CSCI 5527 PROJECT PROGRESS REPORT: GENERATING MOBILITY FLOWS UNDER THE INFLUENCE OF WEATHER USING DEEP LEARNING MODELS

Zhongfu Ma, Yaxuan Zhang, Xiaohuan Zeng

1 WHAT PROBLEM?

Predicting human mobility flow is an important topic in transportation and urban planning fields to understand how humans move in the urban environment. The lack of historical mobility flow may require researchers to develop models to forecast flows considering the features of origins and destinations (e.g., demographic characteristics, POI, and land use). Compared to other statistical models or shallow neural networks, deep learning models can introduce nonlinearities and build more complex representations of the input geographic features by adding more hidden layers. More importantly, the potential spatial interaction between neighboring areas of interest may account for the social and urban configurations overlooked in the models. Therefore, our project aims to incorporate distance decay functions as gravity models into deep learning models to capture the unobserved geographic features and relationships/interactions between two areas of interest when predicting flows. Moreover, we will explore how the external factors (e.g., temperature and precipitation) in addition to social and urban configurations can influence the model accuracy for predicting human mobility flow.

2 WHY INTERESTING

This project is interesting for several reasons:

- It is relevant to transportation and urban planning. Predicting human mobility flow is a crucial aspect of transportation and urban planning, as it helps in understanding how people move within urban environments. This information is vital for designing efficient transportation systems and urban infrastructure.
- The use of deep learning can capture complex nonlinear relationships within data and create more intricate representations of geographic features.
- The project recognizes the importance of spatial interactions between neighboring areas, which can play a significant role in understanding human mobility patterns. This aspect can uncover social and urban configurations that might be overlooked by conventional deep learning models.
- Additionally, the project aims to consider external factors like temperature and precipitation, which can influence human mobility. We are interested to explore that mobility is affected not only by social and urban factors but also by environmental conditions.

3 PREVIOUS WORK

Flow generation problem (Luca et al., 2021) refers to generating the spatial interaction flow, such as the movement of individuals, goods, and currency, between geographic locations, taking into account the characteristics of each location—such as population density, presence of Points of Interest (POIs), land use patterns, and distance to other locations, without historical information of actual flow.

Flow generation has attracted interest for a long time. Among numerous flow generation models, the gravity model is one of the widely employed models, which draws an analogy with Newton's law of universal gravitation. This model is based on the assumption that the number of travelers between two locations (flow) increases with the locations' populations while decreasing with the distance between them (Barbosa et al., 2018). The radiation model is another branch of the flow generation models, which estimates mobility among locations as the number of travelers or trip counts that radiate from an origin i and are then absorbed by each destination j. The process by which absorption occurs is based solely on population distribution, without considering the distance. Stefanouli & Polyzos (2017) compared the radiation model with the traditional gravity model and found parameter-free radiation model gives more competitive results, especially for large scales.

With the development of the deep learning model, researchers got a chance to revisit the flow generation problem, uncovering the complex and nonlinear relationship between the characteristics of places and the flows. However, there is limited literature tackling flow generation. Simini et al. (2021) proposed the DeepGravity model to extend the gravity model by considering multiple attributes, such as land use, POIs, and so on, with the deep learning method. Based on the case study conducted in England, Italy, and New York State, the deep gravity model proposed in their research demonstrates a notable enhancement in the generation of mobility flows. The SI-GCN (Spatial Interaction Graph Convolutional Network) introduced by Yao et al. (2020) is a structured model comprising three core components:(i) A spatial representation layer that constructs local graphs, applies negative sampling and organizes features. (ii) An encoder employing graph convolutions to create latent space representations of geographic units. (iii) A decoder that predicts unobserved flows using these latent representations. SI-GCN's performance is validated on the T-Drive dataset using metrics such as RMSE, MAPE, and CPC to ensure accuracy in flow prediction. Liu et al. (2020) proposed GML (Geocontextual Multitask Embedding Learner) harnesses spatial correlations using geographic context and employs dual attention-based graph neural networks (GATs). These GATs are dedicated to deriving an embedding which is subsequently inputted into a gradient-boosting model responsible for flow generation. Evaluation of GML is conducted on New York City commuting data, utilizing MAE, RMSE, and CPC as metrics.

