Literature Review

1. Introduction

The development and popularization of digital technology over the last decade have ushered a new era for both data science and social science. Current digital data, which directly or indirectly embodies individual-level human behaviors and socio-demographic information, are being accumulated in an unprecedented speed. It thus brings many new opportunities to social science research. Among various types of data, GPS data has proven to be one of the most promising types, especially in accident and risk analysis for insurance strategy (Karapiperis et al., 2015). This is because it can provide three critical sets of always-on information pertinent to accident and risk prediction: overall human mobility pattern, individual-level moving behavior and latent socio-demographic characteristics. More importantly, these sets of information are usually not included in traditional insurance survey (Jin, Deng, & Jiang, 2018), and hence they can provide supplementary information to improve accident and risk prediction.

In this literature review, I will first review the literature that can demonstrate the possibility of extracting these kinds of information from GPS data, and then discuss the existing studies directly relative to accident prediction with GPS or telematics[1] data and their limitations. In other words, the possibility in the first part can shed light on where we can improve the topic in the second part.

2. Extracting information from GPS data in social science research

The current usage of GPS data in social science research can be generally summarized into three categories: 1) mapping human mobility patterns; 2) predicting socioeconomic or human behavioral

[1] Vehicle telematics is a method of monitoring a vehicle by combining GPS navigation system with a set of communication technology.

characteristics; 3) and extracting human behavioral features from GPS data and use them for prediction or optimization.

Mobile phone data is the most widely used data in social sciences due to the popularization of smartphones in both developing and developed countries, and research based on this kind of data is usually highly relevant to population dynamics and mobility. Lu, Bengtsson, and Holme (2012) map the population movements over three months after the 2010 earthquake in Haiti solely relying on GPS data. Deville et al. (2014) demonstrate that, in Portugal and France, not only is the population mapping solely based on GPS data as accurate as that based on census, but it can measure population dynamics almost in real time. Jiang et al. (2016) adopt a more complex method for human mobility prediction. They first extract individual features from mobile phone GPS data and then build a mechanistic modeling framework to simulate individual daily mobility with fine resolution in Boston. In addition to the area of traditional social sciences, mobile phone GPS data has also been used in public health for risk analysis. For example, three public health studies find that human aggregation or high human mobility can drive rubella transmission in Kenya (Wesolowski, Metcalf, et al., 2015), cholera transmission in Senegal (Finger et al., 2016) and dengue transmission in Pakistan (Wesolowski, Qureshi, et al., 2015).

Moreover, the combination of GPS data and mobile phone usage records or other totally disparate datasets are also used in socioeconomic and other human behavioral research. Blumenstock, Cadamuro, and On (2015) infer some socioeconomic features from mobile phone data and use them to reconstruct the distribution of wealth in Kigali. Pokhriyal and Jacques (2017) combine disparate data sources, including mobile phone data, environment data census data and poverty index, for improved poverty prediction and mapping in Senegal. It is worth noting that these studies simply implicate that GPS data embodies socioeconomic information, especially when it is combined with other existing data, rather than that we can use the predicted socioeconomic features to further predict accident risk. This is because a second-step prediction will lose or distort too much information. Therefore, at least for prediction problems, researchers should regard GPS data as latent socioeconomic information and put them into predictive models according to the social, political and economic layout of a city.

In addition to mobile phone GPS data, researchers also use the GPS data from vehicle telematics. However, most studies in this field heavily focus on the GPS data of public transportation and taxi rather than that of private cars. Therefore, the information from GPS data here reflects collective, rather than individual, socioeconomic and behavioral attributes. Besides, researchers in this field are more interested in optimization and pure prediction problems. They use human mobility patterns inferred from GPS data for traffic planning (C. Chen et al., 2013; Z. Chen, Gong, & Xie, 2017; Kong et al., 2016; Rahmani, Jenelius, & Koutsopoulos, 2015; Tang, Liu, Wang, & Wang, 2015) and taxi service strategy (Xu, Zhou, Liu, Xu, & Zhao, 2015; Zhang et al., 2014), or combine GPS trajectory data with spatial socio-economic information, based on economic layouts of cities, to predict land-use (Pan, Qi, Wu, Zhang, & Li, 2013) or calibrate retail trading model (Yue et al., 2012).

