An Energy-Efficient Mobile Recommender System

Yong Ge

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS

Outline

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS

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

Application Background

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS

Research Motivation

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS

- Traditional recommender system
  - Prediction performance (MSE/RMSE)
  - Implicit/Explicit rating
- Mobile recommender system
  - Business success metrics
  - Location-based recommendation

Outline

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS

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

Problem Formulation

The MSR Problem

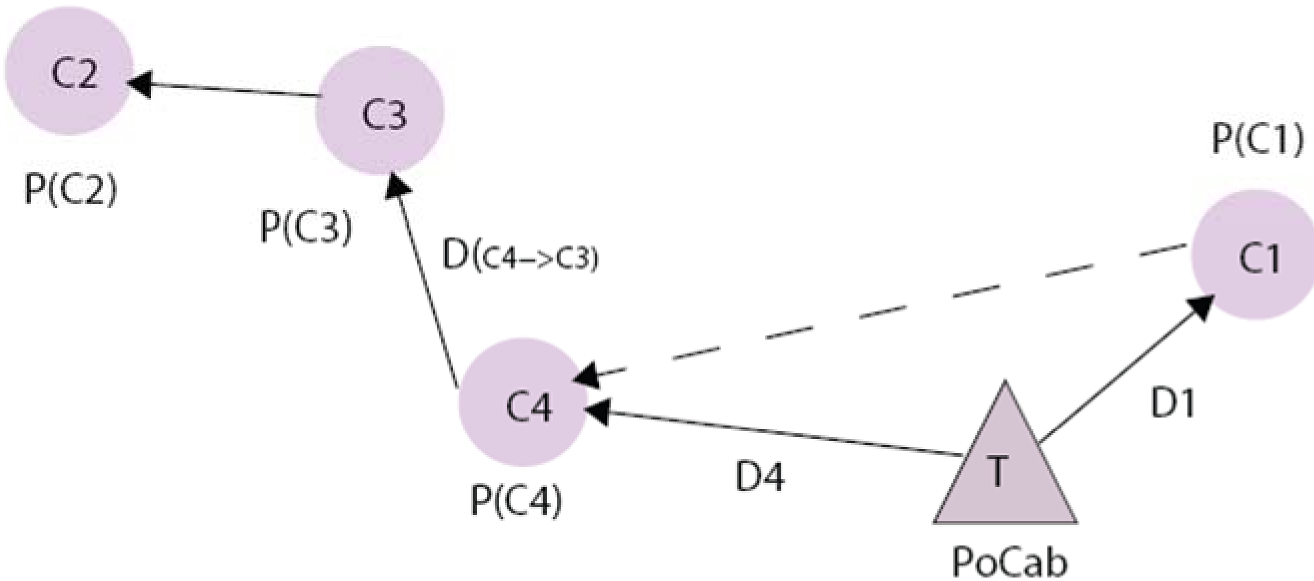
Given: A set of potential pick-up points $\mathcal{C}$ with $|\mathcal{C}| = N$, a probability set $\mathcal{P} = \{P(C_1), P(C_2), \cdots, P(C_N)\}$, a directed sequence set $\overrightarrow{\mathcal{R}}$ with $|\overrightarrow{\mathcal{R}}| = M$ and the current position ($PoCab$) of a cab driver, who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

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

$$\min_{\overrightarrow{R_i} \in \overrightarrow{\mathfrak{R}}} \mathcal{F}(PoCab, \overrightarrow{R_i}, \mathcal{P}_{\overrightarrow{R_i}}) \quad (1)$$

- $\mathcal{P}_{\overrightarrow{R_i}}$: all probabilities of all pick-up points contained in $\overrightarrow{R_i}$
- PTD: Potential travel distance

An Example



An Illustration Example.

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

Two Challenges

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS

- 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* $|\mathcal{C}| = N$, $1 \leq L_{\overrightarrow{R_i}} \leq N$ *and* $Cox(\mathcal{F}) = 1$, *the complexity of searching an optimal directed sequence from* $\overrightarrow{\mathcal{R}}$ *is* $\mathcal{O}(N!)$

MSR Problem with Constraints

The MSR Problem with a Length Constraint

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

$$\min_{\overrightarrow{R_i^{\mathcal{L}}} \in \overrightarrow{\mathfrak{R}}} \mathcal{F}(PoCab, \overrightarrow{R_i^{\mathcal{L}}}, \mathcal{P}_{\overrightarrow{R_i^{\mathcal{L}}}})$$

$\overrightarrow{R_i^{\mathcal{L}}}$ denote the recommended route with a length of $\mathcal{L}$

