GPS Data Analytics

Rathachai CHAWUTHAI

rathachai.creatier.pro

Rathachai Chawuthai

- Assistant Professor at Computer Engineering, KMITL
- Ph.D in Informatics from SOKENDAI (JP)$^{@2016}$
  - M.Eng from AIT (TH)$^{@2012}$, B.Eng from KMITL (TH)$^{@2006}$
- Working Experience$^{(2012\text{-}2016)}$:
  - Thomson Reuters (TH)^{4yrs}, Punsarn Asia (TH)^{1yr}, NII (JP)^{3yrs}
- Fields of Interest
  - AI, Machine Learning
  - GPS Data Analytics
  - Software Engineering
  - Semantic Technology


ASIAN INSTITUTE OF TECHNOLOGY
1959
117th AIT Graduation Ceremony
1 June 2012
DST + HCISDM
ICADL 2012
Conference Program
Rathachai Chawuthai
Asian Institute of Technology
Research Assistant

Agenda

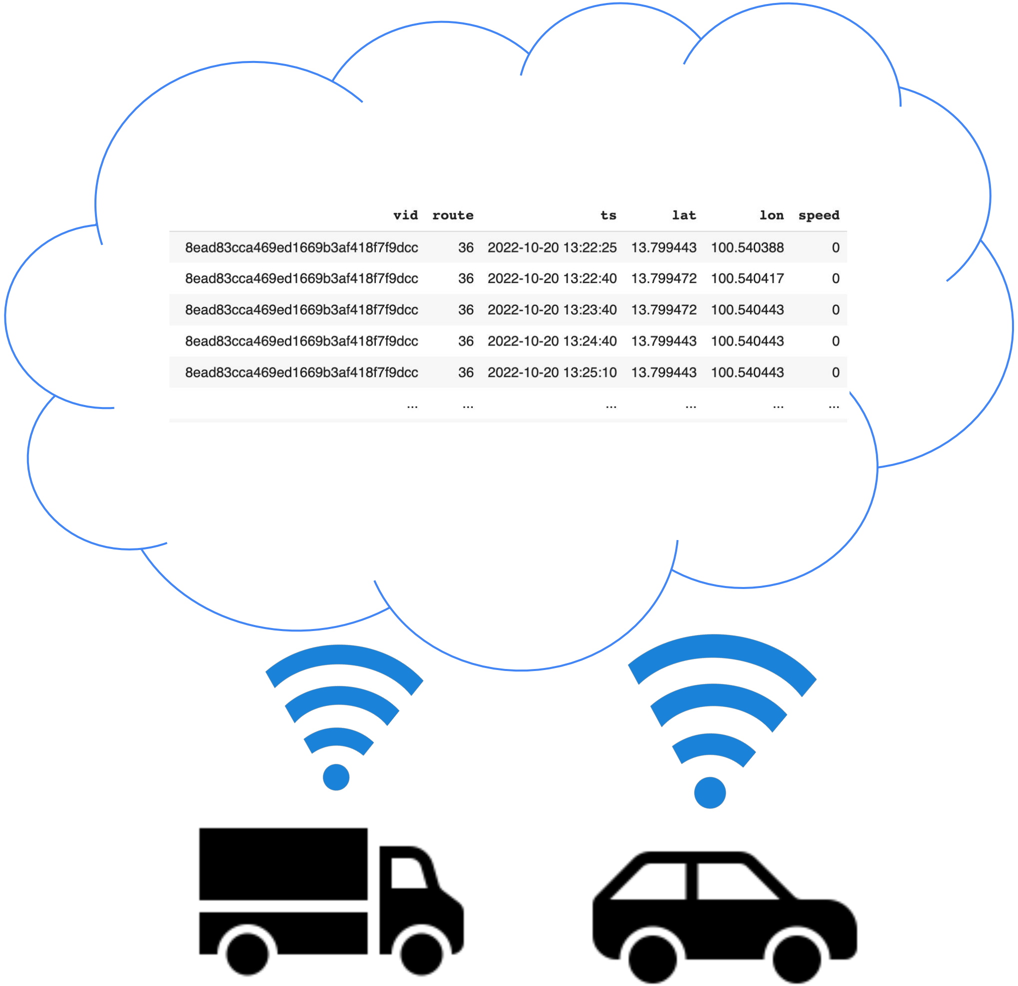
- Bus Quality
- Next Stop Prediction
- Travel Time Prediction
- Speed Prediction

INTRODUCTION

GPS Log

Minutely Transaction
from many vehicles

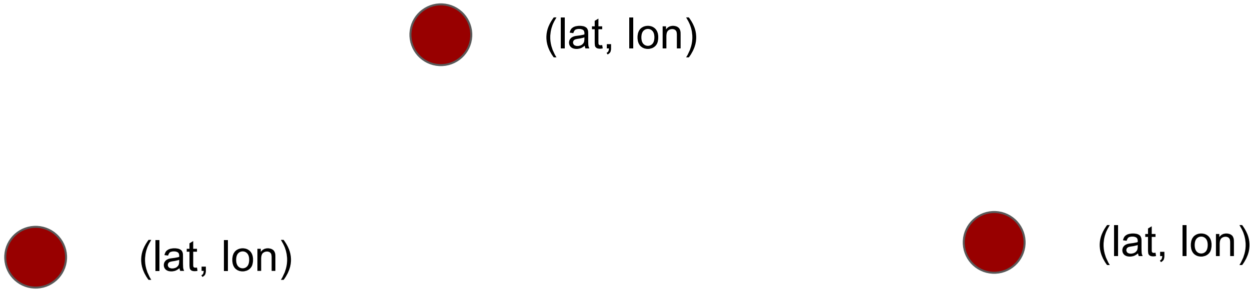
- **vid** : vehicle id
- **ts** : timestamp (date-time)
- **lat** : latitude
- **lon** : longitude
- **speed** : km/h



	vid	route	ts	lat	lon	speed
	8ead83cca469ed1669b3af418f7f9dcc	36	2022-10-20 13:22:25	13.799443	100.540388	0
	8ead83cca469ed1669b3af418f7f9dcc	36	2022-10-20 13:22:40	13.799472	100.540417	0
	8ead83cca469ed1669b3af418f7f9dcc	36	2022-10-20 13:23:40	13.799472	100.540443	0
	8ead83cca469ed1669b3af418f7f9dcc	36	2022-10-20 13:24:40	13.799443	100.540443	0
	8ead83cca469ed1669b3af418f7f9dcc	36	2022-10-20 13:25:10	13.799443	100.540443	0
	...	...	...	...	...	...

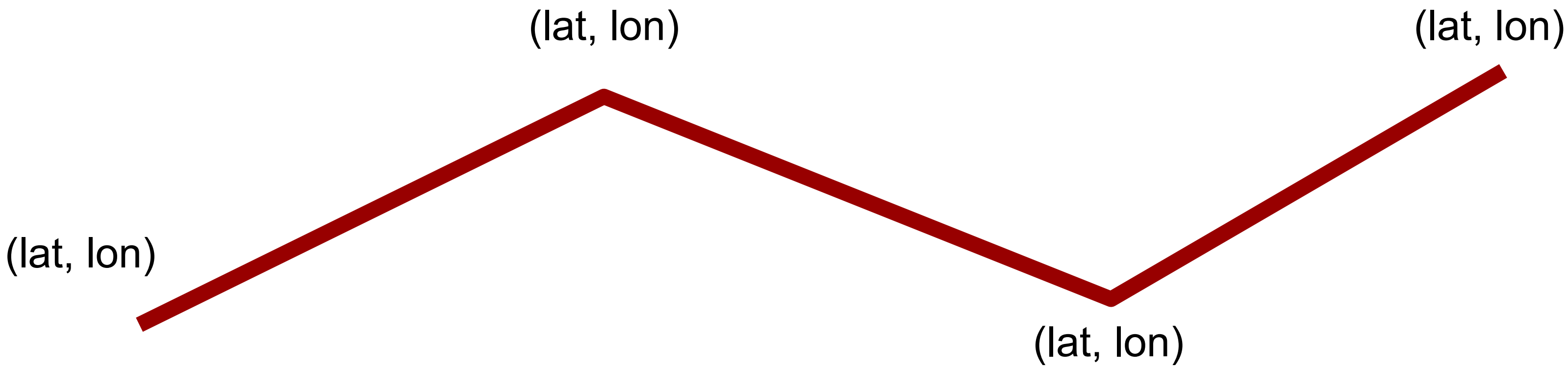
Points

coor = (13.345, 100.434)

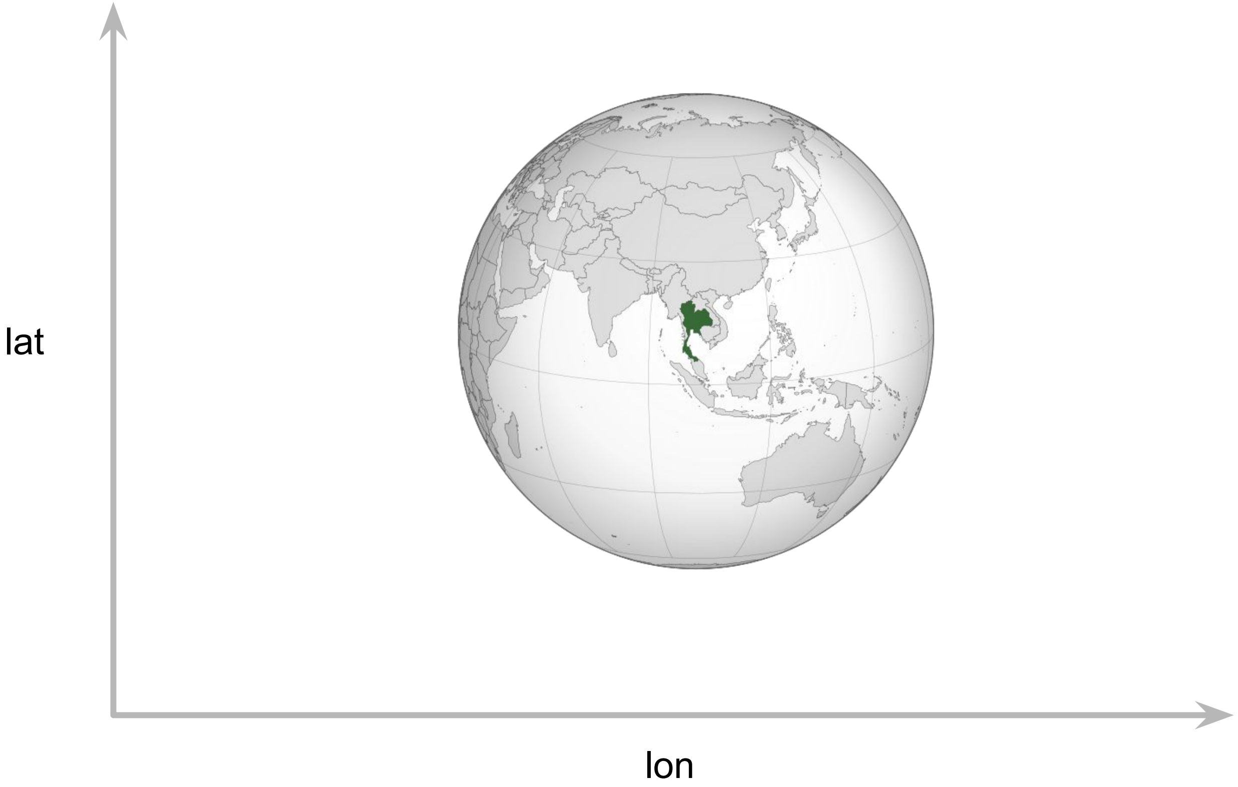


Polylines

coor = [(13.234, 100.244), (13.235, 100.345), (13.583, 100.533), ….]



Projection: latitude and longitude

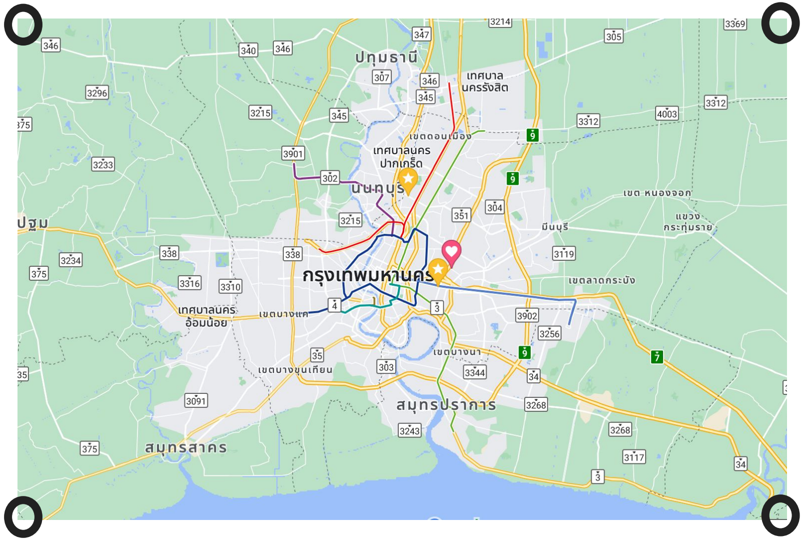


EPSG:3857

Projection: latitude and longitude

(14.03116, 100.23174)

(14.03116, 100.91143)



(13.545157, 100.23174)

(13.545157, 100.91143)

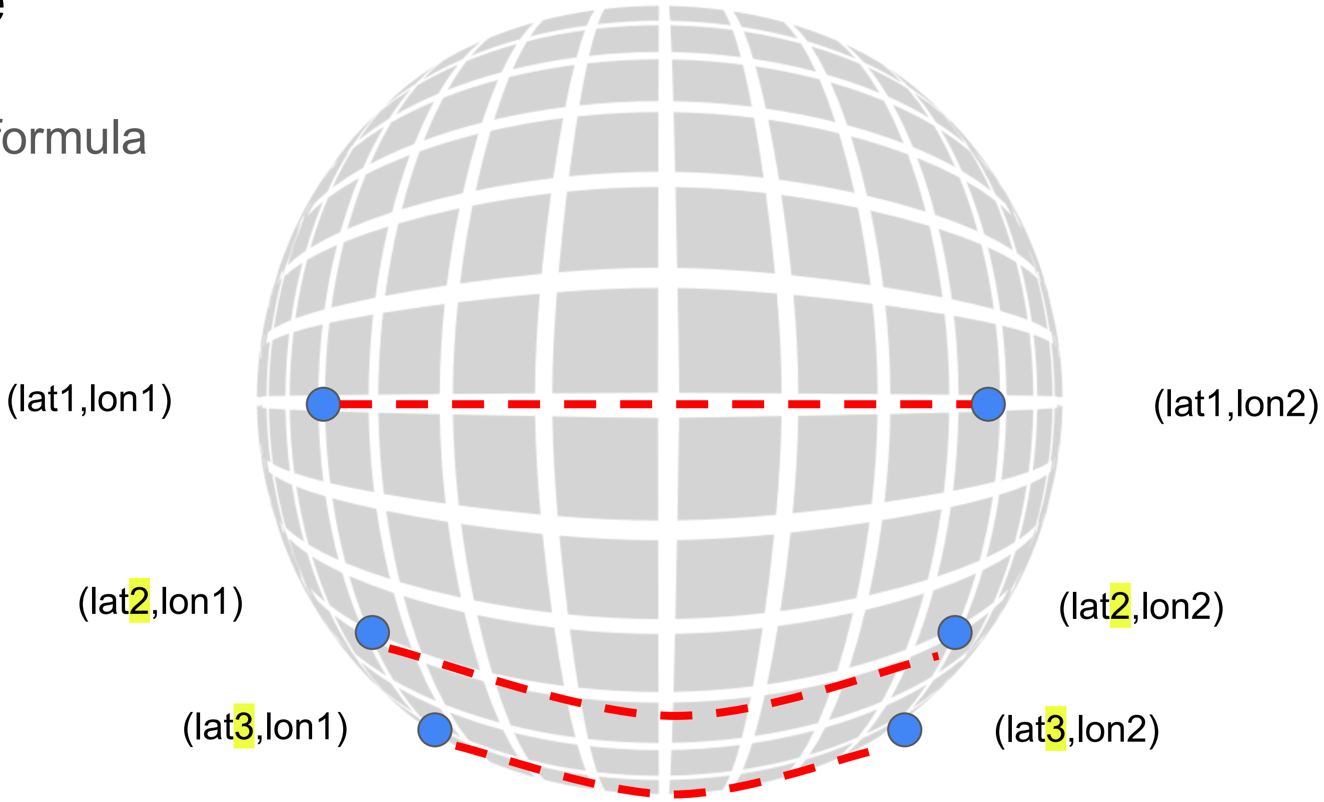
lat

lon

EPSG:3857

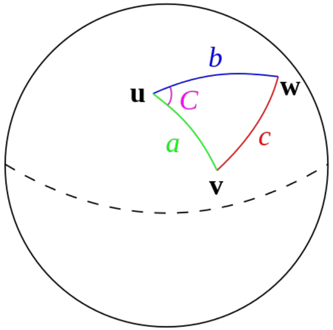
Distance

Haversine formula



Distance

Haversine formula



Ref: https://en.wikipedia.org/wiki/Haversine_formula

Let the central angle θ between any two points on a sphere be:

$$\theta = \frac{d}{r}$$

where:

- d is the distance between the two points along a great circle of the sphere (see spherical distance),
- r is the radius of the sphere.

