Frequent Pattern-based Map-matching on low sampling rate trajectories

Yukun Huang† Weixiong Rao § Zhiqiang Zhang § Peng Zhao § Mingxuan Yuan† Jia Zeng†

†Huawei Noah's Ark Lab, Hong Kong § School of Software Engineering, Tongji University, China

{yuan.mingxuan, zeng.jia}@huawei.com wxrao@tongji.edu.cn

Background

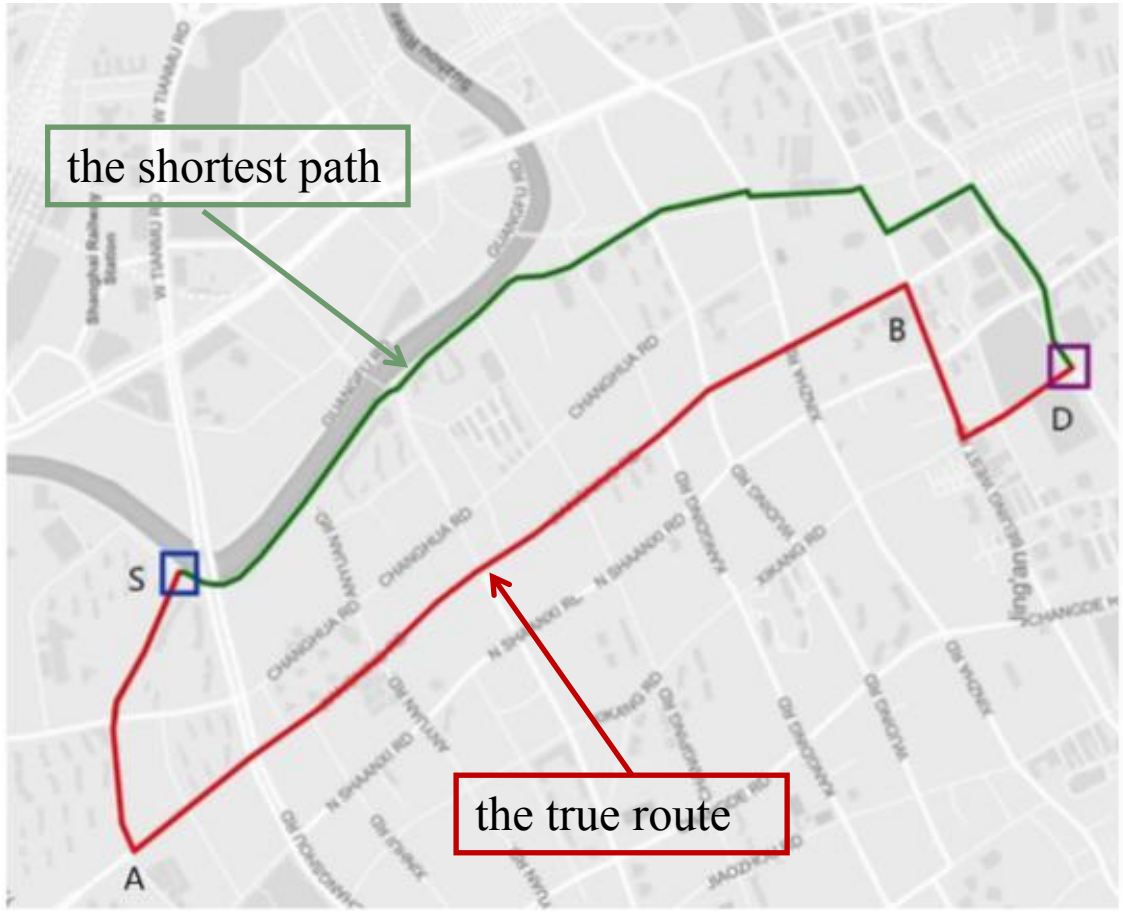
- Map-matching is an important preprocessing task for many location-based services (LBS).

