Highway Tollgates Traffic Flow Prediction

Task 1. Travel Time Prediction

Yide Huang

School of Mathematical Sciences
Zhejiang University, Hangzhou, China
hydantess@163.com

Outline

1 introduction

Task:

To estimate the average travel time from designated intersections to tollgates:

a. Routes from Intersection A to Tollgates 2 & 3
b. Routes from Intersection B to Tollgates 1 & 3
c. Routes from Intersection C to Tollgates 1 & 3

Given:

The road network topology , vehicle trajectories , historical traffic volume at tollgates and weather data
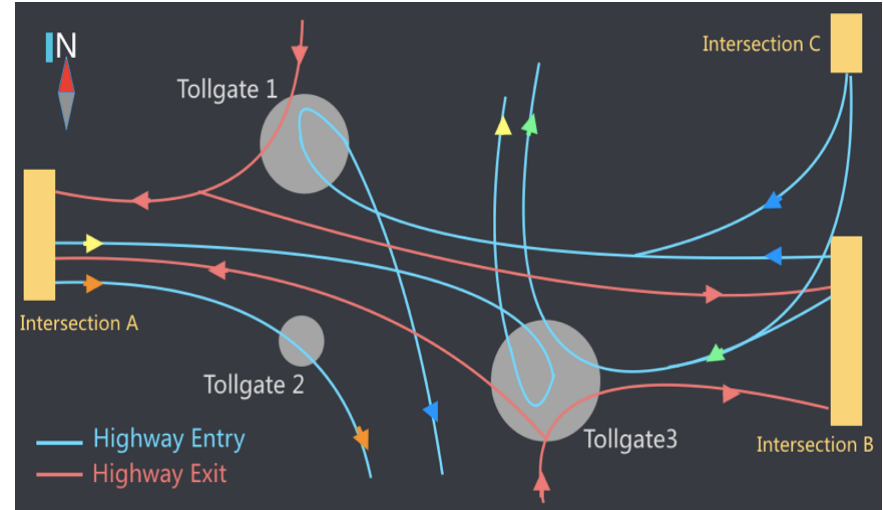

Evaluation Metrics:

$$MAPE = \frac{1}{R}\sum_{r=1}^{R}\left(\frac{1}{T}\sum_{t=1}^{T}\left|\frac{d_{rt} - p_{rt}}{d_{rt}}\right|\right)$$

