Learning and Prediction over Massive Spatio-temporal Data

Dong Wang
Institute for Interdisciplinary Information Sciences
Tsinghua University, China

清華大學 Tsinghua University

My Profile

- **Research Interests**
  - Deep Learning, Machine Learning, Spatio-temporal data mining

- **Awards**
  - Rank 2 / 1648, Didi Supply-Demand Challenge Competition 2016
  - The Most Potential Prize, Didi Supply-Demand Challenge Competition 2016
  - Rank 3 / 1956, Datacastle Travel time estimation competition 2017

- **Publications**
  - **A-level** conferences: IJCAI 2017 (submitted), ICDE 2017, ICDE 2016, Ubicomp 2016
  - **B-level** conference: DASFAA 2016

Spatial Temporal data

Traffic

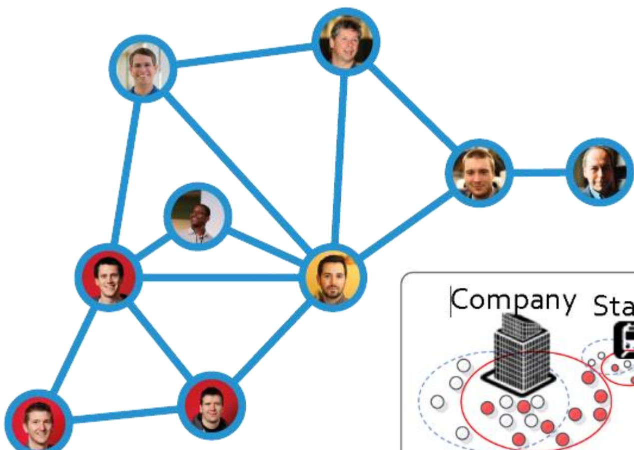
- Location and time information
- Navigation, traffic management etc.

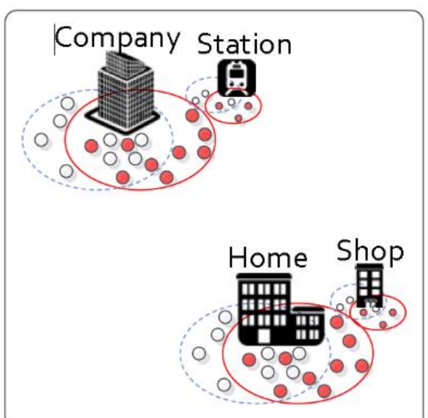
Economy

- Store Site Selection

Spatial Temporal data

Social
- Check-in data
- Infer or recommend the friend to users

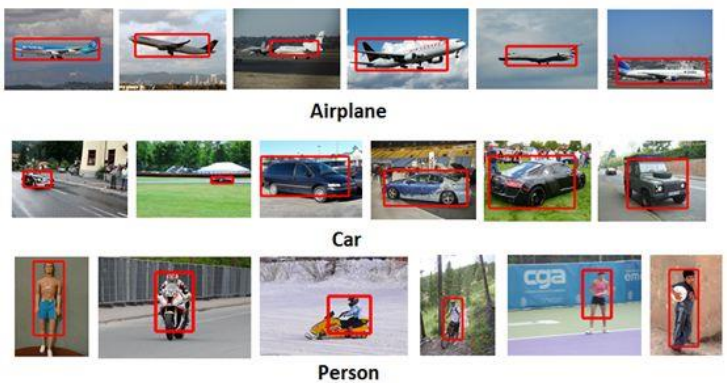




Warehouse management
- Pick requests
- Delivering data

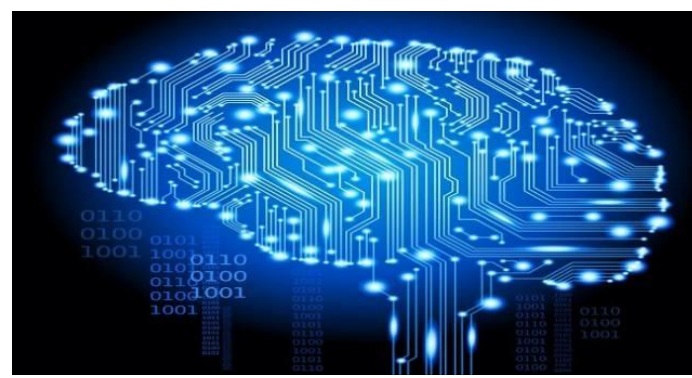
Deep Learning

- The hottest topic in ML / DM

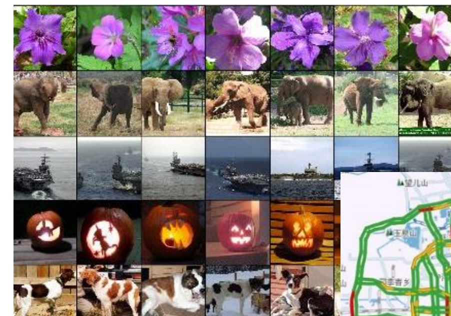


- Standard Architecture
  - Images/Videos —— CNN
  - Text/Speech —— RNN
  - Playing Games/Auto Drive —— DQN

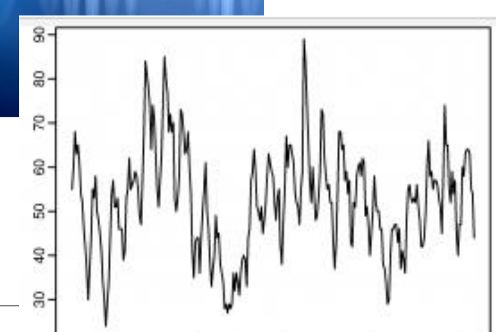
No standard architecture for spatio-temporal data.



Characteristic



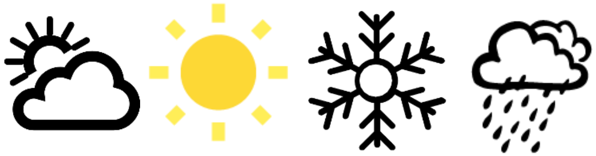
- Spatial dependence
  - different locations interact on each other
  - compare with images:
    - city level scale, sensitive to the granularity



- Temporal dependence
  - past states affect the future
  - compare with texts/speech:
    - periodicity in multi-granularity
    - highly affected by sudden event (raining, traffic accident)

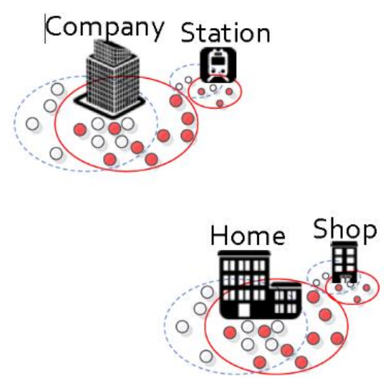
Characteristic

- Diverse data sources
  - Mobile phones, online car-hailing orders, weather, POIs, etc.
- Massive, highly noisy



PHD Work

- **Supply-demand Prediction**
  - Online Car-hailing Services
- **When will you arrive?**
  - Estimating Travel Time Based on Recurrent Neural Networks
- **Social relationship detection**
  - Automatic User Identification across Heterogeneous Data Sources
- **Traffic condition Prediction**
  - Traffic Condition Prediction System



Supply-Demand Prediction for Online Car-hailing Services using Deep Neural Network

- **Objective**
  - Predict the gap between the car-hailing supply and demand in a certain area in the next 10 minutes.[1]
- *This problem is from [Di-tech Algorithm Competition 2016]*
- **Motivation**
  - Balance the supply-demand by scheduling the drivers in advance
  - Adjust the price dynamically

Tsinghua University

Definitions

Car-hailing order

valid (invalid)

1. Date
2. Timeslot
3. Passenger ID
4. Star area ID
5. Destination area ID

Environment data:

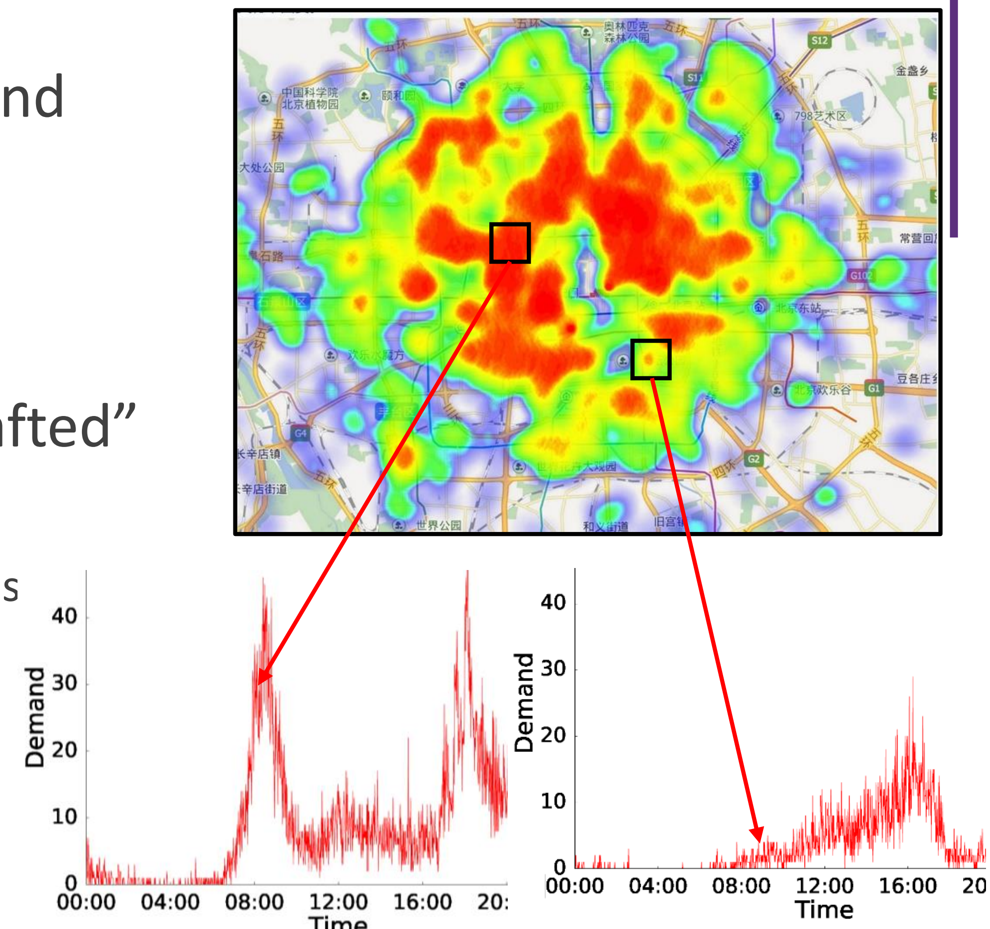
1. weather
2. traffic condition

Objective

Predict the supply-demand gap (e.g., the number of invalid orders) of a certain area, in the next 10 minutes.

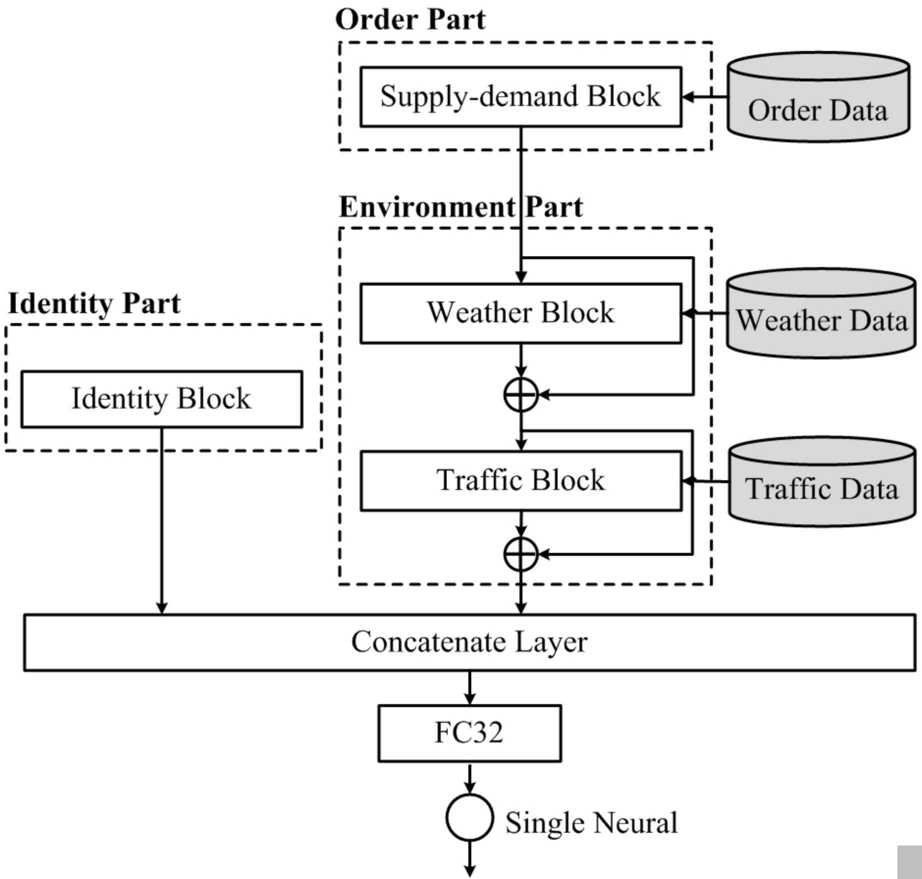
Challenges

- The car-hailing supply-demand varies dynamically
  - geographic locations
  - time intervals.
- Standard models +“hand-crafted” features
- Logistic regression, SVM, random fores gradient boosting
- Various data types
  - Order, date, weather, traffic
- Various data sources



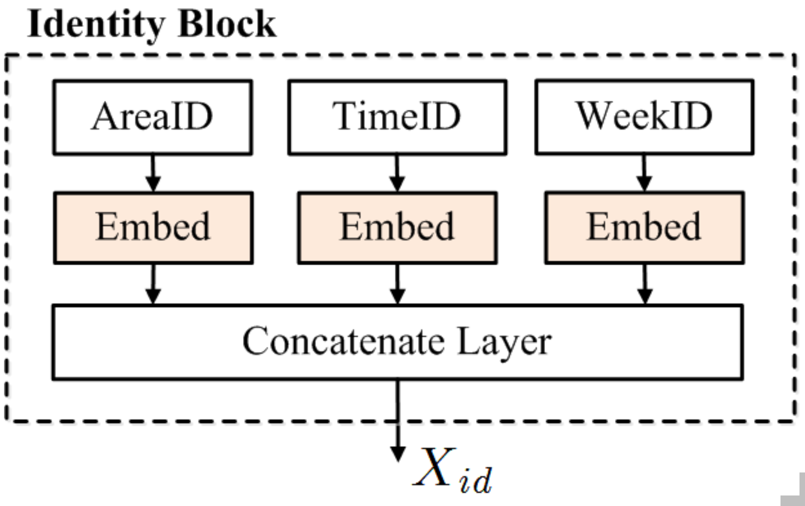
Framework

1. General blocks
2. Using embedding to "cluster" similar areas and timeslots
3. Learning the useful feature vector from the order data
4. Connecting different blocks with residual link
5. End-to-end model



Identity Part

- Different areas at different time can share similar supply-demand patterns.
- Prior work clusters the similar data :
  - Manually design the distance measure
  - Build several sub-models (business area, residential area, etc.)



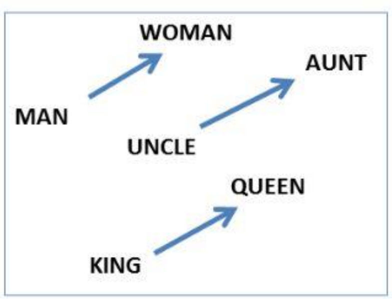
Embedding

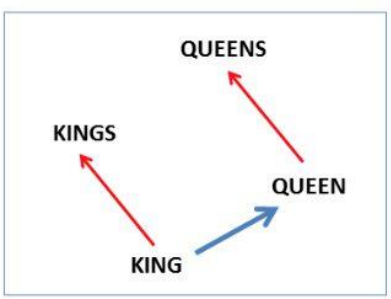
- Categorical value -> real vector

$$y_t = x_t \cdot W$$

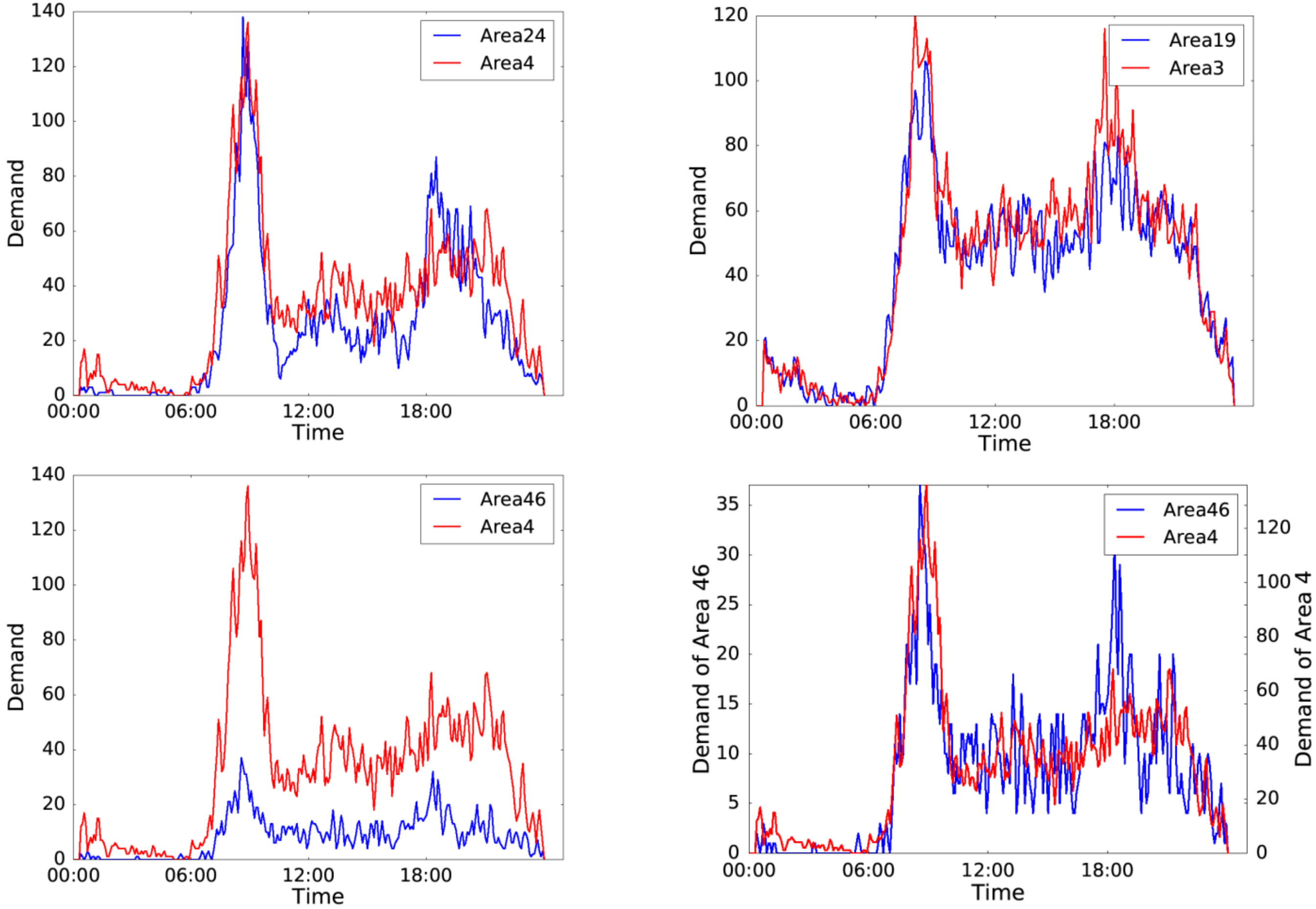
$$y_t = (-0.2, 0.4, 0.1) \qquad x_t = (0, 0, 1, 0, 0)$$