The *haversine formula* allows the haversine of θ (that is, $\operatorname{hav}(\theta)$) to be computed directly from the latitude (represented by φ) and longitude (represented by λ) of the two points:

$$\operatorname{hav}(\theta) = \operatorname{hav}(\varphi_2 - \varphi_1) + \cos(\varphi_1)\cos(\varphi_2)\operatorname{hav}(\lambda_2 - \lambda_1)$$

or, to avoid using cosines which cause resolution degradation at small angles[*dubious – discuss*]:

$$\operatorname{hav}(\theta) = \operatorname{hav}(\varphi_2 - \varphi_1) + (1 - \operatorname{hav}(\varphi_1 - \varphi_2) - \operatorname{hav}(\varphi_1 + \varphi_2)) \cdot \operatorname{hav}(\lambda_2 - \lambda_1)$$

where

- φ_1, φ_2 are the latitude of point 1 and latitude of point 2,
- λ_1, λ_2 are the longitude of point 1 and longitude of point 2.

Finally, the haversine function $\operatorname{hav}(\theta)$, applied above to both the central angle θ and the differences in latitude and longitude, is

$$\operatorname{hav}(\theta) = \sin^2\left(\frac{\theta}{2}\right) = \frac{1 - \cos(\theta)}{2}$$

The haversine function computes half a versine of the angle θ.

To solve for the distance d, apply the archaversine (inverse haversine) to $h = \operatorname{hav}(\theta)$ or use the arcsine (inverse sine) function:

$$d = r \operatorname{archav}(h) = 2r \arcsin\left(\sqrt{h}\right)$$

or more explicitly:

$$\begin{aligned} d &= 2r \arcsin\left(\sqrt{\operatorname{hav}(\varphi_2 - \varphi_1) + (1 - \operatorname{hav}(\varphi_1 - \varphi_2) - \operatorname{hav}(\varphi_1 + \varphi_2)) \cdot \operatorname{hav}(\lambda_2 - \lambda_1)}\right) \\ &= 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \left(1 - \sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) - \sin^2\left(\frac{\varphi_2 + \varphi_1}{2}\right)\right) \cdot \sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right) \text{[9]} \\ &= 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\varphi_2 - \varphi_1}{2}\right) + \cos\varphi_1 \cdot \cos\varphi_2 \cdot \sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right). \end{aligned}$$

Distance

Haversine

in Python

```
def distance_km(location1, location2):
    lat1, lon1 = location1
    lat2, lon2 = location2
    radius = 6371 # km

    dlat = math.radians(lat2-lat1)
    dlon = math.radians(lon2-lon1)
    a = math.sin(dlat/2) * math.sin(dlat/2) + math.cos(math.radians(lat1)) \
        * math.cos(math.radians(lat2)) * math.sin(dlon/2) * math.sin(dlon/2)
    c = 2 * math.atan2(math.sqrt(a), math.sqrt(1-a))
    dist = radius * c

    return dist
```

```
distance_km((13.6666,100.4289),(13.7500,100.5318))
>> 14.47642361652062
```

Distance

Estimated Distance in Thailand

(Error about 15 m)

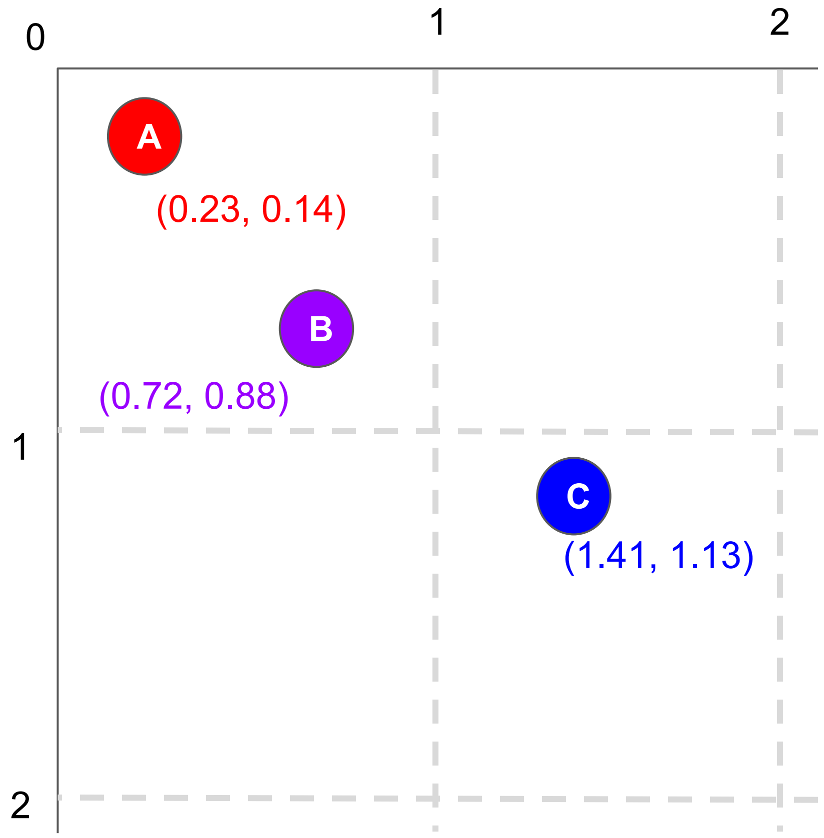
```
#error in Thailand about 15 m

def th_dist_km(lat1, lon1, lat2, lon2):
    return 110*((lat1-lat2)**2 + (lon1-lon2)**2)**0.5
```

```
th_dist_km((13.6666,100.4289),(13.7500,100.5318))
>> 14.56990174984072
```

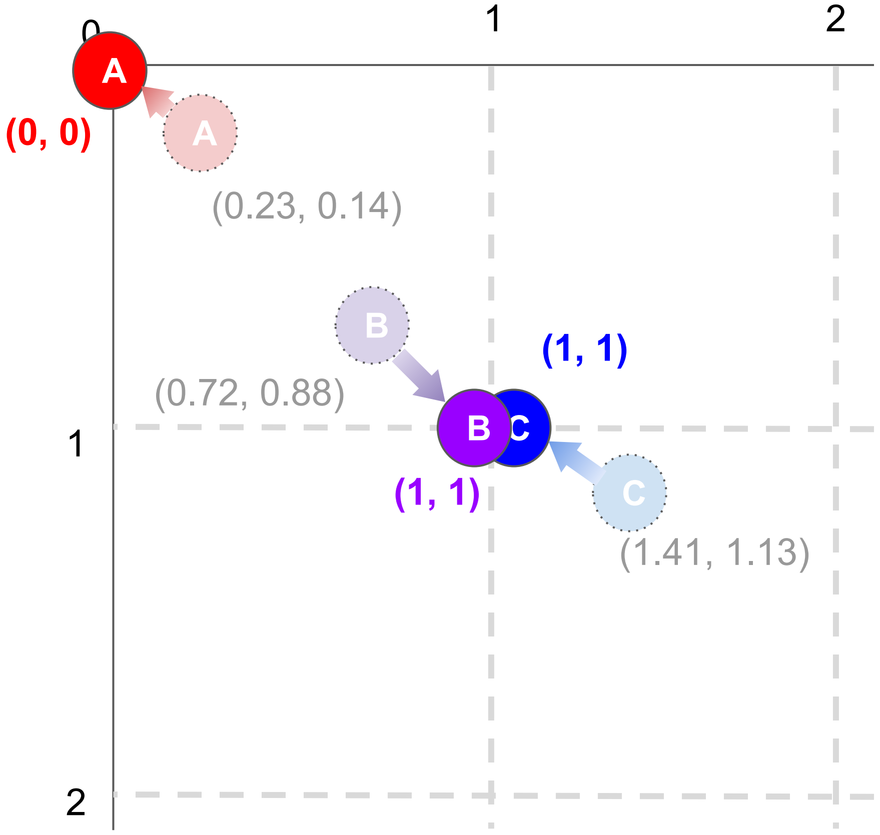
Grid

round() method



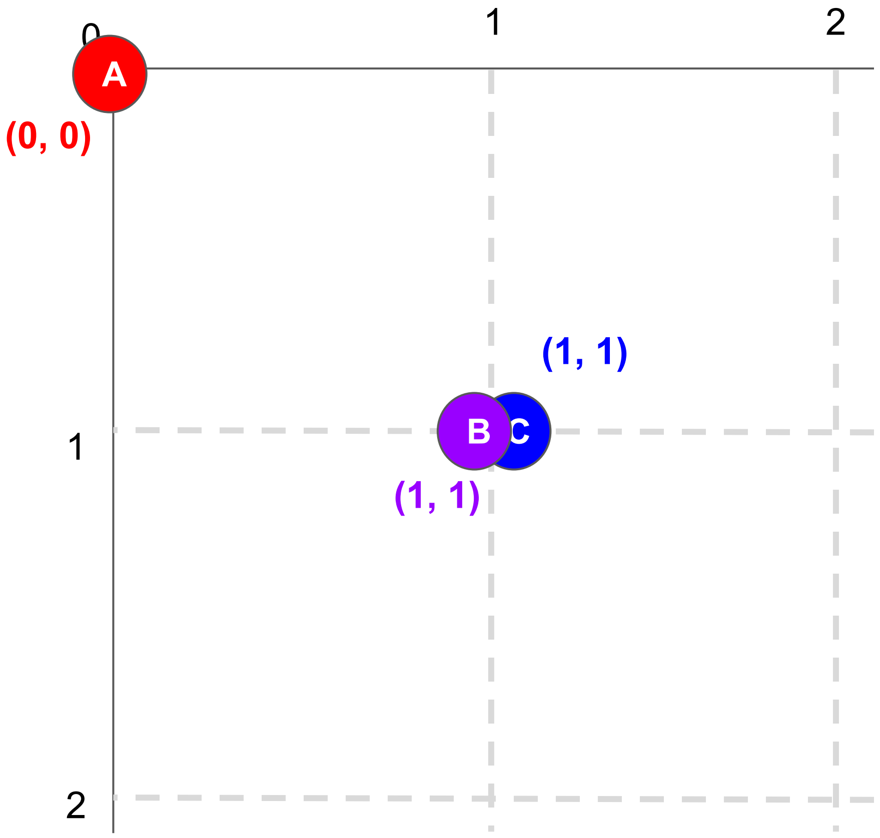
Grid

round() method



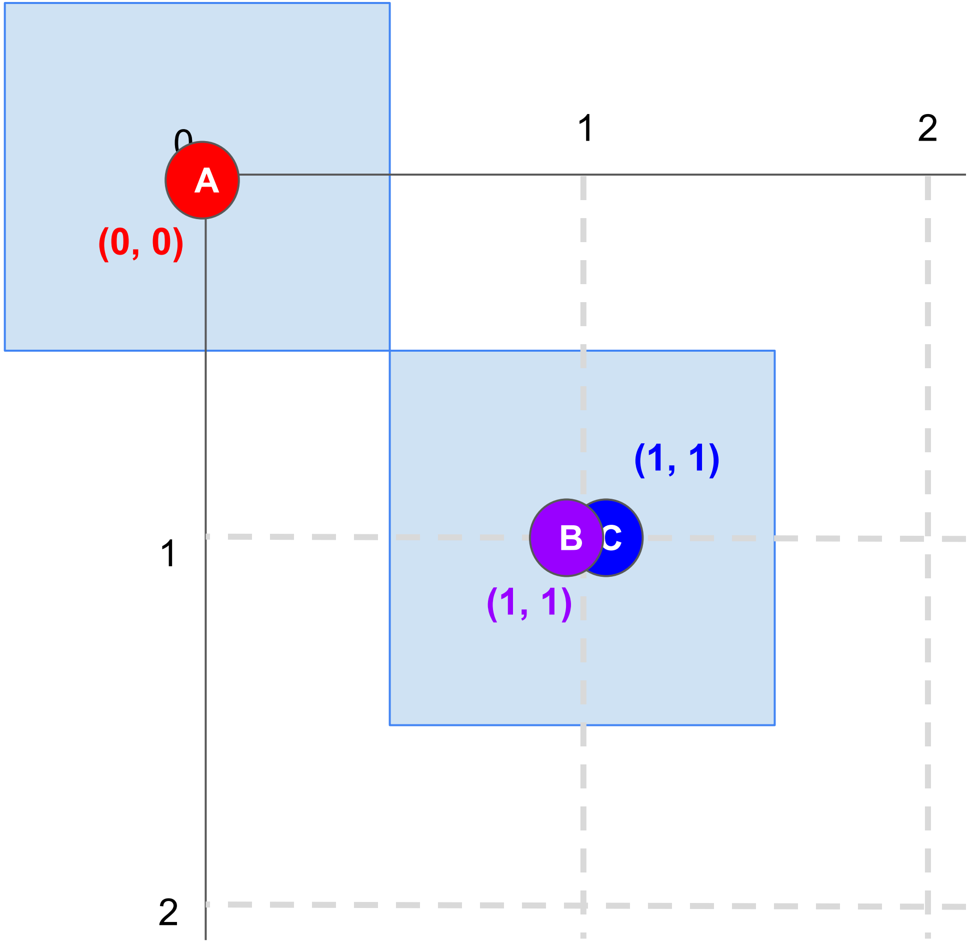
Grid

round() method



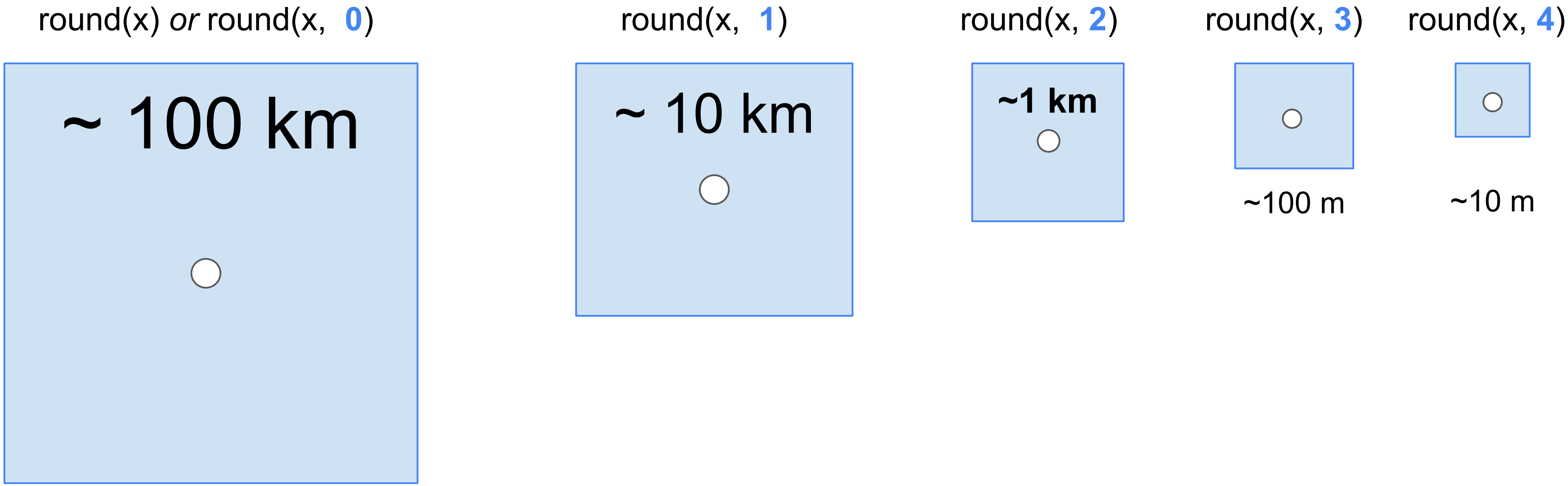
Grid

round() method



Grid

Grid Size



Bus Quality Assessment

R. Chawuthai, A. Sumalee, T. Threepak:
"GPS Data Analytics for the Assessment of
Public City Bus Transportation Service Quality in Bangkok".
In: Sustainability, MDPI, 15(7), 5618 (2023)