d_{rt}: actual average travel time for route r during time window t

p_{rt}: predicted average travel time for route r during time window t

R: the number of routes

T: number of to-predict time windows

2 Problem Understanding

2.1 Influence factor of travel time

1   Weather conditions

2   Time of the day

3   Holidays

4   Traffic conditions

5   Road network topology

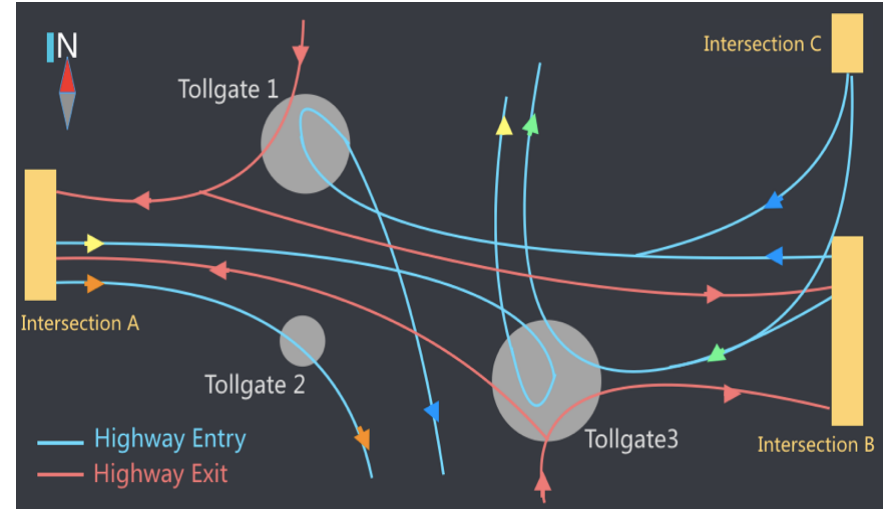# 2 Problem Understanding

2.1 Influence factor of travel time

5 Road network topology
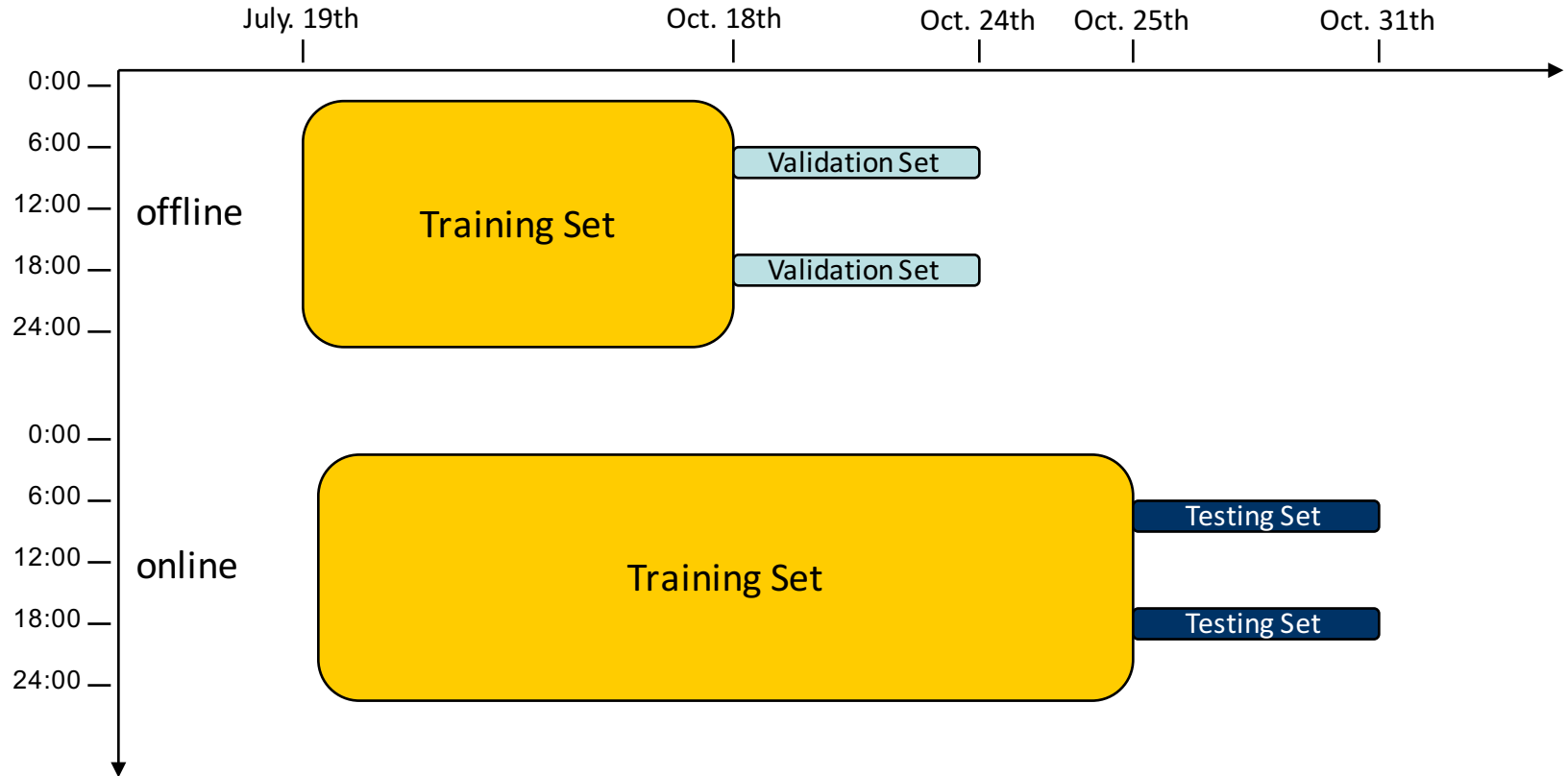
1 Sharing link or not

(A-3, B-3, C-3)

2 Link information

(length, width and number of lanes)

2 Problem Understanding

2.2 Data set partition

3 Features

3.1 Feature Engineering

1 Weather

> 1 Human comfort index
> 2 Precipitation
> 3 Their statistical features (mean, sum)

