HAL open science

Payment Behavior Prediction and Statistical Analysis for Shared Parking Lots

Qingyu Xu, Feng Zhang, Mingde Zhang, Jidong Zhai, Jiazao Lin, Haidi Liu, Xiaoyong Du

▶ **To cite this version:**

Qingyu Xu, Feng Zhang, Mingde Zhang, Jidong Zhai, Jiazao Lin, et al.. Payment Behavior Prediction and Statistical Analysis for Shared Parking Lots. 17th IFIP International Conference on Network and Parallel Computing (NPC), Sep 2020, Zhengzhou, China. pp.288-293, 10.1007/978-3-030-79478-1_25 . hal-03768744

HAL Id: hal-03768744

https://inria.hal.science/hal-03768744v1

Submitted on 4 Sep 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

This document is the original author manuscript of a paper submitted to an IFIP conference proceedings or other IFIP publication by Springer Nature. As such, there may be some differences in the official published version of the paper. Such differences, if any, are usually due to reformatting during preparation for publication or minor corrections made by the author(s) during final proofreading of the publication manuscript.

Payment Behavior Prediction and Statistical Analysis for Shared Parking Lots

Qingyu Xu[1], Feng Zhang[1]*, Mingde Zhang[1], Jidong Zhai[2], Jiazao Lin[3], Haidi Liu[3], Xiaoyong Du[1]

[1] Key Laboratory of Data Engineering and Knowledge Engineering (MOE), and School of Information, Renmin University of China
[2] Department of Computer Science and Technology, Tsinghua University
[3] Zhongzhi Huaching (Beijing) Technology Co., Ltd.
*Corresponding Author: fengzhang@ruc.edu.cn

Abstract. As the sharing economy is booming in China, many intelligent shared parking lots appear. Since more and more Chinese households own cars, it is necessary to study the payment behavior of shared parking lots, which may represent the entire sharing economy. In detail, we analyze the factors that influence users' payment and predict users' payment behavior of whether and when users will deliver parking bills after parking. We use 29,733 real parking records provided by Huaching Tech, a top smart parking company in China, in our study. After a comprehensive statistical analysis, we use decision tree model to predict users' payment behavior. Experiments show that the decision tree model can reach 79% accuracy.

Keywords: Payment Behavior · Payment Time Prediction · Shared Parking Lots

1 Introduction

For intelligent shared parking lots, users' untimely payment behavior has become a serious problem. Fortunately, based on users' parking and payment records, we can predict their payment time and provide timely action to reduce delayed payment behaviors. To achieve this goal, we need to analyze the parking dataset containing parking records and user data with parking and payment history.

Parking related behavior prediction has become a hot research topic in recent days. Alho and others [1] proposed a prediction method for urban freight parking demand using linear regression and generalized linear models. We also explored the relations between parking and weather information [4, 13]. In terms of payment prediction, Wen and others [12] explored the customer behavior and developed a probabilistic graphical model to predict the consumption behavior of a single user.

Compared with the previous work, we predict the payment behavior with limited information, which poses large difficulties in our research. In detail, currently, most users have a small number of parking records; even worse, the users

who have delay payments usually have only one or two parking records. We not only need to clean and preprocess the data, but also need to design appropriate classification criteria.

We provide an effective solution to payment behavior prediction. We first analyze the factors that may cause users' delay payment, and extract relevant features from the obtained payment records. Second, we divide the features into two categories: numerical features and category features, and then we preprocess these two types of features. Third, we use decision tree model [10] and 10-fold cross-validation [6] to predict payment behavior. Experiments show that we can predict the payment types of orders and gain good prediction results of 79.73% accuracy.

2 Background and Motivation

We introduce the delayed payment behavior and decision tree model in this section.

Delayed Payment The purpose of delayed payment is to provide customers with a time buffer for convenience. However, by analyzing the payment records provided by Huaching Tech (http://www.huaching.com/), we find that 4.22% users prefer to pay after three days, and some users forgot to pay the parking fee, which could be a potential problem in the sharing economy. Fortunately, if we can predict the user payment behavior, the parking system can provide better services, such as payment reminders, which is also the motivation of this work.

Decision Tree The decision tree model is a predictive model, and it is composed of a tree-like model and possible results [10]. According to the different types of target variables, the decision trees can be categorized as classification trees, regression trees, and CART (Classification And Regression Trees) [10]. In our work, we use payment records to train and generate CART for predicting the users' payment time. We choose this model due to its simplicity and effectiveness.