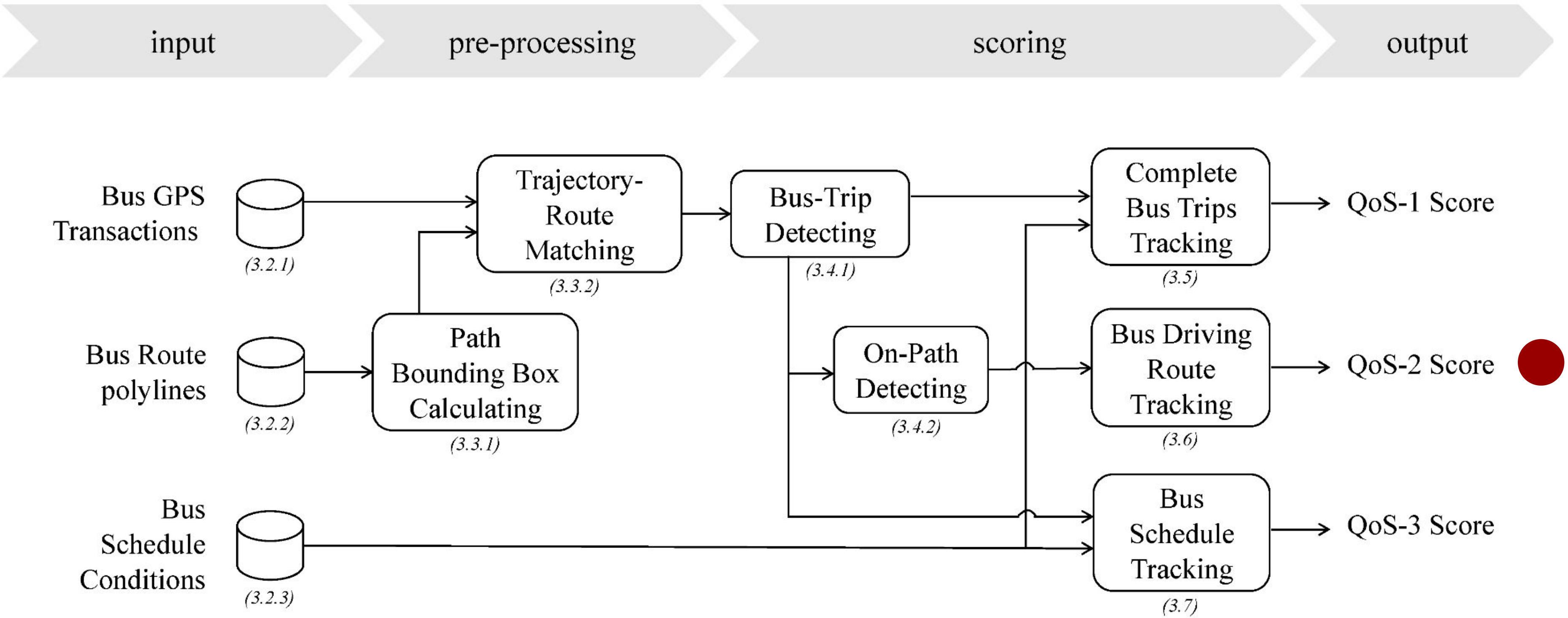
Abstract

Evaluation of the **quality of service (QoS) of public city buses** is generally performed using surveys that assess attributes such as accessibility, availability, comfort, convenience, reliabilities, safety, security, etc. Each survey attribute is assessed from the subjective viewpoint of the service users. This is reliable and straightforward because the consumer is the one who accesses the bus service. However, in addition to summarizing personal feedback from humans, using data analytics has become another useful method for assessing the QoS of bus transportation. This work aims to use global positioning system (GPS) data to measure the **reliability**, **accessibility**, and **availability** of bus transportation services. There are three QoS scoring functions for tracking complete trips, on-path driving, and on-schedule operation. In the analytical process, GPS coordinates rounding is adopted and applied for detecting trips on each route path. After assessing the three QoS scores, it has been found that most bus routes have good operations with high scores, while some bus routes show room for improvement. Future work could use our data to create recommendations for policy makers in terms of how to improve a city's smart mobility.

Objectives

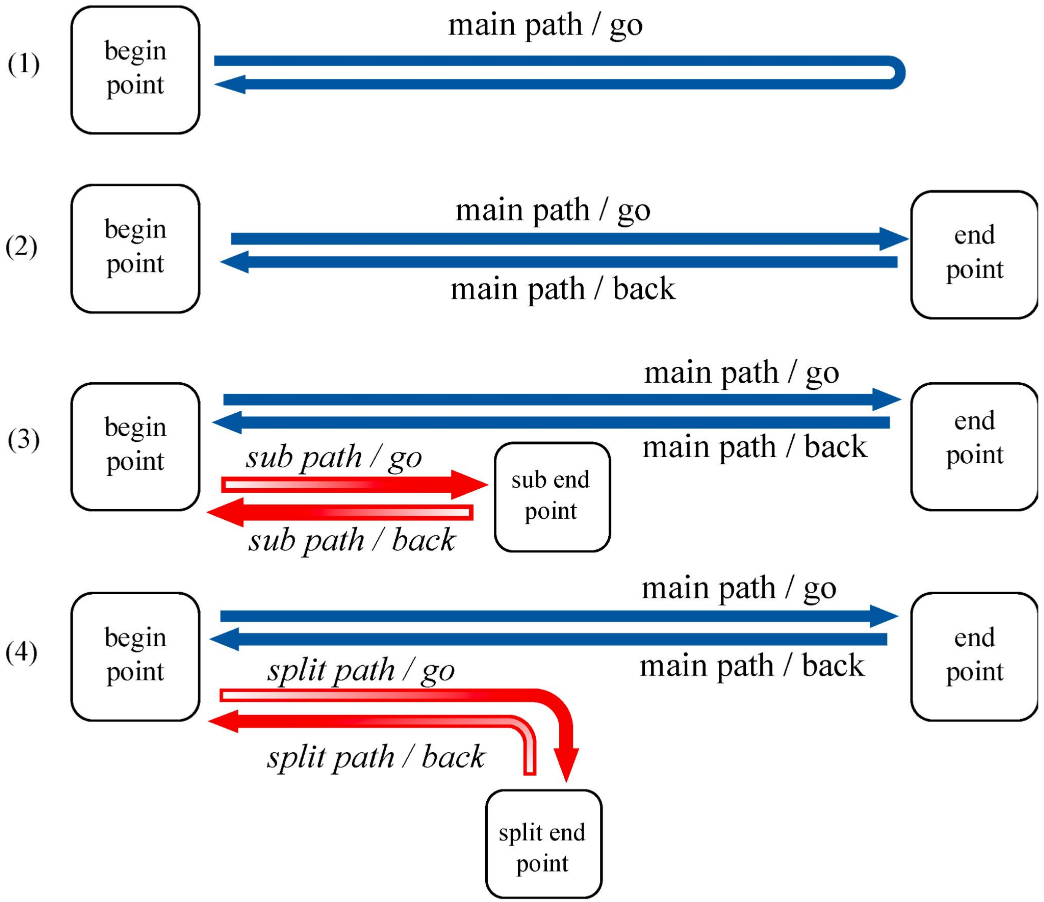
- To track complete bus trips based on commitments.
- **To track on-path driving following operated routes.**
- To track on-schedule operation according to schedule conditions.

Process



input
pre-processing
scoring
output
Bus GPS Transactions
(3.2.1)
Bus Route polylines
(3.2.2)
Bus Schedule Conditions
(3.2.3)
Trajectory-Route Matching
(3.3.2)
Path Bounding Box Calculating
(3.3.1)
Bus-Trip Detecting
(3.4.1)
On-Path Detecting
(3.4.2)
Complete Bus Trips Tracking
(3.5)
Bus Driving Route Tracking
(3.6)
Bus Schedule Tracking
(3.7)
QoS-1 Score
QoS-2 Score
QoS-3 Score

Behaviors of bus route paths in Thailand

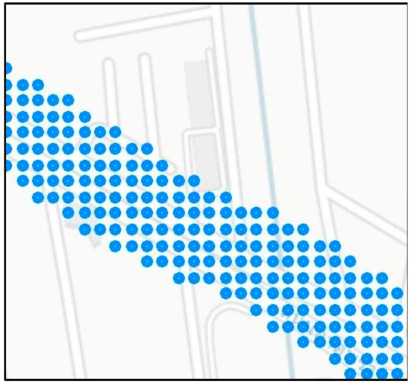


Bus Route Polyline

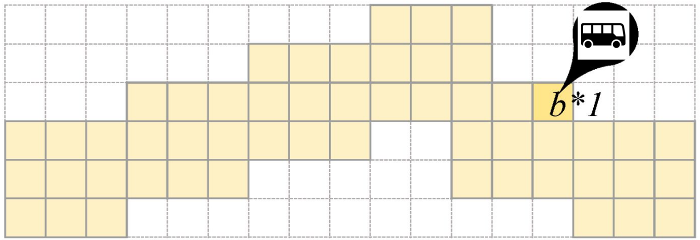
route	path_id	path_type	direction	begin_point	end_point	polyline
R7234	R7234.00	Main	go	(13.81196, 100.54976)	(13.59013, 100.59738)	[(13.81196, 100.54976), (13.81106, 100.54943),...
R7234	R7234.01	Split	go	(13.76977, 100.64184)	(13.60081, 100.74983)	[(13.76977, 100.64184), (13.76865, 100.64196),...
R7234	R7234.02	Split	back	(13.60081, 100.74983)	(13.76977, 100.64184)	[(13.60081, 100.74983), (13.60068, 100.74984),...
R7234	R7234.03	Sub	go	(13.76977, 100.64184)	(13.59004, 100.59742)	[(13.76977, 100.64184), (13.76946, 100.64187),...
R7234	R7234.04	Sub	back	(13.59004, 100.59742)	(13.76977, 100.64184)	[(13.59004, 100.59742), (13.59013, 100.59738),...
R8190	R8190.00	Main	go	(13.74004, 100.49846)	(13.82723, 100.73943)	[(13.74004, 100.49846), (13.74012, 100.49822),...
R8190	R8190.01	Main	back	(13.82723, 100.73943)	(13.74004, 100.49846)	[(13.82723, 100.73943), (13.82581, 100.74775),...
...	...	...	...	...	...	...

Polyline to Grids

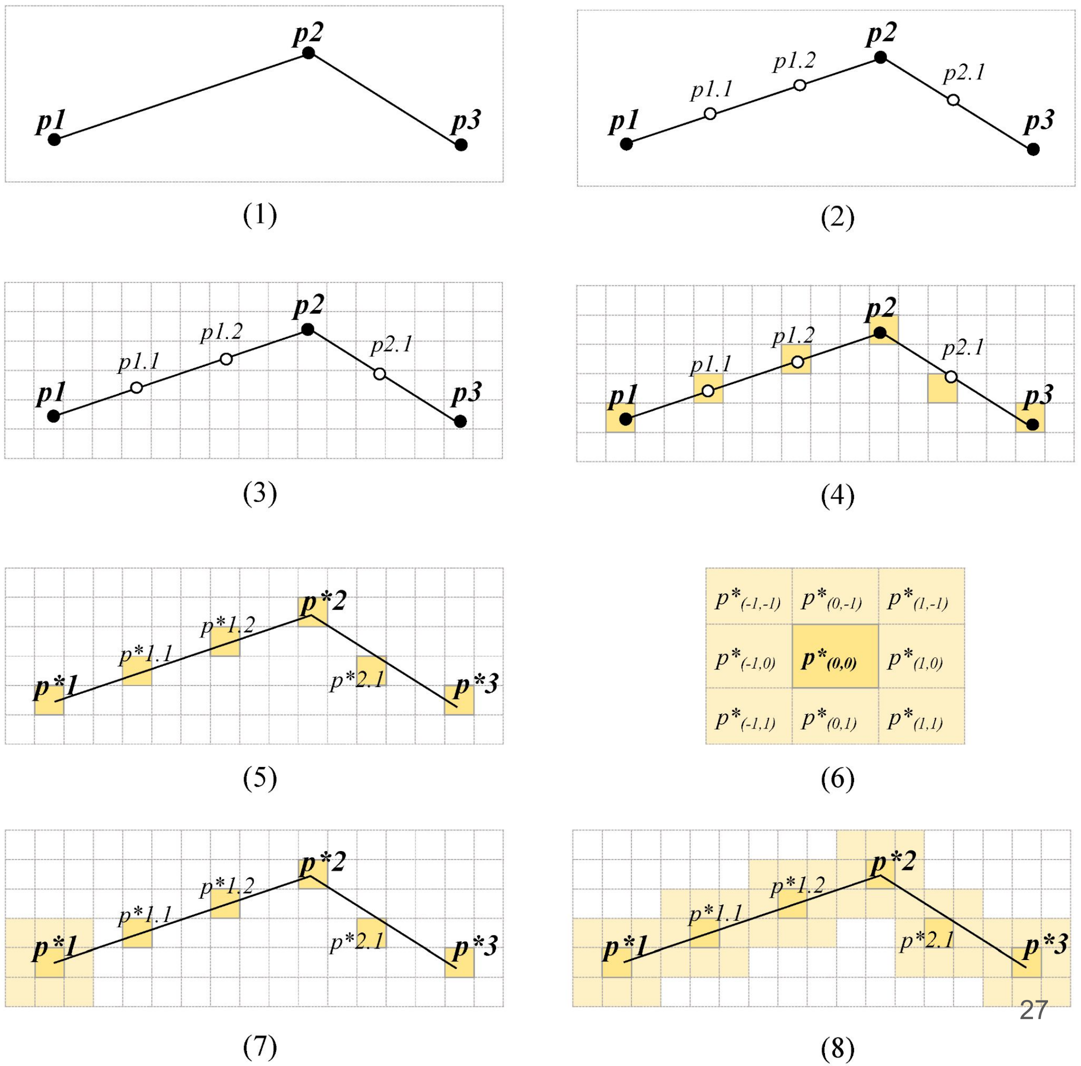
(1)



(2)



Polyline to Grids



GPS Data

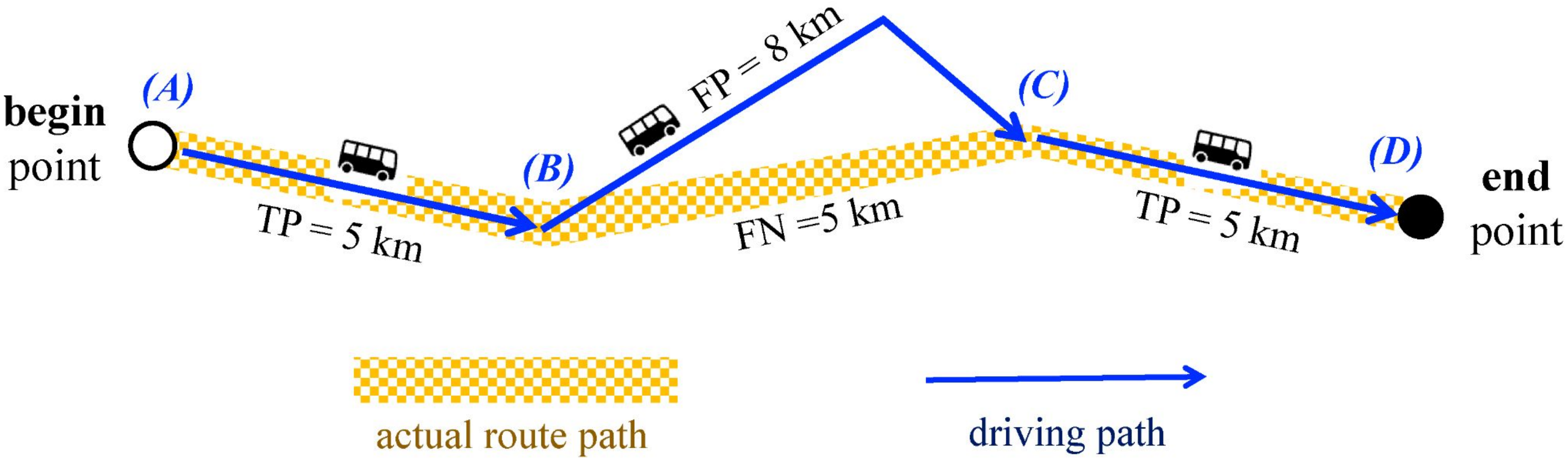
Table 2. Example GPS data of a bus on route R7234. In this table, the bid is a bus identifier, route is a route number, ts is a timestamp, lat is a latitude, lon is a longitude, and speed is a speed in kilometers per an hour.

Bid	Route	Ts	Lat	Lon	Speed
8ead83c5	R7234	2021-10-20 09:39:21	13.729222	100.641610	33
8ead83c5	R7234	2021-10-20 09:40:36	13.721500	100.642138	31
8ead83c5	R7234	2021-10-20 09:41:36	13.713388	100.643667	63
8ead83c5	R7234	2021-10-20 09:42:21	13.709083	100.644722	57
8ead83c5	R7234	2021-10-20 09:42:36	13.706860	100.645250	51
...	...	...		...	...

Bus Trip Detection

$$Jaccard = \frac{TP}{TP + FP + FN}$$

- True-positive (TP): the distance of a bus driving on a route path.
- False-positive (FP): the distance of a bus driving outside of a route path.
- False-negative (FN): the distance of a route path without a bus driving on it.



