# Risk Assessment of Pedestrian Accident Area Using Spatial Analysis and Deep Learning

#### Ilyoung Hong\*, Hanseung Choi, Songpyo Hong

Department of GIS Engineering, Namseoul University, Republic of Korea.

\* Corresponding author. Tel.: 041-580-2307, +82; email: ilyoung.hong@nsu.ac.kr Manuscript submitted November 10, 2017; accepted November 30, 2017. doi: 10.17706/ijeeee.2018.8.4.248-256

**Abstract:** The purpose of this study is to construct the spatial characteristics of pedestrians' accident points as learning materials and to develop a model for deep learning them and to predict the risk of similar areas. In order to construct learning materials for pedestrian accident points, the characteristics of geometric factors related to pedestrian accidents at the accident points were quantitatively analyzed through spatial analysis. As a result of the study, in the analysis of the deep learning pattern, the prediction results could be improved by changing the middle layer and adjusting the parameters. It is expected that this result will contribute to lowering the accident rate of pedestrians by preparing measures for high accident rate points by reviewing vulnerable traffic accident sites in the future.

Key words: Spatial analysis, deep learning, transportation data, pedestrian accident.

## 1. Introduction

According to Korea's 2014 traffic accident statistics, the number of traffic accidents was 4,762, while the number of deaths during the walk was 1,910, accounting for about 40.1% and about two times higher than the OECD average of 19.5%. In particular, the number of deaths from traffic accidents in Seoul in 2014 is 400, and the number of deaths during walking is 220, 55%. As such, pedestrian safety problems in Seoul are serious and need to be managed continuously.

In the meantime, studies have been conducted to analyze the traffic accident data and to find out the cause of the accident in order to identify and prevent pedestrians' major accidents [1]-[3]. They identified the factors related to the accident and analyzed how these factors are distributed in the pedestrian accident area. Through these studies, pedestrian traffic accidents are known to be affected by geometrical environmental factors and regional characteristics of roads [4], [5]. In other words, pedestrian traffic accidents emphasize that there is a high correlation with the environmental factors of the accident site.

On the other hand, information on many traffic accidents that are happening every year is being collected in a database and becoming big data. In recent years, machine learning has attracted attention as a method of analyzing these big data [6]. Especially, deep learning, one of artificial intelligence techniques, is making various use cases through higher analytical ability [7]. The data related to traffic accidents are also a kind of big data and it is necessary to analyze the characteristics of massive data. However, there is little research to apply the method of machine learning in the analysis of traffic data including traffic accidents.

In this study, we applied the deep learning method, which is one of the machine learning methodologies, to analyze the pedestrian accident risk area. In order to utilize the spatial variables in the selection of

learning materials for machine learning, the factors selected through the spatial analysis technique were derived and the data set was constructed. In addition, we applied deep learning which can derive appropriate parameters in the complex structure of various data, and then derived a model using features of pedestrian's accident hazard area through the learned model. This study is meaningful in that it presents a method of analyzing vast amount of data in future pedestrian traffic accident analysis. In addition, it is expected that more advanced analysis will be possible if the problems of analysis data presented in future research are secured.

## 2. Methodology

### 2.1. Data Collection

The purpose of this study is to develop a model for machine learning based on spatial characteristics of pedestrian accident points, and to predict the risk of similar areas. In order to construct the learning data for the pedestrian accident points, the characteristics of the geometric elements related to the pedestrian accident at the accident point were quantified through the spatial analysis and used as independent variables.

The geometric factors affecting pedestrian accidents can be divided into the street environment of pedestrian accident points and the characteristics of street facilities. First of all, the probability of pedestrian traffic accidents is higher when the number of lanes is larger and the lane width is larger, and the higher the speed limit, the higher the probability of occurrence. Therefore, lane number, lane width, and speed limit were selected as independent variables related to the street environment.

Second, it is known that the number of crosswalks and bus stops increases the number of accidents [8]. In addition, it is found that the street facilities such as subway stations and intersections have significant values, and especially if the subway stations are densely populated, it is confirmed that the pedestrian accident probability is high because the flow population is large [9]. The number of pedestrian crossings, bus stops, subway stations, intersections, average distances, and standard deviations of distances were added as independent variables for pedestrian accidents. Here, the pedestrian accident point uses the data of the pedestrian accidents point provided by the traffic safety information management system as the dependent variable. Table 1 shows the independent variables and dependent variables composed of the street environmental variables and the street facility variables. The data collected for these independent variables and dependent variables were collected using public data source.