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

Outline

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS

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

Recommending Point Generation

THE STATE UNIVERSITY OF NEW JERSEY RUTGERS

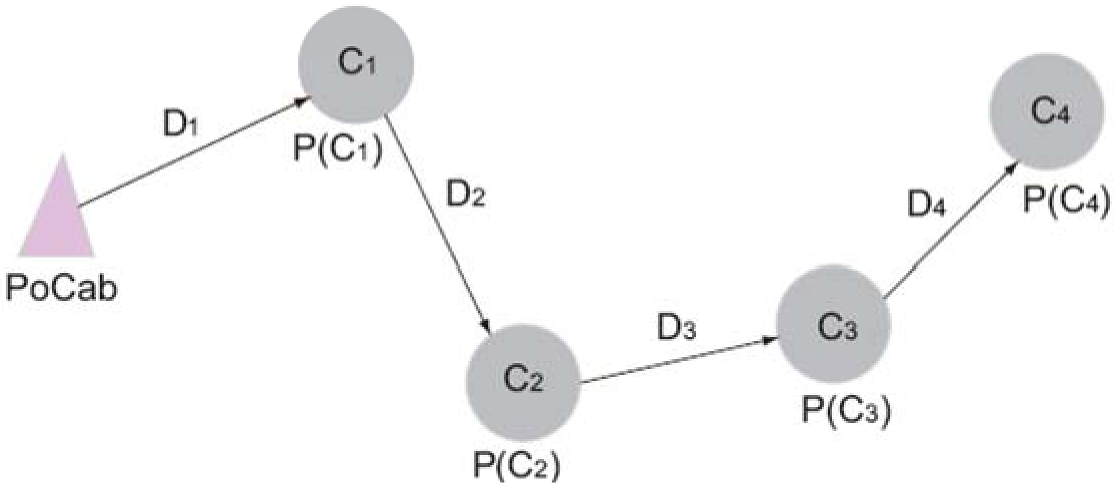
- High-performance Drivers
  - Sufficient Driving Hours
  - High Occupancy Rate
- Clustering based on Driving Distance
  - Clustering close pick-up points into one pick-up cluster
- Probability Calculation
  - Ratio of Pick-up Events Happening when cab travels across pick-up clusters

Outline

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS

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

Potential Travel Distance Function: $\mathcal{F}(PoCab, \overrightarrow{R_i^{\mathcal{L}}}, \mathcal{P}_{\overrightarrow{R_i^{\mathcal{L}}}})$



A Recommended Driving Route.

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

Two Vectors:

$$\mathcal{D}_{\overrightarrow{R^{\mathcal{L}}}} = \langle D_1, (D_1 + D_2), (D_1 + D_2 + D_3), (D_1 + D_2 + D_3 + D_4), D_\infty \rangle$$

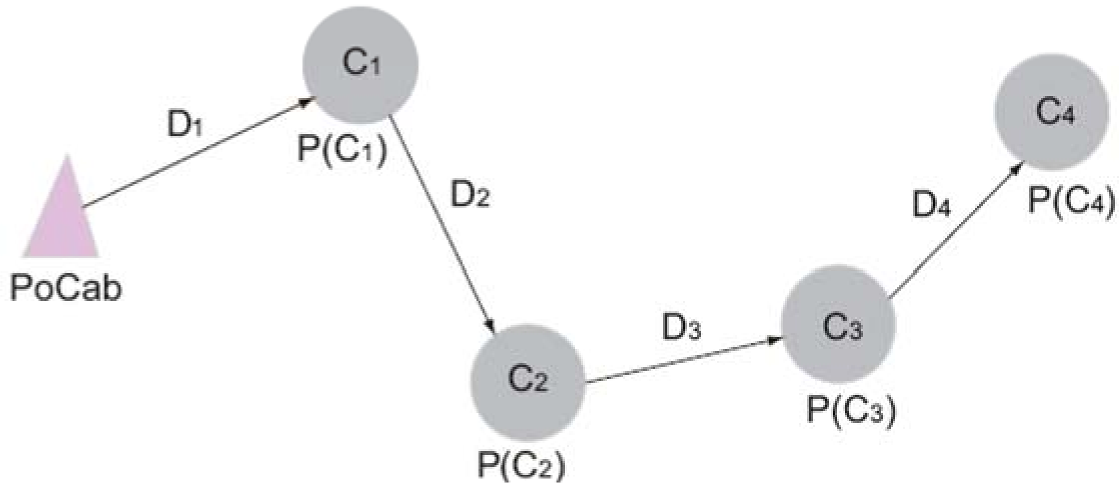
$$\mathcal{P}_{\overrightarrow{R^{\mathcal{L}}}} = \langle P_1, \overline{P(C_1)} \cdot P(C_2), \overline{P(C_1)} \cdot \overline{P(C_2)} \cdot P(C_3), \overline{P(C_1)} \cdot \overline{P(C_2)} \cdot \overline{P(C_3)} \cdot P(C_4), \overline{P(C_1)} \cdot \overline{P(C_2)} \cdot \overline{P(C_3)} \cdot \overline{P(C_4)} \rangle$$

PTD Function:

$$\mathcal{F}(PoCab, \overrightarrow{R_i^{\mathcal{L}}}, \mathcal{P}_{\overrightarrow{R_i^{\mathcal{L}}}}) = \mathcal{D}_{\overrightarrow{R^{\mathcal{L}}}} \cdot \mathcal{P}_{\overrightarrow{R^{\mathcal{L}}}}$$

where $\cdot$ is the dot product

Potential Travel Distance Function: $\mathcal{F}(PoCab, \overrightarrow{R_i^{\mathcal{L}}}, \mathcal{P}_{\overrightarrow{R_i^{\mathcal{L}}}})$



A Recommended Driving Route.