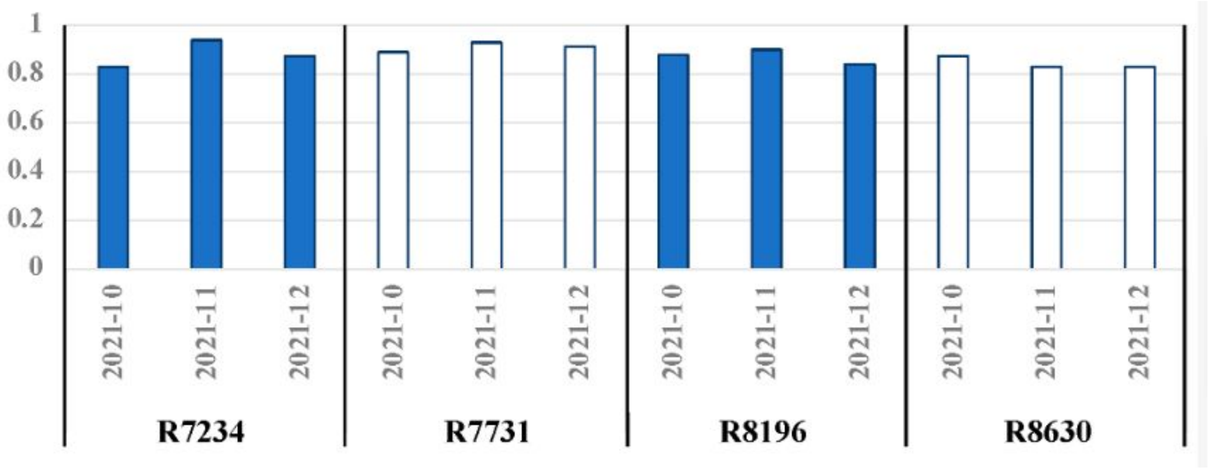
Result

Table 8. Daily QoS scores of the route R8155 in the 4th quarter of 2021.

Route	Date	QoS_1	QoS_2	QoS_3
R7234	2021-10-01	0.83	0.85	0.72
R7234	2021-10-02	0.78	0.73	0.76
R7234	2021-10-03	0.77	0.80	0.83
R7234	2021-10-04	0.83	0.86	0.68
R7234	2021-10-05	0.81	0.74	0.73
R7234	2021-10-06	0.92	0.75	0.82
R7234	2021-10-07	0.84	0.87	0.68
R7234	2021-10-08	0.83	0.76	0.80
	…	…	…	…
R7234	2021-12-28	0.77	0.91	0.81
R7234	2021-12-29	0.75	0.83	0.67
R7234	2021-12-30	0.76	0.74	0.82
R7234	2021-12-31	0.64	0.82	0.83

Result

QoS-2 Score



Next Stop Prediction

R. Chawuthai, N. Chankaew, T. Threepak:
"A Hybrid Method for Predicting a Potential Next Rest Stop of Commercial Vehicles".
In Proceeding of International Symposia of Transport Simulation and the International Workshop on Traffic Data Collection and its Standardisation (ISTS and IWTDCS). Elsevier. (2018)

Abstract

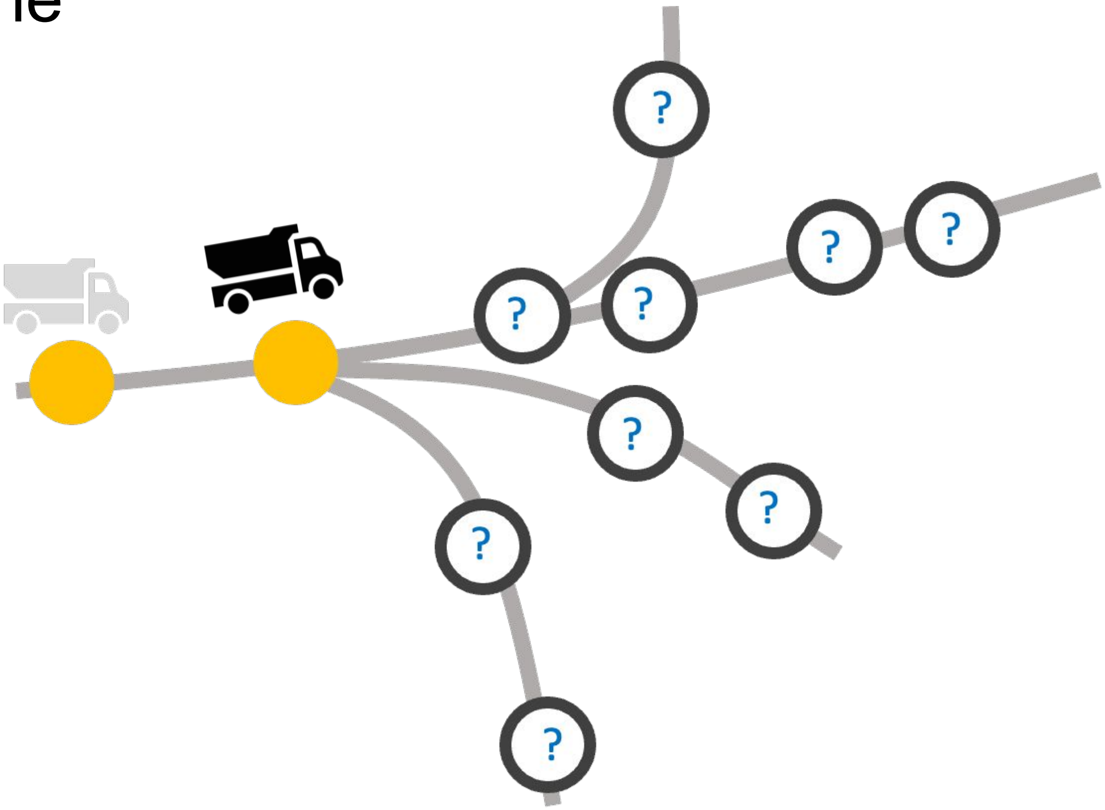
Long-distance trips such as freight and passenger transports over cities can create driver fatigue, so drivers prefer to get a rest for a while during their long-time driving. In Thailand, there are rest stops along main roads between cities, such as petrol stations, travel plazas, wayside parks, and scenic areas. In order to provide a better service to customers, the rest stops must have a good management, so the prediction of the number of potential vehicles in a period of time is primarily needed. One important task is **to predict the next rest stop of every car at a period of time.** Due to this requirement, this paper aims to introduce a prediction model for predicting the next rest stop of a vehicle by analyzing the global positioning system (GPS) tracking data of all commercial vehicles in Thailand. The proposed prediction model is a **hybrid model** that comprises of three scoring functions depended on the **frequent pattern** of connected rest stops, the **direction** of connected rest stops in a route, and the **popularity** of the rest stops. The experimental result shows that the proposed prediction model gives high accurate result in terms of the area under the receiver-operating-characteristic curve (AUC). This predicted result is also useful for a government department and rest stops' owner to improve transportation, road safety, and other service.

Data

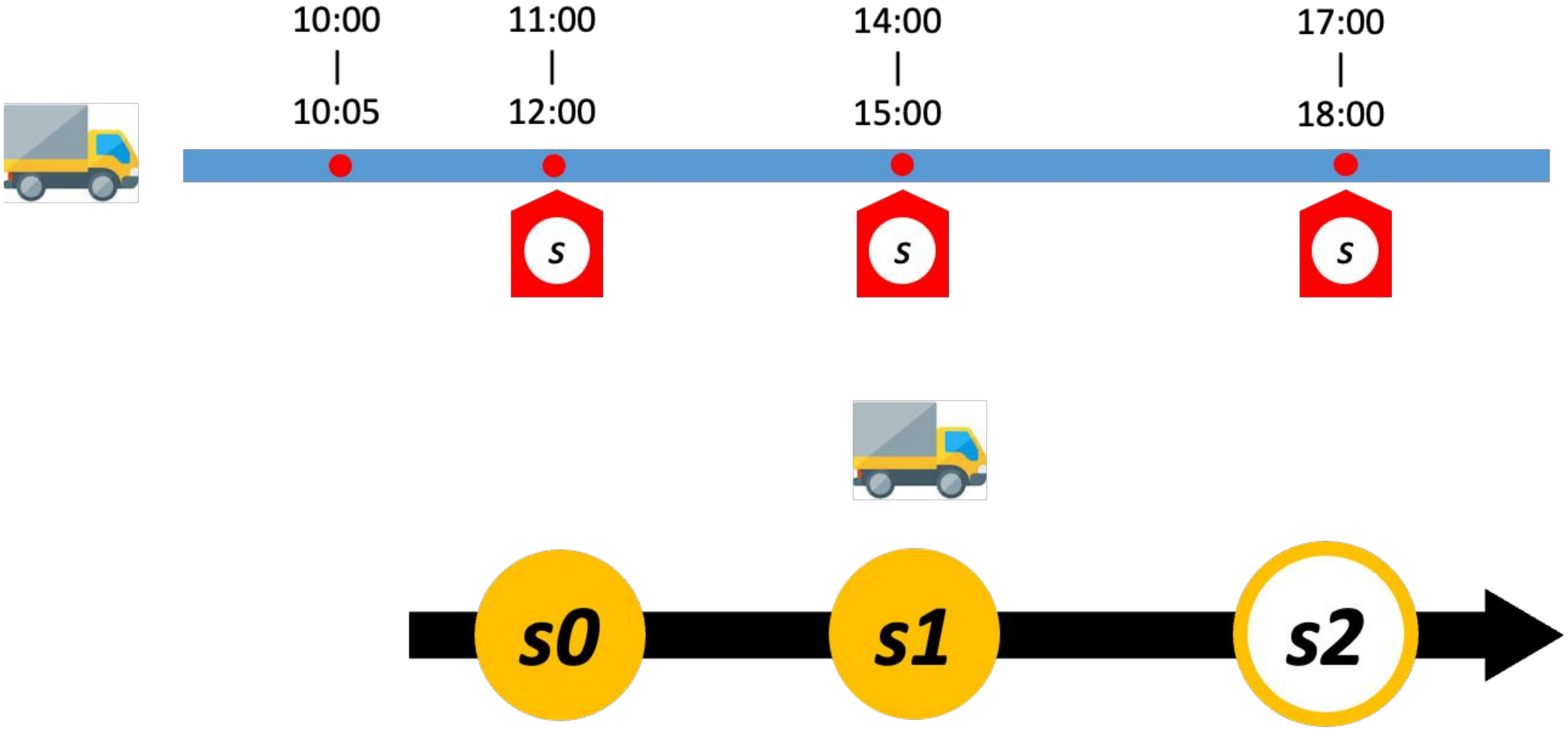
The Department of Land Transport in Thailand collects GPS tracking data from about 70,000 commercial vehicles in every minute

unit_id	ts	lat	lon	vtype
01228308814	2017-10-06 01:30	9.638598	99.12456	Coach
0103025803133	2017-10-06 01:30	7.867501	98.35564	Coach
011007532731	2017-10-06 01:30	7.903763	98.30505	Bus
011007953838	2017-10-06 01:30	8.558653	99.34329	Truck
011007958969	2017-10-06 01:30	8.120136	98.33496	Coach
01835026330329	2017-10-06 01:30	9.6489	99.12369	Bus
01228308814	2017-10-06 01:30	9.638598	99.12456	Truck

Where should be the Next Rest Stops



Data Pre-Processing

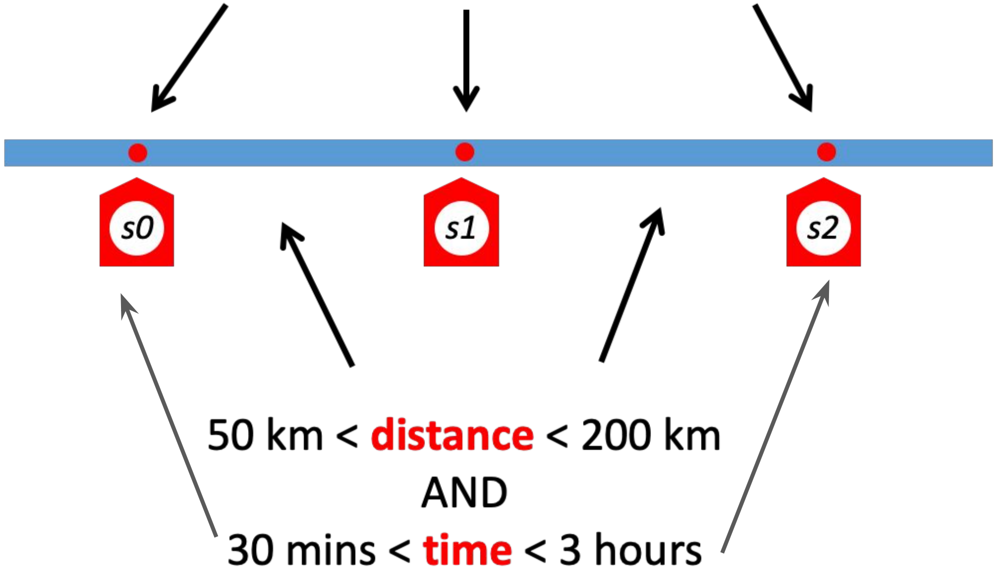


Data Pre-Processing

group **Locations** by two decimal places of lat&lon

AND

select one having vehicles stop more than 100 times in a day



50 km < **distance** < 200 km

AND

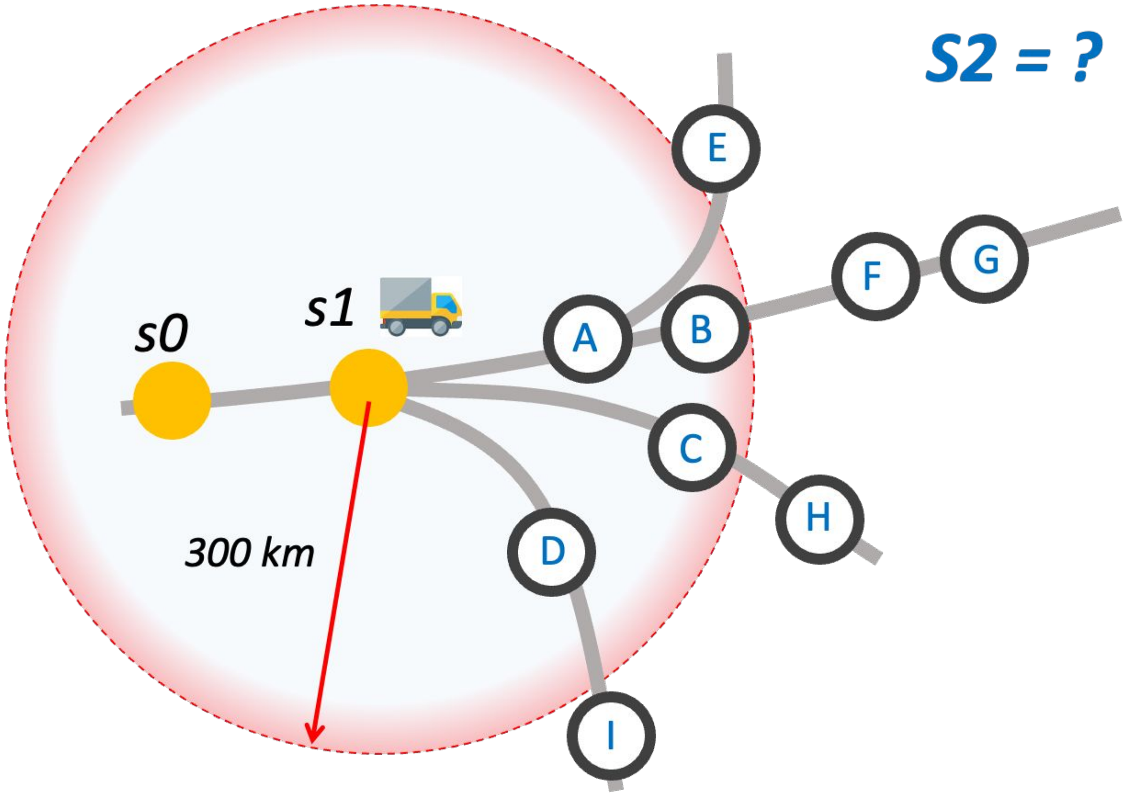
30 mins < **time** < 3 hours

Dataset

$$(s0 \rightarrow s1 \rightarrow s2,\ unit\_id,\ vtype,\ time1)$$

s0	s1	s2	unit_id	vtype	time1
(13.73, 100.76)	(13.12, 100.91)	(13.73, 100.76)	011524	Bus	2017-10-06 10:37
(13.05, 100.89)	(13.73, 100.76)	(13.05, 100.89)	011524	Bus	2017-10-11 12:47
(13.73, 100.76)	(13.10, 100.90)	(13.73, 100.76)	00211	Coach	2017-09-28 03:02
(13.73, 100.76)	(13.05, 100.89)	(13.74, 100.76)	00211	Coach	2017-10-07 09:18
(13.10, 100.90)	(13.12, 100.91)	(13.10, 100.90)	00211	Coach	2017-09-20 19:02
(13.10, 100.90)	(13.05, 100.90)	(13.10, 100.90)	00211	Coach	2017-09-25 21:28
(13.73, 100.76)	(13.75, 100.76)	(13.73, 100.76)	011524	Bus	2017-09-28 10:24
(13.71, 100.76)	(13.65, 100.78)	(13.71, 100.76)	06558511	Truck	2017-09-22 14:35
(13.12, 100.91)	(13.73, 100.76)	(13.05, 100.89)	00211	Coach	2017-09-28 09:15

Potential Next Rest Stops



Scoring Functions

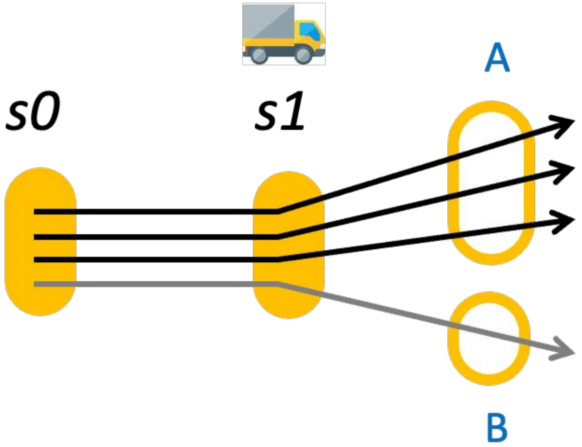
As our pre-analytics, some data characteristics was revealed and the following scoring functions become:

- **F**requent Pattern between rest stops → *SF*
- **D**irection of connected rest stops → *SD*
- **P**opularity of rest stops → *SP*

SF

Scoring Function based on the Frequent Pattern between Rest Stop

$$SF(s2|s0 \rightarrow s1) = \frac{NV(s0 \rightarrow s1 \rightarrow s2)}{NV(s0 \rightarrow s1)}$$



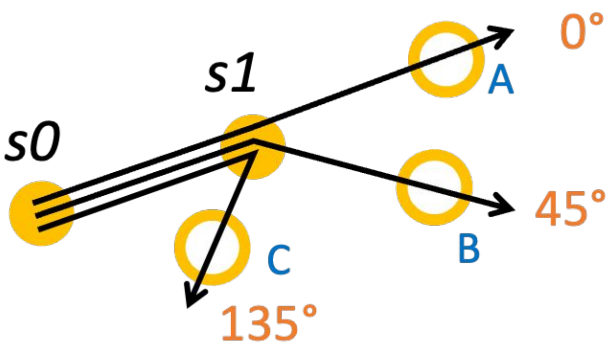
SF (A|s0→s1) = *3/4* = *0.75*

SF (B|s0→s1) = *1/4* = *0.25*

SD

Scoring Function based on the Directions of Connected Rest Stops

$$SD(s2|s0 \rightarrow s1) = \frac{\overrightarrow{s0\,s1} \cdot \overrightarrow{s1\,s2}}{\left\|\overrightarrow{s0\,s1}\right\| \cdot \left\|\overrightarrow{s1\,s2}\right\|}$$



SD(**A**|s0→s1) = *1.0*

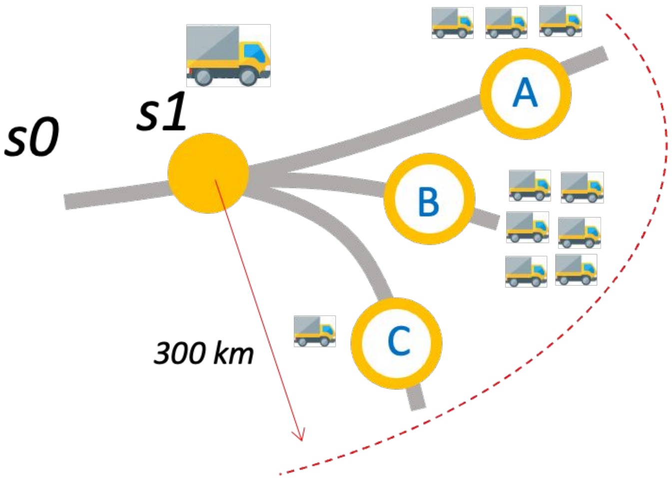
SD (**B**|s0→s1) = *0.7*

SD (**C**|s0→s1) = *-0.7*

SP

Scoring Function based on the Popularity of Rest Stops

$$SP(s2|s0 \rightarrow s1) = \frac{NVD(s2)}{NVD_R(s1)}$$



SP(A|s0→s1) = *3/10 = 0.3*

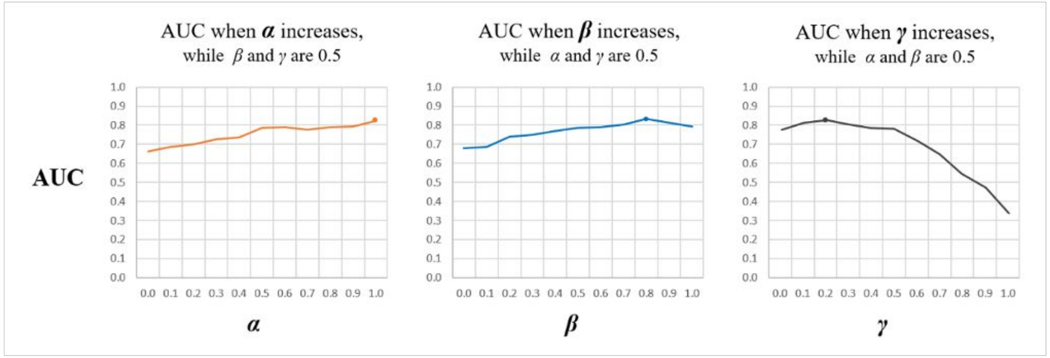
SP(B|s0→s1) = *6/10 = 0.6*

SP(C|s0→s1) = *1/10 = 0.1*

Hybrid Model

$$
\begin{aligned}
\boldsymbol{P}^{\boldsymbol{Next}}(s2|s0 \rightarrow s1) = \quad & \boldsymbol{\alpha} \cdot \boldsymbol{SF}(s2|s0 \rightarrow s1) \\
& + \boldsymbol{\beta} \cdot \boldsymbol{SD}(s2|s0 \rightarrow s1) \\
& + \boldsymbol{\gamma} \cdot \boldsymbol{SP}(s2|s0 \rightarrow s1)
\end{aligned}
$$

Result



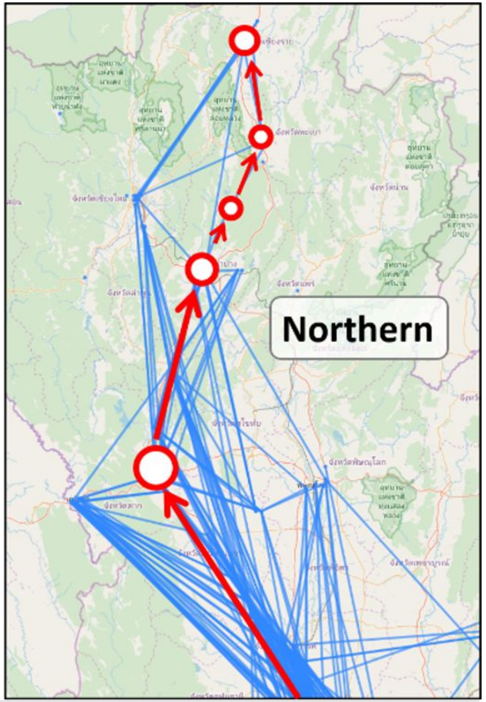
Hybrid Model	AUC
SF	0.623
SD	0.511
SP	0.337
SF + SD	0.782
SF + SP	0.681
SD + SP	0.662
SF + SD + SP	0.786

$$\boldsymbol{P}^{\boldsymbol{Next}}(s2|s0 \rightarrow s1) = 0.5 \cdot \boldsymbol{SF}(s2|s0 \rightarrow s1) + 0.4 \cdot \boldsymbol{SD}(s2|s0 \rightarrow s1) + 0.1 \cdot \boldsymbol{SP}(s2|s0 \rightarrow s1)$$

0.848

Result

Type of Vehicles	AUC		
	Daytime	Nighttime	Whole Day
Limousine Bus	0.834	0.837	0.836
Air-conditioned Bus	0.835	0.832	0.833
Bus	0.889	0.872	0.880
Coach	0.861	0.825	0.843
Pickup truck	0.841	0.846	0.844
Container truck	0.872	0.897	0.884
Liquid Truck	0.938	0.772	0.855
Hazardous Truck	0.871	0.840	0.856
Special-purpose truck	0.833	0.835	0.834
Pick-Up Truck with Labor-Saving Device	0.818	0.812	0.815
Average	0.859	0.837	0.848



Travel Time Prediction

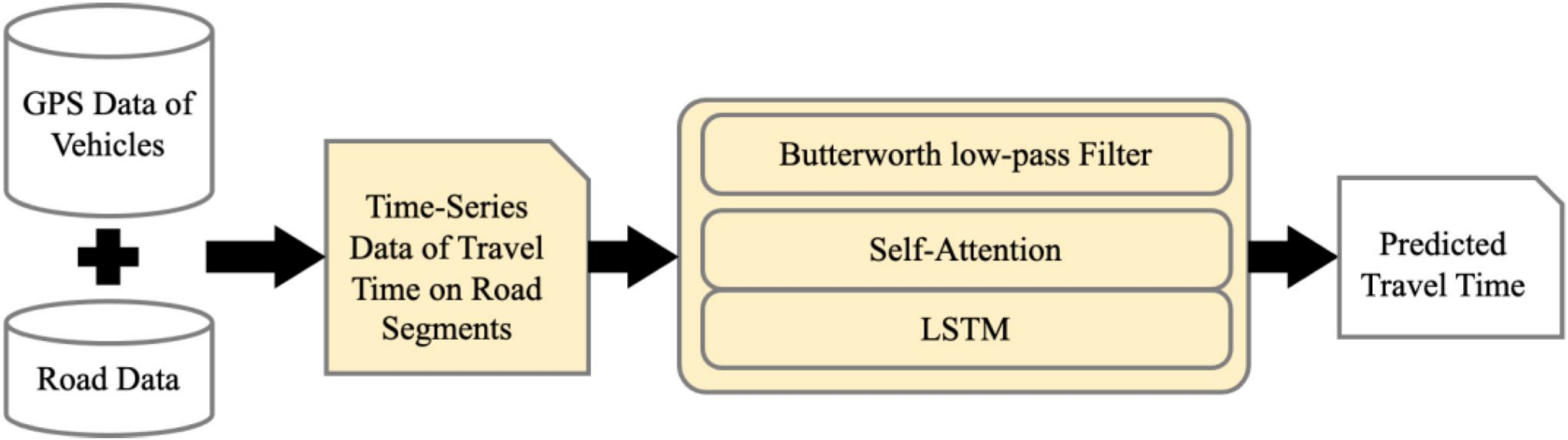
R. Chawuthai, N. Ainthong, S. Intarawart, N. Boonyanaet, A. Sumalee:
"Travel Time Prediction on Long-Distance Road Segments in Thailand"*.*
In: Applied Science, MDPI, 12(11), 5681 (2022)

Abstract

This study proposes a method by which to predict the **travel time of vehicles** on long-distance road segments in Thailand. We adopted the **Self-Attention Long Short-Term Memory (SA-LSTM) model with a Butterworth low-pass filter** to predict the travel time on each road segment using historical data from the Global Positioning System (GPS) tracking of trucks in Thailand. As a result, our prediction method gave a Mean Absolute Error **(MAE) of 12.15 min per 100 km**, whereas the MAE of the baseline was 27.12 min. As we can estimate the travel time of vehicles with a lower error, our method is an effective way to shape a data-driven smart city in terms of predictive mobility.

Process

Our overall approach. The time-series data, which are created from Global Positioning System (GPS) data and road data, were passed through a filter to be used as an input of the Self-Attention Long Short-Term Memory (LSTM). The predicted travel time was the output.

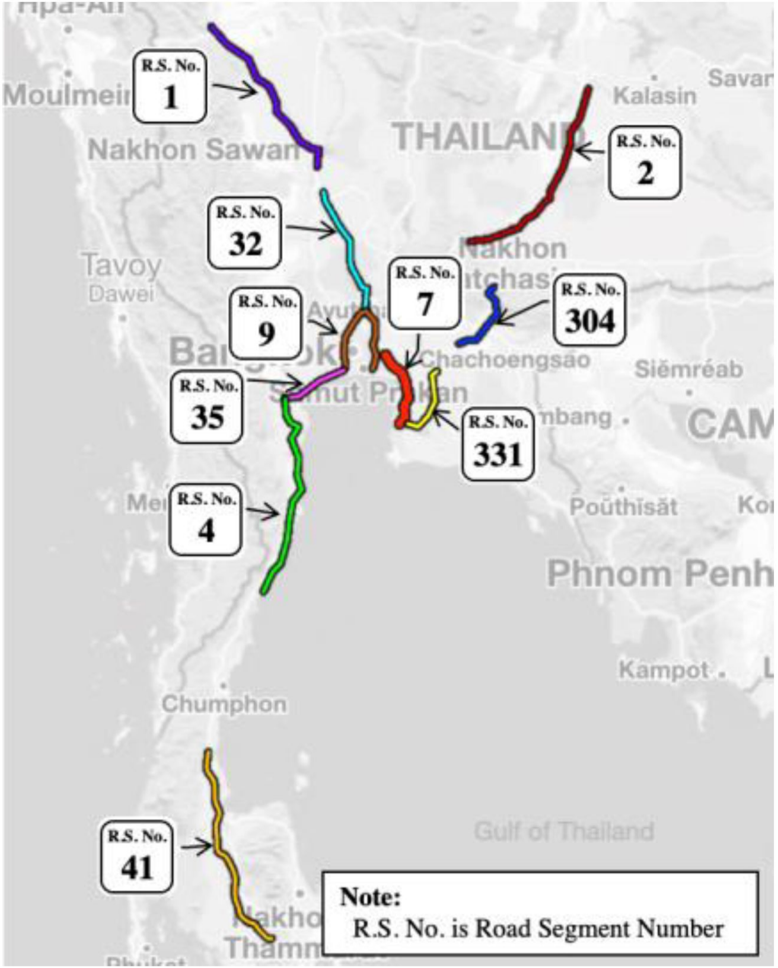


Data

time_stamp	vid	lat	lon
2019-03-01 00:00:35	0000000000	13.702763	100.581581
2019-03-01 00:00:32	0000000001	19.950153	99.236827
2019-03-01 00:00:39	0000000002	17.526147	100.260691
2019-03-01 00:00:31	0000000003	13.747755	100.761312
2019-03-01 00:00:59	0000000004	12.804558	101.136750
2019-03-01 00:00:58	0000000005	15.213062	103.091829
2019-03-01 00:00:45	0000000006	13.617246	100.653880
2019-03-01 00:00:30	0000000007	13.504666	101.038202
2019-03-01 00:00:43	0000000008	12.923107	101.060660
2019-03-01 00:00:46	0000000009	14.572481	100.786637

Selected Road Segments

Road Segment	Distance (km)	Date	Hour Duration	Travel Time (Seconds)
Highway No. 1	332	2019-02-14	00:00–00:59	18,421.14
Highway No. 1	332	2019-02-14	01:00–01:59	16,412.50
Highway No. 1	332	2019-02-14	02:00–02:59	15,809.55
Highway No. 1	332	2019-02-14	03:00–03:59	15,206.60
Highway No. 1	332	2019-02-14	04:00–04:59	16,552.00
Highway No. 1	332	2019-02-14	05:00–05:59	17,110.75
Highway No. 1	332	2019-02-14	06:00–06:59	14,216.83
Highway No. 1	332	2019-02-14	07:00–07:59	15,873.13
Highway No. 1	332	2019-02-14	08:00–08:59	15,494.25
Highway No. 1	332	2019-02-14	09:00–09:59	20,464.50
Highway No. 1	332	2019-02-14	10:00–10:59	14,278.20
Highway No. 1	332	2019-02-14	11:00–11:59	16,780.00
Highway No. 1	332	2019-02-14	12:00–12:59	14,343.60
...	...	...	...	...

