KDD CUP 2017
Travel Time Prediction

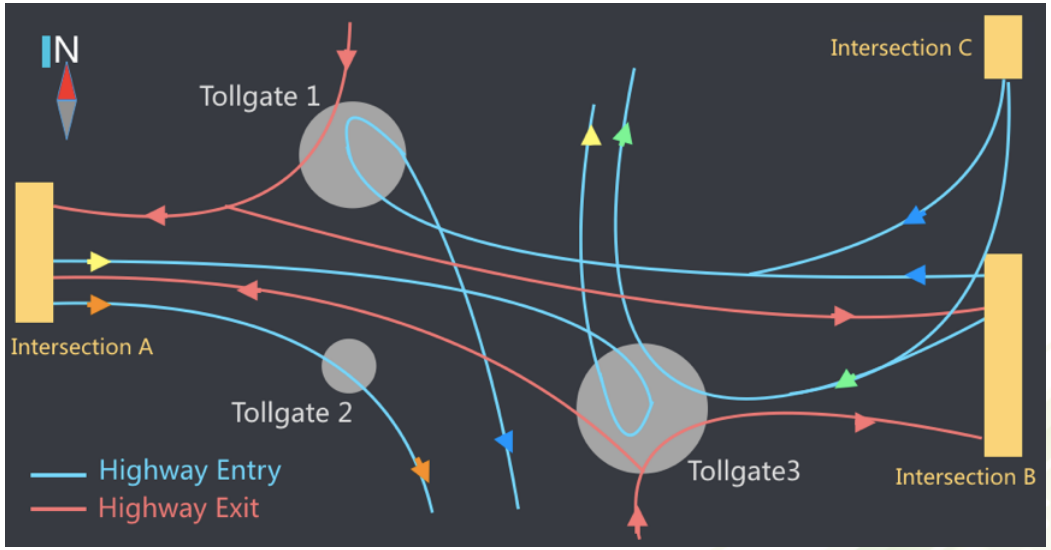
Warriors

Team Members

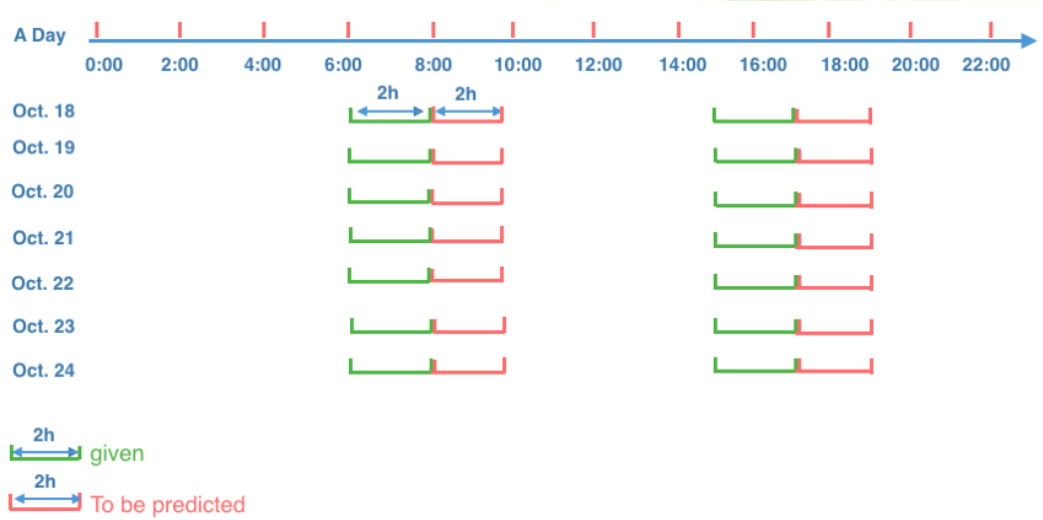
Hengxing Cai, Sun Yat-sen University, Guangdong Key Laboratory of Intelligent Transportation Systems (Team Leader)
Runxing Zhong, Beihang University
Chaohe Wang, Southwest Jiaotong University, Intel
Ruihuan Zhou, Southwest Jiaotong University
Kejie Zhou, University of Chinese Academy of Sciences
Hongyun Lee, University of Chinese Academy of Sciences
Kele Xu, National University of Defense Technology
Zhifeng Gao, Peking University
Renxin Zhong, Sun Yat-sen University, Guangdong Key Laboratory of Intelligent Transportation Systems
Jiachen Luo, Sun Yat-sen University, Guangdong Key Laboratory of Intelligent Transportation Systems
Yao Zhou, Chongqing University of Posts and Telecommunications, Tencent
Ming Ding, Beijing China-Power Information Technology Co. Ltd
Lang Li, ChinaTelecom Bestpay Co. Ltd
Qiang Li, Fudan University
Da Li, Beihang University
Nan Jiang, Beihang University
Xu Cheng, China Mobile Communications Corporation
Shiwen Cui, Harbin Engineering University
Hongfei Ye, Shanghai Jiao Tong University
Jiawei Shen, Shanghai China-Cubee Information Technology Co. Ltd

