Road Risk Evaluation based on the Analysis of Crime and Urban Factors: For the Short and Safe Route

1st Daye Kim
Software Convergence
Kyunghee University
Seoul, Republic of Korea
rlaek4793@khu.ac.kr

2nd Juwon Baek
Computer Science and Electrical Engineering
Handong Global University
Pohang, Republic of Korea
21700340@handong.ac.kr

3rd Jihu Yang
Computer Science and Electrical Engineering
Handong Global University
Pohang, Republic of Korea
zihooy@handong.ac.kr

4th Hyun Roh
Software Convergence
Kyunghee University
Seoul, Republic of Korea
yeshyun@khu.ac.kr

5th Heewon Jeong
Computer Engineering
Chungbuk National University
Cheongju, Republic of Korea
jhjmo0719h@chungbuk.ac.kr

6th Bryanna Ruiz
Computer and Information Technology
Purdue University
West Lafayette, IN, USA
ruiz114@purdue.edu

***Abstract*—While certain cities in the United States have higher crime rates, many areas, such as rural areas, have lower crime levels. Given this fact, the focus of this research will be large on creating recommended safe paths to avoid locations with high crime rates. However, major navigation applications do not consider the dangers of a route, only consider the distance of the route. In order to overcome this restriction, this paper proposes a safe path recommendation algorithm that considers not only crime occurrence but also the urban factors that can appear with criminal activity, such as city facility and service request data. (TBD: the result or very simple implementation explanation). Through the algorithm that this paper proposed, a positive effect on reducing extra crime in the United States is expected.**

***Index Terms*—public safety, road risk evaluation, safe route, correlation, criminal activity and urban factors**

I. Introduction

Public safety is one of the most significant concerns in the United States since crime rates are higher than in other OECD countries. According to the OECD BetterLifeIndex, the homicide rate (the number of murders per 100,000 inhabitants) in the United States is 6, which is 2.6 higher than the average of the OECD countries [1]. Although the OECD countries had a downward trend in crime rates since 2010, the United States had an increase in crime rates. However, having a high crime rate in the United States does not mean it is dangerous in every division; for example, South Chicago accounts for 40 percent of Chicago's total crime rate [ref TBD]. Therefore, it is essential to recommend safe routes to avoid dangerous regions.

Google Maps is one of the navigation applications used by most people since it has acquired almost 64 million users [3]. However, Google Maps prioritizes short routes by using Dijkstra's and A* algorithm rather than an algorithm that considers safety [3]. Moreover, the previous research analyzed only the crime data to find a safe route [4], and it did not consider the population or the level of crime in the data analysis. The previous research did not consider the possibilities of future crimes and had flawed logic when analyzing the occurrence of crimes happening in divisions.

This research suggests an algorithm that recommends routes by avoiding dangerous sections using both crime data and the data that could affect the crime rate. The data used in this research includes two types of Chicago crime data, six types of city facility data, such as police stations, and five types of service request data, such as sanitation complaints [5]–[17]. These data sets were analyzed in terms of density through Kernel Density Estimation (KDE) and distance to the facility and service request location. In addition, the crime data was analyzed considering the population in the division and the level of each crime. After analyzing the 13 types of data, the riskiness of each street was calculated and applied to the algorithm. In the algorithm, it avoids the streets that are highly related to the occurrence of crime, even if it is the fastest path to the destination, and considers the streets that are lowly related to the crime. By calculating the riskiness of each feature, the algorithm has the strength to find the safest path.

The rest of the paper is organized in the following manner: Section II will discuss the related work. Section III will introduce the collected and analyzed data, the definition of riskiness, and the data analysis methods. Section IV will discuss the algorithm considering the riskiness of the street, reflecting the data analysis and the implementation of the application using a Google Map API. Section V will evaluate the performance of the proposed algorithm. Finally, Section VI will summarize all previously mentioned and suggest future plans.