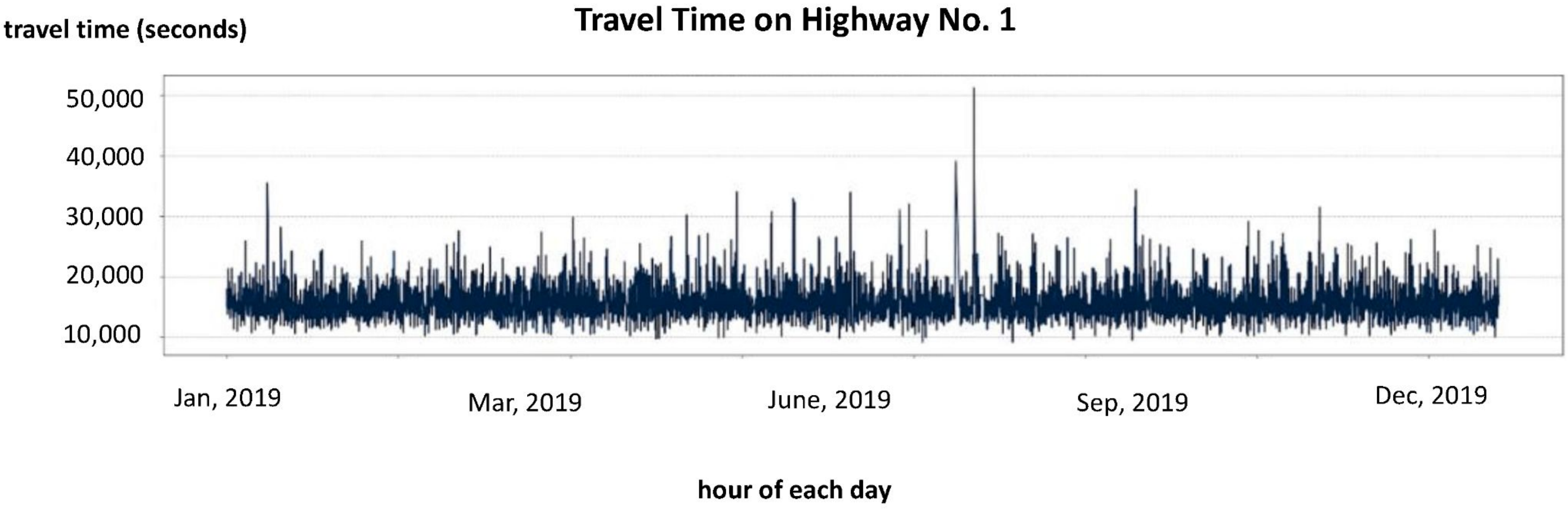


Selected Road Segments

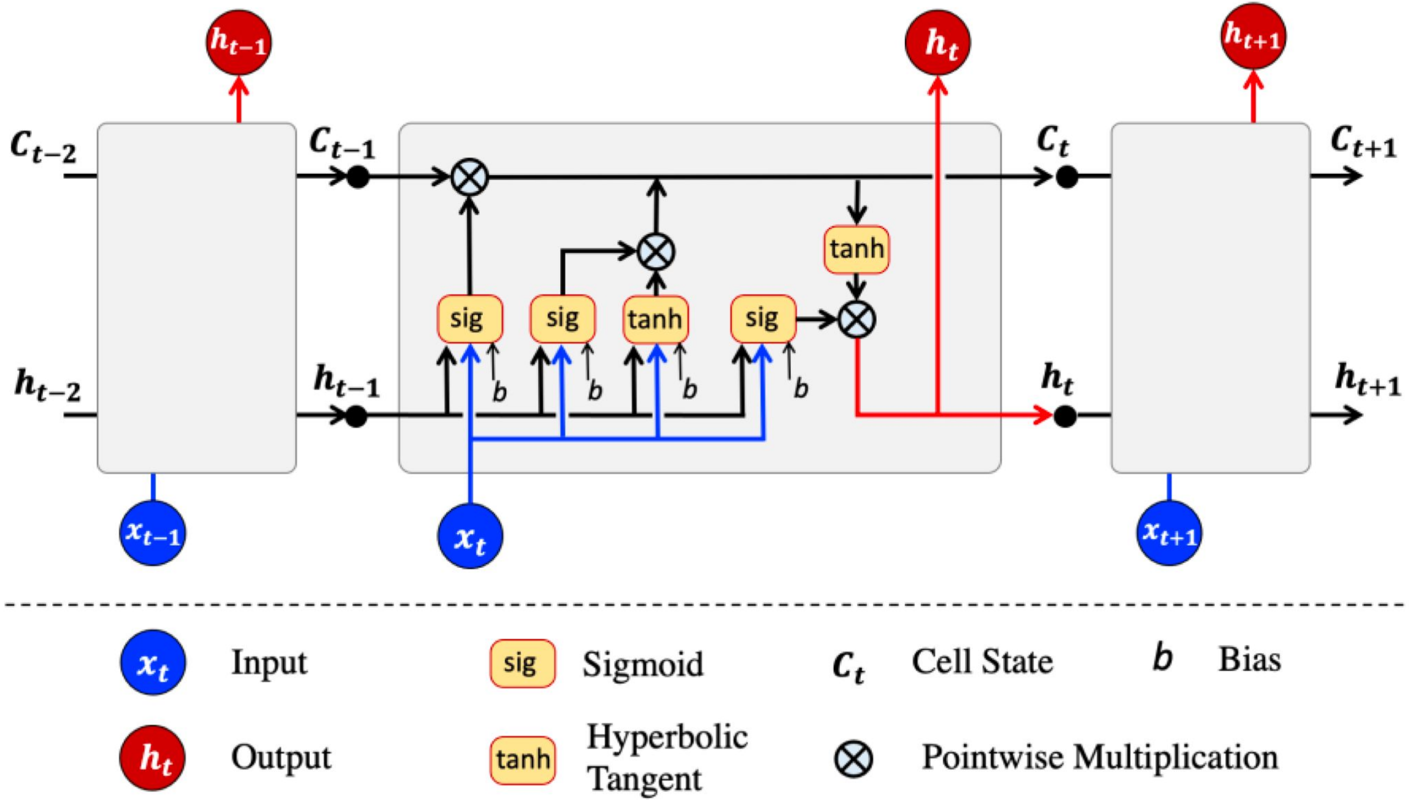
Road Segment	Distance (km)	Number of Trucks	Number of Trips	Average Travel Time (mins)
Highway No. 1	332	9079	71,416	259.96
Highway No. 2	232	8147	53,087	336.68
Highway No. 4	219	15,837	141,157	287.76
Highway No. 7	95	14,018	297,478	220.28
Highway No. 9	149	48,109	755,804	94.78
Highway No. 32	143	10,368	92,304	171.67
Highway No. 35	75	25,827	317,787	96.47
Highway No. 41	242	8676	72,072	293.36
Highway No. 304	104	9741	95,855	138.94
Highway No. 331	80	11,245	113,310	130.48
Summary	**167.1** (Average)	**81,068** (Unique Count)	**2,010,270** (Total)	**203.04** (Average)

Timeseries Data

Line chart of the average hourly travel times from Table 1, where the x-axis shows intervals of hours in chronological order, and the y-axis is average travel times of each hour in seconds.

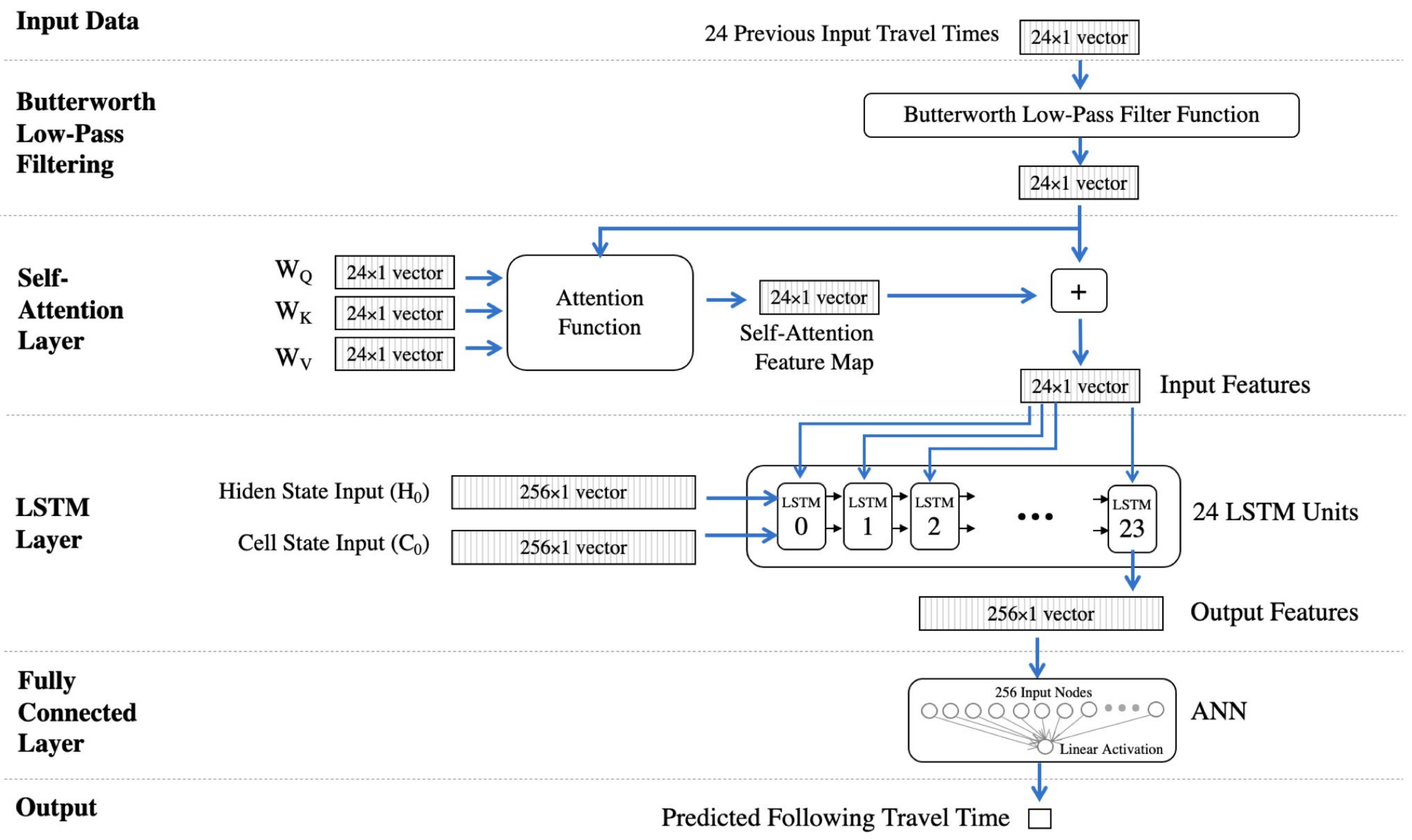


LSTM



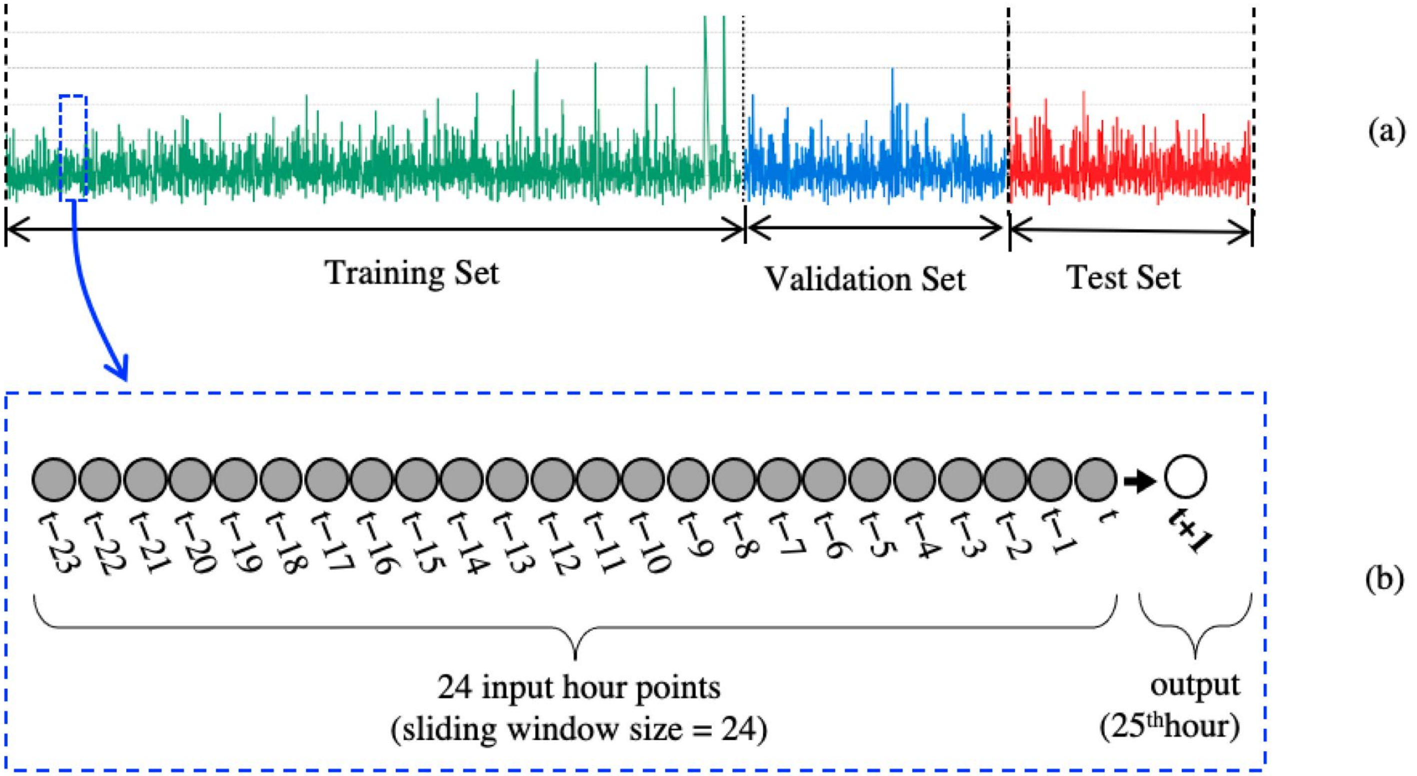
h_{t-1}
h_t
h_{t+1}
C_{t-2}
C_{t-1}
C_t
C_{t+1}
h_{t-2}
h_{t-1}
h_t
h_{t+1}
sig
sig
tanh
sig
tanh
b
b
b
b
x_{t-1}
x_t
x_{t+1}
x_t Input
h_t Output
sig Sigmoid
tanh Hyperbolic Tangent
C_t Cell State
b Bias
Pointwise Multiplication

Method



Input Data
24 Previous Input Travel Times
24×1 vector
Butterworth Low-Pass Filtering
Butterworth Low-Pass Filter Function
24×1 vector
Self-Attention Layer
W_Q
W_K
W_V
24×1 vector
24×1 vector
24×1 vector
Attention Function
24×1 vector
Self-Attention Feature Map
+
24×1 vector
Input Features
LSTM Layer
Hiden State Input (H_0)
256×1 vector
Cell State Input (C_0)
256×1 vector
LSTM 0
LSTM 1
LSTM 2
LSTM 23
24 LSTM Units
256×1 vector
Output Features
Fully Connected Layer
256 Input Nodes
Linear Activation
ANN
Output
Predicted Following Travel Time

Training



(a)
Training Set
Validation Set
Test Set
t−23
t−22
t−21
t−20
t−19
t−18
t−17
t−16
t−15
t−14
t−13
t−12
t−11
t−10
t−9
t−8
t−7
t−6
t−5
t−4
t−3
t−2
t−1
t
t+1
24 input hour points
(sliding window size = 24)
output
(25th hour)
(b)

Baseline ?

Table 3. Performance report of the time-series prediction techniques: Autoregressive Integrated Moving Average (ARIMA), Support Vector Regression (SVR), Long Short-Term Memory (LSTM), and Convolutional LSTM (ConvLSTM), evaluated by Mean Absolute Error (MAE).

Techniques	MAE (minutes) per 100 km
ARIMA	31.25
SVR	29.96
LSTM	**27.12**
ConvLSTM	28.32

Self-Attention ?

Table 4. Performance report of comparison of the LSTM and SA-LSTM.

Techniques	MAE (minutes) per 100 km
LSTM (only)	27.12
SA-LSTM (Self-Attention LSTM)	**26.89**

Filter

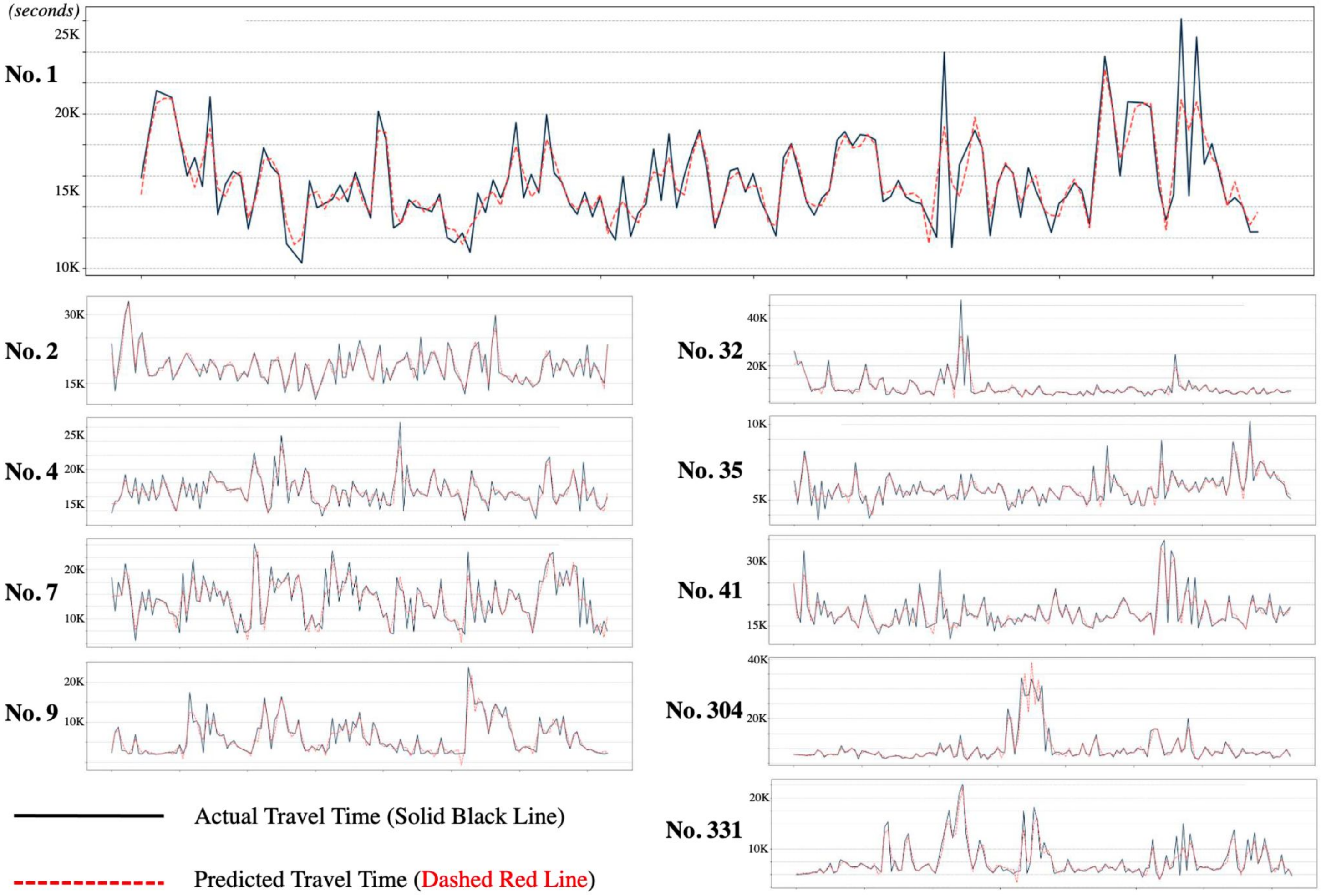
Table 5. Performance report of the comparison of SA-LSTM and SA-LSTM with a filter technique.