4 Project goal

This project belongs to the category of novel application and will explore the application of deep learning models in flow generation. The first goal is to implement a state-of-the-art neural network architecture with PyTorch to get decent performance in flow generation. The second goal is to investigate how spatial interaction attributes and external location features like weather, transportation infrastructure, and land use impact model performance.

5 Methods and Results

The code to reproduce our results can be found `https://github.com/ZhongfuMa/CSCI5527_final_project/tree/main`.

5.1 Data Preparation

5.1.1 Mobility Flow Data

A device-level mobile positioning dataset in the Twin Cities in July 2021 from PlaceIQ, a location data and technology company for place intelligence, is used in this project to create flow ground truth at the census track level. Each record is an event-based visitation cluster of the individual, which contains the device ID, coordinates, duration seconds, zip code, and Unix timestamp of visitation. To simplify the data preprocessing, this project assumes that every two consecutive visitations by the same device on the same day constitutes a flow count (Figure 1 (a)) and we sum all the flow counts from one 3km cell to the other (Figure 1 (b)). The mobility flow data from July 1st to 8th is the training set, and the data on July 9th is the test set.

5.1.2 External Weather Data

The daily temperature (temp_min and temp_max) and precipitation data are pulled from Iowa State University The Iowa Environmental Mesonet (IEM) API (https://mesonet.agron.iastate.edu/). The weather data is originally formatted as GeoJSON with property and geometry. The weather data

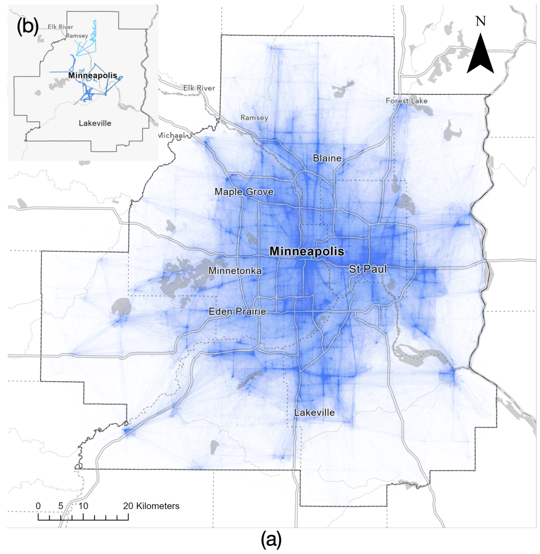

Figure 1: The device-level mobility data and spatial grids in our study area. (a) The individual trips were constructed based on visit records on July 17. (b) Trajectories of three sampled individuals on July 17 with multiple trips throughout the day.

were measured at various station points across Minnesota. To get weather data at unknown locations, spatial interpolation is required. In this project, we decided to use the Inverse Distance Weighting (IDW) method to keep consistency with the deep gravity model. IDW is a spatial interpolation method used to estimate values at unmeasured locations based on surrounding sampled points. The basic idea is to assign weights to known points based on their distance to the target location, with closer points receiving higher weights. The weighted values are then combined to estimate the unknown value at the target location. The IDW formula is given by:

$$Z_0 = \frac{\sum_{i=1}^{n} \frac{z_i}{d_i^p}}{\sum_{i=1}^{n} \frac{1}{d_i^p}}$$

Where Z_0 is the estimated value at the target location, Z_i is the known value at the ith sampled point, d_i is the distance between the target location and the ith point, p is a user-defined power parameter that controls the influence of distance on the weights, and n is the number of sampled points. In this project, the fishnet/grid data at a 3 km resolution are used to represent the geographic area. The fishnet data is a regular arrangement of rectangular or square cells overlaid across the Minnesota Metropolitan 7-county area. The interpolated daily weather data are then assigned to each grid. (Figure 2 shows an example during the interpolation process. (Figure 2 (a) shows the interpolated maximum temperature on July 1st, 2021 and (Figure 2 (b) shows the assigned interpolated value for the fishnet data. The final weather data is formatted as temperature and precipitation values with the latitude and longitude of each grid centroid and the date in July 2021.

