2014 IEEE 17th International Conference on
Intelligent Transportation Systems (ITSC)
October 8-11, 2014. Qingdao, China

Mining freight truck's trip patterns from GPS data

Jun Huang[†] Li Wang[†] Chen Tian[⋆†] Fan Zhang[†] Chengzhong Xu[†§]
[†]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China
[⋆]Department of Electronics and Information, Huazhong University of Science and Technology, China
[§]Department of Electrical and Computer Engineering, Wayne State University, MI, USA
[†]{jun.huang, wangli2, zhangfan, cz.xu}@siat.ac.cn, [⋆]tianchen@hust.edu.cn

***Abstract*— Land carriage is important for nowadays large goods transportation. There are three major roles in a land carriage order: guests, companies and trucks. A significant problem, for logistics companies, is the lack of trip control. Continuous monitoring the trips of freight trucks is necessary. First of all, it does help logistics companies to prevent the fraud behaviours. But more importantly, it help us understanding the trip patterns of freight truck transportation. The development of Global Position System(GPS) and wireless communication together enables the possibility to analyze large scale freight trips. In this paper, we study a large GPS trajectory dataset of 14654 freight trucks from a 3rd-party company, which helps logistics companies monitoring those freight trucks. We propose a method to extract trips from the GPS trajectories of freight trucks and mine the travel patterns, both collectively and individually. To the best of our knowledge, this is the first mining work of large scale freight truck trajectory data.**

I. INTRODUCTION

Land carriage is important for nowadays large goods transportation. Compared with train-based approach, the truck-based transportation is mostly more flexible, convenient and fast. Hence for medium sized demand, freight truck is the major choice of business owner.

There are three major roles in a freight truck operation: guests, companies and trucks. The guests, distributed all over the country, place orders to logistics companies to deliver their goods. A company decomposes the orders to trips, and allocates trips to each freight truck. Note that in many countries (for example, China), a logistics company and the truck owners are usually loosely coupled: the company does not own any freight truck; there are small transport fleets and individual truck owners; the business relationship is somewhat similar to the connections between the taxi drivers and taxi companies.

A significant problem, for logistics companies, is the lack of trip control. The contract of a trip is usually determined: origin, destination, delivery time and payment. However, the real execution of a trip can be affected by many out-of-control factors: congestion condition, weather, etc. For example, a trip is delayed and the company is fined by the guest; the reported cause by the truck driver (also the owner) is congestion somewhere in the middle; while the real reason is that the driver detours around a tolled highway, which he is supposed to take, to save money for himself.

Continuous monitoring the trips of freight trucks is necessary. First of all, it does help logistics companies to prevent the fraud behaviours mentioned above. But more importantly, it helps us understanding the trip patterns of freight truck transportation. The benefits are multi-fold: mining existing trip patterns can help logistics planning such as path choosing [1]; the business transportation flows also open a new field for economic condition analysis [2].

The development of Global Position System(GPS) and wireless communication together enable the possibility of large scale freight trip data analysis. Currently, GPS is widely equipped in freight trucks in China; the collected information is transmitted back to the company center in real-time via 3G/4G.

In this paper, we study a large freight truck GPS dataset from a 3rd-party company, which helps logistics companies monitoring those freight trucks. We mine trips and analyze the mining results of both freight trucks collective and individual patterns.

To the best of our knowledge, this is the first mining work of large scale freight truck trajectory data. The main contributions of this paper are:

- We propose a method to extract trips from the GPS trajectory of freight trucks. To cluster trips into paths, we propose a data-customized measure for clustering. From the analysis results, it can be concluded that the trip patterns are reasonable and the algorithm performs well.
- We find Freight truck drivers works at weekends if needed; they have rest for one day at the end of a month. For a longer holiday, such as Chinese Spring Festival, they can have rest for over half a month.
- The operational patterns of freight truck drivers vary according to the distance between the order's origin and destination cities. Drivers who accept intra-province orders can finish their work in one day, and usually they come back home early. While for inter-province orders, drivers stay on truck overnight, and usually they finish their trips at evening or midnight.
- Truck drivers' knowledge could really help inter-city trip planning. Given two cities, freight trucks have a default path to finish an order; we can consider it the best path to connect the two cities. When there are special limitations such as midway unload/reload, they instead choose a different path.