Background and Task Description

Travel time prediction plays a very important role in traffic status monitoring.



Estimate the next 2 hours average travel time based on historical and current traffic data.



Framework

Preprocessing → Feature engineer → Model → Ensemble

Preprocessing	Feature engineer	Model	Ensemble
Remove outliers	Feature construction	Simultaneous prediction	Weighing
trend filtering	Feature selection	Rolling Prediction	
		Time series	

Preprocessing

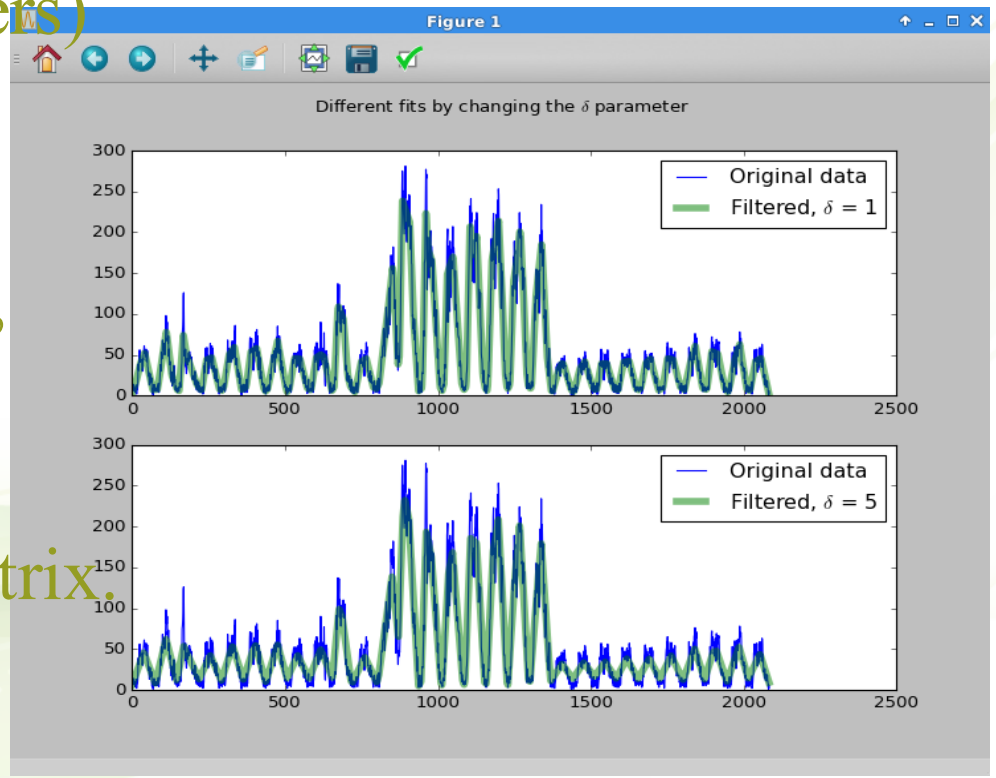
- Remove Outliers (National Day: 10.1-10.7)
- Trend Filtering (missing data, outliers)