5.2 Graviity Model

The gravity model is a commonly used method in transportation planning and geography to predict the flow amount between two locations. This project will use the gravity model as a benchmark to compare with the deep learning model. As an analogy to the principle of gravity in physics, the assumption in flow generation is that the flow amount $F_{i,j}$ will decrease as distance $Dist_{i,j}$ increases. The flow amount will increase if the population at the origin Pop_i or destination Pop_j increases:

$$F_{i,j} = \frac{K \cdot {Pop_i}^{\alpha} \cdot Pop_j \beta}{{Dist_{i,j}}^{\gamma}}$$

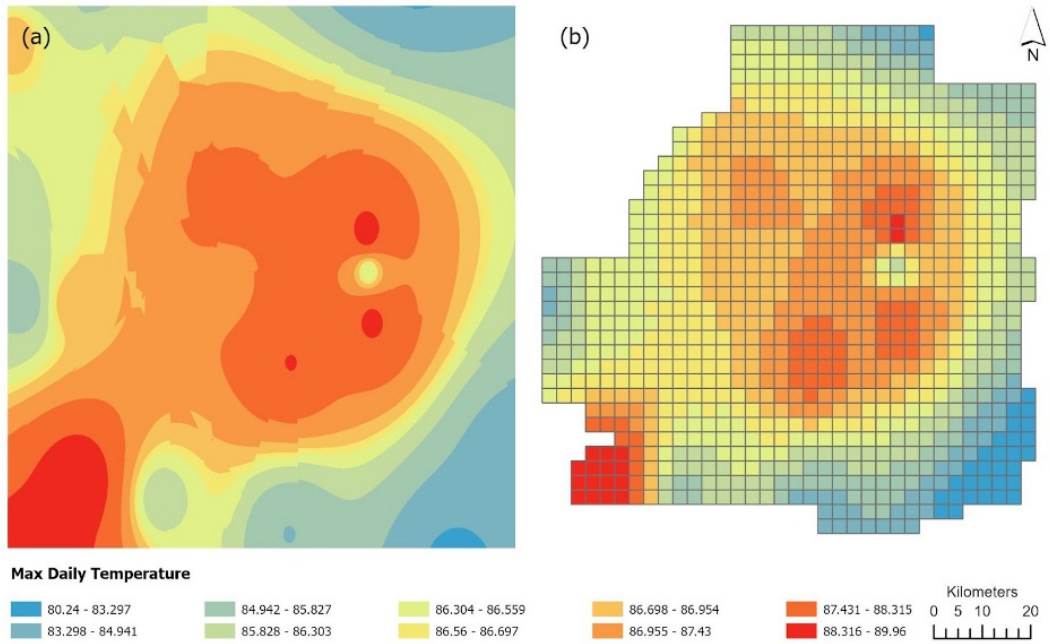

Figure 2: Weather Data (a) Interpolation Plot (b) Assigned Temperature Value

After calibrating the gravity model with non-linear least squares, we obtain the parameters that best fit our data: $\alpha = 0.62$, $= 0.62$, $\gamma = 2.89$, and $K = 0.53$. Figure 3 displays the observed flows in our data and predicted flows using the calibrated gravity model. It seems that the gravity model cannot identify the flows that occur at the boundary of the study area.

