Drive Smartly as a Taxi Driver

Yu Zheng, Jing Yuan, Wenlei Xie, Xing Xie
Microsoft Research Aisa, 4F Sigma Building, Haidian District, Beijing 100190, China
{yuzheng, v-jinyua, xingx}@microsoft.com

Guangzhong Sun
University of Science and Technology of China, Hefei, China
gzsun@ustc.edu.cn

***Abstract*—GPS-equipped taxis are mobile sensors probing the traffic flow on road surfaces, and taxi drivers are experienced drivers who can usually find out the fastest path to a destination based on their knowledge. In this demo, we provide a user with the practically fastest path to a destination at a given departure time in terms of taxi drivers' intelligence mined from historical GPS trajectories of taxis. We build our system, called T-Drive, by using a real trajectory dataset generated by over 33,000 taxis in a period of 3 months, and conduct both synthetic experiments and in-the-field evaluations. As a result, our method outperforms the real-time-traffic-based (RT) and the speed-constraint-based (SC) approaches in both efficiency and effectiveness.**

Keywords- Path finding, driving directions, T-Drive, GPS trajectory, landmark graph.

I. INTRODUCTION

A fast driving path saves not only the time of a driver but also energy consumption (as most gas was wasted in traffic jams). In practice, big cities with serious traffic problem usually have a large number of taxis traversing on road surfaces. For the sake of management and security, these taxis have already been embedded with a GPS sensor, which enables a taxi to report on its present location to a data center in a certain frequency. Thus, a large number of time-stamped GPS trajectories of taxis have been accumulated and easy to obtain.

Intuitively, taxi drivers are experienced drivers who can usually find out the fastest path to send passengers to a destination based on their knowledge [6] (we believe most taxi drivers are honest although a few of them might give passengers a roundabout trip). When selecting driving directions, besides the distance of a route, they also consider other factors, such as the time-variant traffic flows on road surfaces, traffic signals and direction turns contained in a route, as well as the probability of accidents. These factors can be learned by experienced drivers but are too subtle and difficult to incorporate into the existing routing engines. The Therefore, these historical taxi trajectories, which imply the intelligences of experience drivers, provide us a valuable resource to learn practically fast driving directions.

In this paper, we propose to mine smart driving directions from real-world historical GPS trajectories of taxis. This is a real application (termed T-Drive) that follows the strategy of "*Mobile*"+"*Cloud*", which has been regarded as the new trend of Internet Services. As shown in Figure 1, GPS trajectories are generated by (*Mobile*) taxis and aggregated in the *Cloud* where we can mine taxi drivers' intelligence from the trajectories. Then, the *Cloud* is able to leverage the intelligence to answer the queries from ordinary drivers using *Mobile* devices or other Internet users. Given a start and destination, our method can suggest the practically fastest path according to a user's departure time (refer to Figure 2)). Moreover, the *Cloud* can guarantee the suggested fastest route is the state-of-the-art by updating the data timely.

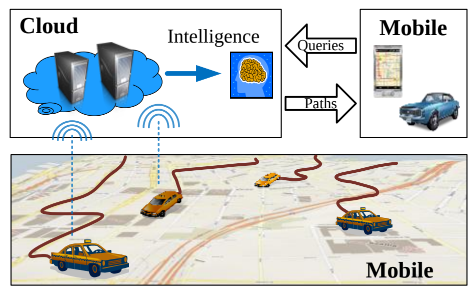


Figure 1. Application scenario of T-Drive

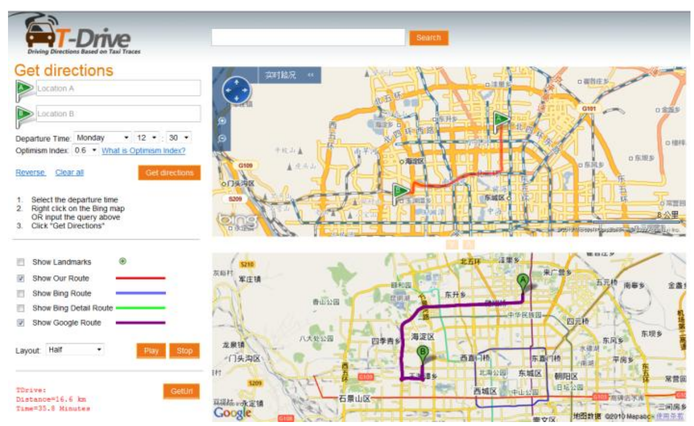


Figure 2. A web-based user interface of T-Drive

The contribution of this work lies in the following three aspects:

- We perform "*sensor data* → *driving direction*" instead of "*sensor data* → *traffic information* → *driving direction*", i.e., we do not need to explicitly build speed estimation models (for each road) that may introduce errors to routing processes.
- We propose the notion of landmark graph that can well model human knowledge of taxi drivers based on the real taxi trajectories and improve the online computation of path-finding.

- We build our system by using a real-world trajectory dataset, and evaluate the system by conducting both synthetic experiments and in-the-field evaluations.

II. ARCHITECTURE

As shown in Figure 3, the architecture of our system consists of three major components: Trajectory Preprocessing, Landmark Graph Construction, and Route Computing. The first two components operate offline and the third is running online.

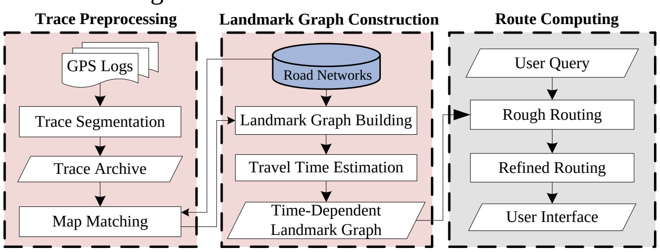
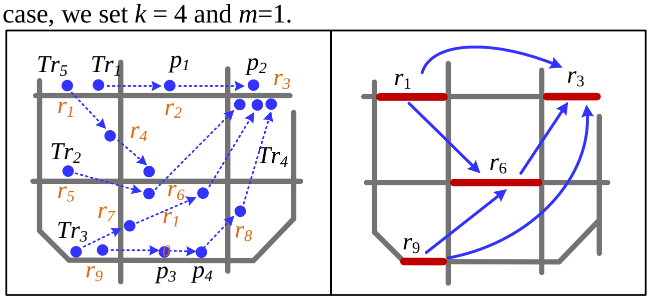


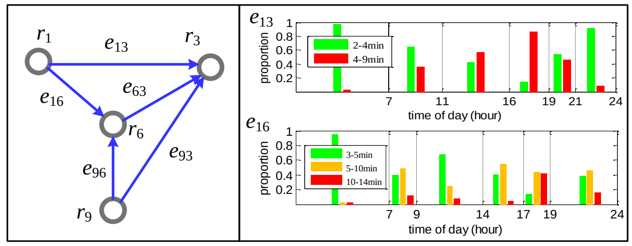
Figure 3. System overview

Trajectory Preprocessing: This component first segments GPS trajectories into effective trips, then matches each trip against the road network. 1) In practice, a GPS log may record a taxi's movement of several days, in which the taxi could send multiple passengers to a variety of destinations. Therefore, we partition a GPS log into some taxi trajectories representing individual trips according to the taximeter's transaction records. 2) Map matching: We employ our IVMM algorithm [5], which has a better performance than existing map-matching algorithms when dealing with the low-sampling-rate trajectories [7], to map each GPS point of a trip to the corresponding road segment where the point was recorded. As a result, a taxi trajectory is converted to a sequence of road segments.

Landmark Graph Construction: We separate the weekday trajectories from the weekend ones, and build a landmark graph for weekdays and weekends respectively. When building the graph, we first select the top-k road segments with relatively more projections (i.e., being frequently traversed by taxis) as the landmarks. Then, we connect two landmarks with a landmark edge if there are at least m trajectories passing these two landmarks. Later, we estimate the distribution of travel time of each landmark edge by using the VE-clustering algorithm. Now, a time-dependent landmark graph is ready for the online computation. Figure 4 illustrates an example of building a landmark graph. In this case, we set $k = 4$ and m=1.



