# Learning Geo-Embeddings for Commuting Flow Prediction

**AAAI 2020**

**ZHUGUOWEI 20201205**

# Introduction

## Commuting flow

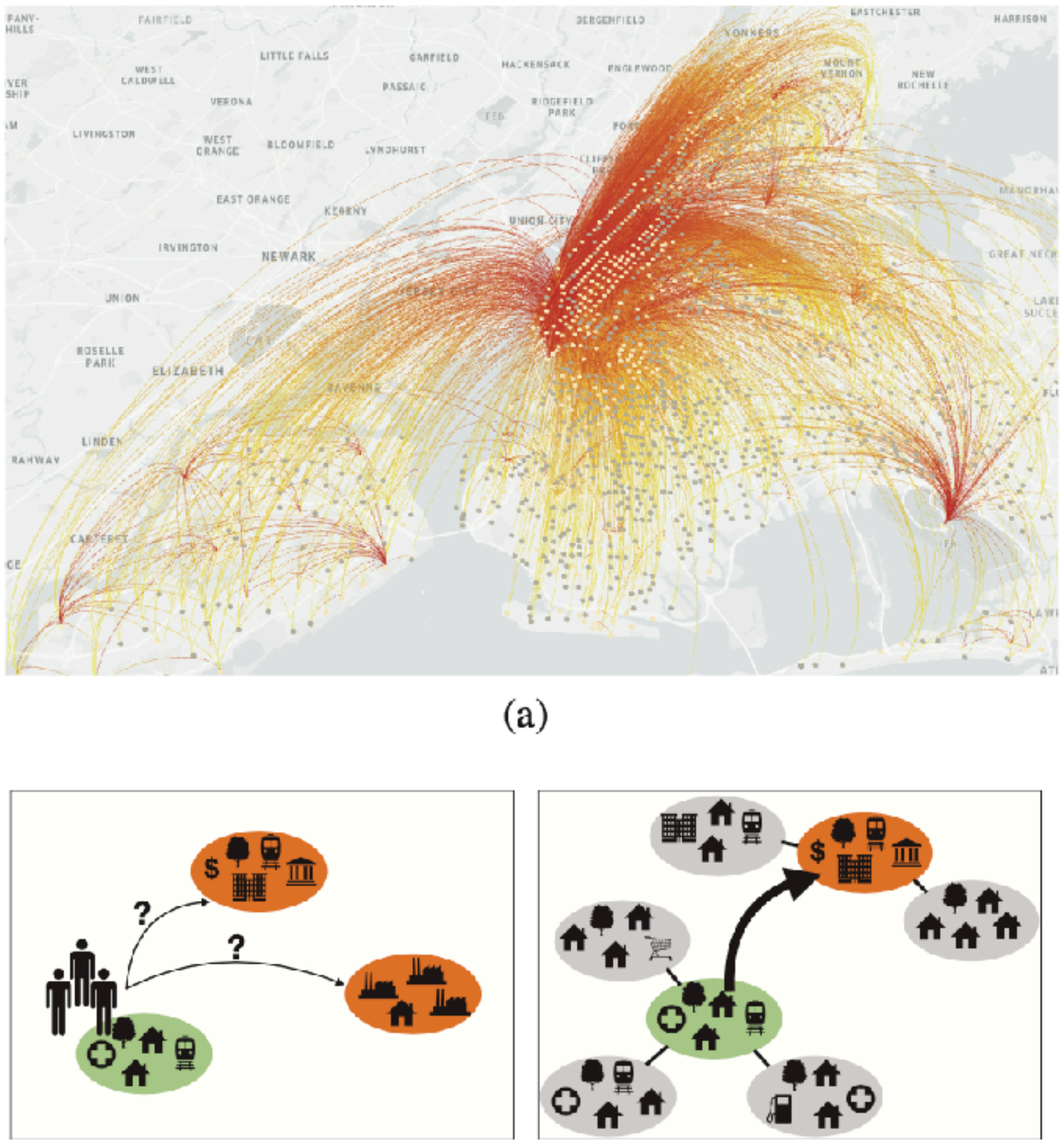
- The commute of people from home to work is a phenomenon that has shaped society and cities throughout the ages, from ancient Egypt to modern New York City
- These daily recurrent movements form a complex network that is highly correlated with the socioeconomic factors of cities
- In order to have more efficiently planned cities, it is crucial to understand how commuting flows are impacted by infrastructure and land use.
- As such, commuting flow prediction is one of the fundamental problems for urban planning in that it reveals the spatial interactions of supply and demand in a city(Rodrigue, Comtois, and Slack 2016)

# Introduction

## Commuting flow

- **Traffic OD forecasting** is essentially a time series prediction problem where *the historical movements will be used as input features*