The rest of this paper is organized as follows. In section II we give a description about the dataset and our analyse methods. Section III demonstrates the daily life of freight truck drivers through several collective patterns of logistics trips. In section IV, a special line that connects the south and

978-1-4799-6077-4/14/$31.00 ©2014 IEEE

TABLE I
FREIGHT TRUCK DATA DESCRIPTION

Field	Sample	Memo
truck ID	7077BPGFSU	
Longitude	115839512	10^{-6} Degree
Latitude	31814818	10^{-6} Degree
Mileage	30913	10^{-1} Kilometre
Speed	0	KM/h
Direction	50	0 to 359 Degree, north is 0
Height	0	Height above sea level
Time	14-5-30 0:1:23	

east China is deeply investigated; the purpose is to show how freight truck drivers plan the path of a trip. We also list some related works at section V. The paper is conclued at the last section.

II. DATASET AND TRIP MINING

A. DATASET

Our research is based on a real time dataset from an information service company which focus on the transparent management of logistics. They help logistics companies equip in-vehicle GPS devices on freight trucks for monitoring and security assurance. For each freight truck, a positioning record would be sent from the vehicle through wireless network in every 30 seconds , thus the whole trajectory of the truck could be constructed. The GPS data description and sample is showed in table I. Each day we receive $\sim$ 6G GPS record data from 35,000 freight trucks since October 1st, 2013.

B. TRIP MINING

As there is no direct sign shows when does a driver start or end a trip for his order, we develop a trip mining algorithm to obtain goods delivery activity between cities from the trajectory of each truck.

We think two factors are signals to separate two consequence trips of a freight truck:

- More than 2 hours' parking behaviour, as it indicates loading/unloading activity or rest.
- More than half an hour's disconnection of GPS device, as it is probably resulted by flame-out.

Using factors listed above we could cut a trajectory of a freight truck into several consequence record sets, each consequence record set could be considered as a trip candidate. We then drop those trips that meet following conditions:

- Trips that are the first or last one in the trajectory of each truck in the dataset, as they are not complete.
- Trips whose average speed exceed (30, 120) Kilometres/Hour as in practice they are not a valid trip.
- Trips whose duration are less than 1 hour.

We apply this method on half year's data and drop trips that start at the first and last month as they might not be complete. In addition trucks that focus on intra-city logistics would not be discussed in this paper. Totally we track 189555 trips of 14654 freight trucks in four months that cover 320 cities in China. To demonstrate the whole logistics network we could track, we plot an origin-destinatnion(OD) network in figure 1. The node of the network represents a city of China, colored by administrative district of the country. If there are trip flow connects two cities, an edge will be placed between the two nodes, and the edge width is proportion to the number of trips. We think the result of trip extraction is acceptable as the network covers most cities in Southeast part of China, which is the most developed area of the country.

