ANALYSIS METHOD FOR A TRAFFIC ACCIDENT USING MOTORCYCLE PROBE DATA

OSAMU ITO, TAKAYUKI KAWBUCHI, HIDEO KADOWAKI, YUJI TAKAGI1
Honda R&D Co., Ltd.
Japan

HIROKI TANAKA
Honda Motor Co., Ltd.
Japan

Paper Number 23-0136

ABSTRACT

To reduce the number of the fatalities among the motorcyclist in Asian countries, it is necessary to analyze and clarify the cause of the accident, however, the accident data are insufficient in these countries for the accurate analysis. To compensate for insufficient accident data, the authors approached to analyze the accident using the probe data obtained from vehicles.

The investigation was conducted by the riding data acquired from the 50 cc motorcycles, including the location information in 1 second cycle, the vehicle speed and the throttle opening signals in 0.2 seconds cycle acquired from the Global Navigation Satellite System (GNSS) and the Electronic Control Unit (ECU), respectively. The time historical data from GNSS and ECU were divided into 5798 trips, separated by the time interval longer than 1 minute. During all trips, there was only one accident. The acquired data were processed by the autoencoder model to extract the characteristics of the trips and riding behavior. The autoencoder model has the latent space between the encoder and decoder to analyze the trips and riding behavior. The information of trips and riding behavior in the latent space was quantified using Kernel Density Estimation to express the anomaly of the trips and riding behavior. In addition, riding simulations were conducted based on GNSS and ECU information to validate the results of abnormality detection by the autoencoder.

The results showed that the accident data were classified as abnormal behavior. The anomalies could be expressed as changes with time history. It proved that the riding abnormalities appeared 30 seconds before the accident occurred. When the simulation was also performed to reconstruct the accident, it was observed that the rider was riding dangerously such as slipping past the car or accelerating and decelerating rapidly.

The authors devised a method to analyze the causes of traffic accidents by using the autoencoder model and riding simulation. This method is expected to improve the efficiency of accident data collection and analysis in regions where accident data for motorcycles is lacking, such as in developing Asian countries.

INTORDUCTION

According to World Health Organization report [1], accidents involving motorcycles account for 28% of all traffic fatalities worldwide, which is second only to automobiles; countermeasures against traffic accidents involving motorcycles are an important research issue to reduce the number of fatalities. In general, to reduce traffic accidents, it is necessary to analyze the actual conditions of accidents and elucidate the causes of accidents. Since accidents involving motorcycles are more prominent in Asian developing countries such as Thailand [1], it is important to analyze accidents in these countries. Previous studies [2] showed the factors of motorcycle accidents by the investigation of the motorcyclist accident data in Thailand. However, the data contains the following residual issues and that precluded an elaborate analysis. First, the volume of the data was insufficient for the investigation and the accuracy of the analysis was low. Moreover, since various data aggregators, which are the police, hospitals, and insurance companies, investigated individually, it is hard to comprehend the relationships among respective data [3]. Second, the numerous accident investigation was engaged manually, which resulted in the inaccurate data collection

due to the error in the descriptions or lack of information. Therefore, these manual works require the large amount of costs to ensure the data accuracy.

To collect more adequate data, it is necessary to create the efficient data collection process by eliminating the individual works. In recent years, the application of Intelligence Technology Systems has promoted the use of vehicle probe data and enable the effective estimation of road conditions and traffic accident risks [4]. In addition, Matsuo et al. improved the accuracy of collision risk estimation for vulnerable traffic by using probe data [5]. The objective of this study is the proof of the concepts applying the probe data obtained from motorcycles to traffic accidents analysis without any investigation reports by the third-party organizations.

METHODS

Motorcycle probe data
The investigation was conducted by the riding data acquired from the Global Navigation Satellite System (GNSS) instrument and the Electronic Control Unit (ECU) installed on the 50 cc motorcycles, including the location information in 1 second cycle, the vehicle speed and the throttle opening signals in 0.2 seconds cycle, respectively. The time historical data from GNSS and ECU were divided into 5798 trips, which was separated by the time interval longer than 1 minute. During the period of all trips, there was only one accident. A schematic diagram of the accident trip is shown in Figure 1.

