Taxi or Hitchhiking: Predicting Passenger's Preferred Service on Ride Sharing Platforms

Lingyu Zhang[1], Wei Ai[2,*], Chuan Yuan[3,*], Yuhui Zhang[4,*], Jieping Ye[1]
[1] AI Labs, Didi Chuxing, Beijing, China
[2] School of Information, University of Michigan, Ann Arbor, MI, USA
[3] School of Software Engineering, Huazhong University of Science and Technology, Wuhan, Hubei, China
[4] School of Software Engineering, Beijing Jiaotong University, Beijing, China
zhanglingyu@didichuxing.com,aiwei@umich.edu,blackyuanc@163.com,14301027@bjtu.edu.cn
yejieping@didichuxing.com

ABSTRACT

Ride sharing apps like Uber and Didi Chuxing have played an important role in addressing the users' transportation needs, which come not only in huge volumes, but also in great variety. While some users prefer low-cost services such as carpooling or hitchhiking, others prefer more pricey options like taxi or premier services. Further analyses suggest that such preference may also be associated with different time and location. In this paper, we empirically analyze the preferred services and propose a recommender system which provides service recommendation based on temporal, spatial, and behavioral features. Offline simulations show that our system achieves a high prediction accuracy and reduces the user's effort in finding the desired service. Such a recommender system allows a more precise scheduling for the platform, and enables personalized promotions.

ACM Reference Format:
Lingyu Zhang[1], Wei Ai[2,*], Chuan Yuan[3,*], Yuhui Zhang[4,*], Jieping Ye[1]. 2018. Taxi or Hitchhiking: Predicting Passenger's Preferred Service on Ride Sharing Platforms. In *SIGIR '18: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, July 8–12, 2018, Ann Arbor, MI, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3209978.3210153

1 INTRODUCTION

Ride sharing has become an increasingly popular option for people to meet their travel needs. Ride-sharing platforms such as Uber and Didi Chuxing have been growing rapidly in the past few years, completing more than 4 billion[1] and 7 billion[2] rides in year 2017 alone. The growing popularity has also led to an increase in the diversity of users' requests. While some users are cost-sensitive and prefer the cheapest option, some other users would rather pay a high price for a higher-standard service. It is also very common for some users to use the ride sharing app to request a taxi pickup.

In light of the diversity in people's needs, ride sharing companies are providing multiple services for users to choose, which, however, puts an extra burden for users. Sometimes, a user has to swipe several times to find the exact line-of-business (s)he is looking for.

In this paper, we model users' choice of services, and build classifiers to predict the user's choice when opening the app. Intuitively, such choices are correlated with several features: (a) the services that the users have chosen in previous trips, and (b) the current time and location of the users.

For each feature, we first characterize its correlation with the choice of services using mutual information, and then build a predictive model with the feature. We also combine all features based on an ensemble model. Finally, we conduct offline simulation experiments to verify the effectiveness of our models with realworld data. The results show that our ensemble model performs very well. In addition, although all of our models increase the prediction accuracy from the majority-guess baseline, no particular model outperforms all other models for all user groups.

The rest of the paper is organized as follows: Section 2 briefly reviews the related work. In Section 3, we formalize the prediction task, the evaluation metrics, and the experiment setup. In Section 4, we describe the prediction models based on sequential features, spatial features, and temporal features. We compare the performance of different models in Section 5. We discuss the limitation and future work in Section 6 and conclude in Section 7.

2 RELATED WORK

Our work is related to prediction and recommendation problems in transportation systems, especially in modeling the travel need of the passengers. Yet existing works focus more on the destination prediction [5, 8] and demand prediction [4]. The most relevant work is [8], which focuses on the prediction of the destination based on spatio-temporal contexts. Our work adopts a similar approach, but we focus on a novel aspect of the travel need, i.e. the preferred service. The use of contexts in prediction is inspired by the research on context-aware recommender systems [1]. In fact, the use of spatio-temporal contexts in user modeling has been well studied in information retrieval and recommender system community (e.g. [6, 7]).

