Construction and Application of Urban Pseudo People Flow Generation Model Based on Large Language Model

**Kunyi Zhang, Yurong Zhang, Yanbo Pang, Yoshihide Sekimoto**

Background

Human mobility simulation plays a crucial role in urban transportation applications. Recently, people flow simulation frameworks based on Large Language Models (LLM) have been gradually noticed and studied. However, the effective integration of massive individual information with spatial trajectories into such GPT-based models remains challenging. To address the challenge, we propose Mobility Generative Language Model (MobGLM) that enables us to fully leverage transformer-based language models to capture the relationship between agents' mobility patterns and individual attributes.

Objectives

I. Develop a generative AI to forecast human mobility based on LLM of GPT2 using Person Trip survey dataset.
II. Address the long-tailed distribution of sparse non-daily activities based on logits adjustment of SoftMax.

Methodology

$$g_y(x) = f_y(x) + \tau \cdot \log \pi_y$$

Figure 1: The overview of MobGLM.
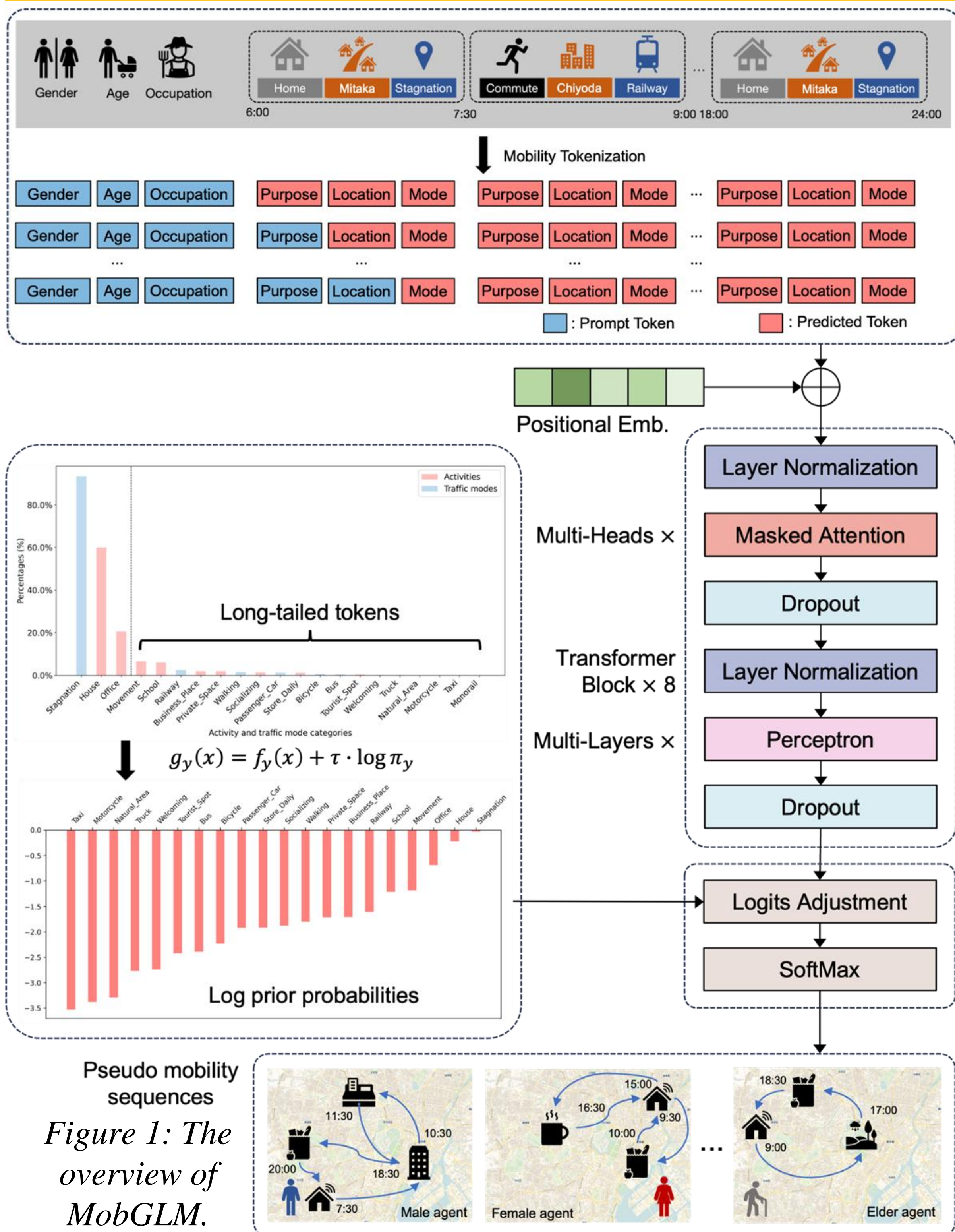
we tokenized over 570000 users' individual attributes, activities, locations and traffic modes in Tokyo Metropolitan from PT survey to generate pseudo mobility sequences. Logits adjustment of activities and traffic modes are introduced with τ as 5 to reduce the generation of major activities.

