Large Language Models for Travel Behavior Prediction

B. Mo[a], H. Xu[b,*], J.-H. Cho[a], D. Zhuang[a], R. Ma[c], X. Guo[a] and J. Zhao[b]

[a] Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
baichuan@mit.edu, jhooncho@mit.edu, dingyi@mit.edu, xtguo@mit.edu

[b] Department of Urban Studies and Planning, Massachusetts Institute of Technology, Cambridge, MA 20139, USA
hanyongx@mit.edu, jinhua@mit.edu

[c] Department of Management Science and Engineering, Stanford University, Stanford, CA 94305, USA
ruoyunma@alumni.stanford.edu

[*] Corresponding author

Extended abstract submitted for presentation at the Conference in Emerging Technologies in Transportation Systems (TRC-30)
September 02-03, 2024, Crete, Greece

April 28, 2024

Keywords: Large language model; Travel behavior prediction; Mode choice; Trip purpose

1 INTRODUCTION

Understanding human travel behavior has always been a crucial part of transportation planning. Travel behavior can be described by many attributes, such as the mode of transportation, the purpose of the travel, the choice of destination, and the time of departure. Conventional travel behavior prediction relies on using numerical data to construct a mathematical model to represent human preferences (Ben-Akiva & Lerman, 1985, Mo *et al.*, 2021, Wang *et al.*, 2021).

Since last year, large language models (LLMs) have generated a tremendous amount of excitement in research due to their advanced language interaction and interpretation capabilities (Zhao *et al.*, 2023). In transportation, studies have shown the abilities of GPT in making mobility predictions which outperforms state-of-the-art time-series machine learning models (Wang *et al.*, 2023b). However, whether LLMs can predict travel behavior well is unclear.

In response to the above inquiry, this research aims to analyze the ability of LLMs to make travel behavior predictions. Through prompt engineering, we establish a zero-shot prompting framework that includes the prediction task, travel characteristics, passenger attributes, and guidance to LLM with transportation domain knowledge. The performance of the framework is tested and compared with classical models such as multinomial logit, random forest, and neural networks.

There are three major contributions: 1) The research presents a framework for predicting travel behavior using LLMs. To the authors' knowledge, this is the first effort to harness the semantic power of LLMs for travel behavior prediction. 2) The research demonstrates the proposed framework's adaptability to new scenarios and effectiveness in cold start situations through two real-world case studies, travel mode choice and trip purpose predictions. Notably, the proposed framework does not require any training samples yet produces competitive results when compared to classical benchmark models. 3) The research proposes a paradigm shift in travel forecasting, moving from reliance on numerical data to leveraging the reasoning abilities of LLMs for ease of understanding the output.

2 METHODOLOGY

Preliminaries. In this study, we introduce the pre-training based on GPT's framework for unsupervised multitask learning (Radford *et al.*, 2019). The objective of pre-training can be expressed as:

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}} \sum_{u\in\mathcal{U}} \sum_{i=1}^{n} \log \mathbb{P}\left(w_i^{(u)} \mid w_1^{(u)}, w_2^{(u)}, \dots, w_{i-1}^{(u)}; \boldsymbol{\theta}\right) \tag{1}$$

where $\mathcal{U}$ is the set of all training corpus, $w_i^{(u)}$ is the i-th token of the u-th sequence. Given the trained parameter $\boldsymbol{\theta}^*$, we can use the model to generate answers for various tasks:

$$s^* = \arg\max_{s} \mathbb{P}(s \mid (\text{Input}, \text{Task}); \boldsymbol{\theta}^*) \tag{2}$$

where s^* is the output sequence with the largest probability (or relatively large probability depending on the searching algorithm and degree of randomness). s^* is generated word by word until the "[End]" token is found.

