

複数都市への転用を志向した深層学習ベースの OD 交通量予測モデルの開発 Development of a deep learning-based OD traffic volume prediction model applicable to multiple cities

鍛本 賢志[†]
Satoshi Kuwamoto

北野 佑[†]
Yu Kitano

1. はじめに

近年、人流データの分析・利活用に注目が集まっている。都市の人流を分析し、人々の移動行動を理解することで、都市の交通需要や混雑状況などを把握でき、都市開発の施策検討に効果的である。またデータ分析だけでなく、人流予測やシミュレーション等を行うことができれば、施策適用時の効果を机上検証でき、施策検討の効率化につながる。一方で予測モデルの構築には、学習データを収集する必要があるが、データ収集のコストが大きい。そこで既に構築したモデルが転用できれば、データ収集コストを抑えることができ、技術適用のハードルを下げるができる。

本研究では、学習データを取得した都市とは異なる都市の人流を予測可能な技術を開発する。具体的には、GPS で測位された人流データを用いて都市の地域間 OD 交通量を予測するタスクを扱う。予測手法として、OD を構成する各地域の属性や地理的な接続性を、GIS データやその他統計データから抽出し、交通量の時系列性と組合せて加味することで、高い表現力を備え汎用的に活用可能な深層学習モデルを検討した。以下で、手法の詳細と実データを用いた予測精度の検証結果について述べる。

2. OD 交通量予測

2.1 概要

本研究では、GPS で測位された人流データを用いて市区町村間の OD 交通量予測に取り組む。ここでいう人流データとは、人の ID、時刻、位置座標から構成されるデータのことである。本研究では株式会社 Agoop が提供する、スマートフォンから取得された位置情報データを用いる[1]。当該人流データをトリップごとに分け、地域ごとの位置・形状を表すジオメトリデータとの交差判定により OD 分析することで、地域間の OD 交通量を算出した。本研究では地域の単位を市区町村とした。

以下で問題設定を述べる。本研究では、ある都市の OD 交通量データをもとに予測モデルを構築し、異なる都市において将来の OD 交通量を予測するタスクを扱う（以下、別都市予測タスクと呼ぶ）。また、モデルの性能検証のため、学習データが取得された都市における予測タスク（以下、同一都市予測タスクと呼ぶ）の検証も行う。ここで都市とは、複数の地域の集合を指し、例えば都道府県などが相当する。本研究では各都市内の移動を対象とし、都市間にわたる移動はスコープ外とする。まず都市における市区町村の組合せで OD の組を定義する。ここで、出発地と目的地が等しい移動も含めるものとする。これは、市区町村内での移動に相当する。そして、1 時間単位で集計して OD 交通量を算出する。予測時には、過去 5 時間分の OD 交通量のデータとその他外部データをもとに将来 1 時間分の OD 交通量を出力するものとした。

2.2 課題

別都市予測タスクに対応するには、都市ごとの性質に依存しない予測モデルを構築する必要がある。OD 交通量の予測手法として、地域間の距離や各地域の人口情報を用いる重力モデルなどの古典的手法や、深層学習を適用し OD 交通量の時系列傾向を LSTM などモデリングする手法が考えられる。しかしながら、上記の手法ではモデルの表現力の低さが問題となる。例えば LSTM などでのモデリングでは、学習データに含まれない時系列パターンが予測対象に含まれると精度劣化をまねく可能性がある。

OD 交通量の傾向は、その時系列性に加え、OD を構成する地域の属性関係や地理的な接続性にも影響を受けると考えられる。例えば平日であれば、朝や夕方の通勤・通学時間帯には都市部と住宅街の間の移動が多く、昼間にはそうした都市部内外の移動が少なくなる傾向があると考えられる。そのような地域属性は、地域に存在する各種施設や交通手段のリソースの多さ、また移動主体の数である人口などのデータに表れると考えられる。また、OD 間の接続性に関しては、OD 間距離が交通量に影響する傾向がある。こうした OD の性質を加味してモデルの表現力を高めるべく、OD 交通量の時系列性に加え、OD を構成する地域の属性や接続性といった空間的性質を合わせて加味するモデルを設計する必要がある。