- while **commuting flow prediction** problem aims at revealing spatial interaction of supply and demand in a city by predicting the edge-level signals (e.g. the volume of the flow), using *only the information of node attributes*, such as infrastructure and land use information



(a)

(b)

(c)

# Preliminaries

- Urban Geographic Unit
- Urban Indicators
- Geo-Adjacency Network
- Distance Matrix
- Commuting Trips

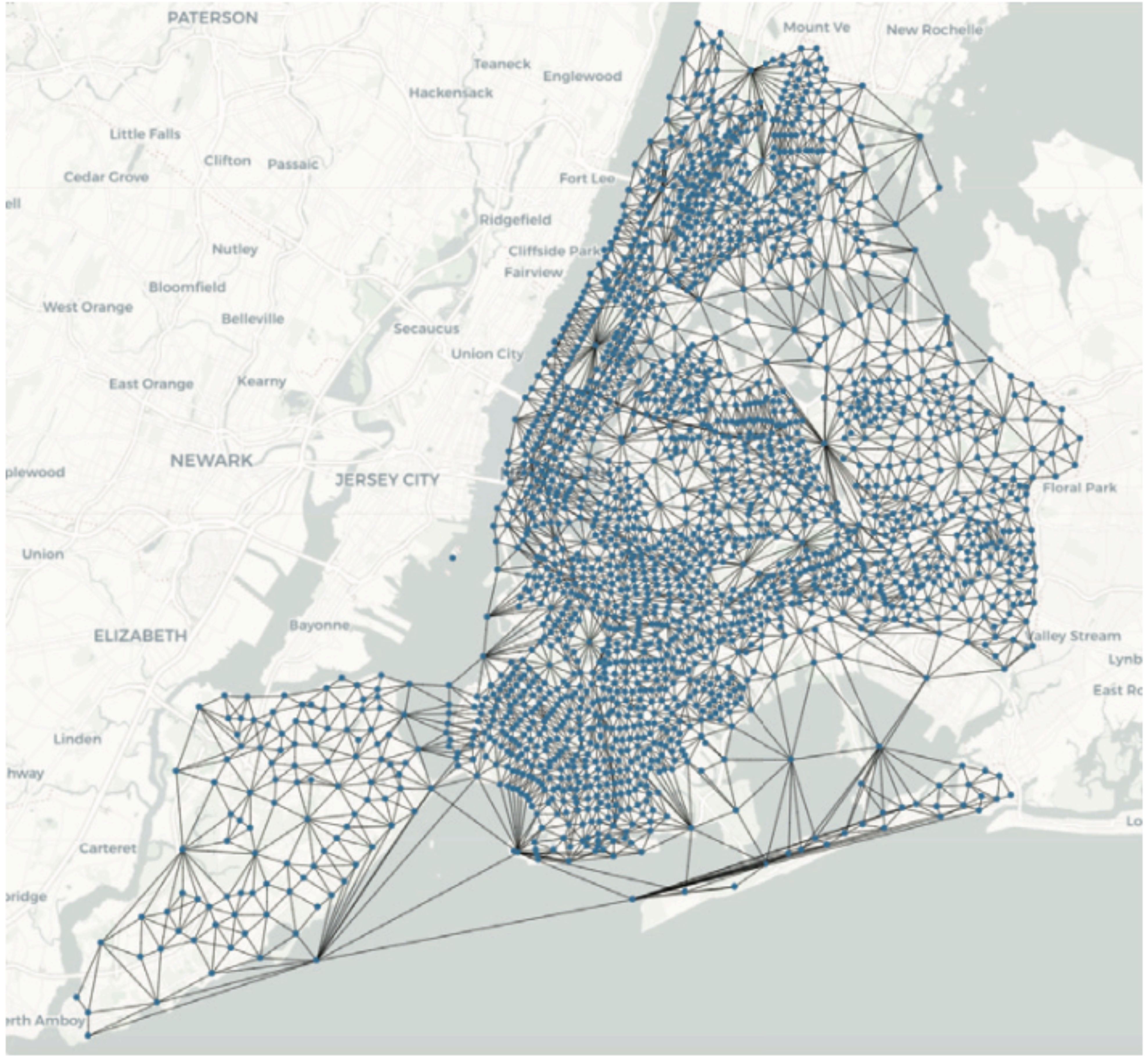



Figure 2: Geo-adjacency network of New York City. The dots represent the centroids of census tracts and the lines represent the edges.