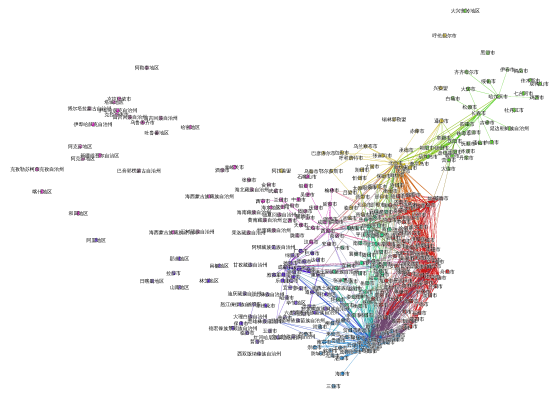


Fig. 1. Freight truck Origin-Destination network

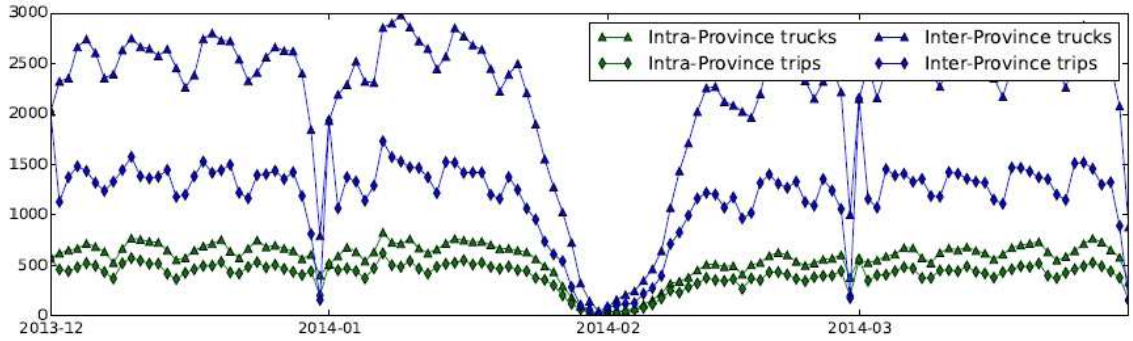
III. COLLECTIVE PATTERN

In this section, we focus on the collective pattern of freight truck trips. Because of the engrossment of logistics company in practise, we will classified our dataset into intra-province trips and inter-province trips by checking whether the origin and destination cities are in the same province in following analysis.

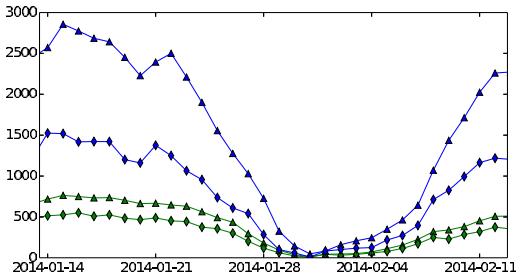
A. Distribution Alone Days

To show the distribution along days. We calculate two types of figure. Firstly we would like to know how the number of trips distributed over a long time. Thus we grouped trips in four months by their start time and count the number. We think it shows the demand of land carriage. In addition we are interested in the supply of drivers. Then we calculate how many freight trucks on service each day. Because normally one freight truck would have fixed number of drivers, we think it reflects the total number of drivers that are working for a day. Figure 2.a displays both the number of trips starts and the number of trucks on-service at each day.

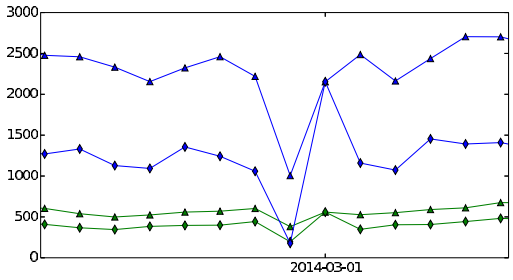
From figure 2.a we could see that the number of trips are affected by two factors: Periodicity and Special event. At December and March, count of trips changes periodically. Given the second day of December is Monday, we could conclude that the figure raises from Monday and peaks at Wednesday, then drops until the end of the week. That means there would be less orders at weekend than weekday. But we should admit that on a usual day the number of trucks on-service would never be less than 2500, so a number of drivers would not have a rest at weekend.



(a) Four months



(b) Chinese new year



(c) The last day of a month

Fig. 2. Trip and Freight truck number over days

But around February it shows a much different number. In figure 2.b a sharp decline could be observed since Jan 22nd in 2014. We think it is because of the Chinese new year starts from Jan 31st. We could see that both the number of inter-province and intra-province trips drop at lest 10 days before the festival as well as the number of trucks on service. At the eve of Chinese new year, all figures reach their minimum, and during the holiday the effect of periodicity will be eliminated.

Another interesting fact of the distribution is that orders at the last day of the month would probably be delayed to the next day, thus usually the maximum and minimum value at a month will be obtained by the first and the last day of the month. Figure 2.c shows the figures around March 1st 2014. The number of trips and the number of on service trucks both drop at February 28th, then meet at the next day. Which means drivers in logistics area would stop accepting order at the last day of the month, then they might have a rest once a month.

Compare to the number trips and the number of trucks on service, it could be concluded that the inter-province trips behave differently from intra-province ones. Each day there are around 1250 trips start for an inter-province order, but in fact there would be nearly 2500 freight trucks are running on their way. That means an order with origin and destination cities in different provinces would usually last for more than one day. However, the intra-province one have similar figures, therefore we think drivers for these orders would not likely to driver overnight.

