model may make a few conservative predictions since the long-tailed part is often caused by rare cases.

7 RELATED WORK

In this section, we summarize the literature works in two related areas to our study: 1) inverse reinforcement learning, and 2) user choice modeling. Learning reward function $R(s, a)$ of an MDP is a problem which has been broadly studied in the past decade, which is called Inverse Reinforcement Learning. The inverse reinforcement learning problem (IRL) is to find a reward function $R(s, a)$, such that the distribution of action and state sequences under a (near-)optimal policy with respect to $R(s, a)$ matches the demonstrated trajectories from an agent [13, 18]. A broadly used solution to IRL problem [5] proposes a model-free method to find the policy. Besides, neural-network-based reward function [7] is considered, which can represents more complex expert behaviors. And the Inverse Reinforcement Learning can be viewed as a special application of GAN [6, 8]. Also, IRL is used to model human's sequential decision-making process and is used to predict human behaviors in urban transit system [16]. Inspiring by these work, we also use Inverse Reinforcement Learning algorithm to extract driver's preference vector. User choice modeling has been extensively studied in the literature with applications, which investigate how users make decisions in various application scenarios. For examples, in [15], they use random utility maximization and random regret minimization to analyze users' choice on park-and-ride lots. In [18], the authors propose a probabilistic approach to discover reward function for which a near-optimal policy closely mimics observed behaviors. However, differing from these works, we employ data-driven approaches to study the unique decision-making process of urban public transit passengers.

8 CONCLUSION

In this paper, we introduce a novel drivers' behavioral prediction framework that can make accurate prediction of driver's future behavior. In this framework, we use driver's historical intra-cycle behaviors to learn driver's preference on his/her decision-making process and then combine the learned preference with inter-cycle features to predict driver's behavior in the future. Our extensive evaluation results based real-world data sets demonstrate that our framework can achieve the prediction accuracy of $MAE = 0.29$, which is on average 13% lower than existing state-of-the-art approaches without using driver's preference. We believe that the idea can benefit domains where both micro (i.e., intra-cycle) and macro (i.e., inter-cycle) decision making process are involved. For example, e-commerce platforms, like Amazon, Alibaba and etc, may apply this methodology to improve the prediction of customers' life time value. This paper gives a simple framework to capture customers' intrinsic preference from the decision making process (search, click, add to cart, ..., search, etc) of each visit, and to combine the learned preference with features aggregated (LSTM, etc) from recent visits.

9 DEPLOYMENT

The driver behavioral prediction algorithm has already been deployed in the production system, used in various applications. For example, if a driver is predicted to be less effective in the next 30 days (less finished orders), we can define a user-specific incentive strategy to increase his willingness to work, based on certain root cause analysis on this driver's recent behavior. We also rely on the prediction system to help evaluate a new pricing strategy, when randomized controlled experiments are restricted by law in the scenario. We apply the new pricing strategy on all the drivers for a couple of days, and the difference between the new pricing strategy versus the old is simply the mean of the observed metrics minus that of the predicted metrics.

The driver behavioral prediction system protects the privacy rights of drivers from three aspects. 1) We avoid using any personally identifiable feature, like social security number, in any of our algorithm. Only features like age or gender that may be attributable to more than one individual are included in the static features of our prediction algorithm. 2) The data are only available to a small group of people who are working on this project, and any one sharing drivers' personal information will break the law. 3) The behavior prediction system only returns the prediction results and related diagnostic information to related users of the system.

REFERENCES

[1] Pieter Abbeel and Andrew Y Ng. 2004. Apprenticeship learning via inverse reinforcement learning. In *ICML*. 1.

[2] Alex Beutel, Paul Covington, Sagar Jain, Can Xu, Li Jia, Vince Gatto, and Ed H. Chi. 2018. Latent Cross: Making Use of Context in Recurrent Recommender Systems. In *Eleventh Acm International Conference*.

[3] Derya Birant. 2011. Data mining using RFM analysis. In *Knowledge-oriented applications in data mining*. IntechOpen.

[4] Michael Bloem and Nicholas Bambos. 2014. Infinite time horizon maximum causal entropy inverse reinforcement learning. In *53rd IEEE Conference on Decision and Control*. IEEE, 4911–4916.

[5] Abdeslam Boularias, Jens Kober, and Jan Peters. 2011. Relative entropy inverse reinforcement learning. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. 182–189.

[6] Chelsea Finn, Paul Christiano, Pieter Abbeel, and Sergey Levine. 2016. A connection between generative adversarial networks, inverse reinforcement learning, and energy-based models. *arXiv preprint arXiv:1611.03852* (2016).

[7] Chelsea Finn, Sergey Levine, and Pieter Abbeel. 2016. Guided cost learning: Deep inverse optimal control via policy optimization. In *International Conference on Machine Learning*. 49–58.

[8] Jonathan Ho and Stefano Ermon. 2016. Generative adversarial imitation learning. In *Advances in Neural Information Processing Systems*. 4565–4573.

[9] Su-Yeon Kim, Tae-Soo Jung, Eui-Ho Suh, and Hyun-Seok Hwang. 2006. Customer segmentation and strategy development based on customer lifetime value: A case study. *Expert systems with applications* 31, 1 (2006), 101–107.

[10] Ugo Lachapelle, Larry Frank, Brian E Saelens, James F Sallis, and Terry L Conway. 2011. Commuting by public transit and physical activity: where you live, where you work, and how you get there. *Journal of Physical Activity and Health* (2011).

[11] Erich L Lehmann and George Casella. 2006. *Theory of point estimation*. Springer Science & Business Media.

[12] Daniel Neil, Michael Pfeiffer, and Shih-Chii Liu. 2016. Phased lstm: Accelerating recurrent network training for long or event-based sequences. In *Advances in neural information processing systems*. 3882–3890.

[13] Andrew Y Ng, Stuart J Russell, et al. 2000. Algorithms for inverse reinforcement learning.. In *ICML*.

[14] Deepak Ramachandran and Eyal Amir. 2007. Bayesian Inverse Reinforcement Learning.. In *29th International Joint Conferences on Artificial Intelligence*, Vol. 7. 2586–2591.

[15] Bibhuti Sharma, Mark Hickman, and Neema Nassir. 2017. Park-and-ride lot choice model using random utility maximization and random regret minimization. *Transportation* (2017).

[16] Pengfei Wang, Yanjie Fu, Guannan Liu, Wenqing Hu, and Charu Aggarwal. 2017. Human mobility synchronization and trip purpose detection with mixture of hawkes processes. In *KDD*.

[17] Jiangchuan Zheng and Lionel M Ni. 2014. Modeling heterogeneous routing decisions in trajectories for driving experience learning. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 951–961.

[18] Brian D Ziebart, Andrew L Maas, J Andrew Bagnell, and Anind K Dey. 2008. Maximum Entropy Inverse Reinforcement Learning.. In *AAAI*.