# Related works

## Commuting Flow Prediction

- Gravity model: assumes the number of commuters traveling from one region to another is proportional to the product of the population of origin and destination and decays with the distance of the trip, as shown below:

- Nonparametric models (such as gradient boosting machine)

  These off-the-shelf machine learning models simply use origin-destination node attributes as input features to fit a regression model, ignoring the influence of nearby regions

- Intervening opportunity model

  These models consider the influence of nearby potential competitors of origin or destination, such as radiation model

Inspired by the idea of intervening opportunity, we propose using the geographic contextual information to develop the regression model where the embeddings of each node is encoded with the influence of nearby regions.

# Related works

## Graph Representation Learning

- GraphSAGE leverages node attribute to generate node embeddings in a message-passing way
- Graph attention network leverages self-attention mechanism to allow messages passed by neighbors to be aggregated with different weights

Motivated by these works, we use the framework of graph attention network and adapt the attention mechanism to our tasks so that our model could capture the geographic context

# Methodology
## Framework

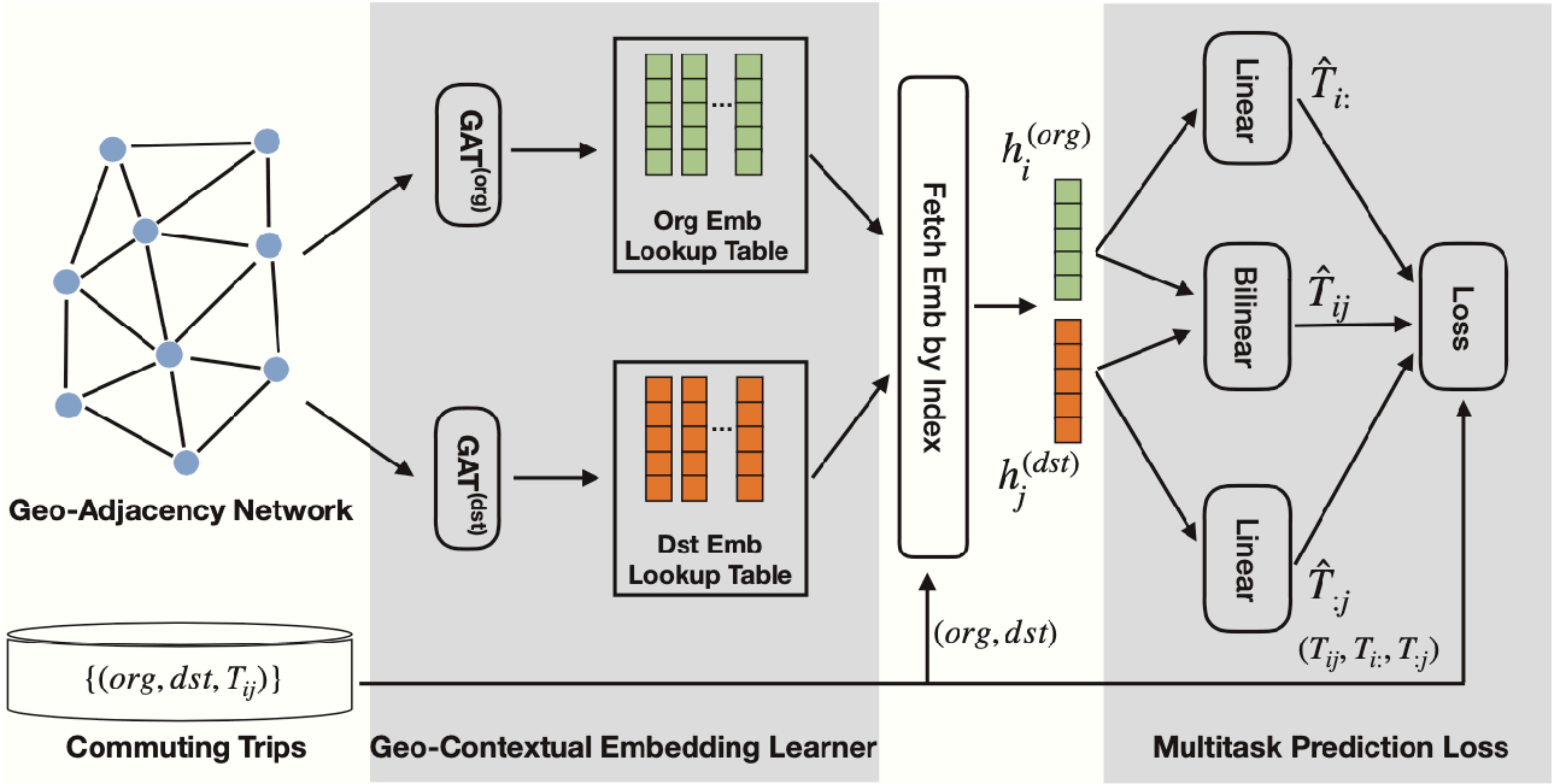
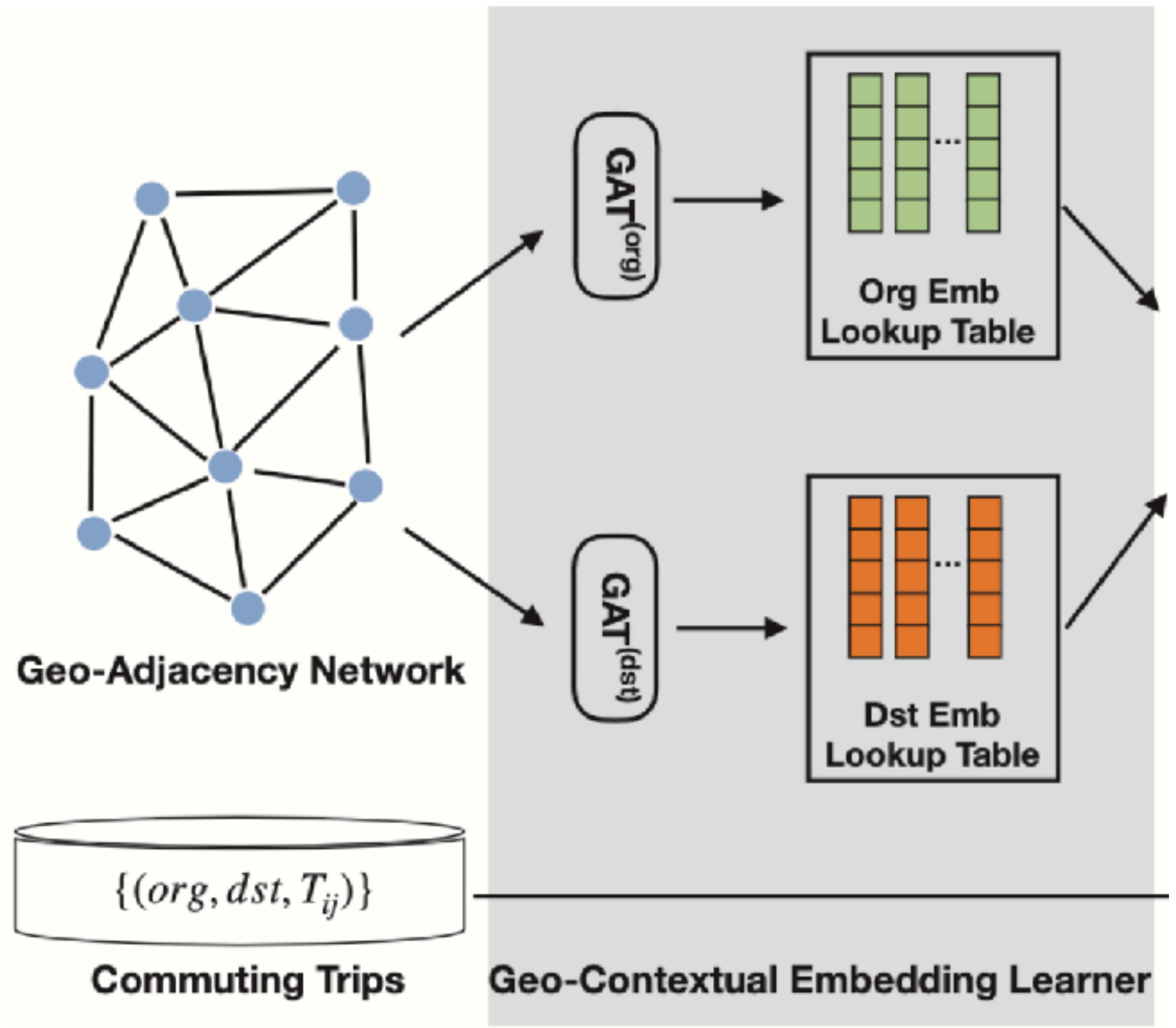