2 Time

> 1 Time o'clock
> 2 Weekday, weekend or holiday
> 3 Whether it's rush hour

3 Road Features

> 1 The number of cars
> 2 The ratio of road's car number
> 3 Road's ETA
> 4 Links' ETA
> 5 The weighted mean of Links' velocity based on their length
> 6 Whether there is an emergency
> 7 The rank feature of Links' velocity
> 8 Traffic volume
> 9 Average capacity of vehicle
> 10 The number and ratio of car that has no ETC
> 11 Road network topology features
> 12 Last week's historical ETA and car number
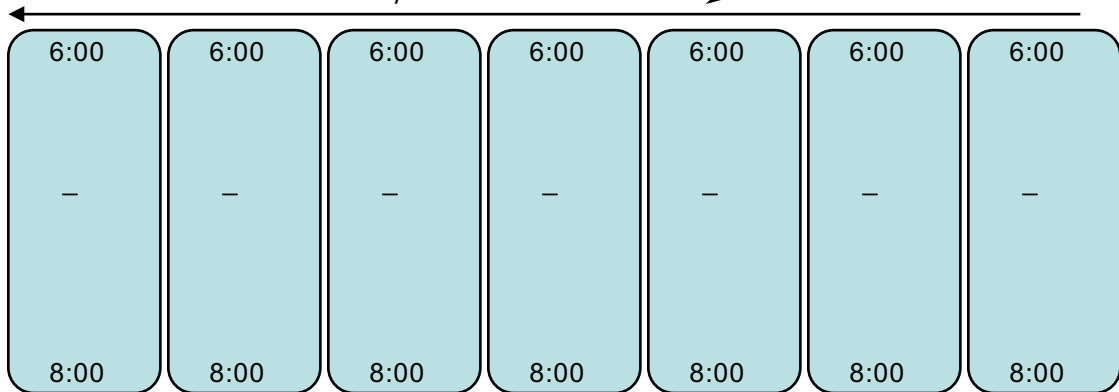> 13 Their statistical features (mean, sum)
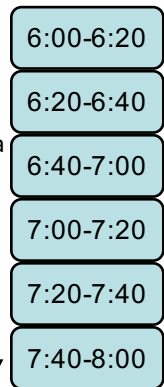
3 Features

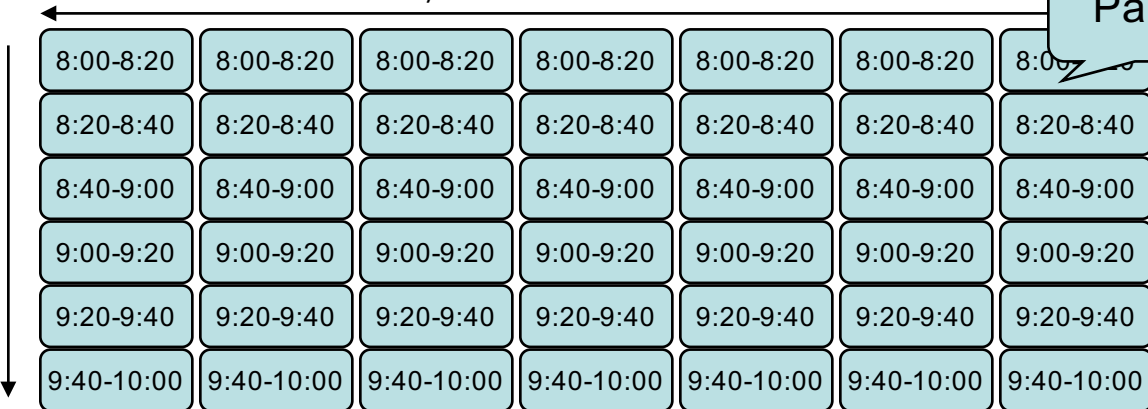## 3.2 Missing data processing

Part 3

7 days' historical data

6:00	6:00	6:00	6:00	6:00	6:00	6:00
_	_	_	_	_	_	_
8:00	8:00	8:00	8:00	8:00	8:00	8:00

7 days' historical ETA

Part 1

Missing data Processing

6:00-6:20
6:20-6:40
6:40-7:00
7:00-7:20
7:20-7:40
7:40-8:00

To predict target time windows one by one, day after day.