$$\frac{1}{2}\|x-y\|_2^2+\delta\|Dx\|_1$$

Where x is output data, y is input data,

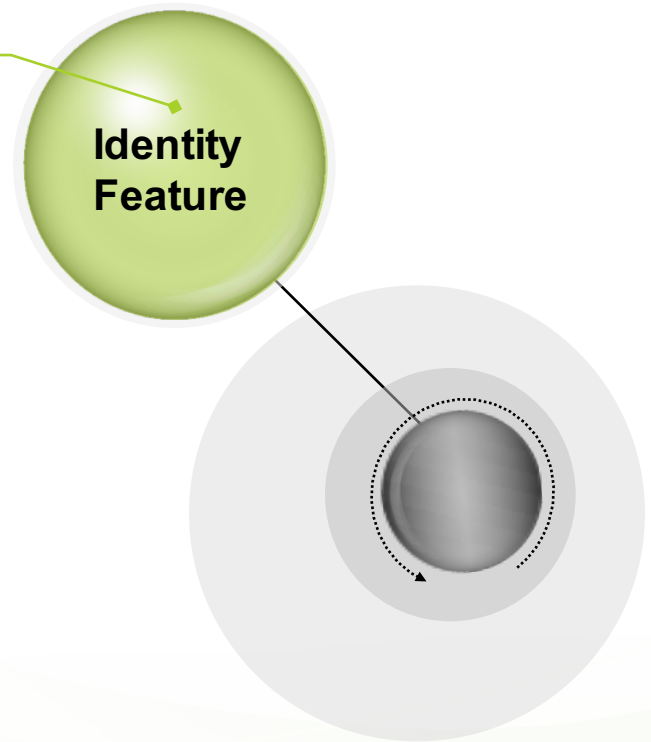
δ is smoothing parameter,

and D is a second-order difference matrix.