Labeling rider behavior
Since riding behaviors cannot be directly observed from the probe data, those were estimated by the locations, the speeds, and the azimuth angles information. Estimated behaviors were defined as going straight, turning right, turning left, stopping, accelerating, decelerating, and cruising based on each state which were listed in Table 1, respectively. The turn direction was defined by the integration of the azimuth angle of travel per unit time within 5-second intervals. The state of acceleration or deceleration was defined based on the comparison of the speed of start and end within 5-second intervals with the average speed. For example, if the start speed was less than average speed and the end speed was greater than average speed, the behavior was defined as acceleration. All of the probe data was separated into 5-second intervals and labeled those riding behaviors according to the definition. In order to represent as various riding behaviors as possible, we defined 10 classes of riding behaviors: "straight + acceleration", "straight + cruise", "straight + deceleration", "right turn + acceleration", "right turn + cruise", "right turn + deceleration", "left turn + acceleration", "left turn + cruise", "left turn + deceleration", "left turn + deceleration", and "stop" by combining [straight, right turn, left turn] with [acceleration, deceleration, cruise] labels. These definitions allow classification of which riding behavior was being performed at a given time in the probe data. This enables analysis of riding behavior until the time of an accident.

Training model of riding history
Probe data contains a vast amount of data on normal riding. In order to analyze accidents, it is necessary to extract only information on the occurrence of accidents. Therefore, a classification model is constructed from the probe data, which can be regarded as the occurrence of an accident.

As the probe data contains a large volume of the data regarding a normal riding behavior without any accidents, it is necessary to extract the part in a short duration related to the traffic accident. The classification model is required to detect the rare incident from the data, however, since there is only one accident data in the probe dataset of this study, it is inappropriate to build the classification model by a supervised learning which generally requires many ground truth data. On the other hand, an anomaly detection model as an unsupervised learning is effective to detect the presence of the error incidents such as the traffic accident by a sparse ground truth data [6]. An anomaly detection model is possible to be trained by various types of data such as complicated images and time historical data [7]. Previous study built the autoencoder model to detect an anomaly taxi route by means of a large amount of vehicle trajectory data [8]. The autoencoder model contains a latent space connecting the input and output variables and the space is observable by visualizing the dimension-reduced vectors. The latent space is the mixture distribution consisting of the mean and variations, therefore, it is possible to determine whether the similarity of the newly obtained data is average or an outliers an anomaly by measuring the distance to a cluster of features in the latent space. This study built the deep anomaly detection model based on an autoencoder to extract the error

incidents from a large amount of historical riding data, which were assumed as that contains the anomaly rider behavior occurring an accident.

Convolutional Neural Networks were used for the encoder and decoder [9]. To ensure that features can be well separated in the latent space, the decoder network was set up to split the trip and riding behavior labels. The trip portion was trained with Mean Square Error loss function, while the riding behavior label portion was trained with Cross Entropy loss function [9]. After training was completed, the riding behaviors were classified into 10-class clusters every 5 seconds. The target riding behavior can be judged as abnormal by measuring the distance from the center of the cluster (Figure 2). To quantitatively measure the distance in the latent space, Kernel Density Estimation (KDE) was used for each cluster [10]. Each cluster's center was defined from the mode of the KDE. The distance from the center was measured in Mahalanobis' distance [11]. For example, riding behaviors of a rider always near the center of the cluster can be considered normal riding, while riding behaviors far from the center of the cluster can be considered abnormal riding. By measuring this distance for riding behavior every 5 seconds, the degree of riding abnormality can be observed in the time history. Figure 3 and Table 2 show the schematic diagram and parameters of the autoencoder model, respectively.

RESULTS

Figure.4 shows the distribution of average Mahalanobis' distance during a trip. The average Mahalanobis' distance during a trip was most often between 0.4 and 0.6. On the other hand, the average Mahalanobis' distance for the accident trip was 1.43. Since the average Mahalanobis' distance was more than 1.4 within 5% of all trips, the anomalies can be classified. Figures 5 and 6 show the latent space and the time history graphs of Mahalanobis' distance for the accident trips. It was found that the Mahalanobis' distance increased about 30 seconds before the timing of the accident. In particular, the distance increased during the actions of accelerating straight, cruising straight, and decelerating straight.

