Imitation Learning from Human-Generated Spatial-Temporal Data

Yanhua Li
Worcester Polytechnic Institute
yli15@wpi.edu

Weixiao Huang
Worcester Polytechnic Institute
whuang2@wpi.edu

ABSTRACT

Human dwellers make daily decisions by their own "strategies" governing their mobility dynamics (e.g., Uber drivers have preferred working regions and times, and urban commuters have preferred routes and transit modes). Understanding and characterizing the unique decision-making strategies of human agents has great potential in promoting their individual well-being. In this paper, we outline a novel spatial-temporal imitation learning (STIL) framework that defines, investigates, and addresses the emerging research challenges of analyzing and learning human decision-making strategies from human-generated spatial-temporal data. We present the state-of-the-art imitation learning algorithms, and the limitations of these algorithms in analyzing human-generated spatial-temporal data. Moreover, we present our preliminary studies, and outline the challenging open questions.

CCS CONCEPTS

• **Information systems** → *Spatial-temporal systems.*

KEYWORDS

human-generated spatial-temporal data, imitation learning, inverse reinforcement learning

ACM Reference Format:
Yanhua Li and Weixiao Huang. 2019. Imitation Learning from Human-Generated Spatial-Temporal Data. In *3rd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery (GeoAI'19), November 5, 2019, Chicago, IL, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3356471.3365229

1 INTRODUCTION

With the fast development of mobile sensing and information technology, large volumes of *human-generated spatio-temporal data (HSTD)* are increasingly collected, (e.g., GPS trajectories from taxis and personal vehicles, passenger trip data from automated fare collection (AFC) devices on buses and subway trains, and working traces from the emerging gig-economy services, such as food delivery (DoorDash [4], Postmates [12]), and everyday tasks (TaskRabbit [15])). Such HSTD capture unique decision-making characteristics of the "data generators" (e.g., gig-workers, and bus passengers), for examples, preferred working regions and time of taxi drivers and gig-workers, favored routes and transit mode of urban travellers, etc. Analyzing HSTD allows us to better understand and dissect how the experienced vs unskilled human agents make decisions under various circumstances, thus, enabling targeted training and incentive mechanisms to promote well-beings of urban dwellers and societies (in income levels, travel and living convenience, etc). In this paper, we outline a **spatial-temporal imitation learning** (STIL) framework that defines, investigates, and addresses the emerging research challenges of analyzing and learning human decision-making strategies from HSTD. As shown in Fig 1, the spatial-temporal imitation learning framework takes demonstration data from a group of (likely interactive) experts as input, and outputs the policies and reward functions employed by the group of expert demonstrators.

GeoAI'19, November 5, 2019, Chicago, IL, USA

ACM ISBN 978-1-4503-6957-2/19/11…$15.00
https://doi.org/10.1145/3356471.3365229

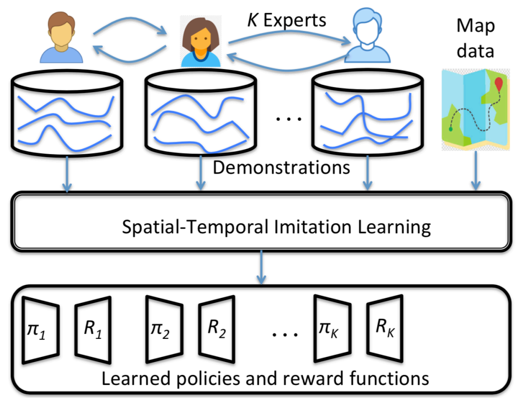

Figure 1: Spatial-temporal imitation learning framework

2 APPLICATION SCENARIOS

Consider taxi services. As a human agent, each taxi driver generates driving trajectories over time which capture her unique strategy of where to find the next passenger. Each taxi driver may only prefer working in certain regions in the city. As a result, passenger demands in some (rural) areas may be significantly under supplied. With the decision-making strategies learned from the drivers' trajectories, service provider can offer targeted incentives to motivate some drivers to cover those under-served areas [21]. More application scenarios include the food delivery service (e.g., DoorDash), and everyday task service (e.g., TaskRabbit), where gig-workers are human-agents making decisions based on their own preferences (or inherent "reward function"). *The goal of spatial-temporal imitation learning is to accurately infer such decision-making preferences from HSTD (e.g., generated by gig-workers).*

