

擬似人流データの生成を目指して： 大規模な日常生活行動を再現するためのPTデータに基づく深層生成モデルの開発

張 育榕, 張 鯤翼, 龐 岩博, 関本 義秀

Background

人の移動データは、新しいインフラや未曾有の災害、大規模な活動が発生した場合の人の行動変化をシミュレーションするなど、都市計画や交通計画において重要な役割を果たしている。しかし、プライバシーの問題や厳しい規制のため、実際の人の流れデータを取得するには、多大な時間、資源、資金が必要である。プライバシーに敏感な携帯電話のデータセットに関するアクセシビリティの問題を克服するために、大規模な人の動きを描写する合成データセットの再現は将来有望なアプローチである。

Object

擬似的な人の流れに向けた、大規模な人々の日常生活行動を再現するための深い生成モデルを開発する。

- 生成された活動の文脈的関連性（人口統計）と現実性を高める。
- 十分に表現されていない活動（買い物、通院）の表現を改善する。

Dataset

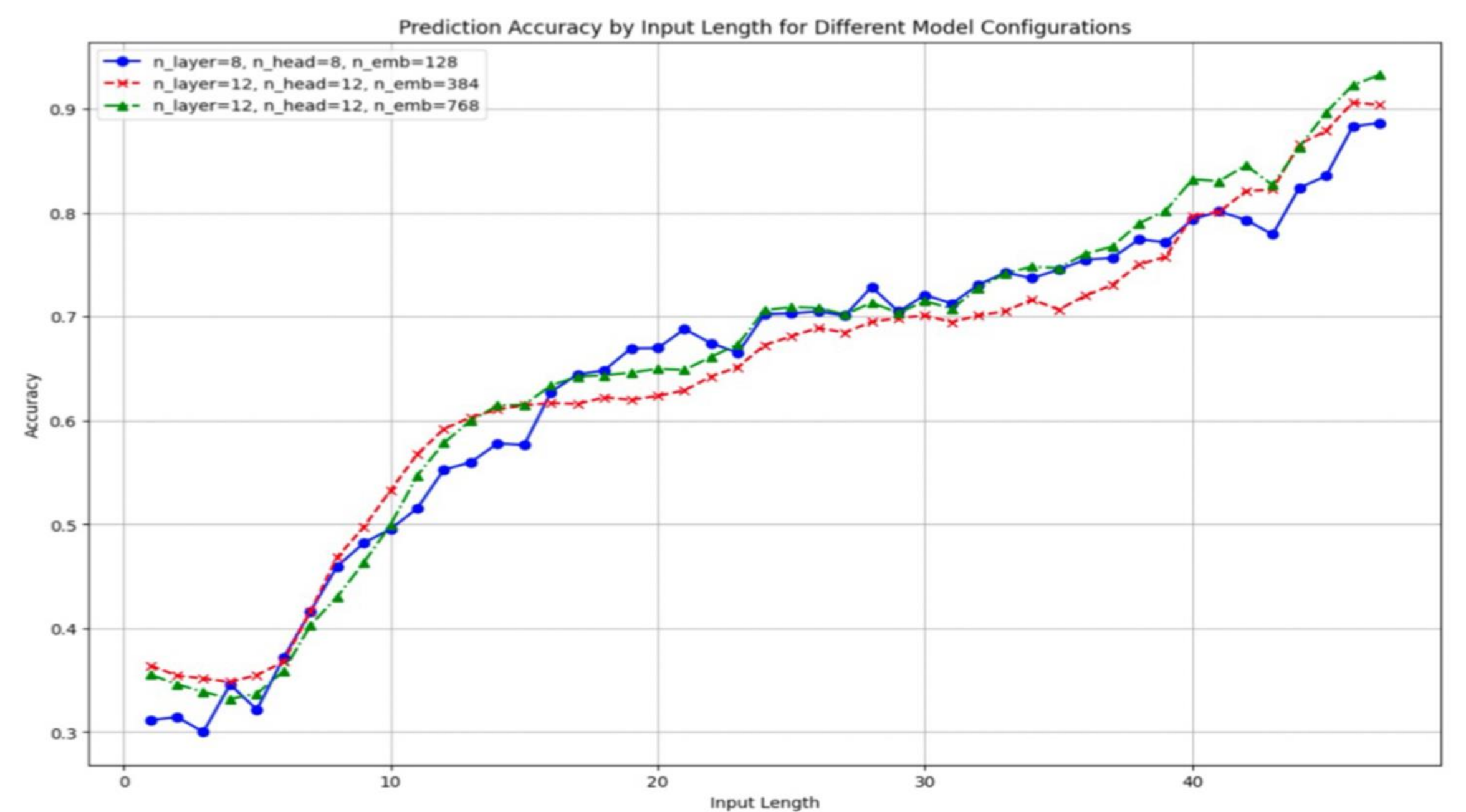
- PT調査で言及された「目的」コードに基づき、各移動区間のアクティビティを導き出し、それに応じた割り当てを行い、各区間は15分と設定する。
- モデルの学習データとして、首都圏をはじめとした日本国内10以上の都市圏のデータを使用する。