- Discover semantic similarity

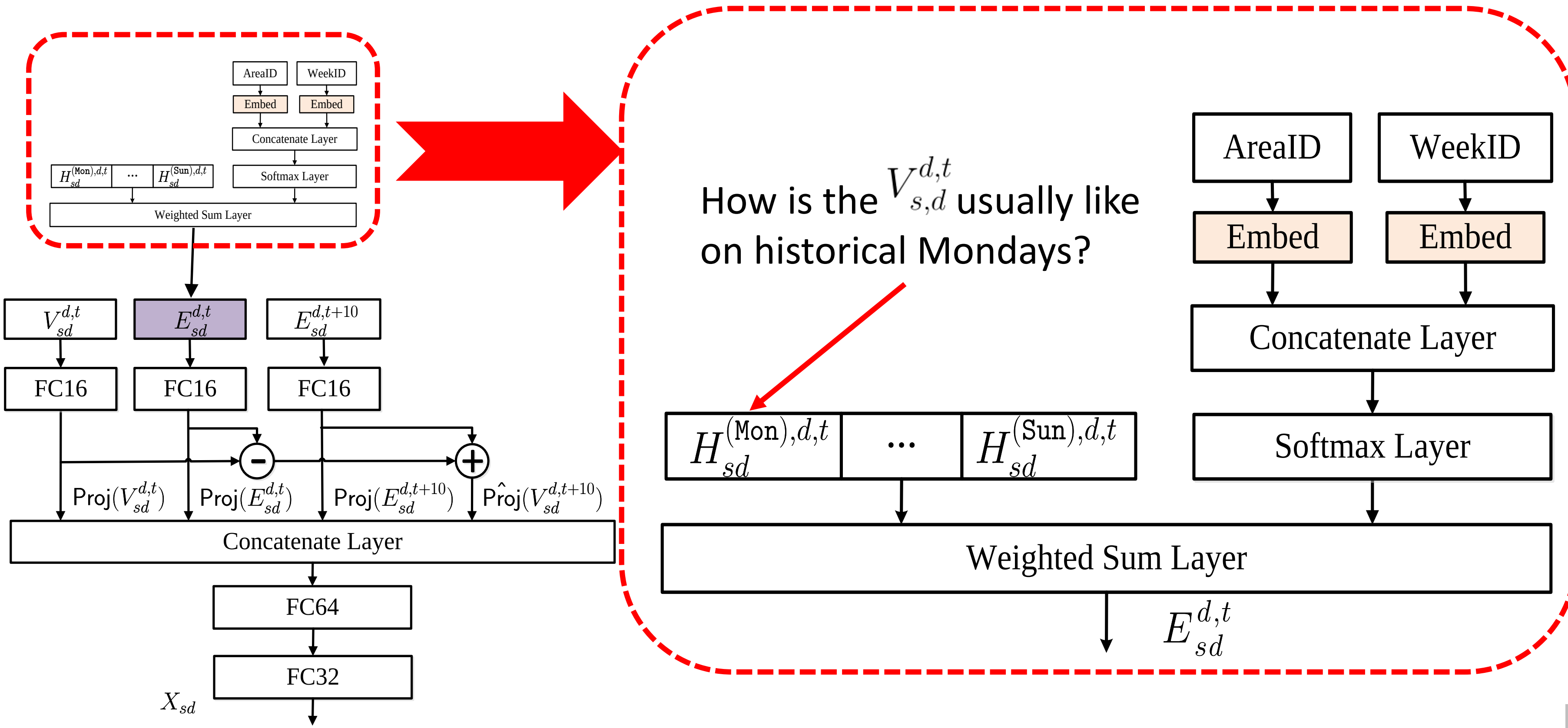




Effects of Embedding

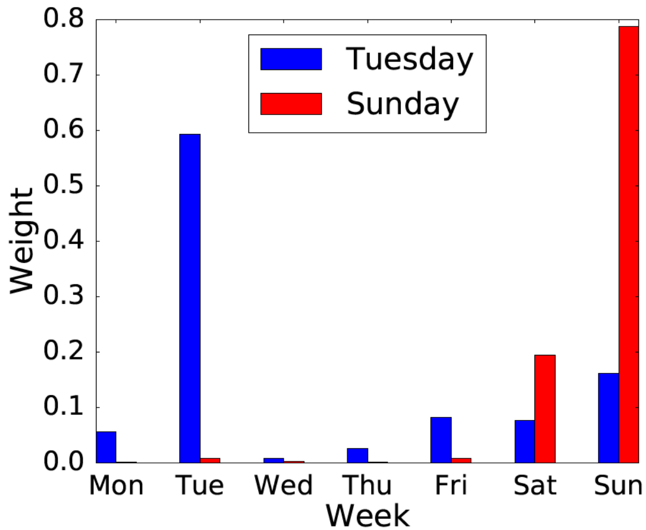


Order Part

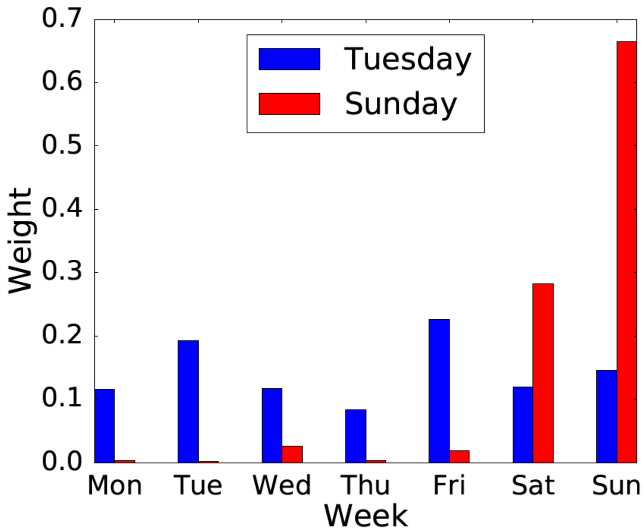


AreaID
WeekID
Embed
Embed
Concatenate Layer
Softmax Layer
$H_{sd}^{(\text{Mon}),d,t}$
···
$H_{sd}^{(\text{Sun}),d,t}$
Weighted Sum Layer
$V_{sd}^{d,t}$
$E_{sd}^{d,t}$
$E_{sd}^{d,t+10}$
FC16
FC16
FC16
$\text{Proj}(V_{sd}^{d,t})$
$\text{Proj}(E_{sd}^{d,t})$
$\text{Proj}(E_{sd}^{d,t+10})$
$\hat{\text{Proj}}(V_{sd}^{d,t+10})$
Concatenate Layer
FC64
FC32
X_{sd}
How is the $V_{s,d}^{d,t}$ usually like on historical Mondays?