Framework for understanding travel behavior. Given LLMs provide a generalized multitask solver, it is possible to use them as a predictor for people's travel behavior prediction. The "(Input, Task)" in Eq. 2 is referred to as **prompt** in LLMs. The framework of the LLM-based travel behavior prediction is shown in Figure 1. Here we use the travel mode choice and trip purpose prediction to illustrate. The input information will be organized and embedded into the prompt (i.e., prompt design). The prompt is fed into an LLM and outputs the prediction and associated reasons. It is worth noting that no supervised training step is needed.

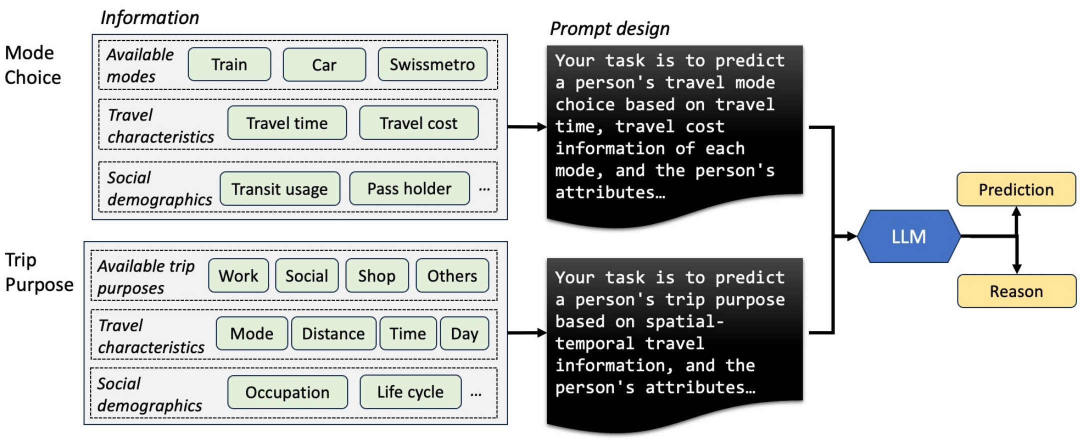

Figure 1 – *Conceptual framework*

Prompt design. Building upon existing prompting strategies, we carefully develop context-inclusive prompts to enhance travel behavior prediction, including data input, semantic guidance, and output specification. The data input includes three sub-components: 1) task descriptions, 2) structural data of travel characteristics, 3) descriptive data of individual attributes. The semantic guidance leads the LLM thinking with domain knowledge, numerical comparison, and strategies like Chain-of-Thought (Wei *et al.*, 2022) and Plan-to-Solve (Wang *et al.*, 2023a). The last part specifies the output with both the predictions and reasoning, which will help us understand the output, improve prompt design, and help LLM to improve its performance. An example of the final prompt for mode choice prediction is shown in Figure 2. The text in square brackets is for reference and not included in the actual prompt.

Example 1: Travel mode choice

[Task description] Your task is to predict a person's travel mode choice from Train, Car, and Swissmetro based on travel time, travel cost information of each mode, and the person's attributes.

[Structure data] The travel time and cost for each mode is expressed as the following dictionary format: {Travel time: {Train: 202, Car: 160, Swissmetro: 97}, Travel cost: {Train: 108, Car: 136, Swissmetro: 140}}

[Thinking guidance with domain knowledge] Swissmetro has the lowest travel time. Choosing it will save 39% travel time compared to Car and save 52% travel time compared to Train. Train has the lowest travel cost. Choosing it will save 21% travel cost compared to Car and save 23% travel cost compared to Swissmetro.

[Descriptive data] The person is not a regular Train user. He/She does not own the Train annual pass.

[Output specification] Please infer what is the mostly likely travel mode that the person will choose. Organize your answer in a JSON object with two keys: "prediction" (the predicted travel mode) and "reason" (explanation that supports your inference).