II. RELATED WORKS

A. In the aspect of riskiness/safety evaluation of roads

Many studies have been conducted to develop algorithms that recommend safe routes [4], [18]-[20]. *Galbrun et.al* solves the safe path problem, resulting in a collection of paths that offer a trade-off between risk and distance using Philadelphia and Chicago crime data [4]. The authors obtained the crime density at different points by applying a KDE model to the data. Using the crime density, the authors defined max and total risk scores. Max risk scores use the highest crime density of the points included in the edge and total risk, which considers the crime density of the whole point. However, the research has the following limitation: it does not consider the other urban factors that can affect the riskiness of the road but only the crime data. In addition, the paper did not consider real-time data calculating the riskiness of the road.

To reduce the limitation of the research, *Goel et.al* evaluates the safety of the road using: accident history, road quality, availability of police stations and public transportation, and such others, and defines a static safety score and dynamic safety score [18]. The static safety score is calculated by the data set obtained by the author. In contrast, the dynamic safety score is calculated by the static safety score, web crawling data, and user feedback data on the safety of the recommended route. However, although the authors tried to reduce the limitations, parts of the data analysis, which could explain the correlation between road safety and the urban factors that affect the roads' safety, were omitted. Furthermore, the paper does not have enough explanation on calculating the safety score, such as the weight of police stations.

Therefore, this paper has the following novelty: (1) Through the various data analysis methods, it is possible to justify the data used in this paper. (2) Based on the result from (1), it is possible to explain the weight, which was used to calculate the riskiness/safety score of the edge.

III. METHODOLOGY

Several urban factors are likely to appear along with the occurrence of crimes, such as specific urban facilities and service requests. These urban factors can play a significant role in recommending safety routes, as they have the possibility to complement searching for places where a crime can occur. The purpose of this research is to make a safe route recommendation algorithm in consideration of criminal activity and urban factors that could affect the occurrence of illegal activity. In implementing the algorithm, this research focused on identifying urban factors that appear most frequently with criminal acts through data analysis and focused on using urban factors with a clear correlation with crime occurrence. Safe route algorithms considering crime activity data alone have a chance to avoid where crimes have occurred but may not be able to detect areas where crimes are likely to occur in the future. On the other hand, by using both crime occurrence and urban factors, it is possible to avoid identified and identifiable risks in advance. This part covers preprocessing for elements used in algorithms and urban factor selection through correlation analysis.

A. Data Collection

Chicago crime data set [5], [6] which reflects reported incidents of crime that occurred in Chicago from 2001 to 2022, was found. The Date, Primary Type, IUCR, Longitude, and Latitude were extracted from the data set, and thirty-six primary types of crime exist, including theft, assault, robbery, and homicide. Six types of city facility data sets were found: Chicago CTA Bus Stops [7], Chicago Police Stations [8], Chicago Fire Stations [9], Chicago Libraries [10], Chicago Public Schools Profile Information [11], and Chicago Family and Support Services Delegate Agencies [12]. Five types of Service request data sets were found: Chicago Vacant and Abandoned Buildings (Service Request) [13]; Data set contains all 311 calls for open and vacant buildings reported to the City of Chicago since January 1, 2010, Chicago Street Lights All Out (Service Request) [14]; All open reports of "Street Lights - All Out" (an outage of 3 or more lights) made to 311 and all requests completed since January 1, 2011, Chicago Sanitation Code Complaints (Service Request) [15]; All open sanitation code complaints made to 311 and all requests completed since January 1, 2011, Chicago Graffiti Removal (Service Request) [16]; All open graffiti removal requests made to 311 and all requests completed since January 1, 2011, and Chicago Shot Spotter Alerts [17]; Data set contains all Shot Spotter alerts since the introduction of Shot Spotter to some Chicago Police Department districts in 2017.

