



Statistics and Prediction on DiDi Traffic Data Based on Neural Network

Discretion Team

Data Structure Project, 2018 Spring

ShanghaiTech University

Yintao Xu

Ruiqi Liu

Qi Qin

Haomin Shi

Zilin Si

Keyi Yuan

June.24th, 2018

Agenda

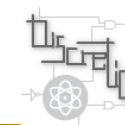
- Analysis
- Visualization
- What is LSTM?
- Application



上海科技大学
ShanghaiTech University



Agenda



- Analysis

- Visualization

- What is LSTM?

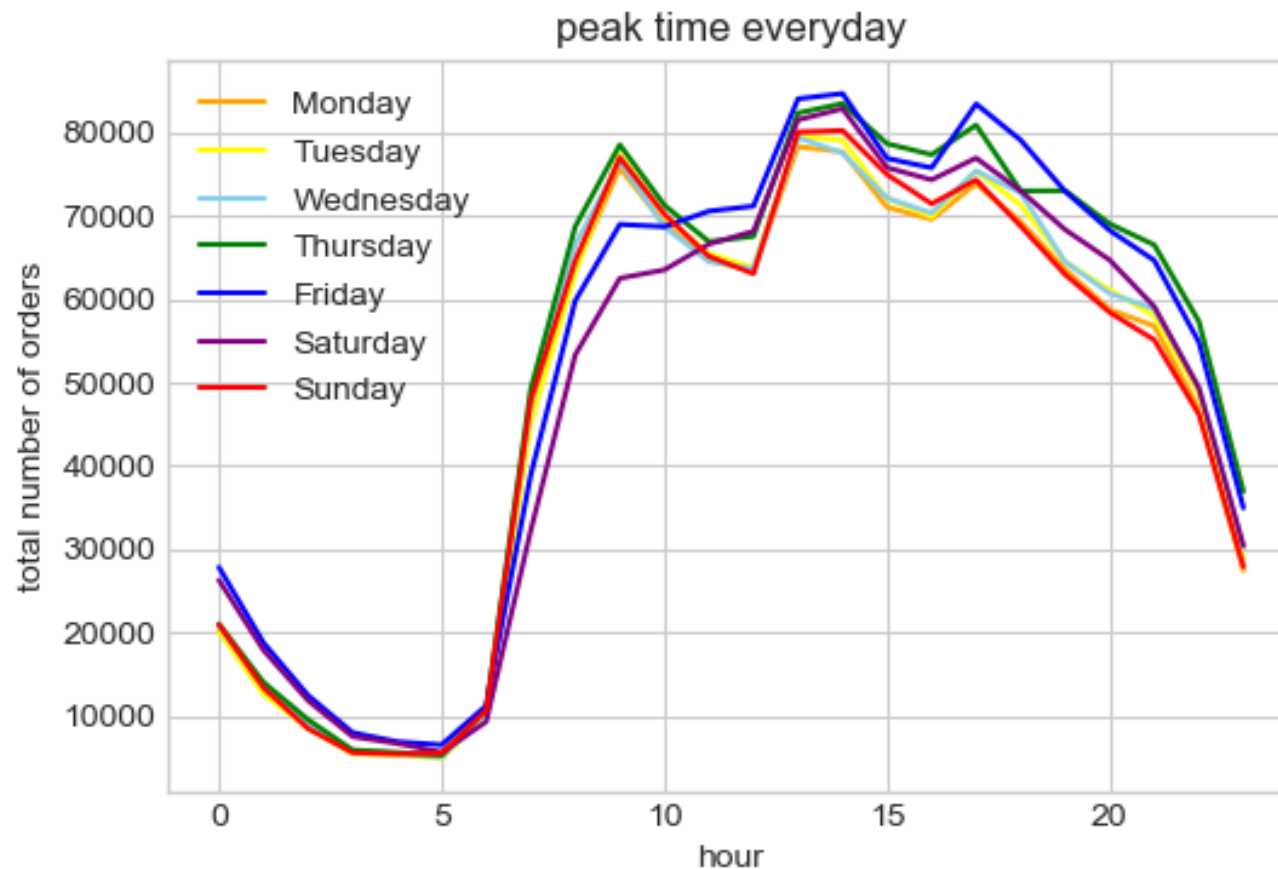
- Application



Order Distribution

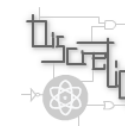


- Tools: Python
- Third Party Package: Pandas, numpys, serborn and matplotlib

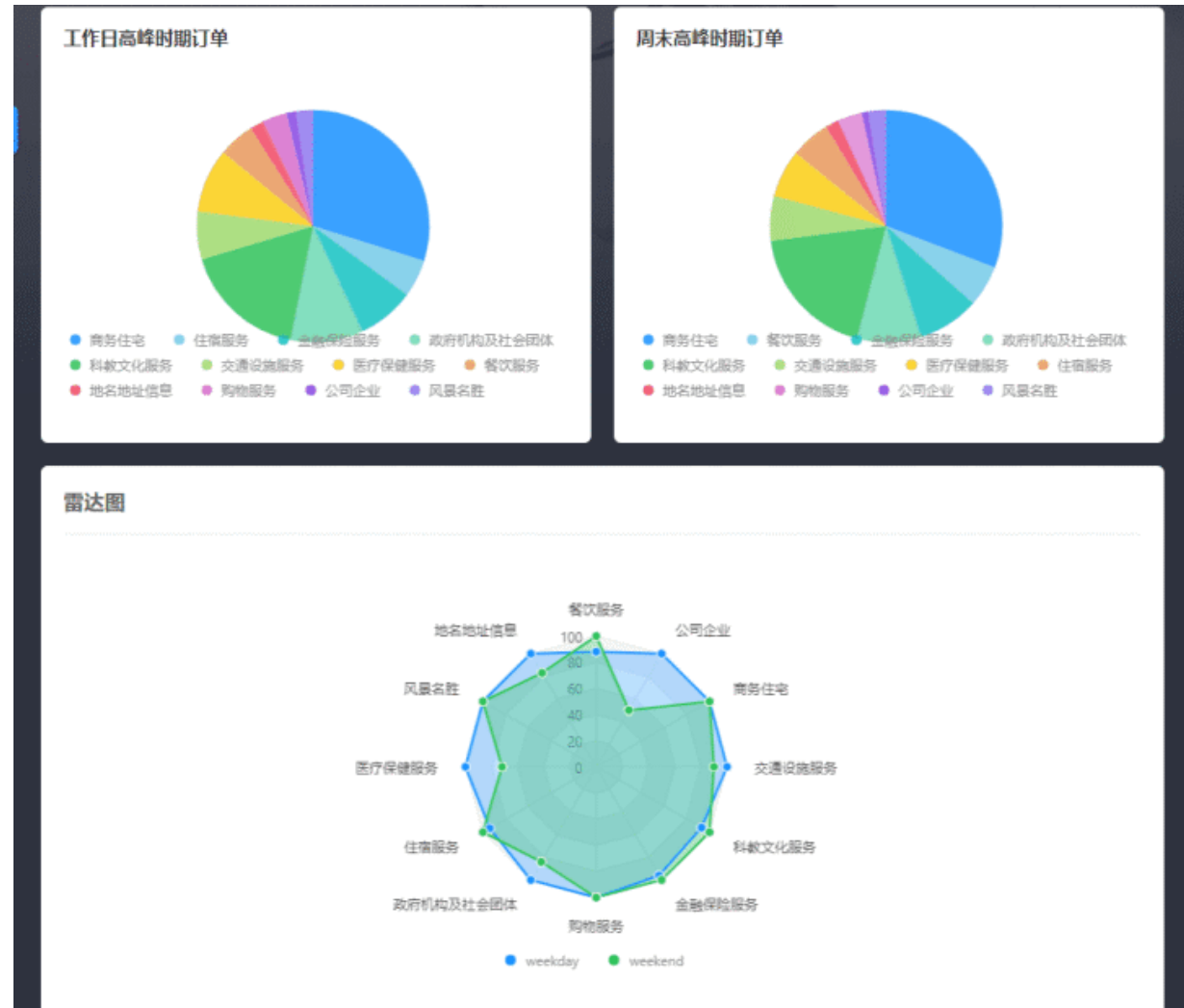


The number of order for every minute from Monday to Sunday.

Where do they prefer?

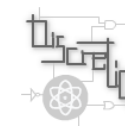


- Tools:
API offered by AMap
Getting **POI**
