# Where to Find My Next Passenger?

Jing Yuan[1] Yu Zheng[2] Liuhang Zhang[1] Guangzhong Sun[1]

[1]University of Science and Technology of China
[2]Microsoft Research Asia

September 19, 2011

# Motivation

- Taxis in big cities (103,000 in Mexico, 67,000+ in Beijing)
- Problems brought by cruising taxis: gas, time, profit

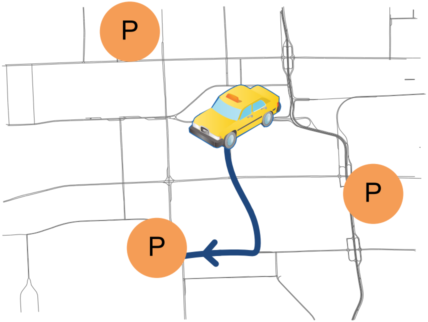
# Motivation

- Taxis in big cities (103,000 in Mexico, 67,000+ in Beijing)
- Problems brought by cruising taxis: gas, time, profit, traffic jams, energy, air pollution

# Motivation

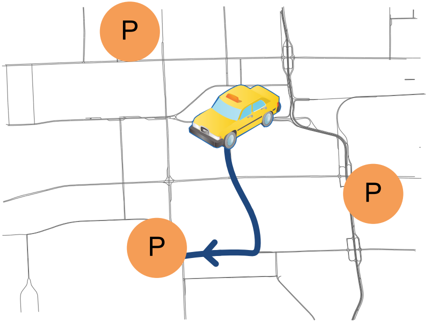
- Taxis in big cities (103,000 in Mexico, 67,000+ in Beijing)
- Problems brought by cruising taxis: gas, time, profit, traffic jams, energy, air pollution
- Passengers are still hard to find a vacant taxi sometimes