Results

Table 1: Comparison of performance between baseline models and proposed MobGenAI.

Model	TMR	LCSS	BLEU	ROUGE-L	DTW
3rd Order MM	0.7912	0.5453	0.6995	0.6762	10338.257
LSTM	0.8552	0.4491	0.8345	0.7718	9123.5013
GRU	0.8560	0.4526	0.8357	0.7769	9071.6620
Transformer XL	0.8309	0.5931	0.8573	0.8006	8728.5932
MobGenAI w/o LA	0.8855	0.6532	0.8962	0.8459	8410.8002
MobGenAI	0.8836	0.6577	0.8955	0.8462	8410.76

12:00 – 16:00 *16:00 – 20:00* *20:00 – 24:00*

(a) Generated business activity of MobGenAI

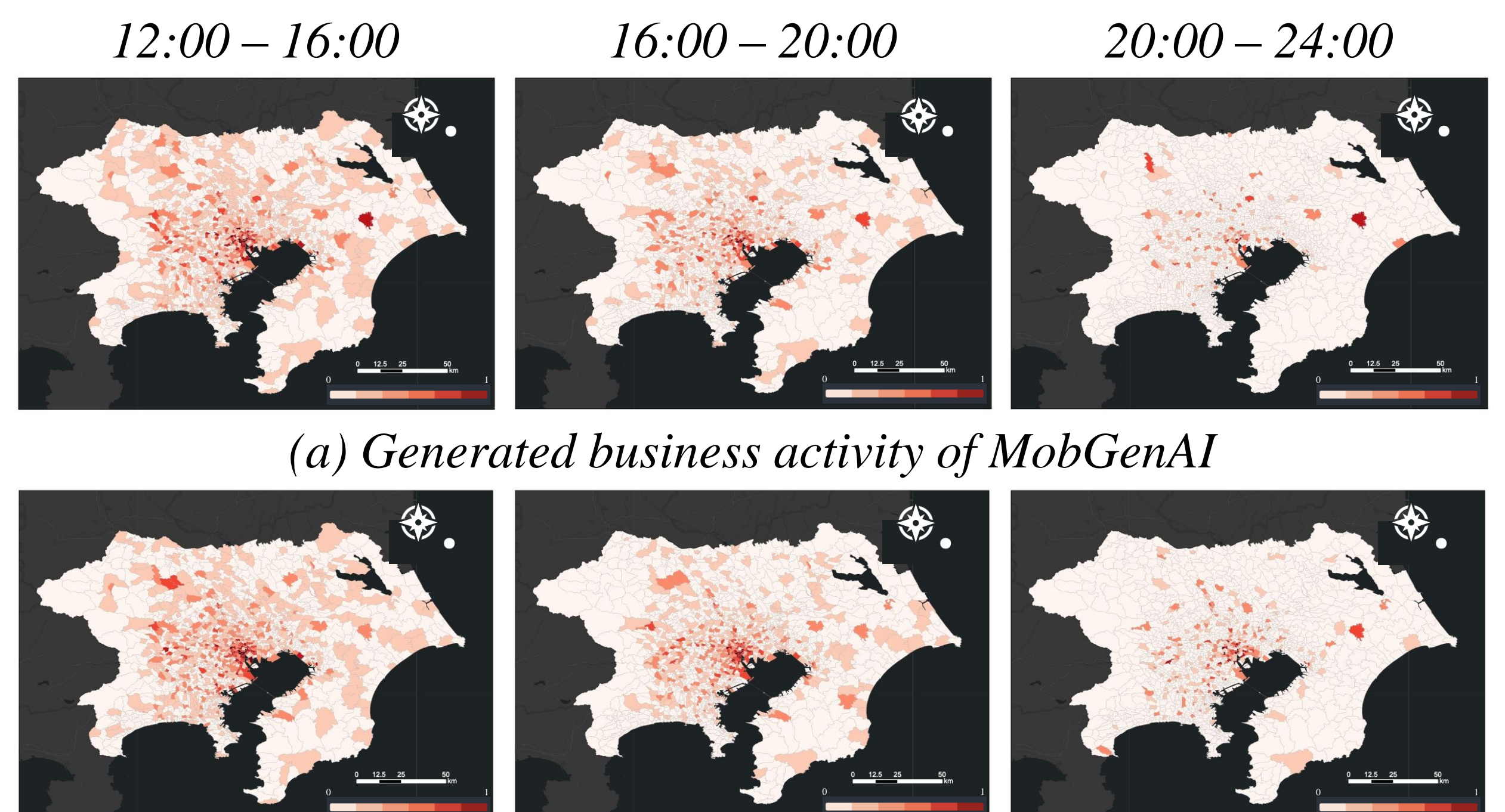(b) Ground truth business activity of PT survey

