Statistics and Its Interface Volume 0 (0000) 1

Usage Based Insurance Model with Point of Interest Data

Yuxuan Zhang, Jinchang Fan†, Rui Pan and Liang Huang

With the rapid advance of the Internet of Vehicles (IOV), IOV data are becoming increasingly available. This kind of data includes plentiful information, which can reflect driving behavior and driving environment. In this work, we try to utilize IOV data for accident prediction. Specifically, we propose a Usage Based Insurance (UBI) model where the response is whether the vehicle is involved in an accident. Mileage, driving behavior as well as Point of Interest (POI) information are incorporated as predictive variables. The estimated model can be further used for driver segmentation.

Keywords and phrases: Internet of Vehicles, Point of Interest, Usage Based Insurance.

1. INTRODUCTION

In the past decade, the whole world witnessed the tremendous development of vehicle industry. According to the statistics published by the International Organization of Motor Vehicle Manufacturers (*www.oica.net*), 94,977 thousands of motor vehicles are produced in 2016 all over the world. In addition, it is projected that there will be around 1.1 billion passenger vehicles by 2030 [1]. The flourish of vehicle industry brings a brand new field, the Internet of Vehicles (IOV). Vehicles equipped with specific communication instruments are incorporated in the IOV. As a result, IOV gathers data from vehicles and makes sure the fulfillment of information communication, environmental protection, energy conservation and safety [2].

In-vehicle data recorders (IVDR) are one of the general resources for IOV data. Thanks to IVDR and various sensors, ample data could be collected from vehicles. These data include but are not limited to, vehicle location, velocity, acceleration, mileage, and engine parameters such as fuel consumption, engine rotation speed, and many others. In the rest of this article, we refer to the data collected by IVDR as IOV data. It is noteworthy that IOV data are typically recorded every second and contain plentiful information in great detail. This leads to large scale of the data, which might be thorny for analysis. For example, the data adopted in this work contain multifarious information of 1,453 cars, accumulating records approximate 14 million kilometers.

Statistical analysis of IOV data could be widely used in different fields, contributing to transportation department, drivers, and insurance business. For instance, collision avoidance technologies could serve to reduce the chance of accident by collision warning system and driver assistant system [3, 4]. In addition, drivers could attain better driving experience by finding local services more conveniently [3, 5]. Another important application of IOV data lies in actuarial decision-making. Insurance companies could profit by offering more reasonable insurance products for policyholders. One of the most prevalent insurance products is Usage Based Insurance (UBI).

UBI is an innovative insurance product, where the premiums are based directly on how much the vehicle is driven during the policy term [6]. This kind of insurance is economical, because drivers are encouraged to mind their behavior in order to save money. In addition, UBI attracts growing interest, since it achieves additional social benefits including traffic safety and environmental objectives [7, 8]. Some insurance companies now offer UBI options, e.g., Progressive (*https://www.progressive.com*), Insure The Box (*https://www.insurethebox.com*), and many others. Nevertheless, UBI could not be implemented without IOV data. IOV data provide information on usage and driving behavior, which are the basis of UBI pricing. By analyzing IOV data, the accident prediction model is established and the probability of accident could be calculated [9, 10, 11].

Generally, there are two types of work on UBI study. The first type only considers mileage as the explanatory variable. Literature includes analyzing the relationship between mileage and accident occurrence [12], mileage and crash fatalities [6] as well as mileage and insurance claims [13, 14]. Later, nonlinear relationship particularly for low-mileage drivers is further studied [15, 16]. Recently, mileage is proved to be a significant predictor of accident cost [17]. Mileage and other classic rating factors together will improve explanatory power significantly. The second type of research takes vehicle velocity into consideration [18, 19, 20]. It is demonstrated that driving performances concerning

*Yuxuan Zhang, Jinchang Fan and Rui Pan are from School of Statistics and Mathematics, Central University of Finance and Economics, Beijing, 100081, P. R. China. Liang Huang is from CIHON (Beijing) New Technology Co., Ltd., Beijing, 110000, P. R. China. The research is supported in part by National Natural Science Foundation of China (NSFC, 11601539, 11631003).
†Corresponding author.

braking and acceleration are significant in distinguishing drivers involved in crashes [19]. Various statistical models are adopted in these studies, including Poisson log-linear model [21], exponential-type model [11], and many others [9, 22].

Although plentiful works concerning UBI have been conducted, several shortcomings still exist. First of all, the sources of data are limited. Two main sources are Virginia Tech Transportation Institute and US Strategic Highway Research Program. As for dataset provided by Virginia Tech Transportation Institute, it merely contains records of 100 vehicles [23]. Second, only GPS observations are available in most of these data, which may cause inaccuracy in calculating velocity and acceleration. On contrary, data recorded by sensors from IVDR are relatively accurate and reliable. Third, predictors used in those works are lack of diversity. Only vehicle information (e.g., mileage and velocity) and driving behavior (e.g., proportion of drowsy driving) are taken into consideration. Nonetheless, important information derived from location data and trajectory data is neglected.

