Traffic Flow Prediction In a U.S. Metropolis

Jingbao Luo

Nanjing University of Science and Technology, China

Problem Definition

The March edition of the 2022 Tabular Playground Series is a prediction project about time series data. We'll forecast twelve-hours of traffic flow in a major U.S. metropolitan area. Time, space, and directional features give us the chance to model interactions across a network of roadways.

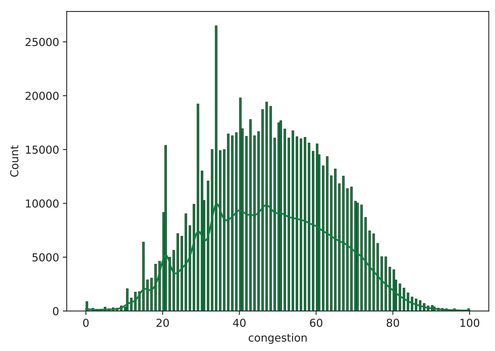
File	Description	Attribution
train.csv	traffic congestion from April through September of 1991	row_id,time,x,y, direction,congestion
test.csv	hourly predictions on the day of 1991-09-30	row_id,time,x,y, direction

Data Processing

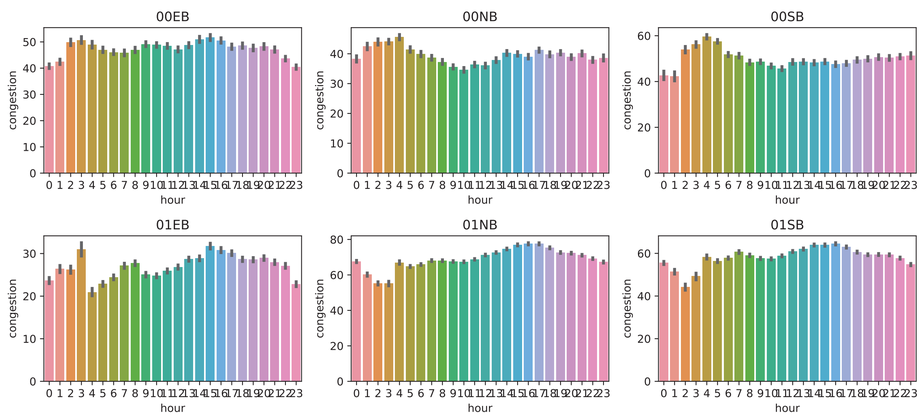
- Split Time data
- Merge x,y,row_id

Data Description

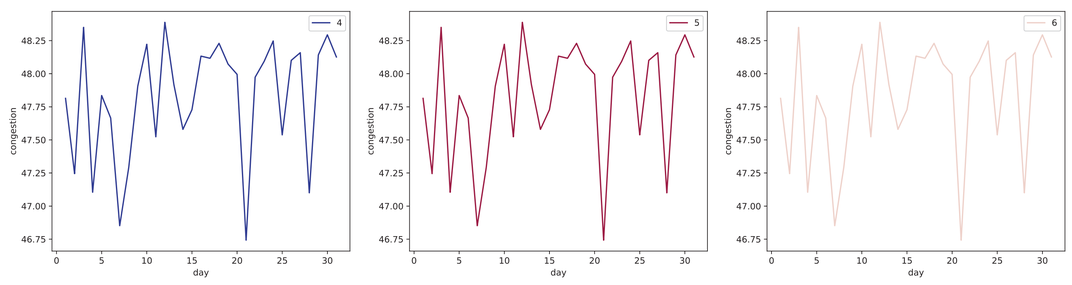
Congestion



The effect of hour on congestion group by road



The effect of day on congestion group by month

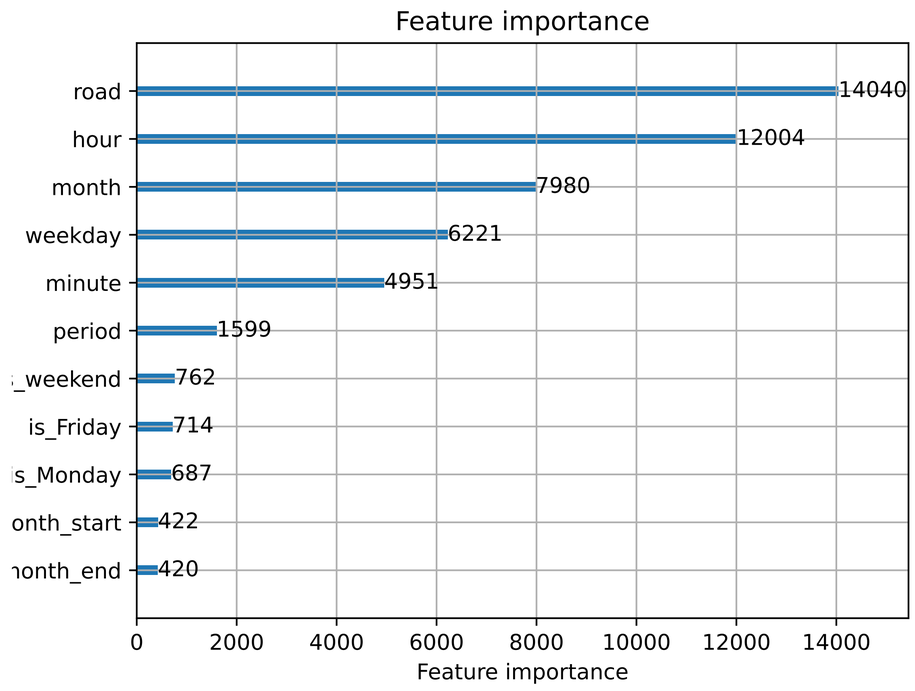


Model Train and Evaluation

- Data feature

  road,hour,month,weekday,minute,period,is_weekend,is_Friday,is_Monday, Month_start,Month_end

- By using sklearn.package ,split train data and evaluation data
- Use ligthgbm model



Evaluation result

Index	Result
explained_variance_score	0.7277243544483329
mean_absolute_error	6.167491947603395
r2_score	0.7277251135484366

Result

Prediction Result

Row_id	Congestion	Row_id	Congestion
848835	47	848836	33
848837	39	848838	54
848839	64	848840	23
848841	28	848842	70
848843	25	848844	47
848845	46	848846	25
848847	69	848848	60

Conclusion

This experiment simply obtains the eigenvalues from the time data and uses the lightgbm model for training, but does not optimize the model.

TULIP *Team for Universal Learning and Intelligent Processing*

The March edition of the 2022 Tabular Playground Series on Kaggle Competition