

OpenStreetMap データからの交通ネットワーク の抽出とグラフ畳込みネットワークによる 交通需要推定

福田 幸二¹・Saeed Maadi²・Kai Shen³・Schmöcker Jan-Dirk⁴

株式会社日立製作所 研究開発本部日立京大ラボ (〒606-8501 京都府京都市左京区吉田本町)

E-mail: koji.fukuda.jf@hitachi.com

グラスゴー大学 計算機科学部 (18 Lilybank Gardens, Glasgow G12 8RZ)

E-mail: Saeed.Maadi@glasgow.ac.uk

京都大学大学院工学研究科 都市社会工学専攻 (〒615-8530 京都市西京区京都大学桂)

E-mail: shen.kai.68w@st.kyoto-u.ac.jp

京都大学大学院工学研究科准教授 都市社会工学専攻 (〒615-8530 京都市西京区京都大学桂)

E-mail: schmoecker@trans.kuciv.kyoto-u.ac.jp

オープンな地図情報プラットフォームである OpenStreetMap から交通ネットワークを抽出し、グラフ畳込みネットワークにより交通需要の OD パタンの推定を行った。

OpenStreetMap (OSM) は、クラウドソース化されたオープンアクセスな地図情報プラットフォームである。本研究では、OSM から交通ネットワークを抽出し、グラフ畳込みネットワーク (GCN) を用いて交通需要の OD パタンの推定を行った。まず、OSM から交通ネットワーク抽出における課題および解決策を説明する。OD パタンの推定は、抽出された交通ネットワークに GCN を適用することで得られる各地点の“雰囲気”をもとに、一般化された重力モデルを用いて教師ありの機械学習を End-to-End で行う。オープンなデータセットである TNTP データを用いて学習を行い、精度を確認した。

Key Words: OpenStreetMap, 交通需要推定, グラフ畳込みネットワーク, 地理情報システム

1. はじめに

OpenStreetMap (OSM) は、クラウドソースのオープンアクセスプラットフォームの地理情報システムである¹⁾。データの品質は地域によって異なるが、世界中の多くの地域において交通ネットワークに関する情報を提供している。とくに、近年、公共交通ネットワークに関する情報の量と質も大幅に改善されている。

ところで、近年の機械学習分野における深層学習技術の進歩は目覚ましいものがある。深層学習は、初期には、主に画像処理に適用され、畳込み層を多数積み重ねる畳込みニューラルネットワーク (Convolutional Neural Network, CNN) が非常に成功をおさめた²⁾。これをうけて、最近、CNN を画像処理以外の分野に適用する試みが行われている。なかでも、グラフ構造に畳込み層を

適用するグラフ畳込みネットワーク (Graph Convolutional Network, GCN) が注目されている^{3,4)}。

本研究では、OSM の地理情報データと、GCN の技術を組み合わせることで、交通需要分布 (OD パタン) の予測を行うことを試みる。具体的には OSM から、道路ネットワークや公共交通ネットワーク、および、住宅・学校・駅・公園といった POI (Point of interest) を抽出して、グラフ構造を生成し、生成したグラフに GCN を適用して学習を行う。交通需要は、出発地および目的地の周辺環境に大きく依存すると考えられるため、GCN によって各地点の周辺環境を要約することで、交通需要の予測が可能となる。TNTP データセット⁵⁾を用いて学習およびテストを行い、交通需要分布を高い精度で推計できることを確認した。

2. OpenStreetMap を用いた交通ネットワーク構造の抽出

(1) OpenStreetMap (OSM) の概要

OpenStreetMap (OSM) はオンライン上のコミュニティで開発・維持されている全世界を対象とするオープンな地理空間情報のデータベースである。OSMのデータは、Open Data Commons Open Database License (ODbL) ライセンスのもとで公開されており、クレジット表示のもとで、自由にコピー、配布、送信、利用することができる。OSM は、多数の Node (緯度・経度で特定された地球上の点) と、複数のノードを順にまとめた Way (地図上の曲線を表す)、複数のノードおよびウェイをまとめた Relation の3種類の要素で構成されている。各要素には、それが何であるかを示す、あるいは、名前、住所などの情報を付随する属性値を表現する複数のタグが付加されている。多少の例外はあるが、要素の種類に応じて付加すべきタグの種類がルール化されており、機械的な処理が可能である。