In our work, improvements are achieved in all the three aspects. Dataset used in this work is offered by a famous vehicle manufacturing group. Detailed and diverse information regarding vehicles is included, providing much evidence of driving behavior. In addition, IVDR records vehicle velocity and acceleration through specific sensors so that the accuracy of data is warranted. At last, Point of Interest (POI) data are considered in this work. It will be discussed below that POI data contain plentiful information concerning driving behavior and environment.

POI is a specific point location which might be useful or interesting to someone. It includes spots such as sightseeing sites, gas stations, restaurants and many others. POI data could be uploaded by users of social networking and online advertising applications like Facebook, Weibo, and Yelp. A piece of POI data normally consists of name, latitude, and longitude of a location. Tags, categories, and other information are also frequently attached to POI data [24]. Furthermore, POI data have a wide range of applications. For example, POI recommendation could be utilized to suggest living facilities, bus stations, restaurants, and etc. [25, 26]. Another application of POI data is pattern identification of population flow [27, 28]. Human daily patterns could be analyzed and discovered through POI data. As for UBI, POI data provide us with a new perspective on predicting accidents since they could reflect driving environment. In fact, several existing pricing strategies including GPS-based pricing have already taken environment into consideration [29]. For instance, it is conceivable that accident is less likely to happen during a trip to a national park, since roads toward national parks are typically uncrowded.

The rest of this article is organized as follows. In Section 2, we introduce our IOV and POI data, together with brief descriptive analysis. In Section 3, logistic regression is applied to build the UBI model. Estimation and prediction results are explained in detail. Some business applications are further discussed in Section 4. Some concluding remarks are given in Section 5.

2. DATA DESCRIPTION

2.1 The IOV Data

In this paper, data are provided by a major Chinese automobile manufacturer, recorded by IVDR. Data are collected from more than 1,400 vehicles, ranging from September 2014 to September 2015. The data cover more than 14 million kilometers in total. Insurance records are also included in the dataset. It is a binary variable indicating whether the vehicle has reported accident (for 1) or not (for 0). There are 49.6% drivers reported accident in our data. Vehicles are identified by Vehicle Identification Number (VIN) and all private information is protected.

IVDR collects and updates records every second and consists of the following information:

- VIN: unique identification number for each vehicle. It is the only identification of a vehicle, which includes no private information, e.g., LLNC1AAA2EA00****.
- Time of record: time stamp of data recorded by IVDR, e.g., 2015-02-12 16:32:48. The interval between adjacent record is one second.
- Latitude & Longitude: instantaneous position of a vehicle. The accuracy of position is around ten meters. The latitude and longitude provide us with the position of a vehicle and POIs near it could be scanned.
- Mileage: accumulated mileage of vehicle till the time point we observe, ranging from 629 to 26,690 kilometers. The mileage is generally adopted as a crucial predictor in UBI. The larger the mileage, the more likely a vehicle is involved in an accident.
- Fuel consumption: instantaneous fuel consumption, measured by liters per 100 kilometers ($L/100\,km$). Higher instantaneous fuel consumption might be related to acceleration or other behavior, thus reflecting driving behavior.
- Velocity: instantaneous velocity of a vehicle, ranging from 0 to 220 kilometers per hour (kph). It is automatically calculated from detected revolution (rev) of wheels.
- Acceleration: longitudinal acceleration, detected by gyroscope that could diminish the influence of tilt when vehicle is on slope. A positive record indicates the vehicle is accelerating and a negative record indicates the vehicle is decelerating. The unit of record is meter per second squared (m/s^2).
- Engine rev: instantaneous engine rotational speed, ranging from 0 to 4,000 revolutions per minute (rpm). It indicates the instantaneous power of engine. The engine rev is related to vehicle velocity. However, considering