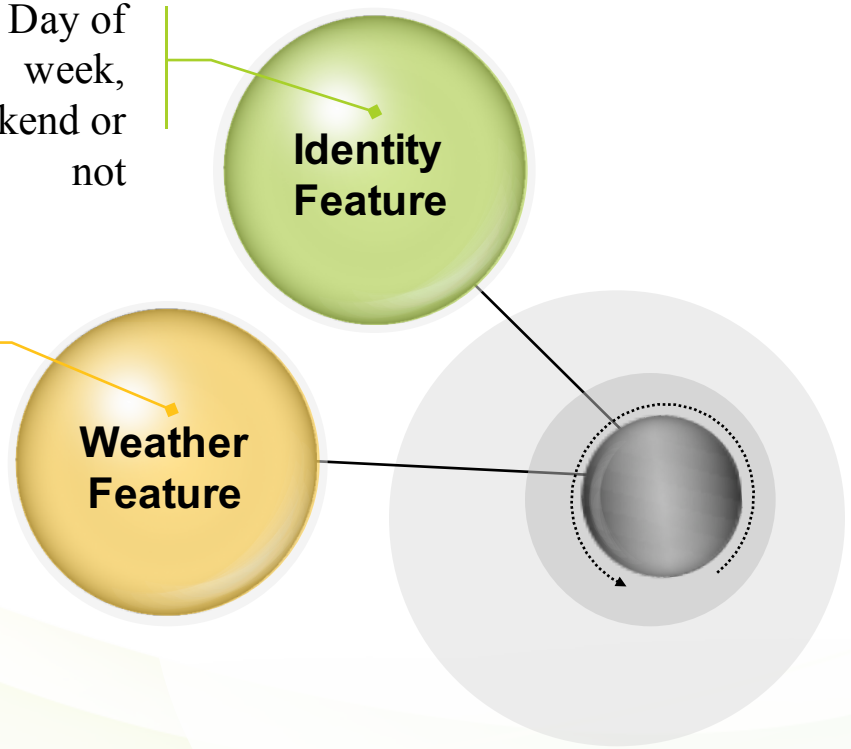
Feature Construction

Road ID, Time of Day (hour, min), Day of week, Tollgate id, weekend or not



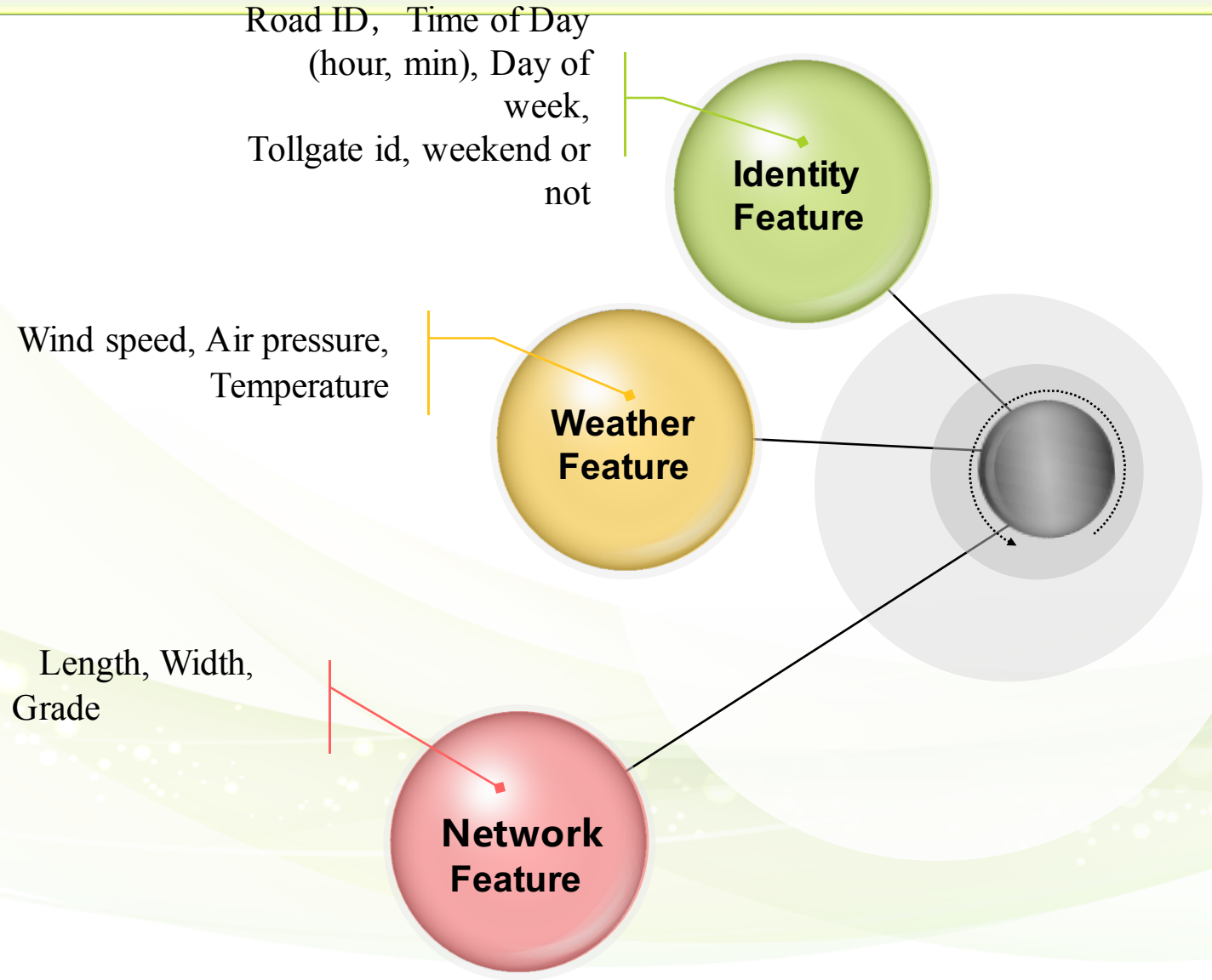
Feature Construction

