Modeling and Predicting Agent Trajectory in Urban Road Networks

Julien Baudru[1][0000−0002−8771−2494] and Hugues Bersini[1][0000−0001−5820−7360]

Université Libre de Bruxelles (ULB), Brussels, Belgium
julien.baudru@ulb.be

1 Introduction

We address next-node prediction in Brussels' urban mobility. Using a custom simulator on the city's road network, agents generate daily trips via five source-target (s–t) models. We train LSTM, Transformer, and GNN models to evaluate their ability to learn spatio-temporal mobility patterns for applications such as transfer point estimation, dynamic car-pooling, or traffic forecasting.

2 Simulation and Prediction Framework

We simulate mobility on the Brussels road network $G = (V, E)$ with $|V| = 18{,}547$ intersections and $|E| = 40{,}890$ segments. Agents travel at edge-specific speeds affected by congestion, modeled via the Bureau of Public Roads (BPR) function [1], with a minimum speed of 10% of normal. Trip initiation follows a 24-hour pattern (6 AM–6 AM) with morning/evening peaks. Five s–t selection models reflect diverse behaviors: **Random** (uniform s–t), **Activity-Based** (home/work patterns by agent type), **Zone-Based** (3×3 grid, 60% intra-zone, 40% inter-zone [2]), **Gravity** (probability $\propto$ centrality / distance), and **Hub-and-Spoke** (70% via high-centrality hubs). We train three neural architectures for *next-node prediction*, i.e., given a fixed-length history of visited nodes, predict the next: **LSTM:** 3-layer bidirectional (512 units/dir) with 256-dim embeddings, dropout 0.3, softmax output. **Transformer:** 6-layer encoder, 8 heads, 512-dim embeddings with sinusoidal positions, causal masking, pre-norm. **GNN:** 4-layer GAT (4 heads) to learn 128-dim node embeddings, followed by 2-layer bidirectional LSTM (256 units/dir). All use cross-entropy loss with 0.1 label smoothing, AdamW with cosine annealing, early stopping ($patience = 10$), and gradient clipping (maxnorm = 1.0). Hyperparameters: LSTM ($batch = 64$, $lr = 0.001$), Transformer ($batch = 32$, $lr = 0.0001$), GNN ($batch = 48$, $lr = 0.0005$).

3 Results and Conclusion

The three prediction models were trained and evaluated independently using trip data from the five s–t selection models with 1000 agents simulated over a 24-hour period. Each training instance consists of an input sequence of 10 consecutive road network nodes and the corresponding next node as output. Dataset sizes are reported in Appendix 3. Table 1 reports the validation loss, the final validation and the test accuracies for each model on each dataset.

Table 1. Validation accuracy, train loss, and final validation loss for each model on five s-t selection datasets with 1000 agents over 24-hour simulation.

Dataset	**Model**	**Train Loss**	**Val. Loss**	**Val. Accuracy (%)**
Random	LSTM	1.4619	1.626	92.77
Random	Transformer	1.5143	1.6871	92.23
Random	GNN	1.5247	1.6635	92.36
Activity-Based	LSTM	1.4187	1.5357	93.97
Activity-Based	Transformer	1.4549	1.5745	93.59
Activity-Based	GNN	1.4673	1.562	93.71
Zone-Based	LSTM	1.4746	1.7235	90.88
Zone-Based	Transformer	0.1626	0.613	90.52
Zone-Based	GNN	0.8477	0.9382	85.94
Gravity	LSTM	1.4806	1.7403	90.74
Gravity	Transformer	1.512	1.7933	90.35
Gravity	GNN	1.5209	1.749	90.66
Hub-and-Spoke	LSTM	1.4559	1.6092	93.04
Hub-and-Spoke	Transformer	1.5023	1.6666	92.55
Hub-and-Spoke	GNN	1.513	1.6439	92.67

Across all models, the Activity and Hub-and-Spoke datasets are the most predictable, with validation accuracies between 92% and 94% and relatively low losses, suggesting their s–t patterns are easier to learn. In contrast, the Zone dataset proves most challenging, especially for GNNs, which drop to 85.94%, likely due to its smaller size and complex spatial structure. Interestingly, dataset size alone does not determine performance: the smaller Activity dataset consistently outperforms the much larger Random dataset ($\approx$ 92%), indicating that complexity and pattern regularity have greater influence than the sample size. Training losses are consistently below validation losses, showing mild overfitting, but the small gap suggests models retain good generalization.

Future work will enhance the predictive setting by incorporating richer node features, explicit spatial–temporal context, and real traffic conditions, aiming to improve accuracy in more complex scenarios like the Zone dataset. More robust evaluation through K-Fold cross-validation and model averaging is also planned, though these were not feasible here due to the computational demands of training.

Appendix

Table 2. Dataset sizes used for training, evaluation and testing

Dataset	**Train. Size (60%)**	**Val. Size (20%)**	**Test Size (20%)**
Random	348,885	116,295	116,294
Activity	330,516	110,171	110,171
Zone	252,849	84,282	84,282
Gravity	259,719	86,573	86,573
Hub-and-Spoke	342,494	114,164	114,164

References

1. Sven Maerivoet and Bart De Moor. Transportation planning and traffic flow models, 2005.
2. Luis M. Martínez, José Manuel Viegas, and Elisabete A. Silva. A traffic analysis zone definition: a new methodology and algorithm. *Transportation*, 36(5):581–599, 2009.