- Weekdays: Commercial houses
- Weekends: Restaurant



The pie chart and radar map of the orders.

Where do they prefer?



上海科技大学
ShanghaiTech University

The TOP 4 places most people will start from.

POI: Unique!

Rank	Place
1	China construction bank (Chengdu people's north road sub-branch)
2	Languang Jinhuahe International Fashion City
3	Chengdu Wanxing Family Apartment
4	Far way dream Teen Apartment

The TOP 4 places most people will go to.

Rank	Place
1	ChunXi Road Hotel Group
2	Sichuan people's government culture and history research institute
3	Chunxi shanghai square
4	Shu du building hotel conference center

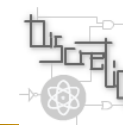
Agenda

- Analysis

- Visualization

- What is LSTM?

- Application



上海科技大学
ShanghaiTech University

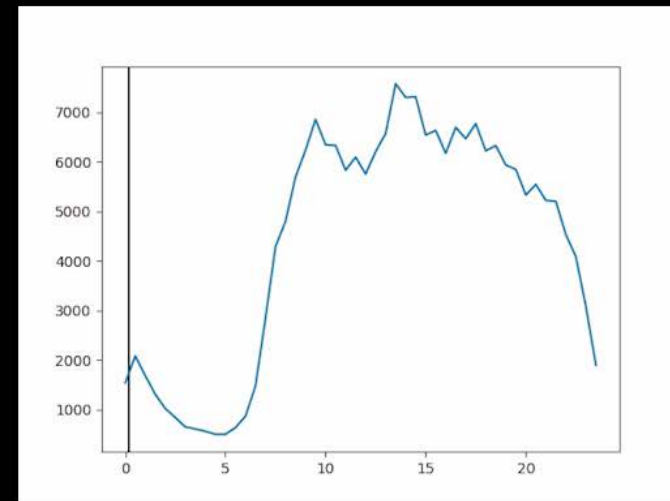
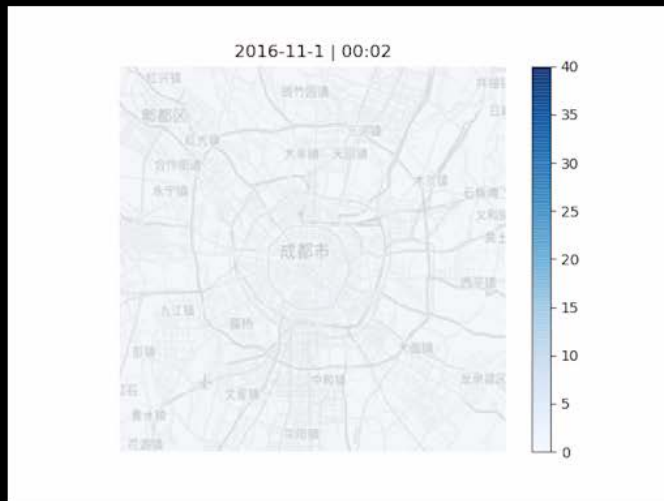