# Recommender Scenario

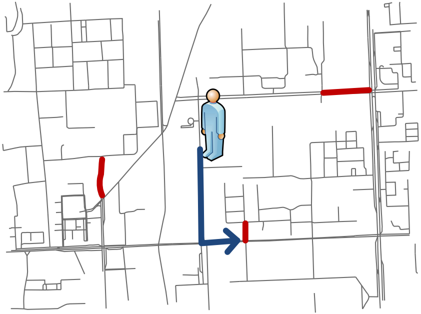



A) Taxi recommender

# Recommender Scenario

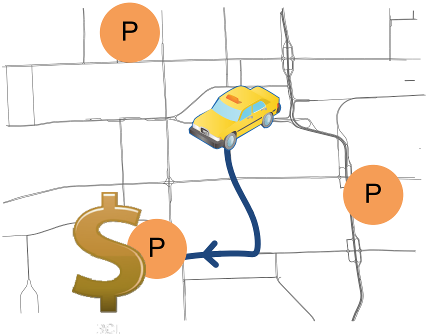



A) Taxi recommender

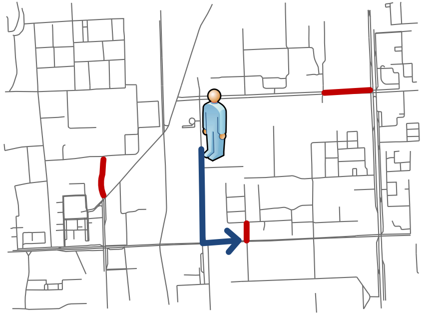


B) Passenger recommender

# Recommender Scenario

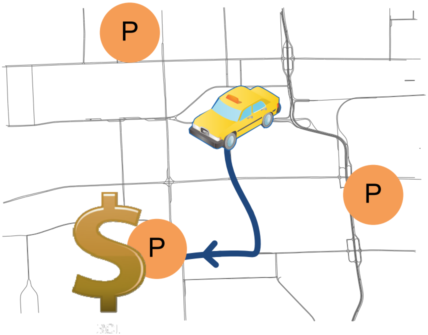



A) Taxi recommender

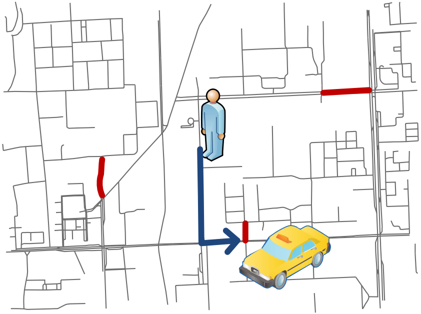


B) Passenger recommender

# Recommender Scenario

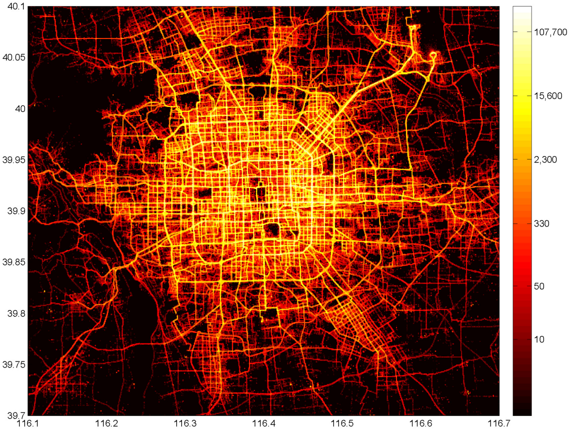



A) Taxi recommender



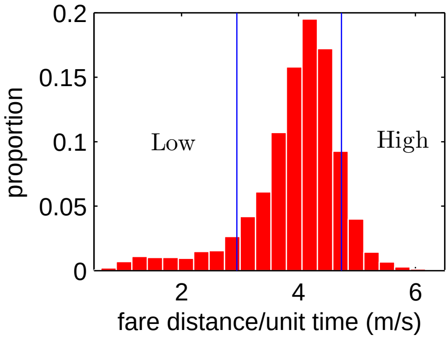
B) Passenger recommender

# Data

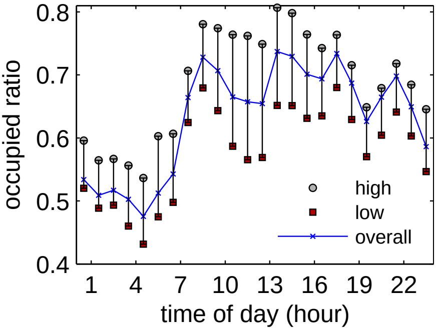



- Beijing Taxi Trajectories
  - 33,000 taixs in 3 month
  - Total distance: 400M km
  - Total number of points: 790M
  - Average sampling interval: 3.1 minutes, 600 meters
- Beijing Road Network
  - 106,579 road nodes
  - 141,380 road segments

# Profit-variant taxi drivers




(a) Distribution of profit




(b) Occupied ratio during a day

Figure 1: Statistics on the profit distribution and occupied ratio

# Cruise More, Earn More?

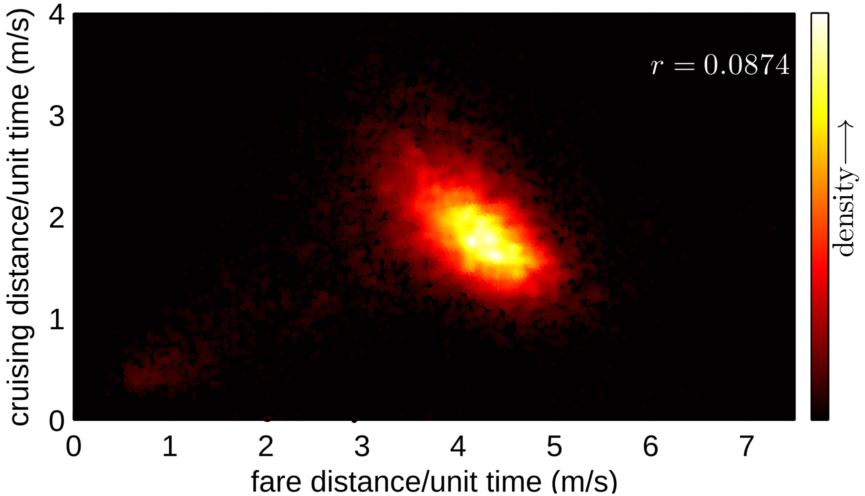
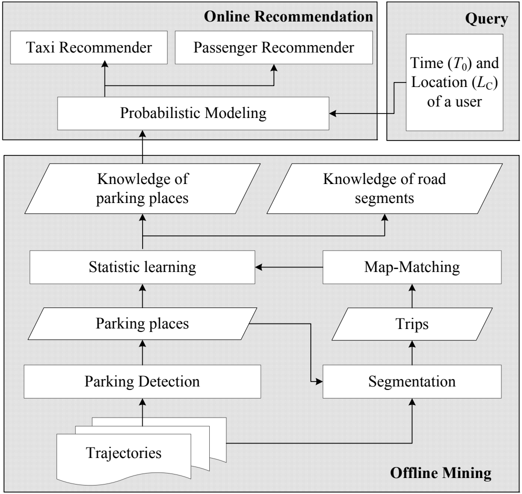
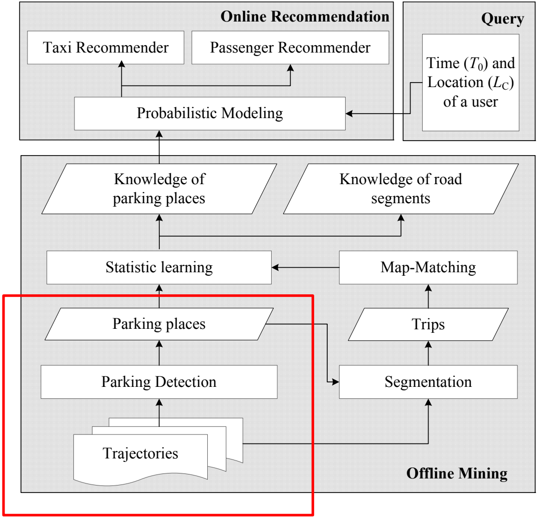