DISCUSSION

In order to analyze how the rider was doing before the timing of the accident, the riding reconstruction simulation was conducted. Motorcyclemaker by IPG was used for the simulation [12]. The vehicle model was simulated only by the exterior shape, and location and time information was input to reconstruct simple riding. The roads were reconstructed by downloading Keyhole Markup Language files of the surrounding roads ridden from Google Maps and inputting them into Motorcyclemaker [13]. The objects such as sidewalks, buildings, traffic signals, and signs were reconstructed by using the 3D city model opened by the Ministry of Land, Infrastructure, Transport and Tourism [14] and applying textures to the objects with reference to Google Street View [15]. Figure 7 shows the picture of the riding trajectory on the road obtained by the simulation. Figure 8 shows the schematic diagram of the travel trajectory. These figures show that the vehicle seems to stop slightly behind the stop line at the intersection and then move forward on the roadway boundary before the intersection. After passing through the signal intersection, the vehicle was traveling at speeds fluctuating between 35 km/h and 40 km/h. Although the speed limiter limits the upper speed limit to about 40 km/h, the vehicle's riding behavior was unnatural, with repeated rapid acceleration and deceleration. Since the accident report noted the presence of a car ahead, we considered the rider to have repeatedly acted in a hurry to keep a short distance from the car ahead. Figure 9 shows the setup with the other vehicles, placed on the reconstruction simulation based on the above assumptions. From the results of the accident reconstruction simulation, the relationship between the Mahalanobis' distance time history and riding behavior is discussed and the results are shown in Figure 10. In addition, the capture of events between the time of arrival before the intersection and the occurrence of the accident is shown in Figure 11. In this accident case, the following three factors are the causes of the accident.

- The rider was slipping past the car at the intersection.

- The rider was accelerating and decelerating rapidly to keep a short distance from the vehicle ahead.

- The rider changed lanes and immediately returned to the original lane.

In this study, the anomaly detection model using the probe data was able to identify the abnormal riding behaviors that led to the accident. Furthermore, by conducting the simulation to reconstruct the accident, we were able to find the insights into the behavior just prior to the accident, which were not recorded in the accident reports. This allowed us to analyze riding behavior about accidents, without the need to conduct on-site investigations. However, the data in this study is limited and the number of accidents is small. We believe that expanding the collection of probe data and validating the methodology of this study will enable reliable analysis of traffic accidents involving motorcycle vehicles in the future.

REFELENCES

[1] World Health Organization (WHO). Global status report on road safety 2018: Summary (No. WHO/NMH/NVI/18.20); 2018

[2] Suriyawongpaisal P. KEY FACTS ON ROAD SAFETY SITUATIONS IN THAILAND 2012-2013; 2015

[3] Japan International Cooperation Agency (JAICA). Project Research "Road Safety Initiatives in Developing Countries", Final Report; 2016. Available at: https://openjicareport.jica.go.jp/pdf/12260915.pdf, Accessed August 2, 2022. (in Japanese)

[4] Najumudeen K, Bin MEA. Traffic accident analysis and prediction using the NPMRDS.; 2020.

[5] Matsuo K, Chigai N, Chattha MI, Sugiki N. Vulnerable road user safety evaluation using probe vehicle data with collision warning information, Accid Anal Prev.2022;165:06528.

[6] Aggarwal CC. An introduction to outlier analysis. In Outlier analysis, Springer. 2013;1-40

[7] Chalapathy R, Chawla S. DEEP LEARNING FOR ANOMALY DETECTION: A SURVEY: 2019. Available at: https://arxiv.org/abs/1901.03407, Accessed August 2, 2022.

[8] Liu Y, Zhao K, Cong G, Bao Z. Online Anomalous Trajectory Detection with Deep Generative Sequence Modeling; 2020 IEEE 36th International Conference on Data Engineering (ICDE),949-960.

