Mining Human Activities by GPS Trajectories

Minjie Chen
SIPU group
Univ. of Eastern Finland

Introduction

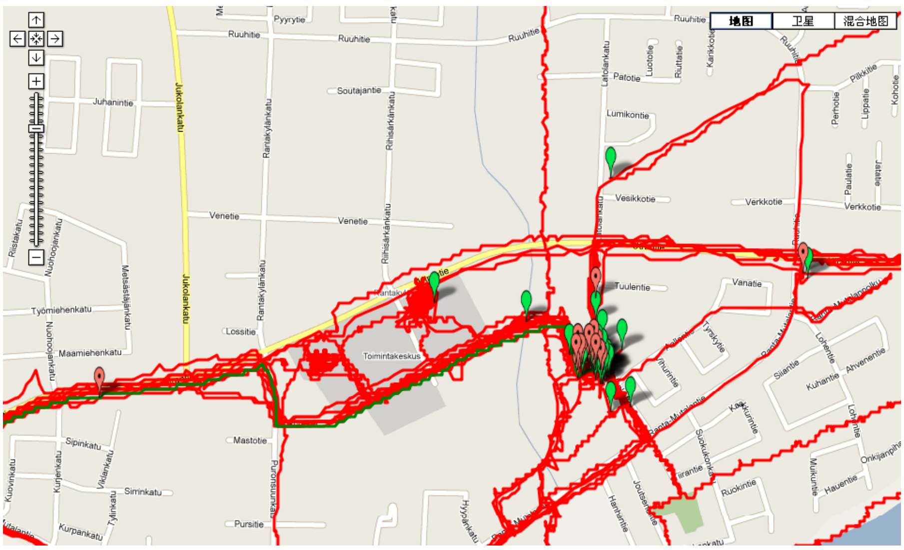
- A number of trajectories/routes are collected of users' position information uses a mobile phone with built-in GPS receiver.
- The focus of this work is to analyze the human behaviour based on the collected GPS data.
- The collected routes are divided into several segments with different properties (transportation modes), such as stationary, walking, biking, running, or car driving.

Methodology

Our approach consists of three parts:

- GPS signal pre-filtering
- A change-point-detection for route segmentation
- An inference algorithm for classification the properties of each segments.

GPS signal pre-filtering



GPS signal has an accuracy around 10m, design efficient filtering algorithm is important for route analysis task

Existing Method

- Remove outlier points based on given criterion

  (Number of satellite, speed, speed variance, HDOP, angle change)
- Weighting method for adjacent point

  (shrinkage problem)
- Map-Matching

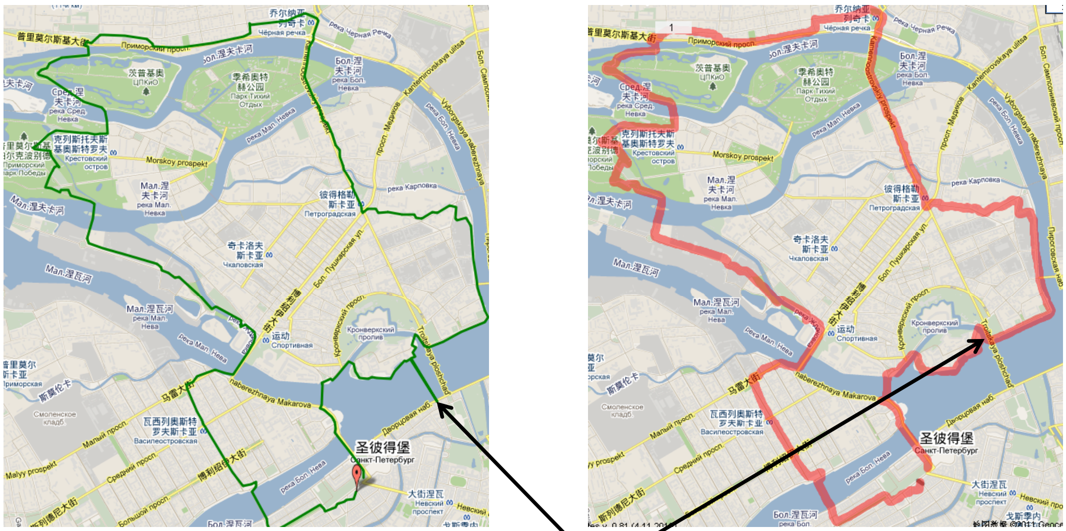
  (A road network is needed)