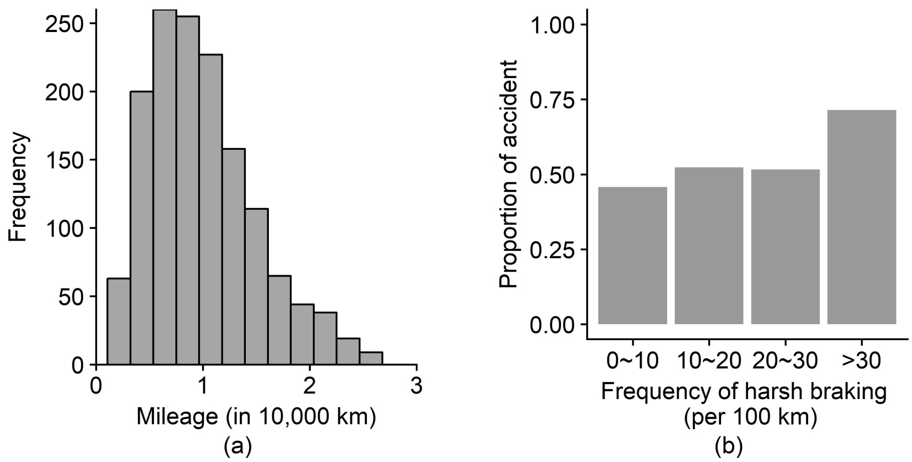

Figure 1. Figure (a): histogram of mileage. A right skewed shape can be detected, indicating the existence of extreme large values in mileage. Figure (b): proportion of accident in different groups of harsh braking frequency. Accidents happen more for those who make harsh breaking more 30 times per one hundred kilometers.

that it takes the size of gear as a multiplier, the relationship is more complicated than linear relationship.

2.2 Descriptive Analysis

We next conduct descriptive analysis on the IOV data. Figure 1 (a) shows the histogram of mileage of all vehicles. The mileage ranges from 629 to 26,690 kilometers, with an average of 9,978 kilometers. It is obvious that mileage is right skewed, which indicates that a few vehicles run extremely long distance. Figure 1 (b) divides drivers into four groups according to their harsh braking habits. The height of the bar indicates the proportion of accidents within each group. As it is shown, frequency of harsh braking could serve as a predictor for accident prediction.

In the rest of this work, we define a trip as a consecutive driving behavior from a start to a destination. Figure 2 (a) represents the path of a trip, according to latitude and longitude recorded by IVDR. Note that the exact records of position are much denser than points showed in the figure, but simplified for clearer distinguish. Accumulative mileage, velocity, and acceleration are also presented in Figure 2. Data recorded every second could bring more insight for us. For example, abnormal driving behavior like harsh braking and rapid acceleration can be accurately detected.

2.3 POI Data

POI data adopted in this work are from Sina Weibo's open API (*http://open.weibo.com*), where Sina Weibo is one of the largest social media platforms in Chinese. Since a POI always contains latitude and longitude, it indicates a point on the map, which is illustrated in Figure 3. For each trip of a vehicle, the position of the destination is derived. Next, all POIs within 500 meters of the destination are scanned and the number of POIs is counted within each POI category. Results for all trips of one particular vehicle are averaged to attain POI data of this vehicle. Thus, the frequencies of each POIs for each vehicle are derived. For simplification, we classify POIs into 117 different labels. According to the POI data, the mean frequency of all POIs in all vehicles is 0.320, which indicates that each POI appears 0.320 time on average in each trip. The most welcomed POI is Residential area.Restaurant with mean frequency of 3.422. On the contrary, the least welcomed POI is Pet supply store.Building gate with mean frequency of 0.001. We will discuss how POI information is further made use of in the next section.

3. USAGE BASED INSURANCE MODEL

The major concern of this work is to discern drivers with different probability of accident involvement. Logistic regression model is adopted to score each driver based on their driving behavior and POI data. Let $Y_i \in \{0, 1\}$ represents the response, where $1 \leq i \leq n$, and n is the total sample size. If $Y_i = 0$, the ith vehicle did not report any accident or insurance claim during the observed period, otherwise $Y_i = 1$. Let $X_i = (x_{i1}, \ldots, x_{ip})^\top \in \mathbb{R}^p$ be the covariate vector. The probability of binary response variable Y_i to be one is modeled as

$$Y_i \sim Bernoulli(p(X_i^\top \beta)), \tag{1}$$

where $\beta = (\beta_1, \cdots, \beta_p)^\top \in \mathbb{R}^p$ is the p-dimensional regression coefficient, and $p(t) = e^t/(1 + e^t)$ is logistic link function, which keeps the probability lying between 0 to 1. In addition, maximum likelihood estimation is conducted, which is denoted as $\hat{\beta} \in \mathbb{R}^p$

In most existing literature, only IOV data are incorporated in the model. Our model incorporates both IOV data