Thermodynamic Diagram



上海科技大学
ShanghaiTech University

- Method: Kernel Smooth, Gaussian Radial Basis Function



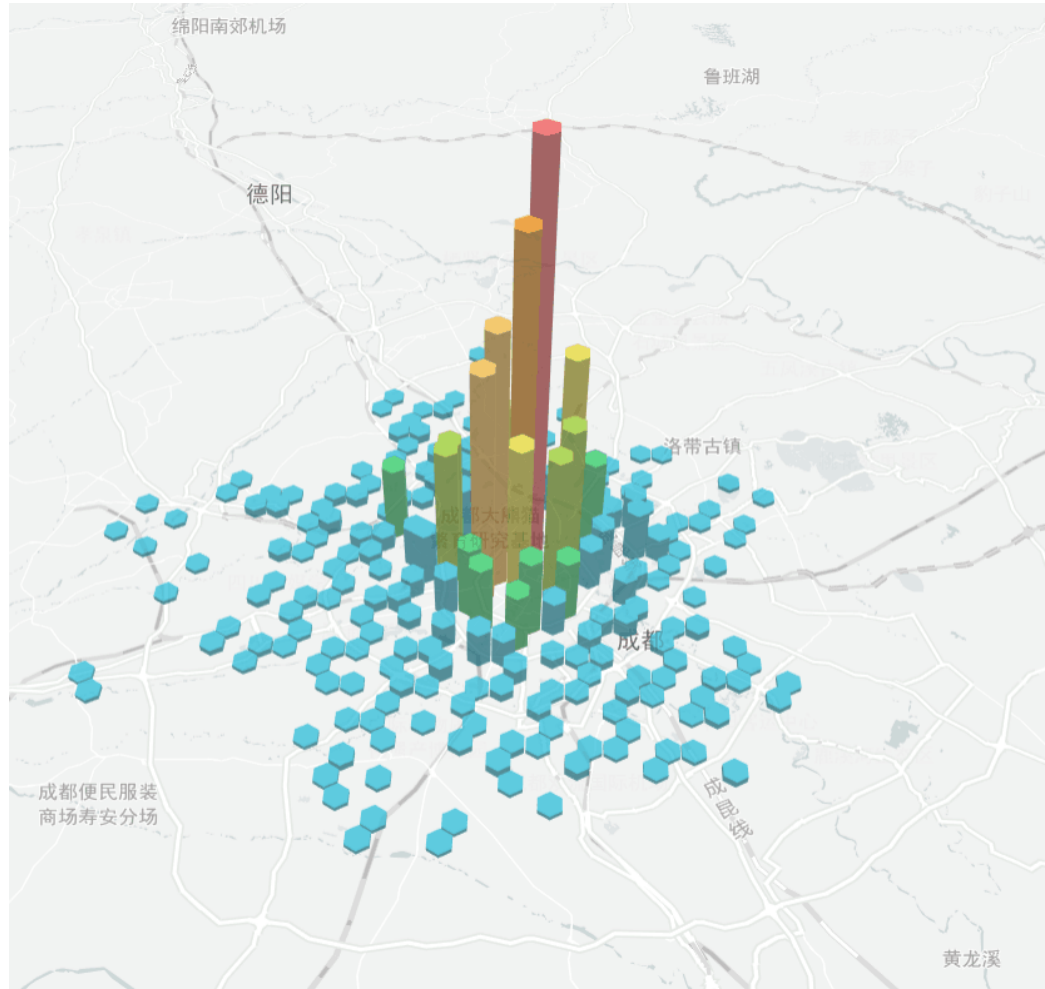
Video: A continuous thermodynamic diagram of order at a specific time.
Notice: In this video, the specific time is Nov. 1st, 2016

3D Thermodynamic Diagram



上海科技大学
ShanghaiTech University

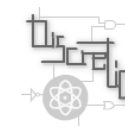
- Method: Amap Source Code, Superposition



3D thermodynamic diagram about the number of DiDi orders

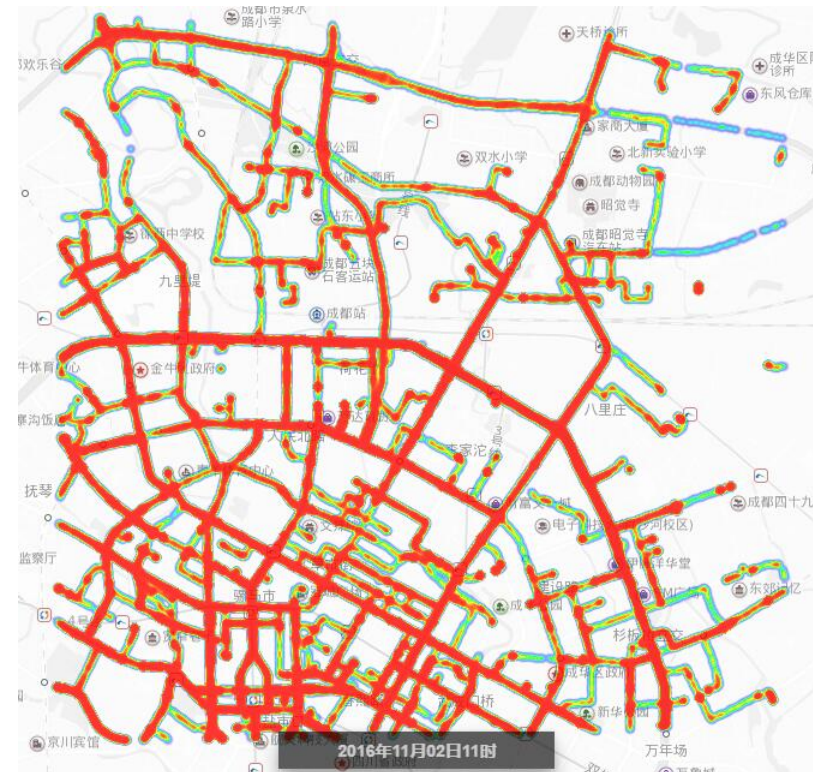
Notice: In this figure, the specific time is 8:22, Nov. 1st, 2016