3 Payment Behavior Prediction and Statistical Analysis

We show our prediction method and statistical analysis in this section.

Payment Records. We randomly select a parking lot in Hangzhou, China. Our users' parking dataset spans 14 months from April 9th, 2019 to June 29th, 2020, which consists of 29,733 parking records. We divide the records into three categories: 1) arrears, whose payment durations are more than three days, 2) late payments, whose payment durations are longer than one day but less than three days, and 3) timely payments, whose payment durations are less than one hour.

To better understand the correlation between payment data and predict payment behavior, we analyze the correlations between payment time and related features. The related features can be divided into two categories: categorical

features and numerical features. The categorical features include year, month, and user's identity (whether a user has a membership). The numerical features include 1) the money calculated as order amount minus deposit amount, 2) the total number of parkings, 3) the parking duration of this record, and 4) the average payment duration in days.

Categorical Features. We first explore the relationship between user late payment ratio and different category features. We show part of the relation between dates and payment behaviors in Figure 1. We can see that the ratio of orders to be paid has an upward trend, possibly because the number of users is increasing year by year. We can also observe that there is a sharp decrease in February 2020, which is due to the coronavirus pandemic (COVID-19) [7].

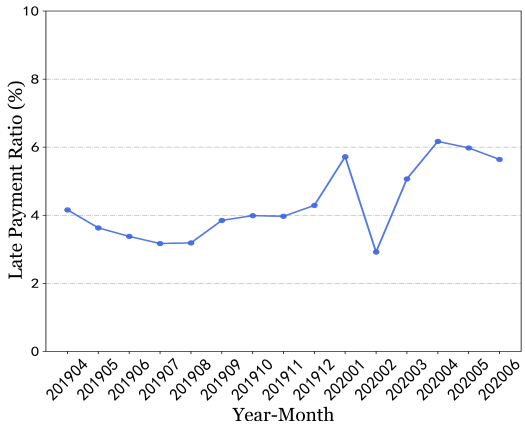

Fig. 1. Relation between dates and payment behaviors.

Numerical Features. We next analyze the numerical features and possible influence factors. We first clean up the outliers. We regard the records with the parking duration less than 5 minutes as noise data. From our analysis of these numerical features, we find that there is a certain linear correlation between the user's past payment time and the payment time of this order. We also observe that the users who have unpaid records have only one or two parking times.

4 Experiment

In this section, we evaluate our payment prediction method and analyze the results.

Evaluation Method. We use the decision tree model [10], as discussed in Section 2, to predict users' payment behavior. We train and test the model using 10-fold cross-validation [6]. The evaluation metrics we use are the average F1-macro [11] shown in Equation 1 and accuracy [8] shown in Equation 2. In the 10-fold cross-validation, the datasets are divided into ten equal parts. There are ten rounds, and one part is used as a test set and the rest as a training set in one round, which means that we use 26,760 records as training data and the rest as

test data in one round. Our validation results(F1-macro, accuracy) are averaged over the rounds.

$$\text{Average F1} - \text{macro} = \sum_{j=0}^{10} \sum_{l=1}^{3} \frac{2 * \left(\text{precision}_{jl} * \text{recall}_{jl}\right)}{\text{precision}_{jl} + \text{recall}_{jl}} / 30 \tag{1}$$

$$\text{Average accuracy} = \sum_{j=0}^{10} \frac{\text{CR}_j}{\text{AL}_j} / 10 \tag{2}$$

In Equation 1, l denotes the classes to be classified. In Equation 2, CR stands for the number of correctly predicted results, and AL denotes the number of all test samples.

Decision Tree. We use the decision tree, CART [10], to predict the orders' payment type. As described in Section 2, decision tree is a supervised model of classification and it learns decision rules from features. We develop the model by using the *DecisionTreeClassifier* module in SKlearn [9] with python. The depth of the decision tree is automatically set by SKlearn.

Results. We use the features mentioned in Section 3 and use the CART model for our prediction. We demonstrate our partial results in Figure 2; other parts of the results are similar. To distinguish between unpaid orders and paid orders, we change the payment time of unpaid orders to negative values. From Figure 2, We have the following observations. First, our method can predict most cases; on average, it achieves 79.73% accuracy. Second, 80.26% of cases belong to timely payments, 2.96% of cases belong to late payments, and 16.79% of cases belong to arrears, which implies that the current payment situation is fine. Third, our method is good at identifying unpayable orders, which is important to shared parking scenarios.