(2) 交通ネットワークグラフの構築

OSM には、道路ネットワークに加えて、鉄道、バスといった公共交通機関のルート情報が格納されている。したがって、OSM データから交通ネットワークを抽出してグラフを構築することができる。本研究では、Julia 言語の OSM 処理ライブラリである `OpenStreetMapX.jl`⁹ を拡張して、OSM データから周辺環境を含んだ交通ネットワークグラフを構築するソフトウェアを開発した。

開発したソフトウェアは、以下の5つのモジュールより構成される。

1. 道路網モジュール： OSM から道路を抽出し、交差点をノード、交差点間をつなぐ道路をエッジとするグラフを構築する。
2. 鉄道網モジュール： OSM から鉄道路線を抽出し、駅をノードとして鉄道網を表すグラフを構築する。
3. バス網モジュール： OSM からバス路線を抽出し、バス停をノードとしてバス網を表すグラフを構築する。
4. POIモジュール： OSM から POI を抽出して、POI をノードとする孤立点グラフ (ノードのみでエッジが存在しないグラフ) を構築する。
5. グラフ併合モジュール： 上記モジュール群で生成したグラフを併合して、1つのグラフにする。

なお、各モジュールによって生成されるグラフのノードは地図上の地点を表しており、必ず、緯度・経度が付されている。グラフ併合モジュールでのグラフ併合の処理

は、ノードの緯度・経度をもとに、道路網モジュールで構築した道路網グラフのノードに、他のグラフのノードを併合することで行った。具体的には、道路網グラフ以外のグラフ上の各ノードを、道路網グラフ上の最も近くのノード (交差点) に併合する。ただし、最も近い交差点が 100m 以上離れているノードは、併合せずに削除した。

グラフ併合モジュールによる併合処理により、道路網グラフをもとにして、グラフ上のノード (交差点) に隣に存在する POI 情報が付加される。また、鉄道網およびバス網は、道路網グラフ上の離れたノード (交差点) 間を直接結ぶエッジ (スキップコネクション) として表現されることになる。これにより、交通ネットワークおよび周辺環境 (周辺に存在する POI の情報) を全て含んだグラフを構築することができる。

(3) OSM から交通ネットワークグラフ構築における課題

OSM データから交通ネットワークグラフを構築する場合、OSM データの構造に由来する課題が存在する。

OSM データでは、前述のように、駅やバス停のような地点は Node として表現され、鉄道路線 (線路) やバスルートのような地図上の曲線は Way として表現されている。問題は、駅やバス停を表す Node と、鉄道路線やバスルートを表す Way とが互いに関連づけられていない場合が多いことである。このため、特に、多くの鉄道路線やバス路線が存在し、多くの駅やバス停が密集する都市部では、ある駅やバス停が、どの鉄道路線やバスルートに属するのかが明らかでない場合がある。

実は、OpenStreetMap の標準の指針としては、Node と Way を関連付ける Relation を使って、駅と鉄道路線、あるいは、バス停とバスルート、を関連づけることが規定されているのであるが、現状、実際のデータにおいては、この指針が適用されていないことが多い。このようなデータ不整合の問題は、OpenStreetMap の、ボランティアによって維持・編集されるという性質上、避けられない課題であると考えられる。

本研究では、駅やバス停は、直線距離で 50m 以内に存在する全ての、鉄道路線およびバスルートに属するとした。こうすることで、複数の鉄道路線やバスルートに属する駅 (乗り換え駅) を抽出できる。もし、駅やバス停から 50m 以内に鉄道路線やバスルートが存在しない場合には、データの不整合として、当該の駅やバス停を削除した。

しかしながら、実際には、例えば、バス停がバスルートであるメインの道路から離れて、病院などの正面玄関前に設置される (バスが病院の敷地内に入る) 場合もしばしばある。このような場合を、どのように扱うかは、今後の課題である。

3. グラフ畳み込みネットワーク (GCN) による交通需要予測

(1) 地理情報からの交通需要分布(OD分布)の推定

交通需要分布 (OD 分布) は、都市設計における基本的な情報である。交通需要分布は、通常、対象地域内の任意の2点 (起点 Origin と、終点 Destination) 間の交通需要量を 2次元の行列形式で表示する OD 表の形で表される。

交通需要は、土地利用パターン (Land-use pattern) に大きく影響されると考えられる。例えば、住宅密集地からは、近隣の工業地帯への通勤交通、および、商業区域への買い物交通があると考えられる。したがって、OSM 等の地理情報データを用いて土地利用パターンを適切に抽出することで OD 分布の推計ができると考えられる。

