Learning Predictive Representations of Human Mobility Flows

Sid Nguyen, Rongchao Xu, Guang Wang

1. Introduction

Human movement through cities appears complex and noisy. However, large-scale mobility data consistently reveals recurring patterns in how people transition between locations.

Understanding these patterns is important for:

Urban transportation planning

Traffic forecasting

Resource allocation

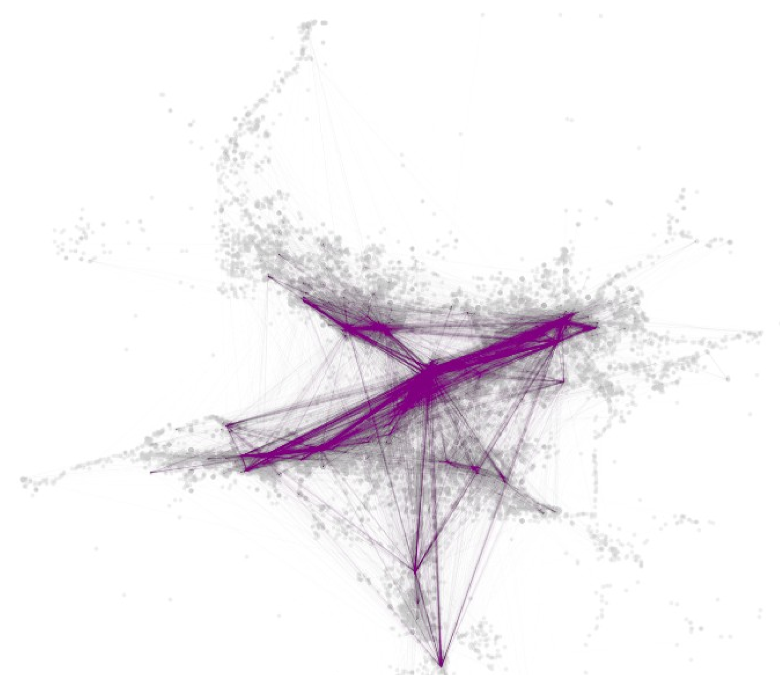
Location recommendation systems

Infrastructure design

Most traditional prediction systems estimate a single "most likely" next location.

However, real-world movement is inherently uncertain — multiple future paths are often plausible.

This project explores wether we can model human mobility as probabilistic flows between regions over time. Rather than predicting one outcome, we aim to capture the distribution of possible next movements.



2. Methods

Research on mobility prediction spans several approaches:

- Markov Models

Early work modeled movement as state transitions where the next location depends only on the current one.

Simple and interpretable but has limited memory of longer movement history.

- Sequence Modeling (RNNs, LSTMs, Transformers)

Learn dependencies across multiple past steps to capture temporal dynamics

Demonstrates that human movement contains predictable sequential structure.

However, many approaches focus on a single most likely outcome and uncertainty is often not explicitly modeled.

- Uncertainty-Aware & Generative Models

Human movement is not deterministic so multiple future states may be plausible.

Modeling probability distributions is more realistic than single-point estimates.

