Article

Estimation of Driving Direction of Traffic Accident Vehicles for Improving Traffic Safety

MinJe Cho [1,2], JeongMin Park [3], Siyoon Kim [3] and Yonggeol Lee [4,*]

1   Police Science Institute, Korean National Police University, Asan 31539, Republic of Korea
2   Department of Urban Engineering, Chungbuk National University, Cheongju 28644, Republic of Korea
3   Department of Information Science and Telecommunication, Hanshin University,
    Osan 18101, Republic of Korea
4   School of Computing and Artificial Intelligence, Hanshin University, Osan 18101, Republic of Korea
*   Correspondence: pattern@hs.ac.kr; Tel.: +82-31-379-0656

Abstract: Traffic accidents are widely recognized as being a significant societal disaster that causes severe human, economic, and social damage. While various factors, such as environmental, road, and human factors are commonly associated with traffic accidents, this study proposes a novel approach to consider the directionality of traffic accidents. It examines its impact on traffic safety in Chungcheongnam-do province, South Korea. Using police reports, we estimated the directionality of each accident by calculating the angle between its occurrence site and destination. Our experimental results revealed a statistically significant association between the directionality of traffic accidents and their incidence rates in Chungcheongnam-do province, with a higher frequency of accidents occurring in a Southbound direction. These findings suggest that considering the directionality of traffic accidents could enhance the effectiveness of traffic safety policies and interventions. Further research is needed to explore this approach's underlying mechanisms and potential applications. Nevertheless, this study provides important insights for policy makers and practitioners to develop targeted and practical measures for improving traffic safety.

Keywords: traffic safety; direction estimation; named entity recognition; visualization; Korean National Police Agency

1. Introduction

According to previous research [1–8], traffic accidents are perceived as a social disaster that causes human, economic, and social damages. Every year, many people experience human losses such as death or injury due to traffic accidents, causing emotional distress to accident victims, their families, and society [9]. In addition, traffic accidents entail various expenses, such as medical bills, rehabilitation costs, and insurance payouts for accident settlement. Indirectly, traffic accidents can cause congestion, leading to fuel waste and secondary accidents for surrounding vehicles, further increasing the economic burden [10]. Moreover, traffic accidents generate recovery costs at the national level and require expenses for policy-making and implementation related to traffic safety, posing significant financial and social burdens to society [11]. Therefore, reducing the scale of damages through effective traffic accident prevention measures is crucial to alleviate the overall burden on society and the nation.

Numerous studies have been conducted to analyze factors related to traffic accidents. Generally, traffic accidents are associated with various factors, including environmental factors (vehicle and road conditions), road characteristics, human factors, weather, and lighting conditions, which are widely recognized. In [12], It has been identified that the severity of injuries among motorcycle riders is associated with factors such as speed limits, collision-related factors, driver-related factors, road characteristics, and weather conditions. In one paper [13], an analysis of the leading causes of road traffic accidents in Hail revealed

that 67% were caused to human factors, 29% by road conditions, and 4% by vehicle defects. Ref. [14] analyzed the traffic accident mortality rate on roads in Selangor and Perlis states based on data from 2013 to 2017 and explained that traffic volume, physical road structure, and user concentration are the main factors leading to traffic accidents in certain areas. In another study [15], the authors explored the behavioral factors that pose a higher risk of causing severe or fatal injuries to drivers, selecting four factors related to traffic accidents: vehicle, road, environment, and human factors. In Ref. [16], the study investigated the effects of various factors on nighttime traffic accidents and identified low-light conditions as a significant factor. Specifically, the severity of nighttime vehicle accidents was found to be influenced by accident location, accident type, and the presence of median barriers.

On the other hand, directionality can be considered to be one of the factors associated with traffic accidents. All roads have a Northbound and Southbound direction, and traffic accidents can be biased toward one of these directions. In cases where there is a significant disparity in the ratio of traffic accidents between two opposing directions (Northbound and Southbound) at a specific location, it can provide an opportunity to analyze the factors contributing to the difference in the number of accidents by direction. This analysis helps improve traffic accident safety by exploring additional environmental factors. In certain advanced countries, such as the United States, investigators record directional information for all traffic accidents. However, many countries, including South Korea, still do not collect data on the driving direction of traffic accident vehicles. Consequently, the absence of such data prevents the establishment of traffic policies based on the Northbound and Southbound directions of the roads. It can lead to inefficient allocation of national budgets and police resources, resulting in inefficiencies within traffic safety policies. Moreover, it may cause delays in policy-making for high-risk areas.

In this study, we propose the driving direction of traffic accident vehicles as a new factor. Currently, the Korean National Police Agency (KNPA) in South Korea does not collect information on the driving direction of traffic accidents. Therefore, it is necessary to estimate the driving direction of the vehicles involved in traffic accidents using the data collected related to the directionality of traffic accidents. First, we set the latitude and longitude of the vehicle involved in the accident as the accident location. Second, since the KNPA traffic accident data does not include information on the destination or destination point, we extract this information from the *"Report"* field, which records the accident details. Since the pattern of writing a report is generally fixed, we use Named Entity Recognition (NER) to extract the destination. Third, we convert the destination to latitude and longitude coordinates through geocoding. Finally, we estimate the directionality (Norhthbound or Southbound) of the vehicle involved in the traffic accident by calculating the angle between the accident location and the destination. In the experimental results, we collected 5181 data items reported to the police in the Chungcheongnam-do region, which had a high incidence of traffic accidents, and estimated the directionality for all incidents. We also visualized the direction-specific occurrence of traffic accidents in each Administrative District (AD) using a cumulative histogram on a map. Analyzing the histograms by AD, it was found that in Chungcheongnam-do, accidents occurred more frequently in the Southbound direction.

The remainder of this paper is organized as follows. Section 2 explains the analysis of the KNPA's traffic accident dataset. Section 3 describes the proposed method and each procedure in detail. The experimental results are described in Section 4, and the discussion and conclusions follow.

2. Analysis of Traffic Accident Data from KNPA

In collecting traffic accident data, the KNPA relies on traffic investigators who collect information related to the accident, such as the coordinates of the accident location, photographs, and a general summary. Given that traffic accidents are typically the result of a complex set of factors, the investigation process records data related to the environmental, temporal, vehicle, and human-related factors (Table 1).