GPS Visualization



上海科技大学
ShanghaiTech University

- Method: Sample
- Reason: Huge Data
- Complexity: $O(n)$, where n is the number of data.
- Aim: It shows the degree of congestion at a certain time.
- The more congestion it occurs, the more red it will be shown.

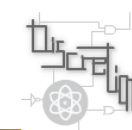


GPS visualization about the degree of congestion

Notice: In these figures, the specific time of left one is 6:00, Nov.2nd, 2016

In these figures, the specific time of right one is 11:00, Nov. 2nd, 2016

Agenda



- Analysis

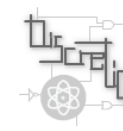
- Visualization

- What is LSTM?

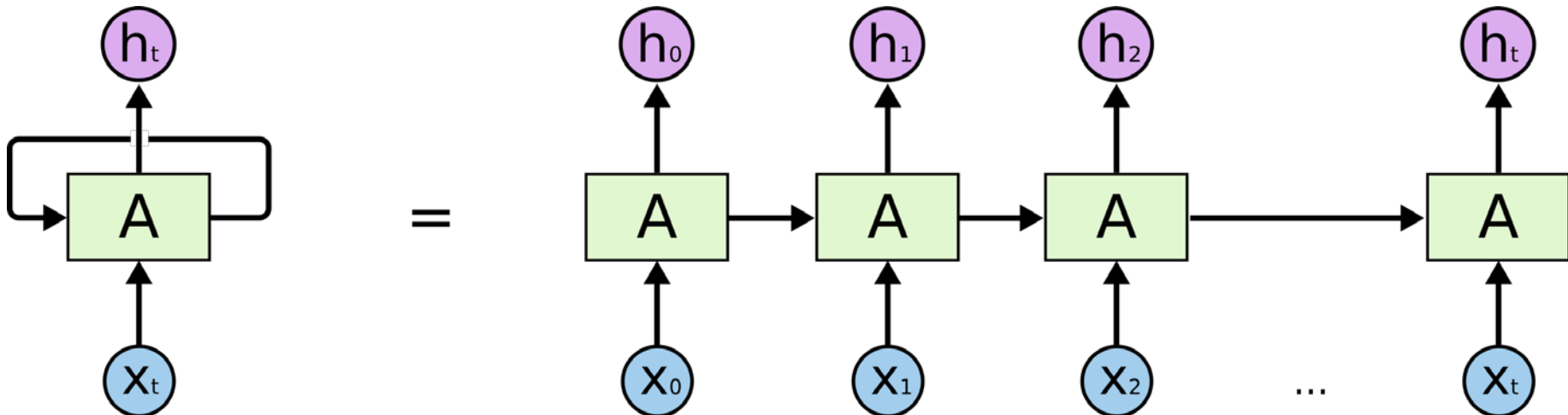
- Application



Recurrent Neural Networks



- Special: You should use something which occurred previously.
- Recurrent neural networks are networks with loops in them, allowing information to **persist**.
- Input: x_t
- Output: h_t
- A loop allows information to be passed from one step of the network to the next.