Road ID, Time of Day (hour, min), Day of week, Tollgate id, weekend or not

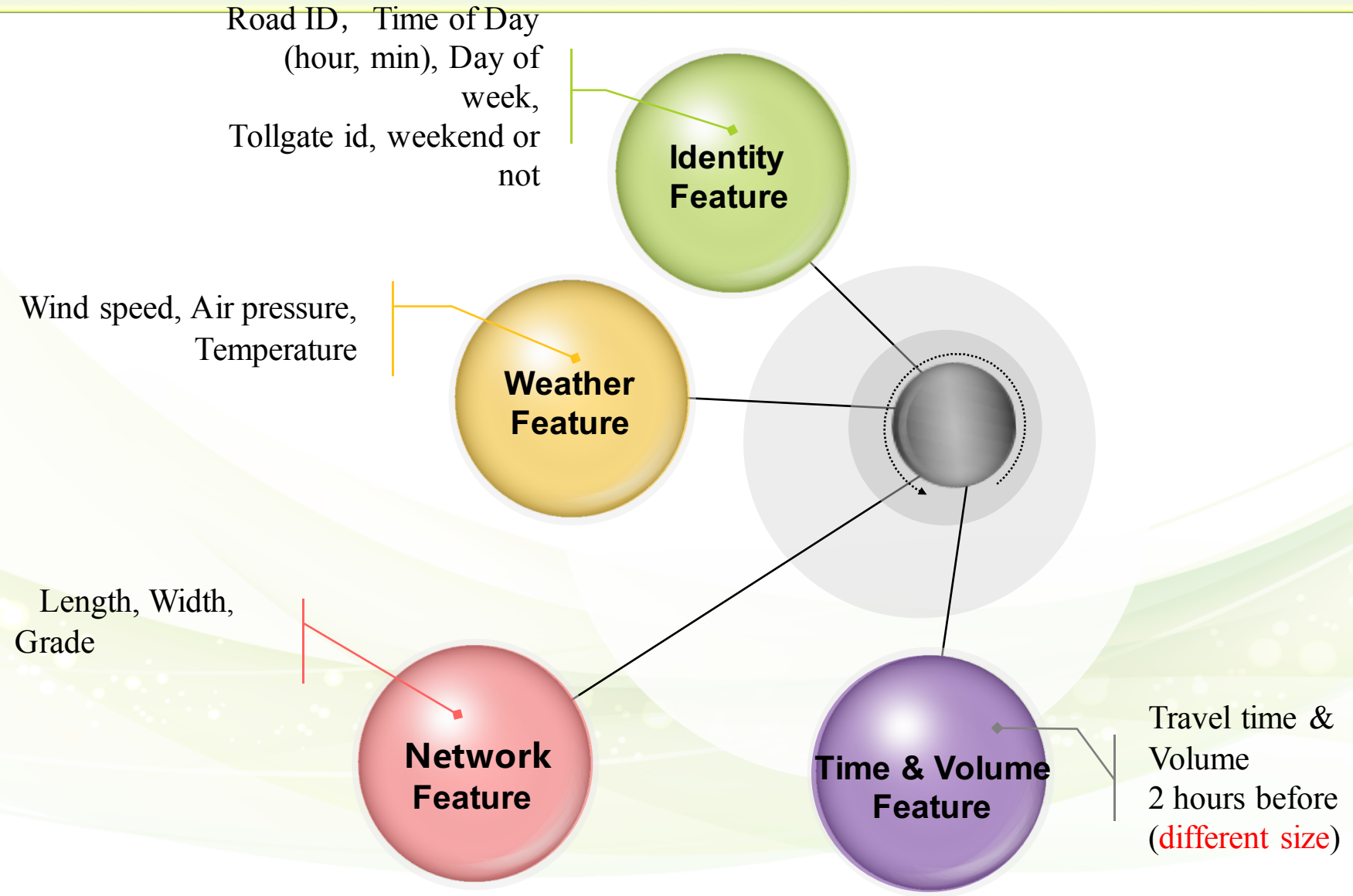


Wind speed, Air pressure, Temperature

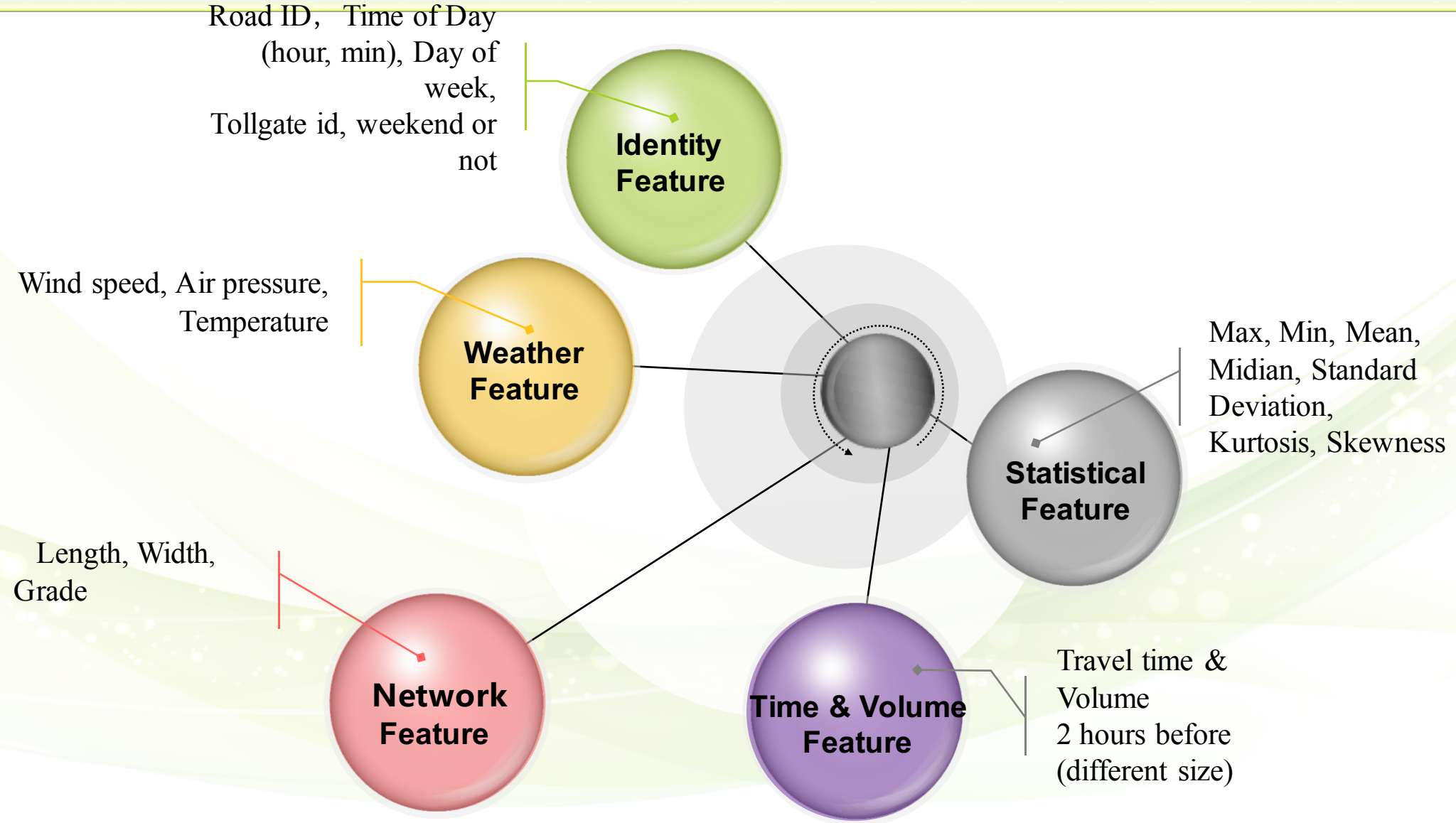