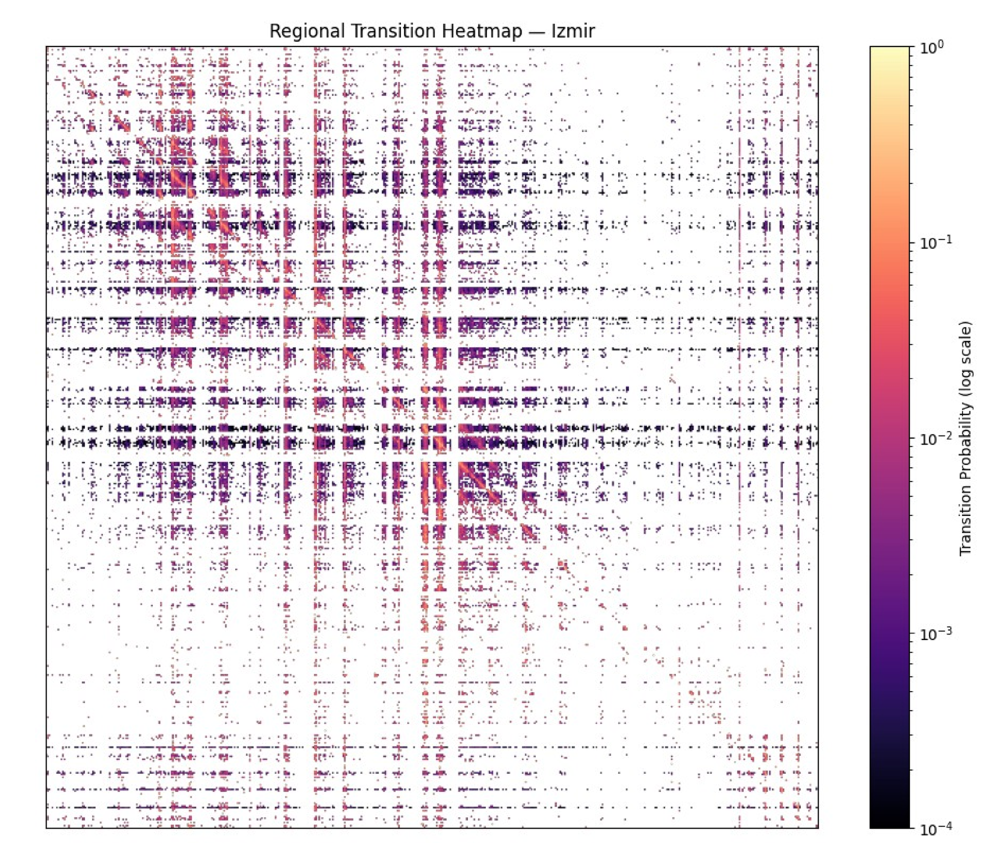
These approaches align more closely with real-world planning needs.

Instead of isolated events, we study ordered movement histories mapped to Region-to-region transitions to visualize the temporal flow patterns

Preliminary analysis reveals:

- Strong recurrence in specific transitions
- Temporal clustering (e.g., daily rhythms)
- Long-tail distributions of rare transitions
- Evidence of structured, non-random mobility behavior

These observations suggest that predictive modeling is feasible.



The area-to-area first-order transition heat map of Izmir

We represent mobility as sequences of flows between regions.

- Convert movement data into ordered transition histories
- Learn a compact representation of recent movement
- Predict a probability distribution over next transitions

Rather than predicting:

Next location = A

We model:

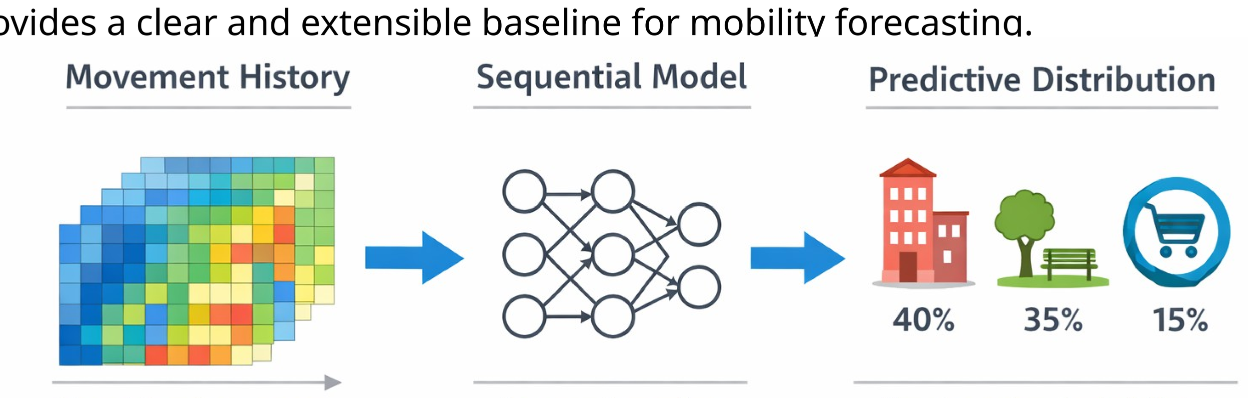
P(next location | history) = {A: 40%, B: 35%, C: 15%, …}

This better reflects the inherent uncertainty in human movement.

