Where to Find My Next Passenger?

Jing Yuan\textsuperscript{1}  Yu Zheng\textsuperscript{2}  Liuhang Zhang\textsuperscript{1}  Guangzhong Sun\textsuperscript{1}

\textsuperscript{1}University of Science and Technology of China
\textsuperscript{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
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 taxis in 3 months
- 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

\[ r = 0.0874 \]
System overview

Online Recommendation
- Taxi Recommender
- Passenger Recommender
- Probabilistic Modeling

Query
- Time ($T_0$) and Location ($L_C$) of a user

Knowledge of parking places
- Statistic learning
- Parking places
- Parking Detection

Knowledge of road segments
- Map-Matching
- Trips
- Segmentation

Offline Mining
- Trajectories

Jing Yuan et al. (USTC, MSRA)
System overview

- **Online Recommendation**
  - Taxi Recommender
  - Passenger Recommender
  - Probabilistic Modeling

- **Query**
  - Time ($T_0$) and Location ($L_C$) of a user

- **Offline Mining**
  - Knowledge of parking places
  - Knowledge of road segments
  - Statistic learning
  - Map-Matching
  - Trips
  - Segmentation

- **Trajectories**
  - Parking Detection
  - Parking places
  - Trajectories
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**

---

Jing Yuan et al. (USTC,MSRA)

Where to Find My Next Passenger?

September 19, 2011 8 / 19
Parking Place Detection

- Candidates Generation
- Filtering
- Density-Based Clustering

Aggregate the candidates belonging to a single parking place
Parking Place Detection

Candidates Generation ▶ Filtering ▶ Density-Based Clustering

Railway Station

Jing Yuan et al. (USTC, MSRA) Where to Find My Next Passenger?

September 19, 2011 8 / 19
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)
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)
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)
Where to Find My Next Passenger?

- $r_i$: road segment
- $t_i$: 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 $r_1$

<table>
<thead>
<tr>
<th>$r_i$</th>
<th>road segment $i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_i$</td>
<td>travel time from $r_1$ to $r_i$</td>
</tr>
<tr>
<td>$p_i$</td>
<td>the probability that a taxi picks up a passenger at $r_i$ (at time $T_0 + t_i$)</td>
</tr>
</tbody>
</table>
Situation 1: Pick up during the route at $r_2$

<table>
<thead>
<tr>
<th>$r_i$</th>
<th>road segment $i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_i$</td>
<td>travel time from $r_1$ to $r_i$</td>
</tr>
<tr>
<td>$p_i$</td>
<td>the probability that a taxi picks up a passenger at $r_i$ (at time $T_0 + t_i$)</td>
</tr>
</tbody>
</table>
Situation 1: Pick up during the route at $r_3$

<table>
<thead>
<tr>
<th>$r_i$</th>
<th>road segment $i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_i$</td>
<td>travel time from $r_1$ to $r_i$</td>
</tr>
<tr>
<td>$p_i$</td>
<td>the probability that a taxi picks up a passenger at $r_i$ (at time $T_0 + t_i$)</td>
</tr>
</tbody>
</table>
Situation 2: Pick up at a parking place

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W$</td>
<td>the event that a taxi waits at a parking place</td>
</tr>
<tr>
<td>$t_i$</td>
<td>travel time from $r_1$ to $r_i$</td>
</tr>
<tr>
<td>$p_*$</td>
<td>the probability that a taxi picks up a passenger at a parking place (at time $T_0 + t_n$)</td>
</tr>
</tbody>
</table>
Situation 3: Fail to pick up a passenger

<table>
<thead>
<tr>
<th>$W$</th>
<th>the event that a taxi waits at a parking place</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_i$</td>
<td>travel time from $r_1$ to $r_i$</td>
</tr>
<tr>
<td>$p_*$</td>
<td>the probability that a taxi picks up a passenger at a parking place (at time $T_0 + t_n$)</td>
</tr>
</tbody>
</table>
Cost and Benefit Analysis

- Duration before the next trip \( T \)

\[
\mathbb{E}[T|S] = \mathbb{E}[T_R|S] + \mathbb{E}[T_P|S] \\
= \sum_{i=1}^{n} t_i \Pr(S_i) + t_n \Pr(S_{n+1}) + \Pr(W) \sum_{j=1}^{m} p^*_jt^*_j \\
= \frac{\sum_{i=1}^{n} t_i \Pr(S_i) + t_n \Pr(S_{n+1}) + \Pr(W) \sum_{j=1}^{m} p^*_jt^*_j}{\Pr(S)}.
\] (1)

- Distance of the next trip \( D_N \)
- Duration of the next trip \( T_N \)
Recommendation Strategies

1 Taxi Recommender

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

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

S3. \( Topk_{\text{max}} \{ \Pr(S) : \mathbb{E}[D_N|S]/\mathbb{E}[T + T_N|S] > F_\theta \} \).
   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. \( \text{Top}_{k_{\text{max}}} \{ \mathbb{E}[D_N | S] / \mathbb{E}[T + T_N | S] : \Pr(S) > P_\theta \} \).

most profitable, given a probability guarantee.