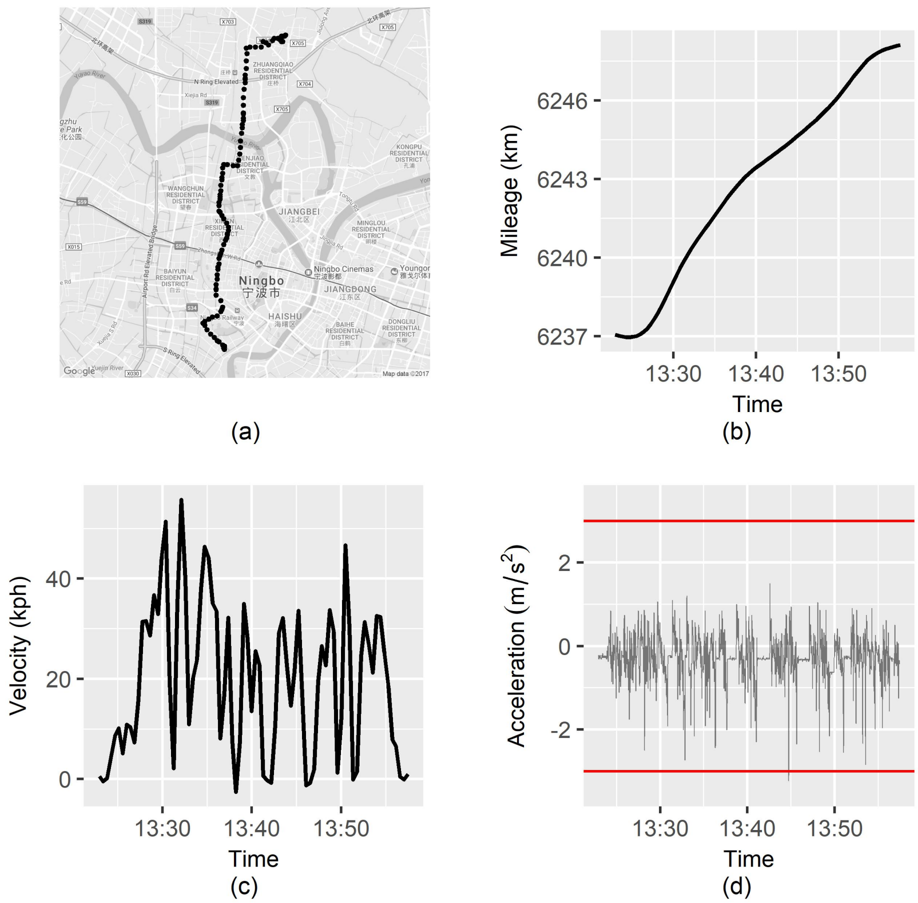

Figure 2. A sample trip. Figure (a) shows the path of the trip. It started from Jiangbei District and went south to Haishu District, experiencing 11 kilometers and around 35 minutes. Figure (b) and (c) show the time series of mileage and velocity of the vehicle during this trip. Low points in (c) indicate several periods of waiting at intersections. Figure (d) shows the time series of acceleration, and two red horizontal lines indicate 3 *and* $-3\,m/s^2$*, which are considered as benchmarks of rapid acceleration and harsh braking.*

and POI data in order to heighten the prediction accuracy. To obtain a more comprehensive understanding, stepwise approach and Akaike information criterion (AIC) [30] are adopted, since not all predictors are significant.

3.1 The Construction of Predictors

Based on the raw data, 10 IOV predictors are derived, as well as 117 POI predictors. Those predictors could be divided into four categories: usage, stability, driving period and POI. Accumulative mileage measures the exposure feature of the vehicle [31]. Predictors belong to usage category except for accumulative mileage are calculated as arithmetic mean of observed data. Predictors in driving period category measure the number of trips occurred during certain time periods. Frequency of rapid acceleration and harsh braking are times of instant longitudinal acceleration over $3\,m/s^2$ or less than $-3\,m/s^2$. It is then divided by the mileage and multiplied by 100, for the frequency of harsh braking instead of count is more suitable to indicate driving habit. IOV predictors are standardized in model estimation. The original data are shown in Table 1.

3.2 Estimation Results

IOV data and POI data are adopted to build the logistic regression model to score drivers. Bidirectional stepwise method with AIC was incorporated to conduct the model selection. In each step, the effect of eliminating or adding each predictor is evaluated by AIC. The selection will stop if adding or eliminating a predictor will cause larger AIC. After the model selection, $\hat{\beta}$ is derived from the model by maximum likelihood estimation method. Each driver was scored and the prediction result is assessed by Area Under Curve (AUC) which is widely adopted as an important indicator of prediction accuracy. Four of ten predictors besides POI predictors are included in the model, including accumulative mileage, average fuel consumption, average engine

Table 1. Predictors involved in different categories (i.e., usage, stability, driving period and POI)

Category	Variable Name	Unit	Range	Detail
Usage	Accumulative mileage	km	629-26690	
	Average velocity	kph	13.30-40.57	
	Average fuel consumption	L/100 km	4.67-8.08	
	Average engine rev	rpm	1019.00-1459.91	
Stability	S.D. of steer position		52.49-123.37	
	Frequency of rapid acceleration		0-2.65	Times per 100 kilometers
	Frequency of harsh braking		1.45-35.90	Times per 100 kilometers
Driving Period	Morning		7-294	From 7:00 to 10:00
	Evening		2-329	From 17:00 to 19:00
	Night		0-53	From 0:00 to 4:00
POI	Residential area, Bus station, Entertainment, Hospital, etc.			