[9] Goodfellow I,Bengio Y, Courville A. Deep Learning, MIT Press: 2016, Chapter 6 and 9.

[10] Chen Y. A Tutorial on Kernel Density Estimation and Recent Advances: 2017, Available at: https://arxiv.org/abs/1704.03924, Accessed August 2, 2022

[11] Prasanta CM. On the generalized distance in statistics, Proceedings of the National Institute of Sciences of India.1936: 2 (1).

[12] IGP Automotive, Motorcyclemaker, Available at: https://ipg-automotive.com/, Accessed August 2, 2022.

[13] Google LLC. Keyhole Markup Language, Available at: https://developers.google.com/kml, Accessed August 2, 2022.

[14] Ministry of Land, Infrastructure, Transport and Tourism (MLIT), PLATEAU, Available at: https://www.mlit.go.jp/plateau/, Accessed August 2, 2022.

[15] Google LLC. Google Street View, Available at: https://www.google.co.jp/maps/preview, Accessed August 2, 2022.
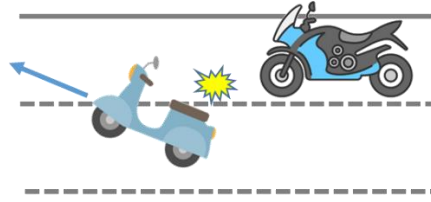
When changing lanes, the motorcycle collided
with the following vehicle.
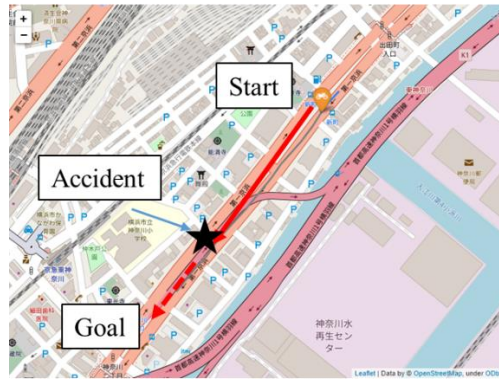
© OpenStreetMap contributors

Figure 1. Schematic diagram of the accident trip

Table 1. Riding behavior label defined from location information and velocity at 5 seconds intervals.

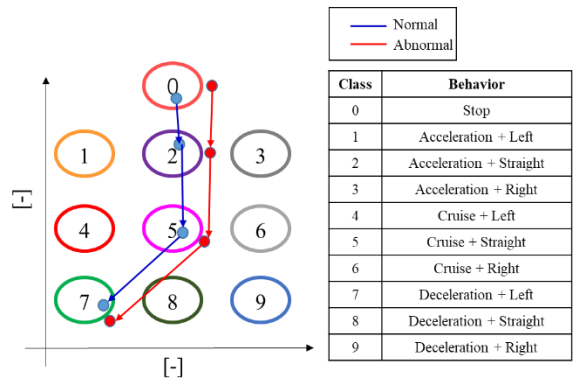| Label	Definition
Go Straight	The integrated value of the azimuth angle is within ± 20 deg.
Turn Left	The integrated value of the azimuth angle is under - 20 deg.
Turn Right	The integrated value of the azimuth angle is over + 20 deg.
Stop	The average velocity is under 5km/h.
Acceleration	Start speed is less than average speed and end speed is greater than average speed.
Deceleration	Start speed is greater than average speed and end speed is less than average speed.
Cruise	Other than acceleration and deceleration conditions.

Class	Behavior
0	Stop
1	Acceleration + Left
2	Acceleration + Straight
3	Acceleration + Right
4	Cruise + Left
5	Cruise + Straight
6	Cruise + Right
7	Deceleration + Left
8	Deceleration + Straight
9	Deceleration + Right