[Thinking guidance with domain knowledge] Please consider the following aspects:

1. People are more likely to choose a travel mode with less travel cost and travel time, especially those with significant cost or time saving. The trade off between time and cost can be quantified using value of time.
2. Regular Train users may prefer to use Train.
3. Owners of Train annual pass are more likely to choose Train.

Figure 2 – *Example complete prompts for travel behavior prediction*

3 RESULTS

Data. Mode choice prediction is assessed on the Swissmetro stated preference survey dataset, where users select from train, car, and Swissmetro given the corresponding travel attributes. Trip purpose prediction is evaluated using the 2017 US National Household Travel Survey, which tracks daily trip, socio-demographic features, and trip purposes like working, social, shopping, and others. We use balanced sampling to randomly select 1,000 entries for the large training set, 10 for the small training set (training sets are for benchmark models only), and 200 for the testing dataset. Results are derived from the average of 5 random samples.

Baselines. The LLM-based prediction model is compared against three benchmark models: 1) multinomial logit (MNL), 2) random forest (RF), and 3) neural networks (NN). MNL is the canonical travel behavior prediction tool, while random forest and NNs are proven to perform well empirically (Wang *et al.*, 2021).

Experimental Setup. The LLM employed is GPT-3.5 (version gpt-3.5-turbo-1106), which is one of the most advanced and widely used LLMs with open APIs. The prompts used follow the structure of Figure 2. We set the temperature to 0 to avoid randomness in the output. The benchmark models are implemented using the Python scikit-learn package with the default hyper-parameters.

Travel Mode Choice and Trip Purpose Estimations. All models are evaluated based on prediction accuracy and weighted F1-score. Results are shown in Table 1, with the best results highlighted in boldness in each training set size. Without using any training samples, the LLM-based prediction has shown competitive performance compared to benchmark models trained on large training sets and outperforms them when their training sample size is limited.

Table 1 – *Combined Model Prediction Results*

Model	Training Set Size	**Mode Choice**		**Trip Purpose**	
		Accuracy	F1 Score	Accuracy	F1 Score
MNL	large	0.604 ± 0.037	0.595 ± 0.034	0.418 ± 0.043	0.401 ± 0.043
RF	large	0.617 ± 0.025	0.613 ± 0.026	0.459 ± 0.046	0.451 ± 0.047
NN	large	$\mathbf{0.641 \pm 0.030}$	$\mathbf{0.637 \pm 0.028}$	$\mathbf{0.463 \pm 0.047}$	$\mathbf{0.455 \pm 0.049}$
MNL	small	$\mathbf{0.469 \pm 0.075}$	$\mathbf{0.457 \pm 0.080}$	0.284 ± 0.058	0.275 ± 0.052
RF	small	0.455 ± 0.068	0.441 ± 0.074	$\mathbf{0.338 \pm 0.030}$	$\mathbf{0.329 \pm 0.038}$
NN	small	0.462 ± 0.044	0.455 ± 0.046	0.322 ± 0.077	0.311 ± 0.065
LLM	zero-shot	$\mathbf{0.590 \pm 0.051}$	$\mathbf{0.580 \pm 0.054}$	$\mathbf{0.393 \pm 0.020}$	$\mathbf{0.332 \pm 0.019}$

Table 2 – *Example case of using LLMs for travel mode choice prediction*

Input:	Travel time: Train: 230, Car: 247, Swissmetro: 111; Travel cost: Train: 56, Car: 169, Swissmetro: 66. The person is a regular Train user. He/She does not own the Train annual pass.
Prediction and Actual Labels:	Train (MNL prediction is Swissmetro) vs. Train
Reason:	The person is a regular Train user and does not own the Train annual pass. Although Swissmetro has the lowest travel time, the person's regular use of Train suggests a preference for this mode of transportation. Additionally, Train has the lowest travel cost, which may also influence the decision.