Urban planning

Intelligent transportation

Observation

- The shortest path computation may not work well for very low sampling rate trajectory



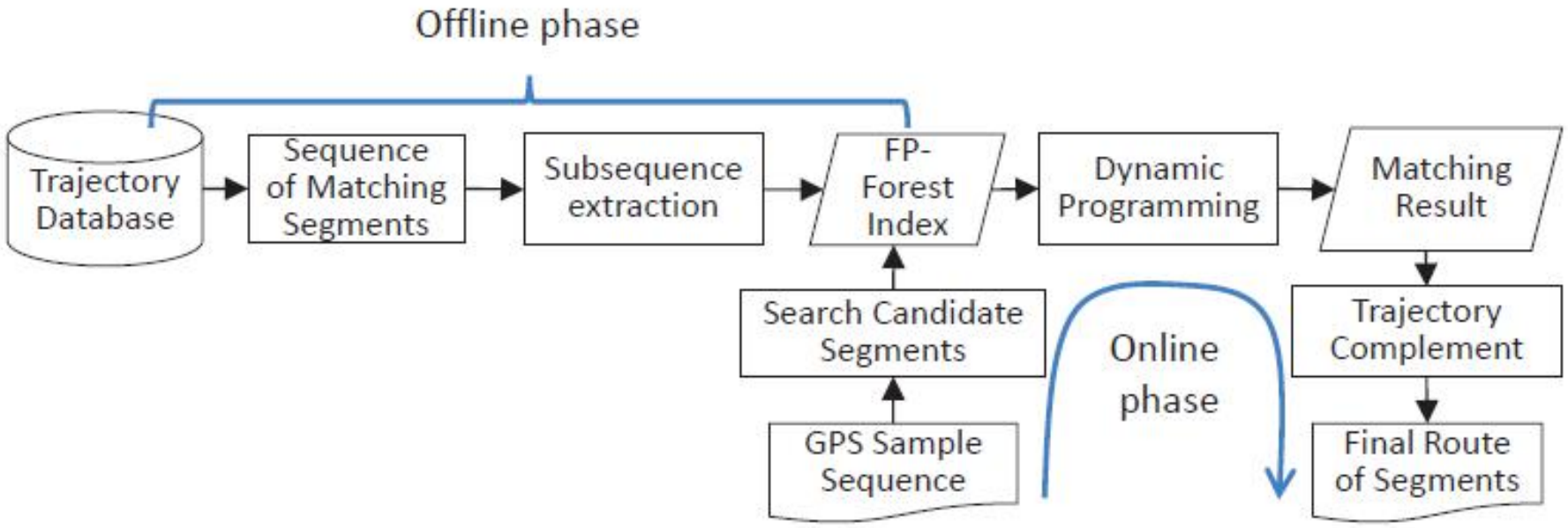
Contribution

➢We propose a frequent pattern-based method (FP-matching) to optimize the map-matching precision for low sampling rate trajectories.

- identify frequent patterns (FPs) from a third-party historical trajectory data

- find those FPs which most match a trajectory and link such found FPs as the map-matching result