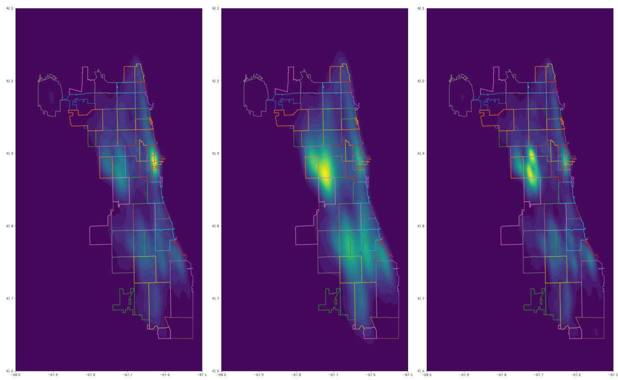
Since Chicago crime data will be preprocessed considering the number of people living in 2018, only the data from 2018 was extracted. Longitude and Latitude were extracted from each city facility and service request data sets.

B. Find the Relation with each Data set

1) Crime Data Preprocessing: To find a relationship between crime occurrence and urban factors, it is especially notable that each crime has a different risk level. In an effort to reflect this, two methods were conducted. One way is to extract only the violent data from the crime data set, and the other is to give different weight to each crime data based on severity. For the first way, data corresponding to the violent index specified by the Chicago Police Department were extracted [21]. The extracted primary data types were Homicide, Criminal Sexual Assault, Robbery, Aggravated Assault, and Aggravated Battery. For the second way, weight is given for each data point of the crime data set based on their sentence. IUCR (Illinois Uniform Crime Reporting) number, which means specific descriptions of the crime types, was used to assign the sentence for each data. The same IUCR numbers were grouped in each primary type, and IUCR numbers that had the maximum count were extracted. With the extracted IUCR numbers, each primary type's class was defined based on the Illinois General Assembly by the Legislative Information System [22]. The relative risk of the primary type was calculated according to the maximum sentence for each class, as shown in Table 1.

TABLE I: Relative Risk of each class

Class Type	Maximum Sentence	Weight
Class C	Under 30 days	0
Class B	Under 6 months	0
Class A	Under 1 year	1
Class 4	Under 3 years	3
Class 3	Under 5 years	5
Class 2	Under 7 years	7
Class 1	Under 15 years	15
Class X	Under 30 years	30
Class M	Under 60 years	60



(a) Original data (b) Violent data (c) Weighted data

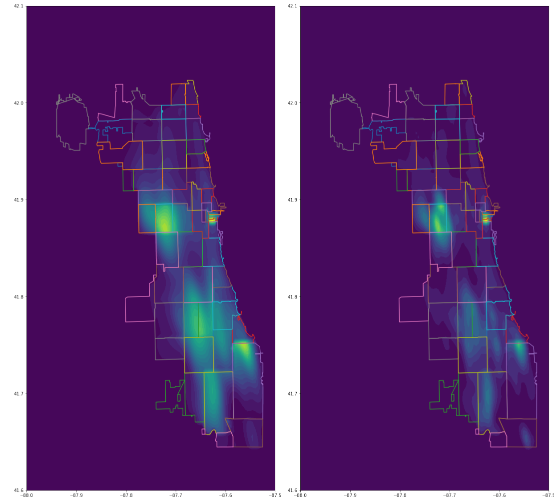
Fig. 1: KDE for crime density in Chicago

With the data from methods 1 and 2, the crime density of the entirety of Chicago was estimated since it could be a safety score. Gaussian Kernel Density Estimation (KDE) was applied to the Chicago crime data set's geographic coordinates to obtain the crime's spatial density. KDE is the non-parametric method for estimating an unknown probability density function with observational data using the kernel function K. Given x points of each crime, the Gaussian kernel function K estimates crime density based on bandwidth h on the 2-dimensional map at the point as follows:

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \quad (1)$$

In other words, the objective of KDE is to find the probability of a density function for a given data set by replacing each data with a kernel function. In KDE, Gaussian sigma is called the bandwidth parameter h because it dictates the spread of the kernel. Bandwidth defines the smoothness of the estimated density. The smaller the h value, the more detailed it is captured and the sharper it looks, while the large values of h lead to smoother estimation. In this study, KDE was applied to crime data from both violent and weighted methods. For data from the weighted method, additional data points were newly generated by weights assigned to each crime data point of the same coordinate.