Figure 2: Density scatter of cruising distance/unit time w.r.t. profit

# System overview



# System overview

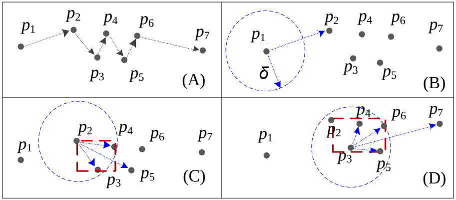


# Parking Place Detection

Parking Place: the places where the taxis frequently wait for passengers. (not a parking slot).

- Candidates Generation
- Filtering
- Density-Based Clustering

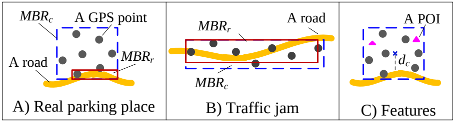
- Candidates Generation
  A group of points satisfying $\delta, \tau$; connect them if overlap exists




- Filtering
- Density-Based Clustering

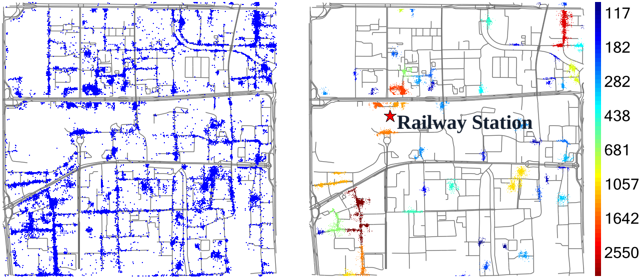
# Parking Place Detection

- Candidates Generation
- Filtering
  Distinguished from traffic jams (bagging classifier)
  features used: spatial-temporal($d_c$,MBR...), POI, ...