System Overview

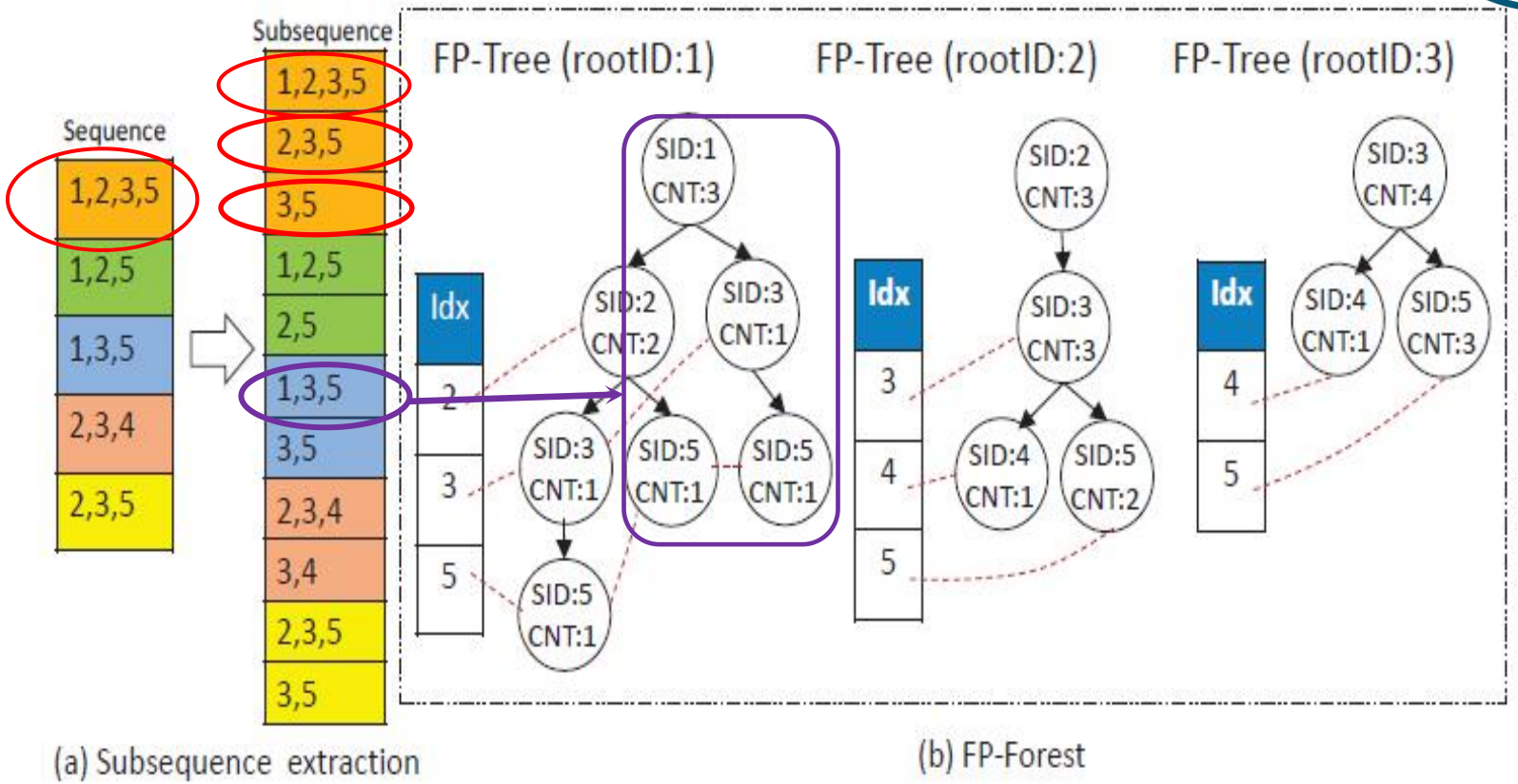


FP-FOREST

- SID:
  the road segment ID

- CNT:
  the count of such trajectories matched to the road segment

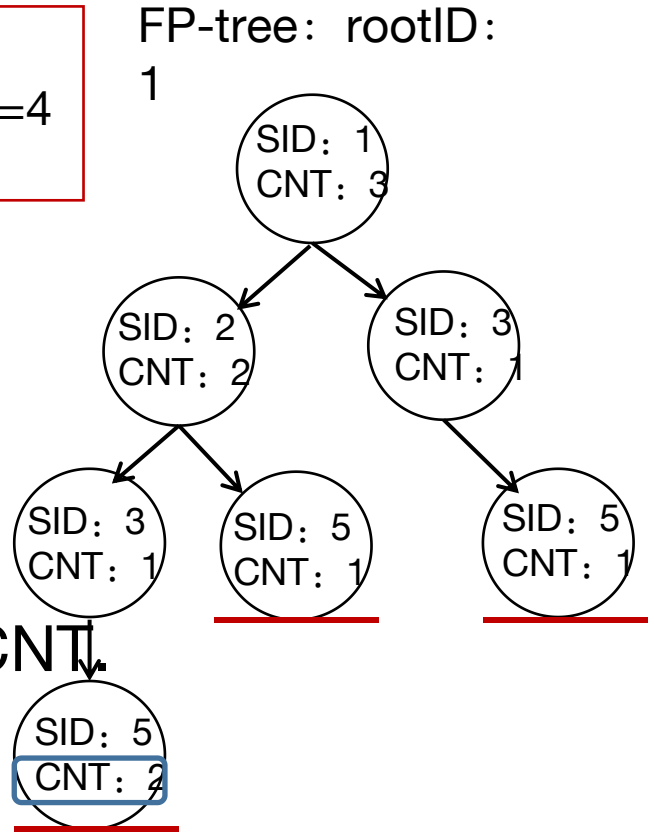
- Creating FP-forest:
  - subsequence extraction
  - Subsequence insertion



FP-Forest used for what?