Proposed Method

- Our proposed algorithm has two steps: outlier removal and route smooth
- No prior information is needed (e.g. road network)

Outlier removal

- For those points with impossible speed and speed variance is detected as outlier and removed.



Outlier point is removed after filtering

Smooth the GPS trajectory

- We assume the speed of adjacent points is similar

- It is converted as an optimization problem by optimizing both the position distortion and speed change.

Smooth the GPS trajectory

Define **p**=[**x**,**y**]^T is coordinate list, **s** is coordinate difference, **v** is speed difference, we have:

$$H = \begin{pmatrix} 1 & 0 & & \\ & \ddots & 0 & \\ 1 & 1 & & \\ & & 1 & 0 \\ & 0 & & \ddots \\ & & 1 & 1 \end{pmatrix} \qquad L = \begin{bmatrix} 0 & -1/\Delta t_1 & 1/\Delta t_2 & & \\ & \ddots & \ddots & & \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ & & & -1/\Delta t_{n-2} & 1/\Delta t_{n-1} \end{bmatrix}$$

$$p = \begin{pmatrix} x \\ y \end{pmatrix}, \quad p = Hs, \text{ where } s = \begin{pmatrix} x_1 \\ dx_1 \\ \vdots \\ dx_n \\ y_1 \\ dy_1 \\ \vdots \\ dx_{n-1} \end{pmatrix}, v = Ls,$$

Smooth the GPS trajectory

We want to solve **s** by:

$$\hat{s} = \arg\min \| p - Hs \|_2^2 + \lambda \| Ls \|_k, \text{k=1 } L_1\text{norm,k=2 } L_2\text{norm}$$

$$\hat{p} = Hs$$

Position distortion Speed difference

For L_2norm, it can be solved by:

$$\hat{s} = \frac{\lambda H^T p}{\lambda L^T L + H^T H}$$

For L_1norm, it can be solved by IRLS(Iterated Reweighted Least Square)

$$\hat{s} = \frac{H^T p}{\lambda L^T (diag(| L s_{t-1} |))^{-1} L + H^T H}$$

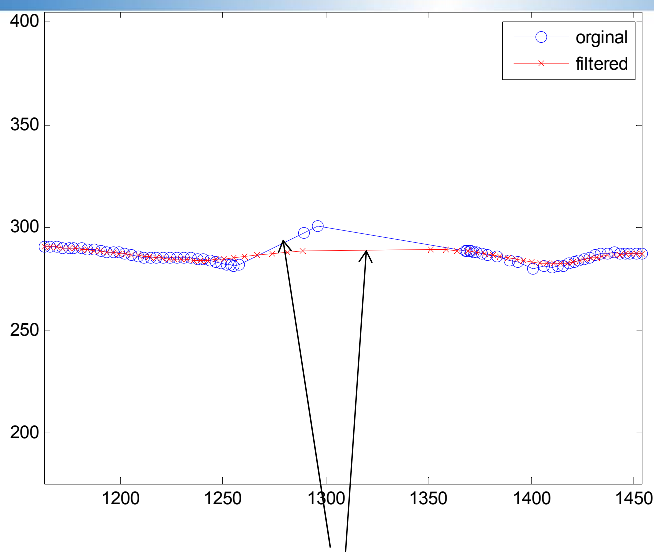
Smooth the GPS trajectory

Save computation cost

- Sub-segments with 2 minutes are processed separately.
- A sliding window is used to process these overlapped sub-segments separately.
- The final result is the averaging of the filtering result of sub-segments.

