Data Center

▌Latest Data

KDD CUP 2020 Data Set (New)

Data Description

This data set is from the trajectory data and order data (including idle transition probability, order cancellation probability and hexagon grid data) of DiDi Express and DiDi Premier drivers within the Second Ring Road of Chengdu City, which measures 65 square–kilometers in total. The measurement interval of the track points is approximately 2–4 seconds. The track points were bound to physical roads so that the trajectory data and the actual road information are matched. The driver and trip order information were encrypted and anonymized.

Data Size

The original data is about 100GB, and about 8GB after compression.

Range of Geographic Coordinates of Driving Trajectories

[30.727818,104.043333],[30.726490,104.129076]
[30.655191,104.129591],[30.652828,104.042102]

Date Range

From November 1 to November 30, 2016

⊕ Apply Again

▌Copyright and Reference

In your paper or industry report based on your analysis of the shared data, please attribute the source of the data to https://gaia.didichuxing.com, and include "Data source: DiDi Chuxing GAIA Open Dataset Initiative" in the acknowledgment section.

▌Trajectory Data

Trajectory Data Accuracy: 2–4s

Field	Type	Sample	Comment
Driver ID	String	glox.jrrlltBMvCh8nxqkt dr2dtopmlH	Anonymized
Order ID	String	jkkt8kxniovlFuns9qrrlvs t@iqnpkwz	Anonymized
Time Stamp	Int	1501584540	Unix timestamp, in seconds
Longitude	Float	104.04392	GCJ–02 Coordinate System
Latitude	Float	104.04392	GCJ–02 Coordinate System

▌Ride Request Data

Field	Type	Sample	Comment
Order ID	String	mjiwdgkqmonDFvCk3n	Anonymized

		tBpron5mwfrqvl	
Ride Start Time	Int	1501581031	Unix timestamp, in seconds
Ride Stop Time	Int	1501582195	Unix timestamp, in seconds
Pick–up Longitude	Float	104.11225	GCJ–02 Coordinate System
Pick–up Latitude	Float	30.66703	GCJ–02 Coordinate System
Drop–off Longitude	Float	104.07403	GCJ–02 Coordinate System
Drop–off Latitude	Float	30.6863	GCJ–02 Coordinate System
Reward Units	Float	12.75	Model prediction value

Idle Transition Probability

The transition probability of idle drivers discretized using a hexagon grid system which is defined in the Hexagon Grid table:

Field	Type	Sample	Comment
Hour	Int	14	Hour of the day
Origin Grid ID	String	272ed2593c3a2fff	Anonymized
Destination Grid ID	String	ed0375d91f493664	Anonymized
Transition Probability	Float	0.23	Probability of transition from origin to destination

Order Cancellation Probability

The probability of cancellation of an order given different pick–up distances

Field	Type	Sample	Comment
Order ID	String	mjiwdgkqmonDFvCk3n tBpron5mwfrqvl	Anonymized
Cancellation Probability (200m)	Float	0.01	Order cancellation probability when pick–up distance is 200m
Cancellation Probability (400m)	Float	0.07	Order cancellation probability when pick–up distance is 400m
...	...	...	...
Cancellation Probability (2000m)	Float	0.87	Order cancellation probability when pick–up distance is 400m

Hexagon Grid Data

The location of each hexagon grid including the longitudes and latitudes of its six vertices

Field	Type	Sample	Comment
Grid ID	String	0c196347357fa044	Anonymized
Vertex 1 Longitude	Float	104.20157	GCJ–02 Coordinate System
Vertex 1 Latitude	Float	30.54293	GCJ–02 Coordinate System
Vertex 2 Longitude	Float	104.20566	GCJ–02 Coordinate

Vertex 2 Longitude	Float	104.20363	System
Vertex 2 Latitude	Float	30.54709	GCJ–02 Coordinate System
Vertex 3 Longitude	Float	104.20363	GCJ–02 Coordinate System
Vertex 3 Latitude	Float	30.55275	GCJ–02 Coordinate System
Vertex 4 Longitude	Float	104.19751	GCJ–02 Coordinate System
Vertex 4 Latitude	Float	30.55424	GCJ–02 Coordinate System
Vertex 5 Longitude	Float	104.19341	GCJ–02 Coordinate System
Vertex 5 Latitude	Float	30.55008	GCJ–02 Coordinate System
Vertex 6 Longitude	Float	104.19544	GCJ–02 Coordinate System
Vertex 6 Latitude	Float	30.54443	GCJ–02 Coordinate System

Applied

Data Name	Application Time	Result	Details
KDD CUP 2020 Data Set (New)	2020–06–13 18:58	Passed	Download