Table 1. Traffic accident data.

Field	Descriptions
Date of occurrence	Year, Monday, Sunday, hour, minute, day, night, day
Occurrence Location	City, county, police station, and traffic accident location coordinates (latitude and longitude)
Occurrence Environment	weather conditions, road conditions
Contents of accident	Report of death, serious injury, minor injury, number of vehicles involved
Type of accident	car-to-person, car-to-car, vehicle alone
Cause of accident	The human condition, vehicle condition, road environment condition

First, environmental factors that may affect traffic accidents include road-related factors such as the absence of road markings, faulty signs or traffic lights, slippery road surfaces, and poor lighting. The conditions of the road, such as its width, length, slope, and position of intersections, can also influence the likelihood of an accident [17–21]. Weather conditions such as rain, snow, fog, wind, and dust can reduce driver visibility and make vehicle control difficult, ultimately increasing the risk of accidents [17,19,21]. Additionally, animals on the road, particularly at night, can create unpredictable situations for drivers and increase the risk of accidents. The KNPA records information related to environmental factors such as road surface conditions (paved-dry, paved-wet/moist), weather conditions (clear, rainy, cloudy), and others.

Second, when traffic volume is high during rush hour, the risk of accidents increases [22,23]. Moreover, drivers commuting during sunrise or sunset are vulnerable to accidents. Sunlight is very intense and can directly shine in drivers' eyes, reducing their visibility and increasing the likelihood of accidents. Additionally, nighttime driving reduces visibility, making it difficult for drivers to detect sudden obstacles, pedestrians, or other vehicles. Furthermore, drivers tend to be less focused and more stressed at night, which can lead to fatigue and impaired concentration, ultimately leading to an increased risk of accidents [24]. The KNPA records information related to time factors, such as the date, time, day of the week, and whether the accident occurred during the day or at night.

Thirdly, the size and type of vehicles involved in accidents are also related to accident risk [25]. Collisions between large and small vehicles often result in more significant damage to the smaller vehicle. Additionally, technical failures such as faulty brakes, steering systems, tires, lights, seat belts, and other vehicle-related issues can lead to accidents. The KNPA records information related to the types of vehicles involved in accidents, including construction equipment, pedestrians, passenger cars, buses, bicycles, motorcycles, and others.

Finally, human factors related to the driver include physical, cognitive, and psychological factors. Physical factors include the driver's health status, sleep deprivation, physical discomfort, and illness [26]. Cognitive factors include the driver's attention, concentration, experience, and judgment. Psychological factors include stress, anxiety, depression, anger, etc. These factors interact in complex ways, ultimately increasing the risk of accidents. The KNPA records information related to human factors, including attention status, environmental factors, psychological factors, driving behavior, vehicle operation errors, and mental and physical health status.

3. Proposed Method

In this paper, we propose a method to estimate the driving direction of traffic accident vehicles. To estimate the driving direction, basic information such as the location and destination of the accident site is required. Currently, there are no separate fields related to origin and destination in the traffic accident data collected by the KNPA, except for the accident location information (address, latitude, longitude). Therefore, we utilize NER by extracting the destination from the text content of the *"Report"* field written by the investigator. We also use geocoding to convert the destination in text format to

latitude and longitude coordinates. Finally, we calculate the angle between the incident location coordinates and the goal coordinates to estimate whether the person was traveling Northbound or Southbound. The overall process of the proposed method is shown in Figure 1.
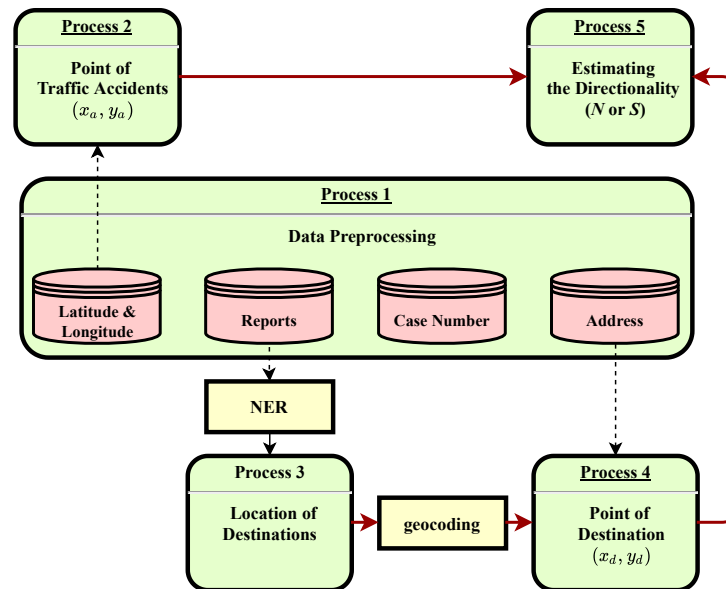

Figure 1. Overall procedure of proposed method.

3.1. Problem Definition

All roads have both Northbound and Southbound lanes, and depending on the environment of the road, traffic accidents can occur more frequently in specific directions. In Figure 2, the cars in the Southbound direction can drive without any problem. Conversely, in the Northbound direction, unpaved roads can serve as potential factors leading to traffic accidents. On paved roads, drivers rely on their line of sight to confirm the presence of vehicles on unpaved roads and make a right turn. Similarly, on unpaved roads, drivers must rely solely on their line of sight to merge onto paved roads. If obstacles obstruct the line of sight, such as trees, seeing other vehicles on different roads can be challenging.
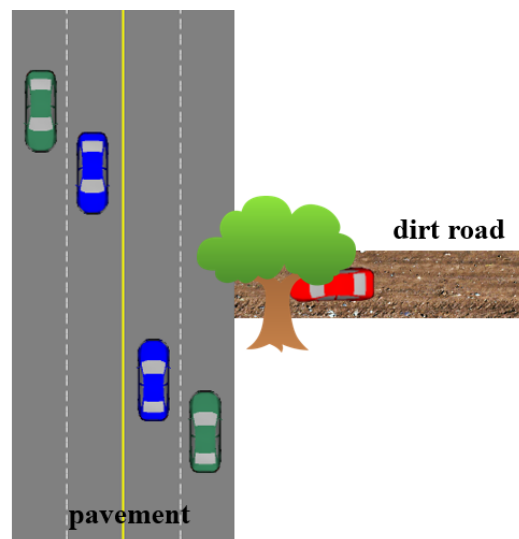

Figure 2. Accident risk case.