Figure 2: Comparison of the spatial distribution between generated mobility and ground truth.
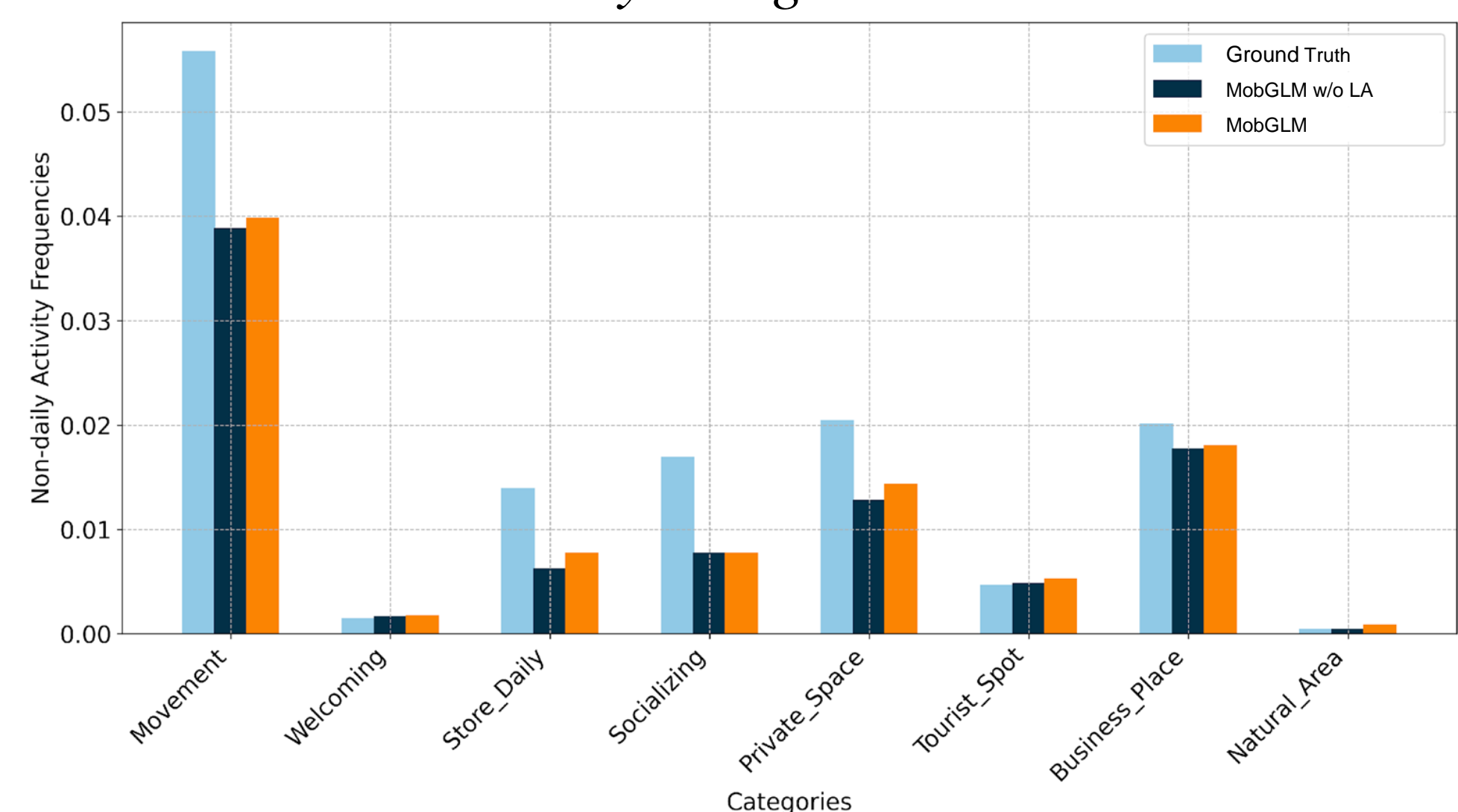

Figure 3: Daily activity frequencies comparison.

Conclusion

We proposed a novel generative model for the sequence forecast of regional human mobility based on the optimized GPT2. Our study clarified the possibility of generating pseudo-people flow and reproducing the real life of individuals.