ATLANTIS PRESS

Advances in Economics, Business and Management Research, volume 91

1st International Symposium on Economic Development and Management Innovation (EDMI 2019)

Study of Piecewise UBI Pricing Strategy based on the Risk Probability

Bo Huang [a], Junyuan Bian [b], Nanjie Liu [c]

College of Telecommunications and Information Engineering Nanjing University of Posts and Telecommunications Nanjing, China

[a]huangbo@njupt.edu.cn, [b] 1017010326@njupt.edu.cn, [c]songm@cpsdna.com

Abstract. With the rapid growth of car ownership in China, vehicle insurance has become an important branch of the insurance industry. The traditional and single vehicle insurance premium pricing strategy is not conducive to the long-term development of the vehicle insurance industry. So, the Usage-Based Insurance (UBI) based on driving behavior analysis is put on the agenda. Based on the data collected by the real vehicle in the UBI pilot city where Dina Technology and an insurance company cooperate, this paper proposes a segmented UBI pricing strategy based on the distribution of risk probability. First, the user's driving behavior data are collected in real time through Dina Technology's intelligent vehicle On-Board Diagnostics (OBD) terminal. Then, the linear logistic regression machine learning algorithm is used to analyze the risk probability of the measured data, and determine the risk factor coefficients of each driving behavior. Finally, on the basis of the distribution function of the risk probability, the segmentation pricing is differentiated by setting the segmentation penalty factor. By supervising user's driving behavior to reward or punish, this strategy can improve the user's driving habits, improve the user's driving safety, and reduce the risk probability.

Keywords: vehicle insurance, driving behavior analysis, the risk probability, probability distribution.

1. Introduction

China's automobile industry has developed rapidly in recent years. At the end of 2017, the number of vehicle ownership reached 209 million, which increased by 23.32 million compared with 2016. The surge in the number of cars has promoted the development of the vehicle insurance industry. However, China's traditional vehicle insurance model adopts vehicle-based premium pricing, only based on vehicle purchase price, vehicle type, displacement, car purchase period, historical accident situation and other information. This pricing method is extremely single and flawed, without taking into account the impact of user driving behavior on traffic accidents [1]. Only about 10% of car owners have frequent traffic accidents, and the remaining 90% safe drivers bear high insurance costs for the traffic accidents of the 10% dangerous drivers, which damages the interests of safe drivers.

With the development of the Internet of Vehicles and Big Data, the human-centered UBI vehicle insurance pricing model is gradually becoming possible. Since 2009, the United States and Europe have begun to introduce and promote UBI vehicle insurance. In the United States, Metromile has adopted a premium pricing vehicle insurance model based on user driving mileage, which reduces insurance premiums for users with less driving mileage. State Farm, the largest vehicle insurance company in the United States, entered the UBI vehicle insurance market in 2011 and cooperated with Hughes Telematics to evaluate vehicle driving conditions based on driving mileage and part of driving behavior data, adjust premium levels, and differentiate pricing for different users. In this way, users can be offered up to 50% discount. In 2014, Allstate launched the UBI project Drivewise, which helps identify and teach safe driving behavior, collecting driver speed, time, emergency braking events and locations for discount pricing [2]. In 2016, the Chinese enterprise Changan Auchan, China Life and the Pingjia Technology carried out the cooperation of UBI vehicle insurance, to achieve the differential pricing of vehicle insurance premiums.

Traffic accidents are largely caused by incorrect driving behavior of the user. Through the intelligent vehicle terminal equipment of Jiangsu Dina Digital Technology Co., Ltd. to monitor the driving status of the car, we can collect the driver's driving behavior, such as speed, position, mileage, driving time and so on [3]. The collected driving behavior data are transmitted to the cloud through the Internet of Vehicles technology and are combined with the historical accident situation and the driver's identity information such as age, driving age, gender, occupation, etc. Under the analysis of big data and machine learning, we can know the impact of different driving behaviors, predict the risk probability, and finally differentiate the segmentation pricing of vehicle insurance premium according to the risk probability.

The adoption of a new vehicle insurance model based on driving behavior data can improve the accuracy of vehicle insurance pricing, differentiate the pricing of different users, protect the interests of safe driving users, and punish dangerous driving users, which can improve the rationality of vehicle insurance pricing and the satisfaction of vehicle insurance policyholders. At the same time, through reward or punishment to improve the user's driving behavior, it can reduce traffic accidents, reduce losses, and protect the interests of insurance companies and car owners [4].