Figure 2. Schematic diagram of analysis method using latent space.
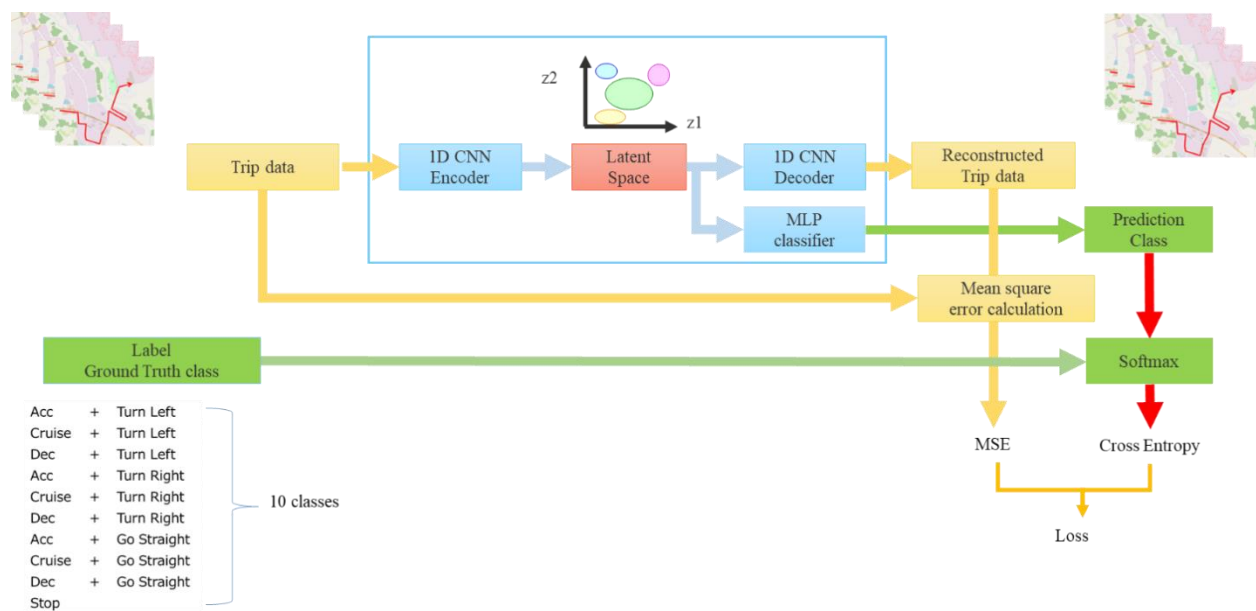
Figure 3. Schematic diagram of the Autoecndoer model

Table 2. Parameters of the Autoecndoer model

Encoder part

Layer	Cin	Lin	Cout	Lout	Kernel	Padding	Stride	Activation Function	Batch Norm
Conv1d	9	500	32	200	5	-	5	Relu	Use
Conv1d	32	200	64	32	5	-	5	Relu	Use
Conv1d	64	32	128	8	5	-	5	Relu	Use
Conv1d	128	8	256	1	4	-	5	Relu	Use
Linear	256	-	2	-	-	-	-	Tanh	Not Use

Decoder part

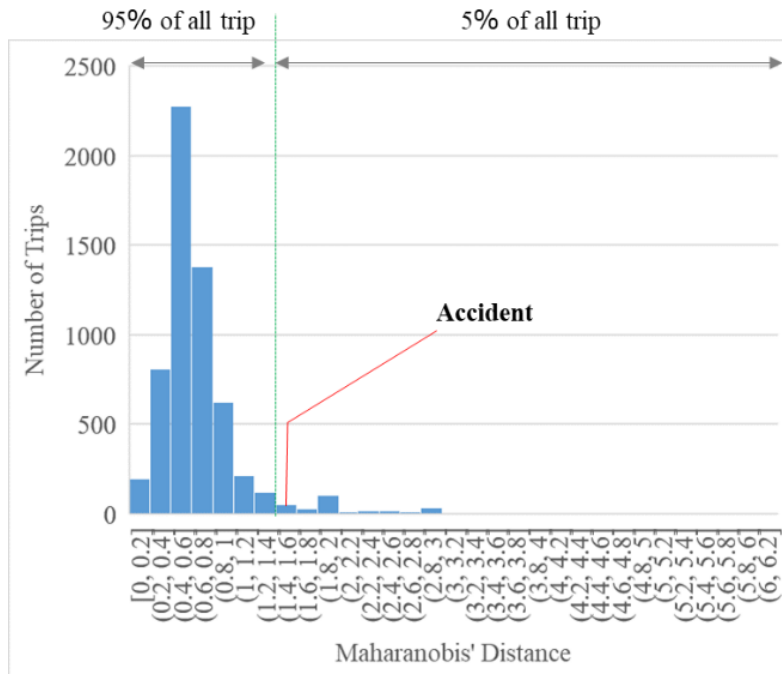| Layer	Cin	Lin	Cout	Lout	Kernel	Padding	Stride	Activation Function	Batch Norm
Linear	2	-	256	-	-	-	-	Leaky ReLU	Use
ConvTranspose1d	256	1	128	5	5	-	5	Leaky ReLU	Use
ConvTranspose1d	128	5	64	25	5	-	5	Leaky ReLU	Use
ConvTranspose1d	64	25	32	125	5	-	5	Leaky ReLU	Use
ConvTranspose1d	32	125	9	500	4	-	4	HardTanh	Not Use

