Towards Constructing a Driver Management System Based on Large-scale Driving Operation Records

Daisaku Yokoyama and Masashi Toyoda
The University of Tokyo
Meguro-ku Komaba 4-6-1, Tokyo, Japan, 153-8505
Email: {yokoyama, toyoda}@tkl.iis.u-tokyo.ac.jp

***Abstract*—We introduce our developing system which can analyze drivers' driving behavior collected from vehicle recorder and other datasources such as weather reports and road maps. We show some performance issues while (pre-)processing large-scale data and discuss the requirements for the practical system.**

I. Introduction

Driver management system has played an important role in the transportation industry. Transport companies typically manage their drivers by using demographic information to estimate their safety; however, such information overlooks the current condition and improvements in the skill of the driver.

We are developing a system that analyzes driver characteristics based on their driving behavior. Many transportation companies have introduced dashboard cameras (dashcams) and/or vehicle data recorders (which collect GPS, velocity, and acceleration data) into their fleets. Our system collects vehicle recorder information and combines them with other datasources such as weather reports and road maps. Kinematic variables (maximum velocity, acceleration, etc.) related to their driving operations (braking, steering, etc.) are collected by vehicle recorders. Many kinds of transportation-related information, such as weather, road structure, the degree of traffic congestion, are also available nowadays. Combining these circumstantial data with driving records will be helpful to investigate driving behavior. We are constructing a scalable analyzing platform for the large-scale driving and circumstantial records which can analyze them in the multiple points of view (geographical, historical, individual, etc.).

We already proposed a method that classifies drivers based on long-term driving records, which successfully reflects drivers' accident history [1]. We try to apply our methods to help driver management process. If our method successfully indicates some clues for the potential risk related to accidents, fleet managers can focus on the driver and may be able to give some advice for keeping safety drive.

In this paper, we introduce our developing system for analyzing driving behavior. Typical driving related records should be cleaned and pre-processed to analyze and it requires massive computing power. We apply parallel computing to handle large-scale records practically. Many studies have been developed to handle large-scale geo-temporal data effectively [2][3][4].

We show some examples of analysis on the large-scale vehicle recorder records and discuss some practical issues of the system.

II. Driving record analysis system

While analyzing large-scale practical driving data, we found several issues for efficient computing. We explain such issues and discuss the requirements for the system in this section.

A. System design

We stored large-scale driving operation records and trajectories from vehicle recorder and X-band Radar data that has fine-grained weather information. Before analysis, vehicle records should be cleaned up inevitable noise and matched to the road map; analysis methods often need circumstantial information about the driving operations such as road structure or the number of lanes. Vehicle records are also be matched to the X-band Radar data; each driving operation record contains GPS data and time information, thus we could perform spatial- and temporal- matching with the other datasources.

After the pre-processing, we can utilize driving records to many problems. Understanding drivers' behavior is one example. Long-term driving records enable to make a classifier that can separate drivers based on their accident history. Understanding dangerous road structure[5] is another example. We match each driving operation records to the road segments or intersections and analyze its distributions to understand the relationship between driving behaviors and accident history at the area.

For utilizing driving record information to many problems, we should construct flexible analysis platform; we should analyze data from the combination of multiple – individual, geographical, and temporal – point of view. Combining heterogeneous datasources is also desirable; weather information and accident records are helpful to understand practical problems. Scalability is inevitably required to treat

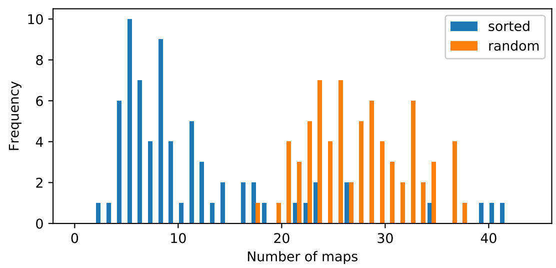

Figure 1. Number of required maps during map matching

large-scale data for practical applications; pre-process algorithms and geo-temporal analysis are both require a lot of computing resources, therefore we constructed distributed computing framework to handle them practically.