2. UBI System Structure

The UBI pricing system adopted in this paper consists of three parts: intelligent vehicle terminal, vehicle communication access network and information processing and analysis cloud platform, which is the "carrier-cloud-client" architecture [5].

Client: The intelligent vehicle terminal equipment installed in the car can detect the running status of the car in real time to collect the driving behavior information of the user.

Carrier: It is the vehicle communication access network, which realizes the network access function of the vehicle network and transmits the collected data to the cloud. Through 4G or WiFi, the car can access the Internet anytime and anywhere, and send the collected driving behavior data to the network in real time for the prediction of the risk probability.

Cloud: It is the information processing and analysis cloud platform, which realizes the processing and analysis of data. It predicts the risk probability of users based on driving behavior data, and on this basis, differentiates the pricing of premiums and provides rich vehicle network services.

The "client-carrier-cloud" system architecture makes it possible to use the UBI pricing model based on user driving behavior.

3. Intelligent Vehicle Terminal

The vehicle insurance model based on user's driving behavior needs to obtain the driving state of the vehicle in real time. Therefore, the vehicle driving data needs to be collected and transmitted to the cloud for processing and analysis. This paper collected the driving behavior data of 3,070 users using the intelligent vehicle terminal equipment of Jiangsu Dina Digital Technology Co., Ltd.

The intelligent vehicle terminal acquires data by reading information such as the status of the vehicle collected by the OBD. OBD is a system for detecting the operating parameters of various systems of the car and reading data. It can detect the driving status of the car such as speed, acceleration, position and mileage in real time, which provides a hardware foundation for UBI vehicle insurance [6]. At the same time, OBD monitors the hardware of the car engine through sensors, so that the owner can more easily know the condition of the car, which is convenient for car maintenance.

4. UBI under the Big Data and Internet of Vehicles

Through the Never-Ending Machine Learning (NEML), the vehicle driving data collected by the intelligent vehicle terminal and the historical accident situation are processed, modeled and analyzed. The sample data are trained to determine the corresponding coefficient value of the risk factor of the user behavior, and the risk probability can be predicted. By segmenting users with different risk probabilities, segmented rewards or penalties are applied to users at different segments to carry out corresponding vehicle insurance pricing, and rewards are used to improve driving behavior habits. In addition, the new data generated later will be added to the sample data, continuous training will be carried out to correct the characteristic coefficients of risk factors, and the accuracy of risk probability judgment will be continuously improved to improve the precision of pricing.

4.1 Data Sources

Traditional premium pricing models include: pricing based on vehicle type, pricing based on premiums, and pricing based on mileage. These traditional pricing models cannot truly reflect the risk probability of vehicle, so there are great drawbacks in traditional models. The premium pricing model based on user behavior can predict the risk probability of new users based on data collected from “people + vehicle + environment” to set insurance premium [7].

The information collected from “people” contains the identity information of the driver, such as age, gender, etc. The information collected from " vehicle " includes times of the “four-sudden” (sudden acceleration, sudden deceleration, sudden turn, sudden braking), speed, driving mileage, driving time and so on. The data collected from “environment” are mainly the location of the car, road conditions and so on [8]. To a certain extent, these risk factors can affect the risk probability. The data used in this paper are the actual driving behavior data of 3070 users measured by Jiangsu Dina Digital Technology Co., Ltd.

4.2 Data Processing

The pre-processing scheme mentioned in this paper is to screen out the invalid data, and average the data to facilitate modeling and analysis [9].

The risk factor data needed include driver's age, sex, average daily number of times of the “four-sudden”(average daily sudden acceleration, average daily sudden deceleration, average daily sharp turn, average daily sudden braking), average daily mileage, average daily driving time, average daily night driving time, average daily speed, and fuel consumption per hundred kilometers [10]. Among them, driver's age and sex can reflect different driving population, and then reflects the driver's driving skills and psychological quality, which has a certain impact on the risk probability. The average daily number of times of the “four-sudden” can reflect the driver's driving habits, and driving habits greatly affect the risk probability. The average daily mileage and the average daily driving time can reflect whether the driver has fatigue driving phenomenon, which has a great influence on the risk probability. At the same time, the longer the mileage and the driving time, the higher the risk probability. Because of poor visibility at night, it is a period of high accidents, and the increase in the average daily driving time will directly lead to the increase of risk probability. The average daily speed and the fuel consumption per 100 kilometers can reflect the quality of the road conditions. The worse the road conditions, the higher the risk probability. The effect of the pre-processed data on the risk probability is more perceptual intuition and convenient for later modeling and analysis.