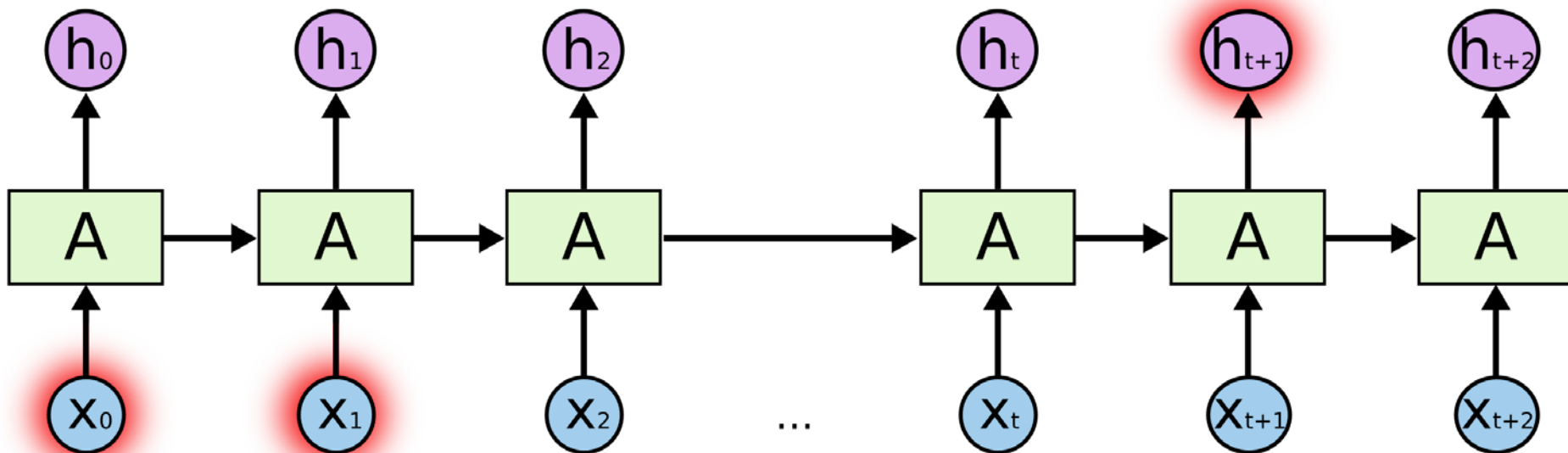


Recurrent Neural Networks have loops.
An unrolled recurrent neural network.

Disadvantage of RNNs



- Input: x_t , where is the heat value of a certain block on the map at the time $t = 0, 1, 2, \dots$
- Output: h_t , where is the prediction heat value of a certain block on the map at the time $t = 0, 1, 2, \dots$
- Disadvantage: As the gap grows the relevant information and the point, RNNs become unable to learn to connect the information.

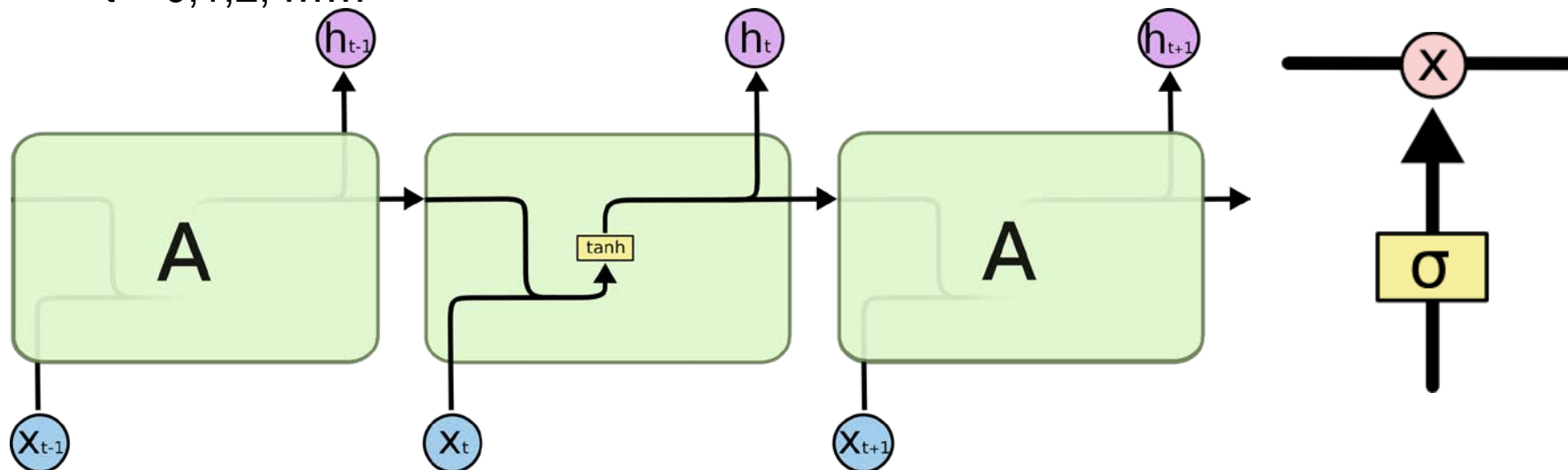


RNN-longtermdependencies

LSTM Networks



- LSTM -- Long Short Term Memory networks
- The key to LSTMs is the cell state. It runs straight down the entire chain, with only some minor linear interactions.
- Input: x_t , where is the heat value of a certain block on the map at the time $t = 0, 1, 2, \dots$
- Output: h_t , where is the heat value of a certain block on the map at the time $t = 0, 1, 2, \dots$



The repeating module in a standard RNN contains a single layer.

Gates are a way to optionally let information through.

An LSTM has three of these gates, to protect and control the cell state.

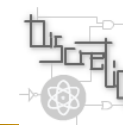
Agenda

- Analysis

- Visualization

- What is LSTM?

