WORKING HOURS AND TRAFFIC ACCIDENT INJURIES: CASE STUDY IN BARCELONA

Ahmad Aiash
PhD student, Civil Engineering School, Technical University of Catalonia – BarcelonaTech (Spain)
Francesc Robusté
Professor of Transportation, Civil Engineering School, Technical University of Catalonia – BarcelonaTech (Spain)

ABSTRACT

Working hours in Spain is generally last for 8 hours per day with a maximum of 40 hours per week. Working hours, therefore, can impact the traffic flow and characteristics due to the intensity of road usage by different road users. On the other side, traffic accidents can also be impacted by several temporal factors that may lead to a higher number of traffic accidents or may increase the level of injury alongside other risk factors. In this study, these timings are examined to comprehend the influence of the temporal factors represented by the working hours scheme on traffic accident injuries in Barcelona. Another temporal factor which is the season of the year is included to provide a wider and clearer image for the conducted results and the current situation. The data is collected from the open data service provided by the City Hall of Barcelona. Data preparation and segregation to include the categories of working hours that may lead to a different level of injury resulted from a traffic accident is, firstly, carried. Then, a machine learning model is applied to classify the correlations for both the temporal factors and traffic accidents. Eventually, a Tree Augmented Naïve Bayes model is applied. The results showed that both working hours timing and summer season have higher probabilities of having traffic accidents slight injuries with different medical care assistance provided compared to other timings.

1. INTRODUCTION

Spain has set a comprehensive labor law to organize and regulate the whole working employees' and employers' relationships with covering other aspects including health and safety at work, Social Security, the procedural law, and special relationships conditions for employment (de Vivero, 2019). As part of these regulations and rules, the maximum number of hours per week has been set to be 40 hours which is estimated based on an average over an annual period. Generally, the working time periods each day is set to start at 9:00 a.m. and end at 7:00 p.m. with having two hours break during the working period. This break period usually starts at 2:00 p.m. and ends at 4:00 p.m. for most employees. These timings, indeed, can be a reason for traffic congestion or flow intensity on different roads due to the fact that several groups of people are heading to different destinations at the same time throughout the day.

Real- time traffic data can be affected by different temporal factors similar to the working hours timings and other temporal factors besides the spatial factors. Several studies have examined the effect of the temporal factors on traffic real-time data. In some European countries, holidays' periods shown lower traffic compared to other seasons (Stathopoulos & Karlaftis, 2001). Peak hours have shown to be another influential factor related to traffic flow. Based on a study that is conducted to examine traffic flow patterns in Shanghai, China (Yang, Wu, Xu, & Yang, 2019), the results showed that peak hours had worse traffic congestion compared to non-peak hours.

Traffic accidents, on the other hand, can also be correlated with different temporal factors that can increase the occurrence of them or affect the level of injury. A study (Kashani & Zandi , 2020) shown that weekends can be correlated with a higher number of traffic accidents compared to weekdays.

The same study also revealed that hot timing had more accidents during weekdays. These timings were found to be 8 a.m. and 2 p.m. While for weekends, these timings differ from weekdays' hot timings. 1 p.m., 8 p.m., and 10 p.m. were the hot timings for accidents during weekends. Slight injuries accidents were found to be more likely to happen when the traffic flow is high (Quddus, Wang, & Ison, 2010). Similar to traffic flow, traffic accidents were found to vary according to the season of the year (Harirforoush, 2017) (Le, Liu, & Lin, 2019). Morning, afternoon, or late-night timings were found to have lower probabilities of having fatal injuries compared to early morning timing in Cartagena, Colombia (Cantillo, Márquez, & Díaz, 2020). Consistently, time was found to be an important independent variable related to traffic accidents severities (Li, Prato, & Wang, 2020). As mentioned earlier, traffic congestion can be highly affected by the timing of the day. Therefore, a study (Hyodo & Todoroki, 2018) showed congested and the mixed flow state can increase the risk of having traffic accidents different types including property damages and slight injury accidents.