Importantly, we focus on:

- Lightweight, interpretable models
- Probabilistic outputs
- Flow-based representations

This provides a clear and extensible baseline for mobility forecasting.



Graphic representation of our process. Taking the movement flow, modeling, and prediction

3. Conclusion

This project explores human mobility as a structured yet inherently uncertain flow process. Through an extensive review of prior work and preliminary data analysis, we find strong evidence that large-scale movement data contains meaningful and recurring transition patterns. Even before full quantitative evaluation, exploratory analysis suggests that mobility behavior is far from random: transitions cluster in predictable ways, exhibit temporal rhythms, and follow heavy-tailed flow distributions.

Building on established sequence modeling techniques, we frame mobility prediction as a probabilistic task rather than a deterministic one. Instead of estimating a single most likely next destination, our approach models a distribution over possible future transitions. This perspective better reflects the reality of human movement, where multiple outcomes are often plausible. By emphasizing interpretable, lightweight sequential models and flow-based representations, this work establishes a clear and extensible baseline for uncertainty-aware mobility forecasting. It lays the conceptual and technical foundation for richer spatiotemporal modeling that bridges machine learning and real-world urban systems.

4. Future direction

Moving forward, several key improvements can significantly extend this framework.

First, expanding from individual check-in style data to aggregated traffic flow datasets will allow modeling at larger spatial and temporal scales. Incorporating richer contextual signals — such as time of day, seasonal variation, special events, and weather — can further improve predictive realism. These factors strongly influence movement patterns and may help models distinguish routine behavior from anomalies.

Second, integrating spatial structure directly into the modeling process is a promising direction. Graph-based representations of regions and their connectivity could better capture geographic relationships, enabling models to understand not just sequences of transitions but the spatial topology underlying them. This would strengthen applications in traffic forecasting and regional planning.

Third, enhancing uncertainty modeling remains central. More expressive probabilistic or generative approaches could simulate multiple plausible future mobility scenarios rather than producing a single forecast. Such capabilities are especially valuable in urban planning, where decision-makers must account for variability and rare but impactful events.

Ultimately, this line of research supports broader real-world applications. Reliable probabilistic mobility forecasting can inform transportation system optimization, infrastructure development, emergency response planning, and intelligent routing systems. By improving how we represent and predict human movement flows, we move closer to building adaptive, data-driven urban environments that respond effectively to dynamic population behavior.

5. Resources

Chen, Wei, et al. "Trajectory Data Management and Mining: A Survey from Deep Learning to the LLM Era." ArXiv.org, 2024, arxiv.org/abs/2403.14151.

Guo, Baoshen, et al. "Language Models Meet Urban Mobility: A Data-Centric Review." TechRxiv, Dec. 2025, https://doi.org/10.36227/techrxiv.176703984.41856875/v1.

Liang, Yuebing, et al. "Exploring Large Language Models for Human Mobility Prediction under Public Events." Computers Environment and Urban Systems, vol. 112, Elsevier BV, Sept. 2024, pp. 102153–53, https://doi.org/10.1016/j.compenvurbsys.2024.102153.

Manvi, Rohin, et al. "Large Language Models Are Geographically Biased." ArXiv.org, 4 Feb. 2024, https://doi.org/10.48550/arXiv.2402.02680.

Zhang, Qianru, et al. "A Survey of Generative Techniques for Spatial-Temporal Data Mining." ArXiv.org, 2024, arxiv.org/abs/2405.09592.