An Energy-Efficient Mobile Recommender System

Yong Ge

Rutgers, the State University of New Jersey
Outline

- Introduction
- Mobile Sequential Recommendation
- Recommending Point Generation
- LCP and SkyRoute Algorithms
- Experimental Results
- Conclusions
Application Background
Research Motivation

- Traditional recommender system
  - Prediction performance (MSE/RMSE)
  - Implicit/Explicit rating

- Mobile recommender system
  - Business success metrics
  - Location-based recommendation
Outline

- Introduction
- Mobile Sequential Recommendation
- Recommending Point Generation
- LCP and SkyRoute Algorithms
- Experimental Results
- Conclusions
The MSR Problem

Given: A set of potential pick-up points \( C \) with \( |C| = N \), a probability set \( \mathcal{P} = \{ P(C_1), P(C_2), \ldots, P(C_N) \} \), a directed sequence set \( \mathcal{R} \) with \( |\mathcal{R}| = M \) and the current position (PoCab) of a cab driver, who needs the service.

Objective: Recommending an optimal driving route \( \mathcal{R} \) (\( \mathcal{R} \in \mathcal{R} \)). The goal is to minimize the PTD:

\[
\min_{\mathcal{R}_i \in \mathcal{R}} \mathcal{F}(PoCab, \mathcal{R}_i, \mathcal{P}_{\mathcal{R}_i}) \tag{1}
\]

- \( \mathcal{P}_{\mathcal{R}_i} \): all probabilities of all pick-up points contained in \( \mathcal{R}_i \)
- PTD: Potential travel distance
An Example

PoCab -> C1 -> C4 or PoCab -> C4 -> C3 -> C2?
Two Challenges

- Mining reliable pick-up point with probability information
- Computation challenge to search the optimal route

Lemma 1. Given a set of pick-up points $C$, where $|C| = N$, $1 \leq L_{\overrightarrow{R_i}} \leq N$ and $\text{Cox}(\mathcal{F}) = 1$, the complexity of searching an optimal directed sequence from $\overrightarrow{R}$ is $O(N!)$
The MSR Problem with a Length Constraint

**Objective:** Recommending an optimal sequence \( \overrightarrow{R^L} (\overrightarrow{R^L} \in \overrightarrow{R}) \). The goal is to minimize the PTD:

\[
\min_{\overrightarrow{R^L} \in \overrightarrow{R}} \mathcal{F}(PoCab, \overrightarrow{R^L}, \overrightarrow{P_{R^L}})
\]

\( \overrightarrow{R^L} \) denote the recommended route with a length of \( L \)

**Computational complexity:** \( \mathcal{O}(N^L) \)
Outline

- Introduction
- Mobile Sequential Recommendation
- Recommending Point Generation
- LCP and SkyRoute Algorithms
- Experimental Results
- Conclusions
Recommending Point Generation

- High-performance Drivers
  - Sufficient Driving Hours
  - High Occupancy Rate

- Clustering based on Driving Distance
  - Clustering close pick-up points into one pick-up cluster

- Probability Calculation
  - Ratio of Pick-up Events Happening when cab travels across pick-up clusters
Outline

- Introduction
- Mobile Sequential Recommendation
- Recommending Point Generation
- LCP and SkyRoute Algorithms
- Experimental Results
- Conclusions
Potential Travel Distance Function: \( \mathcal{F}(PoCab, R_i^L, P_{R_i^L}) \)

PoCab: the position of a cab
C1: one pick-up point
P(C1): the probability of pick-up event
D1: the driving distance

A Recommended Driving Route.

Two Vectors:
\[
\begin{align*}
\mathcal{D}_{R_i^L} &= \langle D_1, (D_1 + D_2), (D_1 + D_2 + D_3), (D_1 + D_2 + D_3 + D_4), D_\infty \rangle \\
\mathcal{P}_{R_i^L} &= \langle P_1, P(C_1) \cdot P(C_2), P(C_1) \cdot P(C_2) \cdot P(C_3), P(C_1) \cdot P(C_2) \cdot P(C_3) \cdot P(C_4), \frac{1}{P(C_4)} \rangle
\end{align*}
\]

PTD Function:
\[
\mathcal{F}(PoCab, R_i^L, P_{R_i^L}) = \mathcal{D}_{R_i^L} \cdot \mathcal{P}_{R_i^L}
\]

where \( \cdot \) is the dot product
Potential Travel Distance Function: \[ \mathcal{F}(PoC\text{ab}, R_i^C, P_{R_i^C}) \]

**PoCab:** the position of a cab

**C1:** one pick-up point

**P(C1):** the probability of pick-up event

**D1:** the driving distance

**A Recommended Driving Route.**

**One Vectors:**
\[ \mathcal{D}\mathcal{P} = \langle D_1, \overline{P(C_1)}, D_2, \overline{P(C_2)}, D_3, \overline{P(C_3)}, D_4, \overline{P(C_4)} \rangle \]

Generally, for \( R_i^C \), \[ \mathcal{D}\mathcal{P} = \langle DP_1, \ldots, DP_i, \ldots DP_{2C} \rangle \]

**PTD Function:**
\[ \mathcal{F}(PoC\text{ab}, R_i^C, P_{R_i^C}) = \mathcal{F}(\mathcal{D}\mathcal{P}) \]
Lemma 3. The Monotone Property of the PTD Function $\mathcal{F}$. The PTD Function $\mathcal{F}(\mathcal{DP})$ is strictly monotonically increasing with each attribute of vector $\mathcal{DP}$, which is a $2L$-dimensional vector.