3 SPATIAL-TEMPORAL IMITATION LEARNING: STATE-OF-THE-ART WORK AND OPEN CHALLENGES

Imitation learning has been extensively studied to inversely learn human demonstrators' decision-making strategies [1, 3, 7, 14, 16, 18, 19, 23, 24], which is an effective computational tool playing essential roles in a variety of applications, including autonomous vehicle and robot control [5, 6], human motion analysis [8, 9], etc. When performing imitation learning over HSTD (in short, spatial-temporal imitation learning), it leads to unique applications, such as traffic estimation and prediction [13, 25], human mobility and intention analysis [11, 17], urban planning [19, 20], and incentive mechanism design [2, 21].

The vast majority of the existing work on imitation learning is essentially designed for general demonstration data from human agents [1, 3, 7, 18, 19, 24], with strong assumptions of invariant decision-making strategies over time and space, small scale scenarios with good spatial-temporal coverage, and optimal decision-making strategies used by human agents. Below, we briefly introduce our preliminary studies to address these challenges, and outline the open challenges to be investigated.

3.1 Preliminary investigations.

In [10, 11], we employ relative entropy inverse reinforcement learning [3] to study the diverse decision-making preferences of taxi drivers, when they look for passengers. In [20], we model the transit mode choice and transit stop selection problem of urban travelers as a Markov decision process (MDP), and inversely learn the passengers' unique decision-making strategy using maximum-entropy inverse reinforcement learning approach [24]. In [22], we develop a conditional generative adversarial imitation learning framework (cGAIL) to address the spatial dynamics of human agents' reward function and the spatial sparsity of collected HSTD. Moreover, in [21], we further investigate how to utilize the decision-making preferences learned from urban decision-makers to improve their decision-making strategies, by developing a targeted incentive design mechanism.

3.2 Open Challenges.

Beyond our preliminary investigations, there are still many crucial and open challenges in conducting preference learning from HSTD listed below.

- **Interactions among agents.** Human agents (e.g., in urban area) are not making decisions independently. Instead, they are making interactive decisions, and their decisions influence each other, (e.g., traffic slows down on a route chosen by more drivers, which affects other drivers choices to avoid the slowdown). It is challenging that how to model such interactions as a multi-agent game, and how to develop game-theoretical approach to solve the preference learning problem.
- **Scalable and online algorithm design.** The real world inverse preference learning problem usually involves a large space of states and actions, (e.g., the locations an urban traveler may visit and the transit routes/stops a traveler may choose). It is thus challenging to develop imitation learning approach that scales well, while learning the decision-making preferences in an online and incremental fashion.

ACKNOWLEDGMENTS

Yanhua Li and Weixiao Huang were supported in part by NSF grants CNS-1657350 and CMMI-1831140, and a research grant from DiDi Chuxing Inc.