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

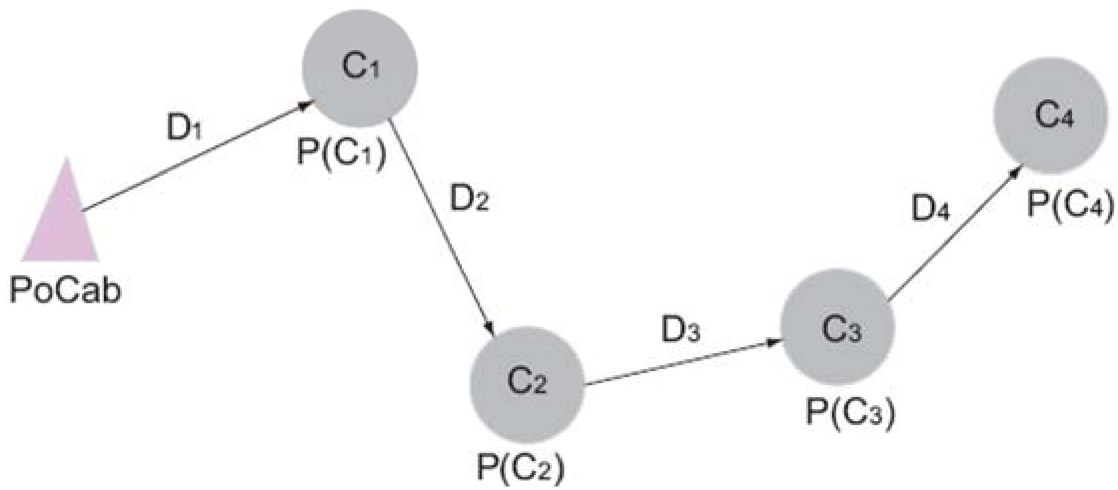
One Vectors: $\mathcal{DP} = \langle D_1, \overline{P(C_1)}, D_2, \overline{P(C_2)}, D_3, \overline{P(C_3)}, D_4, \overline{P(C_4)} \rangle$

Generally, for $\overrightarrow{R_i^{\mathcal{L}}}$, $\mathcal{DP} = \langle DP_1, \cdots, DP_l, \cdots DP_{2\mathcal{L}} \rangle$

PTD Function: $\mathcal{F}(PoCab, \overrightarrow{R_i^{\mathcal{L}}}, \mathcal{P}_{\overrightarrow{R_i^{\mathcal{L}}}}) = \mathcal{F}(\mathcal{DP})$

Important Property of PTD Function

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 $2\mathcal{L}$-dimensional vector.*



A Recommended Driving Route.

$$\mathcal{DP} = \langle D_1, \overline{P(C_1)}, D_2, \overline{P(C_2)}, D_3, \overline{P(C_3)}, D_4, \overline{P(C_4)} \rangle$$

Route Dominance

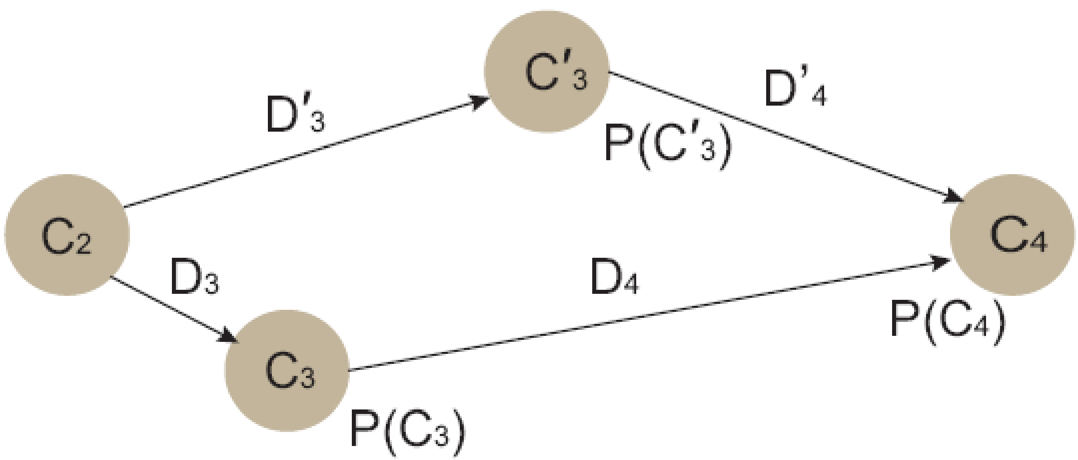
THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS

DEFINITION 1. **Route Dominance.** *A recommended driving route* $\overrightarrow{R^{\mathcal{L}}}$*, associated with the vector* $\mathcal{DP}$*, dominates another route* $\overrightarrow{\widetilde{R}^{\mathcal{L}}}$*, associated with the vector* $\widetilde{\mathcal{DP}}$*, iff* $\exists 1 \leq l \leq 2\mathcal{L}$, $DP_l < \widetilde{DP}_l$ *and* $\forall 1 \leq l \leq 2\mathcal{L}$, $DP_l \leq \widetilde{DP}_l$. *This can be denoted as* $\overrightarrow{R^{\mathcal{L}}} \Vdash \overrightarrow{\widetilde{R}^{\mathcal{L}}}$.