Methodology

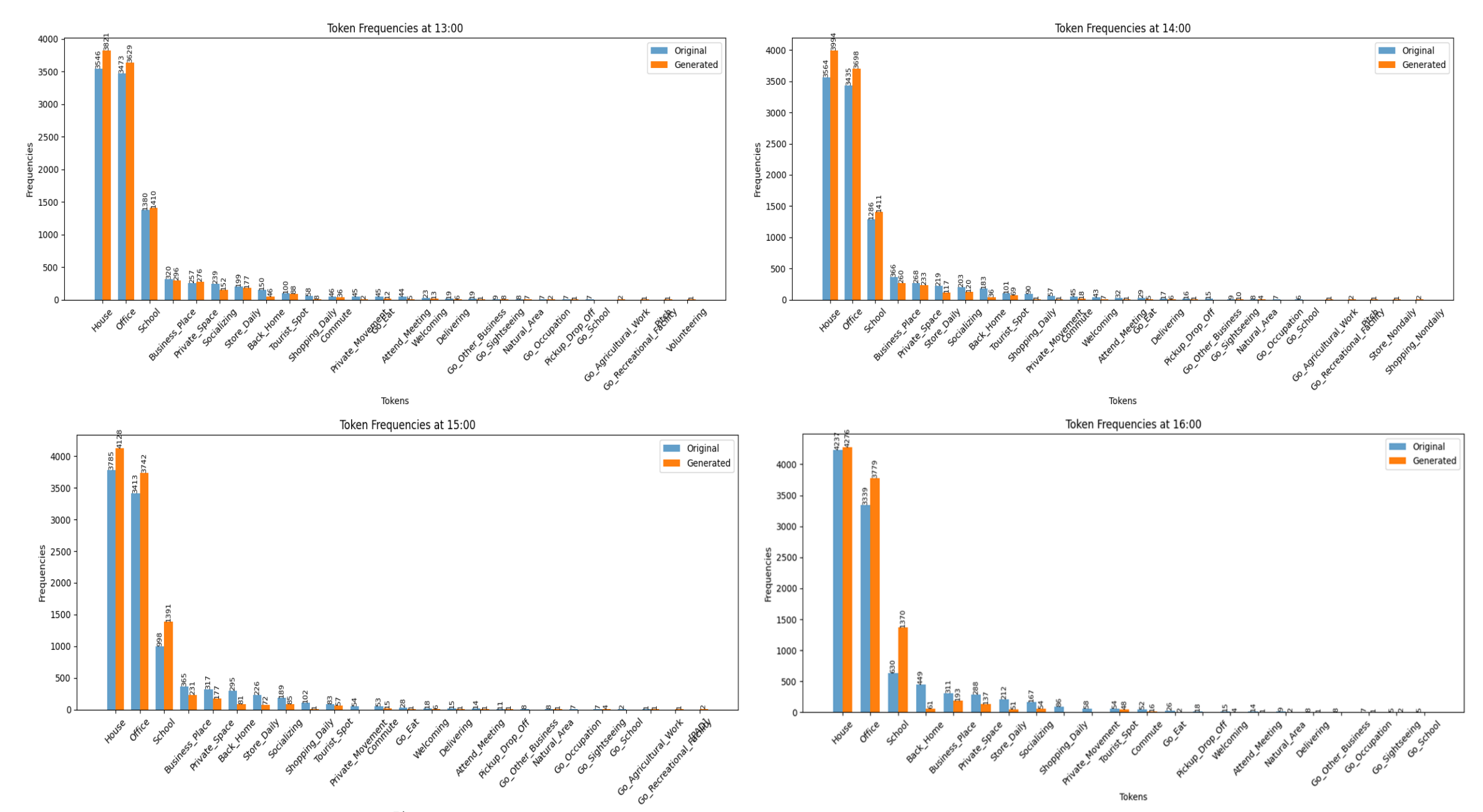
- **トレーニング段階：**
人間の移動データを2つのセグメントに分ける。例えば、24または36の行動データを入力データとして使用し、モデルの訓練と学習を行う。その後、モデルはデータに基づいて判断し、日常生活行動を完成させ、実データと比較して学習する。
- **調整段階：**
モデルの予測評価能力に基づいてモデルを最適化し、より現実的で鮮明な日常生活行動データを生成する。