[1] https://www.recode.net/2018/1/5/16854714/uber-four-billion-rides-coo-barney-harford-2018-cut-costs-customer-service, retrieved Feb. 13th, 2018.
[2] http://www.didichuxing.com/en/aboutus/milestones, retrieved Feb. 13th, 2018.

*This work is done when these authors were interns at Didi Chuxing.

SIGIR '18, July 8–12, 2018, Ann Arbor, MI, USA

ACM ISBN 978-1-4503-5657-2/18/07…$15.00
https://doi.org/10.1145/3209978.3210153

3 THE PREDICTION TASK

3.1 Background

Our analysis is conducted in Didi Chuxing [3], which provides world-leading ride sharing services. The most popular four are:

Express Express service offers the affordable mobility service where passengers can request for individual or pooled rides. Express made up more than 70% of the trips on RidsS.

Premier Premier service offers high-end mobility experience with luxury vehicles and drivers trained with highest service standards.

Taxi Taxi service works with city-registered taxis and sends passengers' pickup requests to nearby taxis.

Hitch Hitch provides a social carpooling platform that helps commuters find/provide carpool service to save the cost.

Each line of business is presented as a tab in the app. Once the user decides on a line of service, (s)he can input her/his origin, destination, and scheduled departure time (if not immediate). The user will then see an estimated price and can send the request. The prediction task takes place at the moment a user launches Didi. The user would land our predicted service tab upon opening the app, while (s)he can always switch to other tabs at any time.

3.2 Prediction Task

The prediction task we tackle in this paper is given as follows: *Given the historical behavior of a user and the spatio-temporal context, how well can we predict the service which the user will request.* In particular, we base our analysis on the following historical data and spatio-temporal contexts:

- The time and location of each request the user has submitted in the past, together with the service chosen by the user.
- Current time and location at the time of prediction.

Arguably, there are other useful features, such as the intended destination. However, for the current design of the application, the user has to choose the service tab before entering the destination, so the destination of the current request would not be available at the time of prediction. The destinations of previous trips could be helpful, and we plan to incorporate such information in future work.

3.3 Evaluation Metrics and Experiment Setup

Since the prediction task is a classic multi-class classification problem, we would use accuracy as the evaluation metric. However, since more than 70% trips are made with *Express*, the classification task suffers from imbalanced data. Therefore, we will also report the macro-F1 score, in order to make sure the classifier is fair to all services. This is crucial in assuring the product team to adopt our model without harming any particular line-of-business.

The experiment is conducted on a metropolitan area with over 10 million population. In consecutive 3-month time, over 4 million active users have made one or more trips in this city using Didi. However, a majority of them have only made a few trips. In this work, we focus on those more active users of the app, that is, users with 10 or more trips in the same 3-month period. As shown in Figure 1, although these users are merely 18% of the aforementioned active users, they contributed more than 67% of the total orders.

[3] https://www.didichuxing.com (referred to as Didi thereafter), retrieved May 1st, 2018.

Table 1: Descriptive statistics of the data set.

# Users	19,985
# Trips	661,195
% of trips in *Express*, *Premier*, *Taxi*, *Hitch*	72.9%, 3.8%, 8.3%, 14.9%

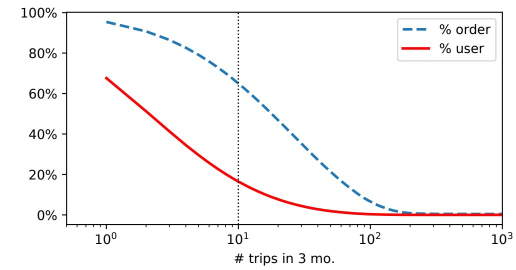

Figure 1: Distribution of the users' activity level. Red-solid line: % of users having made $\geq x$ trips. Blue-dash line: % of orders made by users who have made $\geq x$ trips.