The objective of conducting this study is to examine the influence of certain temporal factors represented by the working hours and the season of the year in Barcelona, Spain. Data collected and preparation process is carried. Then, a Bayesian network is applied to the exploited data to extract the results.

2. METHODOLOGY

2.1 Data description

The data the is exploited in this study is collected from the Barcelona open data service which is called Barcelona's City Hall Open Data Service (Ajuntament de Barcelona's open data service, 2019). The data that is gathered consists of different traffic injuries. In this study, slight injury levels with different categories are selected for this purpose that occurred in Barcelona in 2019.

Two temporal factors are selected for the objective of this study including the working hours timing and the season of the year. Beginning with the season potential temporal risk factor, the four seasons are included in the examined period and categorized based on their climate (The main climate data of Barcelona, 2020). The summer season is considered to start in May and end in August. Followed by the autumn season that begins in September and ends in November. The winter season is considered to start in December and end in February. Spring, therefore, is considered to start in March and end in April. For the working hours, two categories are included: during work and the other rest time. Working hours are considered to start at 9 a.m. and end at 7 p.m. with having two hours break. These two hours are considered to start at 2 p.m. and end at 4 p.m. with considering them with the other rest category when the initiation of the analysis part is established. However, during the data analysis, this time-period is considered to start from 2:01 p.m. to 4 p.m.

These previously mentioned categorizations are for the independent variables, while for the dependent variable, three categories are included. The first category is represented by the person who had a slight injury with medical assistance. The second category is represented by the person who had a slight injury but rejected health care. The third and last category is represented by the person who had a slight injury with having a hospitalization up to 24 hours. Table 1 is displaying the two different independent variables' general statistics. The total number for both predictors is 11620 accidents that occurred in 2019 in Barcelona and belonged to the slight injuries that required different levels of medical care. As mentioned earlier, the working hours variable consists of two categories, while the season has three categories. Both predictors are considered as categorical type variables alongside the dependent variable when the analysis part is carried.

	Count	Mean	Min	Max	Range	Variance	Standard Deviation	Standard Error of Mean
Working hours	11620	1.55	1	2	1	0.247	0.497	0.005
Season	11620	2.68	1	4	3	1.405	1.185	0.011

Table 1: Main independent variables statistics.

2.2 Bayesian network

Bayesian network is a member of probabilistic graphical models (GM)s that provides laconic descriptions of the distribution of joint probability for the given random variables. Two methods can be exploited when applying Bayesian network models when using IBM Watson Studio software platform with utilizing SPSS modeler. Tree Augmented Naïve Bayes and Markov Blanket estimation are both methods that can be utilized. In this study, Tree Augmented Naïve Bayes is implemented to classify the correlation between the two predictors and the level of medical care that is provided for persons that are involved in traffic accidents.

The reason for choosing Tree Augmented Naïve Bayes is for its simplicity as the number of variables is only two and the aim of this study is only to examine the correlation with considering the prediction part. Moreover, the two independent variables are classified under the same category which is the temporal variables category. The classifier of Tree Augmented Naïve Bayes that is exploited through the IBM software platform is based on this book (Friedman, Geiger, & Goldszmidt, 1997). The conditional probabilities are calculated, in general, by SPSS modeler as follows:

$$\{[Pr(Y_i|X_1 = x_1^j, X_2 = x_2^j, \dots, X_n = x_n^j) = (Pr(Y_i)(X_1 = x_1^j, X_2 = x_2^j, \dots, X_n = x_n^j|Y_i))/ (Pr(X_1 = x_1^j, X_2 = x_2^j, \dots, X_n = x_n^j))], [\propto \Pr(Y_i) \prod_{k=1}^{n} Pr(X_k = x_k^j|\pi_k^j, Y_i)]\} \quad (1)$$

For $d_j = (x_1^j, x_2^j, \dots, x_n^j)$ where d is the data set. d_j is the case that is classified to which it belongs to i^{th} target category, which in this study, the target Y_i is the medical care three levels provided for slightly injured persons. x is the predictor. n is the number of predictors which is two in this study. K is the number of non-redundant parameters. π_k is the parent set of the independent variable alongside the dependent variable, it maybe empty for Tree Augmented Naïve Bayes. The conditional probability is $Pr(X_k = x_k^j|\pi_k^j, Y_i)$ that is associated with each node, which in this study, there are two nodes for predictors.