- Application



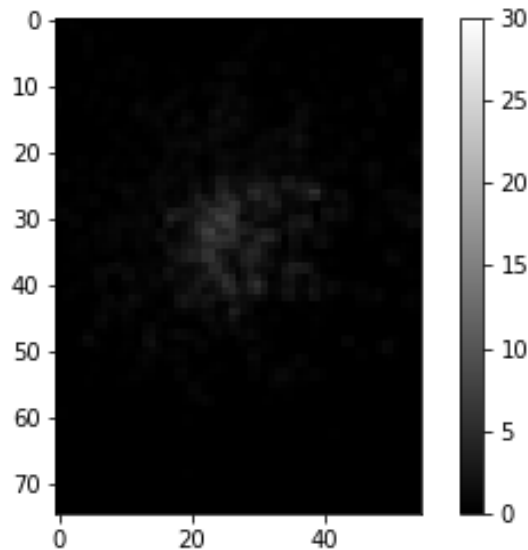
上海科技大学
ShanghaiTech University



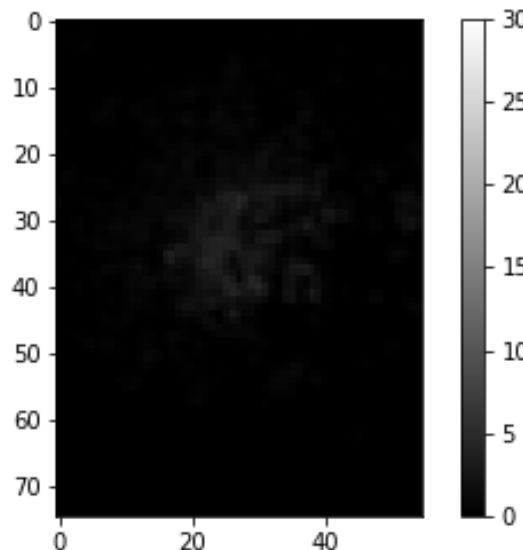
Relate DiDi to LSTM



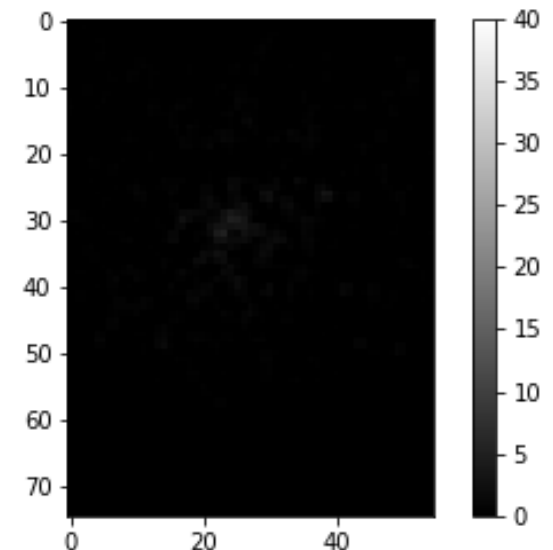
- Since in a short time period the tour is related to themselves which is in the last short time period or at the last seconds, we can think it as a LSTM networks.
- Assuming that the place where we start from exists demands and the place where we get off exists supply.
- We can use this to predict the gap of supplies and demands in a short future time.



The real grey-scale map of the gap between supply and demands for **future 10 minutes**.

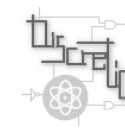


The prediction grey-scale map of the gap between supply and demands for **future 10 minutes**.