To prevent traffic accidents, departments in charge of road safety, such as the police, need to analyze which direction traffic accidents are more likely to occur based on the

environmental conditions in the Northbound and Southbound directions. Some countries, such as the North Carolina Department of Transportation's Division of Motor Vehicles in the United States, record the directionality of traffic accidents based on compass direction (east–west-south–north) [27]. On the other hand, some countries, including Korea, do not collect such directional information; they only collect data such as the date, location, environmental conditions, description, type, and cause of the occurrence in order to collect statistics necessary for developing traffic safety measures and assessing the risk of traffic accidents. In these countries, as there is no policy requirement to collect directional data, there is a need to estimate the directionality from the available information in other fields being collected.

3.2. Data Preprocessing

Since investigators compile traffic accident data, it can contain duplicate entries, missing data, and erroneous information. In addition, even for a single accident, multiple entries may be recorded depending on the specific circumstances of the accident. To address these issues, duplicate data with the same *"Case Number"* is removed, and missing data due to omitted *"Report"* information is excluded from the analysis. For example, suppose the *"Report"* field is missing. In that case, information about the destination cannot be obtained and, therefore, must be excluded from the analysis.

3.3. Query for Information on the Location of Vehicle Traffic Accidents

The location of a traffic accident can be extracted from the *"Address"*, *"Latitude"*, and *"Longitude"* fields. The *"Address"* field contains textual address information about the actual location of the traffic accident. The *"Latitude"* and *"Longitude"* fields include direct coordinates information (x_{acc}, y_{acc}) about the location of the traffic accident. Geocoding is required to convert the address information in the *"Address"* field into coordinates. Geocoding is a crucial part of text geoparsing, which involves several tasks. Specifically, it refers to converting the textual value of identified places into latitude and longitude coordinates on a map [28]. The following steps are performed to obtain coordinate values using geocoding: First, a search is made of various address databases of cities and buildings to find the most similar address. Then, the geocoding engine converts that address to latitude and longitude coordinate values. Therefore, in this paper, the traffic accident location (x_{acc}, y_{acc}) was defined by directly using the *"Latitude"* and *"Longitude"* fields, which can be used directly without the need for a separate conversion process.

3.4. Rule-Based Named Entity Recognition of Arrival Destinations

In a traffic accident, a police officer arrives at the scene with coordinates (x_{acc}, y_{acc}) and records information related to the accident. Detailed records of the accident are recorded in the *"Report"* field. The following are actual *"Report"* cases recorded by a police officer:

- (Case 1) Car #1 was traveling on a one-way road from the direction of the Yeokmal Intersection toward Bukil on the first lane of the road. At the accident site, it collided with the rear of Car #2, which was traveling ahead of it, with the right front of Car #1. As a result, car #1 caused a traffic accident resulting in personal and property damage and then fled the scene without taking any action.
- (Case 2) Car #1 was traveling from Gongju toward Seocheon on the second lane of the two-lane road when it collided with the left side of Car #2, which was parked on the second lane of the road due to a single preceding accident, with the front of Car #1. The accident caused the driver of Car #2, who was sitting in the driver's seat outside the car, to be hit.
- (Case 3) Car #1 was traveling on the first lane (excluding the left turn lane) from Cheonan toward Gongju, and when it collided with the rear of Car #2, which was traveling ahead of it, Car #2's driver fell onto the first lane, causing Car #3 to collide with Car #2's driver.

- (Case 4) Car #1 was traveling on the third lane of the three-lane one-way road toward Seoul direction when it collided with the rear of Car #2, which was traveling ahead of it, and then came to a stop on the third lane and the shoulder of the road at around 1 o'clock (first accident). Car #3, traveling on the third lane toward the left side of Car #1 and its driver, collided with Car #1, who had exited the car and proceeded on foot.
- (Case 5) Car #1 collided with Pedestrian #2, who was crossing the crosswalk toward the direction of K-Mart from the direction of Shin-gwan Elementary School, with its right front side while traveling toward the direction of OO Bank from the direction of OO Mart.
- (Case 6) Car #1 collided with Car #2, coming from the opposite direction and traveling straight through the signal while turning left.

In the given cases, the *"Report"* includes keywords such as 'banghyang (direction)', 'bangmyeon (direction)', and 'jjog (side)', which allow us to estimate the accident vehicle's path. Using these keywords, we can identify the information about the starting and ending points of the accident vehicle in the previous sentence (case 1 to case 3). The starting and ending points are expressed in various forms: such as location, business, building, and road names.

However, there are cases where the information about the starting or ending point is missing. In case 4, for example, information about the starting point is omitted. However, the information about the ending point is included. Second, there are cases where the location of the starting or ending point could be more explicit. In case 5, for example, it is difficult to estimate the exact destination because the country has identical place names, such as "OO Mart". Therefore, it is necessary to estimate the destination based on the nearest location name to the accident site. The last extreme case is when important information is completely missing due to investigator negligence. In case 6, all information related to the path is omitted. In addition, cases where investigators write very brief reports or do not write reports at all, should be excluded from the analysis.

In this study, we use Rule-based Named Entity Recognition to estimate the destination. As seen in the case analysis, the destination is located between the 'banghyang/bangmyeon/jjog' (noun) + 'eseo' (postposition) and the 'banghyang/bangmyeon/jjog' (noun) + 'eulo' (postposition) associated with the destination. Algorithm 1 shows the method for extracting the destination from the *"Report"*. First, we search for candidate phrases between the text-based list BD related to the starting point and the text-based list ED related to the destination after removing various symbols from the report. Finally, we set the shortest phrase among all candidates as the final destination.

3.5. Extracting Accident Vehicle Destinations Using Geocoding

This paper estimates the destination coordinates (x_{dst}, y_{dst}) using the text-based destination information extracted from NER. This paper estimates the destination point through two stages of geocoding processes.