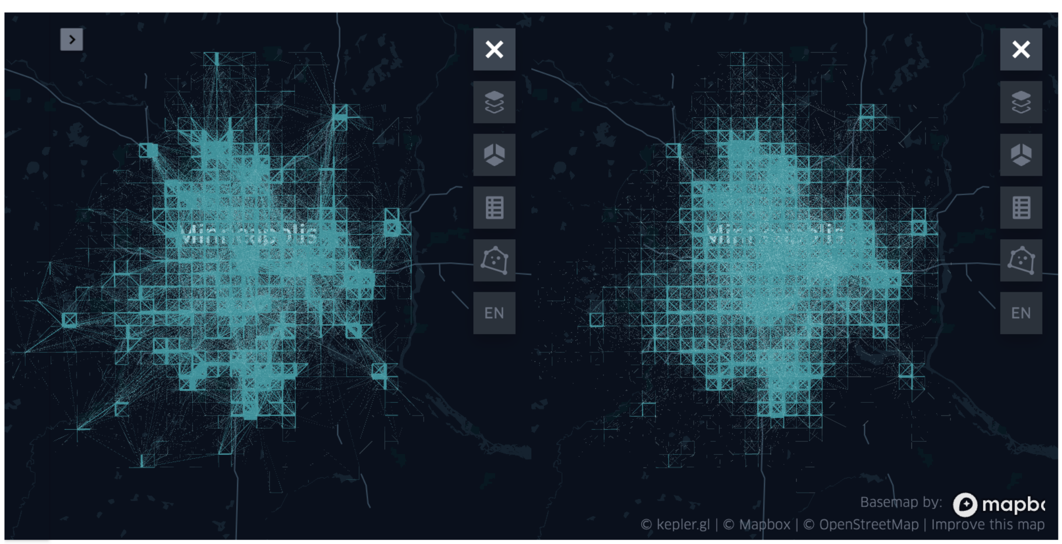

Figure 3: Obseved flow (left) and predicted flow using gravity model (right).

5.3 Deep Learning Model

The model receives input from two distinct categories of location features. The initial category encompasses place characteristics, such as land use type, which involves quantifiable aspects like the count of various infrastructural elements such as restaurants, schools, hospitals, and bus stations.

The second category involves weather-related factors, specifically temperature and precipitation. Illustrated in Figure 4 Simini et al. (2021), the model is composed of multiple parallel structures. The first structure involves the concatenation of features, where, for each spatial flow, the place characteristics and weather effects of both the origin and destination, alongside the distance between the two locations, are combined into input vectors. Subsequently, these input vectors $x(l_i, l_j)$ are concurrently fed into an identical feed-forward neural network. The outcome of the final layer produces a scalar $s(l_i, l_j) \in [-\infty, +\infty]$, referred to as the score. A higher score for a location pair (l_i, l_j) corresponds to an elevated likelihood of observing a trip from l_i to l_j according to the model. The conversion of scores into probabilities is accomplished through a softmax function. The resultant flow between two locations is determined by multiplying the probability (i.e., the model's output) with the origin's total outflow.

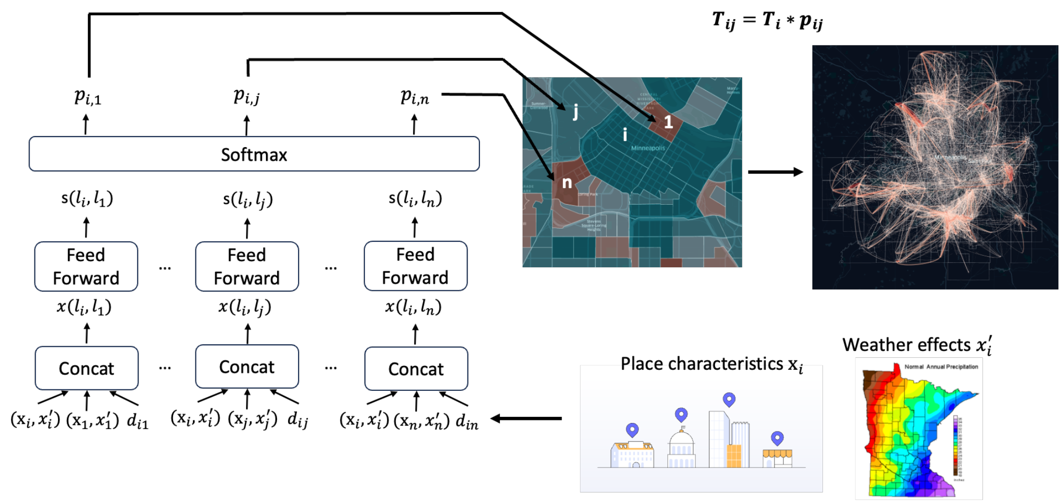

Figure 4: Deep learning model architecture

Figure 5 displays the predicted flows using the deep learning model with (left) or without (right) weather data. It seems that graphically no obvious difference can be seen from the plots, however, both deep learning models seem to do a better prediction at the boundary of the study area than the gravity model.