Table 2. Coefficient estimation of driving behavior predictors, the standard error and p-value are also reported

Predictor	Coefficient	Standard Error	p-value
Intercept	0.062	0.093	0.507
Accumulative mileage	0.141	0.060	0.018
Average fuel consumption	-0.162	0.058	0.005
Average engine rev	-0.190	0.064	0.003
Frequency of harsh braking	0.230	0.057	<0.001

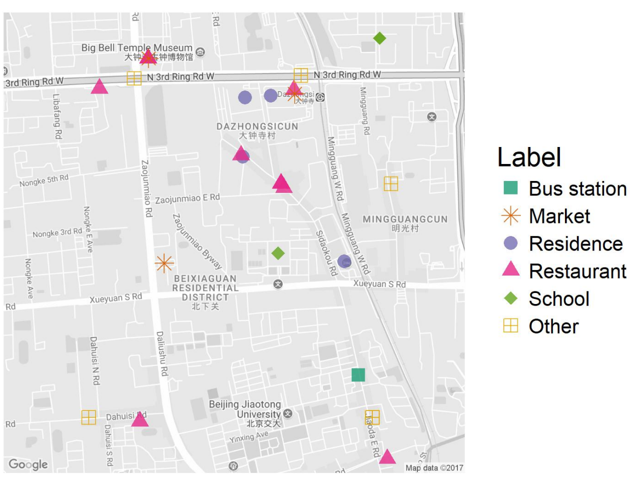

Figure 3. An instance of map with POI. Different shapes and colors indicate different categories of POI, including bus station, restaurant, school, etc..

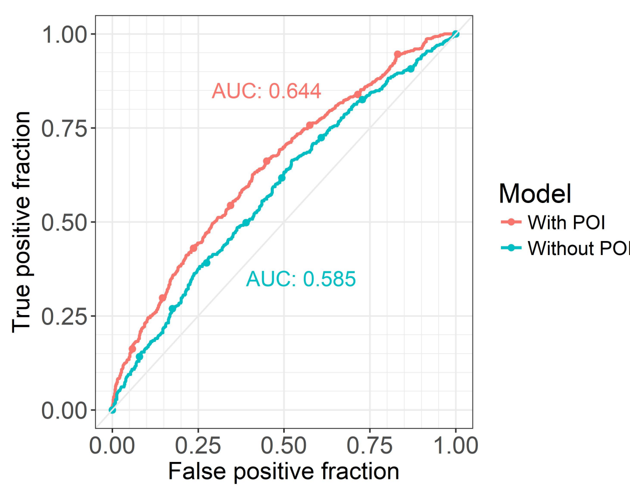

Figure 4. Receiver operating characteristic curve of UBI model. The AUC of the UBI model is 0.644, while that of the model without POI data drops to 0.585.

rev and frequency of harsh braking. They are represented in Table 2. The coefficients of accumulative mileage and frequency of harsh braking are positive. While the coefficients of average fuel consumption and average engine rev represent that higher fuel consumption and engine rev are related to less possibility of accident. The sign of estimated coefficient of accumulative mileage is in accord with most previous literature. Other predictors are also easy to explain. Much harsh braking is related to less stability, and lower engine rev is a sign of low velocity and congestion, which might bring about more car accidents.

As for POI predictors, 22 POI predictors are incorporated in the UBI model. It is shown in the coefficients (Figure 5) that some POIs including Youth hotel and Museum.Training contribute to the occurrence of accident. One likelihood is that roads around these places are commonly crowded especially on weekends so that complex traffic condition would bring about more accidents. Other POIs like III A hospital, Special hospital, and Mosque.Tourism consultation are related to lower likelihood of being in accidents. It is conceivable that these POIs usually are less crowded like mosque or directed by more traffic wardens like hospital, which indicates fewer accidents.

The AUC of the model is 0.644, while AUC of the model without POI data is 0.585 (Figure 4). Obviously, POI data could be served as additional and informative predictors to improve the prediction result. As it is stated above, POI data reflect driving habits of drivers and take the environment factor into consideration. Different POIs imply various preference that might be related to the driving habits. In addition, different POIs represent different traffic conditions and risk of being involved in accidents. Thus, POI predictors present a new perspective on predicting accidents.

