., *School Bus Related Deaths And Injuries In New South Wales* Med J Austr;166(2), pp.07-108,1997. PMID: 8709875

[14] Darrell.S, Dana.H, *Gender, structural disadvantage, and urban crime:do macro-social variables also explan female offending rates* Criminology,volume38,number 2, pp.403-438,2000. http://dx.doi.org/10.1111/j.1745-9125.2000.tb00895.x

[15] Judith R.Blau, Peter M.Blau, *The cost of inequality: metropolitan structure and violent crime* American sociological review1982, Vol47,pp:114-129,1982. http://dx.doi.org/10.2307/2095046

[16] Adam Krasuski, *A framework for Dynamic Analytical Risk Management at the emergency scene. From tribal to top down in the risk management maturity model* Proceedings of the 2014 Federated Conference on Computer Science and Information Systems, ACSIS,Vol.2,pp.323–330.,2014. http://dx.doi.org/10.15439/2014F371

[17] Yau, K. K. W., *Risk factors affecting the severity of single vehicle traffic accidents in Hong Kong* Accident analysis and prevention,36(3), pp.333-340,2004. http://dx.doi.org/10.1016/S0001-4575(03)00012-5

[18] ODonnell, C. J., Connor, D.H., *Predicting the severity of motor vehicle accident injuries using models of ordered multiple choic* Accident Analysis and Prevention, 28(6), pp. 739-753.,1996. http://dx.doi.org/10.1016/S0001-4575(96)00050-4