Techniques	MAE (minutes) per 100 km
SA-LSTM	26.89
SA-LSTM + Kalman filter	20.95
SA-LSTM + Savitzky–Golay filter	20.96
SA-LSTM + Butterworth low-pass filter	**12.15**

Result

Road Segment	Distance (km)	MAE (minutes) per 100 km		% Improvement
		Baseline	Our Approach	
Highway No. 1	332	9.19	4.04	56.08%
Highway No. 2	232	22.06	9.15	58.53%
Highway No. 4	219	13.70	6.01	56.15%
Highway No. 7	95	49.66	25.41	48.83%
Highway No. 9	149	29.89	12.88	56.92%
Highway No. 32	143	31.65	14.29	54.86%
Highway No. 35	75	17.35	8.23	52.57%
Highway No. 41	242	14.16	6.36	55.11%
Highway No. 304	104	19.47	8.16	58.07%
Highway No. 331	80	64.03	26.98	57.87%
Average		**27.12**	**12.15**	**55.20%**
Standard Deviation (STD)		**17.47**	**8.01**	**0.03**

Result


(seconds)
No. 1
25K
20K
15K
10K
No. 2
30K
15K
No. 4
25K
20K
15K
No. 7
20K
10K
No. 9
20K
10K
No. 32
40K
20K
No. 35
10K
5K
No. 41
30K
15K
No. 304
40K
20K
No. 331
20K
10K
Actual Travel Time (Solid Black Line)
Predicted Travel Time (Dashed Red Line)

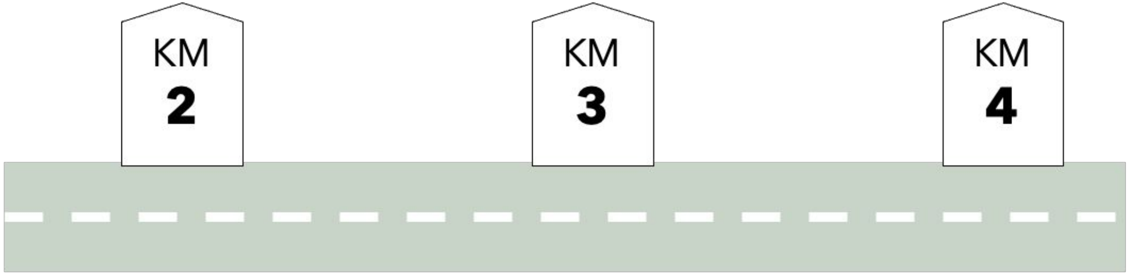
Traffic Speed Prediction

R. Chawuthai, K. Pruekwangkhao, T. Threepak:
*“**Spatial-Temporal Traffic Speed Prediction on Thailand Roads**”.*
In 2021 International Conference on Engineering,
Applied Sciences, and Technology (ICEAST). IEEE. (2021)

Abstract

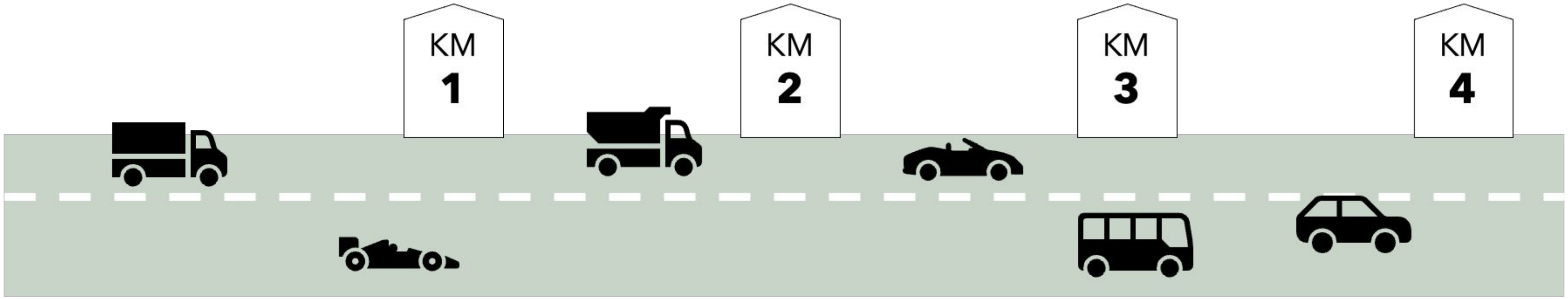
The ultimate goal of our research mission is to build a traffic model for travel time prediction in Thailand in order to improve the mobility domain of the smart city. To achieve our mission, this piece of research places important on a traffic speed prediction of any reference points at an incoming time that is one significant part of the travel time prediction. In this study, we employ a **linear model for predicting traffic speed of some kilometer stones in the next several minutes.** Our prediction model performs less root-mean-squared-error score under some spatial-temporal conditions. In addition, the temporal-lagged associations among kilometer stones, which were extracted during the feature selection process, are observed as a traffic-dependent network on roads for analyzing the traffic congestion spreads in the future.

KM Stone

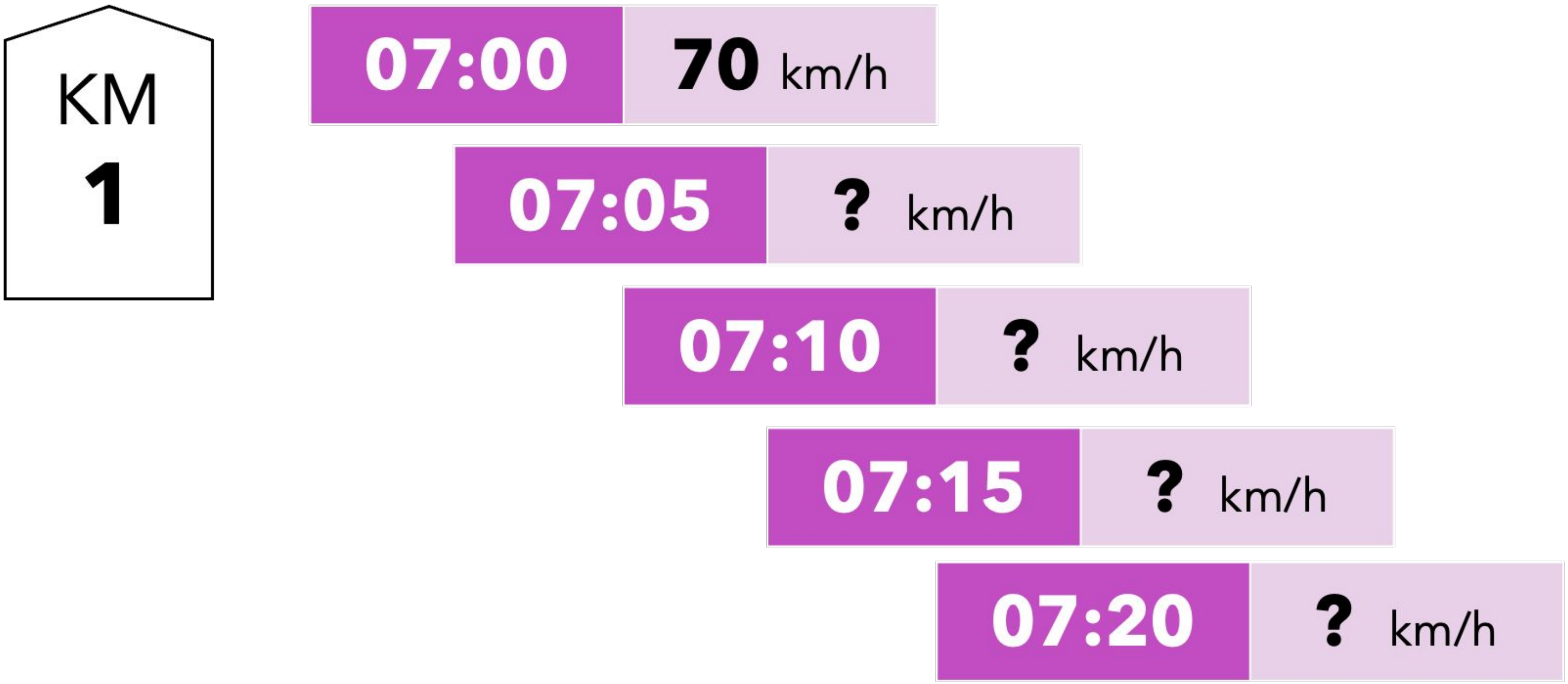


KM Stones

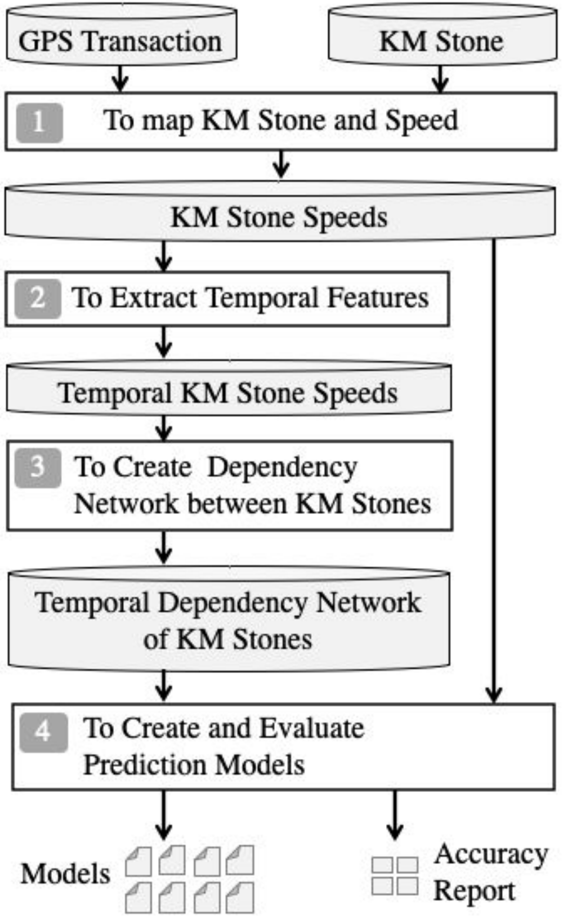
Time	KM # 1	KM # 2	KM # 3	KM # 4
07:00	75 km/h	70 km/h	75 km/h	66 km/h
07:01	60 km/h	55 km/h	46 km/h	86 km/h
07:02	65 km/h	55 km/h	76 km/h	45 km/h



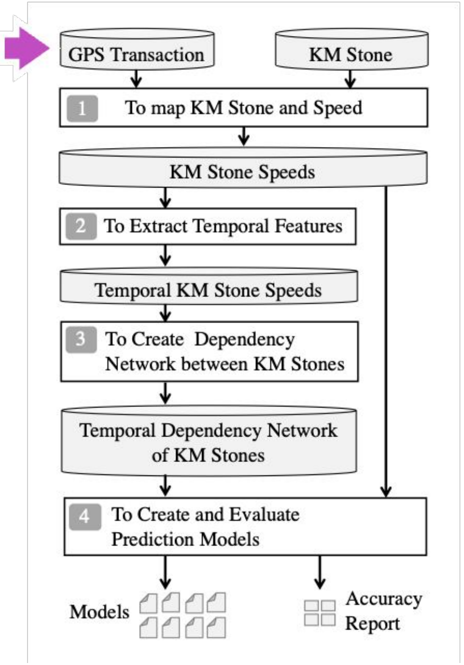
To Predict



Steps

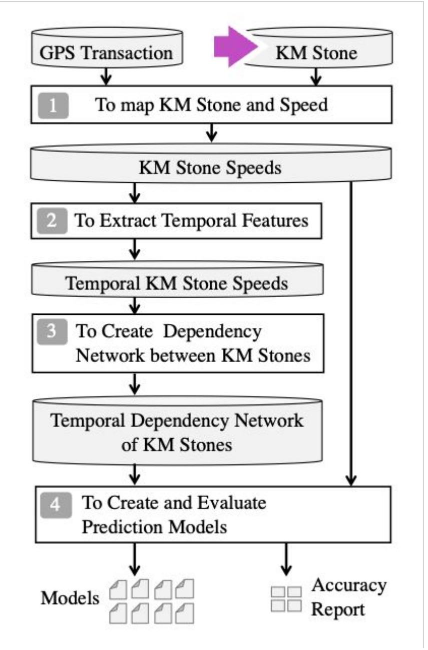

GPS Transaction
KM Stone
1 To map KM Stone and Speed
KM Stone Speeds
2 To Extract Temporal Features
Temporal KM Stone Speeds
3 To Create Dependency Network between KM Stones
Temporal Dependency Network of KM Stones
4 To Create and Evaluate Prediction Models
Models
Accuracy Report

GPS Transaction



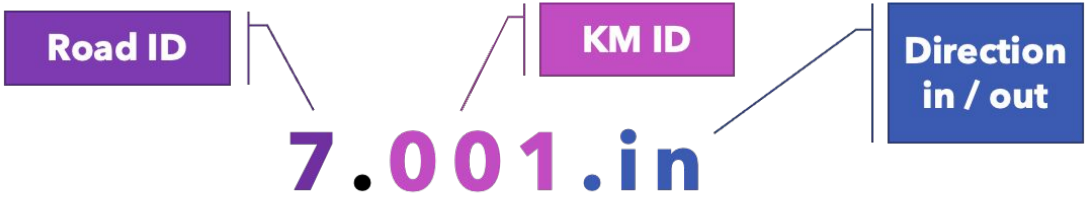
bid	ts	lat	lon	speed
X0014	2020-11-15 09:04:14	14.956732	100.376327	51
X0014	2020-11-15 09:05:15	14.962827	100.371528	50
X0014	2020-11-15 09:06:13	14.968565	100.367037	47
X0014	2020-11-15 09:07:15	14.974672	100.362252	48

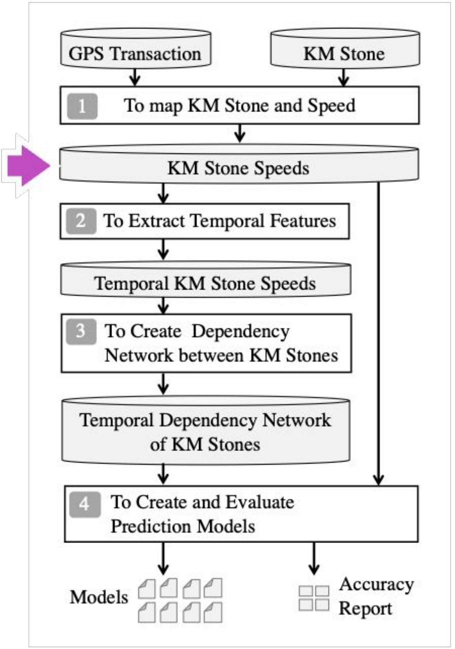
KM Stones



km_label	lat	lon
7.001	13.1161688001	100.975973441
7.002	13.1144100683	100.966835646
7.003	13.1126516511	100.957683748
7.004	13.1108964263	100.948545118
7.005	13.1104979633	100.939314030
7.006	13.1099031085	100.930087223
7.007	13.1061030342	100.921608368

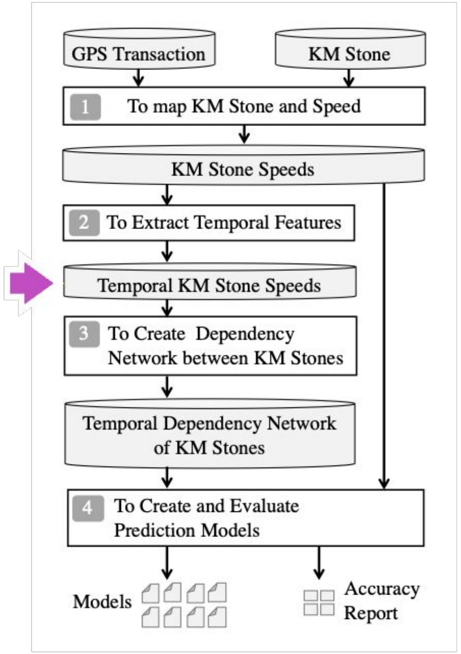
KM Stone Speed





km_label	date	time	speed	num
7.001.in	2020-11-15	09:00	23.25	80
7.001.in	2020-11-15	09:05	22.56	77
7.001.in	2020-11-15	09:10	25.44	103
7.001.in	2020-11-15	09:15	27.21	97
7.001.out	2020-11-15	09:00	34.46	145
7.001.out	2020-11-15	09:05	35.67	160
7.001.out	2020-11-15	09:10	37.44	207
7.001.out	2020-11-15	09:15	36.16	132
7.002.in	2020-11-15	09:00	24.11	66
7.002.in	2020-11-15	09:05	23.42	78
7.002.in	2020-11-15	09:10	25.25	73
7.002.in	2020-11-15	09:15	23.67	105
7.002.out	2020-11-15	09:00	36.12	103
7.002.out	2020-11-15	09:05	35.21	136
7.002.out	2020-11-15	09:10	38.43	155
7.002.out	2020-11-15	09:15	37.12	198