B. An example issue for processing

Map matching is a key component for analyzing driving records. It requires high computing costs[6], thus we split driving records and apply parallel processing. The map matching component requires storing road map information in the memory for efficient computing. It requires geographical index structures, therefore the footprint of the map information tend to become large.

Figure 1 shows the number of required maps (each of them occupies 10km square area) to match every 10,000 trajectories of parcel delivery trucks. If the trajectories are randomly selected, the number of required maps tend to become large (See histogram for 'random'.). We change trajectory selection methods with a heuristic approach. A driver is assigned a fixed area to deliver parcels, therefore driving routes made by one driver usually occupy similar areas. When the trajectories are sorted by the drivers and then split, the number of maps has become small (See histogram for 'sorted'.). Geographical clustering of the trajectories will help further improvement.

When merging weather information, however, requires different scheduling policy to compute; X-band radar information is collected every minute, thus all the weather data in Japan at the moment are stored in the same file; if we pursue efficiency a tiny fragment of each driving records at the moment should be collected and combined with the weather information at the moment. We consider data structures and computing strategies to compensate such inefficiency.

As we can see in Figure 2, execution time does not have a high correlation with the number of required maps. On the current map matching components, execution time is mainly affected by the number of driving records, not by how large area they drive. We can use this knowledge to balance loads in parallel processing.

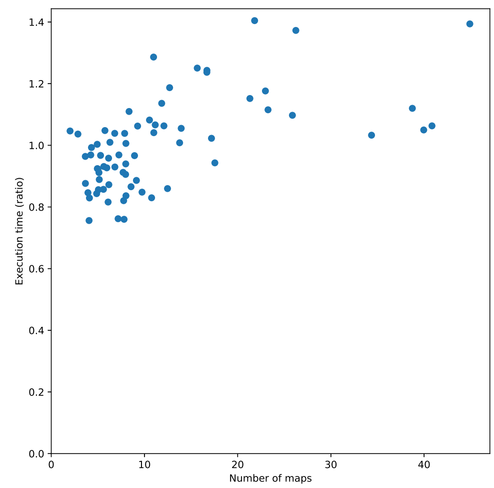

Figure 2. Relationship between execution time and the number of required maps. Execution time is normalized by the average.

III. Conclusion

In this paper, we introduce our developing system that can handle large-scale driving records and related information for analyzing driving behavior. We show an example of practical performance issue during pre-processing of driving operation records; road map information that is required during map-matching tend to become large, therefore we apply a heuristic approach to reduce the footprint of computing memory.

We plan to collect more driving related information from different industries and utilize them in many problem domains in future.

References

[1] D. Yokoyama, M. Toyoda, and M. Kitsuregawa, "Understanding drivers' safety by fusing large scale vehicle recorder dataset and heterogeneous circumstantial data," in *PAKDD 2017, part 2*, May 2017, pp. 734–746.

[2] J. Bao, R. Li, X. Yi, and Y. Zheng, "Managing massive trajectories on the cloud," in *Proceedings of the 24th ACM SIGSPATIAL*, ser. GIS '16. New York, NY, USA: ACM, 2016, pp. 41:1–41:10.

[3] "PostGIS: Spatial and geographic objects for PostgreSQL." [Online]. Available: https://postgis.net/

[4] "SpatialHadoop: A mapreduce framework for spatial data." [Online]. Available: http://spatialhadoop.cs.umn.edu/

[5] M. Itoh, D. Yokoyama, M. Toyoda, and M. Kitsuregawa, "Visual interface for exploring caution spots from vehicle recorder big data," in *2015 IEEE International Conference on Big Data (Big Data)*, Oct 2015, pp. 776–784.

[6] P. Newson and J. Krumm, "Hidden markov map matching through noise and sparseness," in *Proceedings of the 17th ACM SIGSPATIAL*, ser. GIS '09. New York, NY, USA: ACM, 2009, pp. 336–343.