4.3 Data Modeling and Analysis

The modeling and analysis of the data are implemented by NEML. The pre-processed sample data are used to train and to determine the coefficient of each risk factor, and then to predict the risk probability.

The algorithm used in the machine learning model is linear logistic regression. Let the eigenvalue of the risk factor of the user i be the characteristic coefficient (weight) of the risk factor, then the weighted summation result (total eigenvalue) of the eigenvalue is:

$$x_i = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i} \tag{1}$$

Taking the total eigenvalue as an independent variable into the sigmoid function formula, the risk probability of the user i is [11]:

$$\begin{aligned} p_i\left(y=1 \mid x\right) &= \frac{1}{1+e^{-\left(\beta_0+\beta_1 x_{1,i}+\cdots+\beta_k x_{k,i}\right)}} \\ &= \frac{e^{\beta_0+\beta_1 x_{1,i}+\cdots+\beta_k x_{k,i}}}{1+e^{\beta_0+\beta_1 x_{1,i}+\cdots+\beta_k x_{k,i}}} \end{aligned} \tag{2}$$

Then the probability that the user does not risk is:

$$\begin{aligned} p_i(y=0\mid x) &= 1-p_i(y=1\mid x) \\ &= 1-\frac{e^{\beta_0+\beta_1x_{1,i}+\cdots+\beta_kx_{k,i}}}{1+e^{\beta_0+\beta_1x_{1,i}+\cdots+\beta_kx_{k,i}}} \\ &= \frac{1}{1+e^{\beta_0+\beta_1x_{1,i}+\cdots+\beta_kx_{k,i}}} \end{aligned} \tag{3}$$

The ODDS defined as the ratio of the probability of occurrence to the probability of no occurrence is:

$$\begin{aligned} ODDS &= \frac{p_i(y=1\mid x)}{p_i(y=0\mid x)} \\ &= \frac{p_i(y=1\mid x)}{1-p_i(y=0\mid x)} \\ &= e^{\beta_0+\beta_1x_{1,i}+\cdots+\beta_kx_{k,i}} \end{aligned} \tag{4}$$

The logarithmic occurrence ratio is equal to the total eigenvalue:

$$\begin{aligned} \ln ODDS &= \ln\frac{p_i(y=1\mid x)}{1-p_i(y=0\mid x)} \\ &= \beta_0+\beta_1x_{1,i}+\cdots+\beta_kx_{k,i} \end{aligned} \tag{5}$$

It can be seen from (5) that β_k (the characteristic coefficient of the k-th risk factor) is positively correlated with the logarithmic occurrence ratio $\ln ODDS$, while all other variables remain unchanged. In addition, $\ln ODDS$ is positively correlated with the user risk probability, so β_k is positively correlated with the risk probability.

Through the preprocessing of risk factor and historical accident situation data, machine learning training is carried out to determine the characteristic coefficient corresponding to each risk factor, and on this basis, the risk probability is calculated. The distribution curve of the risk probability is shown in Fig. 1.

In the Fig. 1, the overall risk probability distribution curve is indicated by a solid line, the risk distribution curve of the accidental users is represented by a dashed line, and the risk distribution curve of the no-accidental users is indicated by a dot-shaped line. It can be seen from the Fig. 1 that the risk probabilities of the accidental users are relatively high. Moreover, the risk probabilities of no-accidental users are relatively low. The conclusions obtained from the analysis in the Fig. 1 are similar to the actual accident situation. Moreover, after statistical analysis, the actual overall risk probability of the sample data is 0.1042. Correspondingly, the average risk probability predicted by the model is 0.1046. The prediction error is less than 1%, so the vehicle insurance model based on user behavior has reasonableness and accuracy, which can be applied to the vehicle insurance pricing system.

