Profiling Driver Behavior for Personalized Insurance Pricing and Maximal Profit

Bing He*, Dian Zhang†, Siyuan Liu‡, Hao Liu§, Dawei Han¶, Lionel M. Ni*
*Department of Computer and Information Science, University of Macau, Macau SAR, China
†College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China
‡Smeal College of Business, Pennsylvania State University, Pennsylvania, USA
§The Business Intelligence Lab, Baidu Research, Beijing, China
¶Auto Insurance Department, China Pacific Insurance Company, Shenzhen, China
Email: {mb65479,ni}@umac.mo, zhangd@szu.edu.cn, sxl68@psu.edu, liuhao30@baidu.com, iamhandawei@163.com

***Abstract*—Profiling driver behaviors and designing appropriate pricing models are essential for auto insurance companies to gain profits and attract customers (drivers). The existing approaches either rely on static demographic information like age, or model only coarse-grained driving behaviors. They are therefore ineffective to yield accurate risk predictions over time for appropriate pricing, resulting in profit decline or even financial loss. Moreover, existing pricing strategies seldom take profit maximization into consideration, especially under the enterprise constraints. The recent growth of vehicle telematics data (vehicle sensing data) brings new opportunities to auto insurance industry, because of its sheer size and fine-grained mobility for profiling drivers. But, how to fuse these sparse, inconsistent and heterogeneous data is still not well addressed. To tackle these problems, we propose a unified `PPP` (Profile-Price-Profit) framework, working on the real-world large-scale vehicle telematics data and insurance data. `PPP` profiles drivers' fine-grained behaviors by considering various driving features from the trajectory perspective. Then, to predict drivers' risk probabilities, `PPP` leverages the group-level insight and categorizes drivers' different temporal risk change patterns into groups by ensemble learning. Next, the pricing model in `PPP` incorporates both the demographic analysis and the mobility factors of driving risk and mileage, to generate personalized insurance price for supporting flexible premium periods. Finally, the maximal profit problem proves to be NP-Complete. Then, an efficient heuristic-based dynamic programming is proposed. Extensive experimental results demonstrated that, `PPP` effectively predicts the driver's risk and outperforms the current company's pricing strategy (in industry) and the state-of-the-art approach. `PPP` also achieves near the maximal profit (difference by only 3%) for the company, and lowers the total price for the drivers.**

***Index Terms*—Driver behavior profiling, personalized insurance pricing, company profit, trajectory data mining**

I. Introduction

Insurance is designed to protect the people and things we value most. Among it, auto insurance is one important category that provides financial protection against damages and liability resulting from car accidents. How to profile driver behaviors and devise pricing models, plays an essential role for insurance companies to gain profits and attract customers (drivers).

There has been considerable research on understanding the driver behavior and auto insurance pricing in the last decade. Traditional approaches, e.g., generalized linear models [1], [2], rely on drivers' static demographic information (e.g., age, gender and vehicle type) to compute the insurance price, but usually neglect the driving risk. Usage-Based Insurance (UBI) [3] based methods [4], such as Pay-As-You-Drive model [5] and Pay-How-You-Drive model [6], are introduced to model the driver mobility factors like time, mileage and speed for improving insurance pricing.

TABLE I: Drivers' claim statistics.

Driver Type	Driver Percentage	Accident Percentage
Claim count $= 0$	72%	0%
Claim count $= 1$	11%	20%
Claim count ≥ 2	**17**%	**80**%

However, the above solutions have the following drawbacks. 1) They are able to model only coarse-grained driving behaviors, resulting in inappropriate pricing, and incurring potential profit decline or even financial loss. As reported in 2016, over 70% auto insurance companies in China were in financial loss [7]. Table I shows the real-world claim data offered by a mainstream insurance company (due to the privacy concern, we omit the name) for the year 2016. Note that 17% of the drivers cause 80% of the accidents and claim indemnity. These drivers' risk behaviors necessitate further investigation at a finer-grained level. 2) The existing approaches cannot capture the time-variant driving risk. According to the survey conducted by the same company mentioned above, the number of overlapping accident-involved drivers between 2016 and 2017 is only about 3% (2.8% between 2015 and 2016). This indicates that driver risk behaviors often change over time. Thus, capturing the temporal risk patterns is crucial to build an accurate pricing model, which is required by the company as personalized and flexible. 3) Traditional models fail to link driver behaviors and pricing models with the ultimate goal of maximizing company profits, especially under the real-world enterprise constraints.

Recently, the rapid development of telematics [2], [8] in auto insurance industry has enabled to collect large amounts of fine-grained mobility data, like vehicle speed, acceleration, engine speed and so on, to better profile drivers' risk for pricing. With these telematics data, traditional methods [9] are usually leveraged to compute the insurance price, e.g., Pay-How-You-Drive model [10]. Although the mass of new telematics data has great potential to model driving behaviors more accurately and improve the granularity of risk prediction, it also poses new research challenges. First, telematics data