Effects of Embedding



Area 1

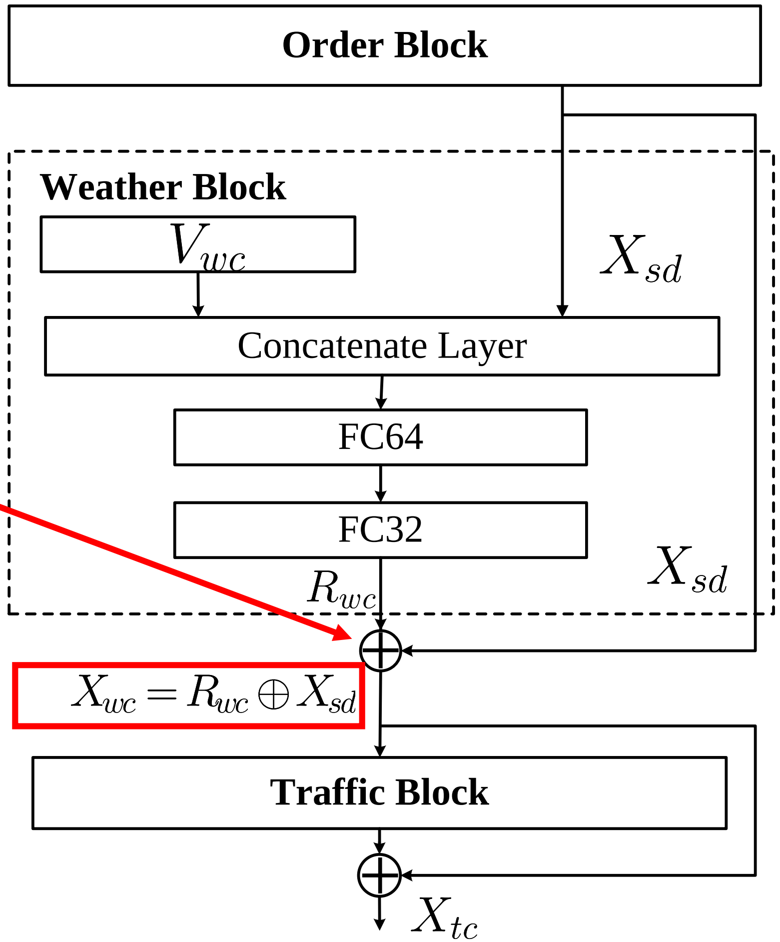


Area 26

We visualize the weight vectors in two different areas at Tuesday and Sunday.

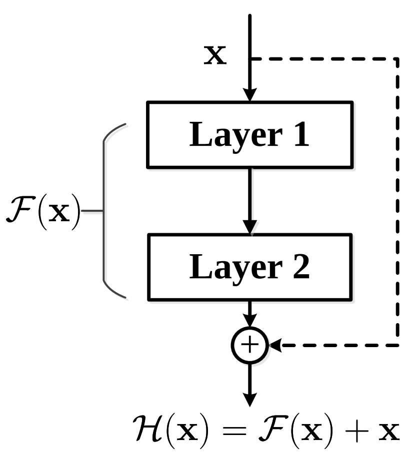
Residual links

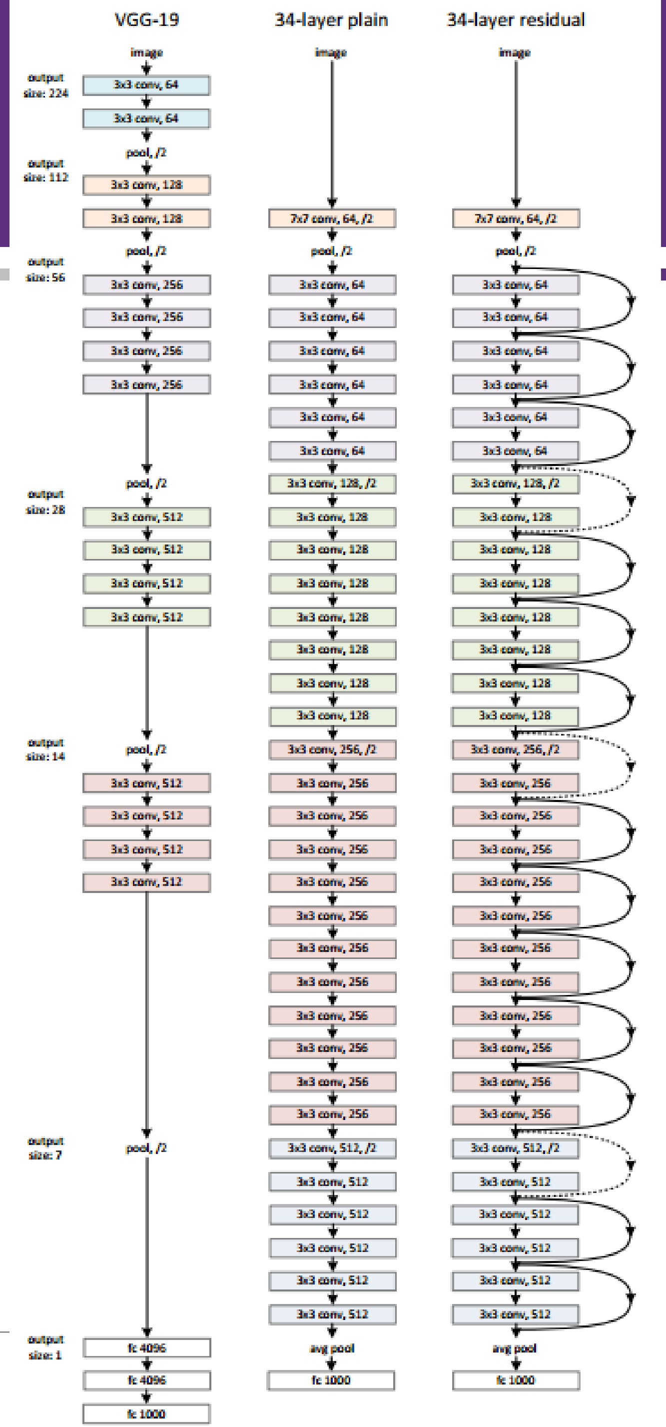
- Weather Block
- Residual link
  - Take the output of weather block as the “residual”
  - Makes the model more flexible to incorporate new data



Deep Residual Networks[1]

- Train very deep neural network
  - Gradient vanishing/exploding problem
- Add connections between layers

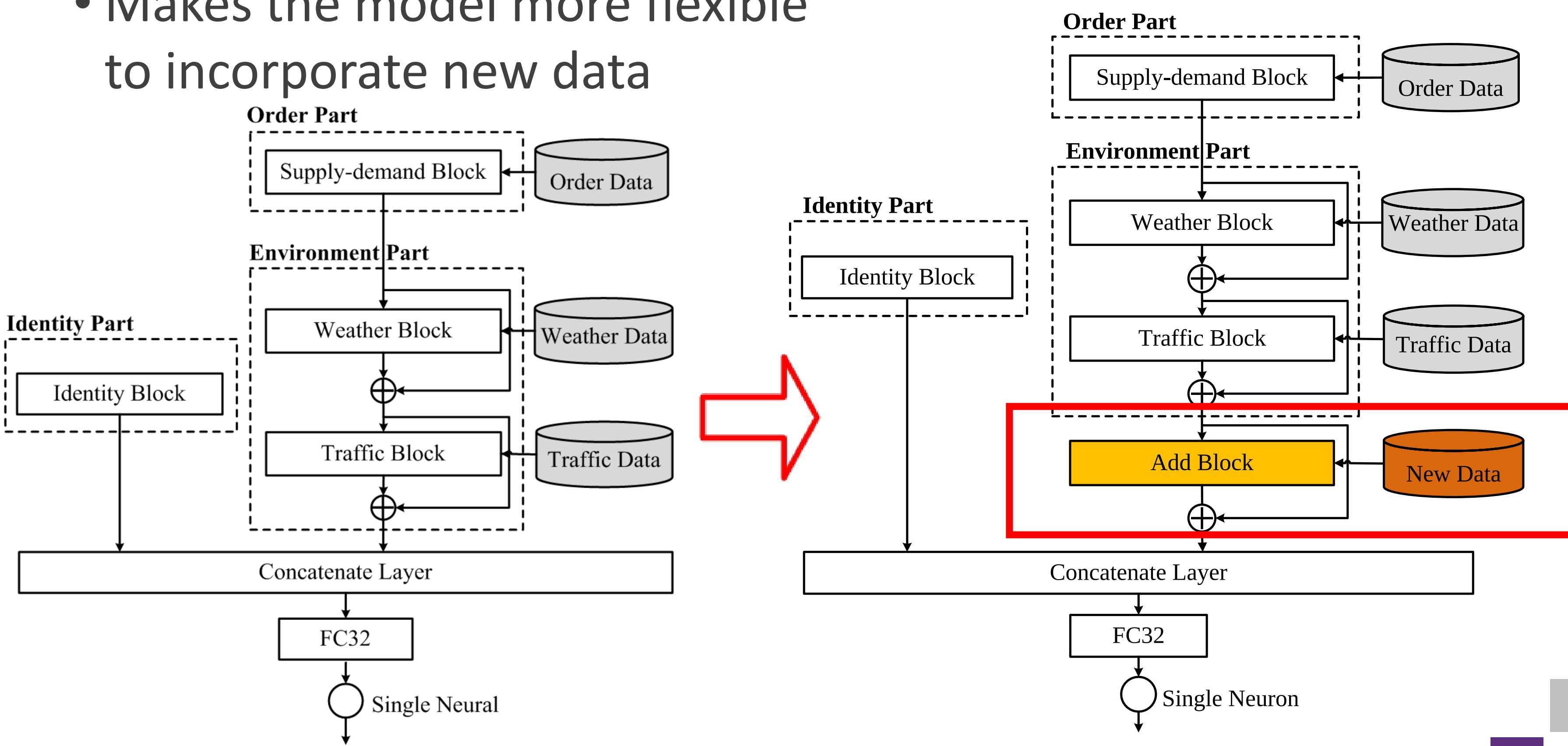




[1] He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016.