In the first step, if the extracted destination from NER represents a specific place name such as 'Busan' or 'Seoul', the geocoding engine directly outputs the corresponding coordinates (x_{dst}, y_{dst}). However, there are multiple locations with the same name, such as apartments or stores. In that case, obtaining specific coordinates from the geocoding engine becomes impossible. In such cases, it is reasonable to assume that the destination is located in the nearest area to the accident location.

In the second step, geocoding is performed in a restricted area by providing additional information about the destination. The reverse-geocoding engine provides text address information for the accident location coordinates (x_{acc}, y_{acc}). By adding the "do" and "si" information about the area to the destination and inputting it into the geocoding engine, reliable location information can be obtained. For example, assuming a traffic accident in "358, Sagiso-dong, Dangjin-si, Chungcheongnam-do", if the extracted destination from the NER results in "Himart", the geocoding engine cannot return any results. Therefore, by reverse-geocoding the accident information (x_{acc}, y_{acc}), "Chungcheongnam-do" and "Dangjin-si"

can be additionally obtained. A new query, "Himart, Dangjin-si, Chungcheongnam-do" is created. If multiple candidate destination points are returned, the geocoding engine sets the closest point to (x_{acc}, y_{acc}) as the final (x_{dst}, y_{dst}) using the second geocoding. Figure 3 illustrates the entire geocoding process.

Algorithm 1 Rule-based Named Entity Recognition.

1: **procedure** NER(*report*)
2:     $BD \leftarrow$ ["banghyang-eseo", "bangmyeon-eseo", "jjog-eseo", "eseo"]
3:     $ED \leftarrow$ ["banghyang-eulo", "bangmyeon-eulo", "jjog-eulo", "eulo"]
4:     $entities \leftarrow \varnothing$
5:     **for** $i \leftarrow 1$ to len(*report*) **do**
6:         **for** $j \leftarrow 1$ to len(BD) **do**
7:             **if** $report[i - \text{len}(BD[j]) + 1 : i] = BD[j]$ **then**
8:                 **for** $k \leftarrow i + 2$ to len(*report*) **do**
9:                     **for** $l \leftarrow 1$ to len(ED) **do**
10:                         **if** $report[k : k + \text{len}(ED[l]) - 1] = ED[l]$ **then**
11:                             $entities \leftarrow entities \cup \{(report[i + 1 : k - 1])\}$
12:                         **end if**
13:                     **end for**
14:                 **end for**
15:             **end if**
16:         **end for**
17:     **end for**
18:     $destination \leftarrow entities[0]$
19:     **for** $i \leftarrow 1$ **to** len(*entities*) **do**
20:         **if** len(*entities*[*i*]) < len(*destination*) **then**
21:             $destination \leftarrow entities[i]$
22:         **end if**
23:     **end for**
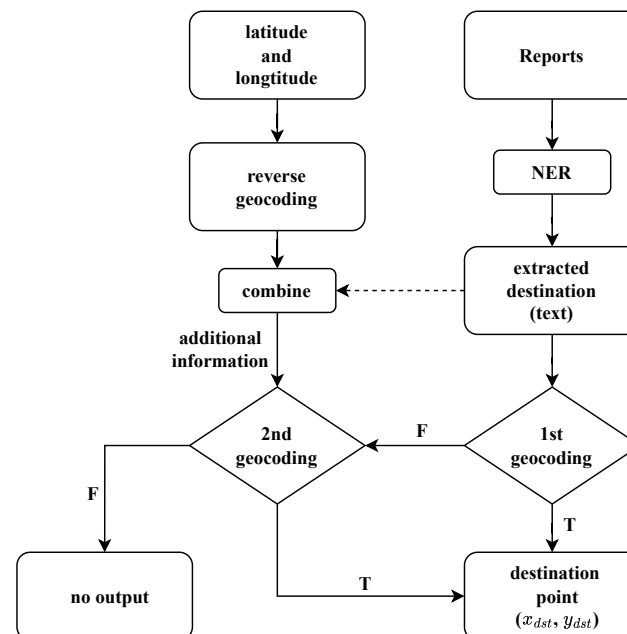24:   **return** *destination*
25: **end procedure**

Figure 3. Geocoding flowchart.

3.6. Driving Direction Estimation of Vehicles Involved in Traffic Accidents

Finally, we calculate the angle between (x_{acc}, y_{acc}) and (x_{dst}, y_{dst}) to estimate the driving direction that includes the Northbound and Southbound directions.

The arctan obtains the angle using the slope values of $||x||, ||y||$ coordinates. However, when calculated by the slope (x_{dst}, y_{dst}), the relative position of the destination location that changes according to the sign of the coordinate value is not reflected. As a result, the relative position $(- \text{ or } +)$ of (x_{dst}, y_{dst}) in (x_{acc}, y_{acc}) is not known precisely, so it is impossible to distinguish between Northbound and Southbound. Therefore, we use the *arctan*2 function, the inverse function of *tan*, and provide the angle in the range of $-\pi$ to π to accurately determine which quadrant the angle belongs to [29]. The angle θ between the two points is defined as follows:

$$\theta = \frac{arctan2(||x||, ||y||)}{\pi} \times 180 \tag{1}$$

In Equation (1), the value obtained from *arctan*2 is in radians, so we multiply it by $180/\pi$ to convert it to degrees. The distances between the latitudes and longitudes of the two points, $||x||$ and $||y||$, are calculated as follows:

$$||x|| = x_{dst} - x_{acc}$$
$$||y|| = y_{dst} - y_{acc} \tag{2}$$

Finally, the final driving direction is determined as follows:

$$direction = \begin{cases} N, & 0 \leq \theta < 180 \\ S, & -180 \leq \theta < 0 \end{cases} \tag{3}$$

