Apollo Obstacle Trajectory Prediction

This document describes the data format and usage instructions for the obstacle trajectory prediction in the Baidu Apollo Project.

1. Introduction to dataset

This obstacle trajectory prediction dataset provides 20,000 obstacle movement data, including 10,000 training data and 10,000 test data, each of which represents the current movement characteristics and the behavioral intention of an obstacle. The training data can be used to train the algorithm model; the test data can be used to test the algorithm. In addition, the dataset provides 100 downloadable data for debugging algorithms, data analysis and data visualization.

2. Data Labeling

The dataset is sampled from the real road scene, and consists of the real motor vehicle obstacles on the road. It contains the historical data of the vehicle multi-frame information, and obstacle features and lane features generated by a series of perception processing. The labeling of the data is based on whether the obstacle is conducted automatically by the status of the vehicle – whether it will be within the current lane in 1 second. Thus, a binary classification problem is formed: positive labeling means the obstacle will be within the specific lane in 1 second. Non-negative labeling means the obstacle will not in the specific lane in 1 second. The following table shows some details.

Label	Introduction
-1	The vehicle is not within the specific lane and will not cut in there in 1 second.
0	The vehicle is within the specific lane but it will not keep this lane in 1 second.
1	The vehicle is within the specific lane and it will keep this lane in 1 second.
2	The vehicle is not within the specific lane but will cut in there in 1 second.

2.1 Sensors Integration

The data is generated from a variety of sensors, as shown in Figure 1. It includes Lidar (Velodyne HDL-64E S3), two cameras (one uses a 25mm focal length lens and the other uses a 6mm focal length lens), high definition map and localization system. For more details, please refer to other data platform in Apollo.

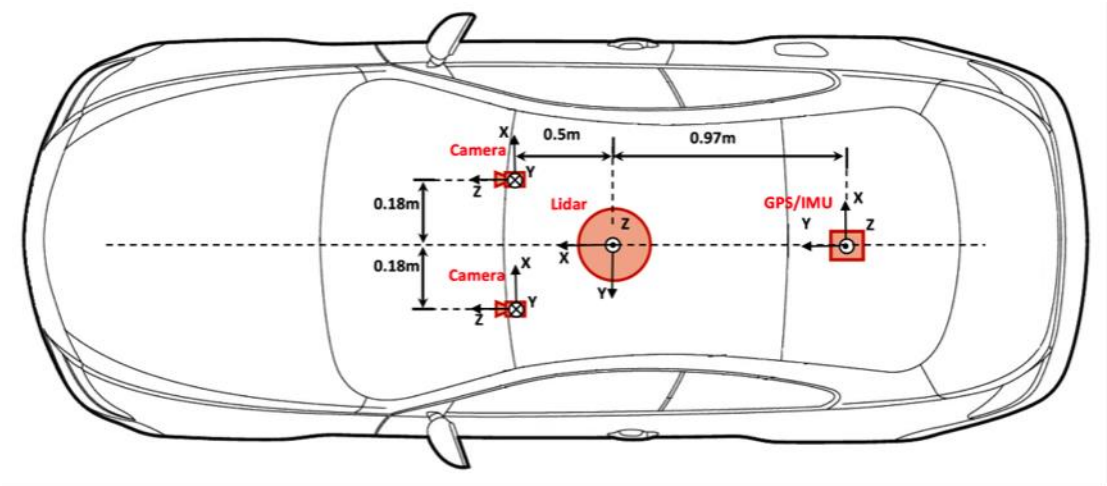

Figure 1: Sensors Integration

3. Data Format

Prediction dataset and labels are in the form of h5 files. Each data has 63 dimensions, where the first 62 dimensions are the features of obstacles and the last dimension is the label. The features are divided into two categories, obstacle-related features and lane-related features. Every feature is of type float which has precision of 32-bit. The coordination is lane SL coordination system. For more details, please refer to Table 1 and Figure 2.

Table 1 Introduction to features

Category	Dimension	Introduction
Obstacle	1~5	Current θ, historical average θ, difference between current θ and historical average θ, difference between previous θ and current θ, and its derivative.
	6~10	Current L, historical average L, difference between current θ and historical average L, difference between previous L and current L, and its derivative.
	11	Average speed.
	12	Current acceleration.
	13~15	Distance to left lane bound, average rate of approaching to left bound, and current rate of approaching to left bound.
	16~18	Distance to right lane bound, average rate of approaching to right bound, and current rate of approaching to right bound.
	19~22	If the lane type is straight, left turn, right turn, U turn, respectively. 1 means true, 0 means false.
Lane	23~62	10 groups of lane points features, each group has 4 features. In each group, the four features are: • The sine value of the difference between vehicle heading and lane point heading. • The L coordinate value of the lane point. • The heading of the lane point. • The difference between vehicle heading and lane point heading.

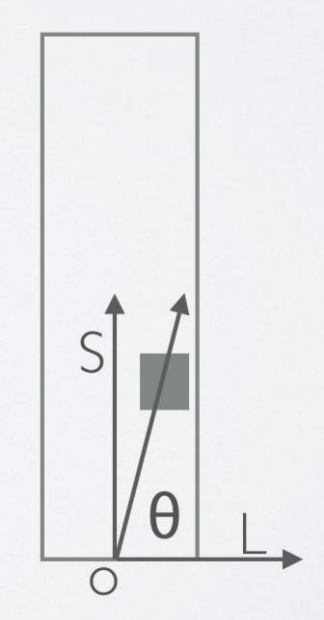

图 2 SL 坐标系

3.1 Data Usage

This dataset can be used to develop and test prediction algorithms for obstacle vehicles. A machine learning algorithm can be constructed using the provided training set for model training and using the test set for evaluation. Rule-based algorithms can directly use the test set for the effectiveness of testing.

3.2 Prediction Results

The user needs to predict whether the obstacle will be within the specific lane in 1 second given the obstacle features and lane features or not.

3.3 Evaluation

Based on the model provided by the user, we will use the test set to evaluate the model's classification precision and recall the level of the algorithm