By this definition, if a candidate route A is dominated by a candidate route B, A cannot be an optimal route

Constrained Sub-route Dominance

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.* $\overrightarrow{R}_{sub} \Vdash \overrightarrow{R'}_{sub}$.



$\overrightarrow{R}_{sub}$ is $C_2 \rightarrow C_3 \rightarrow C_4$

$\overrightarrow{R'}_{sub}$ is $C_2 \rightarrow C'_3 \rightarrow C_4$

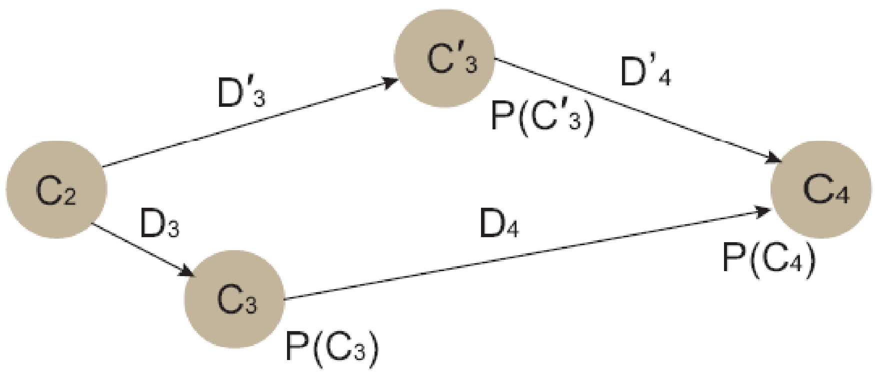
$$\mathcal{DP}_{sub} = \langle D_3, \overline{P(C_3)}, D_4, \overline{P(C_4)} \rangle$$

$$\mathcal{DP}'_{sub} = \langle D'_3, \overline{P(C'_3)}, D'_4, \overline{P(C_4)} \rangle$$

Illustration: the Sub-route Dominance

The $\mathcal{LCP}$ Algorithm

PROPOSITION 1. $\mathcal{LCP}$ **Pruning.** *For two sub-routes A and B with a length $\mathcal{L}$, 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 $\mathcal{L}$ which contain sub-route A will be dominated and can be pruned in advance.*



If $\mathcal{L} = 3$ 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

Skyline Route

DEFINITION 3. **Skyline Route.** *A recommended driving route* $\overrightarrow{R^{\mathcal{L}}}$ *is a skyline route iff* $\forall \overrightarrow{R_i^{\mathcal{L}}} \in \overrightarrow{\mathcal{R}}$, $\overrightarrow{R_i^{\mathcal{L}}}$ *cannot dominate* $\overrightarrow{R^{\mathcal{L}}}$ *by Definition 1. This is denoted as* $\overrightarrow{R_i^{\mathcal{L}}} \not\Vdash \overrightarrow{R^{\mathcal{L}}}$.

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

- 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

One Pruning Illustration

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS

- If SubRoute R_1 (PoCab→C_1→C_3) dominates
  SubRoute R_2 (PoCab→C_2→C_3)
  by DEFINITION 2

  Then Any candidate like
  (PoCab→C_1→C_3→C_i→…→C_L) dominates
  candidate as
  (PoCab→C_2→C_3→C_i→…→C_L),
  where C_i→…→C_L is one of all possible sub-routes

Outline

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS

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

Experimental Setup

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS

- 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

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS

- 6-7PM
- L=3:

  $PoCab \rightarrow C1 \rightarrow C3 \rightarrow C2$
- L=4:

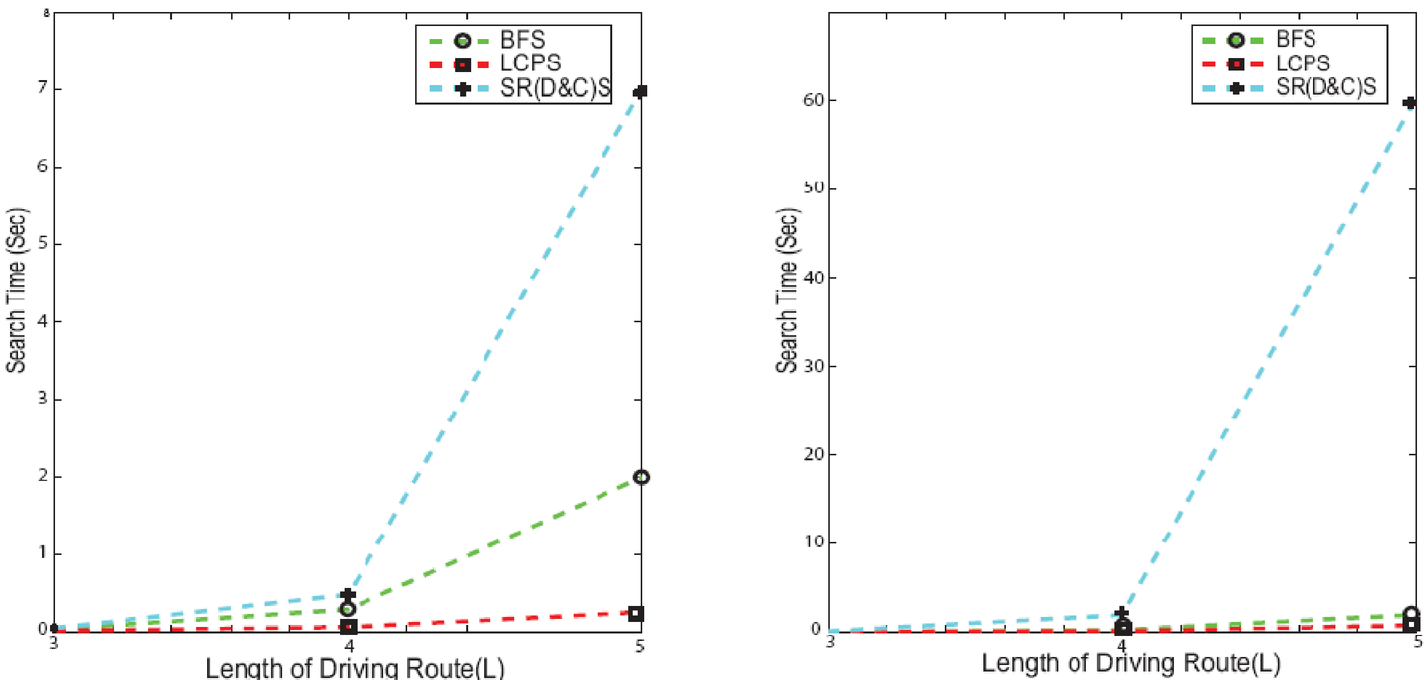
  $PoCab \rightarrow C1 \rightarrow C3 \rightarrow C2 \rightarrow C7$
- L=5

$PoCab \rightarrow C4 \rightarrow C1 \rightarrow C3 \rightarrow C2 \rightarrow C7$

An Overall Comparison

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS

A Comparison of Search Time



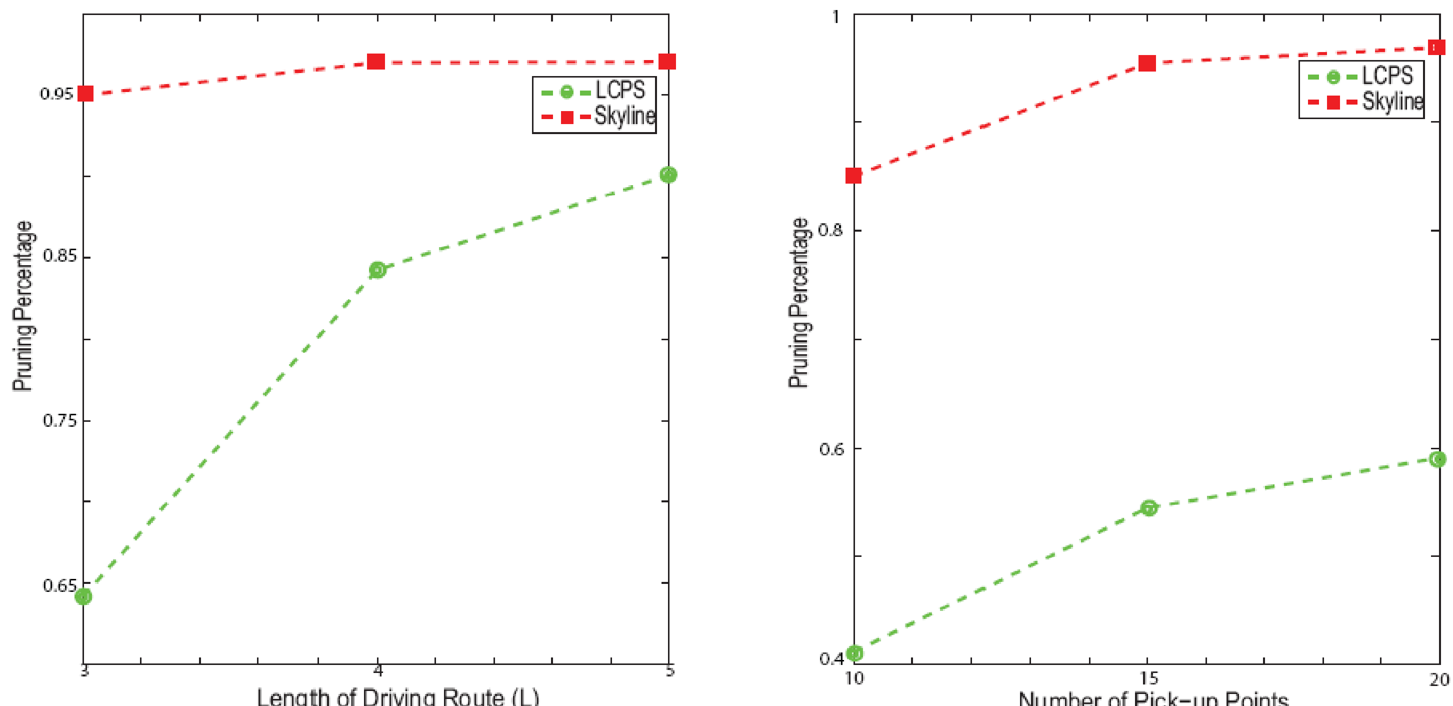
(a) Comparisons on Real Data (6–7PM)   (b) Comparisons on Synthetic Data (10 Clusters)

Some Acronyms.