Incorporate New Data

- Makes the model more flexible to incorporate new data

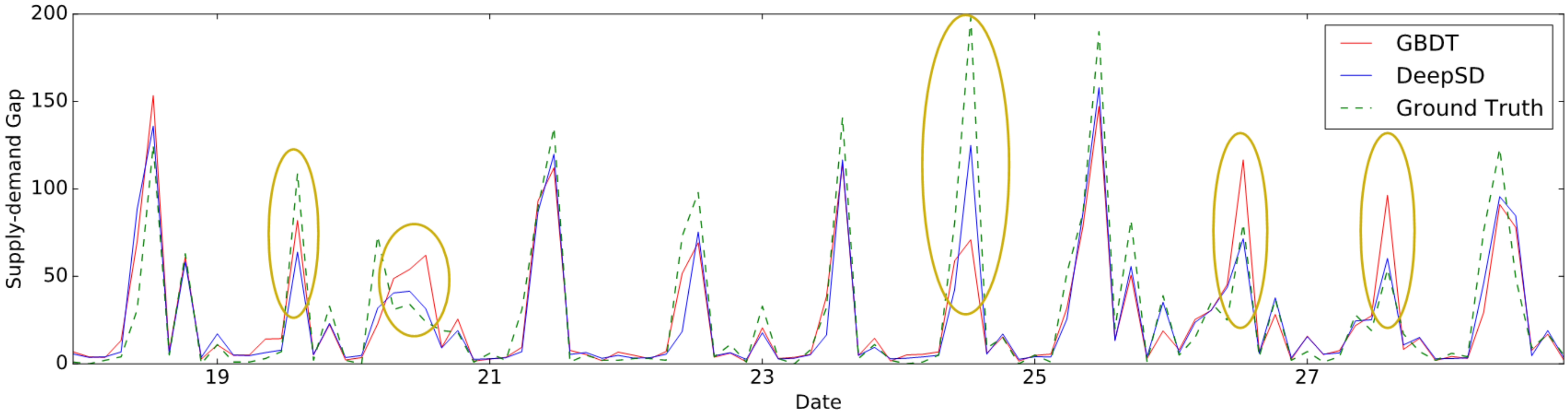


Experiment

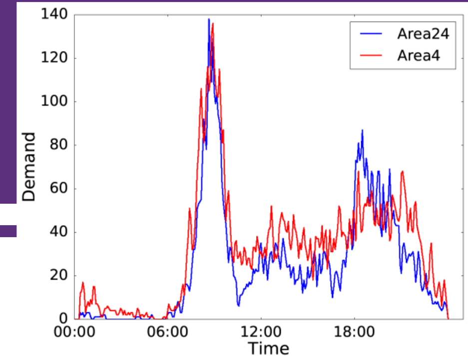
Table: Performance Comparison

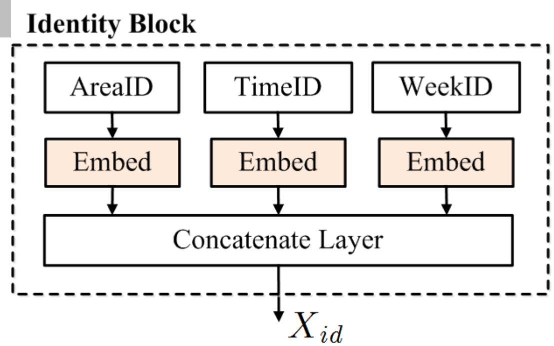
Model	Error Metrics	
	MAE	RMSE
Average	14.58	52.94
LASSO	3.82	16.29
GBDT	3.72	15.88
RF	3.92	17.18
Basic DeepSD	3.56	15.57
Advanced DeepSD	**3.30**	**13.99**

Experiment

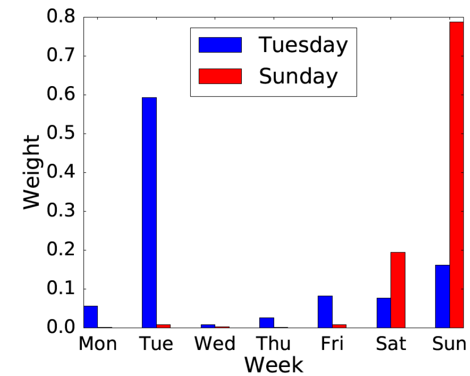

GBDT
DeepSD
Ground Truth
Supply-demand Gap
Date
200
150
100
50
0
19
21
23
25
27

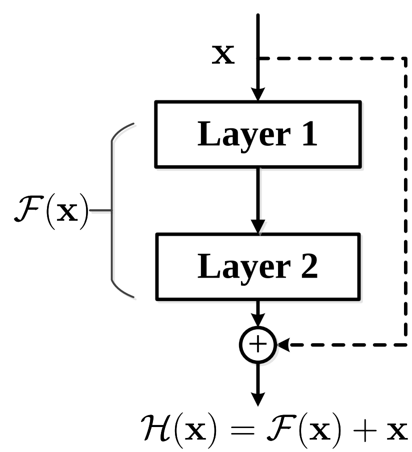
Conclusion

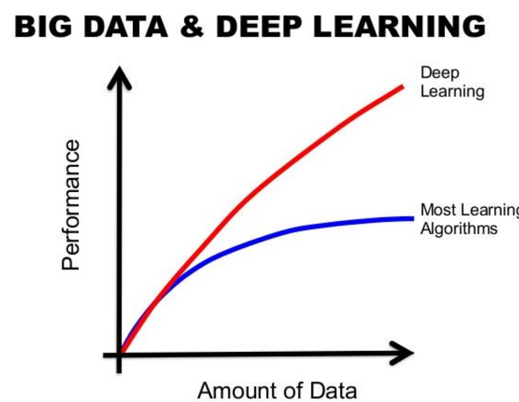




1. End-to-end model
2. Design a general block
3. Learn the useful feature vector from the order data
4. Involve in new external data easily
5. Great potential







Tsinghua University

Estimating Travel Time Based on Recurrent Neural Networks

When will you arrive?[1]

Motivation

- Routes planning, Navigation
- Traffic dispatching

Previous work

- Estimate for each individual road
- Road intersections and traffic lights
- No driving habits

1. This problem is from [DataCastle 2017]

Definitions

Objective

Given: 1. path 2. driver 3. start time

Estimate:

the travel time for the given path.

Train data

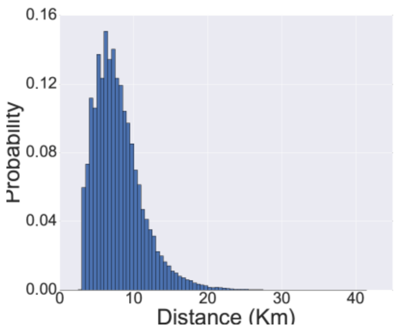
GPS trajectory

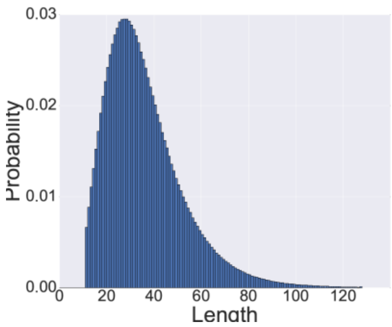
Sample points from path

Challenges

- The travel time of a specific path can be very different
  - ✓ Peak/Non-peak hour
  - ✓ The day of the week

- Diverse values of trajectory length/distance.