Reasoning Ability. LLMs offer a distinctive advantage by articulating the reasoning behind their predictions. As demonstrated in Table 2, the LLM not only accurately estimates the individual's travel mode preference but also provides a logical explanation including regular usage of travel modes and their cost efficiency, an aspect where traditional models like MNL may not perform as effectively.

4 DISCUSSION

In this study, we show the potential of using LLM for the complex travel behavior prediction task without training samples. The performance of LLM-based predictions is competitive with traditional models like MNL, RF, and NNs, especially when training data is scarce. This makes LLMs ideal for rapid adaptations to new scenarios. Besides, LLM can also output explanations for inference despite some reasoning errors and hallucination problems.

This research presents a preliminary result, there is a lot of ongoing and future work. 1) Ablation studies can be conducted to test the effectiveness of different components in the prompt design. 2) The current prompt is a zero-shot prompt. Future studies can include some representative samples in the training data set as a few-shot prompt for in-context learning. The few-shot prompt needs to guide the LLMs to learn to quantify the value of time (or other important concepts for prediction) using the limited samples. 3) More experiments can be conducted with other prompting strategies and variants of LLMs.

References

Ben-Akiva, Moshe, & Lerman, Steven R. 1985. *Discrete choice analysis: Theory and application to travel demand.* Cambridge, MA: MIT Press.

Mo, Baichuan, Wang, Qing Yi, Moody, Joanna, Shen, Yu, & Zhao, Jinhua. 2021. Impacts of subjective evaluations and inertia from existing travel modes on adoption of autonomous mobility-on-demand. *Transportation Research Part C: Emerging Technologies*, **130**, 103281.

Radford, Alec, Wu, Jeffrey, Child, Rewon, Luan, David, Amodei, Dario, Sutskever, Ilya, *et al.* 2019. Language models are unsupervised multitask learners. *OpenAI blog*, **1**(8), 9.

Wang, Lei, Xu, Wanyu, Lan, Yihuai, Hu, Zhiqiang, Lan, Yunshi, Lee, Roy Ka-Wei, & Lim, Ee-Peng. 2023a. Plan-and-Solve Prompting: Improving Zero-Shot Chain-of-Thought Reasoning by Large Language Models. *Page 2609–2634 of:* Rogers, Anna, Boyd-Graber, Jordan, & Okazaki, Naoaki (eds), *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Toronto, Canada: Association for Computational Linguistics.

Wang, Shenhao, Mo, Baichuan, Hess, Stephane, & Zhao, Jinhua. 2021. Comparing hundreds of machine learning classifiers and discrete choice models in predicting travel behavior: an empirical benchmark. *arXiv preprint arXiv:2102.01130.*

Wang, Xinglei, Fang, Meng, Zeng, Zichao, & Cheng, Tao. 2023b. Where Would I Go Next? Large Language Models as Human Mobility Predictors. Aug. arXiv:2308.15197 [physics].

Wei, Jason, Tay, Yi, Bommasani, Rishi, Raffel, Colin, Zoph, Barret, Borgeaud, Sebastian, Yogatama, Dani, Bosma, Maarten, Zhou, Denny, Metzler, Donald, Chi, Ed H., Hashimoto, Tatsunori, Vinyals, Oriol, Liang, Percy, Dean, Jeff, & Fedus, William. 2022. Emergent Abilities of Large Language Models. Oct. arXiv:2206.07682 [cs].

Zhao, Wayne Xin, Zhou, Kun, Li, Junyi, Tang, Tianyi, Wang, Xiaolei, Hou, Yupeng, Min, Yingqian, Zhang, Beichen, Zhang, Junjie, Dong, Zican, Du, Yifan, Yang, Chen, Chen, Yushuo, Chen, Zhipeng, Jiang, Jinhao, Ren, Ruiyang, Li, Yifan, Tang, Xinyu, Liu, Zikang, Liu, Peiyu, Nie, Jian-Yun, & Wen, Ji-Rong. 2023. A Survey of Large Language Models. June. arXiv:2303.18223 [cs].