Figure 4. Histogram of average Mahalanobis' distance for all trips.
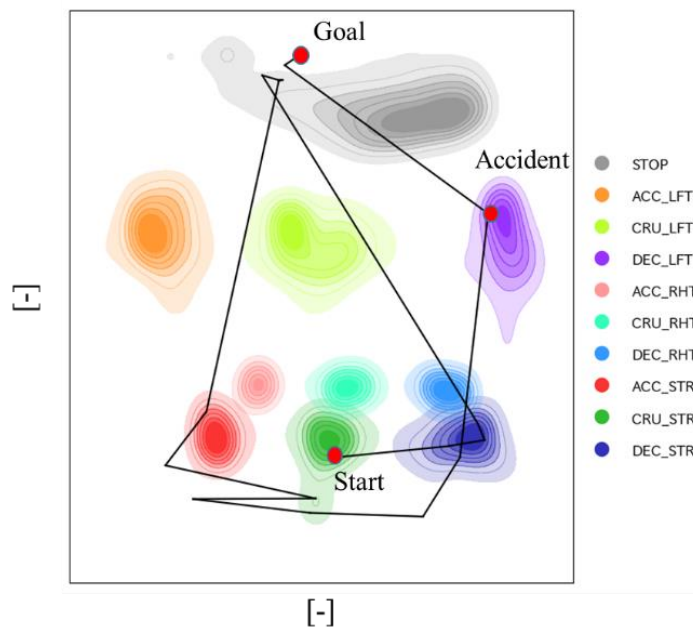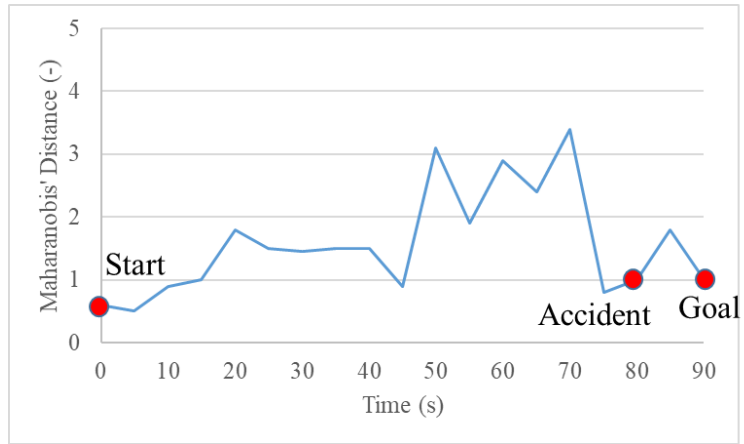