B. Operational Duration and Mileage Distance

We are interested in how much time does a trip cost and how long should a trip go, and their relationship. Thus we calculate the operational duration of each trip using its start time and end time. As well we use the ”Mileage” field of the dataset to obtain the mileage distance of a trip.

In practice we observe wide range of operational duration and distance of trips. The longest trip of the dataset runs 5611 kilometres, starting from Kashgar city in the north-west of China, ending at Foshan city which is located at the south-eastern part of the country. The average distance and duration of all trips are listed as follow:

TABLE II
AVERAGE FIGURES OF DATASET

	Distance(KM)	Duration(Hour)
Intra-province	309.7	10.4
Inter-province	948.1	22.5

To show the distribution of duration of the whole dataset, we plot the cumulative-distribution of duration of our dataset in figure 3. In this figure we found that most intra-province trips would have a limited time up to 24 hours, which means drivers focusing on one province can come back home every day. However, trips that have origin and destination in different province has a more dispersive distribution. It shows that when drivers accept an inter-province order, they usually should stay overnight in the freight truck in group

for alternate driving.

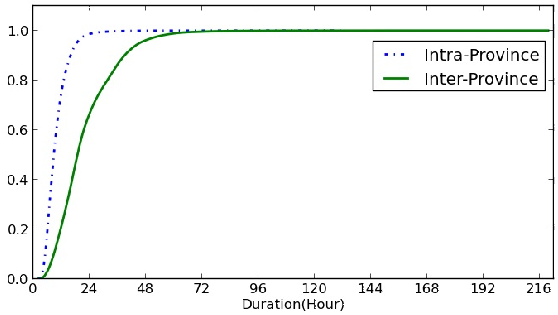


Fig. 3. CDF of trip duration

To give a straightforward understanding of the correlation between distance and duration of trips. We plot a sample of dataset on figure 4. From the scatter diagram we find a lower bound of trip order. Most points lie on the upper side of the line on the figure, which fits the 50 KM/h, indicating the companies in logistics area would order a maximum time of 1 hour for each 50 kilometres, in spite of load/unload and rest stop.

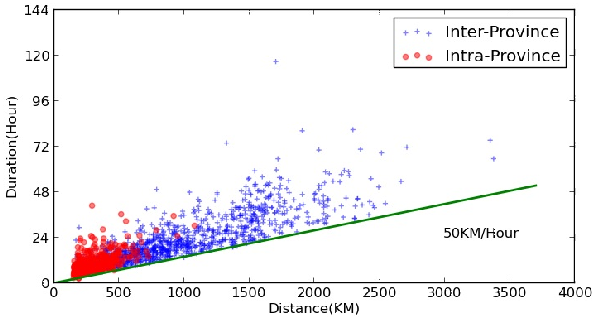


Fig. 4. Distance and Duration of a sample of trip dataset

C. Distribution in one day

We also wonder when a driver start and end a trip at a day while getting an order. To calculate an unbiased figure we use the average counting of trips starts through December 2nd to 29th as they do not include the first and last day of the month. Figure 5 shows how start time of trips locate in one day by hour. We could conclude that most drivers should start their work around 7 o'clock. However, drivers of inter-province trips would not probably start a trip in the evening.

But the arrival time gives another figure on both types of trip. There are two peaks at a day for freight truck drivers to hand over goods: the noon peak around 11 AM and midnight peak around 2 AM. We think the midnight peak is caused by the policy that some heavy trucks are forbidden at daytime in downtown. It seems that drivers for intra-province trips would be more likely to end their trips at noon rather than at night, which contradicts to those for inter-province that drive longer distance.