BFS:	Brute-Force Search .
$\mathcal{LCPS}$:	Search with $\mathcal{LCP}$
SR(BNL)S:	Search via Skyline Computing algorithm $SkyRoute + BNL$.
SR(D&C)S:	Searching via Skyline Computing Algorithm $SkyRoute + D\&C$.

An Overall Comparison

The Pruning Effect

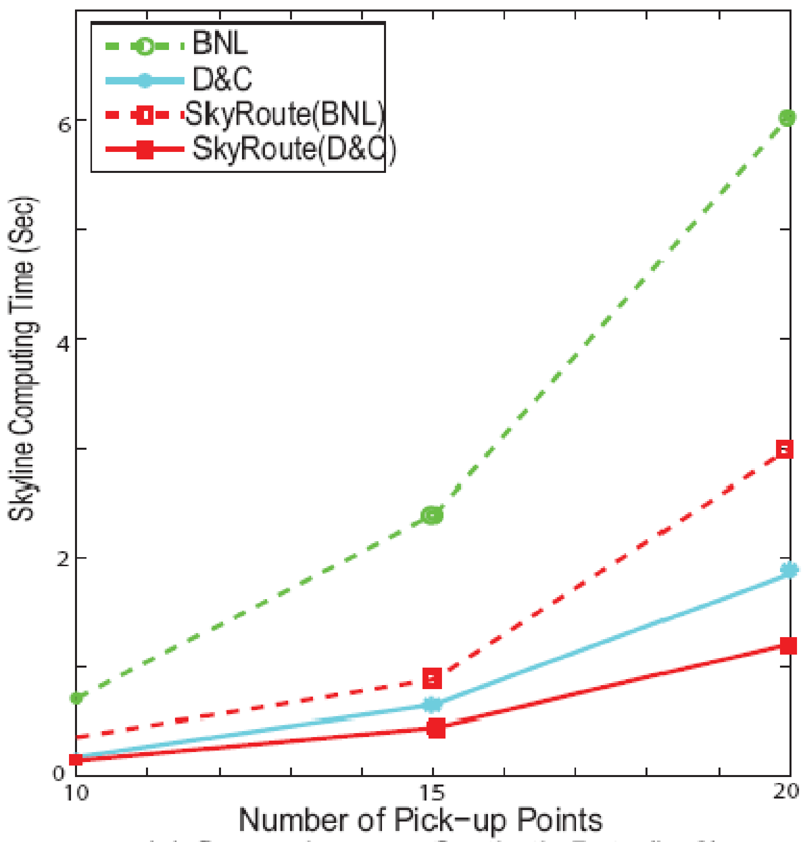


(a) The Pruning Effect on Real Data (6−7PM)

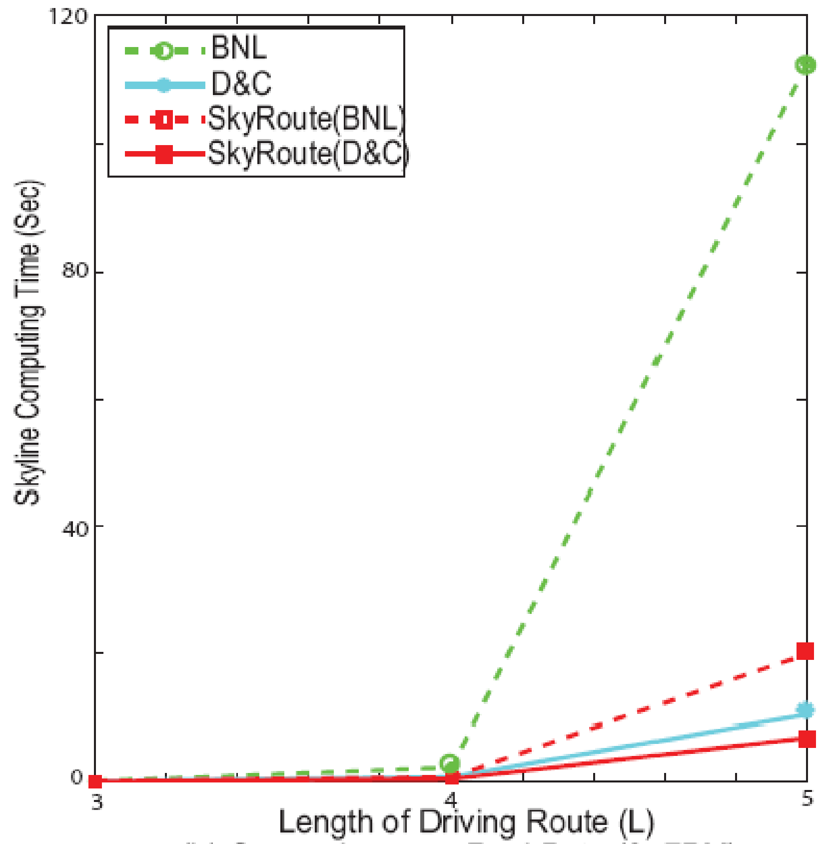
(b) The Pruning Effect on Synthetic Data (L=3)