5.4 Evaluation Metrics

This project applies four different measurements to compare the performance of different models in flow generation: the Common Part of Commuters (CPC), the Pearson correlation coefficient(r), the Root Mean Squared Error (RMSE), the Jensen-Shannon divergence (JSD) Simini et al. (2021). CPC computes the similarity between the observed flow $f_{i,j}^p$ and generated flow $f_{i,j}^o$ and it is a positive value ranging from 0 to 1. The value one suggests a perfect match between the observation and predictions, while zero means no match at all and bad model performance. Pearson correlation measures the linear dependence between the observation and predictions, ranging from -1 (strongest negative correlation) to 1. Lower values of RMSE indicate better performance. JSD assesses the dissimilarity between the two variables and ranges in [0,1]. The value zero suggests two variables from the same distribution and good model performance.

Table 1 shows the performance of different flow generation models, where the best performance for each metric is bold. More specifically, the deep learning model with weather performs the best for the metrics of CPC and r, while the gravity model performs the best for the metrics of RMSE and JSD. The deep learning model with weather data is always better than the model without weather data.

$CPC = \frac{2 \cdot \sum_{i,j} min(f_{i,j}^p, f_{i,j}^o)}{\sum_{i,j} f_{i,j}^p + \sum_{i,j} f_{i,j}^o}$

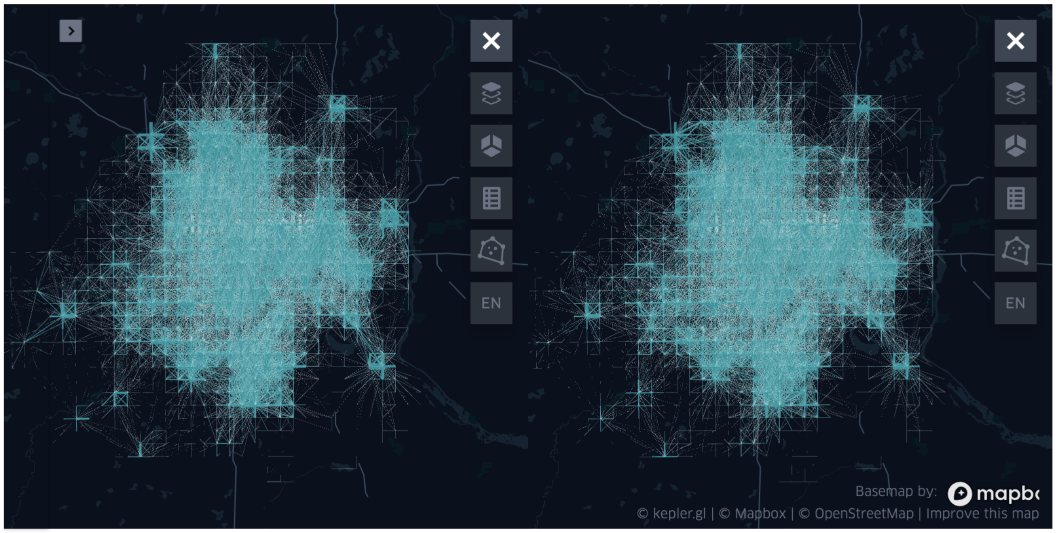

Figure 5: Predictions using deep learning model: with weather data (left) and without weather data (right).

Table 1: Performance of different models to generate flows

Metric	**Gravity**	**Deep Learning**	**Deep Learning (Weather)**
CPC	0.741	0.739	**0.755**
r	0.917	0.933	**0.946**
RMSE	**12.206**	13.110	12.439
JSD	**0.063**	0.075	0.066

$$r = \frac{\sum_{i,j}(f^p_{i,j} - \bar{f}^p) \cdot (f^o_{i,j} - \bar{f}^o)}{\sqrt{\sum_{i,j}(f^p_{i,j} - \bar{f}^p)^2} \cdot \sqrt{\sum_{i,j}(f^o_{i,j} - \bar{f}^o)^2}}$$

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i,j}(f^p_{i,j} - f^o_{i,j})^2}$$

Figure 6 displays how the metrics of r and CPC change across epochs for the training and test set. Again, the performance of the deep learning model with weather outperforms the model without weather info.