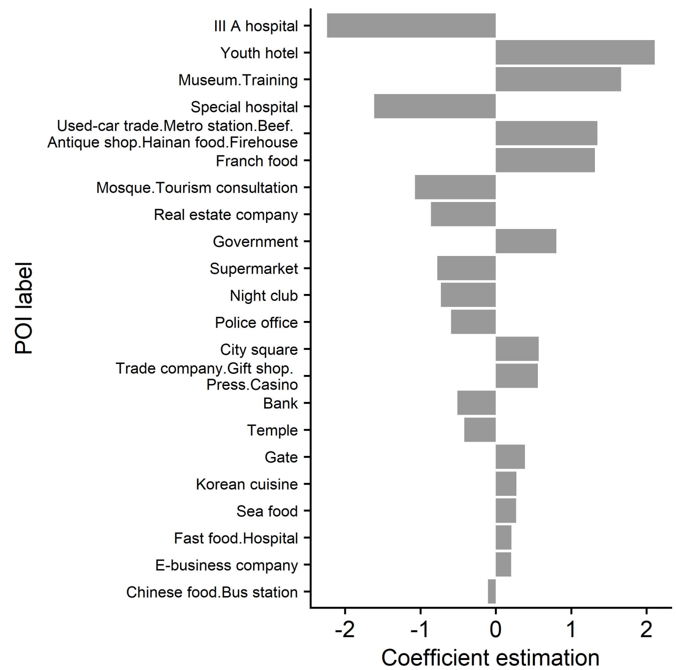

Figure 5. Coefficient estimation of POI predictors in UBI model. III A hospital is the POI predictor with the largest absolute value of coefficient.

4. BUSINESS APPLICATION

4.1 Driver Segmentation

With more convenient and advancing facilitates to acquire IOV data, more and more insurance companies now offer UBI. However, POI data that are informative and easy to acquire have not been fully used of. The key of UBI is discerning drivers with different risk, and the accuracy of it could be improved by POI data. According to the model above with IOV and POI data, each driver could be scored based on $X_i^\top \hat{\beta}$ which is the predicted probability of getting involved in car accident. Furthermore, driver segmentation could be achieved according to the scores, and insurance companies could offer different premiums for drivers in different classes. To illustrate, we divide drivers in the dataset into four different risk classes as demonstrated in Figure 6 and Figure 7. The accident risk of the first and the fourth classes is obviously different from those in other classes, so special premium strategy could be adopted to them in order to profit more in UBI market. For example, insurance companies could charge drivers in high-risk class higher premiums, while they can charge low-risk drivers less premiums to attract more customers. That strategy on average would reduce the premiums and profit more. Since market of vehicle insurance is a fully developed market, this pricing advantage, despite tiny, would make great contribution.

4.2 Pricing Strategy

There are mainly three UBI pricing options now, including mileage rate factor (MRF), per-mile premiums and GPS-based pricing. As for MRF, annual mileage served as a key rating factor in pricing. Drivers with low annual mileage will be provided with discount, and vice versa. GPS-based pricing uses IVDR to identify position of cars, and the premiums are based on when and where drivers drive. MRF only take mileage into consideration and many MRF systems rely on self-reported estimates of future mileage, which might lead to less accuracy in pricing [32].

Driving behavior rate factor (DBRF) which includes time, speed and other driving behavior serves as an expanded version of MRF, but the accuracy of DBRF can also be further improved [33]. By incorporating POI predictor into DBRF, more information about environment and driving habits are considered as it is shown in sections above. Based on driver segmentation result, linkage model could decide the rate adjustment coefficients based on scores of drivers [33]. Furthermore, the scores of drivers could be used for UBI pricing directly, since it not only reflects relative likelihood of accidents but also gives the probability of getting involved in accident for each driver.

5. CONCLUDING REMARKS

In this paper, we build a UBI model with IOV data and POI data. The data are collected by a major Chinese automobile manufacturer. Logistic regression model together with AIC is used to conduct the model selection. It is found that accumulative mileage, average fuel consumption, average engine rev, frequency of harsh braking and other 22 POI predictors are incorporated in the final model, which lead to an AUC of 0.644. In addition, the business application of the model in driver segmentation and UBI pricing is also discussed. To conclude this work, we have the following

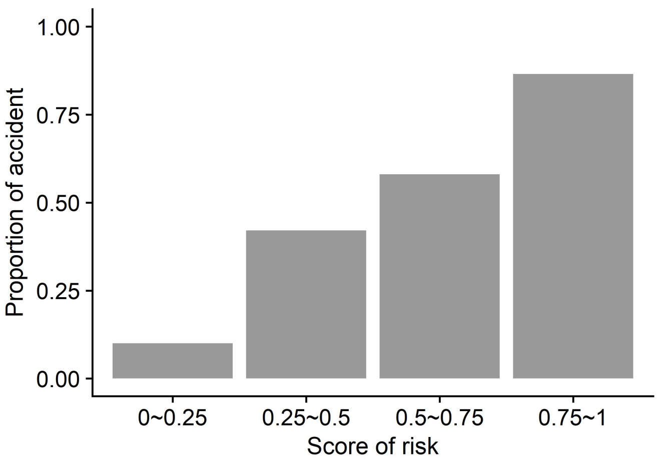

Figure 6. Proportion of accident in different groups according to scores of drivers. The predicted risk ranges are approximately in accord with actual proportion of accident.