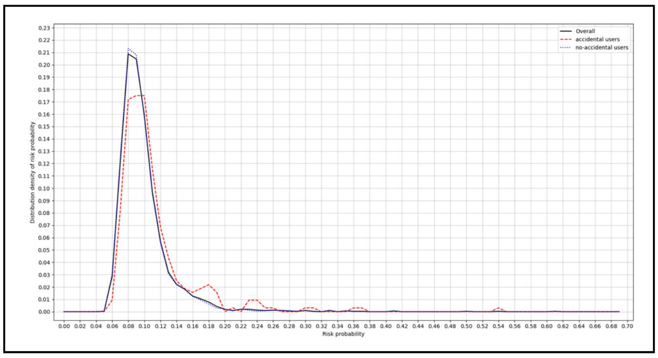

Fig 1. Risk probability distribution of the overall, accidental, and no-accidental users

4.4 Predicting Scheme of Vehicle Insurance

Using the sample data for training, the characteristic coefficient corresponding to each risk factor are determined. When a new user is insured, the user's driving data can be collected, the risk probability will be predicted, and the differential pricing of the vehicle insurance premium is based on the predicted risk probability. Moreover, the driving data and the historical accident situation collected from the new user can also be trained as new sample data to continuously improve the accuracy of the risk probability prediction. In the vehicle insurance pricing model based on user behavior, insurance companies divide insurance users into different groups according to the risk probabilities [12]. As shown in Fig. 2.

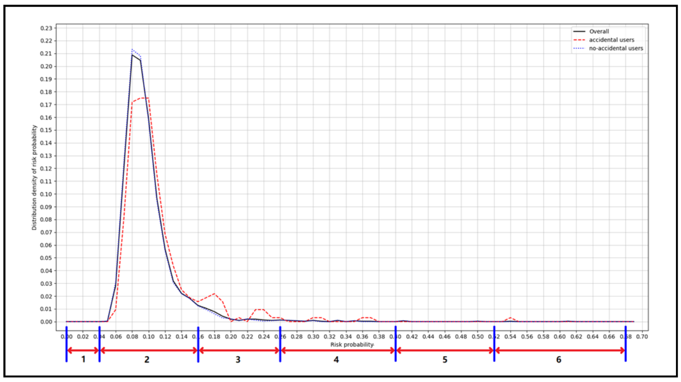

Fig 2. Segmentation of risk probability distribution

As the curve in the Fig. 2 is divided into different ranges, the users are divided into different groups, and the segment pricing model is adopted according to different groups of people. For users in segment 1, the risk probabilities are low, so the insurance company can retain the quality customer by lowering the premium. For users in segment 2, the proportion of users is large and the risk probabilities are medium, so normal pricing can be made based on the driving behaviors. In this way, it can provide auto insurance services to as many users as possible while ensuring the profits of insurance companies. For users in segment 3 or 4, the risk probabilities are slightly higher, but it is still within the insurance company's tolerance. So, the insurance company can increase the premium for these users to avoid the company's losses. For users in segments 5, 6 and with higher risk probability, these users have a high-risk probability, and the proportion of these users is small. If they are insured, it may bring losses to the insurance company. Therefore, it can be considered to give up this part of users.

In the UBI vehicle insurance model, the insurance company's predicted premium is:

$$\Pr e_{ubi} = \Pr e_{base} \times \eta \times \left(1 + \mu p\right) \tag{6}$$

In (6), $\mathrm{Pre}_{\mathrm{base}}$ is the basic premium, η is the discount according to the historical accident situation, μ is the segmentation penalty factor according to the risk probability, and p is the risk probability after zeroed by the mean. When the base premium is fixed, η times $(1+\mu p)$ is the proportional coefficient of the relative base premium of different users, so Pre_{ubi} is the insurance premium based on user behavior and historical accident situation .

The premium price given by the insurance company to the agent is:

$$Pre_{agent} = \left[Pre_{base} \times \eta \times (1+\mu p) \right] \times \lambda \tag{7}$$

In (7), λ is the discount given by the insurance company to the agent.

The reward mechanism can protect the interests of quality customers and retain quality customers for insurance companies. On the premise of ensuring no loss, the incentive rewards to cultivate users' behaviors are:

$$Pre_{client} = \left\{ \left[Pre_{base} \times \eta \times (1+\mu p) \right] \times \lambda \right\} \times \delta \tag{8}$$

In (8), δ is the budget ratio, and the maximum bounty that the user receives does not exceed the bonus budget Pre_{client} .