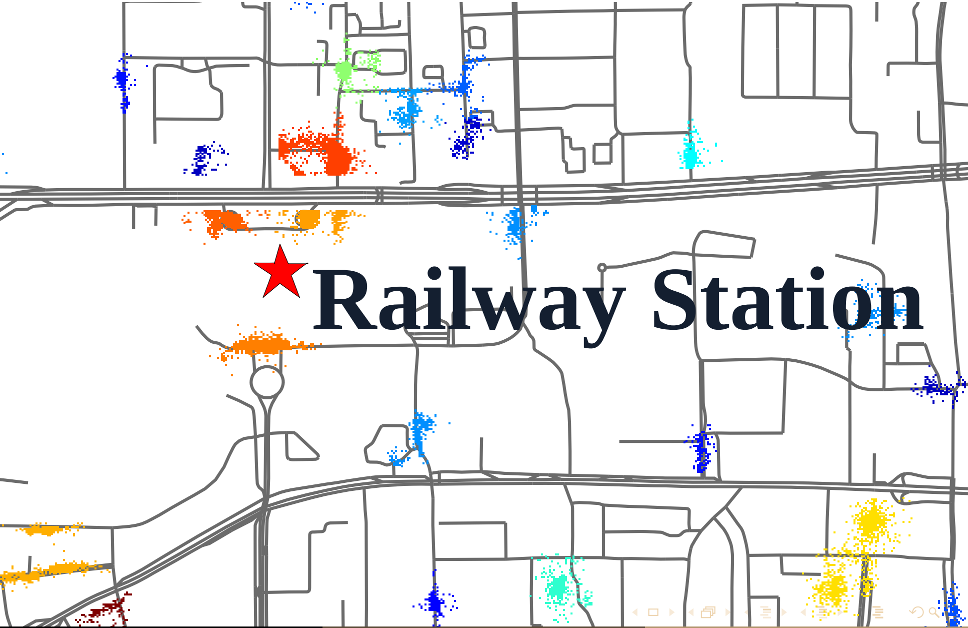



- Density-Based Clustering

- Candidates Generation
- Filtering
- Density-Based Clustering
  Aggregate the candidates belonging to a single parking place



# Parking Place Detection



A “good” parking place (to go towards):

- the probability to pick up a passenger ⇑ (Possibility)
- the expected duration from $T_0$ to the time the next passenger is picked up ⇓ (Cost)
- the distance/duration of the next trip ⇑ (Benefit)

# Taxi Recommender

A "good" parking place (to go towards):

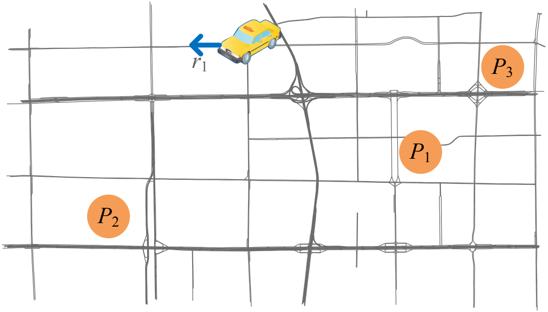
- the probability to pick up a passenger $\Uparrow$ (Possibility)
- the expected duration from $T_0$ to the time the next passenger is picked up $\Downarrow$ (Cost)
- the distance/duration of the next trip $\Uparrow$ (Benefit)

# Taxi Recommender

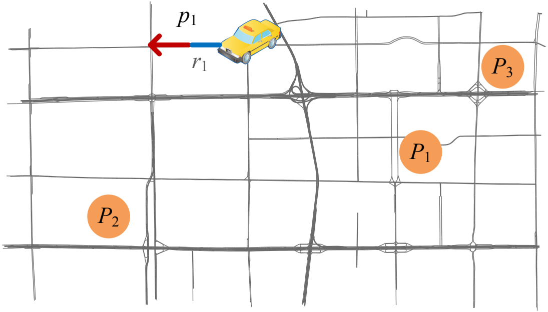
A “good” parking place (to go towards):

- the probability to pick up a passenger $\Uparrow$ (Possibility)
- the expected duration from $T_0$ to the time the next passenger is picked up $\Downarrow$ (Cost)
- the distance/duration of the next trip $\Uparrow$ (Benefit)

# Probability



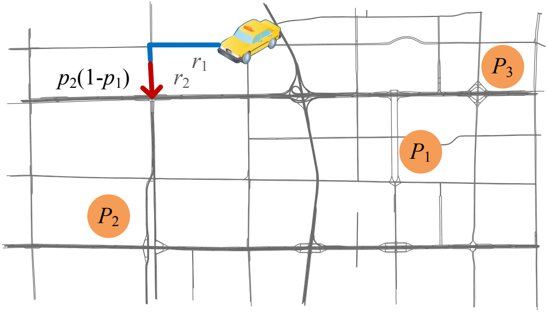
# Probability




Situation 1: Pick up during the route at $r_1$

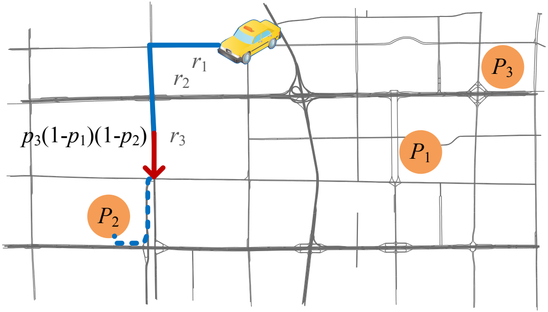
| | |
|---|---|
| $r_i$ | road segment $i$ |
| $t_i$ | travel time from $r_1$ to $r_i$ |
| $p_i$ | the probability that a taxi picks up a passenger at $r_i$ (at time $T_0 + t_i$) |