Pruning percentage: the number of pruned candidates divided by the number of all original candidates

A Comparison of Skyline Computing



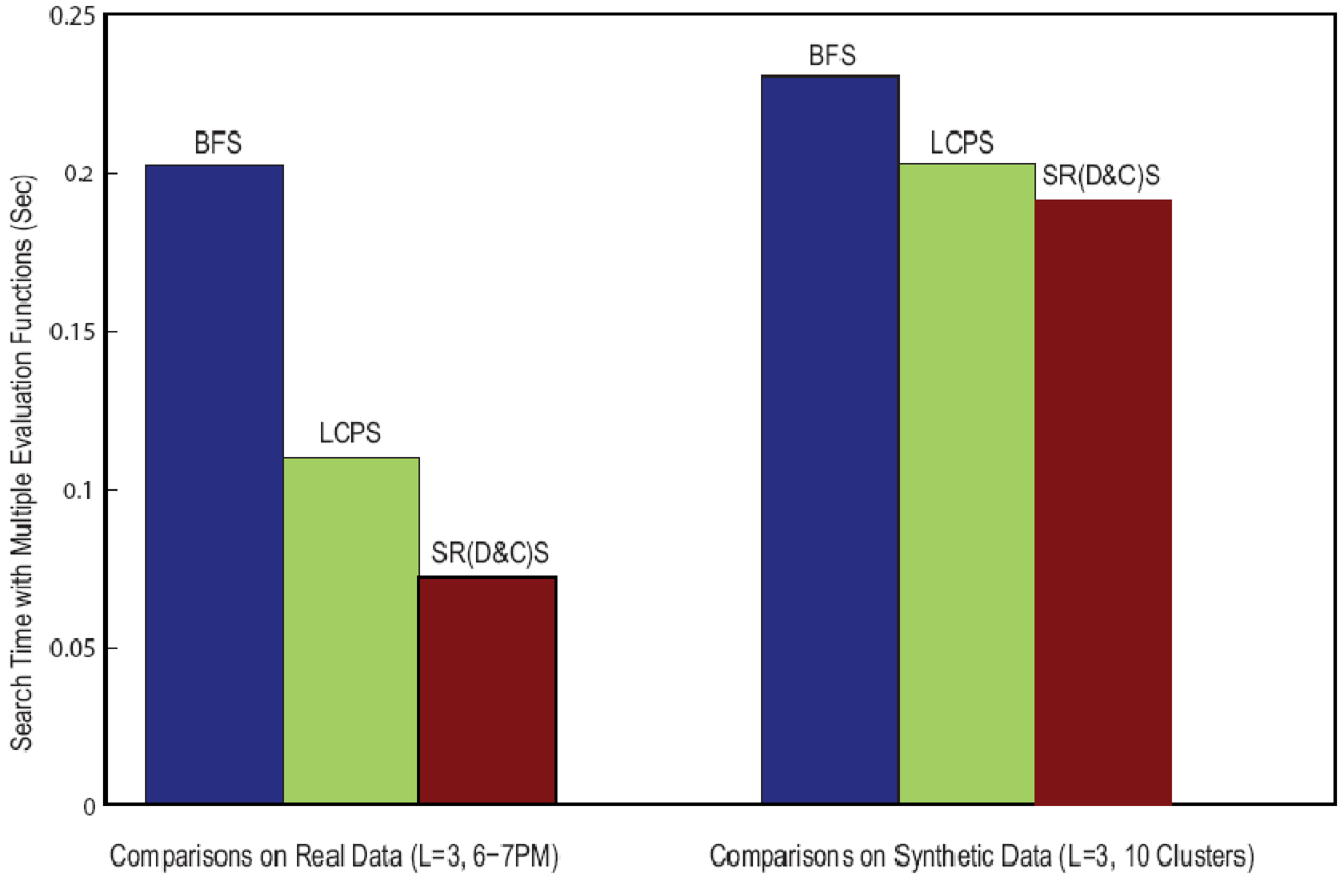
(a) Comparisons on Synthetic Data (L=3)



(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

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS

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

Conclusion Remarks

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS

- An energy-efficient mobile recommender system
- Potential Travel Distance (PTD) Function
- Algorithms: $\mathcal{LCP}$ and $\mathrm{SkyRoute}$

Thank You !