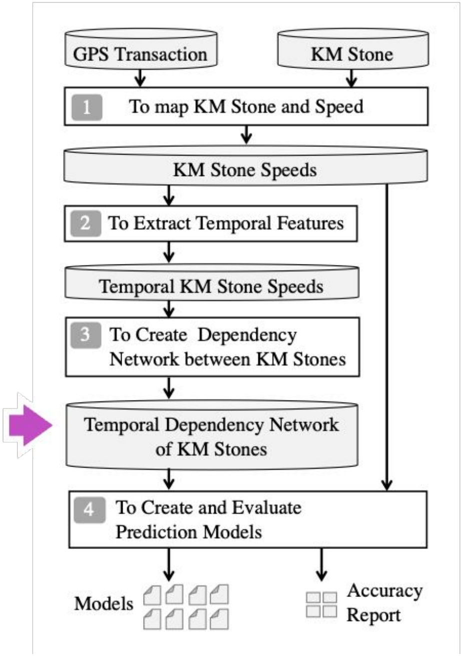
Temporal KM Stone Speed

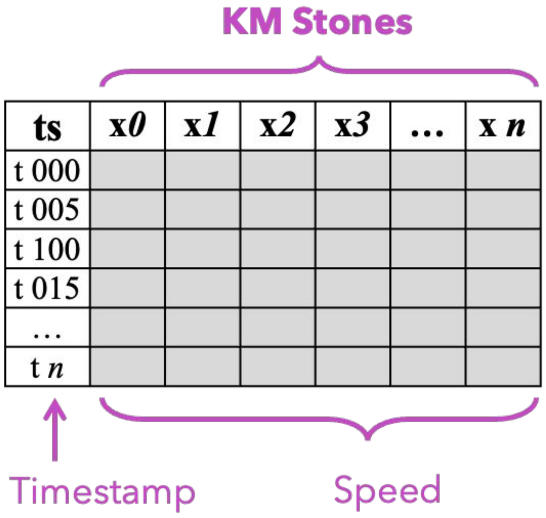


sat.sun / mon / fri / tue.wed.thu

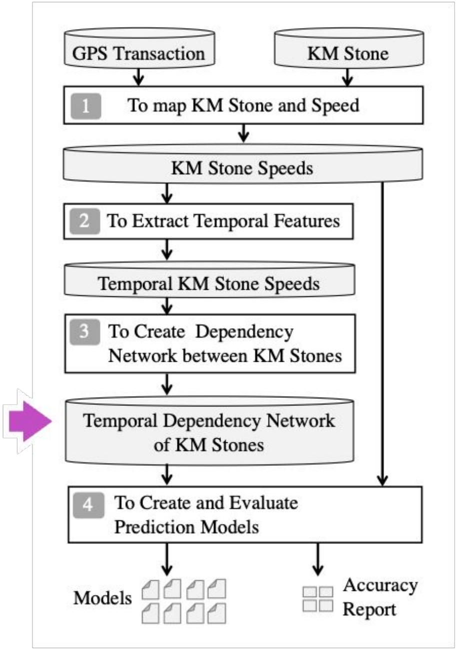
km_label	date	time	speed	num	weekday	hour
7.001.in	2020-11-15	09:00	23.25	80	sat.sun	9
7.001.in	2020-11-15	09:05	22.56	77	sat.sun	9
7.001.in	2020-11-15	09:10	25.44	103	sat.sun	9
7.001.in	2020-11-15	09:15	27.21	97	sat.sun	9
7.001.out	2020-11-15	09:00	34.46	145	sat.sun	9
7.001.out	2020-11-15	09:05	35.67	160	sat.sun	9
7.001.out	2020-11-15	09:10	37.44	207	sat.sun	9
7.001.out	2020-11-15	09:15	36.16	132	sat.sun	9
7.002.in	2020-11-15	09:00	24.11	66	sat.sun	9
7.002.in	2020-11-15	09:05	23.42	78	sat.sun	9
7.002.in	2020-11-15	09:10	25.25	73	sat.sun	9
7.002.in	2020-11-15	09:15	23.67	105	sat.sun	9
7.002.out	2020-11-15	09:00	36.12	103	sat.sun	9
7.002.out	2020-11-15	09:05	35.21	136	sat.sun	9
7.002.out	2020-11-15	09:10	38.43	155	sat.sun	9
7.002.out	2020-11-15	09:15	37.12	198	sat.sun	9

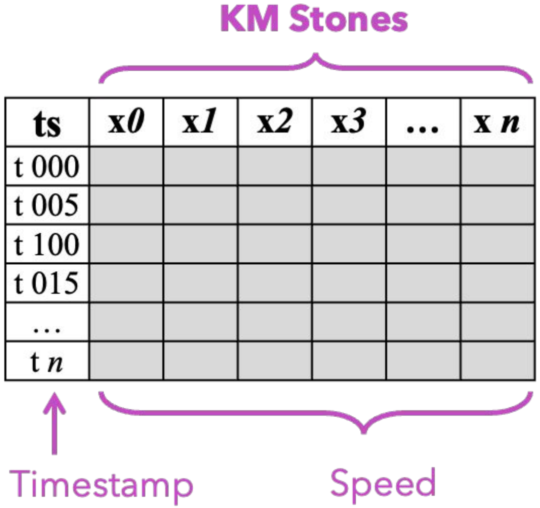
Temporal Dependency Network of KM Stone

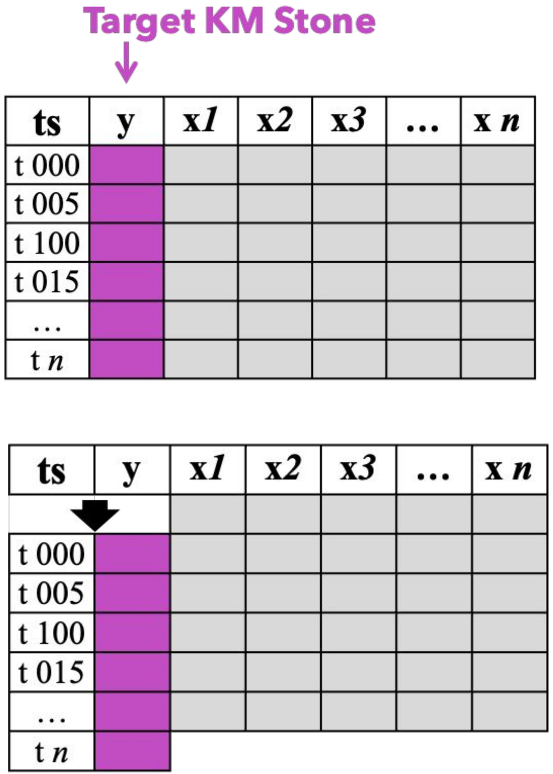




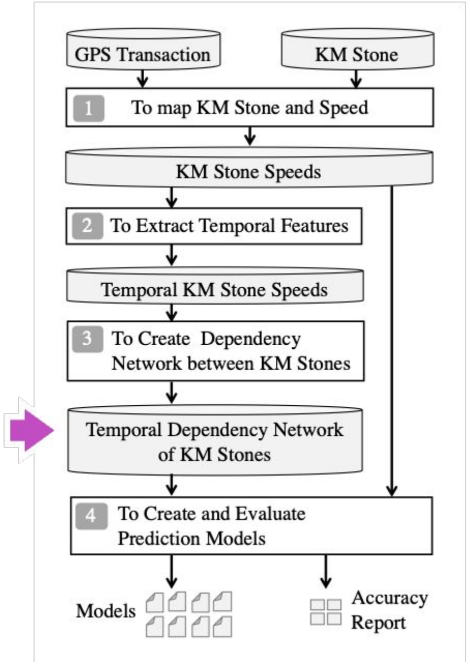
Temporal Dependency Network of KM Stone







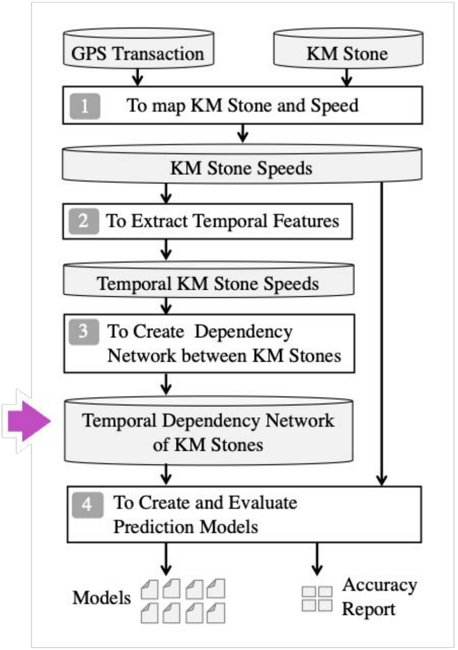
Temporal Dependency Network of KM Stone



Correlation

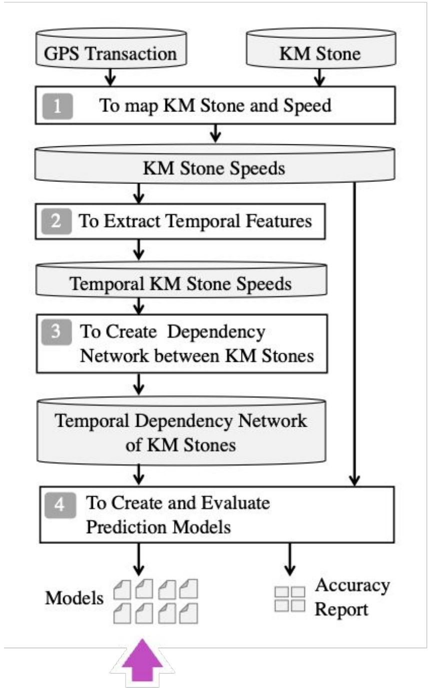
ts	y	x1	x2	x3	…	x n
t 005						
t 100						
t 015						
…						
t n						

Temporal Dependency Network of KM Stone



km_y	km_x	lag_mins	hours	weekdays	support
34.036.out	34.035.out	5	6	fri	1.00
31.009.in	31.010.in	10	15	mon	1.00
34.019.in	34.018.in	15	15	tue.wed.thu	0.82
35.001.out	35.001.out	20	18	tue.wed.thu	1.00
34.036.in	34.035.out	20	6	tue.wed.thu	0.82
31.006.in	31.006.in	5	18	tue.wed.thu	0.82
31.009.out	31.009.in	10	15	tue.wed.thu	0.82
34.036.out	34.035.out	10	6		0.81
34.036.out	34.035.out	15	6		0.81
31.010.out	31.009.in	15	18	mon	0.75
31.006.out	31.006.out	20	18	mon	0.75
304.012.out	31.017.out	10	18	mon	0.75
31.010.in	31.010.out	5	18	mon	0.75

Models

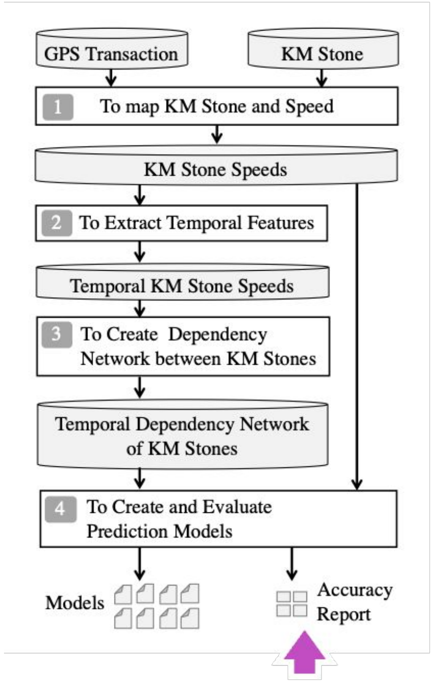


Linear Regresion

to predict speed at

$$\mathbf{km_Y} = a_1 km_{X1} + \ldots + a_n km_{Xn} + C$$

Result



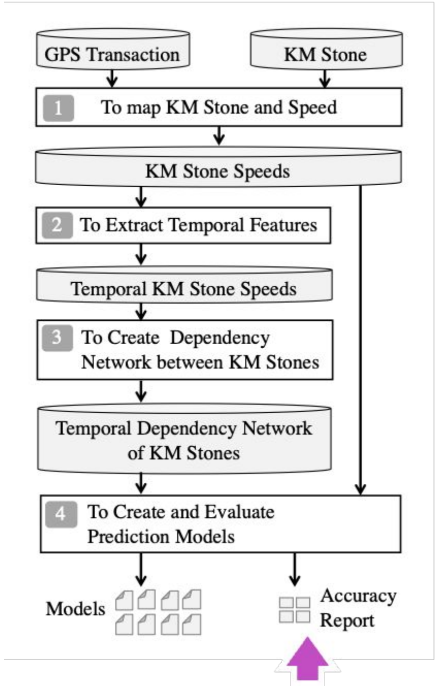
RMSE Result for the Prediction in the next N minutes

Hours	Next *N* Minutes				Total
	5	10	15	20	
06:00 – 08:59	9.19	9.20	9.03	8.59	**9.07**
09:00 – 11:59	8.75	8.83	9.47	9.00	**8.85**
12:00 – 14:59	6.90	6.40	-	-	**6.79**
15:00 – 17:59	9.17	9.26	9.30	9.48	**9.26**
18:00 – 20:59	9.74	9.65	9.98	10.07	**9.82**
Total	**9.27**	**9.33**	**9.56**	**9.51**	**9.37**

Baseline = 14.76, so the improvement is **30.14 %** ↑

(baseline is LM with the whole dataset)

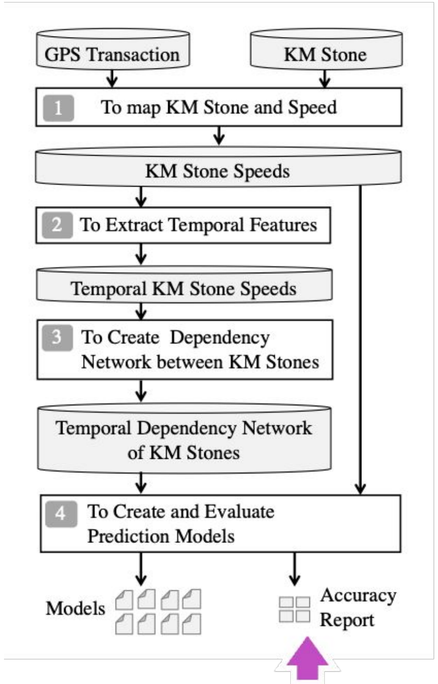
Result



RMSE of Top-10 KM Stones

Target KM Stone Label	Next *N* Minutes				Total
	5	10	15	20	
304.002.in	5.66	4.92	-	-	5.41
304.012.in	5.71	5.73	5.64	5.73	**5.70**
304.014.out	6.06	6.09	6.09	6.09	**6.07**
304.003.in	5.87	5.78	7.48	7.48	**6.25**
304.001.out	6.39	-	-	-	**6.39**
304.012.out	5.61	7.46	9.51	9.51	**6.56**
3.020.in	6.83	6.71	-	-	**6.78**
304.015.in	7.05	7.05	6.72	6.98	**7.01**
304.002.out	7.07	7.07	-	-	**7.07**
304.005.out	7.10	7.15	6.76	7.37	**7.11**

Result



RMSE for the Prediction in next 10 minutes

Hours	Weekday Groups			Total
	Mon	Tue,Wed,Thu	Fri	
06:00 – 08:59	9.91	9.01	8.49	**9.20**
09:00 – 11:59	9.87	8.71	8.73	**8.83**
12:00 – 14:59	-	6.49	6.10	**6.40**
15:00 – 17:59	9.82	9.27	8.33	**9.26**
18:00 – 20:59	10.26	9.22	10.22	**9.68**
Total	**10.05**	**9.08**	**9.06**	**9.34**

Hand-On

Hand-On

- Basic
  - https://colab.research.google.com/drive/1ye6T4fndJ3NdM7YDeMWjBae7zRz-kqUv

- GPS (Bus)
  - https://docs.google.com/presentation/d/1LKedIo6EiJmSy2XhK461xC-iGNGfIraAIVDzn5sjruo

Summary

Summary

- Introduction
  - Point, Polyline, Projection, Distance, and Grid
- Bus Quality
  - Polyline → Grid
- Next Stop Prediction
  - Frequency, Direction, and Popularity
- Travel Time Prediction
  - LSTM, Self-Attention, and Filter
- Speed Prediction
  - Some KM stones are depended on other KM stones

つづく