REFERENCES

[1] Pieter Abbeel and Andrew Y Ng. 2004. Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the twenty-first international conference on Machine learning*. ACM, 1.
[2] Kent C Berridge. 2000. Reward learning: Reinforcement, incentives, and expectations. In *Psychology of learning and motivation*. Vol. 40. Elsevier, 223–278.
[3] 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.
[4] DoorDash. 2019. DoorDash Services. https://www.doordash.com/.
[5] Chelsea Finn, Sergey Levine, and Pieter Abbeel. 2016. Guided cost learning: Deep inverse optimal control via policy optimization. In *ICML*.
[6] Justin Fu, Anoop Korattikara, Sergey Levine, and Sergio Guadarrama. 2019. From Language to Goals: Inverse Reinforcement Learning for Vision-Based Instruction Following. In *ICLR*.
[7] Jonathan Ho and Stefano Ermon. 2016. Generative adversarial imitation learning. In *NeurIPS*. 4565–4573.
[8] Kun Li and Joel W Burdick. 2017. Inverse Reinforcement Learning in Large State Spaces via Function Approximation. *arXiv preprint arXiv:1707.09394* (2017).
[9] Kun Li and Joel W Burdick. 2017. Meta inverse reinforcement learning via maximum reward sharing for human motion analysis. *arXiv preprint arXiv:1710.03592* (2017).
[10] Menghai Pan, Weixiao Huang, Yanhua Li, Xun Zhou, Zhenming Liu, Rui Song, Hui Lu, Zhihong Tian, and Jun Luo. 2019. DHPA: Dynamic Human Preference Analytics Framework – A Case Study on Taxi Drivers' Learning Curve Analysis. *ACM Transactions on Intelligent Systems and Technology (TIST)* (2019).
[11] Menghai Pan, Yanhua Li, Xun Zhou, Zhenming Liu, Rui Song, Hui Lu, and Jun Luo. 2019. Dissecting the Learning Curve of Taxi Drivers: A Data-Driven Approach. In *SDM*. SIAM.
[12] Postmates. 2019. Postmates Services. https://postmates.com/.
[13] Masamichi Shimosaka, Junichi Sato, Kazuhito Takenaka, and Kentarou Hitomi. 2017. Fast inverse reinforcement learning with interval consistent graph for driving behavior prediction. In *AAAI*.
[14] Umar Syed and Robert E Schapire. 2008. A game-theoretic approach to apprenticeship learning. In *Advances in neural information processing systems*. 1449–1456.
[15] TaskRabbit. 2019. TaskRabbit Services. https://www.taskrabbit.com/.
[16] Faraz Torabi, Garrett Warnell, and Peter Stone. 2018. Behavioral cloning from observation. *arXiv preprint arXiv:1805.01954* (2018).
[17] Mohamed Wahba and Amer Shalaby. 2009. Learning-based departure time and path choice modelling for transit assignment under information provision: a theoretical framework. In *12th international conference on travel behaviour research, Jaipur, Rajasthan, India*.
[18] Kevin Waugh, Brian D Ziebart, and J Andrew Bagnell. 2013. Computational rationalization: The inverse equilibrium problem. *arXiv preprint arXiv:1308.3506* (2013).
[19] Guojun Wu, Yichen Ding, Yanhua Li, Jun Luo, Fan Zhang, and Jie Fu. [n. d.]. Data-Driven Inverse Learning of Passenger Preferences in Urban Public Transits. In *CDC 2017*. IEEE.
[20] Guojun Wu, Yanhua Li, Jie Bao, Yu Zheng, Jieping Ye, and Jun Luo. 2018. Human-Centric Urban Transit Evaluation and Planning. In *ICDM*.
[21] Guojun Wu, Yanhua Li, and Jun Luo. 2019. Transforming Policy via Reward Advancement. In *The 58th IEEE Conference on Decision and Control (CDC 2019)*. IEEE, 1–6.
[22] Xin Zhang, Yanhua Li, Xun Zhou, and Jun Luo. 2019. Unveiling Taxi Drivers' Strategies via cGAIL–Conditional Generative Adversarial Imitation Learning. In *2019 IEEE International Conference on Data Mining (ICDM)*. IEEE.
[23] Brian D Ziebart, J Andrew Bagnell, and Anind K Dey. 2011. Maximum causal entropy correlated equilibria for Markov games. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*. International Foundation for Autonomous Agents and Multiagent Systems, 207–214.
[24] Brian D Ziebart, Andrew L Maas, J Andrew Bagnell, and Anind K Dey. 2008. Maximum Entropy Inverse Reinforcement Learning.. In *AAAI*, Vol. 8. Chicago, IL, USA, 1433–1438.
[25] Brian D Ziebart, Andrew L Maas, J Andrew Bagnell, and Anind K Dey. 2009. Human Behavior Modeling with Maximum Entropy Inverse Optimal Control.. In *AAAI*. 92.