The differentiated pricing of vehicle insurance premiums for user driving behavior ensures the fairness of the vehicle insurance industry and reduces the insurance company's claim amount. Moreover, the supervision and reward for user behavior improve the user's driving behavior, reduce the number of accidents, reduce the losses caused by traffic accidents, and protect the interests of insurance companies and users [13].

5. Conclusion

The vehicle insurance premium pricing model of user behavior proposed in this paper analyzes the impact of user driving behavior on the historical accident situation and risk probability. The vehicle networking service and reward mechanism mentioned in the article for user driving behavior can improve the user's driving habits, reduce traffic accidents, avoid casualties and property losses, relieve traffic pressure and reduce pollutant emissions. Therefore, the vehicle insurance model will greatly promote the development of the vehicle insurance industry.

References

[1]. L. Zhou, Q. Chen, Z. Luo, H. Zhu and C. Chen, "Speed-Based Location Tracking in Usage-Based Automotive Insurance," 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS), Atlanta, GA, 2017, pp. 2252-2257.

[2]. C. Tribune. Allstate launches teen driving app for parents [EB/OL]. https:// www.chicagotribune.com/business/ct-xpm-2014-06-06-chi-allstate-launches-teen-driving-app-for-parents-20140606-story.html,2014/06/06.

[3]. J. Cen, Z. Wang, C. Wang and F. Liu, "A System Design for Driving Behavior Analysis and Assessment," 2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Chengdu, 2016, pp. 882-887.

[4]. B. Y. Lee, Y. Y. Shin and H. J. Bae, "Development of insurance server system based on vehicle driving information," 2012 7th International Conference on Computing and Convergence Technology (ICCCT), Seoul, 2012, pp. 156-159.

[5]. M. Yu, N. Liu, H. Zhao and J. Peng, "Intelligent Taxi Service System Based on Carrier-Cloud-Client," 2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE

12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom), Beijing, 2015, pp. 113-118.

[6]. L. Guo, M. Dong, K. Ota, Q. Li, T. Ye, et al. "A Secure Mechanism for Big Data Collection in Large Scale Internet of Vehicle," in IEEE Internet of Things Journal, vol. 4, no. 2, pp. 601-610, April 2017.

[7]. K. Zfnebi, N. Souissi and K. Tikito, "Driver behavior quantitative models: Identification and classification of variables," 2017 International Symposium on Networks, Computers and Communications (ISNCC), Marrakech, 2017, pp. 1-6.

[8]. J. Peng, N. Liu, H. Zhao and M. Yu, "Usage-based insurance system based on carrier-cloud-client," 2015 10th International Conference on Communications and Networking in China (ChinaCom), Shanghai, 2015, pp. 579-584.

[9]. K. Vassiljeva, A. Tepljakov, E. Petlenkov and E. Netšajev, "Computational intelligence approach for estimation of vehicle insurance risk level," 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, 2017, pp. 4073-4078.

[10]. C. D'Agostino, A. Saidi, G. Scouarnec and L. Chen, "Learning-Based Driving Events Recognition and Its Application to Digital Roads," in IEEE Transactions on Intelligent Transportation Systems, vol. 16, no. 4, pp. 2155-2166, Aug. 2015.

[11]. S. Taguchi, S. Inagaki, T. Suzuki, S. Hayakawa and N. Tsuchida, "Modeling and analysis of driver's decision making based on logistic regression model," SICE Annual Conference 2007, Takamatsu, 2007, pp. 2390-2395.

[12]. W. Nai, Y. Chen, Y. Yu, F. Zhang, D. Dong and W. Zheng, "Fuzzy risk mode and effect analyasis based on raw driving data for pay-how-you-drive vehicle insurance," 2016 IEEE International Conference on Big Data Analysis (ICBDA), Hangzhou, 2016, pp. 1-5.

[13]. P. Lertpunyavuttikul, P. Chuenprasertsuk and S. Glomglome, "Usage-based Insurance Using IoT Platform," 2017 21st International Computer Science and Engineering Conference (ICSEC), Bangkok, 2017, pp. 1-5.