Feature Construction



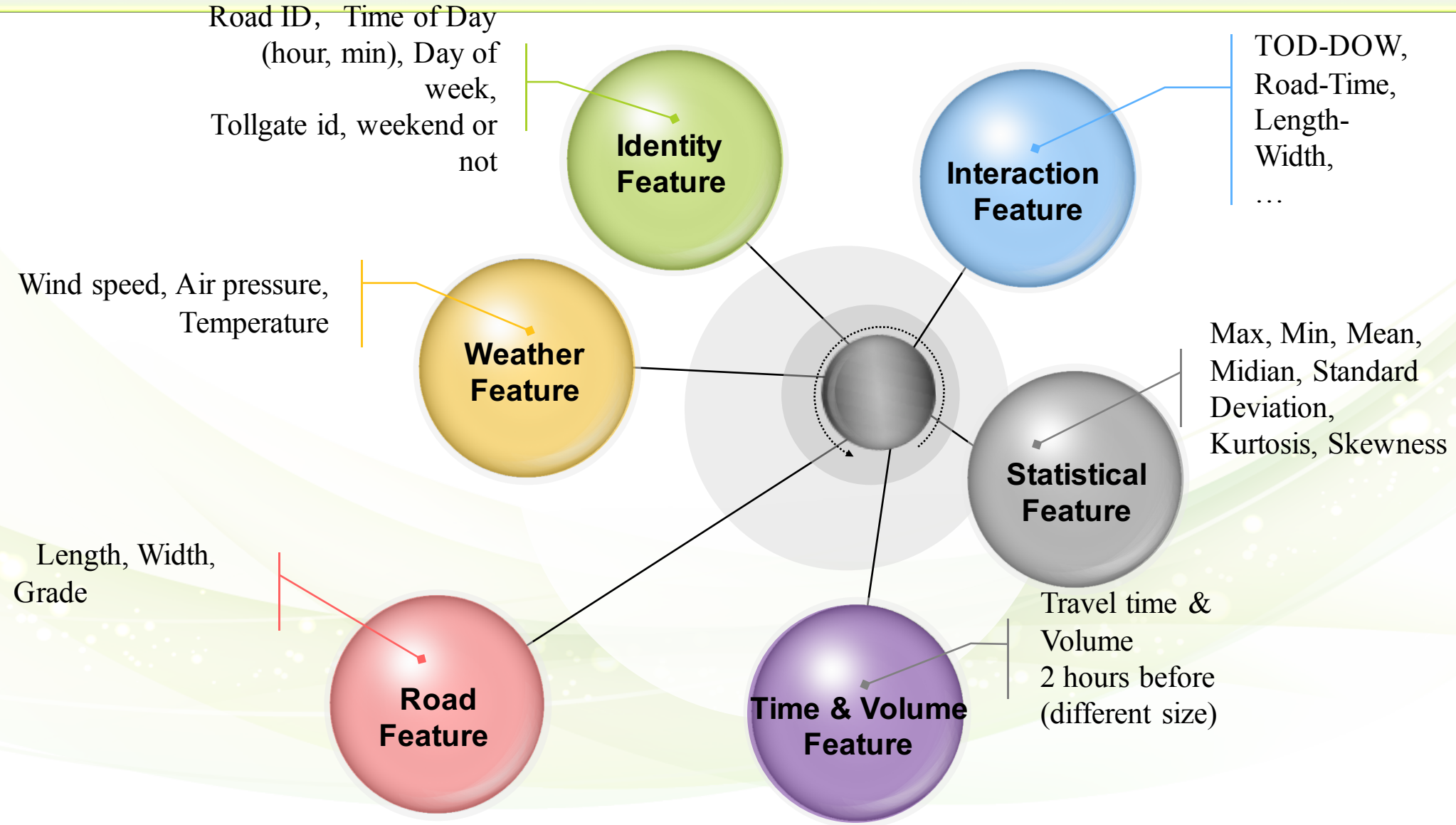
Feature Construction



Feature Construction



Feature Construction



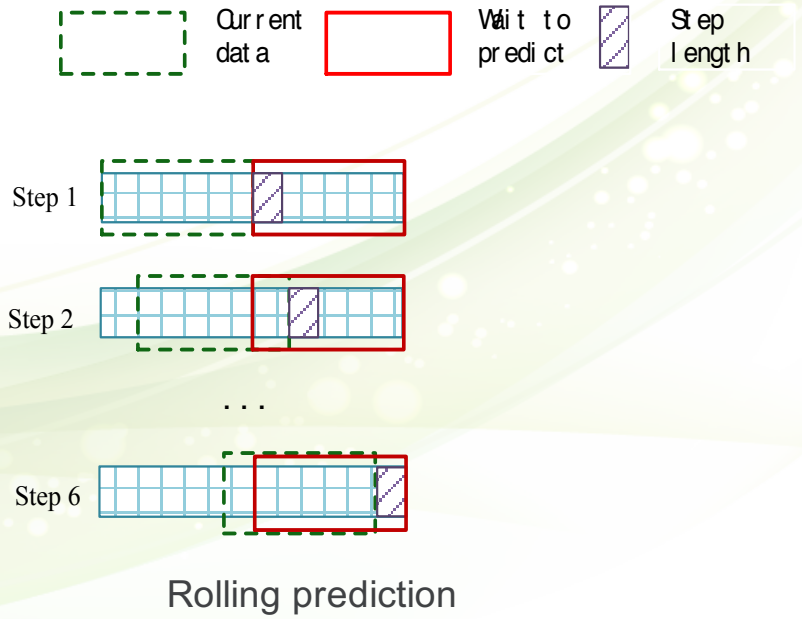