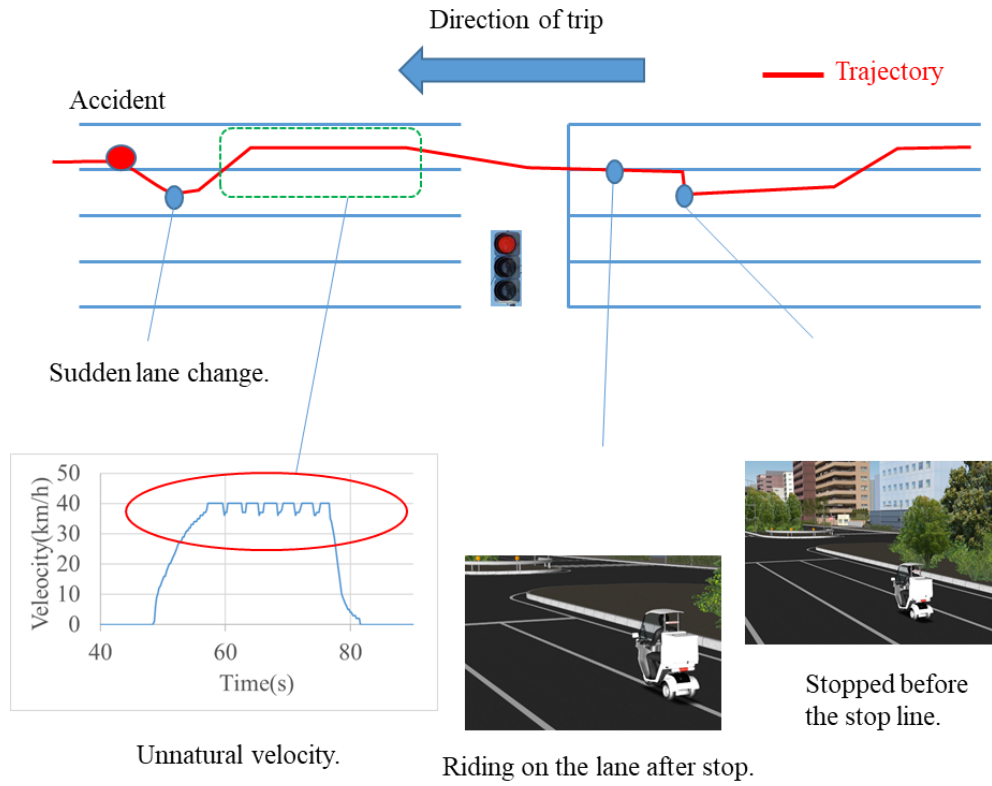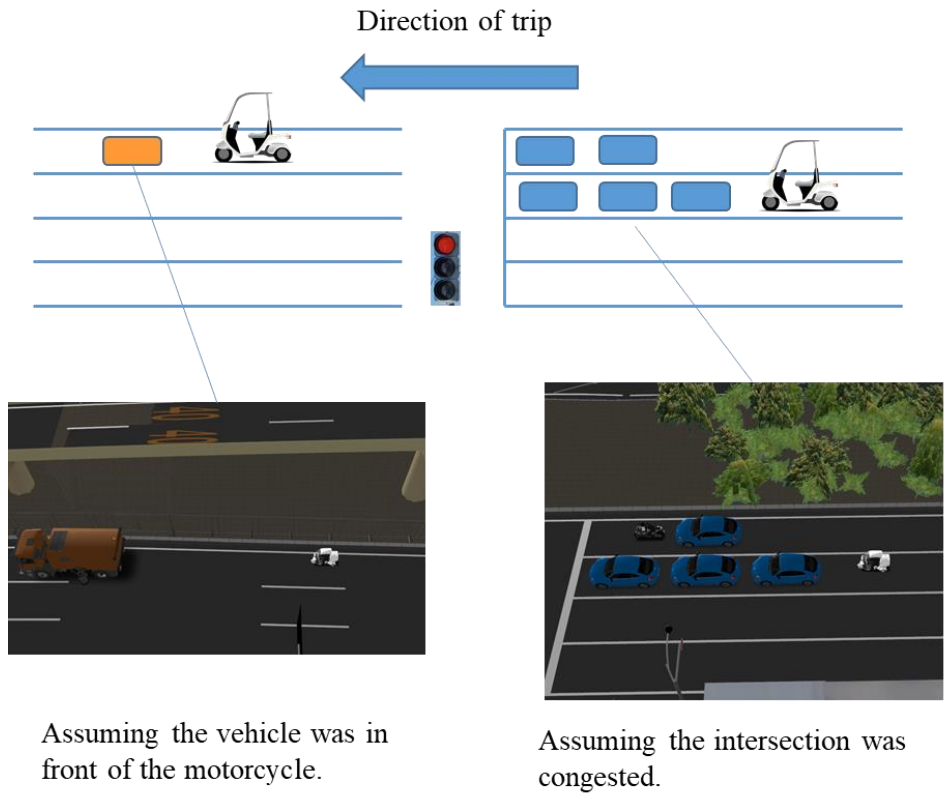**Figure 5. Trajectory of the accident trip for all trips in the latent space.**