A Recommended Driving Route.

\[ \mathcal{DP} = \langle D_1, P(C_1), D_2, \overline{P(C_2)}, D_3, \overline{P(C_3)}, D_4, \overline{P(C_4)} \rangle \]
**Route Dominance**

**Definition 1. Route Dominance.** A recommended driving route $\vec{R}^L$, associated with the vector $\vec{DP}$, dominates another route $\vec{\tilde{R}}^L$, associated with the vector $\vec{D\tilde{P}}$, iff $\exists 1 \leq l \leq 2L$, $DP_l < D\tilde{P}_l$ and $\forall 1 \leq l \leq 2L$, $DP_l \leq D\tilde{P}_l$. This can be denoted as $\vec{R}^L \Downarrow \vec{\tilde{R}}^L$.

By this definition, if a candidate route A is dominated by a candidate route B, A cannot be an optimal route.
**Definition 2. Constrained Sub-route Dominance.**
Consider that two sub-routes $\overrightarrow{R}_{sub}$ and $\overrightarrow{R'_{sub}}$ with an equal length (the number of pick-up points) and the same source and destination points. If the associated vector of $\overrightarrow{R}_{sub}$ dominates the associated vector of $\overrightarrow{R'_{sub}}$, then $\overrightarrow{R}_{sub}$ dominates $\overrightarrow{R'_{sub}}$, i.e. $\overrightarrow{R}_{sub} \models \overrightarrow{R'_{sub}}$.

Illustration: the Sub-route Dominance

\[ \overrightarrow{R}_{sub} \text{ is } C_2 \rightarrow C_3 \rightarrow C_4 \]
\[ \overrightarrow{R'_{sub}} \text{ is } C_2 \rightarrow C'_3 \rightarrow C_4 \]
\[ \mathcal{DP}_{sub} = \langle D_3, \overline{P(C_3)}, D_4, \overline{P(C_4)} \rangle \]
\[ \mathcal{DP'_{sub}} = \langle D'_3, \overline{P(C'_3)}, D'_4, \overline{P(C_4)} \rangle \]
**Proposition 1. LCP Pruning.** For two sub-routes $A$ and $B$ with a length $L$, which includes only pick-up points, if sub-route $A$ is dominated by sub-route $B$ under Definition 2, the candidate routes with a length $L$ which contain sub-route $A$ will be dominated and can be pruned in advance.

If $L = 3$ and $\overrightarrow{R}_{sub}$ dominates $\overrightarrow{R'}_{sub}$

Then $PoCab \rightarrow C_2 \rightarrow C_3 \rightarrow C_4$

dominates $PoCab \rightarrow C_2 \rightarrow C_3' \rightarrow C_4$

- Enumerate all sub-routes with length of $L$
- Prune dominated constrained sub-routes with length of $L$
- Once effort to prune search space offline
Definition 3. Skyline Route. A recommended driving route $\overrightarrow{R^L}$ is a skyline route iff $\forall \overrightarrow{R_i^L} \in \overrightarrow{R}$, $\overrightarrow{R_i^L}$ cannot dominate $\overrightarrow{R^L}$ by Definition 1. This is denoted as $\overrightarrow{R_i^L} \not\succ \overrightarrow{R^L}$.

Lemma 4. Joint Principle of Skyline Routes and the PTD Function $F$. The optimal driving route determined by the PTD function $F$ should be a skyline route. This is denoted as $\overrightarrow{R^L} \in \overrightarrow{R_{Skyline}}$

- First find the skyline routes
- Search the optimal driving route from the set of skyline routes
The SkyRoute Algorithm

- Traditional Skyline computing algorithms
  - time-consuming and large memory

- Pruning for SkyRoute
  - Prune some candidates including dominated sub-routes, at a very early stage
  - The search space will be significant reduced, since lots of candidates containing dominated sub-routes are discarded
  - Gradual effort to continue prune the search space
If \( \text{SubRoute } R_1 \) (PoCab\(\rightarrow\)C\(_1\)\(\rightarrow\)C\(_3\)) dominates \( \text{SubRoute } R_2 \) (PoCab\(\rightarrow\)C\(_2\)\(\rightarrow\)C\(_3\)),

by DEFINITION 2

Then Any candidate like

\((\text{PoCab}\rightarrow\text{C}_1\rightarrow\text{C}_3\rightarrow\text{C}_i\rightarrow\ldots\rightarrow\text{C}_L)\) dominates candidate as

\((\text{PoCab}\rightarrow\text{C}_2\rightarrow\text{C}_3\rightarrow\text{C}_i\rightarrow\ldots\rightarrow\text{C}_L),\)

where \( \text{C}_i\rightarrow\ldots\rightarrow\text{C}_L \) is one of all possible sub-routes
Outline