Table 1. Table Organization of Point, Line and Polygon							
Classification		Data	Data type	Data source	Variable		
Indep ende nt varia ble		Bus stop location Point Topis(http://topis.seoul.go.kr) information (34,091)		Topis(http://topis.seoul.go.kr)	Count Mean distance Standard deviation		
	Street facilitie s	Subway Station location information (366)	Point	Seoul open data plaza(http://data.seoul.go.kr)	Count Mean distance Standard deviation		
		Intersection location information (6,848)	Point	Intelligent Transportation System(http://nodelink.its.go.kr)	Count Mean distance Standard deviation		
		Crosswalk location information (32,133)	Point	Seoul open data plaza(http://data.seoul.go.kr)	Count Mean distance Standard deviation		
	Street enviro nment	Main road	Line	Intelligent Transportation System(http://nodelink.its.go.kr)	Road width		
		Sub road	Line	Seoul open data plaza(http://data.seoul.go.kr)	Road number Maximum speed		

1 0 1

Dependent	Pedestrian accident black spot (750)	Polygon	TMACS (https://tmacs.to2020.kg)	Hagard
variable			I MACS(IIIIpS.//tillacs.ts2020.ki )	nazalu

#### 2.2. Organizing Table with Spatial Analysis

Machine learning is a technique that learns existing data and predicts it by applying it to new data through learned model. In order to construct such a model, learning materials must be constructed, and learning materials are composed of a number of independent variables and one dependent variable. In this study, the independent variables were used to quantify the factors affecting the accident at the pedestrian accident point and the dependent variable was classified into the degree of vulnerability of the accident, upper, middle, and lower. In the composition of the independent variables, the elements for each pedestrian accident point were extracted through spatial analysis and additional variables were constructed using statistical values for extracted information. First, the process of extracting elements for each pedestrian accident point through spatial analysis is as follows. The pedestrian accident point is converted to a polygon, 50 m buffer. Fig. 1 shows the pedestrian accident point, which includes various factors.

In order to extract only the elements within the pedestrian accident point as shown in the right side of Fig. 1, the elements of the facilities including the pedestrian accident points were first extracted and constituted as one shape file for each pedestrian accident point. For example, if there are 20 pedestrian accident points in one shape file, each of them is split into 20 shape files and stored. Thus one file has one pedestrian accident point that has the various element. After this preprocessing, spatial analysis was performed. The spatial analysis process was created using ArcGIS Model Builder, as shown in Fig. 2. ArcGIS Model Builder provides the ability to work with complex spatial analysis at once, and automate repetitive data [10], [11].



Fig. 1. Extraction of elements within pedestrian accident point.

The spatial analysis process can be roughly divided into A and B stages as shown in Fig. 2. Step A is a step of extracting only elements within a pedestrian accident point, and step B is a step of creating a table by assigning a unique point number to each element. For example, as shown in Fig. 3, step A-1 selects the bus stop in the accident point, and step A-2 extracts only the selected bus stop. In step B, a unique point number is entered in the table of the extracted bus stop.

This method is repeated until each separate pedestrian accident point is performed, and one table is created for each element, and the table is accumulated every time it is repeated. The table consists of six tables in total, and it creates four tables in the form of a point and two tables in the form of a line. The table in dot form shows the unique number (Buffer\_ID) and the X and Y coordinates of the pedestrian accident

point classified as shown in Table 2 and the line-shaped table shows the X and Y coordinates as shown in Table 3 and Table 4. Thus, it means that the pedestrian accident points having the same unique number are the same area. In the next calculation step, the average and standard deviation between the same unique numbers are calculated.



Fig. 2. Model builder of spatial analysis.



Fig. 3. Model builder process example.

Table	Table 2. Table Organization of Point Type						
OBJECTID	Х	Y	Buffer_ID				
0	196708.3548	543299.1057	0				
1	196713.9553	543317.8077	0				

Ta <u>bl</u>	e 3. Table Organiza	tion of Line Type	<u>e with Road Wid</u> th
0	BJECTID	Road width	Buffer_ID
0		20	0
1		25	0

OBJECTID	Lane number	Maximum speed	Buffer_ID	
0	3	60	0	
1	4	60	0	

Through the spatial analysis, the learning materials were constructed by using each data converted into the text format. MATLAB programmed the process of batch processing for multiple data in data transformation. The number of each element, the pedestrian accident point, and the distance between each element (pedestrian accident point, pedestrian crossing, intersection) corresponding to the corresponding ID of the pedestrian accident point are matched as shown in Fig. 4. Euclidean distance averages and standard deviations were calculated. The road width, lane, and speed limit were calculated by using the number of main roads and property information.

