# Demo Abstract: Frequent Pattern-based Trajectory Completion


Zhiqiang Zhang
Tongji University, China
zqzhang@tongji.edu.cn

Weixiong Rao
Tongji University, China
wxrao@tongji.edu.cn

Xiaolei Di
Tongji University, China
xldi@tongji.edu.cn

Peng Zhao
Tongji University, China
peng.zhao@tongji.edu.cn

Xiaofeng Xu
Tongji University, China
xfxu@tongji.edu.cn

Fehmi Ben Abdesslem
RISE SICS, Sweden
fehmi.ben.abdesslem@ri.se



## ABSTRACT

GPS sensors have been widely used to track people's everyday life trajectories, generating massive trajectory datasets. The trajectory data typically contains sparse GPS points, and completing trajectories is often necessary. State-of-the-art methods [3, 4] essentially complete the entire route by using a single metric, e.g., either the shortest distance or the fastest driving/walking time. Unfortunately, using a single metric may not always work in real life due to the diversity of mobility patterns. In this demo abstract, we propose a frequent pattern (FP)-based trajectory completion approach, and demonstrate a system prototype to showcase the advantages of our approach over four previous works, in terms of accuracy and running time.



**ACM Reference Format:**

Zhiqiang Zhang, Weixiong Rao, Xiaolei Di, Peng Zhao, Xiaofeng Xu, and Fehmi Ben Abdesslem. 2018. Demo Abstract: Frequent Pattern-based Trajectory Completion. In *SenSys '18: Conference on Embedded Networked Sensor Systems, November 4–7, 2018, Shenzhen, China.* ACM, New York, NY, USA, Article 4, 2 pages. https://doi.org/10.1145/3274783.3275158


## 1 INTRODUCTION

Recent years witness the proliferation of location-based services, such as Uber, Wechat and Baidu Map on mobile devices. Such services use GPS sensors to track and record people's everyday life trajectories, generating massive trajectory datasets. Mobile users often switch off GPS sensors on mobile phones, e.g., for privacy concerns or energy saving, and switch them on when necessary. This results in sparse GPS points and incomplete trajectories. A typical trajectory completion first involves the projection of GPS points onto digital maps, and then the connection of the projected points (e.g., by shortest paths) to recover the entire route. The state-of-the-art methods [3–6] perform the connection step with a Hidden Markov Model (HMM) and its variations. The fundamental assumption of such approaches is that the connected path between two GPS points is likely the shortest path according to *a certain metric*. For example, [3] considers both spatial distance and time interval from one GPS point to another. Unfortunately, such metric may not always work in practice due to the diversity of mobility patterns: some mobile users may favour short distance, while others favour fuel saving by avoiding traffic jams or travel time by driving on high speed roads.

In this demo, we demonstrate a frequent pattern (FP)-based method [2]. The general idea is first to identify frequent patterns (FPs) from historical trajectory data, and then to find those FPs which are most likely to match the sparse GPS points. Such FPs reflect the diversity of mobility patterns for various roads, mobile users and time periods: each FP corresponds to either the most familiar route, the fastest route, or the most time saving route, for example. Hence we can exploit this diversity to recover an entire route by connecting multiple FPs and various metrics to better match the diversity of mobility patterns.

To evaluate the proposed FP approach, we demonstrate a trajectory completion platform to perform various trajectory completion approaches including the proposed approach and four state-of-the-art approaches [3–6]. This platform allows algorithm configuration, trajectory visualization and performance measurement. This demo does not only showcase the advantages of the FP approach but can also be used as a benchmark tool to evaluate the state-of-the-art trajectory completion algorithms.

The remainder of this demo abstract is organized as follows. Section 2 first introduces the proposed FP-based trajectory completion. Section 3 describes both the system prototype and the demo. Section 4 finally concludes the abstract.

## 2 TRAJECTORY COMPLETION

As shown in Figure 1, the FP approach [2] employs online and offline phases to perform the trajectory completion. The offline phase maintains an FP-Forest indexing structure [1], and the online phase evaluates an input of sparse GPS points against the FP-forest to generate the most likely route of connected road segments (representing the complete trajectory).




*SenSys '18, November 4–7, 2018, Shenzhen, China*



ACM ISBN 978-1-4503-5952-8/18/11... $15.00
https://doi.org/10.1145/3274783.3275158

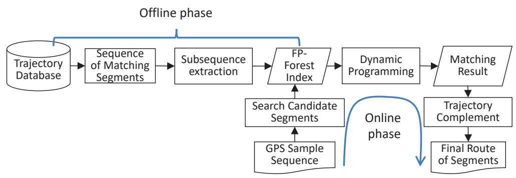


**Figure 1: System Overview**

In the *offline* phase, we pre-process a third-party history trajectory database to maintain an FP-forest structure. Such trajectory database is publicly available nowadays, for example generated by taxi and other mobile devices. Given the database, we first find high sampling rate trajectories. Next, we employ a classic map-matching algorithm, e.g., [4], to find the sequence of matching road segments to which GPS points in the high sampling rate trajectories are projected. Finally, we extract frequent subsequences from such sequences of matching road segments. All such sequences (and subsequences) are then indexed with FP-forest [1].