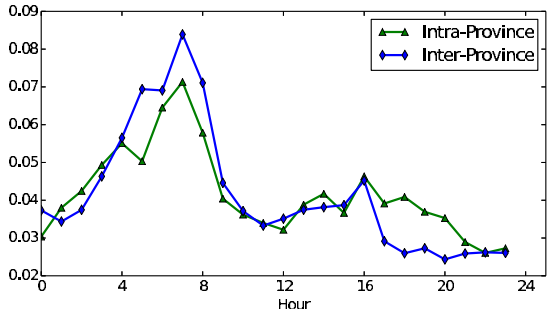


Fig. 5. Number of trip starts over hours in one day

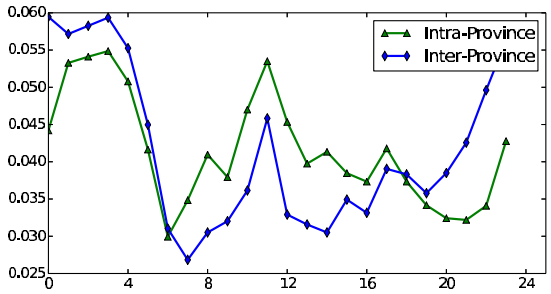


Fig. 6. Number of trip ends over hours in one day

D. Summary

To summary, our research shows the daily life of freight truck drivers. Usually they might have a job everyday, start driving around 7 AM. Someone whose order ends in a nearby city could stop at 11 AM, while some others should handle goods at night. But they are not the hardest one, as a portion of drivers have to drive overnight with their colleagues. Drivers in logistics area usually should work on both weekday and weekend, while at the end of a month they will have a day for rest. The longest holiday might be the Chinese new year festival, at that time freight truck drivers will stop working for more than half month.

IV. INDIVIDUAL ANALYSIS

In this section, we will focus on trips that start from Shenzhen, a city lies beside Hong Kong, and end at Suzhou, which is not far away from Shanghai, to show how a driver make decision on trip planning given an logistics order.

A. Path Diversity

In order to understand how a driver plan his trip for an order, we plot all trajectory points of all trips from Shenzhen to Suzhou in figure 7. As a transportation line that connects the south and the east part of China, it is the most frequent line in our dataset. Averagely, a trip goes from Shenzhen to Suzhou costs 40 hours as well as 1600 kilometres including intra-city delivery and temporary parking.

It could be concluded from figure 7 that a driver might choose various paths to finish an order from Shenzhen to Suzhou. Some drivers will go along the coast across Fujian province, while some drivers would like to select the west path through Jiangxi and Anhui province. Some paths would

$$LCSS(R,S) = \begin{cases} 0, & if \quad R \quad or \quad S \quad is \quad empty, \\ 1 + LCSS(R.tail, S.tail), & if \quad isNearby(R.head, S.head) \\ max(LCSS(R, S.tail), LCSS(R.tail, S)), & otherwise \end{cases} \tag{1}$$

have identical beginning and fork to two different roads and vice versa.

Fig. 7. All Records in Trips from Shenzhen to Suzhou

B. TRIP CLUSTERING

We are interested in the variation of paths planned by drivers. However, we could not directly tell whether two trips go the same path because of two trips of the same path might capture different sampling data point sets. Thus it is necessary to cluster trips into paths for further research.

As the key part of cluster algorithm, the definition of trip similarity should be deeply investigated to fit the characteristics of dataset. We think the best one should have following features:

- Noise tolerant
- Time uncorrelated

In practice we found that noise is inevitable, but map-matching is time-consuming [3], thus it does not fit our large scale dataset. To overcome the outlier we decide to use the Longest Common Sub-Sequence (LCSS) [4], [5] to define trip similarity.

Equation 1 shows how to calculate LCSS value of two trajectory. In this formula R and S are two trip trajectories that are represented as two GPS positioning point sequences, and R.head means the first GPS point of the trajectory while R.tail is the rest part of it. We think each positioning point is like a word in neural language area, and the trajectory could be regarded as a paragraph written by freight truck. If two points from two different trips are not far away from each other, in our opinion the two trucks speak an identical word, then the longer the common subsequence of the two trajectories have, the more likely the two trips go identity path. Because few typos would not sharply affect the LCSS value, therefore it is noise tolerant.

Because we would like to focus on the path of a trip in spatial but ignore the temporal detail, the cluster algorithm should be time uncorrelated. Therefore we design a pre-processing step at the beginning of trip cluster algorithm. For each trip, we only keep the first record in every square kilometre. So when a truck temporally pause at somewhere, only one record will be put into the cluster algorithm.