# Probability



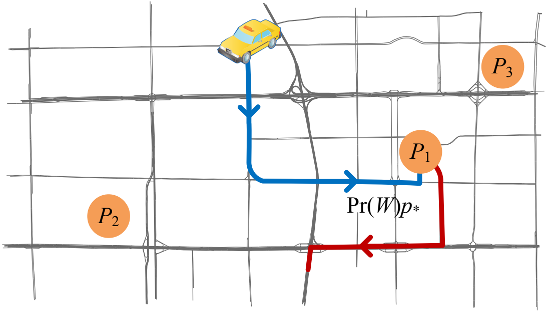

Situation 1: Pick up during the route at $r_2$

| | |
|---|---|
| $r_i$ | road segment $i$ |
| $t_i$ | travel time from $r_1$ to $r_i$ |
| $p_i$ | the probability that a taxi picks up a passenger at $r_i$ (at time $T_0 + t_i$) |



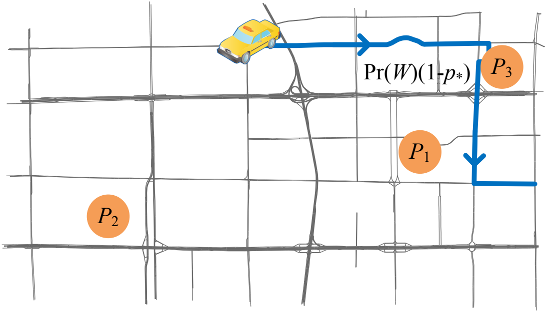

Situation 1: Pick up during the route at $r_3$

| | |
|---|---|
| $r_i$ | road segment $i$ |
| $t_i$ | travel time from $r_1$ to $r_i$ |
| $p_i$ | the probability that a taxi picks up a passenger at $r_i$ (at time $T_0 + t_i$) |




Situation 2: Pick up at a parking place

| | |
|---|---|
| $W$ | the event that a taxi waits at a parking place |
| $t_i$ | travel time from $r_1$ to $r_i$ |
| $p_*$ | the probability that a taxi picks up a passenger at a parking place (at time $T_0 + t_n$) |




Situation 3: Fail to pick up a passenger

| | |
|---|---|
| $W$ | the event that a taxi waits at a parking place |
| $t_i$ | travel time from $r_1$ to $r_i$ |
| $p_*$ | the probability that a taxi picks up a passenger at a parking place (at time $T_0 + t_n$) |

# Cost and Benefit Analysis

- Duration before the next trip $T$

$$\boxed{\begin{aligned} &\mathbf{E}[T|S] \\ =&\mathbf{E}[T_R|S] + \mathbf{E}[T_P|S] \\ =&\frac{\sum_{i=1}^{n} t_i \Pr(S_i) + t_n \Pr(S_{n+1}) + \Pr(W) \sum_{j=1}^{m} p_*^j t_j^*}{\Pr(S)}. \end{aligned}} \quad (1)$$

- Distance of the next trip $D_N$
- Duration of the next trip $T_N$

# Recommendation Strategies

1. Taxi Recommender

   S1. $Topk_{\max}\{\mathbf{E}[D_N|S]/\mathbf{E}[T+T_N|S] : \Pr(S) > P_\theta\}$.
   most profitable, given a probability guarantee.