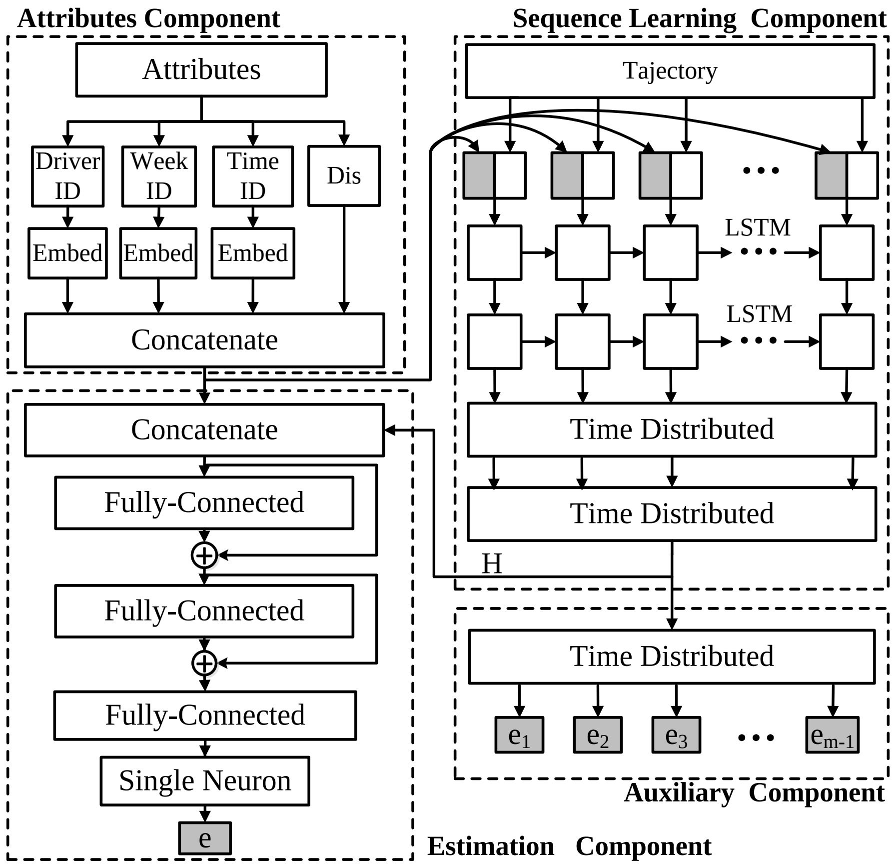




- Different driving habits

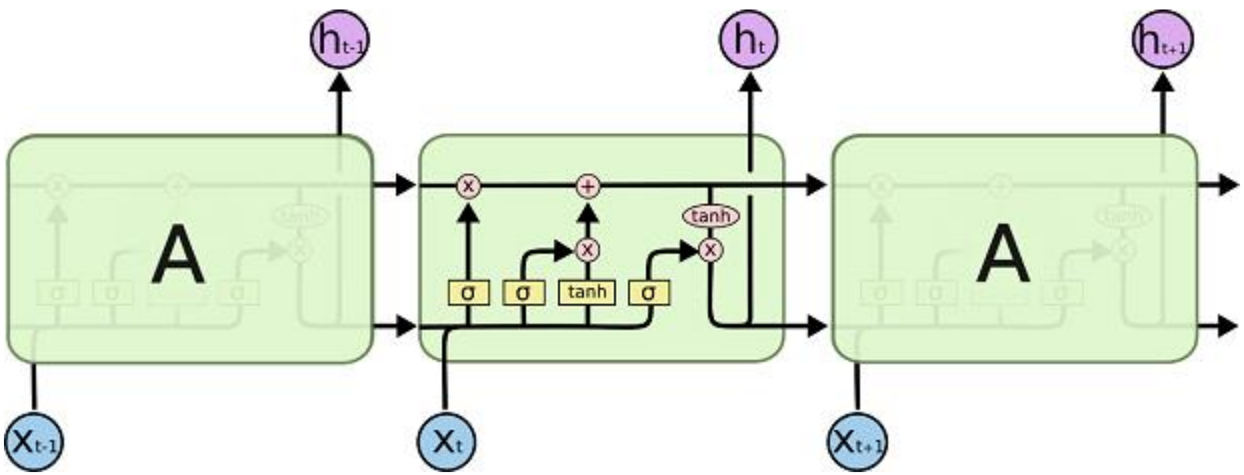
Architecture

1. Use Attributes Component incorporate various factors
2. Use Sequence Learning Component to handle trajectory
3. Use Estimation Component to predict the travel time
4. Extend to multi-task learning by introducing an Auxiliary Component



Sequence Learning Component

- RNN(Recurrent Neural Network)
- LSTM (Long Short Term Memory)
- Time dependence and spatial dependence

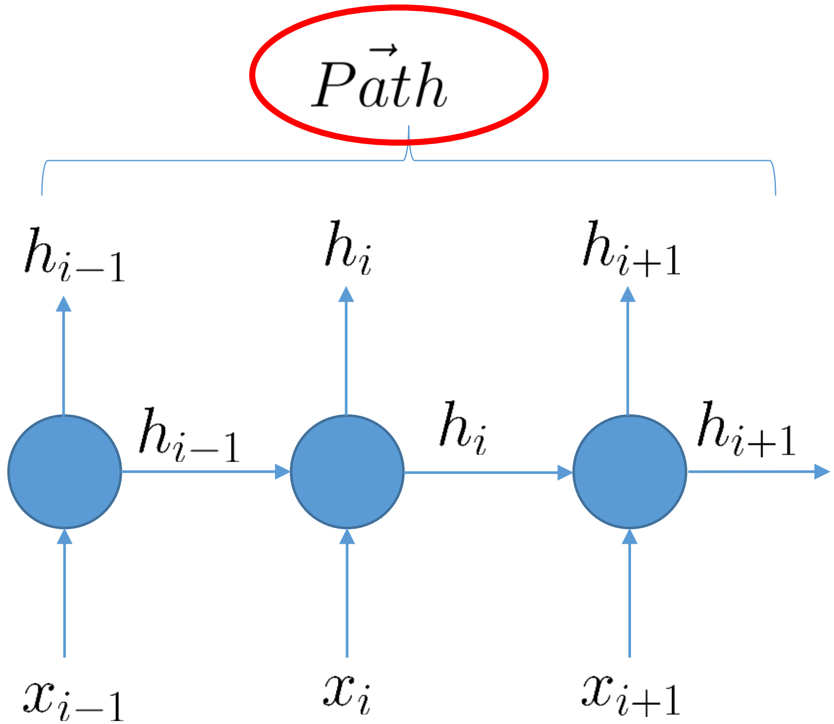


$$x_i = (lng_i, lat_i, lng_{i+1}, lat_{i+1}, d_{i,i+1})$$

清華大学 Tsinghua University

Sequence Learning Component

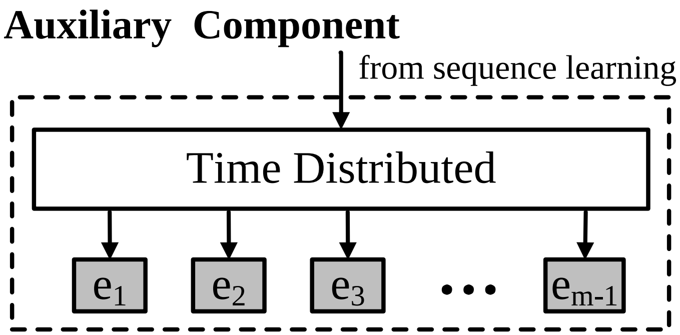
- x_i -> h_i
  - Abstract of the first i points
  - Deal the new point
- Trajectory -> Vector
  - Represent the whole trajectory with all h_i.
- Handling different trajectory lengths
  - Mean Pooling Trick
  - Sampling Trick



Auxiliary Component

To utilize the “local information”

- estimate the travel time of each sub-trajectory
- extend to a multi-task model
- used as the auxiliary output



Model Training

- Evaluate: mean absolute percentage error (MAPE)
  - Estimation Component

$$\mathrm{loss}_{seq} = \left|e - \Delta t_{p_1 \to p_{L_m}}\right| / \Delta t_{p_1 \to p_{L_m}}.$$

  - Auxiliary Component

$$\mathrm{loss}_{aux} = \frac{1}{m-1} \sum_{i=1}^{m-1} \frac{\left|e_i - \Delta t_{p_{L_i} \to p_{L_{i+1}}}\right|}{\Delta t_{p_{L_i} \to p_{L_{i+1}}} + \epsilon}.$$

  - Final loss:

$$\mathrm{loss} = \mathrm{loss}_{seq} + \alpha \cdot \mathrm{loss}_{aux}$$

Experiment

Data Description

- 1.4 billion GPS records of 14,864 taxis in Oct. 2014 in Chengdu.
- Total number of trajectories: 9,653,822. (60GB)
- Use the last 7 days (from 24th to 30th) as the test set and the remaining ones as the training set.

Experiment

Table: Performance Comparison

Model	MAPE
Gradient Boosting	20.32%
MLP-3 layers	16.17%
MLP-5 layers	15.75%
Vanilla RNN	18.85%
DeepTTE	**13.14%**

Experiment

Table: Performance of Different Number of Samples

#Samples	MAPE	Time (per epoch)
DeepTTE-10	15.45%	674s
DeepTTE-30	13.14%	1729s
DeepTTE-70	13.02%	3879s
DeepTTE-100	12.74%	5484s
DeepTTE-Var	12.87%	5841s

Experiment

Effects of Components

- Eliminate Estimation Component, 28.44%;
- Eliminate Auxiliary Component, 13.95%;
- Our entire model, 13.14%.

Experiment

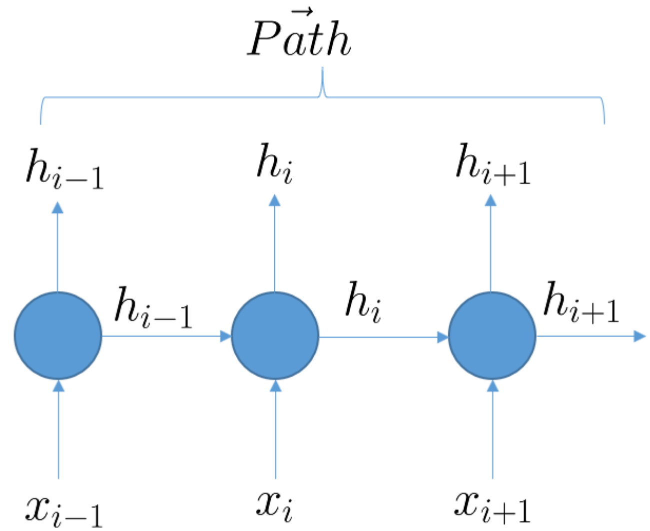
Table: Effects of Attribute Component

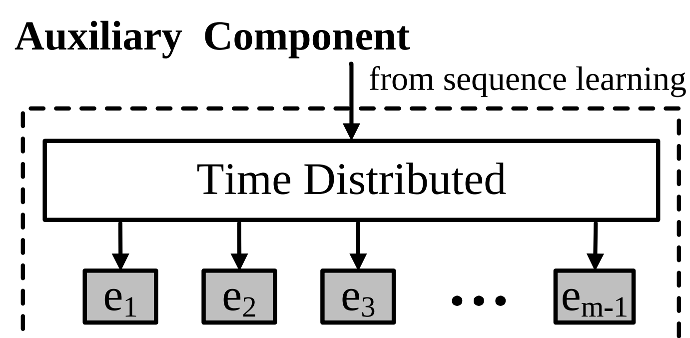
Model	MAPE
DeepTTE-30	**13.14%**
Eliminate driverID	13.37%
Eliminate weekID	13.58%
Eliminate both	13.59%

Tsinghua University

Conclusion

1. New block for handle trajectory (with LSTM)
2. Extend to multi-task learning by introducing an Auxiliary Component





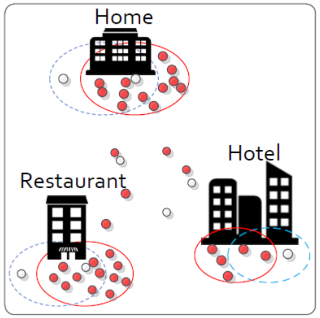
Automatic User Identification across Heterogeneous Data Sources

Goal: Identify the same user from the historical trajectory data set.