3. The use of vehicle telematics data in accident analysis

Although the studies mentioned above have demonstrated the huge potential of GPS data for prediction if various information can be extracted from it, the current accident and insurance analysis has not taken enough advantage of them.

In traditional car insurance industry, the accident prediction is based on some variables usually reported by the drivers. They typically contain annual mileage, drive-related variables such as age, gender, occupation and income, and vehicle-related variables such as vehicle age, type and price

(Jin et al., 2018). Thanks to the commercialization of the concept of Usage-Based Insurance (UBI), also known as Pay-As-You-Drive (PAYD), in the car insurance industry (Karapiperis et al., 2015), there have been some studies that try to extract behavioral features from vehicle telematics data and add them to conventional baseline models for improving prediction of accident risk. These features include the rates of hard accelerations or hard brakes (Bagdadi & Várhelyi, 2011; Handel et al., 2014; Paefgen, Staake, & Fleisch, 2014; Weidner, Transchel, & Weidner, 2017), strategic driving behaviors such as road and time selection (Tselentis, Yannis, & Vlahogianni, 2017), and daily driving behaviors and mobility patterns such as nighttime driving and familiarity with driving routes (Ayuso, Guillén, & Pérez-Marín, 2014, 2016; Jin et al., 2018)[2]. These features have also proven to be, in these studies, more useful than many self-report features in insurance survey in terms of prediction since the features in survey usually cannot reflect the real driver/drivers' behaviors and even socio-demographic information.

However, it is obvious that current studies heavily focus on driving behaviors and patterns that can be directly obtained from vehicle telematics data, but latent socio-economic information contained in GPS data has not attracted enough attention. Therefore, my project will extract more driving behaviors and mobility patterns as well as some latent socioeconomic information based on the political, economic and social layout of Beijing, which are not used in previous studies.

[2] It is worth noting that these researchers are apt to use the annual mileage values in their GPS data even though they are also contained in their insurance data. This is because the self-reported values in insurance survey are usually lower than the actual values (White, 1976).

Reference

Ayuso, M., Guillén, M., & Pérez-Marín, A. M. (2014). Time and Distance to First Accident and Driving Patterns of Young Drivers with Pay-as-you-Drive Insurance. *Accident Analysis & Prevention, 73*, 125 - 131.

Ayuso, M., Guillén, M., & Pérez-Marín, A. M. (2016). Using GPS Data to Analyse the Distance Travelled to the First Accident at Fault in Pay-as-You-Drive Insurance. *Transportation Research Part C: Emerging Technologies, 68*, 160 - 167.

Bagdadi, O., & Várhelyi, A. (2011). Jerky driving—An indicator of accident proneness? *Accident Analysis & Prevention, 43*(4), 1359 - 1363.

Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting Poverty and Wealth from Mobile Phone Metadata. *Science, 350*(6264), 1073 - 1076.

Chen, C., Zhang, D., Zhou, Z.-H., Li, N., Atmaca, T., & Li, S. (2013). *B-Planner: Night bus route planning using large-scale taxi GPS traces*. Paper presented at the IEEE International Conference on Pervasive Computing and Communications (PerCom), San Diego.

Chen, Z., Gong, X., & Xie, Z. (2017). An Analysis of Movement Patterns between Zones Using Taxi GPS Data. *Transactions in GIS, 21*(6), 1341 – 1363.

Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E., . . . Tatem, A. J. (2014). Dynamic Population Mapping Using Mobile Phone Data. *PNAS, 111*(45), 15888 - 15893.

Finger, F., Genolet, T., Mari, L., Magny, G. C. d., Mang, N. M., Rinaldo, A., & Bertuzzo, a. E. (2016). Mobile Phone Data Highlights The Role of Mass Gatherings in the Spreading of Cholera Outbreaks. *PNAS, 113*(23), 6421 – 6426.

Handel, P., Skog, I., Wahlstrom, J., Bonawiede, F., Welch, R., Ohlsson, J., & Ohlsson, M. (2014). Insurance Telematics: Opportunities and Challenges with the Smartphone Solution. *IEEE Intelligent Transportation Systems Magazine, 6*, 57 - 70.

Jiang, S., Yang, Y., Gupta, S., Veneziano, D., Athavale, S., & González, M. C. (2016). The TimeGeo Modeling Framework for Urban Mobility without Travel Surveys. *PNAS, Published online*, E5370–E5378.