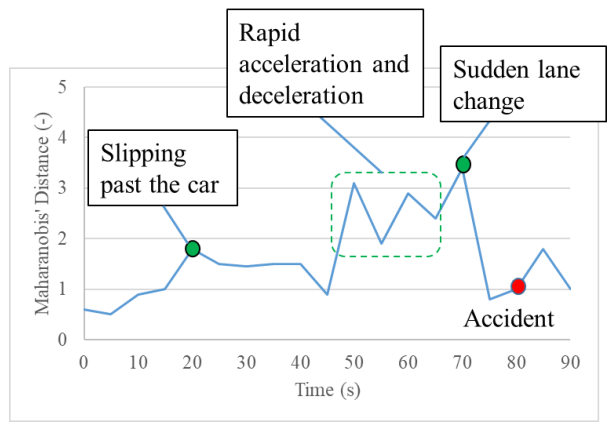**Figure 6. Maharanobis' distance of accident trip.**

Figure 7. Riding trajectory on the road from reconstruction simulation

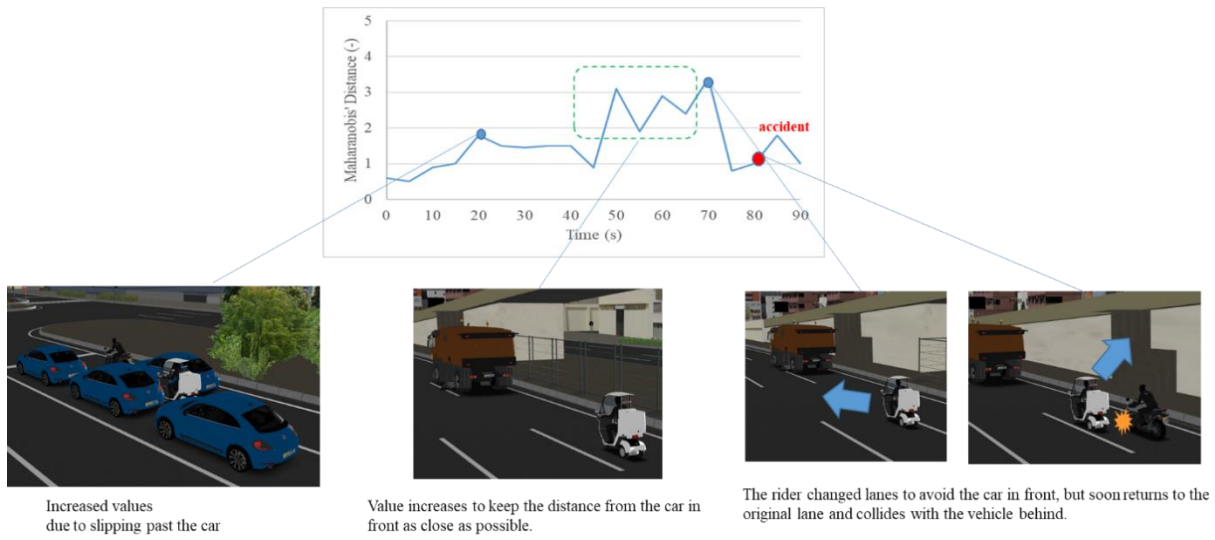**Figure 8. Schematic diagram of the riding trajectory with the features.**

Assuming the vehicle was in front of the motorcycle.

Assuming the intersection was congested.

Figure 9. Assumptions for placement of other vehicles

Figure 10. Relationship between riding events and Mahalanobis' distance.

Figure 11. Relationship between driving behavior and Mahalanobis' distance with simulation results