Result & Conclusion

- 異なるモデル構成において、入力される行動データの長さでモデルの精度の変化を比較した。入力の長さが長くなるにつれて、すべてのモデル構成で予測精度が徐々に向上している。モデル構成が12ヘッド、12レイヤー（緑）の方が高い精度を示している。
- 各時点で様々な行動を行っている人の数を調べることで、モデル（Top_k=0.95, Top_p=10）の予測性能を評価する。時間が遅くなるほど、予測は現実から乖離し、予測できる行動は少なくなる。

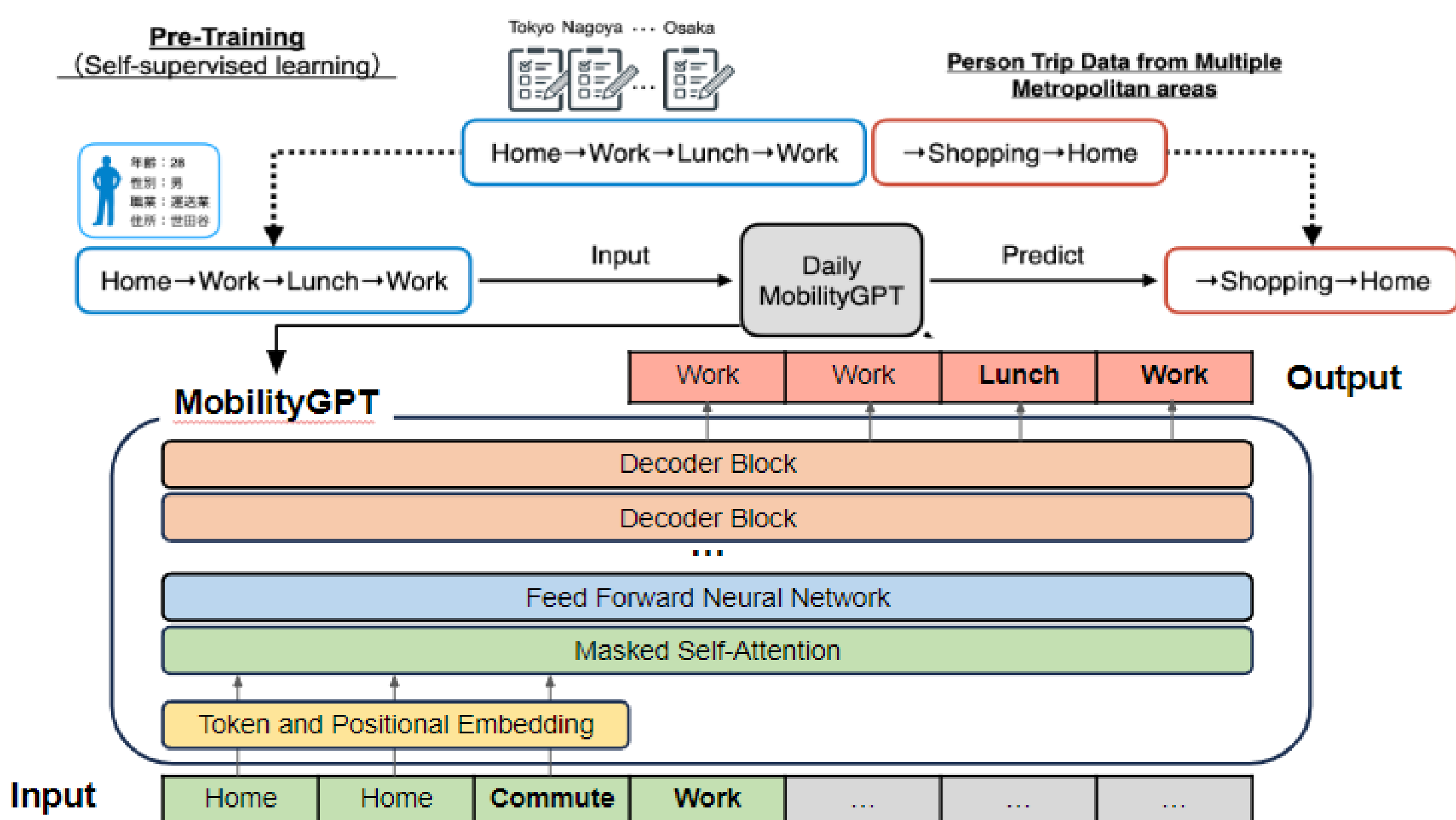


入力データの長さに対する異なるモデル設定の精度比較図



各時点での予測結果と実際のデータ活動数の比較図
(首都圏、入力データ=24)

以上より、GPT2モデルをベースに、日常生活行動に着目した人流データ生成モデルを開発した。しかし、研究目的と比較すると、鮮明で現実的な人流データを生成するためには、さらなる努力が必要である。



提案手法の流れ