Figure 3: Framework of GMEL

# Methodology

## Geo-contextual Multitask Embedding Learner (GMEL)

- GMEL is designed to capture the spatial correlations from geographic context.
- Basically, the geographic context can be viewed as the graph neighborhoods of $G_{adj}$.
- GMEL utilizes Graph Attention Network (GAT) to encode the geographic contextual information into an embedding space.
- To disentangle the supply and demand characteristics that are hidden in infrastructure and land use, GMEL employs two separate GATs to encode the geographic contextual information
- GMEL employs multitask learning framework which imposes stronger restrictions forcing the embeddings to encapsulate effective representation for flow prediction (Caruana 1997).



# Methodology

## Multitask Learning

- **Main Task**: Predicting Commuting Flow

$$\mathcal{L}_{main} = \frac{1}{|T|}\sum_{i,j}(\hat{T}_{ij} - T_{ij})^2 \qquad \hat{T}_{ij} = h_i^{(org)T} W_b h_j^{(dst)}$$

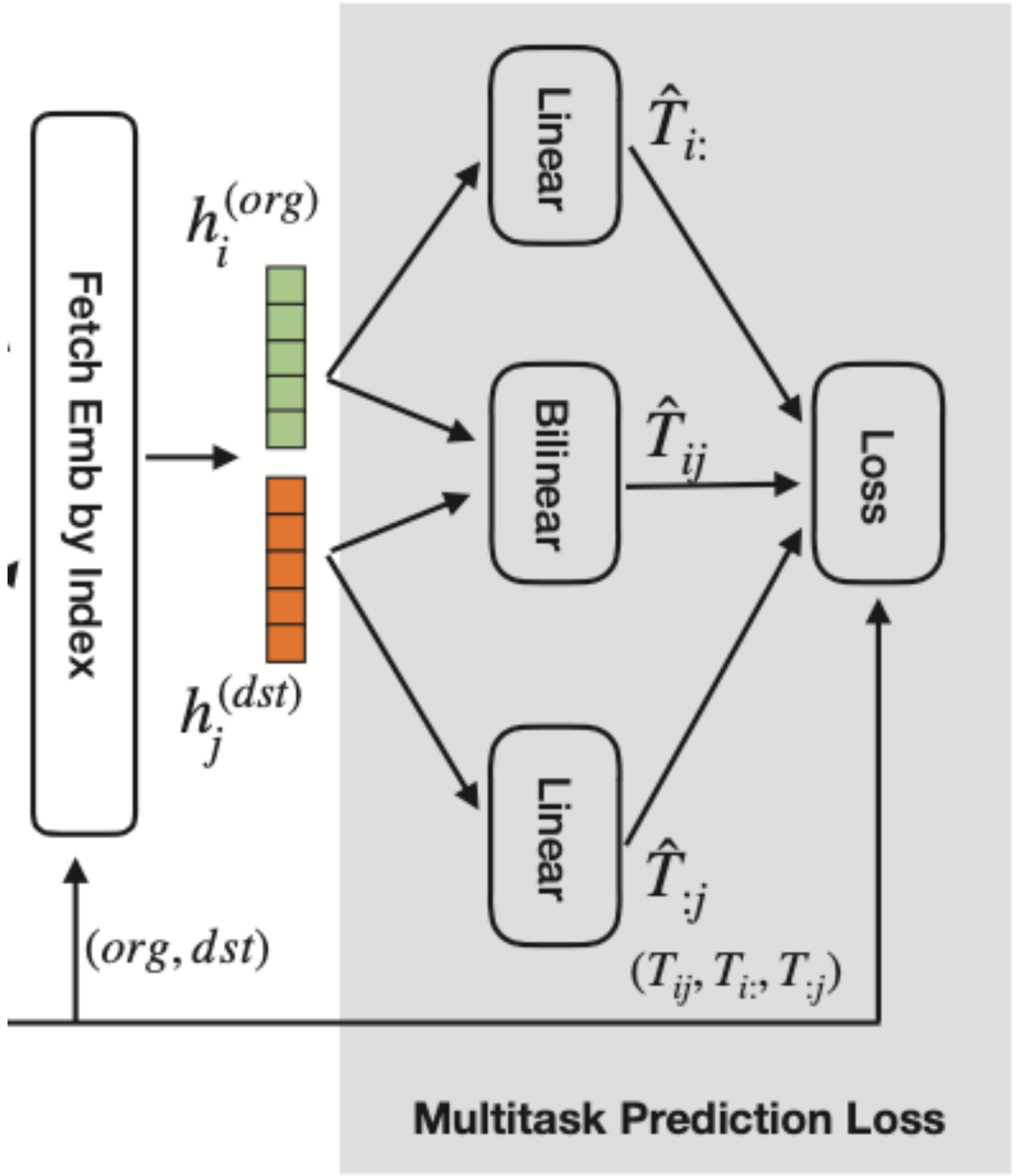
- **Subtasks**: Predicting In/Out Flow

$$\mathcal{L}_{out} = \frac{1}{N}(\hat{T}_{i:} - T_{i:})^2 \qquad \hat{T}_{i:} = \mathbf{w}_{out}^T h_i^{(org)}$$

$$\mathcal{L}_{in} = \frac{1}{N}(\hat{T}_{:j} - \hat{T}_{:j})^2 \qquad \hat{T}_{:j} = \mathbf{w}_{in}^T h_j^{(dst)}$$

- **Overall Loss Function**

$$\mathcal{L}_{GMEL} = \lambda_{main}\mathcal{L}_{main} + \frac{\lambda_{sub}}{2}(\mathcal{L}_{in} + \mathcal{L}_{out})$$



# Methodology

## Flow Predictor

- Most recently proposed machine learning models for commuting flow prediction employ gradient boosting regression tree (GBRT) or random forest as the regression function (Spadon et al. 2019; Pourebrahim et al. 2019; Robinson and Dilkina 2018).