- The time cost is 1s for 5000 points in L_2 in 5s in L_1 in matlab

Example

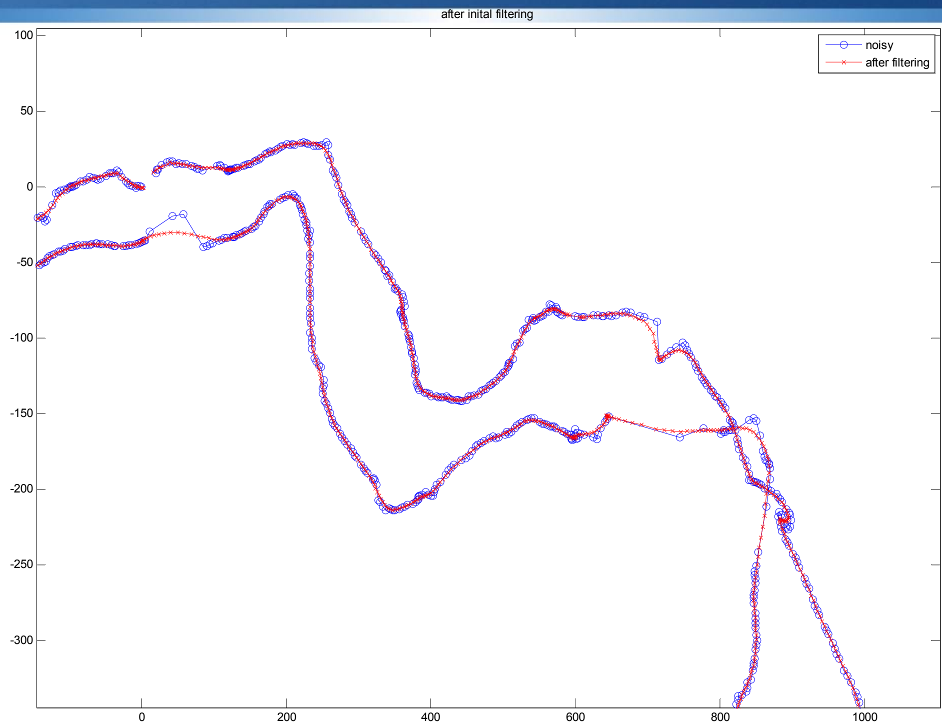


8s interval here, other are also 1s

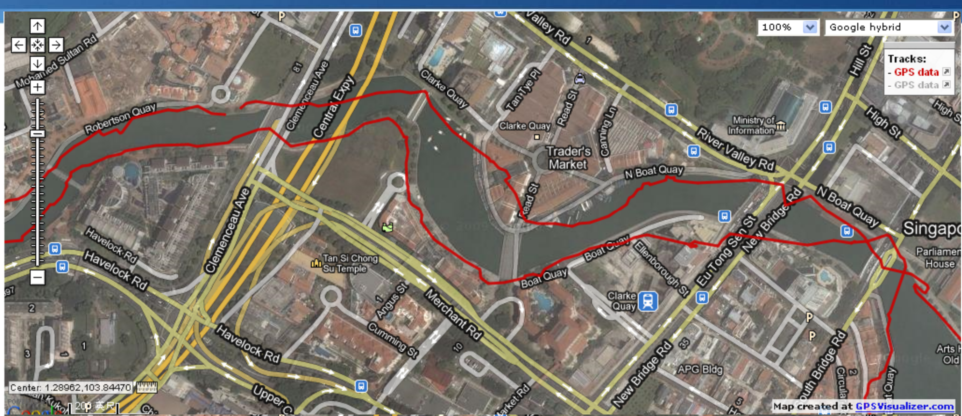
before

after

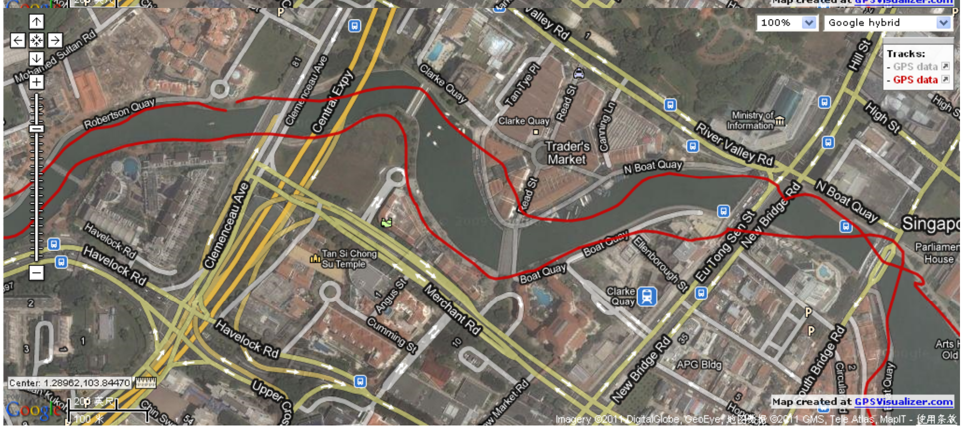
Example



Example

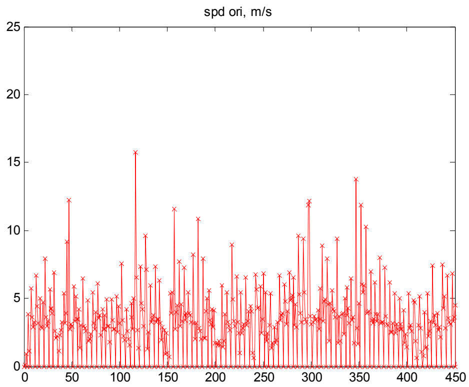


Before

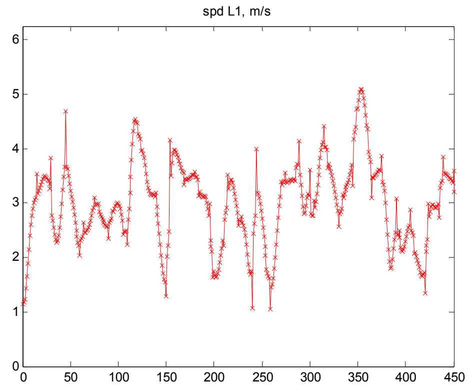


After filtering

Speed calculation



Speed calculation: before



Speed calculation: after

Route Segmentation

- This can be considered as a change-point detection problem