Fig 1 visualizes the estimated crime density for Chicago. The brighter the area, the higher the frequency of crime. The



(a) Violent data (b) Weighted data

Fig. 2: KDE for crime density without population impact

estimated crime density of the original data, which is not reflecting the risk of crime, is considered relatively dangerous in downtown Chicago. However, the estimated crime density of the violent and weighted methods, which reflects the risk of each crime, is considered relatively dangerous in western and southern Chicago.

Since the population in each area could affect the crime occurrence, the population is excluded from the data to find influential urban factors more clearly. For this, the number of people living in each ZIP code is collected [23], and crime density is newly calculated as follows :

$$Adjust\ crime\ density = \frac{Existing\ crime\ density \times area\ of\ zipcode}{number\ of\ people\ living\ in\ zipcode} \quad (2)$$

The plot for the Chicago crime density considering population normalization is shown in Figure 2.

2) Calculate Correlation: Significant urban factors for implementing the algorithm were found through the correlation between urban factor density and crime occurrence density. The tendency was verified by the correlation between distance with the nearest urban factor and crime occurrence density.

The calculated crime density with KDE is a 100x100 matrix, and each cell contains location and crime density information. Cells were used to calculate the relation between crime occurrence and urban factors. With this crime density value, two methods are conducted. First, the correlation is obtained by calculating the Euclidean distance between each cell and each urban factor closest to it. The distance is calculated as the equation below.

$$distance = \sqrt{(x_{cell} - x_{factor})^2 + (y_{cell} - y_{factor})^2} \quad (3)$$

x means longitude, and y means latitude of each data. After getting the Euclidean distance from each 100x100 cell, the Pearson correlation between crime density and the distance of the corresponding cell was calculated. Second, Gaussian KDE was applied to the geographic coordinates of each urban factor to get the spatial density. Similar to crime density, the calculated density of the urban factor is a 100x100 matrix. In each cell of the matrix, Pearson correlation was calculated between the density of each urban factor and the crime density. The results are shown in Table 2.

TABLE II: Results of correlation

Factor\Case	Weighted & KDE	Weighted & Distance	Aggravated & KDE	Aggravated & Distance
Police	0.65	-0.38	0.66	-0.42
Bus	0.6	-0.34	0.68	-0.38
Fire Station	0.61	-0.36	0.65	-0.40
Library	0.61	-0.37	0.65	-0.40
Public School	0.69	-0.36	**0.75**	-0.40
Family Support	**0.71**	-0.36	**0.74**	-0.36
Light Out	**0.73**	-0.35	**0.78**	-0.39
Sanitation	**0.71**	-0.35	**0.76**	-0.39
Graffiti	0.27	-0.34	0.23	-0.38
Shot Spotter	**0.83**	-0.40	**0.89**	-0.45
Abandoned Building	**0.77**	-0.36	**0.93**	-0.40

In terms of results from the Gaussian KDE method, all urban factors have a high positive correlation for weighted and aggravated types of crime densities. This indicates that the higher the density of urban factors, the higher the crime density. In addition, in the case of the Distance method, all urban factors negatively correlate with each type of crime density. This shows that the closer to urban factors, the higher the crime density. The results from the two aspects indicated the same tendency in the end: The place where certain urban factors are located has a possibility to be more dangerous.

The results with a correlation of 70% or more in the Gaussian KDE method are denoted in bold. Urban factors with more than 70% correlation in both weighted and aggregated types of crime density are Shot Spotter, Abandoned Building, Light Out, Sanitation, and Family Support. These five features were used in algorithms to find safe routes.