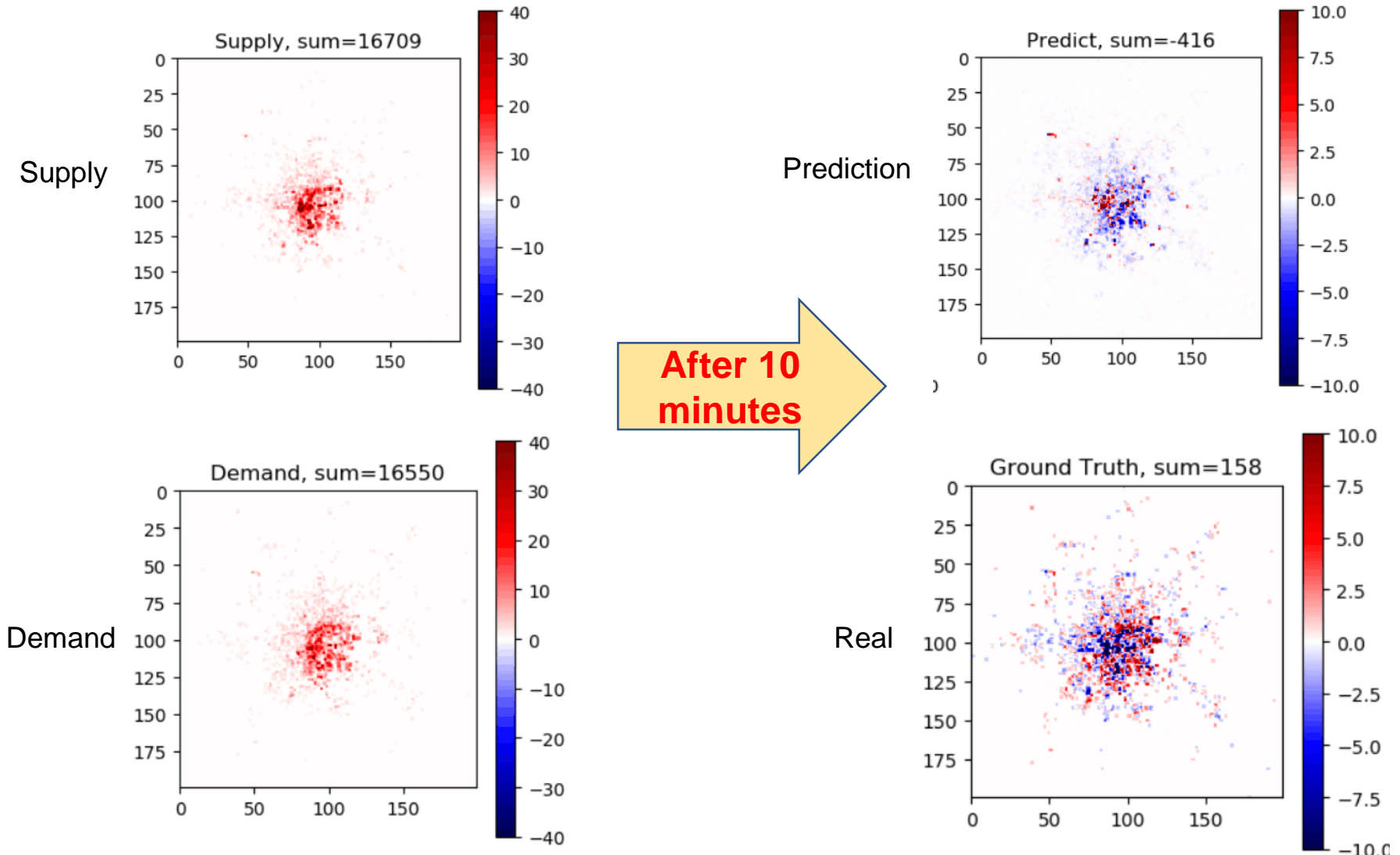


The difference of real and prediction of grey-scale map of the gap between supply and demands for **future 10 minutes**.

Relate DiDi to LSTM



- We can use this to predict the gap of supplies and demands in a short future time.
- It can be seen that the prediction **is very close to** the fact.



Application–The Optimal Place

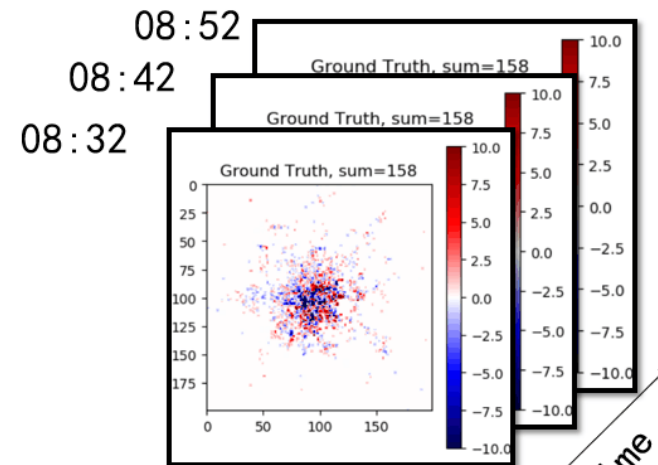


上海科技大学
ShanghaiTech University

- You may meet: long waiting time, failure to taxi, rush hour.....
- Use the previous prediction to help user or driver to choose the best place.
- Input: Time, latitude, longitude.
- Output: Assemble optimal route.

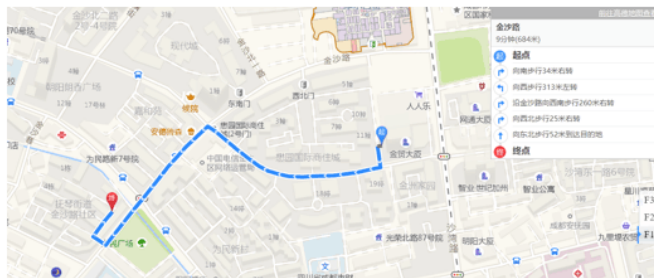
Input

(2016-11-01 08:35:15, 30.69, 104.05)
(Time, latitude, longitude)

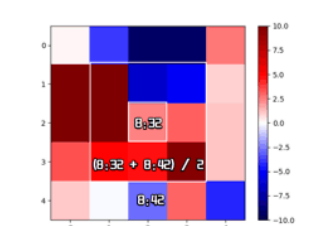
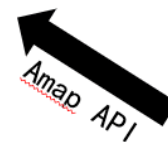