We hold out the last week of the 3 months as the test period, and used the rest as the training period. The week of the test period is a normal week without any national holidays or disruptive events. We sampled 19,985 users who have 10 or more trips in the 3-month period and at least 1 trip in both training and test period. Descriptive statistics is shown in Table1

3.4 Baseline

Before introducing our models, we will first evaluate the performance of a few simple baselines.

Majority: Since *Express* service is the most affordable and accessible among the four services, by majority guessing, we would predict *Express* for all trips. This trivial baseline would result in an accuracy of 0.72 and an F1-score of 0.21.

Local Majority: Although in the city level, the *Express* service is the most popular choice, the case can be different in different local areas. For example, in a high-end residential community or the CBD, most users would prefer the *Premier* service. Therefore, we implement a basic grid structure by partitioning the map in to 1 km^2 square cells and calculate the most popular service in each cell to make prediction. We would then predict the local majority service to all trips within the same sell. This results in an accuracy of 0.74 and an F1-score of 0.31.

4 MODELING USER BEHAVIOR IN CONTEXT

The two baselines perform poorly simply because they ignore the difference between different users. A natural idea is to extend the majority guess to individual level, that is, we use the most frequently used service of each user to predict his/her next trip. A simple simulation shows that such an *Individual Majority* model achieves an accuracy of 0.83 and an F1-score of 0.70.

$$\begin{bmatrix} .8 & .1 & .1 & 0 \\ .1 & .7 & .2 & 0 \\ & & \ddots & \\ & M_{ij} & & .9 \end{bmatrix} \quad \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Markov ***Indi-Majority*** ***Last Used***

Figure 2: *Markov* Model $M_{ij} = P(S_t|S_{t-1})$ and its relationship with the *Individual Majority* model and the *Last Used* model.

4.1 Sequential Models

We can see that the *Individual Majority*'s performance is much higher than the baselines, which suggests that users' choice may be dependent on their previous trips. In fact, if we use S_t to represent the service of the t-th trip of a user. The entropy of the user's choice $H(S)$ is 0.48 on average. Yet the conditional entropy given the user's choice in the last trip is 0.4, suggesting a mutual information [2] of 0.08. Therefore, we propose two models to characterize the relationship between sequential orders.

Last Used: The first model assumes that users will choose the same service as their last trip.

Markov: A more general approach is to use a Markov matrix to model the correlation between the chosen service with the most recent trip. That is, we use a Markov matrix M to characterize the conditional probability of choosing service S_t conditioned on the service chosen last time S_{t-1}.

In fact, the *Last Used* model and the *Individual Majority* models are two special cases. In the *Last Used* model, the transition matrix is essentially an identity matrix, while in the *Individual Majority*, one column of the transition matrix is 1 and the rest is zero. See Figure 2 for illustration.

4.2 Spatio-Temporal Model

One essential element missing in majority guessing and the sequential model is that users take trips for different purposes. For example, Jim uses Didi for all his travel needs. For his morning commute from the suburban area, he may prefer to request a *Hitch*, which is not only affordable, but also available as many nearby residents are heading towards downtown at that time. When he heads out of the CBD in the afternoon to meet a potential customer, he might prefer a *Premier* for business purpose.

This intriguing example suggests the importance of understanding the purpose of the users. The destination of a trip would be the ideal indicator of the purpose. Yet in the current application, without knowing the destination of the trip at the time of prediction, we decide to directly predict the choice of service from the spatio-temporal features of the trip origin.

In fact, such spatial and temporal features are strongly correlated with the choice of service. The mutual information between the service and the location (indexed by the aforementioned grid structure) is 0.288, and the mutual information between the service and the time (indexed by the hour of day) is 0.293, both of which are much larger than that of the most recent used service shown in Section 4.1.

Spatial Model: Although the grid structure greatly reduces the entropy in the training set, creating a "Cell Majority" model for each user would not generalize well in the testing period due to data sparsity. Here, we adopt the decision tree classifier and build a decision tree for each user to model the user-specific relationship between location and service. The model is trained with CART with all default parameters.