5.5 Reflections

Overall, this project explores the application of deep learning models in flow generation. Based on previous research, the neural network model implemented in this project is superior to the traditional gravity model in some indicators in predicting traffic. Additionally, this project found that weather information can help models achieve better predictions. The limitation of this study is that it does not fully exploit the advantages of deep learning models in processing unstructured data. For example, we can consider how to integrate street view photo data into the model.

References

Hugo Barbosa, Marc Barthelemy, Gourab Ghoshal, Charlotte R James, Maxime Lenormand, Thomas Louail, Ronaldo Menezes, José J Ramasco, Filippo Simini, and Marcello Tomasini. Human mobility: Models and applications. *Physics Reports*, 734:1–74, 2018.

Zhicheng Liu, Fabio Miranda, Weiting Xiong, Junyan Yang, Qiao Wang, and Claudio Silva. Learning geo-contextual embeddings for commuting flow prediction. In *Proceedings of the AAAI con-*

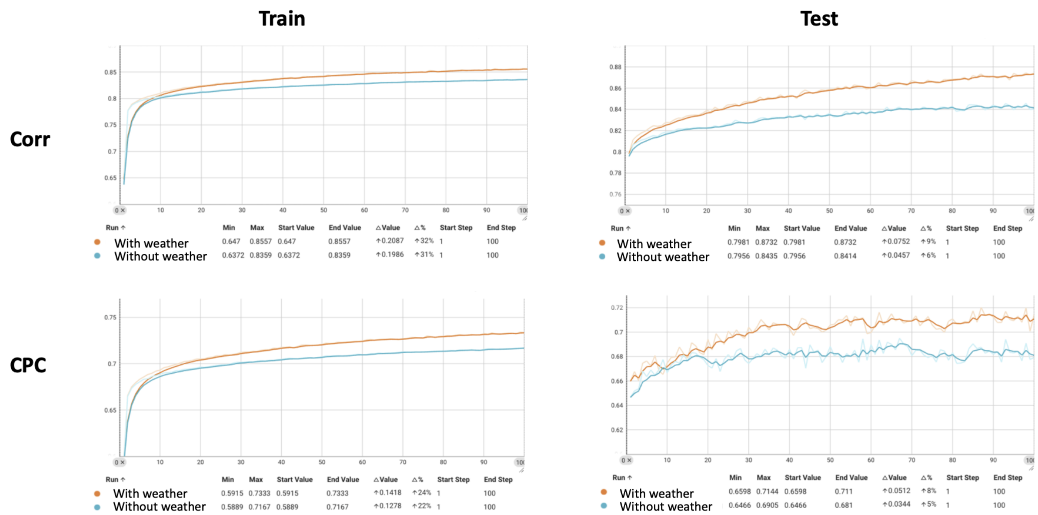

Figure 6: Predictions using deep learning model: with weather data (left) and without weather data (right).

ference on artificial intelligence, volume 34, pp. 808–816, 2020.

Massimiliano Luca, Gianni Barlacchi, Bruno Lepri, and Luca Pappalardo. A survey on deep learning for human mobility. *ACM Computing Surveys (CSUR)*, 55(1):1–44, 2021.

Filippo Simini, Gianni Barlacchi, Massimilano Luca, and Luca Pappalardo. A deep gravity model for mobility flows generation. *Nature communications*, 12(1):6576, 2021.

Maria Stefanouli and Serafeim Polyzos. Gravity vs radiation model: two approaches on commuting in greece. *Transportation research procedia*, 24:65–72, 2017.

Xin Yao, Yong Gao, Di Zhu, Ed Manley, Jiaoe Wang, and Yu Liu. Spatial origin-destination flow imputation using graph convolutional networks. *IEEE Transactions on Intelligent Transportation Systems*, 22(12):7474–7484, 2020.