2.3 提案手法

提案手法のモデル構造を図 1 に示す。本モデルへの入力データは、地域ごとの情報を表すノード特徴量と、OD の情報を表すエッジ特徴量に分けられる。ノード特徴量には、市区町村ごとの発生交通量および集中交通量の時系列データと、各地域の属性データ、時間帯や天候などの補助データを含む。これらノード特徴量をニューラルネットワークなどに入力して地域ごとの潜在表現を獲得し、さらに各 2 地域の潜在表現を組合せて処理することで、地域属性の関係を加味した OD の潜在表現を得る。エッジ特徴量には、

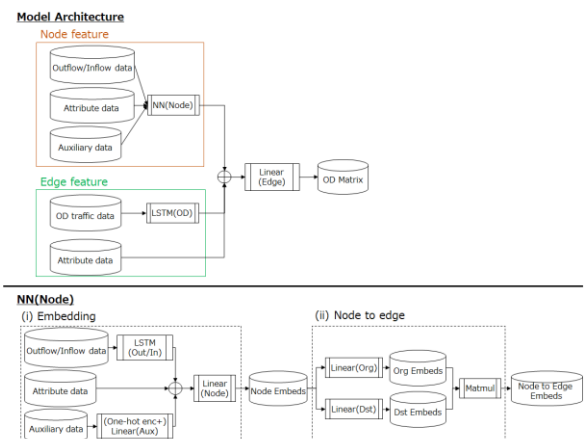


図 1 モデルの全体構造（上）と、NN(Node)コンポーネントの構造（下）

[†] (株) 日立製作所 研究開発グループ
Hitachi Ltd., Research and Development Group

OD 交通量の時系列データと、接続性など OD の属性を表すデータを含む。これらをニューラルネットワークで処理し、OD の属性や交通量の時系列的傾向を加味したベクトルを得る。最後に、ノード特徴量とエッジ特徴量をそれぞれ処理して得た OD の潜在表現を結合し、全結合層に通して OD 行列の予測結果を得る。

本研究では各地域の属性データとして、移動主体の数を表す人口、移動需要を生み出す施設数（事業所・学校・医療機関）、移動手段のリソースを表す駅・バス停留所の数、道路長を含む。OD の属性データとしては各 2 地域間の距離を用いた。これらは国勢調査や経済センサス、国土数値情報をもとに作成した。補助データには時間帯(7 時台、8 時台など)と曜日の情報を用いた。またデータ前処理の際、都市ごとの性質に非依存なモデル構築を目的として、各種データのスケール処理をした。外部データ由来のノード特徴量については、各都市における最大値で割って正規化した。発生交通量と集中交通量については、時間帯ごとの各都市の総交通量で割って正規化した。特徴量および目的変数である OD 交通量に関しては、標準偏差で割ってスケール処理した。

3. 実験

3.1 実験条件

提案手法の性能検証のため、実データを用いた評価実験を行った。OD 交通量の算出は、株式会社 Agoop が提供する人流データを OD 分析した結果を用いた。人流データは、2019 年 10 月 1 日から 10 月 31 日の間に、東京都、神奈川県、札幌市の 3 都市にて計測されたデータを対象とした。各都市の地域単位となる市区町村数は東京都、神奈川県、札幌市のそれぞれで 62, 58, 10 であった。また、計測期間の最後の 1 週間をテストデータとし、それより前の期間のデータを学習データとした。モデル学習時は、学習データの最後の 1 週間分を検証用データとして用いた。本実験において、同一都市予測タスクでは、ある 1 都市の学習データを用いてモデルを構築し、同じ都市のテストデータを用いてモデルの評価実験を行った。別都市予測タスクでは、ある 2 都市の学習データを用いてモデルを構築し、残り 1 都市のテストデータを用いてモデルの評価実験を行った。

