Report of Research Results

A. Title

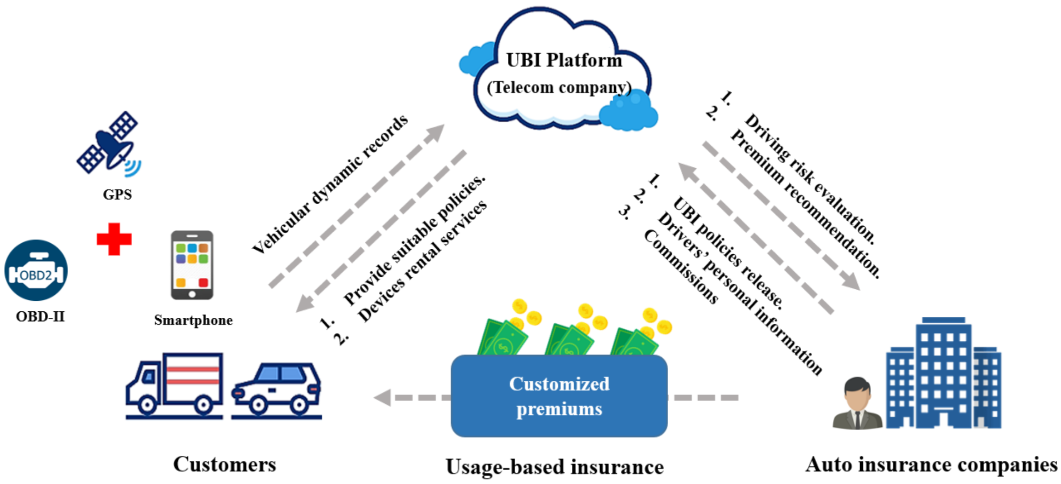
Design of a Usage-Based Insurance Platform for Evaluating Driver's Risk by Big Data Analysis

B. Researchers

Primary Researcher: Hsiao, Chih-Liang randyshiao0723@gmail.com
Co-researcher(s):
Lee, Wei-Hsun leews@mail.ncku.edu.tw
Tsai, Yi-Cheng errnnniie@gmail.com
Lu, Kuan-Hung k5e6n7n8y001@gmail.com

C. Summary

Usage-based insurance is a noticeable trend in auto insurance industry, many researches indicate that UBI can bring a WIN-WIN-WIN situation for auto insurers, customers and the whole society. However, due to the vicious cycle in auto insurance market, insurance companies are now lack of resource and unable to manipulate a complete UBI program. Therefore, we purposed a novel UBI platform which is designed to be provided by a telecom corporation. Telecom corporations possess cost advantage on data transmission and experience on renting devices which make them suitable to manipulate the platform. Furthermore, due to the dump pipe dilemma, digital transformation is a trend in telecom industry which means telecom companies would have motivations to provide the platform.



UBI platform proposed in this research

The core concept of UBI is to determine one's premium based on his/her realistic driving status. Owing to disadvantage of current UBI model, we purposed a new driving feature called "Driving pattern-N" to estimate driver's risk. This work design three experiments to compare the prediction performance of driving pattern-N model, driving pattern model, behavior-centric model and statistic model by using records in January, February and March from HO-HSIN. In addition, we label drivers into three risk levels and six risk levels due to the former is a common setting in previous driving safety studies, but the latter is more suitable for auto insurers to calculate premiums. However, the overall results show that behavior-centric model remain the highest predicted performance and follow by statistic model, driving pattern-N model and driving pattern model.

D. Aim of research

1. Prediction of driver's risk level
   A. Using data mining technique to find out driving pattern-N.
   B. Using driving pattern-N to predict driver's risk level.
   C. Labeling driver's risk level by driving score.
2. Compare performance between different models
   A. Driving pattern-N model vs. Driving pattern model
   B. Driving pattern-N model vs. Behavior-centric model
   C. Driving pattern-N model vs. Statistic model
3. Design an UBI platform manipulating by a telecom company
   A. Providing the risk assessment of user driving behaviors.
   D. Define relationships between stakeholders.
   B. Define the function of the UBI platform.

E. Method of Research & Progression

This research designs a data analysis pyramid to extract driving pattern-N and calculate driver's premium from vehicular dynamic records. The pyramid includes 6 layers: raw vehicular dynamic records, vehicular dynamic records, driving behaviors / near crashes, driving pattern-N, risk level and premiums. We acquired raw vehicular dynamic records, driving behavior records, journey records and employee performance evaluation from HO-HSIN BUS TRAFFIC CO., LTD (HO-HSIN). The records include 356 employees and 48,057 journeys from January to March in 2019. We import all the records into MySQL database through MS SQL database and Open Database Connectivity (ODBC) which spend approximately 24 hours for a month. From layer 1 to layer 2, vehicular dynamic records are pre-processed to be analyzable. 2,615,130 driving behaviors and 264,388 near crash events are extracted from layer 2 to layer 3. Two data mining algorithms: association rule mining and sequential pattern mining are utilized to discover driving patterns in layer 4. In layer 5, we use random forest algorithm to predict driver's risk level based on the results of layer 4. A premium formula is designed to calculate driver's personal premium in layer 6. More detail of each layers and transformation methods are presented in the complete report.

F. Result of Research

This research designs 3 experiments to compare performance of different models under different combination of training data and testing data. We divide training and testing data based on drivers, journeys and months in experiment 1, 2 and 3, respectively. The performance of each model in the three experiments are arranged in table 38 in the complete report. In the three experiments, driving pattern-N models have better performance than driving pattern model but worse performance than behavior-centric model and statistic model. Besides, due to some drivers lack driving pattern-N, we also select top 10 drivers who have the most association rules and sequential patterns within near crashes in every risk level to train and test. However, the result still remains the same.

In addition, the results of driving pattern models are different from Li et al. (2017). The 3 experiments in this research all show that behavior centric models have better performance on predicting driver's risk. Moreover, driving pattern models have the worst performance in experiment 1 and 2 and similar performance with DPN model in experiment 3. The difference between the method of formulating risk levels and definition of driving behaviors in Li et al. (2017) and this work would be the main reasons to cause the opposite results.
On the other hand, behavior centric models always have the best performance no matter in which experiment or risk level. The possible reason may be the two labeling method both imply the concept of behavior frequency which is similar with the features in behavior centric models. Besides, statistic models have much better performance than DPN model and DP model in experiment 2 but the performance is just merely better than DPN model and DP model in experiment 3. To conclude, the driver's dangerous behavior frequency are still the best features to predict driver's risk, which follow by last period risk level, driving pattern-N and driving pattern.

G. Future Areas to take off and going forward

Future research may further add normal behaviors such as “free driving” in layer 3 because we only consider dangerous behaviors instead of normal behaviors in this research. Moreover, using different criteria to label drivers may have different results from this research owing to there are no perfect label method and the label methods used in this research possess similar concept with the features in behavior-centric model. Besides, future research may also focus on the risk estimation of general drivers because they are the major customers of UBI. On the other hand, combining conventional factors (i.e. age, gender and accident records), PAYD and PHYD factors to evaluate driver risk may improve the predicted model performance, furthermore, image records which can represent more realistic driving status of driver may also apply to the prediction task in the future research. Regarding to predict algorithm, by using deep learning algorithms is possible to rise the performance of driving risk prediction (even better than behavior-centric model), due to the training features can be automatically generated.

H. Means of Official Announcement of Research Results

1. Proposed as a dissertation of master’s degree of Telecommunication Management of National Cheng Kung University.
2. Submit the results of the research to the Transport Research Part A: Policy and Practice Journal.