Feature Selection

For interaction feature:

- Removing features with low variance
- Feature selection using tree-based model (top 30%)

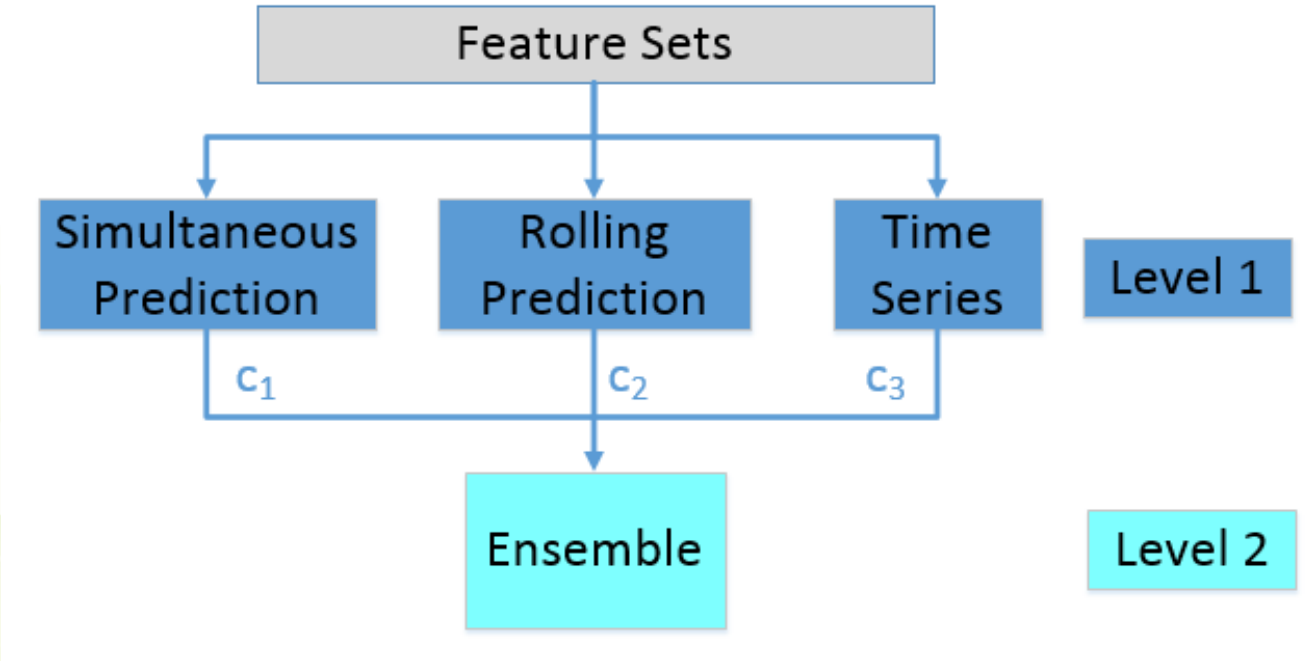
Model

- Model1: Simultaneous Prediction (Xgboost)
- Model2: Rolling Prediction (Xgboost)
- Model3: Time Series Prediction (Arima)



Ensemble learning

- Result = 0.4*Model1 + 0.4*Model2 + 0.2*Model3



THANK YOU