   S2. $Topk_{\min}\{\mathbf{E}[T|S] : \Pr(S) > P_\theta, D_N > D_\theta\}$.
   fastest to find a passenger, given probability and distance guarantee

   S3. $Topk_{\max}\{\Pr(S) : \mathbf{E}[D_N|S]/\mathbf{E}[T+T_N|S] > F_\theta\}$.
   most likely to find a passenger, given profit guarantee

   S4. …

2. Passenger Recommender

$$r = \underset{r\in\Omega}{\operatorname{argmax}} \Pr(C; r|t).$$

$\Omega$: search space within a walking distance

# Recommendation Strategies

1. Taxi Recommender
   S1. $Topk_{\max}\{\mathbf{E}[D_N|S]/\mathbf{E}[T+T_N|S] : \Pr(S) > P_\theta\}$.
   most profitable, given a probability guarantee.
   S2. $Topk_{\min}\{\mathbf{E}[T|S] : \Pr(S) > P_\theta, D_N > D_\theta\}$.
   fastest to find a passenger, given probability and distance guarantee
   S3. $Topk_{\max}\{\Pr(S) : \mathbf{E}[D_N|S]/\mathbf{E}[T+T_N|S] > F_\theta\}$.
   most likely to find a passenger, given profit guarantee
   S4. …
2. Passenger Recommender

$$r = \operatorname*{argmax}_{r \in \Omega} \Pr(C; r|t).$$

$\Omega$: search space within a walking distance

# Recommendation Strategies

1. Taxi Recommender
   - S1. $Topk_{\max}\{\mathbf{E}[D_N|S]/\mathbf{E}[T+T_N|S] : \Pr(S) > P_\theta\}$.
     most profitable, given a probability guarantee.
   - S2. $Topk_{\min}\{\mathbf{E}[T|S] : \Pr(S) > P_\theta, D_N > D_\theta\}$.
     fastest to find a passenger, given probability and distance guarantee
   - S3. $Topk_{\max}\{\Pr(S) : \mathbf{E}[D_N|S]/\mathbf{E}[T+T_N|S] > F_\theta\}$.
     most likely to find a passenger, given profit guarantee
   - S4. …
2. Passenger Recommender

$$r = \operatorname*{argmax}_{r \in \Omega} \Pr(C; r|t).$$

$\Omega$: search space within a walking distance

# Recommendation Strategies

1. Taxi Recommender

   S1. $Topk_{\max}\{\mathbf{E}[D_N|S]/\mathbf{E}[T+T_N|S] : \Pr(S) > P_\theta\}$.
   most profitable, given a probability guarantee.

   S2. $Topk_{\min}\{\mathbf{E}[T|S] : \Pr(S) > P_\theta, D_N > D_\theta\}$.
   fastest to find a passenger, given probability and distance guarantee

   S3. $Topk_{\max}\{\Pr(S) : \mathbf{E}[D_N|S]/\mathbf{E}[T+T_N|S] > F_\theta\}$.
   most likely to find a passenger, given profit guarantee

   S4. ...

2. Passenger Recommender

$$r = \operatorname*{argmax}_{r\in\Omega} \Pr(C; r|t).$$

$\Omega$: search space within a walking distance

# Recommendation Strategies

1. Taxi Recommender

   S1. $Topk_{\max}\{\mathbf{E}[D_N|S]/\mathbf{E}[T + T_N|S] : \Pr(S) > P_\theta\}$.
   most profitable, given a probability guarantee.

   S2. $Topk_{\min}\{\mathbf{E}[T|S] : \Pr(S) > P_\theta, D_N > D_\theta\}$.
   fastest to find a passenger, given probability and distance guarantee

   S3. $Topk_{\max}\{\Pr(S) : \mathbf{E}[D_N|S]/\mathbf{E}[T + T_N|S] > F_\theta\}$.
   most likely to find a passenger, given profit guarantee

   S4. …

2. Passenger Recommender

$$r = \underset{r \in \Omega}{\operatorname{argmax}} \Pr(C; r|t).$$

$\Omega$: search space within a walking distance

# Recommendation Strategies

1. Taxi Recommender
   - S1. $Topk_{\max}\{\mathbf{E}[D_N|S]/\mathbf{E}[T+T_N|S] : \Pr(S) > P_\theta\}$.
     most profitable, given a probability guarantee.
   - S2. $Topk_{\min}\{\mathbf{E}[T|S] : \Pr(S) > P_\theta, D_N > D_\theta\}$.
     fastest to find a passenger, given probability and distance guarantee
   - S3. $Topk_{\max}\{\Pr(S) : \mathbf{E}[D_N|S]/\mathbf{E}[T+T_N|S] > F_\theta\}$.
     most likely to find a passenger, given profit guarantee
   - S4. ...
2. Passenger Recommender