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS

Poster Session II & Demo Session
Date: Tuesday, July 27, 2010
Time: 5:45pm - 8:00pm
Location: Independence Center B, floor 1

The Recommendation Process

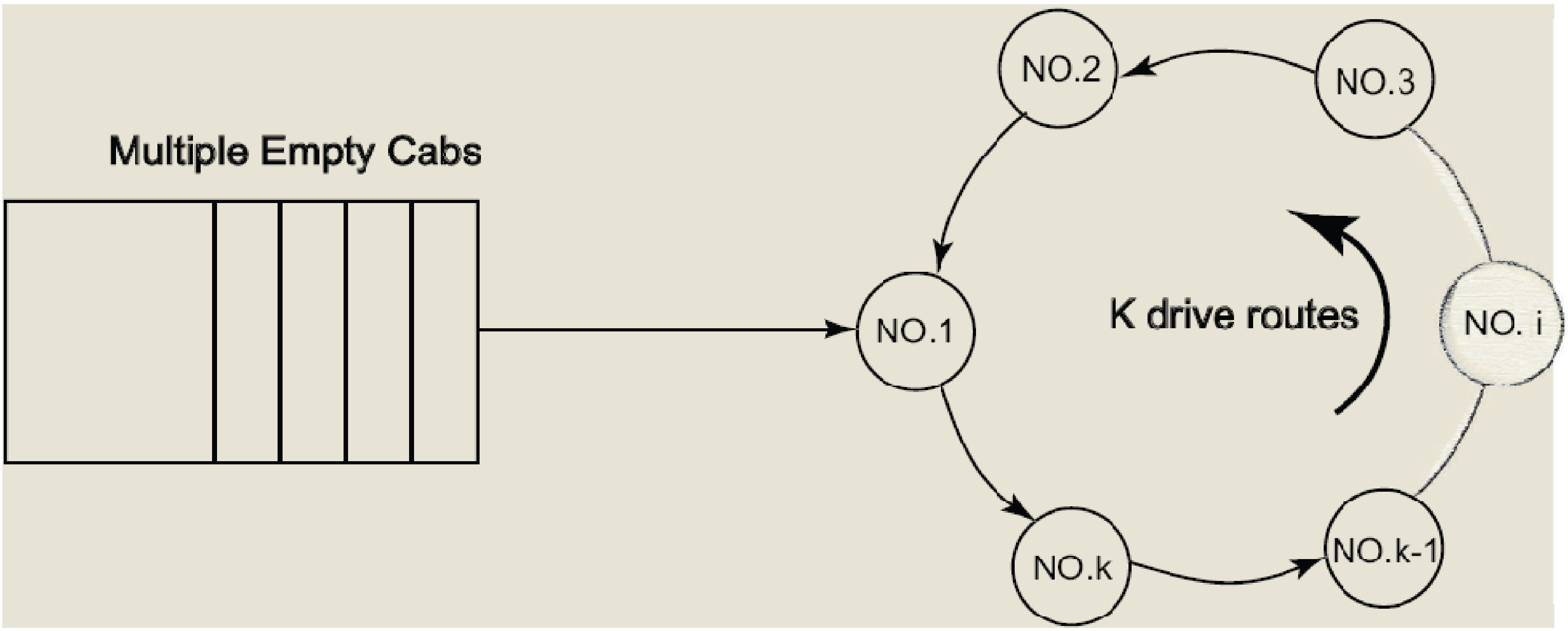


Illustration of the *Circulating Mechanism*

The *SkyRoute* Algorithm

ALGORITHM *SkyRoute*($\mathcal{C}$,$\mathcal{P}$,$Dist$,$\mathcal{L}$,$PoCab$)
Input:
$\mathcal{C}$: set of cluster nodes with central positions
$\mathcal{P}$: probability set for all cluster nodes
$Dist$: pairwise drive distance matrix of cluster nodes
$\mathcal{L}$: the length of suggested drive route
$PoCab$: the position of one empty cab
Output:
$\overrightarrow{R}_{Skyline}$: list of skyline drive routes.

Online Processing
1. Enumerate all candidate routes by connecting $PoCab$ with each sub-route of $\mathcal{R}^{\mathcal{L}}_{sub}$ obtained in step 10 during Offline Processing
2. for $i = 2 : \mathcal{L} - 1$
3. Decide dominated sub-routes with ith 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 $\mathcal{L}$ from the loop above
7. Final typical skyline query to get $\overrightarrow{R}_{Skyline}$ from those candidate routes in step 6

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

The pseudo-code of SkyRoute Algorithm

- First use LCP to prune some candidates
- Gradually prune candidates containing dominated sub-routes

An Overall Comparison

THE STATE UNIVERSITY OF NEW JERSEY
RUTGERS

A Comparison of Search Time (Second) between BFS and $\mathcal{L}CPS$

10 Synthetic Pick-up Clusters			
	$\mathcal{L} = 3$	$\mathcal{L} = 4$	$\mathcal{L} = 5$
BFS	0.051643	0.300211	2.000949
$\mathcal{L}CPS$	**0.043750**	**0.165401**	**0.803290**
15 Synthetic Pick-up Clusters			
BFS	0.142254	1.925054	23.517042
$\mathcal{L}CPS$	**0.095364**	**0.611193**	**4.322053**
Real Data (2-3PM)			
BFS	0.045933	0.297187	1.991507
$\mathcal{L}CPS$	**0.036736**	**0.141536**	**0.622932**

An Overall Comparison