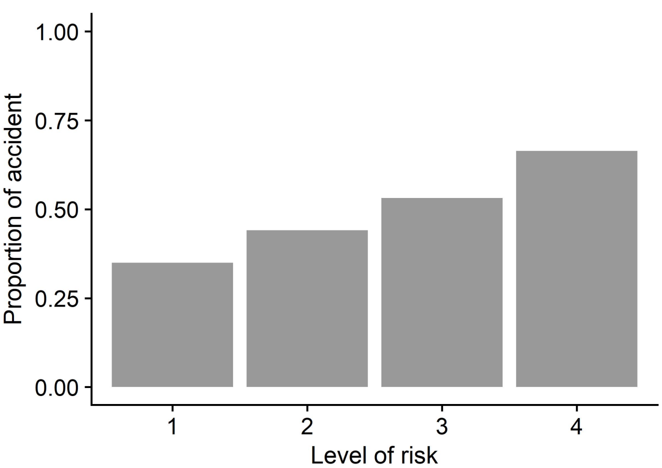

Figure 7. Different segmentation compared to the previous one. As the previous segmentation uses linear scores (actually, 0, 0.25, 0.75, and 1), this figure demonstrates the groups divided by quantiles to make size of each group similar.

discussions about future research topics. First, the POI data are provided by Sina Weibo's open API, which are generated by its users. To further improve prediction accuracy, it is feasible to adopt a larger and more accurate POI dataset. Second, the UBI model takes insurance claim as response variable which is binary. Nonetheless, in practice, different accidents correspond to different amounts of compensation. So the amount of compensation and the type of accidents can be taken into consideration. Third, a threshold of 500 meters for scanning nearby POIs is set in our model for simplicity. As an alternative, it is more flexible to set the distance of scanning as a tuning parameter.

REFERENCES

[1] Henry Lee and Lovellette Grant. Will electric cars transform the u.s. market. *The American Economic Review*, 2011.

[2] N Liu. Internet of vehicles: Your next connection. *Huawei Win-Win*, 11:23–28, 2011.

[3] Yang Fangchun, Wang Shangguang, Li Jinglin, Liu Zhihan, and Sun Qibo. An overview of internet of vehicles. *China Communications*, 11(10):1–15, 2014.

[4] Shane B McLaughlin, Jonathan M Hankey, and Thomas A Dingus. A method for evaluating collision avoidance systems using naturalistic driving data. *Accident Analysis & Prevention*, 40(1):8–16, 2008.

[5] Georgios Karagiannis, Onur Altintas, Eylem Ekici, Geert Heijenk, Boangoat Jarupan, Kenneth Lin, and Timothy Weil. Vehicular networking: A survey and tutorial on requirements, architectures, challenges, standards and solutions. *IEEE communications surveys & tutorials*, 13(4):584–616, 2011.

[6] T Litman. Distance-based vehicle insurance feasibility, costs and benefits. victoria transport policy institute, 2011.

[7] JW Bolderdijk, J Knockaert, EM Steg, and ET Verhoef. Effects of pay-as-you-drive vehicle insurance on young drivers' speed choice: Results of a dutch field experiment. *Accident Analysis & Prevention*, 43(3):1181–1186, 2011.

[8] Ian WH Parry. Is pay-as-you-drive insurance a better way to reduce gasoline than gasoline taxes? *American Economic Review*, pages 288–293, 2005.

[9] Johannes Paefgen, Thorsten Staake, and Elgar Fleisch. Multivariate exposure modeling of accident risk: Insights from pay-as-you-drive insurance data. *Transportation Research Part A: Policy and Practice*, 61:27–40, 2014.

[10] Kun-Feng Wu and Paul P Jovanis. Crashes and crash-surrogate events: Exploratory modeling with naturalistic driving data. *Accident Analysis & Prevention*, 45:507–516, 2012.

[11] Mercedes Ayuso, Montserrat Guillén, and Ana María Pérez-Marín. Analyzing the effect of driving patterns on the risk of accident in young drivers with pay-as-you-drive insurance. *Accident Analysis & Prevention*, 73:125–131, 2014.

[12] Peter F Lourens, Jan AMM Vissers, and Maaike Jessurun. Annual mileage, driving violations, and accident involvement in relation to drivers' sex, age, and level of education. *Accident Analysis & Prevention*, 31(5):593–597, 1999.