- Our solution has two steps: initialization and merging.

Route Segmentation - Initialization

- We assume the speed variance is small inside one segment
- Suppose we want to minimize the sum of speed variance for all segments as follows:

$$f = \sum_{i=1}^{M} \sigma(v_i) \cdot t_i$$

where $\sigma(v_i)$ and t_i are the variance and time of segment i

- This can be solved by dynamic programming

Route Segmentation - Merging

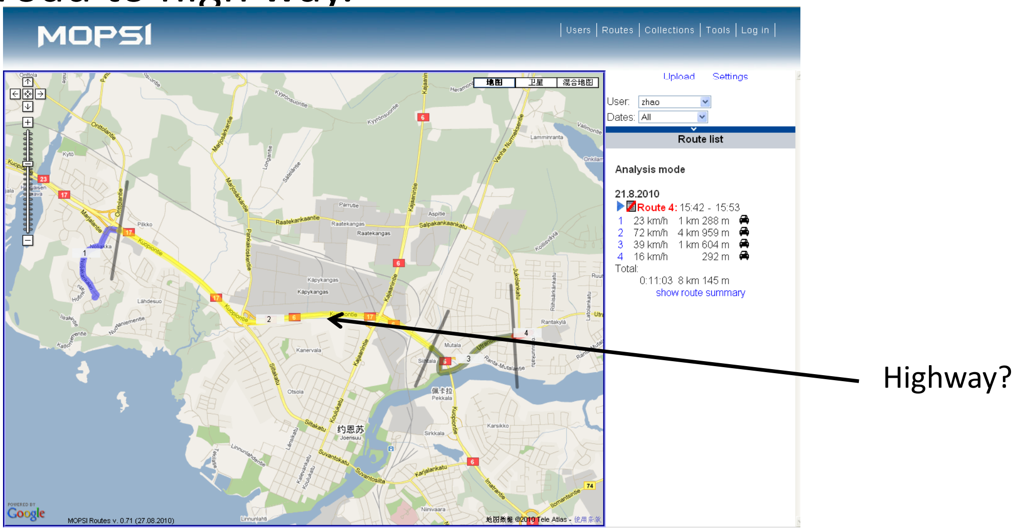
- Merging is done to fine-tune the initial result.
- Adjacent segments with similar properties are merged together by:

$$T = g(v_1, v_2, \sigma(v_1), \sigma(v_2), \ldots)$$

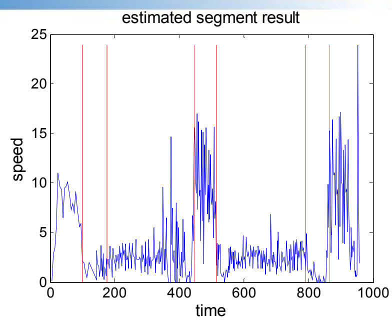
- A classifier is trained to determine if two adjacent segments is merged of not

Route Segmentation

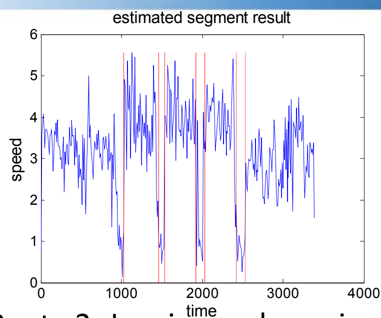
- The proposed algorithm can also detect some speed change on the road, e.g. car is moved from normal road to high way.



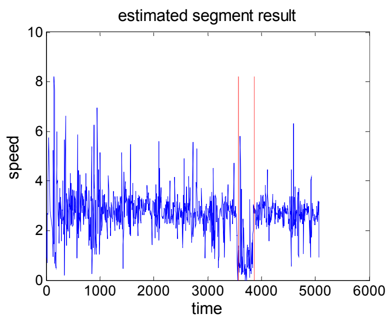
Result



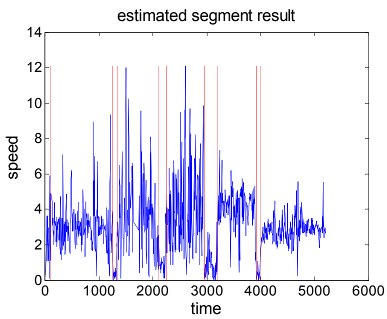
Route 1: ski



Route 2: Jogging and running with non-moving interval



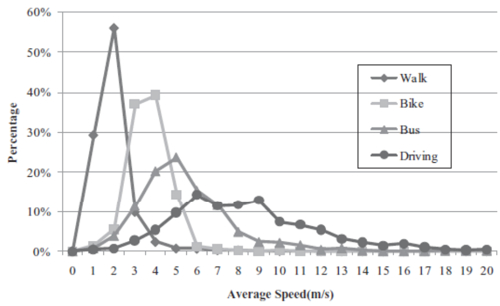
Route 3:Non-moving



Route 4: Jogging and running with non-moving interval

Route Classification

- In classification step, we want to classify each segments with corresponding properties, e.g. stationary, walking, biking, running, or car driving
- A number of features can be extracted for each segments, such as speed, acceleration, time, distance,…, we define it as feature set X.
- However, training a classifier on these features directly is inaccurate, they are overlap.