BUS							Subway				
FID		X Y		B	ufferID	FID		Х	Y	BufferID	
0	196	708.3548	543299.10	57	1	0	183387.1969 196567.785		551311.4071 552447.9794	1	
1	196	713.9553	543317.80	77	1	1					
2	196708.3548		543299.1057		1	2	2 1965		552435.4601	2	
3	196	713 9553	5/13317 80	77	1	:	:		÷	:	
4	196	548.8379	552414.08	24	2		1				
•		: :	:		:						
		-	Bus			Sub	way				
Buffe	erID	Count	Mean	SD	Cour	nt Me	an	SD			
0		0	0	0	0		D 🖡	0			
1		4	25.4	5.5	1	43	3.9	0			
2		1	27.7	0	2	14	1.9	0.2			
:		:	:	:	:		:	:	•••		

Fig. 4. Matching of each element according to the corresponding ID of the accident bundle point.

In this way, training data about independent variables was constructed. Finally, the dependent variable should be divided into three categories: upper, middle, and lower, which are vulnerable to accidents. The statistic is defined as '2' and '0' as '1'. The accident risk vulnerability classification is based on the number of accidents at each branch. As shown in Fig. 5, there are 2 to 47 cases. Analysis of these data showed an average of 10, a median of 9, a mode of 9, and a standard deviation of 4.762. Here, the standard deviation is based on the mode, and the accident risk level is classified based on the mode and the standard deviation. The risk level of the accident is shown in Table 5, and the step separated by the mode and standard deviation is shown as Fig. 5.



Table 5. Risk Step of a Lot of Pedestrian Accidents Location

Fig. 5. Number of a lot of pedestrian accidents location.

Tensorflow was used to perform deep learning analysis by using spatial analysis for independent variables and statistical methods for building dependent learning data. Tensorflow is available as a Python-based deep-learning library in Google's open source format since 2015 and has been successfully used in areas such as image and speech recognition [12]. Tensorflow provide specialized functions for computing the multidimensional arrays of nodes and entities in a graphical form of data flow. Tensor flows have the advantage of graphical representation of the flow of data, various computational structures, and Python-based interoperability. The analysis using deep learning plays an important role in enhancing the accuracy of the analysis data because the accuracy of the learning and prediction results depend on the learning rate of the data input method and the learning data.

#### 3. Experiment

In this study, we applied the multinomial logistic regression, which is the supervisory classification analysis of multi-class classification using multidimensional variables. For the analysis of deep learning, the input of the data can only increase the efficiency of learning and testing if the relative deviation between the input data is small. In this study, data values were normalized to values between 0 and 10. In the experiment, the middle layer was experimented with 5 nodes, and the Relu function was used as the activation function. The environment of this experiment uses the Ubuntu operating system, Python 3.4-based Tensorflow.

Experiments were applied to variables such as bus stop, subway station, intersection, crosswalk, street environment, subway, main road width, lane number, maximum speed, and road environment. A total of 20 variables obtained from the spatial analysis were used to construct the learning data and the analysis using the Tensorflow was performed. Fig. 6 shows the experiment using Tensorflow. First, the value of 76.53 was obtained by setting 2 layers and 5 layers in the middle layer. As shown in Fig. 7, the result shows that the green color is the same as the previously set risk level, and the red color shows the difference with the deep

#### learning results.







Fig. 7. Deep learning results for accident area.

Next, the correlation coefficients were compared while varying the middle layer values of deep learning to 3, 5, and 7. When the variables of the subgroups are removed from the correlation variable values, the prediction results are somewhat improved. In the case of the hidden layer setting of the middle layer, when the middle layer is changed in the order of 3, 5, and 7, the prediction result is slightly improved to 76.3, 76.4, 75.7.