[13] Jason Bordoff and Pascal Noel. Pay-as-you-drive auto insurance: A simple way to reduce driving-related harms and increase equity. *Hamilton Project Discussion Paper*, 2008.

[14] J Ferreira and E Minike. Pay-as-you-drive auto insurance in massachusetts: A risk assessment and report on consumer, industry and environmental benefits. department of urban studies and planning, massachusetts institute of technology. *Massachusetts Institute of Technology (http://dusp. mit. edu/) for the Conservation Law Foundation, http://www. clf. org/, http://www. clf. org/our-work/healthy-communities/modernizing-transportation/pay-as-you-drive-auto-insurance-payd*, 2010.

[15] Loren Staplin, Kenneth W Gish, and John Joyce. 'low mileage bias' and related policy implications—a cautionary note. *Accident Analysis & Prevention*, 40(3):1249–1252, 2008.

[16] Jim Langford, Sjaanie Koppel, Dennis McCarthy, and Sivaramakrishnan Srinivasan. In defence of the 'low-mileage bias'. *Accident Analysis & Prevention*, 40(6):1996–1999, 2008.

[17] Joseph Ferreira Jr and Eric Minikel. Measuring per mile risk for pay-as-you-drive automobile insurance. *Transportation Research Record: Journal of the Transportation Research Board*, (2297):97–103, 2012.

[18] Craig Norman Kloeden, Jack McLean, and Garique Francis Vladimir Glonek. *Reanalysis of travelling speed and the risk of crash involvement in Adelaide South Australia*. Australian Transport Safety Bureau, 2002.

[19] Jungwook Jun, Jennifer Ogle, and Randall Guensler. Relation-

ships between crash involvement and temporal-spatial driving behavior activity patterns: use of data for vehicles with global positioning systems. *Transportation Research Record: Journal of the Transportation Research Board*, (2019):246–255, 2007.

[20] Tom Brijs, Geert Wets, Robin Krimpenfort, and Col Offermans. Impact of hourly measured speed on accident risk in the netherlands: results from exploratory study using geographic information systems. *Transportation Research Record: Journal of the Transportation Research Board*, (1972):85–93, 2006.

[21] Tim Gordon, Lidia Kostyniuk, Paul Green, Michelle Barnes, Daniel Blower, Adam Blankespoor, and Scott Bogard. Analysis of crash rates and surrogate events: unified approach. *Transportation Research Record: Journal of the Transportation Research Board*, (2237):1–9, 2011.

[22] Johannes Paefgen, Thorsten Staake, and Frédéric Thiesse. Evaluation and aggregation of pay-as-you-drive insurance rate factors: A classification analysis approach. *Decision Support Systems*, 56:192–201, 2013.

[23] Thomas A Dingus, Sheila G Klauer, Vicki L Neale, A Petersen, Suzanne E Lee, JD Sudweeks, MA Perez, J Hankey, DJ Ramsey, S Gupta, et al. The 100-car naturalistic driving study, phase ii-results of the 100-car field experiment. Technical report, 2006.

[24] Yu Chen, Rui Pan, and Hansheng Wang. Analyzing beijing point of interest data using grop linked cox process. Working paper, 2017.

[25] Quan Yuan, Gao Cong, Zongyang Ma, Aixin Sun, and Nadia Magnenat Thalmann. Time-aware point-of-interest recommendation. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 363–372. ACM, 2013.

[26] Bin Liu, Yanjie Fu, Zijun Yao, and Hui Xiong. Learning geographical preferences for point-of-interest recommendation. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1043–1051. ACM, 2013.

[27] Marco Veloso, Santi Phithakkitnukoon, and Carlos Bento. Urban mobility study using taxi traces. In *Proceedings of the 2011 international workshop on Trajectory data mining and analysis*, pages 23–30. ACM, 2011.

[28] Anastasios Noulas, Salvatore Scellato, Cecilia Mascolo, and Massimiliano Pontil. An empirical study of geographic user activity patterns in foursquare. *ICwSM*, 11:70–573, 2011.

[29] Todd Litman. Distance-based vehicle insurance feasibility, costs and benefits. *Victoria*, 11, 2007.

[30] Hirotugu Akaike. Factor analysis and aic. *Psychometrika*, 52(3):317–332, 1987.

[31] Arthur C Wolfe. The concept of exposure to the risk of a road traffic accident and an overview of exposure data collection methods. *Accident Analysis & Prevention*, 14(5):337–340, 1982.

[32] Todd Litman. Pay-as-you-drive insurance: recommendations for implementation. 2008.

[33] Zhishuo Liu, Qianhui Shen, Han Li, and Jingmiao Ma. A risky driving behavior scoring model for the personalized automobile insurance pricing. In *Proceedings of the 2nd International Conference on Crowd Science and Engineering*, pages 61–67. ACM, 2017.