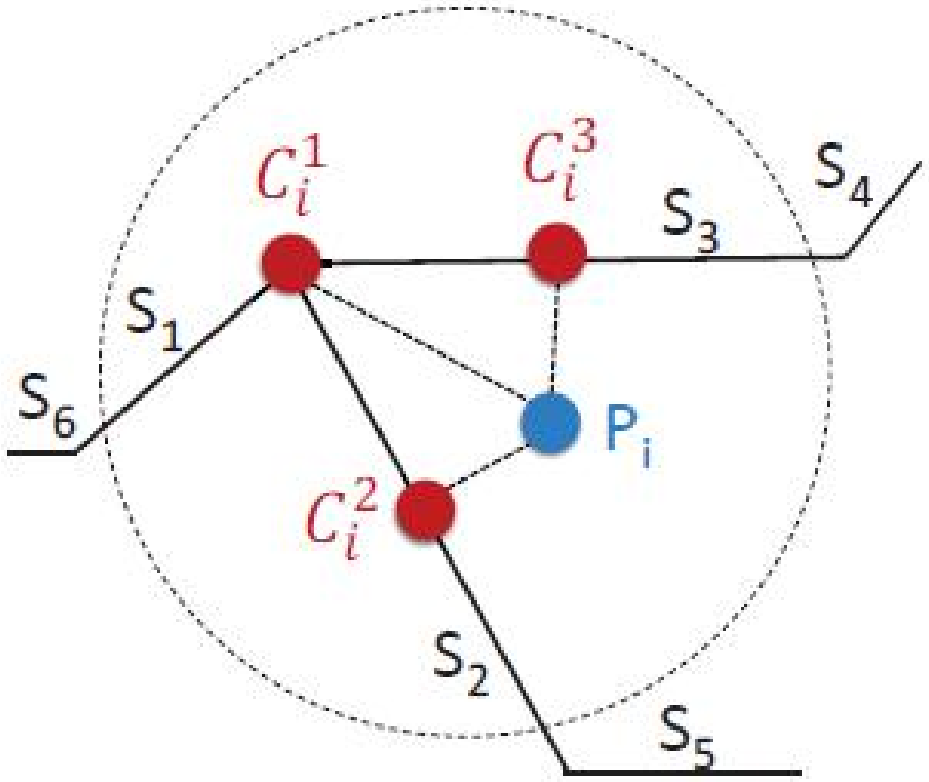
- Frequency Lookup:
  find the count of trajectories passing from a road segment to another
- Trajectory Complement:
  select the most frequent route between two disconnected road segments in the FP-tree by CNT.

Eg: SID 1 To SID 5
$Feq\ (S1 \rightarrow S5) = 2 + 1 + 1 = 4$
Complement:S1,S2,S3, S5