IV. IMPLEMENTATION

Risk perception could vary from person to person. For example, a family with a child is likely more vulnerable to kidnapping than a single person. Taking this into account, the algorithm considers information about the degree of risk perception from a user and recomputes the KDE value of Chicago, which means riskiness in that area. If the user does not enter any information about risk perception, the path is recommended based on the default risk weight this research suggested. The default risk weight is defined based on the Illinois General Assembly by the Legislative Information System [22]. This criterion was discussed in Section 3.

A. Road Networks

Existing road networks in Chicago were extracted from OpenStreetMap (OSM). OSM is an open-source and collaborative platform that aims to make world map data freely available. Many tools in OSM have strength in several ways, including route planning. For our purposes, 'OSMnx' was used, which is the Python library that exports the road network map from an OSM format to a graph-based format. In this format, each node represents an intersection in the street, while each edge corresponds to a load segment that connects the intersections. Each edge and node is annotated with various features. The following are interests for our study:

Edge;

- Length: The physical distance of the corresponding road segment.
- Risk1: The score of calculated risk from the city facility and service request.
- Risk 2: The score of crime occurrence density.
- Geometry: A set of longitude/latitude points from road segments.
- Accessibility flag: The indicators that define accessibility of road segments such as car, bicycle, and foot.

In this research, the pedestrian street network of Chicago was considered. According to the Chicago OSM map, the number of nodes and edges are 28699 and 76160. The average of the edge's geographic distance is 136.52327861081704 m.

B. Following is TBD:

- Additional information with the above will be described.
- About How to combine three features in edge, the Length, Risk1, and Risk2.
- By doing this, One crucial feature is defined, and It is used to build an algorithm.
- determined Node information

B) Algorithm and Implementation:

- Description of algorithms and and tools used in OSM
- Process and results after entering the feature into the OSM tool.
- A description of the Google API we used to develop the app.
- How to select nine nodes from OSM results
- Results of handing over nine waypoints to Google API

V. EXPERIMENTS

TBD: Experiments

VI. CONCLUSION

TBD: Conclusion

REFERENCES

[1] J. Horton, "US crime: Is America seeing a surge in violence?," *BBC News*. [Online]. Available: https://www.bbc.com/news/57581270. [Accessed: 28-Sep-2022].
[2] TBD
[3] R. Dicker, "3 new ways to navigate more sustainably with maps," *Google*. [Online]. Available: https://blog.google/products/maps/3-new-ways-navigate-more-sustainably-maps. [Accessed: 28-Sep-2022].
[4] E. Galbrun, K. Pelechrinis, and E. Terzi, "Urban navigation beyond shortest route: The case of safe paths," *Information Syst.*, vol. 57, pp. 160–171, Apr. 2016.
[5] "Crimes - map: City of chicago: Data Portal," *Chicago Data Portal*. [Online]. Available: https://data.cityofchicago.org/Public-Safety/Crimes-Map/dfnk-7re6. [Accessed: 15-Sep-2022].

[6] "Crime Data Explorer," *FBI*. [Online]. Available: https://crime-data-explorer.fr.cloud.gov/pages/explorer/crime/crime-trend. [Accessed: 28-Sep-2022].

[7] "CTA bus stops: City of chicago: Data Portal," *Chicago Data Portal*, 28-Feb-2017. [Online]. Available: https://data.cityofchicago.org/Transportation/CTA-Bus-Stops/hvnx-qtky. [Accessed: 28-Sep-2022].

[8] "Police Stations: City of chicago: Data Portal," *Chicago Data Portal*, 10-Jun-2016. [Online]. Available: https://data.cityofchicago.org/Public-Safety/Police-Stations/z8bn-74gv. [Accessed: 28-Sep-2022].

[9] C. of Chicago, "Fire stations: City of chicago: Data Portal," *Chicago Data Portal*, 21-Aug-2011. [Online]. Available: https://data.cityofchicago.org/Public-Safety/Fire-Stations/28km-gtjn. [Accessed: 28-Sep-2022].