In Equation (3), if θ is between 0 and 180, it is Northbound; if it is between -180 and 0, it is Southbound. Figure 4 illustrates how the driving direction is determined based on the angle between (x_{acc}, y_{acc}) and (x_{dst}, y_{dst}) in the first quadrant. The driving direction is Northbound for (x_{acc}, y_{acc}) and (x_{dst}^1, y_{dst}^1) in the first quadrant because their relative position is $(+, +)$ and for (x_{dst}^1, y_{dst}^1) in the second quadrant because its relative position is $(-, +)$. On the other hand, the driving direction is Southbound for (x_{dst}^2, y_{dst}^2) in the fourth quadrant because its relative position to (x_{acc}, y_{acc}) is $(+, -)$. Finally, for (x_{dst}^2, y_{dst}^2) in the third quadrant, the relative position is $(-, -)$, so the driving direction is Southbound.
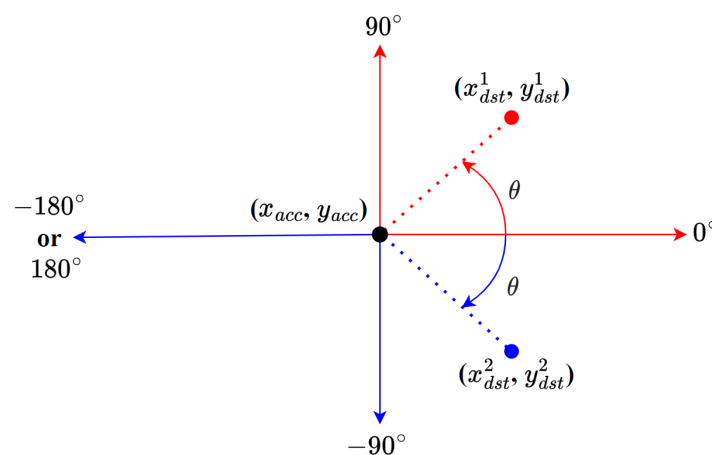

Figure 4. The distinction between Northbound and Southbound according to the degree.

4. Experimental Results

4.1. Experimental Setup

In this section, we detail the experimental setup utilized in our research. Our hardware setup comprised an AMD Ryzen 9 5950X 16-Core CPU, 32GB DDR4 RAM, and an NVIDIA

GeForce RTX 4080 GPU. The software used in our study included Windows 11 Pro as the operating system and Python 3.7.6. For data preprocessing, we utilized pandas version 1.3.5 DataFrames and numpy version 1.12.6 for handling missing data. Furthermore, we utilized Google Maps Application Programming Interface (API) and Kakao Maps API for geocoding. In addition, we utilized folium version 0.12.1.post1 library for Python to visualize maps. Additionally, we employed the Altair API version 4.2.0 to display a pop-up histogram of the frequency of directions on the map.

4.2. Dataset

In this paper, we conducted a study to estimate the driving direction of traffic accident vehicles in Chungcheongnam-do. Despite various prevention activities such as traffic control, publicity campaigns, and facility improvement conducted by the police and related transportation safety agencies to reduce traffic fatalities, Chungcheongnam-do is one of the areas with the highest traffic accidents in South Korea [30]. Chungcheongnam-do is composed of a total of 15 "si" and "gun" areas. Table 2 shows the statistics of traffic accidents reported to the police as of 2021. When comparing the number of traffic accidents, Cheonan-si and Asan-si had the highest traffic accidents. In addition, Cheonan-si and Asan-si showed the highest Traffic Accident Rate per unit area (TAR). TAR is metric that represents the number of traffic accidents per unit area, is typically expressed as "the number of traffic accidents/dimension" and is usually measured as the number of traffic accidents per 1 km^2. Despite the narrow dimension and a low number of reports, Gyeryong-si had a high TAR, indicating that it cannot be considered a traffic accident safety area.

Table 2. Statistics of traffic accidents by AD (city/gun) in Chungcheongnam-do received by the police as of 2021.

Administrative District	Dimension (km^2)	Traffic Accidents	
		Number	TAR
Cheonan-si	636.14	3039	4.7773
Gongju-si	864.19	461	0.5334
Boryeong-si	573.78	424	0.7390
Asan-si	542.19	1185	2.1856
Seosan-si	741.29	667	0.8998
Nonsan-si	554.75	350	0.6309
Gyeryong-si	60.72	56	0.9223
Dangjin-si	704.26	690	0.9798
Geumsan-gun	577.13	111	0.1923
Buyeo-gun	624.52	193	0.3090
Seocheon-gun	365.70	256	0.7000
Cheongyang-gun	479.10	151	0.3152
Hongseong-gun	443.99	244	0.5496
Yesan-gun	542.62	326	0.6008
Taean-gun	515.79	270	0.5235
Chungcheongnam-do	8226.17	8423	1.0239

This paper collected 5181 traffic accident data items in Chungcheongnam-do from 2016 to 2018. First, we removed 418 duplicate, missing, and outlier data items in the data preprocessing stage. Second, in the NER process, we removed 802 data items in which the destination was not output. Finally, we excluded 368 data items that were not returned as destination point coordinates through geocoding from the experiment. As a result, we performed the analysis on a total of 3593 data items.

4.3. Visualization

This section discusses the visualization of AD-specific directional information on the map at the villages/towns/neighborhoods level. First, we define each AD's center

coordinates $(\bar{x}_{AD}, \bar{y}_{AD})$ for visualization. Then, we created a lookup table containing the center coordinates of 207 ADs at the village/town/neighborhood level using Kakao Maps API's reverse geocoding. Second, we matched all accident locations (x_{acc}, y_{acc}) to their corresponding text addresses using geocoding and then matched them to the lookup table. During this process, we accumulated the frequency of Northbound and Southbound directions for each $(\bar{x}_{AD}, \bar{y}_{AD})$ (Figure 5).
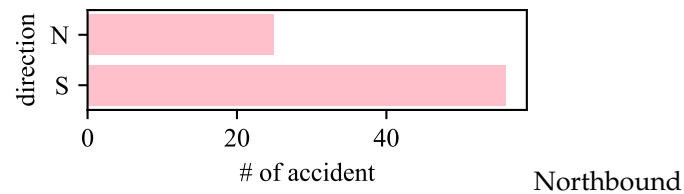
Figure 5. Example of histogram on map.