Part 2

Missing data Processing

8:00-8:20	8:00-8:20	8:00-8:20	8:00-8:20	8:00-8:20	8:00-8:20	8:00-8:20
8:20-8:40	8:20-8:40	8:20-8:40	8:20-8:40	8:20-8:40	8:20-8:40	8:20-8:40
8:40-9:00	8:40-9:00	8:40-9:00	8:40-9:00	8:40-9:00	8:40-9:00	8:40-9:00
9:00-9:20	9:00-9:20	9:00-9:20	9:00-9:20	9:00-9:20	9:00-9:20	9:00-9:20
9:20-9:40	9:20-9:40	9:20-9:40	9:20-9:40	9:20-9:40	9:20-9:40	9:20-9:40
9:40-10:00	9:40-10:00	9:40-10:00	9:40-10:00	9:40-10:00	9:40-10:00	9:40-10:00

Missing label Processing

8:00-8:20
8:20-8:40
8:40-9:00
9:00-9:20
9:20-9:40
9:40-10:00

Target time windows

3 Features

3.2 Missing data processing

Part 1

Missing data
Processing

6:00-6:20
6:20-6:40
6:40-7:00
7:00-7:20
7:20-7:40
7:40-8:00

&

Part 2

Missing data
Processing

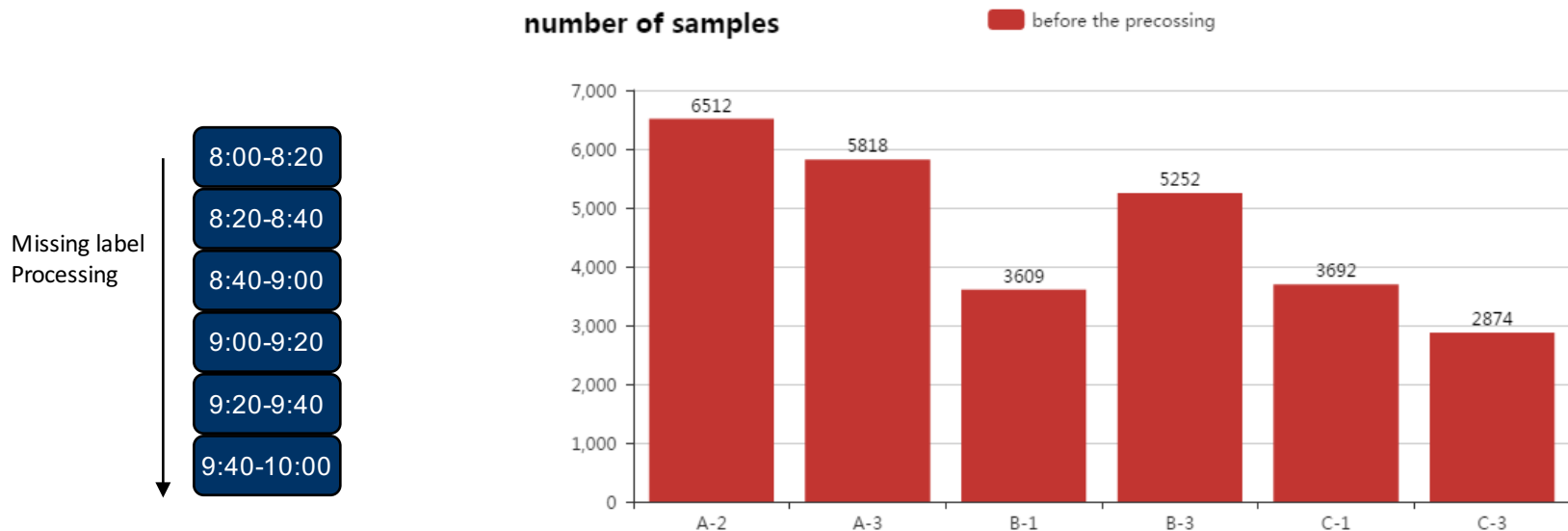| |
|---|
| 8:00-8:20 |
| 8:20-8:40 |
| 8:40-9:00 |
| 9:00-9:20 |
| 9:20-9:40 |
| 9:40-10:00 |
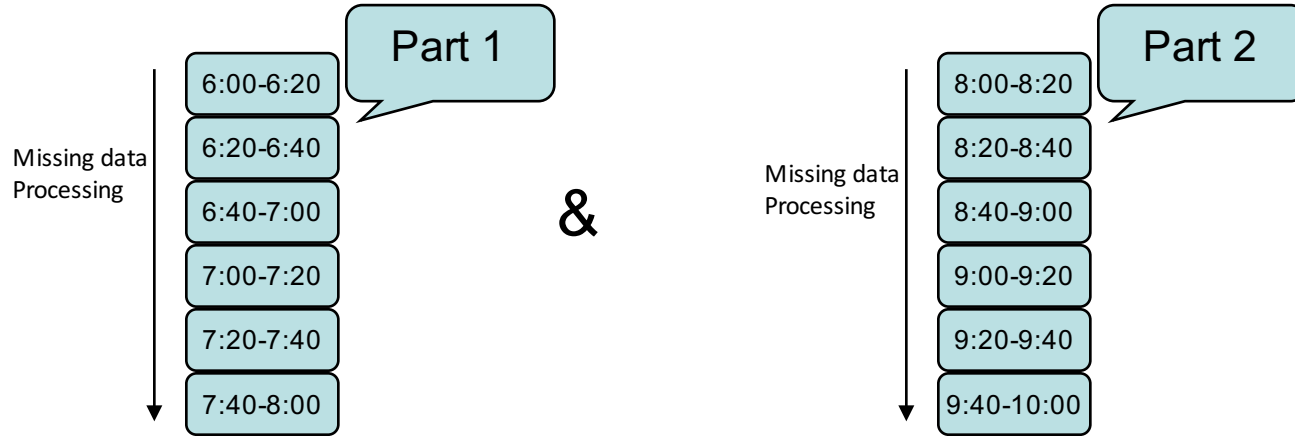
High feature importance: replaced by mean values