- Introduction
- Mobile Sequential Recommendation
- Recommending Point Generation
- LCP and SkyRoute Algorithms
- Experimental Results
- Conclusions
Experimental Setup

- **Data set**
  - Real data set: 500+ taxi drivers, 30 days, San Francisco Bay Area
  - Time period: 2-3PM, 6-7PM
  - 10 pick-up clusters
  - Synthetic data set: 10, 15, 20 Pick-up points

- **Platform**
  - Intel Core2 Quad Q8300 and 6.00GB RAM
  - Windows 7 Professional
  - Matlab2008a
An Illustration of Optimal Driving Route

- **6-7PM**

- **L=3:**
  \[P_{oC ab} \rightarrow C1 \rightarrow C3 \rightarrow C2\]

- **L=4:**
  \[P_{oC ab} \rightarrow C1 \rightarrow C3 \rightarrow C2 \rightarrow C7\]

- **L=5**
  \[P_{oC ab} \rightarrow C4 \rightarrow C1 \rightarrow C3 \rightarrow C2 \rightarrow C7\]
An Overall Comparison

A Comparison of Search Time

Some Acronyms:

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BFS:</td>
<td>Brute-Force Search</td>
</tr>
<tr>
<td>LCPS:</td>
<td>Search with LCPS</td>
</tr>
</tbody>
</table>
An Overall Comparison

The Pruning Effect

Pruning percentage: the number of pruned candidates divided by the number of all original candidates
A Comparison of Skyline Computing

(a) Comparisons on Synthetic Data (L=3)

(b) Comparisons on Real Data (6-7PM)
Case: Multiple Evaluation Functions

A Comparison of Search Time for Multiple Optimal Driving Routes

Five different evaluation functions
Outline

- Introduction
- Mobile Sequential Recommendation
- Recommending Point Generation
- LCP and SkyRoute Algorithms
- Experimental Results
- Conclusions
Conclusion Remarks

- An energy-efficient mobile recommender system
- Potential Travel Distance (PTD) Function
- Algorithms: $LCP$ and SkyRoute
Thank You!

Poster Session II & Demo Session
Date: Tuesday, July 27, 2010
Time: 5:45pm - 8:00pm
Location: Independence Center B, floor 1
Illustration of the *Circulating Mechanism*
The pseudo-code of SkyRoute Algorithm

**ALGORITHM** SkyRoute(\(\mathcal{C}, \mathcal{P}, \text{Dist}, L, \text{PoCab}\))

**Input:**
- \(\mathcal{C}\): set of cluster nodes with central positions
- \(\mathcal{P}\): probability set for all cluster nodes
- \(\text{Dist}\): pairwise drive distance matrix of cluster nodes
- \(L\): the length of suggested drive route
- \(\text{PoCab}\): the position of one empty cab

**Output:**
- \(\overrightarrow{R}_{\text{Skyline}}\): list of skyline drive routes.

**Online Processing**

1. Enumerate all candidate routes by connecting \(\text{PoCab}\) with each sub-route of \(R_{\text{sub}}\) obtained in step 10 during Offline Processing
2. for \(i = 2 : L - 1\)
3. Decide dominated sub-routes with \(i\)th intermediate cluster and prune the corresponding candidates by using proposition 2
4. Update the candidate set by filtering the pruned candidates in step 3
5. end for
6. Select the remained candidate routes with length of \(L\) from the loop above
7. Final typical skyline query to get \(\overrightarrow{R}_{\text{Skyline}}\) from those candidate routes in step 6

**Offline Processing(LCP)**

8. Enumerate all sub-routes with length of \(L\) from \(\mathcal{C}\)
9. Prune and maintain dominated Constrained Sub-routes with length of \(L\) using proposition 3
10. Maintain the remained non-dominated sub-routes with length of \(L\), denoted as \(R_{\text{sub}}\)

- First use LCP to prune some candidates
- Gradually prune candidates containing dominated sub-routes
### A Comparison of Search Time (Second) between BFS and LCPS

<table>
<thead>
<tr>
<th></th>
<th>10 Synthetic Pick-up Clusters</th>
<th></th>
<th>15 Synthetic Pick-up Clusters</th>
<th></th>
<th>Real Data (2-3PM)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \mathcal{L} = 3 )</td>
<td>( \mathcal{L} = 4 )</td>
<td>( \mathcal{L} = 5 )</td>
<td>( \mathcal{L} = 3 )</td>
<td>( \mathcal{L} = 4 )</td>
</tr>
<tr>
<td>BFS</td>
<td>0.051643</td>
<td>0.300211</td>
<td>2.000949</td>
<td>0.142254</td>
<td>1.925054</td>
</tr>
<tr>
<td>LCPS</td>
<td><strong>0.043750</strong></td>
<td><strong>0.165401</strong></td>
<td><strong>0.803290</strong></td>
<td><strong>0.095364</strong></td>
<td><strong>0.611193</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
An Overall Comparison

A Comparison of Search Time \( (L = 3) \) on the Synthetic Data set