FP-MATCHING ALGORITHM

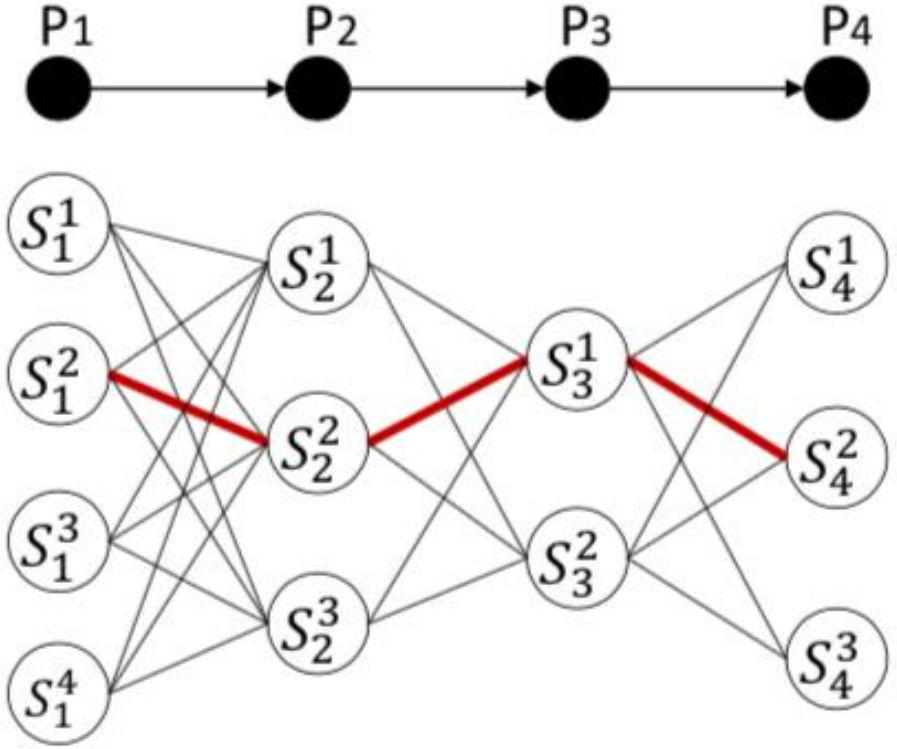
- Searching Candidate



FP-MATCHING ALGORITHM

- Result Matching

$$P = \arg\max_{S_i^{best}} \sum_{i=1}^{n-1} Weight(S_i^{best_i} \rightarrow S_{i+1}^{best_{i+1}})$$



FP-MATCHING ALGORITHM

Calculate Weights

$$Weight(S_i^j \rightarrow S_i^{j'}) = Weight(S_i^j, S_i^{j'}) * Freq(S_i^j \rightarrow S_i^{j'})$$

1. frequency between segments: $Freq(S_i^j \rightarrow S_i^{j'})$
2. weight of road segment: $W(S_i^j) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(dist_i^j - \mu)^2}{2\sigma^2}}$
3. effect of two candidates: $Weight(S_i^j, S_i^{j'}) = \frac{(1+\beta^2) * W(S_i^j) * W(S_i^{j'})}{\beta * W(S_i^j) + W(S_i^{j'})}$

Experiment Evaluate

- Dataset

  road network: the digital map of Shanghai from Open Street Map

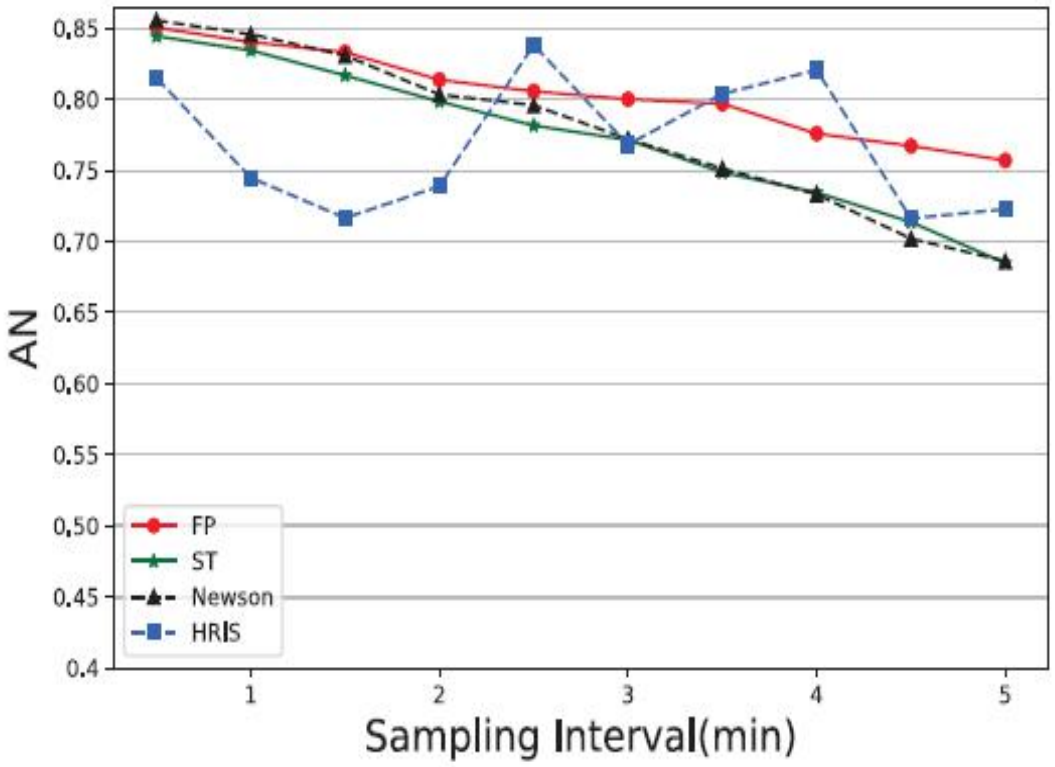
  Real Taxi Data : taxi trajectories in one day in Shanghai (92602 trajectories)