Low feature importance: if the number of time windows which have missing values <=3, missing values will be replaced by mean values

3 Features

3.2 Missing data processing

8:00-8:20
8:20-8:40
8:40-9:00
9:00-9:20
9:20-9:40
9:40-10:00

Missing label Processing

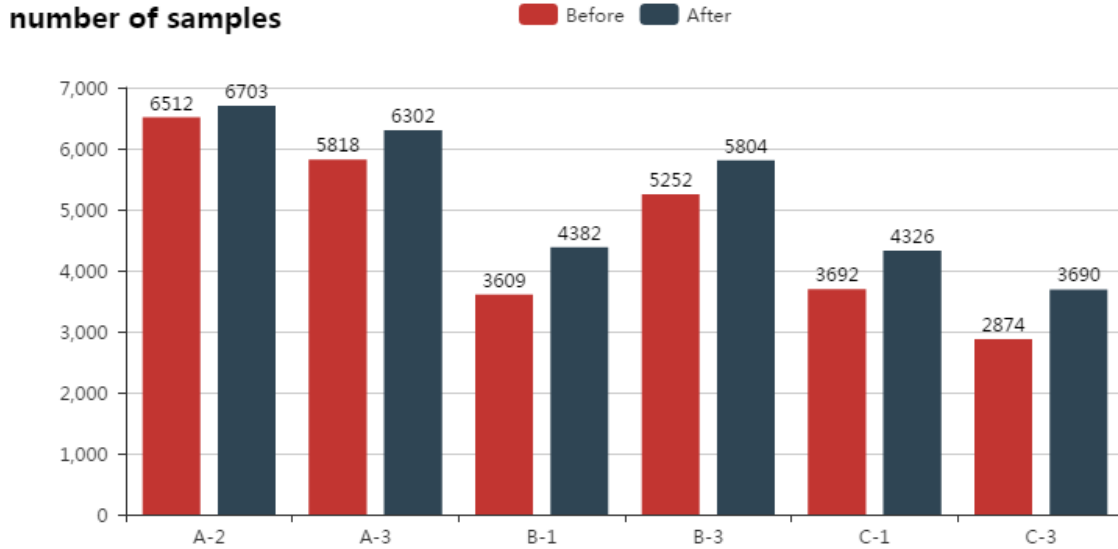**number of samples** ■ before the precossing

Balance: more samples and less noisy data

1 if the number of missing values <=3, missing value will be replaced by mean value
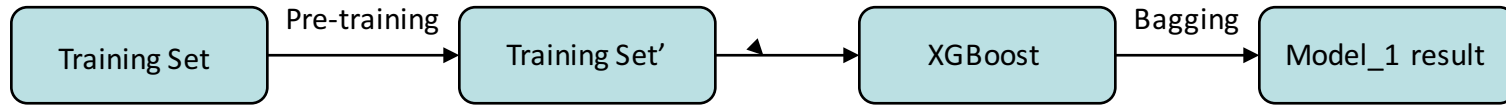
2 Pre-training

3 Features

3.2 Missing data processing

number of samples

Before | After

Route B-1, C-1, C-3: the number of samples increases about 20%

4 Models

4.1 Model_1

```
┌──────────────┐  Pre-training  ┌──────────────┐      ┌──────────────┐  Bagging  ┌──────────────┐
│ Training Set │ ─────────────► │ Training Set' │ ◄──► │   XGBoost    │ ────────► │Model_1 result│
└──────────────┘                └──────────────┘      └──────────────┘           └──────────────┘
```