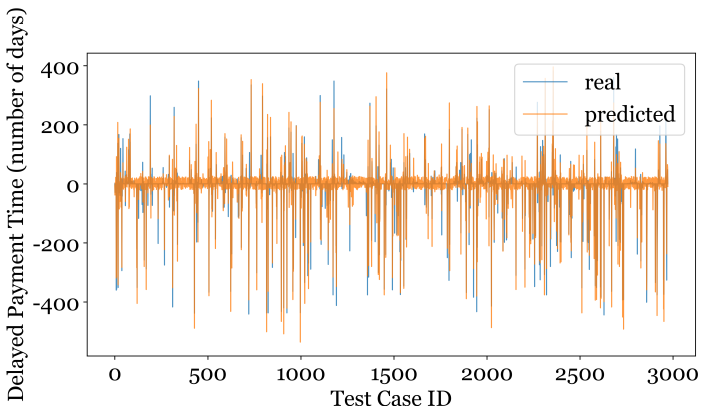

Fig. 2. Predicted payment time vs real payment time.

5 Conclusion

In this paper, we have analyzed payment behavior for a shared parking lot by leveraging related payment features. We exhibit our methods on how to perform feature analysis, and also use the CART model for parking payment prediction. Experimental results demonstrate the advantage of our method, and show that the CART model is effective in payment prediction, which achieves 79.73% accuracy.

6 Acknowledgments

This work is supported by the National Key Research and Development Program of China (No. 2018YFB1004401), National Natural Science Foundation of China (No. U1911203, 61802412, and 61732014), and Beijing Natural Science Foundation (No. L192027). Feng Zhang is the corresponding author of this paper (fengzhang@ruc.edu.cn).

References

1. Alho, A.R., Silva, J.d.A.e.: Freight-trip generation model: predicting urban freight weekly parking demand from retail establishment characteristics. Transportation Research Record **2411**(1), 45–54 (2014)
2. Cahill, M.H., Lambert, D., Pinheiro, J.C., Sun, D.X.: Detecting fraud in the real world. In: Handbook of massive data sets, pp. 911–929. Springer (2002)
3. Chen, X.: Parking occupancy prediction and pattern analysis. Dept. Comput. Sci., Stanford Univ., Stanford, CA, USA, Tech. Rep. CS229-2014 (2014)
4. Feng, N., Zhang, F., Lin, J., Zhai, J., Du, X.: Statistical analysis and prediction of parking behavior. In: Tang, X., Chen, Q., Bose, P., Zheng, W., Gaudiot, J.L. (eds.) Network and Parallel Computing. pp. 93–104. Springer International Publishing, Cham (2019)
5. Florian, M., Los, M.: Impact of the supply of parking spaces on parking lot choice. Transportation Research Part B: Methodological **14**(1-2), 155–163 (1980)
6. Geisser, S.: Predictive inference: an introduction, 1993
7. Mehta, P., McAuley, D.F., Brown, M., Sanchez, E., Tattersall, R.S., Manson, J.J., Collaboration, H.A.S., et al.: COVID-19: consider cytokine storm syndromes and immunosuppression. Lancet (London, England) (2020)
8. Metz, C.E.: Basic principles of roc analysis. In: Seminars in nuclear medicine. vol. 8, pp. 283–298. WB Saunders (1978)
9. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al.: Scikit-learn: Machine learning in python. the Journal of machine Learning research **12**, 2825–2830 (2011)
10. Qiu, Q., Huang, W.: Forward supervised discretization for multivariate with categorical responses. Big Data & Information Analytics **1**(2/3), 217–225 (2016)
11. Sasaki, Y., et al.: The truth of the f-measure. 2007 (2007)
12. Wen, Y.T., Yeh, P.W., Tsai, T.H., Peng, W.C., Shuai, H.H.: Customer purchase behavior prediction from payment datasets. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. pp. 628–636 (2018)
13. Zhang, F., Feng, N., Liu, Y., Yang, C., Zhai, J., Zhang, S., He, B., Lin, J., Du, X.: PewLSTM: Periodic LSTM with Weather-Aware Gating Mechanism for Parking Behavior Prediction. In: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20. pp. 4424–4430. International Joint Conferences on Artificial Intelligence Organization (7 2020)