Route Classification

- We also consider the dependency of the properties in previous and next segments by minimizing:

$$f = \prod_{i=1}^{M} P(m_i \mid X, m_{i-1}, m_{i+1})$$

where m_i = {stationary, walking, biking, running, car } is the classification result

- This is same as:

$$f = \prod_{i=1}^{M} \frac{P(m_i \mid m_{i-1}, m_{i+1}) P(m_i \mid X)}{P(m_i)} = \prod_{i=1}^{M} \frac{P(m_{i+2} \mid m_i, m_{i+1}) P(m_{i+1} \mid X)}{P(m_{i+2})}$$

$P(m_{i+2} \mid m_i, m_{i+1}), P(m_i \mid X), P(m_i)$ are all the prior information

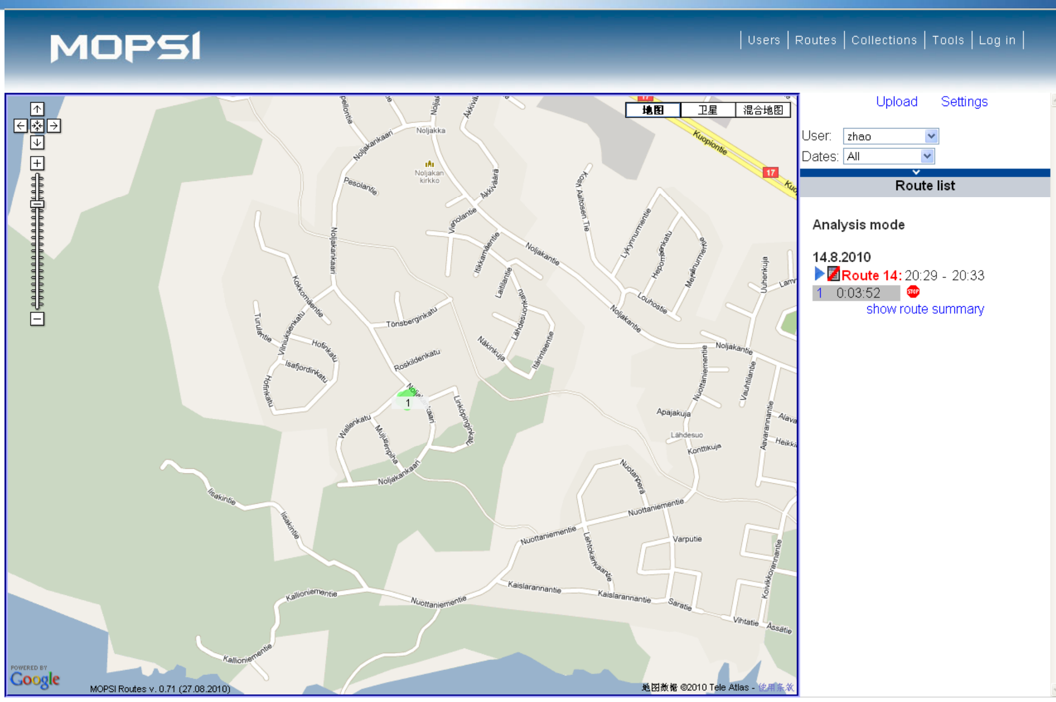
- Viterbi algorithm can be used to solve the problem

Route Classification, Proof

$$
\begin{aligned}
& P(m_i \mid m_{i-1}, m_{i+1}, X) \\
&= \frac{P(m_{i-1}, m_{i+1}, X \mid m_i) P(m_i)}{P(m_{i-1}, m_{i+1}, X)} \\
&= \frac{P(X \mid m_i) P(m_{i-1}, m_{i+1} \mid m_i) P(m_i)}{P(m_{i-1}, m_{i+1}) P(X)} \\
&= \frac{P(m_i \mid X) P(X)}{P(m_i)} \frac{P(m_i \mid m_{i-1}, m_{i+1}) P(m_{i-1}, m_{i+1})}{P(m_i)} \frac{P(m_i)}{P(m_{i-1}, m_{i+1}) P(X)} \\
&= \frac{P(m_i \mid m_{i-1}, m_{i+1}) P(m_i \mid X)}{P(m_i)}
\end{aligned}
$$