grasp s&d gap

Time



Assemble optimal route



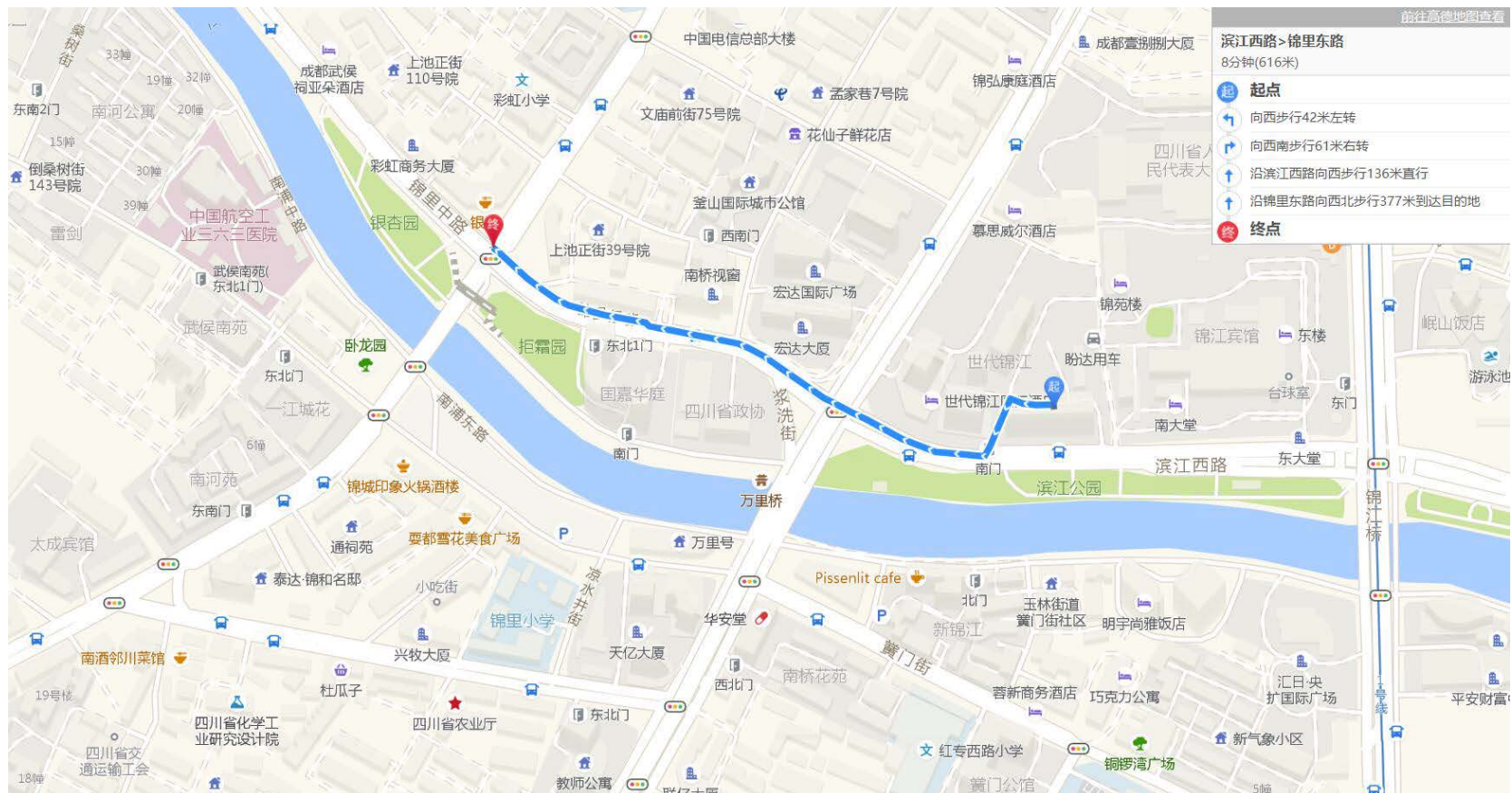
choose optimal location

Application–The Optimal Place



上海科技大学
ShanghaiTech University

- You may meet: long waiting time, failure to taxi, rush hour.....
- Use the previous prediction to help user or driver to choose the best place.
- Input: Time, latitude, longitude.
- Output: Assemble optimal route.



Application–The Optimal Place

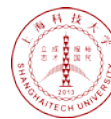


上海科技大学
ShanghaiTech University

We use an **APP** to realize it!



Application-The Optimal Place



上海科技大学
ShanghaiTech University

DiDi Analyst

反馈

帮助



Q7
技术部

Dashboard

订单类型

产品介绍

昨日订单数量

215,028

较前日 ↑ +418 近7天 ↑ +632

昨日活跃司机数

19,164

较前日 ↓ -14 近7天 ↑ +1057

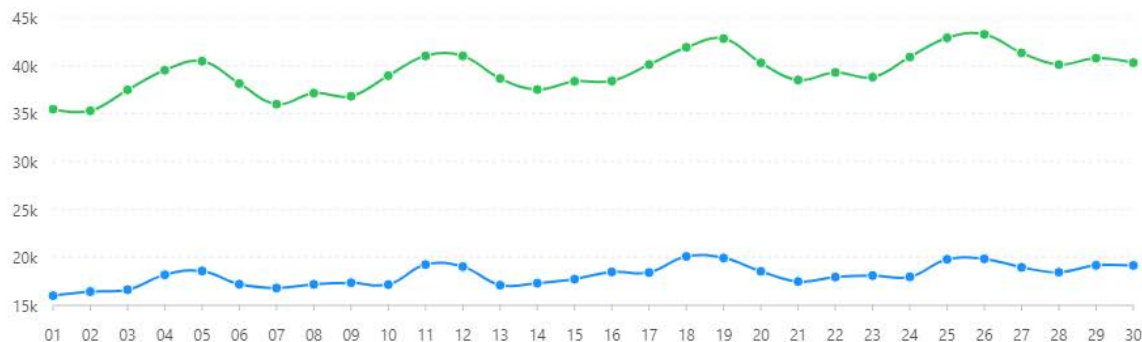
昨日司机数

40,297

较前日 ↓ -474 近7天 ↑ +1512

司机数走势

订单量趋势



昨日司机活跃度排行

司机ID	成交单数	
9f3256e119fbb17fba9c99e7bb04b7ec	32	<div></div>
4d106da5c8badb9e2ee0637dc41f2149	31	<div></div>
1448a1cf5586d57ee535522f0865c2f3	31	<div></div>
f892acd777b661b42618bbfc24e6cdda	31	<div></div>
0824f144aeee496ea83a27a269de8928	31	<div></div>

The Optimal Place based on **LSTM**

基于LSTM的上车地点推荐系统

系统使用长短期记忆网络对供需Gap进行学习，用户可自由选择打车地点，系统将为用户智能推荐最方便的上车地点

场景价值

- 减少用户候车时间
- 增加用户打车成功率

所用技术

- LSTM
- POI大数据

Input
(2016-11-01 08:35:15, 30.49, 104.05)
(Time, latitude, longitude)

grasp s&d gap

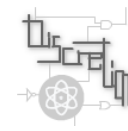
Assemble optimal route

choose optimal location

立即体验



Video: How to use App to get the best place.



Statistics and Prediction on DiDi Traffic Data Based on Neural Network

Discretion Team

Data Structure Project, 2018 Spring

ShanghaiTech University

Yintao Xu

Ruiqi Liu

Qi Qin

Haomin Shi

Zilin Si

Keyi Yuan

June.24th, 2018