The reason why we choose 1 square kilometre is that a freight truck would not likely to run more than 1 kilometre in 30 seconds. For similar reason we define the "isNearby" function as:

$$isNearby(a,b) = Distance(a,b) < 1(kilometre). \tag{2}$$

In practice we use a normalized LCSS value as the similarity of two trips, because the result of LCSS is unbounded. But differ from [5] we think the length of trip is still an important factor as two trips with different length would not likely to be the same path. Therefore the normalized LCSS value is defined as:

$$Similarity(R,S) = \frac{LCSS(R,S)}{max(length(R), length(S))} \tag{3}$$

C. Frequency Distribution

But the path frequency is non-uniform distributed. We calculate the normalized LCSS value of each pair of trips, then construct a trip similarity network. In figure 8, each node represents a trip from Shenzhen to Suzhou, a edge between two nodes indicates that the similarity of them are greater than 0.9. The color of nodes shows the result of trip clustering algorithm.

It could be easily recognized that there is a giant connected component in the network, which means a number of trips would fall into one path. Followed by the big group there are several groups with much smaller scale, as well as some isolate trips connect no-one. To give a quantizing illustration we display the distribution in table III. Clusters whose size are less than 10 are grouped into "Other" category. It could be concluded that when a driver getting an order through the two cities, he is much likely to choose the path represented by cluster 1.

Thus we wonder why so many drivers would like to choose a specific path. In figure 9 we draw the trajectory of the four clusters listed above. Compare with paths of other clusters, the path represented by cluster 1 seems like to be the shortest path between Shenzhen and Suzhou. We think it explains why drivers would like to add this path to their plans: the path is the optimal path from Shenzhen to Suzhou.

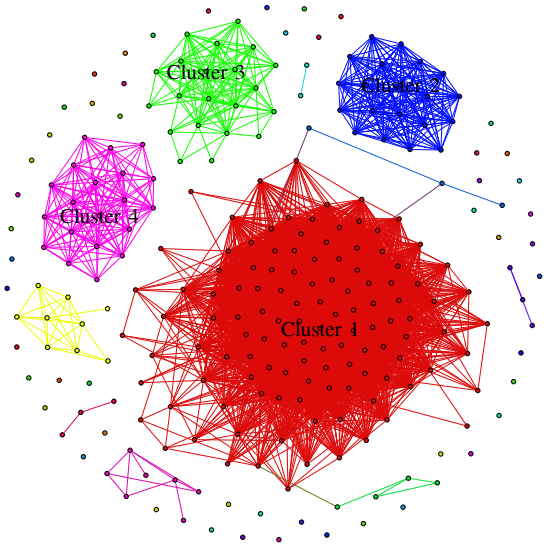


Fig. 8. Trip network

TABLE III
CLUSTER SIZE DISTRIBUTION

Cluster	Proportion
Cluster 1	45.2%
Cluster 2	8.0%
Cluster 3	7.7%
Cluster 4	7.7%
Other	31.4%

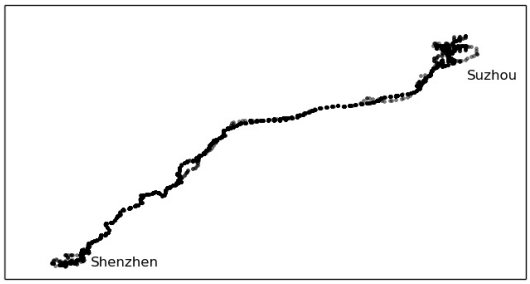
D. Detour Analysis

We are also interested in the reason why freight truck drivers would like to choose a much costly path. Therefore we focus on trips in cluster 2, which draw a much longer trajectory than trips in cluster 1. We capture all temporary parking behaviours as well as their durations of these trips. In figure 10 there are four triangles on the path, each of them represents a parking point, and the size of the triangle is related to the total duration of parking.