Temporal Model: Unlike spatial features where the location of a POI is typically fixed, the temporal distribution of the requests suffers less from sparsity. Therefore, we adopt a Gaussian Mixture Model approach. Exploratory analysis suggests that $k = 2$ components are enough to capture most of the trips of a service used by a user. Note that for the time of the day, 23:59 is very close to 00:01, such variables are called circular quantities [3]. We adapt the approach reported in [8] to calculate the mean and variance of the distribution, and we modify the traditional GMM accordingly to account for the circular quantities.

4.3 Ensemble Model

So far, we have introduced several different models, each of which utilizes a particular feature: The *Markov* model utilizes the sequential feature, the *Spatial* model utilizes the spatial feature, and the *Temporal* utilizes the temporal features. These models allow us to measure how much improvement each feature brings, however different users may benefit from different models since their preference may be more correlated with some features than others. It is desirable to have one unified model that combine all features.

However, an excessively complex model would surely suffer from sparsity. Besides, some features are highly correlated with each other. For example, for users that take *Hitch* from home to work in the morning and *Express* from work to home in the evening, either location or time alone would be sufficient. Thus, we decide to apply the ensemble approach.[9].

Ensemble: The three models, namely *Markov*, *Spatial*, and *Temporal* vote on their best guesses with equal weight and the service with the most vote is the final prediction. In case of tie (all three models predicts differently), the prediction from *Spatial* model is chosen.

5 COMPARISON OF DIFFERENT MODELS

Following the experiment setup described in Section 3.3, we compare the performance of different models. We further break down our analysis on users with different levels of activities, based on their number of trips made in the training periods. In Figure 3, we split users into five bins such that the number of users in each bin is similar. The accuracy of the models is plotted in Figures 3, and we can make several observations:

The *Last Used* model achieves the highest performance for the most inactive users. It even performs better than its more generalized form *Markov* Model. This suggests that these users are more likely to request the same service as they did last time. As we have discussed, the *Last Used* model is a special case of the *Markov* model, yet for the inactive users, the *Last Used* out-performs the *Markov*

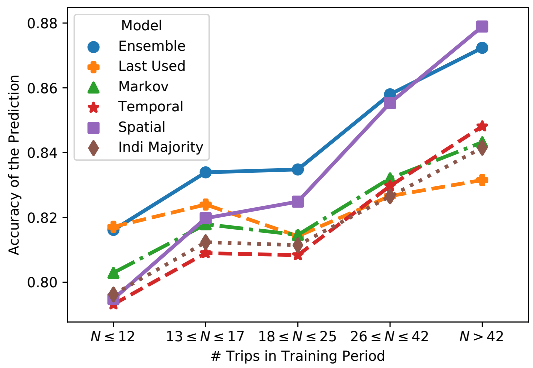

Figure 3: Accuracy of different models. The baseline scores from *Majority* and *Local Majority* are 0.72 and 0.74, which are omitted in this plot.

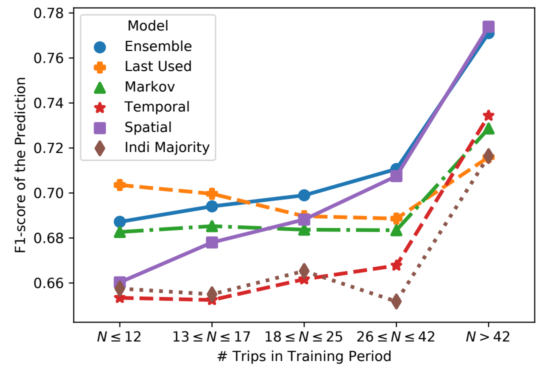

Figure 4: Macro-F1 scores of different models. The baseline scores from *Majority* and *Local Majority* are 0.21 and 0.31, which are omitted in this plot.

model. This is probably because there is not enough training data to estimate the Markov matrix.

The *Ensemble* model achieves the highest performance for users with a medium level of activities. It outperforms the three individual models. These users can be heterogeneous, and a same feature cannot models the behavior of all user in this group.