$$r = \operatorname*{argmax}_{r\in\Omega} \Pr(C; r|t).$$

$\Omega$: search space within a walking distance

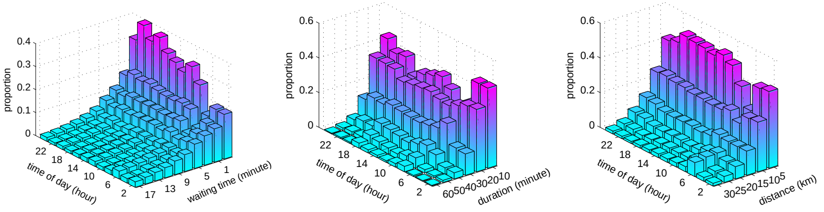
# Evaluation on Parking Place Detection

- Key issue: traffic jams vs. parking places

| Features | Precision | Recall |
|---|---|---|
| Spatial | 0.695 | 0.670 |
| Spatial+POI | 0.716 | 0.696 |
| Spatial+POI+Collaborative | 0.725 | 0.706 |
| Spatial+POI+Collaborative+Temporal | 0.909 | 0.889 |

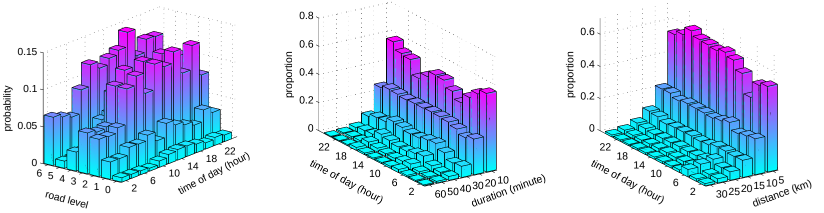
Table 1: Results of parking place filtering

# Evaluation on Knowledge Learning




(a) waiting time (b) duration of the first trip (c) distance of the first trip

Figure 3: Distribution in parking places (overall)




(a) $\Pr(C \rightsquigarrow O)$ (b) duration of the first trip (c) distance of the first trip

Figure 4: Statistics results of road segments (overall)

# Evaluation on Online Recommendation

- Precision (#hits/#recommendations) and Recall (#parking places the drivers actually go to/#suggested parking places)
- NDCG@k
- RME for the hit parking places on $T$, $T_N$ and $D_N$.

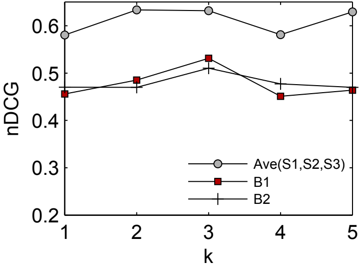



Figure 5: nDCG

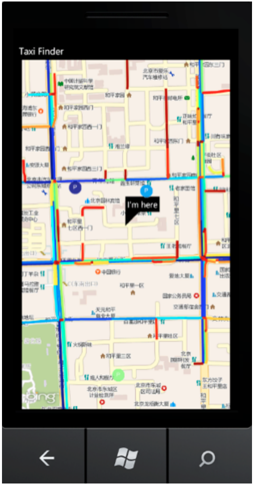
| | S1 | S2 | S3 | B1 | B2 |
|---|---|---|---|---|---|
| Precision | 0.63 | 0.66 | 0.67 | 0.60 | 0.61 |
| Recall | 0.59 | 0.65 | 0.64 | 0.57 | 0.52 |
| $RME(T)$ | | | 0.15 | | |
| $RME(D_N)$ | | | 0.02 | | |
| $RME(T_N)$ | | | 0.03 | | |

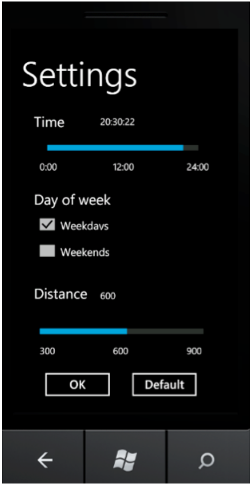
Table 2: RME, precision and recall

# Screenshot of Passenger Recommender

# Screenshot of Taxi Recommender

# Windows Phone 7 APP

# Next Step

- Waiting time modeling for passenger recommender
- Queueing models for parking places
- More in-the-field study

# Thanks!

Jing Yuan
yuanjing@mail.ustc.edu.cn