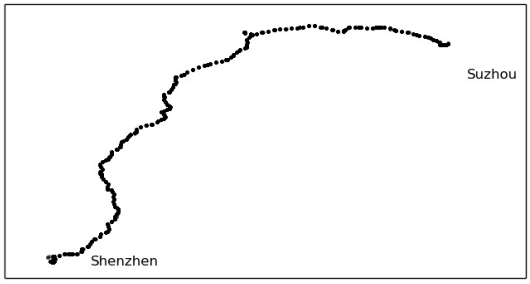
Trips that go this path would temporarily stop at four places. The 1st one locates in Xinyu City in Jiangxi province, the 2nd and 3rd one are in Hefei and Nanjing, the capital city of Anhui Province and Jiangsu Province respectively. The last one lies to the destination city, Suzhou. In this figure it should be pointed out that drivers who choose this path would stop for a related long time at Hefei, which is the anchor point of the path. We think it account for the detour feature of the trip. The order might require special unload/reload at Hefei city, thus freight truck drivers shall add the city to the path while trip planning.

E. Summary

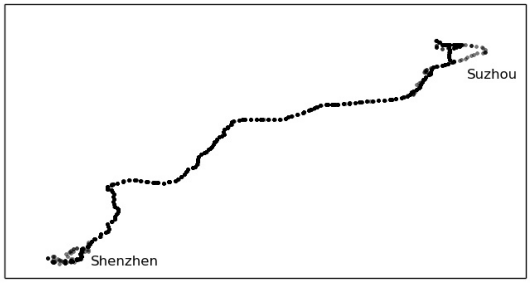
Given an origin and a destination city, freight truck drivers have a default path to finish the order connecting the two cities. We think it is a path obtaining best efficiency, it is the wisdom of crowd. But if some special limitations are given to the order like midway unload/reload, drivers would like



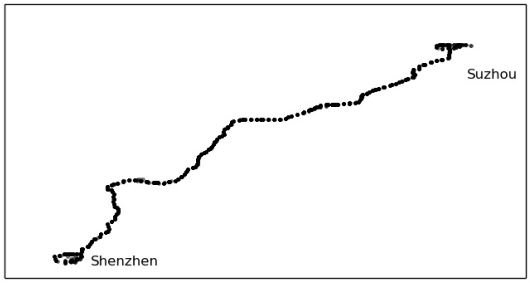
(a) Cluster 1



(b) Cluster 2



(c) Cluster 3



(d) Cluster 4

Fig. 9. Result of trip cluster algorithm

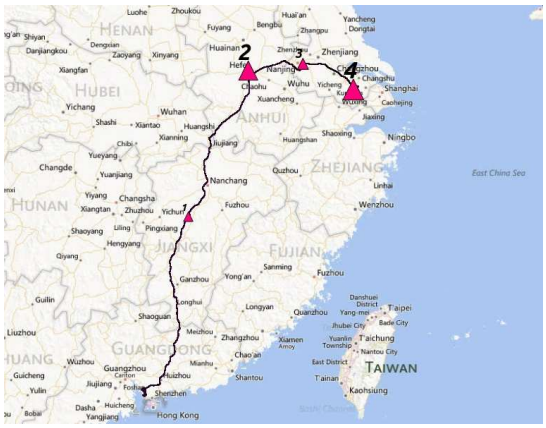


Fig. 10. Cluster 2 with temporary parking

planing another path for the order.

V. Related Work

Because of the availability of lowcost location-aware equipment and data storage, mining vehicle patterns from GPS log data have been widely investigated in recent years, such as Taxi [6] [7], Bus [8], Private truck [9]. But in terms of freight truck in logistics [10], only several jobs could be found. McCormack and his team discuss how to build a gps-based truck measure platform [11], and how to extract transportation metrics from freight truck data [12]. The most related work from our research is done by Liao [13], who presented a case study using one year of truck GPS location data along the I-94/I-90 corridor in USA, giving some statistical figures like truck speed, volume, parking behaviours. But they focused on the on-road feature in the path from Twin Cities to Chicago, instead of national wide trip patterns.

Some probable applications on these data have been discussed. Bassok etc. [14] wonder if freight truck on roadway could be used for freight forecasting. Using the speed extracted from freight truck trajectory, Zhao etc. [15] develop a probabilistic method to identify and rank roadway bottlenecks.