[10] C. P. Library, "Libraries - locations, contact information, and usual hours of operation: City of chicago: Data Portal," *Chicago Data Portal*, 28-Oct-2021. [Online]. Available: https://data.cityofchicago.org/Education/Libraries-Locations-Contact-Information-and-Usual-/x8fc-8rcq. [Accessed: 28-Sep-2022].

[11] "Chicago Public Schools - school locations SY2021 - map: City of chicago: Data Portal," *Chicago Data Portal*. [Online]. Available: https://data.cityofchicago.org/Education/Chicago-Public-Schools-School-Locations-SY2021-Map/9ybt-s3ms. [Accessed: 28-Sep-2022].

[12] C. of Chicago, "Family and Support Services Delegate Agencies: City of chicago: Data Portal," *Chicago Data Portal*, 07-Oct-2015. [Online]. Available: https://data.cityofchicago.org/Health-Human-Services/Family-and-Support-Services-Delegate-Agencies/jmw7-ijg5. [Accessed: 28-Sep-2022].

[13] "311 Service Requests - Vacant and Abandoned Buildings Reported - Historical," *Chicago Data Portal*, 7-Mar-2019. [Online]. Available: https://data.cityofchicago.org/Service-Requests/311-Service-Requests-Vacant-and-Abandoned-Building/7nii-7srd. [Accessed: 28-Sep-2022].

[14] "311 Service Requests - Street Lights - All Out - Historical," *Chicago Data Portal*, 7-Mar-2019. [Online]. Available: https://data.cityofchicago.org/Service-Requests/311-Service-Requests-Street-Lights-All-Out-Histori/zuxi-7xem. [Accessed: 28-Sep-2022].

[15] "311 Service Requests - Sanitation Code Complaints - No Duplicates," *Chicago Data Portal*, 17-Apr-2019. [Online]. Available: https://data.cityofchicago.org/Service-Requests/311-Service-Requests-Sanitation-Code-Complaints-No/rccf-5427. [Accessed: 28-Sep-2022].

[16] "311 service requests - graffiti removal - historical: City of chicago: Data Portal," *Chicago Data Portal*. [Online]. Available: https://data.cityofchicago.org/Service-Requests/311-Service-Requests-Graffiti-Removal/hec5-y4x5/data. [Accessed: 28-Sep-2022].

[17] C. of Chicago, "Violence reduction - shotspotter alerts: City of Chicago: Data Portal," *Chicago Data Portal*. [Online]. Available: https://data.cityofchicago.org/Public-Safety/Violence-Reduction-Shotspotter-Alerts/3h7q-7mdb. [Accessed: 28-Sep-2022].

[18] F. Mata et al., "A mobile information system based on crowd-sensed and official crime data for Finding Safe Routes: A case study of mexico city," *Mobile Information Syst.*, vol. 2016, pp. 1–11, Mar. 2016.

[19] Y. Pang, L. Zhang, H. Ding, Y. Fang, and S. Chen, "Spath: Finding the safest walking path in smart cities," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 7, pp. 7071–7079, May 2019.

[20] Y. Zhao, Y. Xie, and S. Ahvar, "On integration of any urban factor with distance for navigation : Walk safely and fast enough," *2019 IEEE 23rd International Enterprise Distributed Object Computing Workshop (EDOCW)*, 2019.

[21] "Data requests - chicago police department." [Online]. Available: https://home.chicagopolice.org/statistics-data/data-requests/. [Accessed: 6-Oct-2022].

[22] *Illinois General Assembl - legislation*. [Online]. Available: https://www.ilga.gov/legislation/ilcs/ilcs2.asp?ChapterID=53. [Accessed: 6-Oct-2022].

[23] C. of Chicago, "Chicago population counts: City of Chicago: Data Portal," *Chicago Data Portal*, 13-Jun-2022. [Online]. Available: https://data.cityofchicago.org/Health-Human-Services/Chicago-Population-Counts/85cm-7uqa. [Accessed: 13-Oct-2022].