比較のため、OD 交通量の時系列データを LSTM で処理するベースライン手法も評価した。ベースライン手法では、OD 交通量の時系列データを LSTM で処理し得られた潜在変数を、複数の全結合層に通して予測値を算出するものとした。ベースライン手法の設定について、LSTM の層数は 1、全結合層の層数は 3 とした。提案手法の設定は図 1 を参照し説明する。LSTM(OD), LSTM(Out), LSTM(In)の層数はすべて 1 とした。補助データである時間帯、曜日に対する全結合層 Linear(Aux)の層数はいずれも 1、潜在変数の次元はそれぞれ 3, 2 とした。その他の全結合層について、Linear(Node), Linear(Org), Linear(Dst)の層数は 1、Linear(Edge)の層数は 5 とした。各手法とも各全結合層と LSTM の潜在変数次元は同一とし、検証データを用いて最適な次元数を探索した。探索範囲は{16,32,64,128}とした。また、最適化アルゴリズムは Adam とした。学習率は上記と同様に検証データを用いて探索し、探索範囲は{0.01,0.001,0.0001}とした。また、過学習を回避するため L1 正則項を加えた。正則項の係数は 0.0001 とした。さらに EarlyStopping を設定し 100 エポックの間検証用デー

表 1 予測精度検証結果。表中の数値は、左が同一都市予測タスクの結果、右が別都市予測タスクの結果。

対象都市	手法	MAE	MAPE
東京都	提案手法	0.707 / 0.879	22.8% / 30.4%
	ベースライン	0.842 / 0.936	28.6% / 29.7%
神奈川県	提案手法	0.486 / 0.679	22.5% / 32.0%
	ベースライン	0.564 / 0.837	27.7% / 38.8%
札幌市	提案手法	2.282 / 2.696	18.0% / 23.1%
	ベースライン	2.913 / 2.862	23.1% / 23.5%

タに対する損失関数の値が低減しない場合に学習を停止させるものとした。

評価手順を述べる。まず各手法にて OD 交通量の回帰を行った後、出発地域を基準として OD 遷移割合に変換する。そして、発生交通量の実測値に OD 遷移割合を乗じて最終的な OD 交通量の予測値を算出し、精度評価に利用した。評価指標は MAE と MAPE を用いた。なお、OD 交通量が少ない時に MAPE が発散する問題があるため、実測値が 10 人以上のデータのみを対象として MAPE を算出した。

3.2 実験結果

提案手法とベースライン手法について、評価指標の比較結果を表 1 に示す。タスク間で結果を比較すると、別都市予測タスクの方が同一都市予測タスクよりも概ね精度が低い。想定通り、都市ごとにデータの性質が異なるために、別都市予測タスクの方が予測が困難であると見られる。両手法の評価結果を比較すると、両タスクともベースライン手法に対し提案手法による精度改善が確認される。MAE ベースで見ると、同一都市予測タスクでは 3 都市平均 17.2%の精度向上、別都市予測タスクでは 3 都市平均 10.3%の精度向上となった。ただし、東京都を対象とした別都市予測タスクでは、MAPE ベースで 0.7%誤差増と僅かな精度劣化が確認された。MAPE は実測値が 10 人以上のデータを用いて評価していることを踏まえると、交通量が多い場合の予測で課題が残っている可能性が考えられる。

4. おわりに

本研究では、予測モデル構築のためのデータ取得コストを抑え、様々な都市での予測を容易に実現すべく、学習データと異なる都市でも OD 交通量を予測可能な手法を検討した。具体的には、OD 交通量の時系列データに加え、GIS データや統計データを活用し、交通量の時系列性と、OD を構成する地域属性の関係や OD 間の接続性といった空間的性質を合わせて加味する深層学習モデルを開発した。

提案手法の評価のため、東京都、神奈川県、札幌市にて GPS 計測された人流データを用いて、ある都市でのデータをもとにモデルを訓練し、他都市に転用する予測実験を行ったところ、LSTM を用いたベースライン手法に比して MAE が平均 10.3%改善することを確認した。今後は他都市でのさらなる検証を進めるとともに、GCN 等により都市のネットワーク構造を精緻に加味する手法を検討する。

参考文献

- [1] 株式会社 Agoop, 流動人口データ ホームページ, <https://www.agoop.co.jp/service/dynamic-population-data/>