Figure 6 shows the results of mapping the frequency of Northbound and Southbound directions for each $(\bar{x}_{AD}, \bar{y}_{AD})$ on the map. In the visualization, the direction with a higher frequency is displayed. The green upward arrow, red downward arrow, and black bidirectional arrow represent Southbound, Northbound, and no-priority directions, respectively. In addition, a tooltip is used to indicate the cursor's position explicitly.
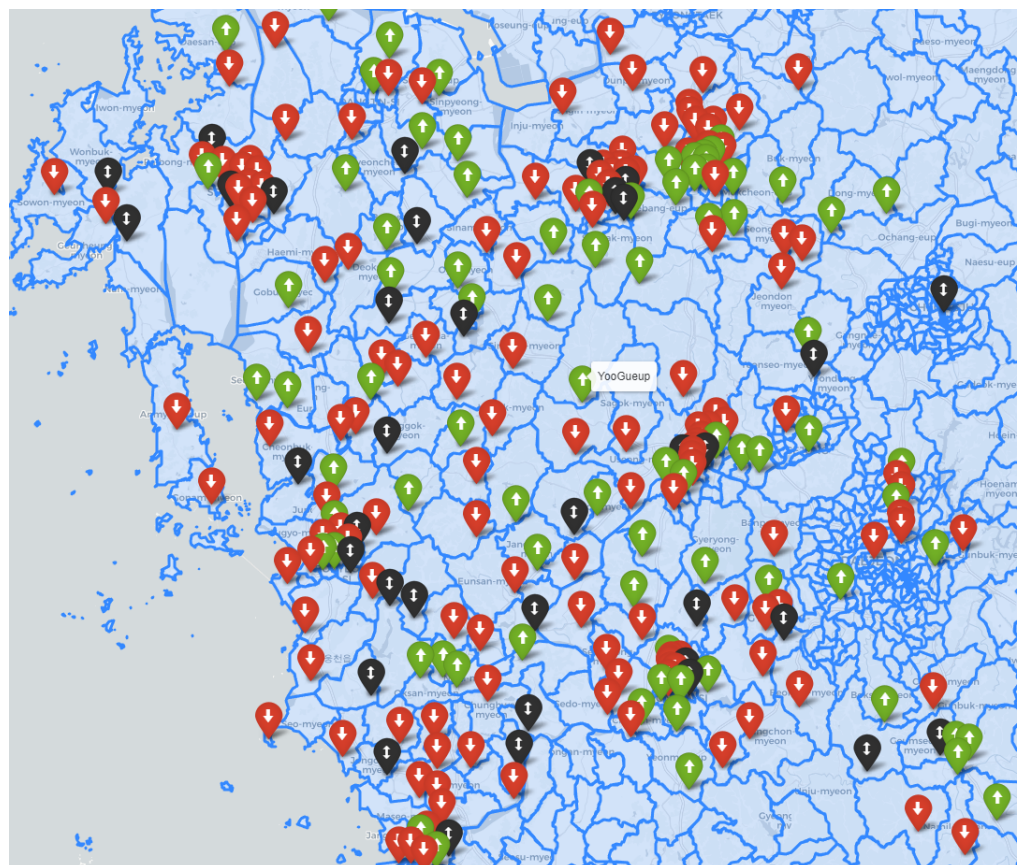

Figure 6. The estimated representative accident directions by dong/eup/myeon based on frequency within the Chungcheongnam-do (green upward arrow: priority given to Northbound direction, red downward arrow: priority given to Southbound direction, black bidirectional arrow: no priority direction).

Table 3 summarizes the number of traffic accidents by city/county in Chungcheongnam-do as of 2021. Since villages/towns/neighborhoods have a minimum unit size and contain much information, we aggregated the data by city/county. Southbound traffic accidents represented 57.25% of the total cases, about 14.5% higher than Northbound. Northbound

was higher only in 'Cheonan-si' and 'Geumsan-gun', with 6.02% and 0.82% higher Northbound ratios, respectively. The remaining 13 ADs all had a higher proportion of Southbound traffic accidents. In order of decreasing Southbound ratio, they were 'Yesan-gun', 'Cheongyang-gun', 'Seosan-si', 'Hongseong-gun', 'Seocheon-gun', 'Nonsan-si', 'Gongju-si', 'Boryeong-si', 'Buyeo-gun', 'Gyeryong-si', 'Asan-si', 'Dangjin-si', and 'Taean-gun', with a difference of about 9.84% to 71.42% higher Southbound ratios. However, the AD with the highest difference in Northbound and Southbound ratios, 'Taean-gun', had minimal traffic accidents and thus low reliability. Excluding this AD, the difference in Northbound and Southbound ratios is approximately 9.84% and 26.66%, respectively. Furthermore, 'Taean-gun', 'Cheonan-si', and 'Dangjin-si' have a significant difference between the number of traffic accidents in Table 2 and the number of traffic accidents in Table 3, indicating that much data were lost due to inaccurate driving direction information in the incident reports.

Table 3. Accident statistics by driving direction aggregated from the proposed method.

Administrative District	Northbound		Southbound		Total
	Number	Percentile	Number	Percentile	
Cheonan-si	176	53.01%	156	46.99%	332
Gongju-si	196	41.26%	279	58.74%	475
Boryeong-si	108	40.75%	157	59.25%	265
Asan-si	132	36.97%	225	63.03%	357
Seosan-si	110	42.31%	150	57.69%	260
Nonsan-si	179	41.53%	252	58.47%	431
Gyeryong-si	23	39.66%	35	60.34%	58
Dangjin-si	11	36.67%	19	63.33%	30
Geumsan-gun	61	50.41%	60	49.59%	121
Buyeo-gun	80	40.20%	119	59.80%	199
Seocheon-gun	106	42.23%	145	57.77%	251
Cheongyang-gun	61	42.96%	81	57.04%	142
Hongseong-gun	118	42.29%	161	57.71%	279
Yesan-gun	174	45.08%	212	54.92%	386
Taean-gun	1	14.29%	6	85.71%	7
Chungcheongnam-do	1536	42.75%	2057	57.25%	3593

Therefore, except for some areas in Chungcheongnam-do, we can see that the proportion of traffic accidents in a Southbound direction is higher than that in a Northbound direction. However, there is a limitation in estimating the direction of traffic accidents due to the inaccurate direction information in the incident reports regarding the destination direction of the vehicles involved.