VI. CONCLUSION

In this paper, we uncover the trip patterns of inter-city roadway goods transportation, by giving an insight understanding on a large scale freight truck trajectory dataset. From a collective perspective, the freight truck drivers shall setting out early and returning late. And they should work at weekend if necessary. Usually they could rest once a month, and for longer holiday, they must wait for the Chinese new year. Focusing on an individual line that achieve maximum frequency, we think the experience of truck drivers would really help improve the performance the performance of inter-city trip planning algorithm. The most frequency path chosen by freight truck drivers should be the most efficient way to connect two cities, it could be recommended to the trip planer. If the path should go through a specific city, these drivers will have another option to fulfil the requirement, thus it could be the customized recommendation.

In terms of future work, we will attempt to utilize the historical truck trajectory data to plan inter-city travel, as we think these roadway transportation experts could really help.

Acknowledgment

The authors would like to thank anonymous reviewers for their valuable comments. This work is partially supported in part by "National Natural Science Foundation of China (No. 61100220, No. 61202107, No. 61202303)" and by the "Fundamental Research Funds for the Central Universities", and by NSF under grant CCF-1016966.

References

[1] J. Yuan, Y. Zheng, X. Xie, and G. Sun, "Driving with knowledge from the physical world," in *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2011, pp. 316–324.

[2] "Keqiang ker-ching," *The Economist*, Dec. 2010. [Online]. Available: http://www.economist.com/node/17681868?story_id=17681868

[3] M. A. Quddus, W. Y. Ochieng, and R. B. Noland, "Current map-matching algorithms for transport applications: State-of-the art and future research directions," *Transportation Research Part C: Emerging Technologies*, vol. 15, no. 5, pp. 312–328, 2007.

[4] L. Chen, M. T. Özsu, and V. Oria, "Robust and fast similarity search for moving object trajectories," in *Proceedings of the 2005 ACM SIGMOD international conference on Management of data*. ACM, 2005, pp. 491–502.

[5] M. Vlachos, M. Hadjieleftheriou, D. Gunopulos, and E. Keogh, "Indexing multi-dimensional time-series with support for multiple distance measures," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2003, pp. 216–225.

[6] T. Jia and B. Jiang, "Exploring human activity patterns using taxicab static points," *ISPRS International Journal of Geo-Information*, vol. 1, no. 1, pp. 89–107, 2012.

[7] Y. Liu, C. Kang, S. Gao, Y. Xiao, and Y. Tian, "Understanding intra-urban trip patterns from taxi trajectory data," *Journal of Geographical Systems*, vol. 14, no. 4, pp. 463–483, 2012.

[8] J. G. Strathman, K. J. Dueker, T. Kimpel, R. Gerhart, K. Turner, P. Taylor, S. Callas, D. Griffin, and J. Hopper, "Automated bus dispatching, operations control, and service reliability: Baseline analysis," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1666, no. 1, pp. 28–36, 1999.

[9] A. Bazzani, B. Giorgini, S. Rambaldi, R. Gallotti, and L. Giovannini, "Statistical laws in urban mobility from microscopic gps data in the area of florence," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2010, no. 05, p. P05001, 2010.

[10] H. Kargupta, K. Sarkar, and M. Gilligan, "Minefleet®: an overview of a widely adopted distributed vehicle performance data mining system," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2010, pp. 37–46.

[11] E. D. McCormack and T. Northwest, "Developing a gps-based truck freight performance measure platform," Transportation Northwest, University of Washington, Tech. Rep., 2010.

[12] E. McCormack, "Developing transportation metrics from commercial gps truck data," Tech. Rep., 2011.

[13] C.-F. Liao, "Using archived truck gps data for freight performance analysis on i-94/i-90 from the twin cities to chicago," 2009.

[14] A. Bassok, E. D. McCormack, M. L. Outwater, and C. Ta, "Use of truck gps data for freight forecasting," in *Transportation Research Board 90th Annual Meeting*, no. 11-3033, 2011.

[15] W. Zhao, E. McCormack, D. J. Dailey, and E. Scharnhorst, "Using truck probe gps data to identify and rank roadway bottlenecks," *Journal of Transportation Engineering*, vol. 139, no. 1, pp. 1–7, 2012.