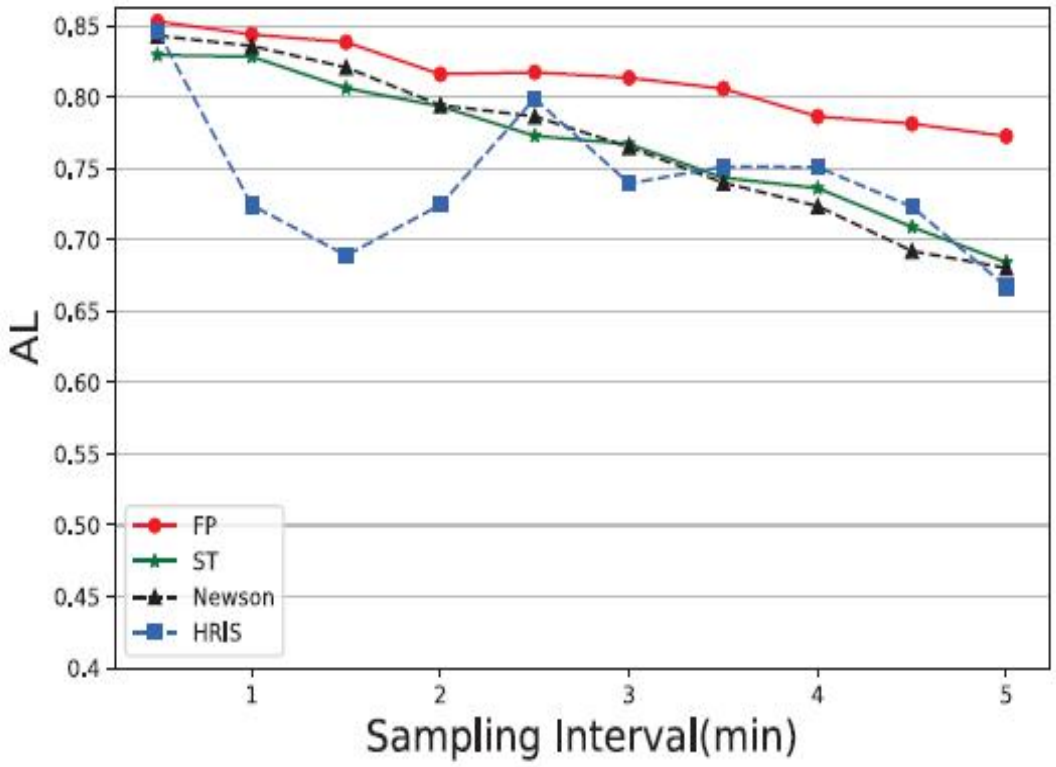
- Performance metrics

$$A_N = \frac{\#\text{correctly matched road segments}}{\#\text{all road segments of the trajectory}}$$