- In particular, we use GBRT in this paper.

# Methodology

## Training Algorithm

1. Train GMEL using stochastic gradient descent method in an end-to-end manner
2. A GBRT is trained as flow predictor based on the concatenation of origin-destination embeddings and travel distance to predict the commuting flow.

---

**Algorithm 1:** Training Algorithm

---

**Input:** Geo-adjacency Network $G_{adj} = (V, E, A)$,
Distance Matrix $D$,
Commuting Trips $T_{train} = \{(v_i, v_j, T_{ij})\}$

**Output:** The learned GMEL,
The learned flow predictor $\hat{f}$

`/* GMEL Learning */`

**repeat**

  $T_{batch} \leftarrow$ Draw a training batch from $T_{train}$

  $\{h_i^{(org)}\} \leftarrow GAT^{(org)}(G_{adj})$

  $\{h_j^{(dst)}\} \leftarrow GAT^{(dst)}(G_{adj})$

  Evaluate $\mathcal{L}_{GMEL}$ by $(\{h_i^{(org)}\}, \{h_j^{(dst)}\}, T_{batch})$ using Equation 13

  $\nabla\mathcal{L}_{GMEL} \leftarrow$ Backpropagate $\mathcal{L}_{GMEL}$

  $w \leftarrow w - \gamma\nabla\mathcal{L}_{GMEL}$ `//` $\gamma$ `is the learning rate`

**until** *stopping criterion is met*;

`/* Flow Predictor Learning */`

$\{h_i^{(org)}\} \leftarrow GAT^{(org)}(G_{adj})$

$\{h_j^{(dst)}\} \leftarrow GAT^{(dst)}(G_{adj})$

$\mathcal{X}_{input} \leftarrow \{\}, \mathcal{Y}_{input} \leftarrow \{\}$

**for** $(v_i, v_j, T_{ij})$ *in* $T_{train}$ **do**

  $\mathcal{X}_{input} \leftarrow \mathcal{X}_{input} \cup Concat(h_i^{(org)}, h_j^{(dst)}, D_{ij})$

  $\mathcal{Y}_{input} \leftarrow \mathcal{Y}_{input} \cup T_{ij}$

**end**

$\hat{f} \leftarrow$ Train GBRT on $(\mathcal{X}_{input}, \mathcal{Y}_{input})$

---

# Experiments

## Dataset

- LODES

It is collected yearly and records the home and employment locations of workers, representing stable commuting flows.

- PLUTO

It records land use and infrastructure information at the tax lot level.This information is aggregated into census tract level (65 urban indicators for each census tract). A summary of the urban indicators is listed in Table 1.

- OSRM

We employ Open Source Routing Machine (OSRM) to measure the travel distance between the centroids of census tracts (Luxen and Vetter 2011).

Table 1: Summary of Urban Indicators

| Categories | # Features | Contents |
|---|---|---|
| Infrastructure | 40 | The number of different types of buildings (25), the density of commercial/residential/etc. units (4), the number of buildings in each built year interval (11) |
| Land Use | 23 | The number of tax lots in different land use (11), the land area ratio of retail/office/etc. (10), statistics of floor area ratio (2) |
| Speciality | 2 | Whether or not the census tract contains landmarks or historic districts (2) |
| Total | 65 | |

# Experiments

## Baselines

- Gravity Model with Power-Law Decay(GM-P)
- Gravity Model with Exponential Decay(GM-E)
- Random Forest(RF)
- Gradient Boosting Regression Tree(GBRT)
- Node2Vec
- GMEL-noMul

remove multitask settings and ***only keep the main task***

- GMEL-noSep

***only one GAT is used*** to generate embeddings and this set is used as both origin and destination embeddings.

# Experiments

## Performance Analysis

Table 2: Performance on Test Set

| *Model* | **RMSE** | **MAE** | **CPC***  |
|---|---|---|---|
| GM-P | 7.035 | 2.236 | 0.589 |
| GM-E | 6.944 | 2.179 | 0.602 |
| RF | 6.273 | 2.436 | 0.638 |
| GBRT | 5.454 | 1.974 | 0.707 |
| Node2vec | 5.455 | 1.994 | 0.704 |
| GMEL-noMul | 5.356 | 1.910 | 0.716 |
| GMEL-noSep | 4.982 | 1.772 | 0.737 |
| **GMEL (ours)** | **4.887** | **1.747** | **0.741** |