5. Discussion and Conclusions

When formulating traffic safety policies, various statistical data on traffic accidents can be considered, including road conditions, weather conditions, traffic volume, and vehicle types. The directional aspect of traffic accidents can also be critical in formulating traffic policies. Analyzing high-risk roads by direction for each local jurisdiction can prevent the waste of budget and law enforcement resources, allowing proactive and direct use in traffic planning and improving transportation facilities. It can also influence the formulation of traffic safety policies. For example, improvements can be made to driving conditions that reduce driver visibility, such as curvy roads or hills. It can be used to proactively remove obstacles, such as trees or facilities that pose a threat to traffic safety.

This paper analyzes the likelihood of traffic accidents occurring based on the environment of the Northbound and Southbound directions. It proposes a method for estimating the driving direction of traffic accident vehicles using accident and destination information. We directly extract the latitude and longitude fields from the traffic accident data to extract the accident location. To extract the destination, we first use Named Entity Recognition to remove destination information from the *"Report"* field written by the investigator. We then

use geocoding technology to convert the destination to latitude and longitude coordinates. Finally, we estimate the driving direction of the Northbound and Southbound directions by calculating the angle between the accident location and destination coordinates. In the experimental results, we analyzed 5181 traffic accident data from Chungcheongnam-do from 2016 to 2018. We examined the directionality of traffic accidents by AD. After the preprocessing stage, we analyzed 3593 data. We visualized the directionality information of the ADs at the village/town/district level on a map. The analysis revealed that Southbound traffic accidents accounted for 57.25%, higher than Northbound traffic accidents. In some areas, Southbound traffic accidents were as high as 71.42%. However, there were limitations, such as the need for more accurate direction information in some areas, making it difficult to estimate the directionality of traffic accidents.

The method proposed in this study can be effectively utilized for traffic safety enhancement and accident analysis, even in countries where traffic accident guidance records are not mandatory. The ideal solution is to establish a policy that obliges the recording of traffic accident directions. However, the proposed method is one of the alternatives available when these policies are not feasible. The proposed method can also be used to generate collision diagrams for traffic accident analysis. The crash diagram is an essential aspect of the crash report as the investigator can show the relationship that existed during the crash between the vehicle and the surrounding environment. According to DMV-349, written by the North Carolina Department of Transportation's Department of Motor Vehicles, it is recommended to include the "direction of travel for each traffic lane" when drawing crash scene diagrams. Therefore, the proposed method can help create a collision diagram by effectively estimating directions by referring to accident points and destination points, even in scenarios where direction information is not documented.

On the other hand, the proposed method has some limitations. First, the utilization of estimated directional information raises concerns about reliability in terms of accuracy. Specifically, the proposed method estimates the destination information from the Report field. It selects the final destination as the closest location to the accident site in cases where a precise location is not found. As a result, the propagation of errors in multiple estimation stages compromises the overall reliability. Additionally, unlike the recording method used by the North Carolina Department of Transportation Division of Motor Vehicles, which includes four directions encompassing north, south, east, and west, the proposed method only considers the Northbound and Southbound directions, resulting in a limitation of limited directional information for traffic accidents. In addition, the proposed method generally overlooked various parameters related to traffic accident risk because it is intended only to estimate the direction.

From this perspective, future work will involve researching estimating directionalities in terms of angle units in addition to the compass directions. To accomplish this, a road navigation API can be utilized to generate a repository of road journeys, enabling the determination of the precise direction of movement from the accident site. Furthermore, there is a need for research on traffic safety from various perspectives, exploring risk levels based on multiple factors by integrating various parameters associated with traffic accidents. For instance, factors such as the speed of the vehicles and time of the day (off-peak and on-peak), and also traffic volumes during peak periods, can be incorporated in this regard. In addition, recent developments in "autonomous driving scenarios" are being researched based on existing traffic accident overview data [31]. Therefore, the method defined in this paper, which uses rule-based NER, is expected to help develop new autonomous driving scenarios.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not available.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AD	Administrative District
API	Application Programming Interface
KNPA	Korean National Police Agency
NER	Named Entity Recognition
TAR	Traffic Accident Rate per unit area