In this experiment, the limitation of the analysis is limited in that it excludes the personal conditions affecting the pedestrian accident and considers only the environmental factors of the accident point. In addition, among the environmental factors affecting the traffic accident, only the most representative factors among the street environment and road characteristics affecting the accident were considered and the parameters for the safety facilities such as bumpers, pedestrian fences and child protection areas were not included. In addition, complementary research is needed to improve the analytical model, such as analysis of errors in the process of analyzing learning data of machine learning and the process of adding learning materials to improve the model. It is expected that better results will be obtained if these points

are supplemented by future analysis.

#### 4. Conclusion

In this study, we analyze the relationship with environmental variables affecting pedestrian traffic accidents and study the deep learning analysis for predicting pedestrian traffic accidents. As a result of the study, in the analysis of the deep learning pattern, the prediction results could be improved somewhat by changing the middle layer and adjusting the parameters. It is expected that this result will contribute to lowering the pedestrian accident rate by preparing measures to prevent accidents by supplementing the safety facilities for high accident points by reviewing vulnerable traffic accident sites in the future.

Through this study, it can be confirmed that it is possible to make a meaningful analysis through the results of the artificial neural network learning of deep learning in the analysis of vulnerability of pedestrian traffic accidents. However, the prediction probabilities are still lower than those of conventional artificial neural networks. This problem requires the necessity of improvement in the selection of variables in the existing data collection and the exploratory analysis process of more variables in the analysis. In addition, if we collect more training materials for analysis and improve the model, we will be able to create a model with a higher level of accuracy.

#### References

- [1] Haleem, K. Alluri, P., & Gan, A. (2015). Analyzing pedestrian crash injury severity at signalized and non-signalized locations. *Accident Analysis & Prevention*, *81*, 14-23.
- [2] Hess, P. M., Moudon, A. V., & Matlick, J. M. (2004). Pedestrian safety and transit corridors. *Journal of Public Transportation*, *7*(*2*), 73-93.
- [3] Mark, J. K. (2009). Illegal pedestrian crossing at signalized intersections: Incidence and relative risk, Elsevier, *Accident Analysis & Preventions*, 485-490.
- [4] Dumbaug, E., & Li, W. (2011). Designing for the safety of pedestrians, cyclists, and motorists in urban Environments, *Journal of the American Planning Association*, *77(1)*, 69-88.
- [5] Ewing, R., & Dumbaugh, E. (2009). The built environment and traffic safety: A review of empirical evidence. *CPL bibliography*, *23*(*4*), 347-367.
- [6] Najafabadi, M. M., Villanustre, F., Khoshgoftaar, T. M., Seliya, N., Wald, R., & Muharemagic, E. (2015). Deep learning applications and challenges in big data analytics. *Journal of Big Data*, *2*(*1*), 1.
- [7] Costa-Jussa, M. R., Allauzen, A., Barrault, L., Cho, K., & Schwenk, H. (2017). Introduction to the special issue on deep learning approaches for machine translation. *Computer Speech & Language*, *46*, 367-373.
- [8] Marshall, W. E., & Garrick, N. W. (2011). Does street network design affect traffic safety? *Accident Analysis & Prevention*, *43*(3), 769-781
- [9] LaScala, E. A., Gerber, D., & Gruenewald, P. J. (2000). Demographic and environmental correlates of pedestrian injury collisions: A spatial analysis, *Accident Analysis & Prevention*, *32(5)*, 651-658.
- [10] Allen, D. W. (2011). *Getting to Know ArcGIS ModelBuilder: 1 Editon*, Esri Press. 380 New York Street, Redlands, California.
- [11] Ding, H., & Tang, L. (2009). The conception and application of spatial processing modeling technology — Implement of spatial data transformation with ArcGIS model builder tools. *Bulletin of Surveying and Mapping*, 1, 64-67
- [12] Abadi, M., Agarwal, A., Barham, P., *et al.* (2015). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *Google Research*. California.



**Ilyoung Hong** is a professor in the Department of GIS engineering at Namseoul University, Cheoan City, South Korea. He holds a Ph.D. in gisciences from State University of New York at Buffalo, USA. His present research interests include spatial analysis, Location-Based Service and machine learning.



**Hanseung Choi** is a master's degree student in the Department of GIS engineering at Namseoul University, Cheoan City, South Korea. His present research interests include spatial analysis, machine learning and photogrammetry.



**Songpyo Hong** is a master's degree student in the Department of GIS engineering at Namseoul University, Cheoan City, South Korea. His present research interests include spatial analysis, machine learning and photogrammetry.