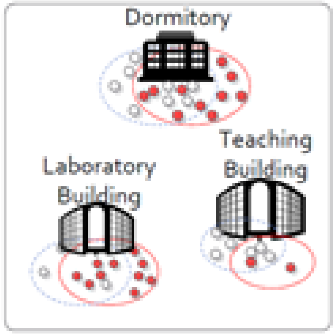
Motivation: human mobility, data integration, improve data quality

Challenges:

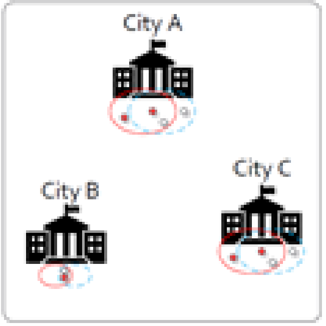
- Very different sampling rates
- Information loss in sparse trajectories
- Temporally disjoint
- Distinguish the overlaps



Same person, different sampling rates



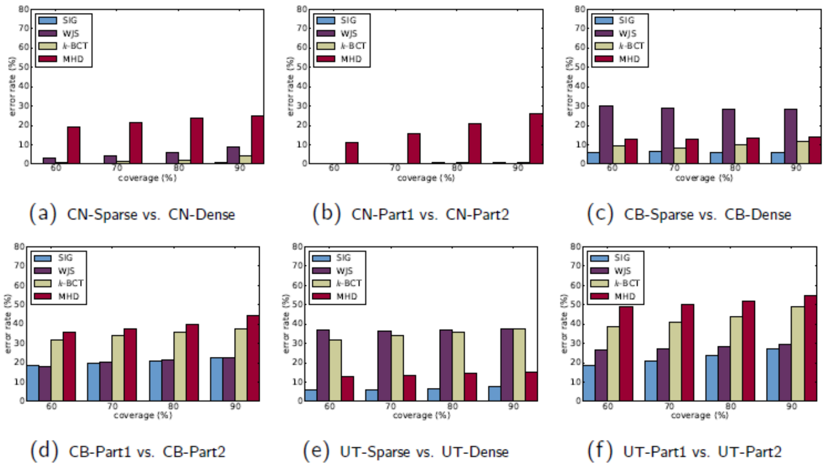
School mates, significant overlap



Same person, sparse rate, occurred in several places

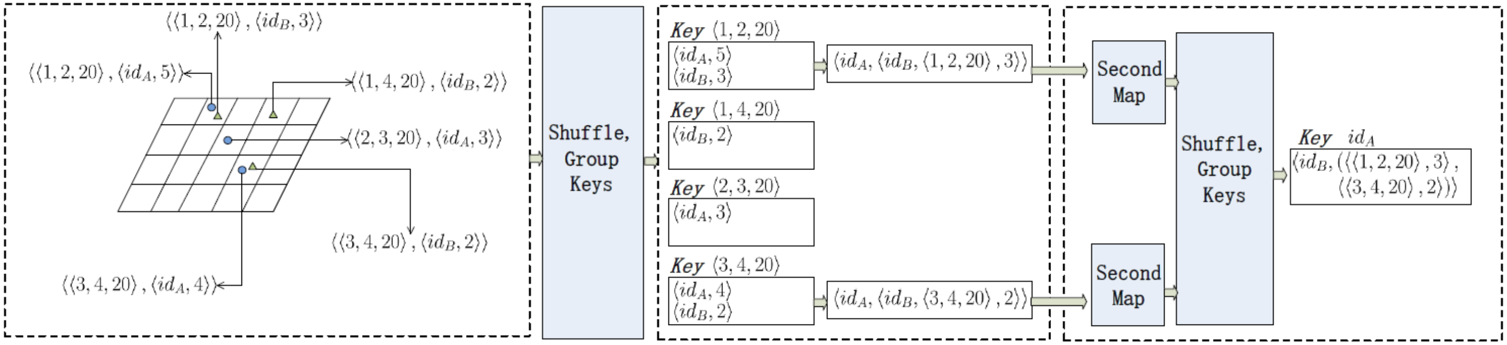
Automatic User Identification across Heterogeneous Data Sources

- **Novel similarity measure based on signals:**
    - *co-occurrence, sampling rate, distance*
    - *Robust performance*
- **Efficient framework based on MapReduce**



(a) CN-Sparse vs. CN-Dense (b) CN-Part1 vs. CN-Part2 (c) CB-Sparse vs. CB-Dense

(d) CB-Part1 vs. CB-Part2 (e) UT-Sparse vs. UT-Dense (f) UT-Part1 vs. UT-Part2



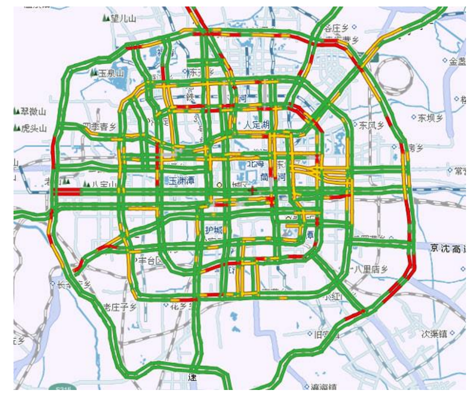
ETCPS: An Effective and Scalable Traffic Condition Prediction System

Goal: Predict the traffic condition of each road in the urban area after 15 minutes

Motivation: traffic management, routing service, taxi ride sharing

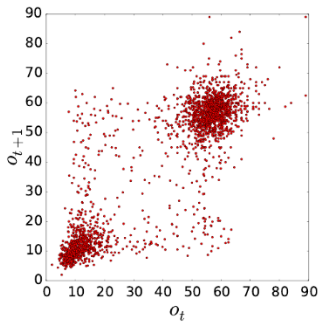
Previous work:

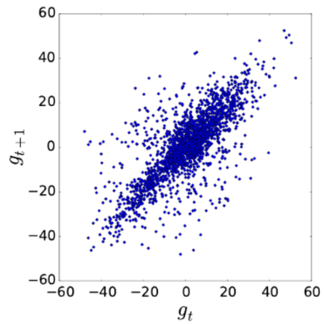
- road side loop sensors data
- GPS data collected from floating vehicles
  - only focused on the arterial roads
  - urban roads not considered

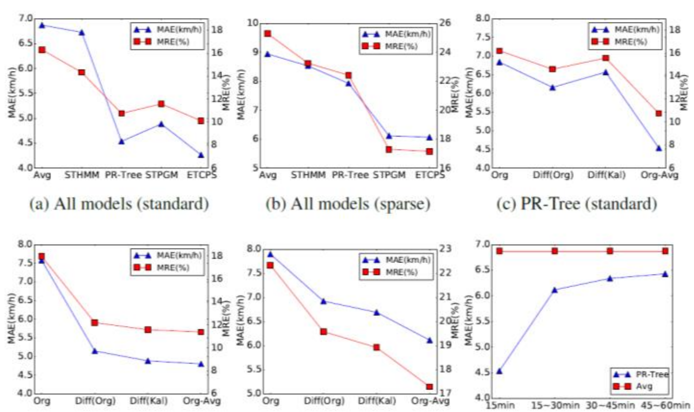


Traffic Condition Prediction System

Relationships observation



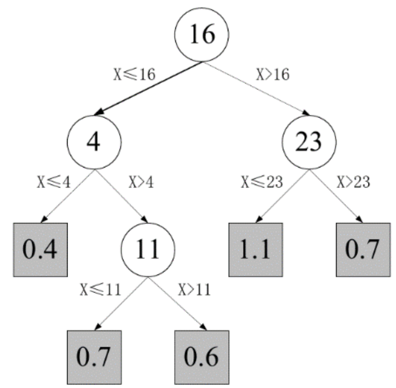


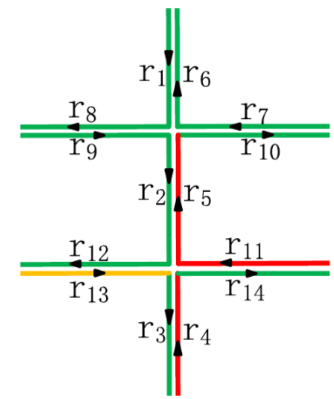


(a) All models (standard) (b) All models (sparse) (c) PR-Tree (standard)

(d) STPGM (standard) (e) STPGM (sparse) (f) Predict longer

- PR-Tree models the traffic condition time series of each individual roads





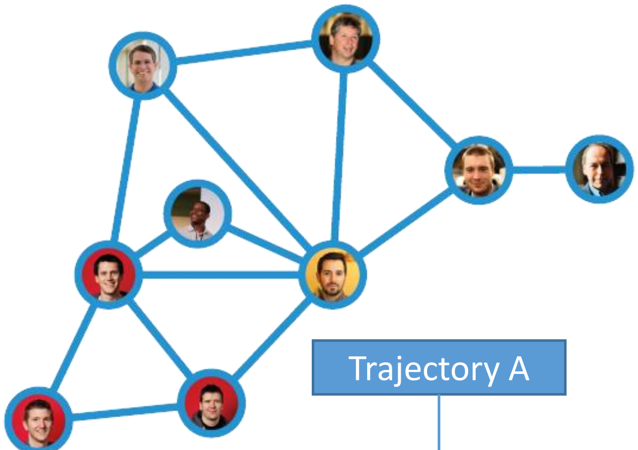
- STPGM (Spatial temporal probabilistic graphic model) models the relationship between different roads
- Our best quality prediction is achieved by a careful ensemble of the two models.