References

1. Gaikwad, D.B.; Wanjari, Y.W.; Kale, K.V. Disaster management by integration of web services with geospatial data mining. In Proceedings of the 2014 Annual IEEE India Conference (INDICON), Pune, India, 11–13 December 2014; pp. 1–6.
2. Chen, S.; Kuhn, M.; Prettner, K.; Bloom, D.E. The global macroeconomic burden of road injuries: Estimates and projections for 166 countries. *Lancet Planet. Health* **2019**, *3*, e390–e398. [CrossRef] [PubMed]
3. Anjuman, T.; Hasanat-E-Rabbi, S.; Siddiqui, C.K.A.; Hoque, M.M. Road traffic accident: A leading cause of the global burden of public health injuries and fatalities. In Proceedings of the International Conference on Mechanical Engineering 2007 (ICME2007), Dhaka, Bangladesh, 29–31 December 2020; pp. 29–31.
4. Wangdi, C.; Gurung, M.S.; Duba, T.; Wilkinson, E.; Tun, Z.M.; Tripathy, J.P. Burden, pattern and causes of road traffic accidents in Bhutan, 2013–2014: A police record review. *Int. J. Inj. Control. Saf. Promot.* **2018**, *25*, 65–69. [CrossRef]
5. Mohammed, A.A.; Ambak, K.; Mosa, A.M.; Syamsunur, D. A review of traffic accidents and related practices worldwide. *Open Transp. J.* **2019**, *13*, 65–83. [CrossRef]
6. Chantith, C.; Permpoonwiwat, C.K.; Hamaide, B. Measure of productivity loss due to road traffic accidents in Thailand. *IATSS Res.* **2021**, *45*, 131–136. [CrossRef]
7. Ladeira, R.M.; Malta, D.C.; Morais, O.L.d.; Montenegro, M.d.M.S.; Soares, A.M.; Vasconcelos, C.H.; Mooney, M.; Naghavi, M. Road traffic accidents: Global Burden of Disease study, Brazil and federated units, 1990 and 2015. *Rev. Bras. Epidemiol.* **2017**, *20*, 157–170. [CrossRef]
8. Ang, B.H.; Chen, W.S.; Lee, S.W.H. Global burden of road traffic accidents in older adults: A systematic review and meta-regression analysis. *Arch. Gerontol. Geriatr.* **2017**, *72*, 32–38. [CrossRef]
9. Tsujimura-Ito, T. State of damage to and support for victims of motor vehicle accidents in Japan. *IATSS Res.* **2019**, *43*, 97–107. [CrossRef]
10. Blincoe, L.; Miller, T.R.; Wang, J.S.; Swedler, D.; Coughlin, T.; Lawrence, B.; Guo, F.; Klauer, S.; Dingus, T. *The Economic and Societal Impact of Motor Vehicle Crashes, 2019*; Technical report; National Center for Statistics and Analysis: Washingtod, DC, USA, 2022.
11. Bachani, A.; Peden, M.; Gururaj, G.; Norton, R.; Hyder, A.A. Road traffic injuries. In *Injury Prevention and Environmental Health*; The World Bank: Washington, DC, USA, 2017.
12. Waseem, M.; Ahmed, A.; Saeed, T.U. Factors affecting motorcyclists' injury severities: An empirical assessment using random parameters logit model with heterogeneity in means and variances. *Accid. Anal. Prev.* **2019**, *123*, 12–19. [CrossRef]
13. Touahmia, M. Identification of risk factors influencing road traffic accidents. *Eng. Technol. Appl. Sci. Res.* **2018**, *8*, 2417–2421. [CrossRef]
14. Kamarudin, M.K.A.; Abd Wahab, N.; Umar, R.; Saudi, A.S.M.; Saad, M.H.M.; Rosdi, N.R.N.; Razak, S.A.A.; Merzuki, M.M.; Abdullah, A.S.; Amirah, S.; et al. Road traffic accident in Malaysia: Trends, selected underlying, determinants and status intervention. *Int. J. Eng. Technol* **2018**, *7*, 112. [CrossRef]
15. Febres, J.D.; Mohamadi, F.; Mariscal, M.A.; Herrera, S.; García-Herrero, S. The role of journey purpose in road traffic injuries: A Bayesian network approach. *J. Adv. Transp.* **2019**, *2019*, 6031482. [CrossRef]
16. Liu, J.; Li, J.; Wang, K.; Zhao, J.; Cong, H.; He, P. Exploring factors affecting the severity of night-time vehicle accidents under low illumination conditions. *Adv. Mech. Eng.* **2019**, *11*, 1687814019840940. [CrossRef]
17. Wu, G.F.; Liu, F.J.; Dong, G.L. Analysis of the Influencing Factors of Road Environment in Road Traffic Accidents. In Proceedings of the 2020 4th Annual International Conference on Data Science and Business Analytics (ICDSBA), Changsha, China, 5–6 September 2020; pp. 83–85. [CrossRef]

18. Xu, C.; Bao, J.; Wang, C.; Liu, P. Association rule analysis of factors contributing to extraordinarily severe traffic crashes in China. *J. Saf. Res.* **2018**, *67*, 65–75. [CrossRef] [PubMed]

19. Angın, M.; Ali, S.A. Analysis of Factors Affecting Road Traffic Accidents in North Cyprus. *Eng. Technol. Appl. Sci. Res.* **2021**, *11*, 7938–7943. [CrossRef]

20. Zhang, Y.; Lu, H.; Qu, W. Geographical detection of traffic accidents spatial stratified heterogeneity and influence factors. *Int. J. Environ. Res. Public Health* **2020**, *17*, 572. [CrossRef] [PubMed]

21. Megnidio-Tchoukouegno, M.; Adedeji, J.A. Machine Learning for Road Traffic Accident Improvement and Environmental Resource Management in the Transportation Sector. *Sustainability* **2023**, *15*, 2014. [CrossRef]

22. Ashraf, I.; Hur, S.; Shafiq, M.; Park, Y. Catastrophic factors involved in road accidents: Underlying causes and descriptive analysis. *PLoS ONE* **2019**, *14*, e0223473. [CrossRef] [PubMed]

23. Delgado-Fernández, V.J.; Rey-Merchán, M.D.C.; López-Arquillos, A.; Choi, S.D. Occupational traffic accidents among teachers in Spain. *Int. J. Environ. Res. Public Health* **2022**, *19*, 5175. [CrossRef]

24. Madleňák, R.; Hoštáková, D.; Madleňáková, L.; Drozdziel, P.; Török, A. The analysis of the traffic signs visibility during night driving. *Adv. Sci. Technology. Res. J.* **2018**, *12*, 92103. [CrossRef]

25. Zheng, Z.; Lu, P.; Lantz, B. Commercial truck crash injury severity analysis using gradient boosting data mining model. *J. Saf. Res.* **2018**, *65*, 115–124. [CrossRef]

26. Zhang, G.; Yau, K.K.; Zhang, X.; Li, Y. Traffic accidents involving fatigue driving and their extent of casualties. *Accid. Anal. Prev.* **2016**, *87*, 34–42. [CrossRef] [PubMed]

27. North Carolina Department of Transportation Division of Motor Vehicles. *North Carolina Crash Report Instruction Manual*; North Carolina Department of Transportation Division of Motor Vehicles: Raleigh, NC, USA, 2006.

28. Fize, J.; Moncla, L.; Martins, B. Deep learning for toponym resolution: Geocoding based on pairs of toponyms. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 818. [CrossRef]

29. De Dinechin, F.; Istoan, M. Hardware implementations of fixed-point Atan2. In Proceedings of the 2015 IEEE 22nd Symposium on Computer Arithmetic, Lyon, France, 22–24 June 2015; pp. 34–41.

30. Youn, C.; Kang, C. An Analysis of the Causes of Traffic Accidents using TCS: Focused on Chungcheong Namdo. *Korea Assoc. Public Manag.* **2019**, *33*, 271–298.

31. Park, S.; So, J.; Ko, H.; Jeong, H.; Yun, I. Development of Safety Evaluation Scenarios for Autonomous Vehicle Tests Using 5-Layer Format (Case of the Community Road). *J. Korea Inst. Intell. Transp. Syst.* **2019**, *18*, 114–128. [CrossRef]