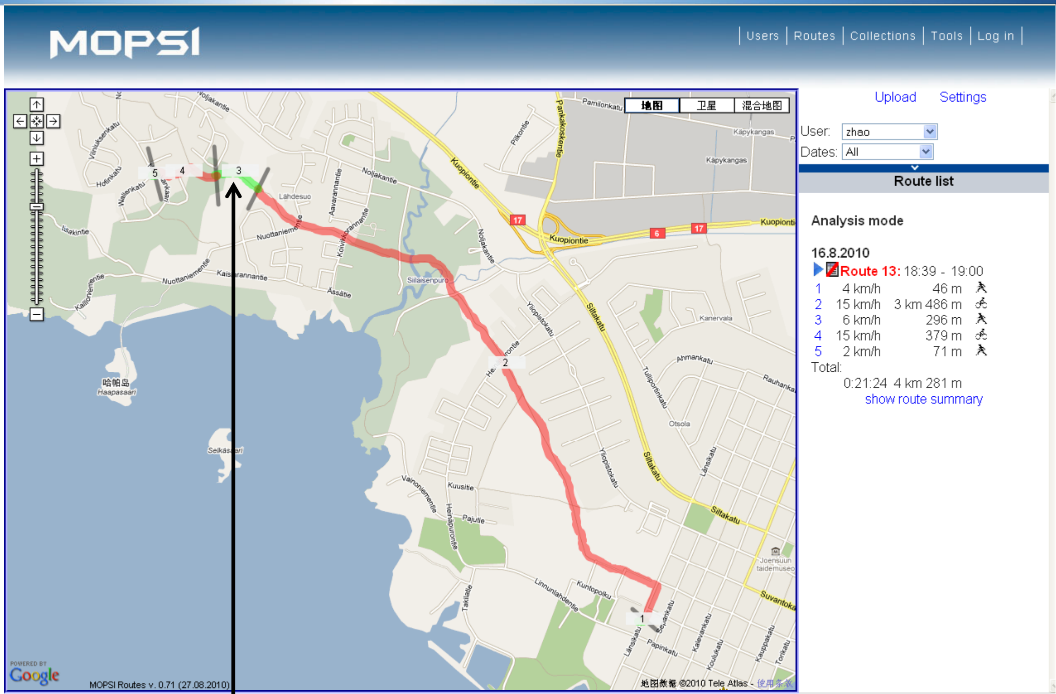
$$
\begin{aligned}
& \prod_{i=1}^{M} \frac{P(m_i \mid m_{i-1}, m_{i+1}) P(m_i \mid X)}{P(m_i)} \\
&= \prod_{i=1}^{M} \frac{P(m_{i+1} \mid m_{i-1}, m_i) P(m_i)}{P(m_{i+1})} \frac{P(m_i \mid X)}{P(m_i)} \\
&= \prod_{i=1}^{M} \frac{P(m_{i+1} \mid m_{i-1}, m_i) P(m_i \mid X)}{P(m_{i+1})}
\end{aligned}
$$