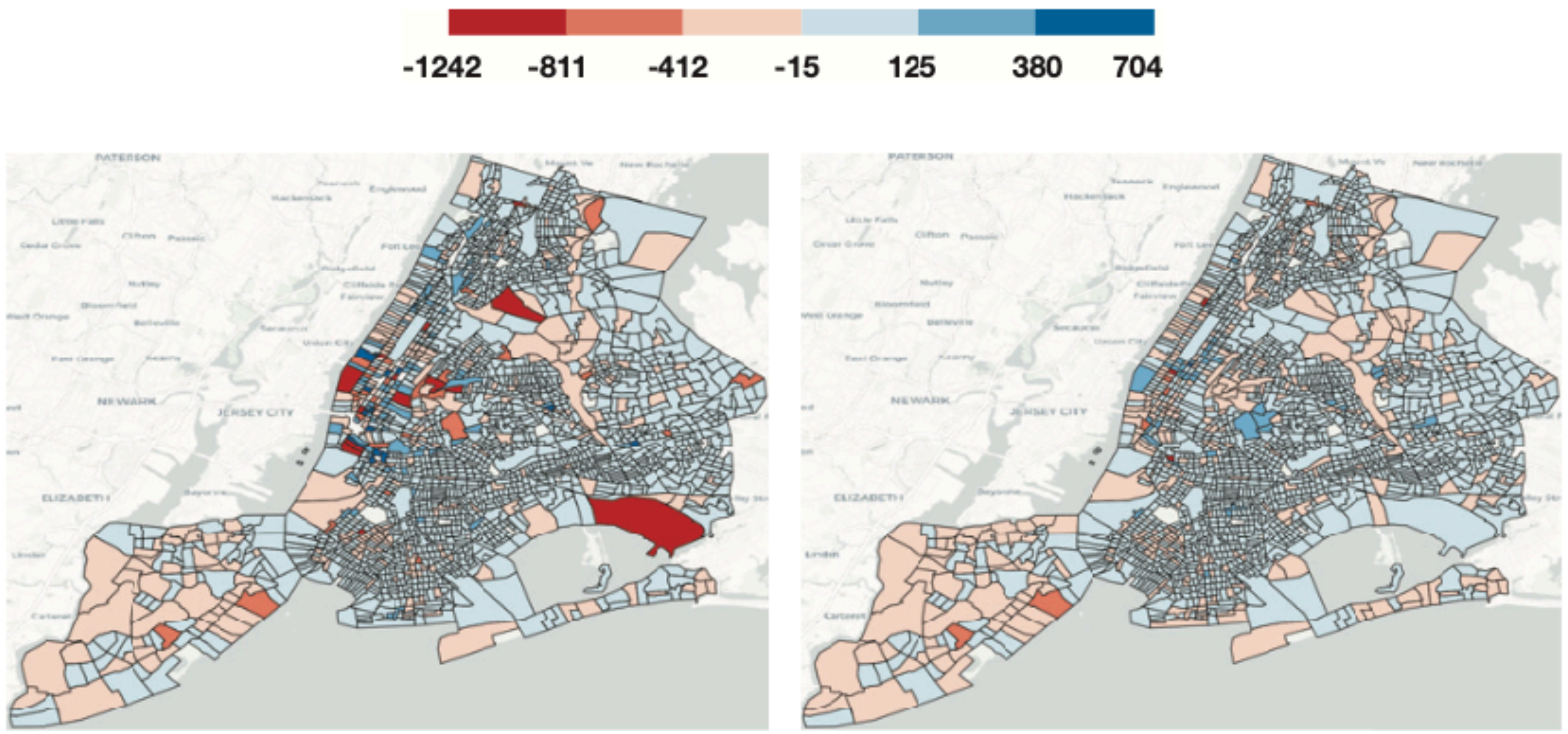
* Higher is better.

All GMEL variants outperform the above baseline models. This verifies the effectiveness of leveraging geographic contextual information for commuting flow prediction.

GMEL outperforms GMEL-noMul and GMEL-noSep. This shows the effectiveness of multitask learning framework and the necessity of modeling supply and demand characteristic separately.