2.3 Results and discussions

The structure of the applied Tree Augmented Naïve Bayes structure consists of three nodes including medical care level, season, and time which is the working hours as shown in figure 1. The parent node for the season independent variable is only the medical care level. For the time node, this node is linked to two nodes including medical care level and the season. This makes sense since the fact that working hours are already part of the season of the year. Time has a higher importance value compared to season based on the applied model.

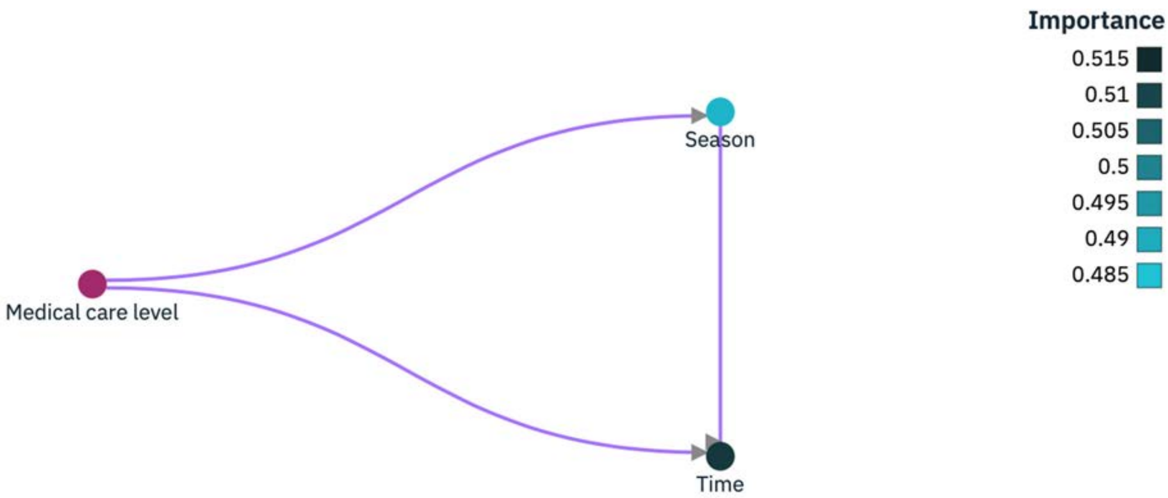

Figure 1: Tree Augmented Naïve Bayes structure.

Table 2 and table 3 are both displaying the estimated conditional probabilities based on the applied Tree Augmented Naïve Bayes for season and time, respectively. For the season, summer has a higher conditional probability compared to all other seasons of the year, followed by autumn, winter, and lastly, the spring season. For the time variable, the working hours category has the highest conditional probabilities during all seasons for all medical care categories that are provided for the slightly injured person except spring season when a person got a slight injury, but this injured person rejected medical assistance.

Parents		Season Probability			
Medical care level	Autumn	Winter	Spring	Summer	Total
With medical assistance	0.245	0.233	0.180	0.341	2,263
Rejected health care	0.243	0.231	0.188	0.338	420
Hospitalization up to 24 hours	0.255	0.235	0.164	0.346	5,428

Table 2: Season conditional probability.

Parents		Time Probability		
Season	Medical care level	Working hours	Rest of the day	Total
Autumn	With medical assistance	0.514	0.486	555
Autumn	Rejected health care	0.520	0.480	102
Autumn	Hospitalization up to 24 hours	0.525	0.475	1,384
Spring	With medical assistance	0.576	0.424	408
Spring	Rejected health care	0.468	0.532	79
Spring	Hospitalization up to 24 hours	0.552	0.448	888
Winter	With medical assistance	0.547	0.453	528
Winter	Rejected health care	0.619	0.381	97
Winter	Hospitalization up to 24 hours	0.558	0.442	1,277
Summer	With medical assistance	0.585	0.415	772
Summer	Rejected health care	0.549	0.451	142
Summer	Hospitalization up to 24 hours	0.559	0.441	1,879

Table 3: Time conditional probability.