1 pre-training : The ratio of samples are preserved : 0.8 - 0.95
2 Model : eXtreme Gradient Boosting
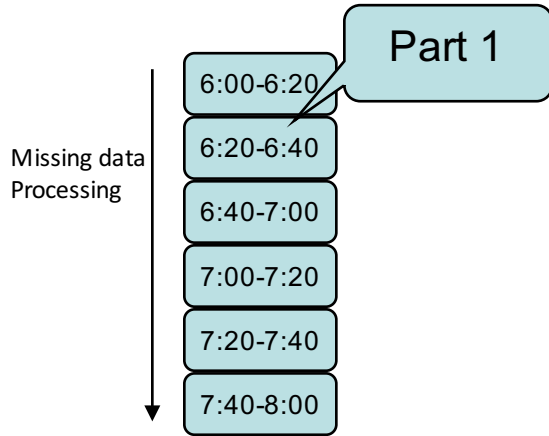3 Bagging : Different parameters, Average value
4 Model_1 result :   stage1:MAPE = 0.1785
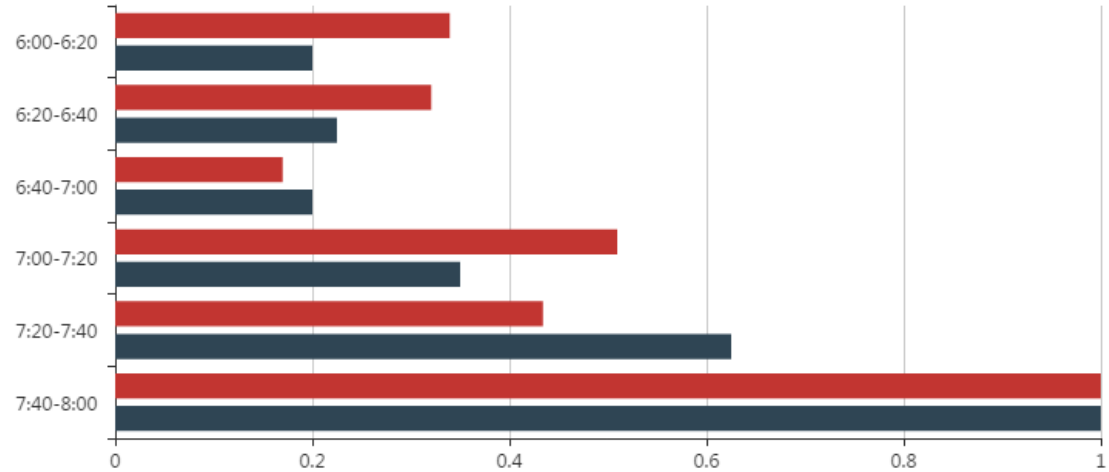   stage2:MAPE = 0.1786
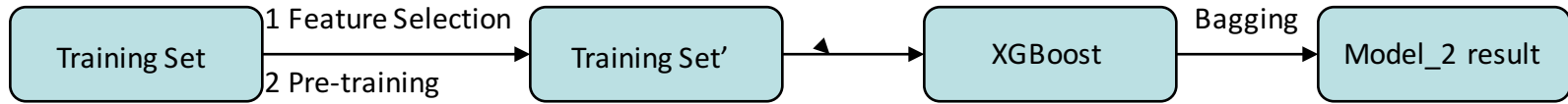
4 Models

4.2 Model_2

The last time window's features is more important than other time windows

4 Models

4.2 Model_2

```
┌──────────────┐  1 Feature Selection   ┌──────────────┐      ┌──────────────┐  Bagging  ┌──────────────┐
│              │ ─────────────────────► │              │ ◄──► │              │ ────────► │              │
│ Training Set │                        │ Training Set'│      │   XGBoost    │           │Model_2 result│
│              │  2 Pre-training        │              │      │              │           │              │
└──────────────┘                        └──────────────┘      └──────────────┘           └──────────────┘
```

1 Feature selection: Preserve the last time window's features and
         delete low-importance features of other five time windows

2 pre-training : The ratio of samples that are preserved : 0.8 - 0.95

3 Model : eXtreme Gradient Boosting

4 Bagging : Different parameters, Average value

5 Model_2 result :   stage1:MAPE = 0.1792
                       stage2:MAPE Unknown

5 Summary

MAPE:

	Model_1	Model_2	Ensemble
Stage1	0.1785	0.1792	0.1763
Stage2	0.1786	Unknown	0.1771

Future work:

1 Fully using of link information

2 Missing data processing

Thanks!

Q&A