The *Spatial* model achieves the highest performance for the most active users. These users have incorporated the Didi app into their daily life. Their preference of service is strongly associated with the location where they launch Didi.

The *Individual Majority* model has an increasing performance as the activity level gets higher. This suggests that the more active an user is, the more concentrated his/her requested service is.

Since the data set is highly unbalanced, we also report the macro-F1 score (unweighted mean of the F1 score of each class) in Figure 4. Although the values of the score is hard to interpret, it can tell us that our model is fair to all four types of services (when compared with the *Majority* and *Local Majority* model). We can see that the comparison of different models is similar to that in Figure 3. The *Last Used* model performs best for in-active users while the *Ensemble* and *Spatial* model perform best for active users.

6 DISCUSSION AND FUTURE WORK

Although our models have shown great improvement over the baselines, there are still a lot to explore to improve our prediction:
Adding More Features: Our model can greatly benefit from including more features. For example, the type of the origin POI (business, residential, shopping, etc.) would allow us to better understand the purpose of the trip, which translates into the preferred line-of-business.
Counting for the Supply: So far, our analysis focuses only on the demand (passenger) side of the ride sharing. Yet the supply (driver/vehicle) side can also affect users' choice. For example, many drivers only work as *Express* drivers during the weekend. Thus, an experienced user knowing such supply pattern would avoid requesting *Express* on weekdays.
Controlling Other Contexts: Other contexts may potentially affect the user's choice. For example, the coupons and promotions provide strong incentives for the users to choose a certain service. For another example, the "default" option offered by the system could also affect the user's preference without the user knowing it. The current app interface implements a *Last Used* model as the default choice. Could that be the reason why the *Last Used* works well for inactive users? We plan to explore these in our future work.

7 CONCLUSION

In this paper, we study a novel problem of predicting the line-of-business that a user is going to use for his/her next ride sharing trip. We model the problem as a multi-class classification and propose several classifiers based on sequential, spatial, and temporal features. Offline simulations show that no particular classifier out-performs other classifiers and an ensemble model achieves the best overall performance. Our work serves as a necessary step to model and predict the users' travel need, which can help ride sharing platforms to do a more effective scheduling.

REFERENCES

[1] Gediminas Adomavicius and Alexander Tuzhilin. 2015. Context-aware recommender systems. In *Recommender systems handbook*. Springer, 191–226.
[2] Thomas M Cover and Joy A Thomas. 2012. *Elements of information theory*. John Wiley & Sons.
[3] Nicholas I Fisher. 1995. *Statistical analysis of circular data*. Cambridge University Press.
[4] Luis Moreira-Matias, Joao Gama, Michel Ferreira, Joao Mendes-Moreira, and Luis Damas. 2013. Predicting taxi–passenger demand using streaming data. *IEEE Transactions on Intelligent Transportation Systems* 14, 3 (2013), 1393–1402.
[5] Kohei Tanaka, Yasue Kishino, Tsutomu Terada, and Shojiro Nishio. 2009. A destination prediction method using driving contexts and trajectory for car navigation systems. In *Proceedings of the 2009 ACM symposium on Applied Computing*. ACM, 190–195.
[6] Hongzhi Yin, Yizhou Sun, Bin Cui, Zhiting Hu, and Ling Chen. 2013. LCARS: a location-content-aware recommender system. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 221–229.
[7] Quan Yuan, Gao Cong, Zongyang Ma, Aixin Sun, and Nadia Magnenat Thalmann. 2013. Time-aware point-of-interest recommendation. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. ACM, 363–372.
[8] Lingyu Zhang, Tao Hu, Yue Min, Guobin Wu, Junying Zhang, Pengcheng Feng, Pinghua Gong, and Jieping Ye. 2017. A taxi order dispatch model based on combinatorial optimization. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2151–2159.
[9] Zhi-Hua Zhou. 2012. *Ensemble methods: foundations and algorithms*. CRC press.