3. CONCLUSIONS

The fact that there are enormous factors that influence traffic accident occurrences and severities is leading to conducting several studies to examine these potential risk factors. Temporal factors are part of these factors that can impact traffic accidents. Working hours and the season of the year are part of these temporal factors that may lead to traffic accidents. Therefore, this study has exploited traffic accident injuries by focusing on the level of medical care that is provided for the slightly injured person in Barcelona in the year 2019.

Tree Augmented Naïve Bayes based on Bayesian network is employed to classify the correlations between the two predictors and the dependent variable.

Four seasons are included to understand its temporal impact on the three levels of medical care. Two timings are included the working hours which starts at 9 a.m. and ends at 7 p.m. with having 2 hours break that is not included in this category which is from 2 p.m. to 4 p.m. period. The results show that the summer season has higher conditional probabilities of having slight injuries with including different medical care assistance compared to other seasons. For the timing, the working hours period, similar to the summer season, has the highest conditional probabilities for traffic accidents with being involved in slight injuries with different medical care assistance. For the future work and based on the concluded results, data from delivery operating firms is needed to grasp the impact of these different timings on traffic accidents occurring while the operation is maintained. Then, similar data analysis can be carried out to detect the impact of working hours on this category of employees as they may have a higher risk of being involved in traffic accidents compared to other categories during these timings. Other levels of injury resulted from traffic accidents may also be considered in the future work.

REFERENCES

AJUNTAMENT DE BARCELONA'S OPEN DATA SERVICE. (2019). Retrieved from https://opendata-ajuntament.barcelona.cat/en/

CANTILLO, V., MÁRQUEZ, L., & DÍAZ, C. (2020). An exploratory analysis of factors associated with traffic crashes severity in Cartagena, Colombia. Accident Analysis & Prevention, 146.

DE VIVERO, S. (2019). Employment law overview Spain 2019-2020. Barcelona: L&E Global.

FRIEDMAN, N., GEIGER, D., & GOLDSZMIDT, M. (1997). Bayesian Network Classifiers. Machine Learning, 29, 131–163.

HARIRFOROUSH, H. (2017). Thesis: An Integrated GIS-Based and Spatiotemporal Analysis of Traffic Accidents: A Case Study in Sherbrooke. Université de Sherbrooke.

HYODO, S., & TODOROKI, T. (2018). An Analysis of Risk Factors for Rear-Ender Accident on Urban Expressway Considering Accident Severity. Transportation Research Procedia, 34, 203-210.

KASHANI, A., & ZANDI, K. (2020). Influence of Traffic Parameters on the Temporal Distribution of Crashes. Transportation Engineering, 24, 954–961.

LE, K., LIU, P., & LIN, L.-T. (2019). Determining the road traffic accident hotspots using GIS-based temporal-spatial statistical analytic techniques in Hanoi, Vietnam. Geo-spatial Information Science, 23(2), 153-164.

LI, L., PRATO, C., & WANG, Y. (2020). Ranking contributors to traffic crashes on mountainous freeways from an incomplete dataset: A sequential approach of multivariate imputation by chained equations and random forest classifier. Accident Analysis & Prevention, 146.

QUDDUS, M., WANG, C., & ISON, S. (2010). Road Traffic Congestion and Crash Severity: Econometric Analysis Using Ordered Response Models. Journal of Transportation Engineering, 136(5), 424-435.

STATHOPOULOS, A., & KARLAFTIS, M. (2001). Temporal and Spatial Variations of Real-Time Traffic Data in Urban Areas. Transportation Research Record: Journal of the Transportation Research Board, 1768(1), 135-140.

THE MAIN CLIMATE DATA OF BARCELONA. (2020). Retrieved from https://www.barcelona.de/en/barcelona-climate-travel-season.html

YANG, S., WU, J., XU, Y., & YANG, T. (2019). Revealing heterogeneous spatiotemporal traffic flow patterns of urban road network via tensor decomposition-based clustering approach. Physica A: Statistical Mechanics and its Applications, 526.