A) Matched taxi trajectories B) Detected landmarks



C) A landmark graph D) Travel time estimation

Figure 4. An example of building a landmark graph

Route Computing: Given a query (q_s, q_d, t_d), we carry out a two-stage routing algorithm to find out the fastest route. In the first stage, we perform a rough routing that search the time-dependent landmark graph for the fastest rough route represented by a sequence of landmarks. In the second stage, we conduct a refined routing algorithm, which computes a detailed route in the real road network to sequentially connect the landmarks in the rough route. Figure 5 demonstrates the routing concept.

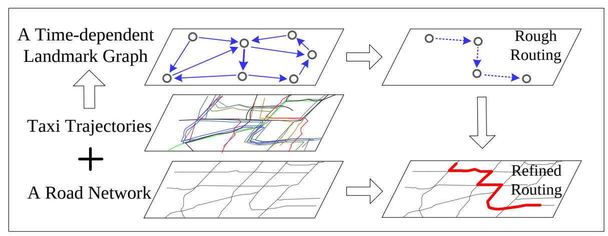
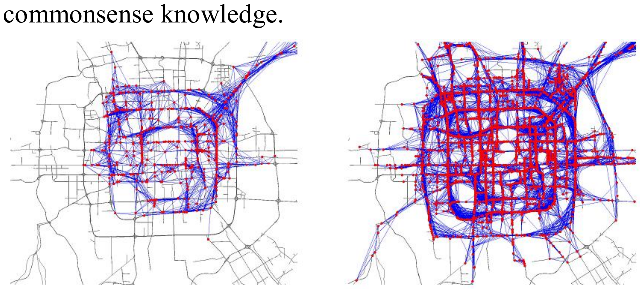


Figure 5. A demonstration of the routing process of T-Drive

III. EVALUATION

We evaluated our approach by using a real trajectory dataset created by over 33,000 taxis in Beijing in a period of three months. Both simulation experiments and in-the-field study have been carried out. When conducting the in-the-field study, we utilize some GPS trajectories (from GeoLife project [11]) recording real users' driving histories in the real world to test the effectiveness of our method. The dataset has been released to the public [1], and can be used freely for research purpose. During the evaluation, we compare our method with the real-time traffic-based (RT) and the speed-constraint-based (SC) approaches.

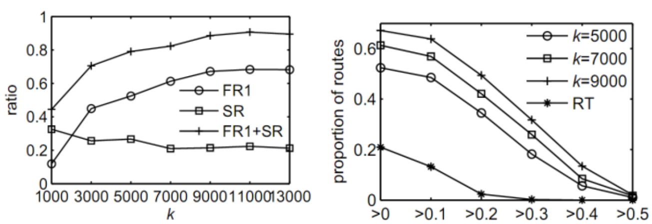
Figure 6 visualizes two landmark graphs when k = 500 and k = 4000. The red points represent landmarks and blue lines denote landmark edges. Generally, the graphs well covers Beijing city, and its distribution follows our commonsense knowledge.



A) k=500 B) k=4000

Figure 6. Landmark graphs (k=500 and 4000)

Figure 7 shows the results of the synthetic evaluation, where FR1 represents how many routes suggested by our method are faster than that of baseline, and FR2 reflects to what extents our routes are faster than the baseline's. Here, both our method and the RT approach use the SC method as a baseline. As depicted in Figure 7 A), over 60% of our routes are faster than that of the SC approach, and about 20% routes share the same results. Meanwhile, FR1 is being enhanced with the increase of k when $k < 9000$, and becomes stable when $k > 9000$. That is, it is not necessary to keep on expanding the scale of a landmark graph to achieve a better performance. Figure 7 B) plots the FR2 of ours and RT. For example, when k = 9000, over 50% routes suggested by our method are at least 20% faster than the SC approach. This clearly outperforms the RT approach, which only has 5% routes WITH the same FR2.