Jin, W., Deng, Y., & Jiang, H. (2018). Latent Class Analysis of Accident Risks in Usage-Based Insurance: Evidence from Beijing. *Accident Analysis & Prevention, 115*, 79-88.

Karapiperis, D., Birnbaum, B., Brandenburg, A., Castagna, S., Greenberg, A., Harbage, R., & Obersteadt, A. (Eds.). (2015). *Usage-Based Insurance and Vehicle Telematics: Insurance Market and Regulatory Implications*.

Kong, X., Xu, Z., Shen, G., Wang, J., Yang, Q., & Zhang, B. (2016). Urban Traffic Congestion Estimation and Prediction Based on Floating Car Trajectory Data. *Future Generation Computer Systems, 61*, 97 - 107.

Lu, X., Bengtsson, L., & Holme, a. P. (2012). Predictability of Population Displacement After the 2010 Haiti Earthquake. *PNAS, 109*(29), 11576 – 11581.

Paefgen, J., Staake, T., & Fleisch, E. (2014). Multivariate Exposure Modeling of Accident Risk: Insights from Pay-as-You-Drive Insurance Data. *Transportation Research Part A, 61*, 27 - 40.

Pan, G., Qi, G., Wu, Z., Zhang, D., & Li, S. (2013). Land-Use Classification Using Taxi GPS Traces. *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, 14*(1), 113 - 123.

Pokhriyal, N., & Jacques, D. C. (2017). Combining Disparate Data Sources for Improved Poverty Prediction and Mapping. *PNAS, Publish online October 31, 2017*, E9783–E9792.

Rahmani, M., Jenelius, E., & Koutsopoulos, H. N. (2015). Non-Parametric Estimation of Route Travel Time Distributions from Low-Frequency Floating Car Data. *Transportation Research Part C: Emerging Technologies, 58*, 343 - 362.

Tang, J., Liu, F., Wang, Y., & Wang, H. (2015). Uncovering Urban Human Mobility from Large Scale Taxi GPS Data. *Physica A: Statistical Mechanics and its Applications, 438*, 140 - 153.

Tselentis, D. I., Yannis, G., & Vlahogianni, E. I. (2017). Innovative Motor Insurance Schemes: A Review of Current Practices and Emerging Challenges. *Accident Analysis & Prevention, 98*, 139 - 148.

Weidner, W., Transchel, F. W. G., & Weidner, R. (2017). Telematic Driving Profile Classification in Car Insurance Pricing. *Annals of Actuarial Science, 11*(2), 213 - 236.

Wesolowski, A., Metcalf, C. J. E., Nathan Eaglea, g., Janeth Kombich, Grenfell, B. T., Bjørnstad, O. N., Lessler, J., . . . Buckee, C. O. (2015). Quantifying Seasonal Population Fluxes Driving Rubella Transmission Dynamics Using Mobile Phone Data. *PNAS, 112*(35), 11114 – 11119.

Wesolowski, A., Qureshi, T., Boni, M. F., Sundsøy, P. R., Johansson, M. A., Rasheed, S. B., . . . Buckee, C. O. (2015). Impact of Human Mobility on the Emergence of Dengue Epidemics in Pakistan. *PNAS, 112*(38), 11887 – 11892.

White, S. B. (1976). On the Use of Annual Vehicle Miles of Travel Estimates from Vehicle Owners. *Accident Analysis & Prevention, 8*(4), 257-261.

Xu, X., Zhou, J., Liu, Y., Xu, Z., & Zhao, X. (2015). Taxi-RS: Taxi-Hunting Recommendation System Based on Taxi GPS Data. *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, 16*(4), 1716 - 1727.

Yue, Y., Wang, H.-d., Hu, B., Li, Q.-q., Li, Y.-g., & Yeh, A. G. O. (2012). Exploratory Calibration of a Spatial Interaction Model Using Taxi GPS Trajectories. *Computers, Environment and Urban Systems, 36*(2), 140-153.

Zhang, D., Sun, L., Li, B., Chen, C., Pan, G., Li, S., & Wu, Z. (2014). Understanding Taxi Service Strategies From Taxi GPS Traces. *IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, 16*(1), 123 - 135.