S2. \( \text{Top}_{k_{\text{min}}} \{ \mathbb{E}[T | S] : \Pr(S) > P_\theta, D_N > D_\theta \} \).

fastest to find a passenger, given probability and distance guarantee.

S3. \( \text{Top}_{k_{\text{max}}} \{ \Pr(S) : \mathbb{E}[D_N | S] / \mathbb{E}[T + T_N | S] > F_\theta \} \).

most likely to find a passenger, given profit guarantee.

S4. . . .

2 Passenger Recommender

\[ r = \arg \max_{r \in \Omega} \Pr(C; r | t). \]

\( \Omega \): search space within a walking distance
Recommendation Strategies

1. Taxi Recommender

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

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

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

   S4. ...

2. Passenger Recommender

   $$r = \arg\max_{r \in \Omega} \Pr(C; r|t).$$

   $\Omega$: search space within a walking distance
Taxi Recommender

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

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

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

S4. ...

Passenger Recommender

$$r = \arg\max_{r \in \Omega} \text{Pr}(C; r|t).$$

$\Omega$: search space within a walking distance.
Recommendation Strategies

1 Taxi Recommender

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

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

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

S4. ...

2 Passenger Recommender

$r = \arg\max_{r \in \Omega} \Pr(C; r|t).$

$\Omega$: search space within a walking distance
Recommendation Strategies

1 Taxi Recommender

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

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

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

S4. ...

2 Passenger Recommender

$r = \arg\max_{r \in \Omega} \Pr(C; r|t)$. 

$\Omega$: search space within a walking distance
Key issue: traffic jams vs. parking places

<table>
<thead>
<tr>
<th>Features</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial</td>
<td>0.695</td>
<td>0.670</td>
</tr>
<tr>
<td>Spatial+POI</td>
<td>0.716</td>
<td>0.696</td>
</tr>
<tr>
<td>Spatial+POI+Collaborative</td>
<td>0.725</td>
<td>0.706</td>
</tr>
<tr>
<td>Spatial+POI+Collaborative+Temporal</td>
<td>0.909</td>
<td>0.889</td>
</tr>
</tbody>
</table>

Table 1: Results of parking place filtering
Evaluation on Knowledge Learning

Figure 3: Distribution in parking places (overall)

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

Figure 4: Statistics results of road segments (overall)

(a) Pr(C⇝O) (b) duration of the first trip (c) distance of the first trip
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 $D_N$.

![Graph showing nDCG values](image)

### Table 2: RME, precision and recall

<table>
<thead>
<tr>
<th></th>
<th>S1</th>
<th>S2</th>
<th>S3</th>
<th>B1</th>
<th>B2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>0.63</td>
<td>0.66</td>
<td>0.67</td>
<td>0.60</td>
<td>0.61</td>
</tr>
<tr>
<td>Recall</td>
<td>0.59</td>
<td>0.65</td>
<td>0.64</td>
<td>0.57</td>
<td>0.52</td>
</tr>
<tr>
<td>$RME(T)$</td>
<td></td>
<td></td>
<td></td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>$RME(D_N)$</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$RME(T_N)$</td>
<td></td>
<td></td>
<td></td>
<td>0.03</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 5: nDCG**
Screenshot of Passenger Recommender

T-Finder

- Point out your location by right clicking on the map, or inputing below
- Location A

- Distance: 500 meters
- Time: Thursday 12:29

- Type: passenger

Results:

Top 3 parking places: (4 in total)
1. distance to parking place 1: 348m
2. distance to parking place 2: 264m
3. distance to parking place 3: 493m

Top 3 road segments:
1. distance to road 1: 245m
2. distance to road 2: 446m
3. distance to road 3: 458m

A parking place means a place where taxis wait for passengers.
A colored road segment means a road where you could find a taxi.
The possibility is indicated by the color.

Most likely to find a vacant taxi
Most impossible to find a vacant taxi
Screenshot of Taxi Recommender

T-Finder
- Point out your location by right clicking on the map, or inputting below
- Location A
- Distance: 500 meters
- Time: Thursday 12:29

Type: taxi driver
- Show all
- Clear all
- Find Taxis

Results:
Top 3 parking places: (5 in total)
1. distance to parking place 1: 419m
2. distance to parking place 2: 431m
3. distance to parking place 3: 350m

A parking place means a place where taxis wait for passengers.
A colored road segment means a road where you could find a taxi.
The possibility is indicated by the color:
- Most likely to find a vacant taxi
- Most impossible to find a vacant taxi
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