A) FR1 and SR B) FR2

Figure 7 Synthetic evaluation result

In Table 1, the symbol Δ stands for the difference value of distance or duration. R1 represents the ratio of our routes outperforming the baseline (Google Map), and R2 denotes to what extent our routes are beyond that of the baseline. For example, 80.8% of the routes suggested by our T-Drive system are faster than that of Google Map and on average our routes saves 11.9% time (T-test: $p < 0.001$).

Table 1: In-the-field evaluation

	T-Drive	Google	△	R1	R2
Distance	13.91km	15.56km	1.65km	0.517	0.106
Duration	25.80min	29.28min	3.48min	0.808	0.119

IV. Related Work

Zheng et al. [8, 9, 10] propose several novel approaches to learn the transportation modes from GPS data. Paper [2, 4] presents a probabilistic based method to predict a driver's destination and route based on historical GPS trajectories. Although paper [2] also uses GPS trajectories generated by 25 taxis, this work aims to predict a driver's destination instead of providing the fastest route that a user can follow. Paper [3] computes the fastest route by taking into account the driving and speed patterns learned from historical GPS trajectories. Our method differs from this work in the following aspects. First, we do not explicitly detect speed and driving patterns from the taxi trajectories. Instead, we use the concept of landmarks to summarize the intelligence of taxi drivers. The notion of landmarks follows people's natural thinking patterns, and can improve efficiency of route finding. Second, our approach is driven by the real dataset while paper [3] is based on the assumption of synthetic data. Actually, the real data causes some challenges, e.g., low sampling rate and sparseness of trajectories. Moreover, we consider the time-variant and location-dependent properties of real-world traffic flows.

V. Conclusion

This demo shows an approach that finds out the practically fastest path to a destination at a given departure time based on taxi drivers' intelligence learned from the historical taxi trajectories. The results show that our method significantly outperforms both the speed-constraint-based and the real-time-traffic-based method in the aspects of effectiveness and efficiency. More than 60% of our routes are faster than that of the speed-constraint-based approach, and 50% of these routes are at least 20% faster than the latter. On average, our method can save about 16% time of a trip, i.e., 5 minutes per 30-minutes driving.

References

[1] GeoLife GPS Trajectories: http://research.microsoft.com/en-us/downloads/b16d359d-d164-469e-9fd4-daa38f2b2e13/default.aspx

[2] B. Ziebart, A. Maas, A. Dey, and J. Bagnell. Navigate like a cabbie: Probabilistic reasoning from observed context-aware behavior. In Proc. Ubicomp, pp. 322-331.

[3] H. Gonzalez, J. Han, X. Li, M. Myslinska, and J. Sondag. Adaptive fastest path computation on a road network: A traffic mining approach. In Proc. VLDB, 2007.

[4] J. Krumm and E. Horvitz. Predestination: Inferring destinations from partial trajectories. LNCS, pp. 243-260.

[5] J. Yuan, Y. Zheng, C. Zhang, and X. Xie. An interactive-voting based map matching algorithm. In: IEEE conference on Mobile Data Management (MDM), pp.43-52. 2010.

[6] J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun., Y. Huang. T-Drive: Driving directions based on taxi trajectories. In Proc. of ACM SIGSPATIAL GIS 2010.

[7] Y. Lou, C. Zhang, Y. Zheng, X. Xie, Y. Huang. Map-Matching for Low-Sampling-Rate GPS Trajectories. In: ACM SIGSPATIAL Conference on Geographic Information System, pp. 352-361(2009).

[8] Y. Zheng. L. Liu, L. W, X. Xie. Learning transportation mode from raw gps data for geographic applications on the web. In Proc.WWW, pp. 247-256, 2008.

[9] Y. Zheng, Q. Li, Y. Chen, X. Xie. Understanding mobility based on GPS data. In Proc. Ubicomp, pages 312-321, 2008.

[10] Y. Zheng, Y. Chen, Q. Li, X. Xie. W. Y. Ma. Understanding transportation modes based on GPS data for Web applications. ACM Transactions on the Web, 4(1):1-36, 2010.

[11] Y. Zheng, X. Xie, W. Y. Ma, GeoLife: A Collaborative Social Networking Service among User, location and trajectory. IEEE Data Engineering Bulletin. 33, 2, 2010, pp. 32-40.