On going work

Social Relationship Detection Based on Sequence to Sequence Learning

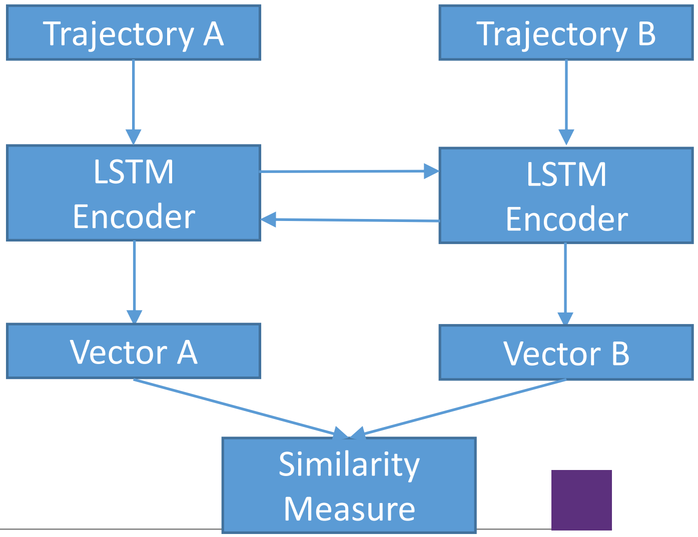
Goal:

- Learn the similarity measure by neural network
- Transform trajectories to vectors



Motivation:

- Faster algorithm for similar user searching
- Behavior prediction
- Social relationship detection



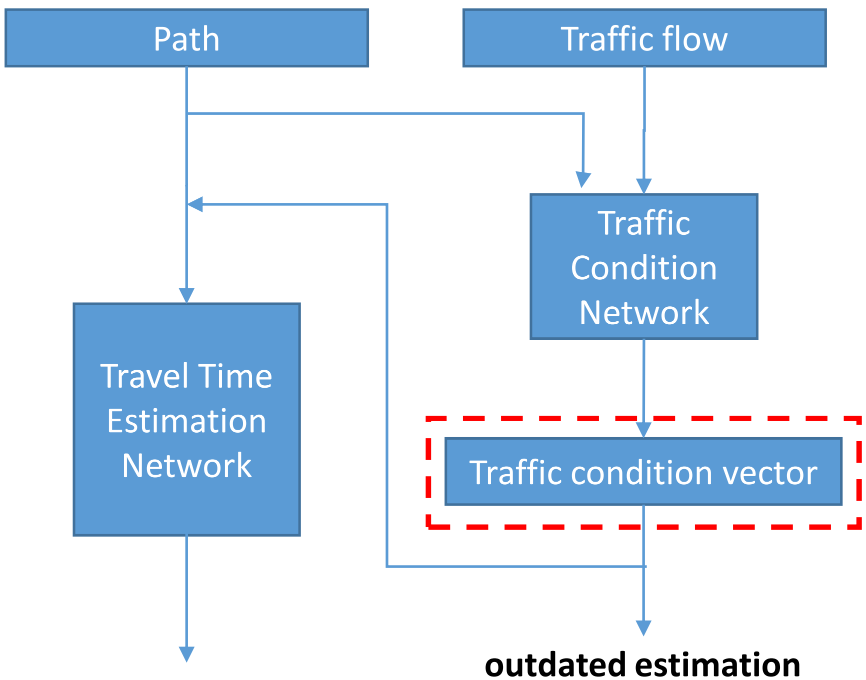
On going work

Characterizing Traffic Conditions using Recurrent Neural Networks

Goal: characterize the traffic condition of any given path in the last 30 min.

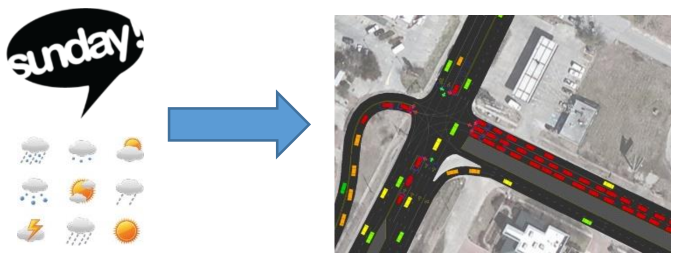
Motivation:

- Enhance accuracy for travel time estimation
- Refinement prediction of travel flow



Future plan

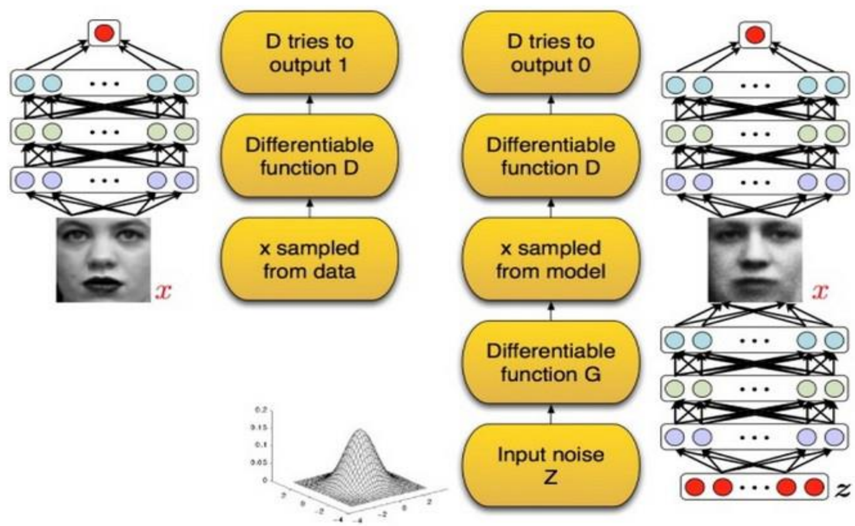
Simulation of complicated environment using generative models



- Spatio-temporal data generator under complicated environment
- Provide information for decision making

Generative Adversary Nets

(e.g. Info Gan)

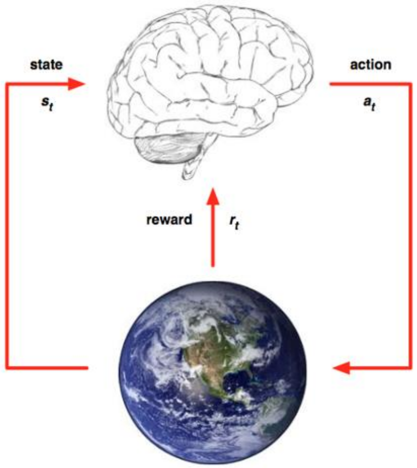


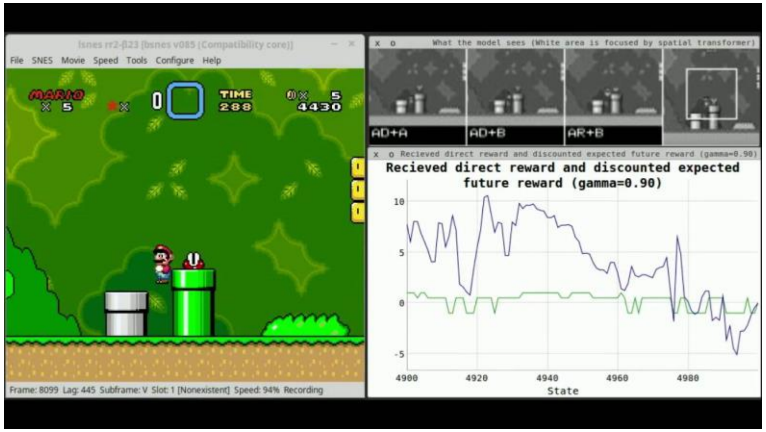
Future plan

Policy designing using the Deep Reinforcement Learning

- Car-dispatching
- Price adjusting
- Storage planning
- ……

起步价	0元
里程费(32.66公里)	45元
时长费(32分钟)	14元
动态加价 ?	39元
优惠券抵扣	10元
合计预估	88元





- Deep Reinforcement Learning
  - Combine prediction & simulator & decision making
  - Deep learning to learn Q-function

Future plan

What if I were fortune enough …

- Set up a deep learning course to attract more students
- Build up a deep learning interest group
  - Research
  - Machine learning competitions
- Apply the deep learning into economics
  - e.g., Spatial economics
  - Cooperation work with the Professors in economics

Thank you