Example



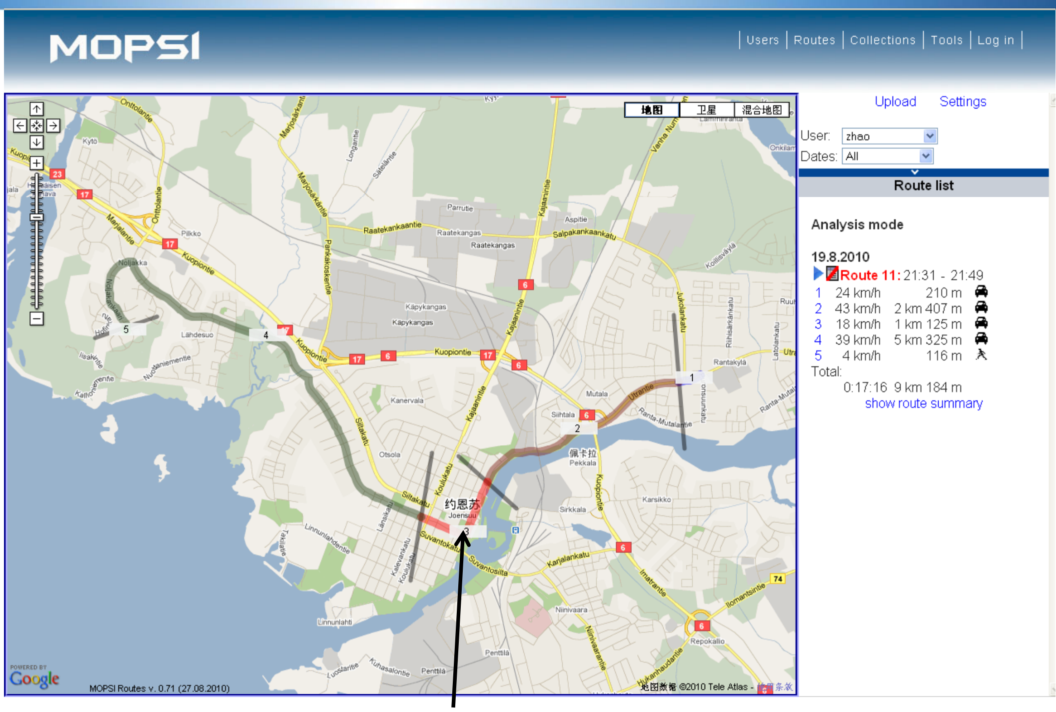
Stopping area

Example



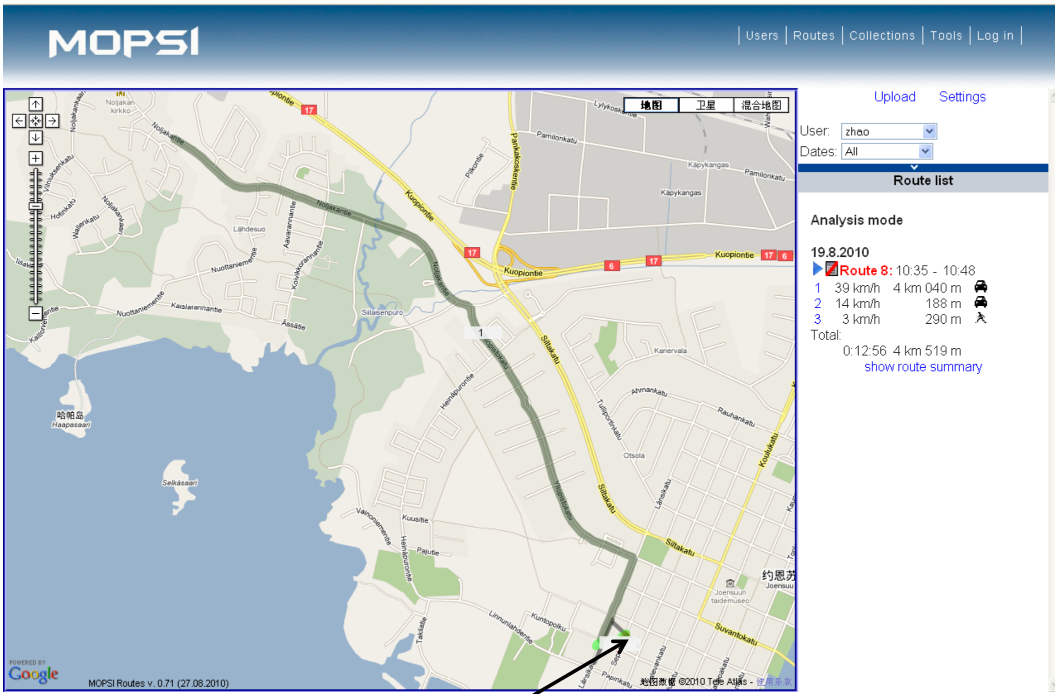
Not flat ground and can’t go by bicycle?

Example



City center?

Example



Parking place?

Conclusion

- We propose an algorithm for human behaviour (Transportation mode) analysis from the GPS data collected by mobile phone
- It consists of three parts: GPS signal pre-filtering, a change-point-detection for route segmentation and an inference algorithm for classification the properties of each segments.

Questions?