In the *online* phase, when given a sequence of sparse GPS points, we first search digital maps to find the nearest road segment candidates with respect to each GPS point. For each road segment candidate, we use the FP-forest to find those frequent subsequences (of road segments) containing the candidate. After that, with a dynamic programming algorithm, we select the path of frequent subsequences to best match the entire input of GPS points. Finally, in case the best matching path contains disconnected subsequences of road segments, we then complement the segments to recover the entire route, e.g., by using a default metric.

## 3 PROTOTYPE DEMONSTRATION

Our prototype contains three components. 1) **Data Store**: including the road network map data, and the raw trajectory history database, 2) **Algorithm Library**: implementing the following algorithms. ST [3] incorporates spatial geometric and topological structures of road networks and temporal/speed constraints of input trajectories. HMM [4] uses a Hidden Markov Model to find the most likely road route represented by a time-stamped sequence of latitude/longitude pairs. FV [5] completes sparse trajectories by Fast Viterbi. HRIS [6] infers possible routes of a low-sampling-rate trajectory by extracting rich information (e.g., reference trajectories) from the historical trajectories. Finally the proposed FP algorithm. 3) **GUI**: To configure the algorithms, display the running results and the completed trajectories.

To demonstrate how the prototype performs trajectory completion, we perform the following steps. First, in the *configuration* page, we load and split a raw trajectory data file into a training set and a test set. We then choose one or more trajectory completion algorithms to process the file. The proposed FP approach takes the training data as the third-party historical data and uses the several sparse GPS points (randomly chosen from the test data) to test the trajectory completion algorithms. When choosing an algorithm, we can tune the associated parameters. For example, to run a ST algorithm, we can configure the number of candidate points and the search radius. After the configuration, we can run the chosen algorithms to show their running result on the *display* page. The prototype can *a*) list the running result (such as accuracy and running time) of each algorithm. For example, on a sample data set, the FP approach reaches 95.9% accuracy in only 49 ms, over-performing the four other algorithms reaching at best 85.7% accuracy in 253 ms. In addition, the prototype can also *b*) show the completed trajectories on a real road network map, together with the ground truth which can be optionally selected by users. The screenshot example in Figure 2 shows a completed trajectory.

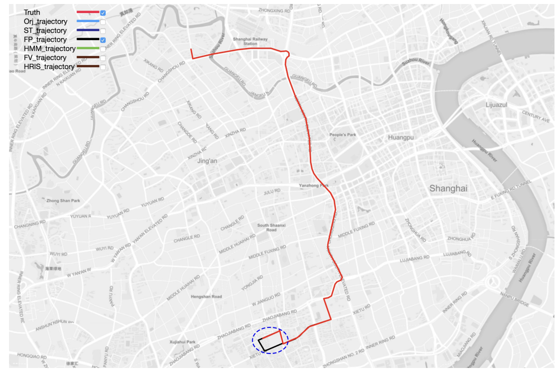


**Figure 2: Trajectory Visualization**

## 4 CONCLUSION

In this demo abstract, we introduce a frequent pattern-based trajectory completion approach, describe a prototype to evaluate trajectory completion algorithms, and demonstrate the advantages of the proposed approach over four previous works in terms of both accuracy and running time.

## ACKNOWLEDGMENTS

This work is partially supported by the National Natural Science Foundation of China (Grant No. 61572365, 61503286) and the Fundamental Research Funds for the Central Universities.

## REFERENCES

[1] Jian Hu and Yang-Li Xiang. 2008. A Fast Parallel Association Rules Mining Algorithm Based on FP-Forest. In *ISNN*. 40–49.
[2] Yukun Huang, Weixiong Rao, Zhiqiang Zhang, Peng Zhao, Mingxuan Yuan, and Jia Zeng. 2018. Frequent Pattern-Based Map-Matching on Low Sampling Rate Trajectories. In *MDM*. 266–273.
[3] Yin Lou, Chengyang Zhang, Yu Zheng, Xing Xie, Wei Wang, and Yan Huang. 2009. Map-matching for low-sampling-rate GPS trajectories. In *SIGSPATIAL*. 352–361.
[4] Paul Newson and John Krumm. 2009. Hidden Markov map matching through noise and sparseness. In *SIGSPATIAL*. 336–343.
[5] Hong Wei, Yin Wang, George Forman, Yanmin Zhu, and Haibing Guan. 2012. Fast Viterbi map matching with tunable weight functions. In *SIGSPATIAL*. 613–616.
[6] Kai Zheng, Yu Zheng, Xing Xie, and Xiaofang Zhou. 2012. Reducing Uncertainty of Low-Sampling-Rate Trajectories. In *ICDE*. 1144–1155.