# Experiments

## Residual Analysis



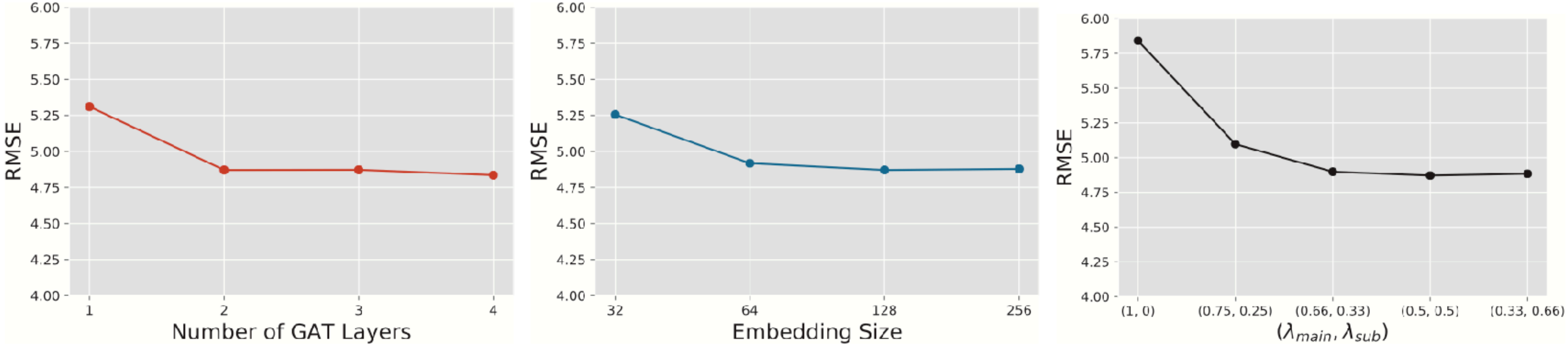

(a) Residuals of GBRT (b) Residuals of GMEL

Figure 5: Spatial distribution of residuals. Red indicates underestimation and blue indicates overestimation. Light blue census tracts indicate the best predictions.

- the residuals of GMEL are spatially smoother than that of GBRT.
- The reason is that GMEL exploits geographic contextual information to capture spatial correlations, and in doing so the prediction will take into account both the characteristics of regions of interest and the influence of nearby regions.

# Experiments

## Parameter Sensitivity Analysis




(a) Effect of number of GAT layers.

(b) Effect of embedding size.

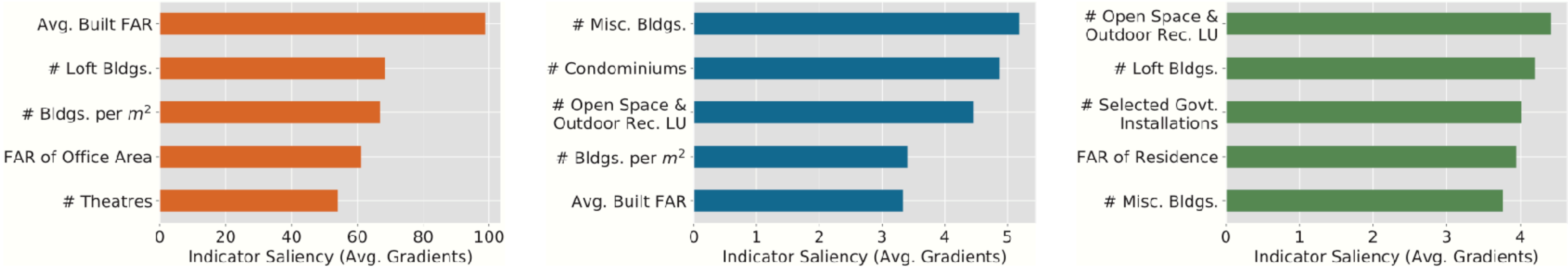
(c) Effect of multitask weights.

Figure 4: Results of different hyperparameter settings.

- The effect of **number of GAT layers**: When the number of GAT layers is greater than or equal to two, the performance doesn’t fluctuate too much
- The effect of **embedding size**: the performance increases as the embedding size increases from 32 and saturates at the size of 128
- The effect of **multitask weights**: when the weight of subtasks increases, the performance of the main task keeps increasing until the weights of the main task and subtasks are equal, i.e. $\lambda_{main} = 0.5$, $\lambda_{sub} = 0.5$.

# Experiments

## Feature Sensitivity Analysis




(a) Edge-level flow. (b) In-flow. (c) Out-flow.

Figure 6: Top-5 salient urban indicators.

- We evaluate the impact of the urban indicators by computing the saliency map of GMEL
- A larger absolute value of the saliency map points to a more prominent urban indicator.
- These salient urban indicators present the supply and demand characteristics for different kind of flows.

For example, the number of buildings per square meters, indicating job opportunities, is salient for in-flow, meanwhile, floor area ratio of residence, indicating the density of regular residences, is salient for out-flow.

# Conclusions

- Different from conventional gravity model and recently proposed machine learning methods, we propose the use of geographic contextual information for commuting flow prediction.
- As such, an end-to-end embedding learning framework based on graph attention network is proposed to learn geo-contextual embeddings of the geographic units.
- The results show that introducing geographic contextual information can greatly improve the accuracy of prediction and our model outperforms all baseline methods including the state of the art.

# Thanks!