Construction of Collision Type Prediction Model Based on Pre-crash Data for Advanced Driver Assistance Systems

Wei Junhao [1)] Yusuke Miyazaki [1)] Koji Kitamura [2)] Fusako Sato [3)]

1) Tokyo Institute of Technology, 2-12-1-W8-2, O-okayama, Meguro,Tokyo,152-8552, Japan (E-mail:wei.j.ac@m.titech.ac.jp)
2) National Institute of Advanced Science and Technology.
10-3 Gobancho, Chiyoda, Tokyo, 102-0076, Japan
3) Japan Automobile Research Institute.
2530 Karima, Ibaraki, 305-0822, Japan

KEY WORDS: **Safety, Collision Prediction, Machine Learning** Collision Prediction Fields [C1]

The number of fatalities and injuries in traffic accidents in Japan has been decreasing, however the pace of decline has slowed in recent years, while more safety measures are required than ever before. In previous studies, occupant injuries were estimated based on information such as the types of collision, and the relationship between these injuries was clarified. However, more advance prediction of collision types is highly required, which is supposed to be applied in Autonomous Driving Systems (ADS) to prevent serious accidents from occurring, or to minimize damage of collisions.

Therefore, the aim of this research is to construct a model that predicts the collision types between vehicles that may occur during driving, based on information on the traffic environment, vehicle and driver that might be obtained prior to the accident.

The data source is NASS-CDS and CISS databases from United States Department of Transportation, which conclude detail collision data for over 15 years. To make a pre-crash dataset, pre-crash features are extracted from the database; furthermore, those features are able to be divided into physical environment related, vehicle related and driver related data. As for physical environment related features, lanes, traffic flows, related to intersecton, speed limit, roadway-alignment and light condition are chosen; for vehicle related ones, pre-crash movement, relative force direction, pre-crash location are chosen; moreover, in order to recognize the obstacle in the collision, a new variable ‘obstacle’ is defined by origin collision type. As for driver related features, the age of driver and driver distraction information are extracted.

The predicted feature, collision type, is divided into 5 categories: Rear End, Head On/Side Swipe, Turn/Across from Opposite, Turn/Across from The Same Side, and Straight Path, which are mainly from NHTSA.

Strict refinement is required to obtain a reliable dataset, which includes the following: 1. Cases between involving only 2 vehicles; 2. Road surface conditions and vehicle types are restricted; 3. Deficiency values are removed.

While a dataset with only 6212 cases remains after refinement, in which the predicted feature presents a severe imbalance, thus, resampling method such as SMOTE and ROS are used for data augmentation; furthermore, grouping is performed for machine learning.

To construct a prediction model by pre-crash information, several machine learning methods such as Logistic Regression, SVM, Random Forest and LightGBM are used to compare total prediction performance by weighted F1 score.

The result shows that compared with other models, model based on LightGBM algorithm holds the best prediction performance of a F1 score over 92%, , which is shown in Fig.1.

Furthermore, cross-validation performance attains better after resampling. Thus, it is clear the prediction model by LightGBM shows a good performance on the dataset, which means it is possible to predict collision type by using only pre-crash information, while this type of prediction model is capable to be applied in ADS to prevent serious accidents from occurring or minimize damage of collisions.

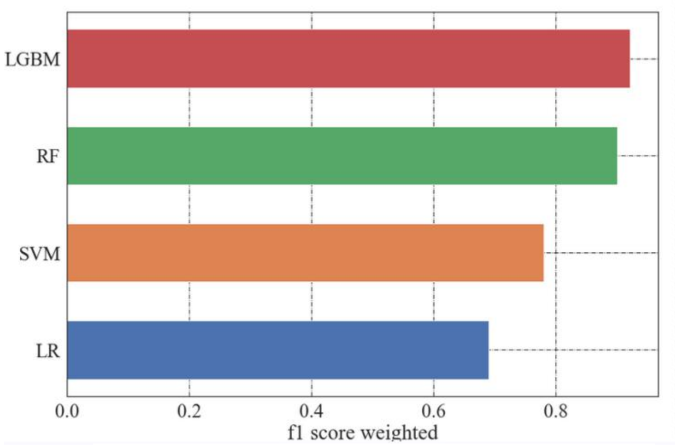

Fig.1 A Bar Graph of Comparison of models