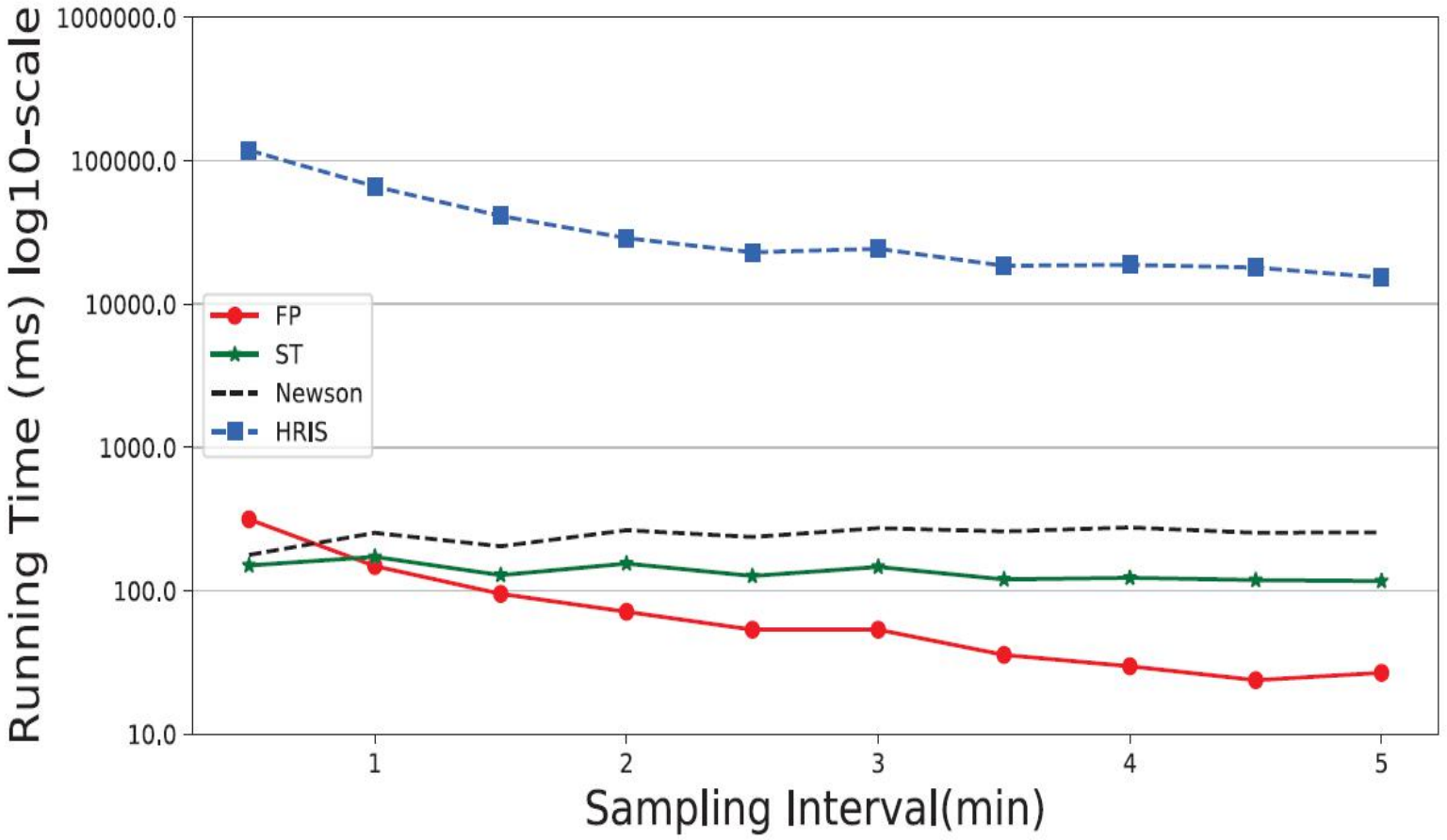
$$A_L = \frac{\sum \text{the length of matched road segments}}{\text{the length of the trajectory}}$$

Comparison of three algorithm





Efficiency


Running Time (ms) log10-scale
1000000.0
100000.0
10000.0
1000.0
100.0
10.0
FP
ST
Newson
HRIS
1
2
3
4
5
Sampling Interval(min)

FUTURE WORK

- we plan to compress FP-forest for small space overhead.
- we are going to design an effective approach to fill the missed trajectories
- we are interested in how to extend our algorithm in an online manner.

Thanks

Offline discussion email：
zhangzhiqiang@tongji.edu.cn