地点 i から地点 j までの交通需要 T_{ij} が、関数 f および g を使用して、以下のような一般化された重力モデルとして表されると仮定する。

$$T_{ij} = g(f(a_i), f(a_j), d_{ij}) \quad (1)$$

ただし、 a_i 、 a_j は、それぞれ、地点 i と地点 j の「雰囲気」すなわち周辺の土地利用パターンを要約した情報を表す d 次元のベクトルであり、 d_{ij} は、地点 i から地点 j までの一般化された「近さ」すなわち移動コストを要約した情報を表す e 次元のベクトルである。

(2) グラフ畳み込みニューラルネットワーク (GCN)

近年の機械学習分野における深層学習技術の進歩は目覚ましいものがある。深層学習は、初期には、主に画像処理に適用され、畳み込み層を多数積み重ねる畳み込みニューラルネットワーク (Convolutional Neural Network, CNN) が非常に成功をおさめた²⁾。これをうけて、最近、CNN を画像処理以外の分野に適用する試みが行われている。なかでも、グラフ構造に畳み込み層を適用するグラフ畳み込みネットワーク (Graph Convolutional Network, GCN) が注目されている^{3,4)}。本研究では、GCN を交通ネットワークグラフに適用し、式(1)の関数 f および g をエンドツーエンドで学習する。

交通ネットワークグラフ $G = (\mathcal{V}, \mathcal{E}, \mathcal{R})$ を考える。ただし、 $v_i \in \mathcal{V}$ は交通ネットワークグラフのノード (交差点) を、 $(v_i, r, v_j) \in \mathcal{E}$ は、交通ネットワークグラフのノード v_i からノード v_j へのエッジを表す。各エッジには、リンクの種類 (道路、鉄道路線、バス路線) を表すラベル $r \in \mathcal{R}$ が付加されている。また、各ノード v_i には、その交差点周辺 100m 以内に存在する POI の数を種類 (住宅・学校・お店・駅・公園・図書館...) 毎に表す 1次元のベクトル (POI ベクトル) x_i が対応している。

交通ネットワークグラフ $G = (\mathcal{V}, \mathcal{E}, \mathcal{R})$ に以下で定義

される Relational-GCN⁵⁾ を L 回適用する。

$$h_i^{(l+1)} = \sigma \left(\sum_{r \in \mathcal{R}} \sum_{j \in \mathcal{N}_i^r} W_r^{(l)} h_j^{(l)} + W_o^{(l)} h_i^{(l)} \right) \quad (2)$$

ただし、 $h_i^{(0)} = x_i$ すなわち、1 層目への入力、は、POI ベクトルである。ここで、 $\mathcal{N}_i^r = \{j \mid (v_i, r, v_j) \in \mathcal{E}\}$ はノード v_i からラベル r のエッジを通じてつながるノードの集合を、 σ は ReLU 関数を、 $W_r^{(l)}$ と $W_o^{(l)}$ は学習パラメータである畳み込みの重みを表す。

Relational-GCN の 1 層目への入力 x_i は、それぞれのノード (交差点) のすぐ近く (100m 以内) に存在する POI (住宅・学校・お店・駅・公園・図書館...) のみの情報であるのに対して、交通ネットワークグラフに

Relational-GCN を L 回適用した出力 $h_i^{(L)}$ は、 L ホップ先のノードの情報まで集約したものとなっている。Relational-GCN をモデルに組み込んで、交通需要分布を学習させることで、宅地・工業地帯・商業地帯といった、各地点の周辺の「雰囲気」を抽出することができ、精度の高い推計が可能になると期待される。

3. 実験

(1) TNTP データセット

実際に、交通需要分布の学習を行うには、教師データ、すなわち、実際の交通需要分布のデータが必要である。本研究では、Transportation Network Test Problems (TNTP) データセット⁶⁾を用いて学習およびテストを行った。TNTP データセットは、ネゲヴ・ベン＝グリオン大学のチームが研究用に公開している交通需要分布のデータセットであり、世界中のいくつかの都市における交通ネットワーク (各リンクの長さや太さの情報を含む) と、ネットワーク上の主要なノード間の交通需要量 (OD パターン) とで構成されている。本データセットは、基本的に、交通配分問題の研究に用いる想定で公開されているものであるが、いくつかの都市については、ネットワークの各ノードに緯度・経度の情報が付されており、交通需要分布の推計問題の教師データとして用いることができる。

今回、TNTP データセットの中から、Berlin-Mitte-Prenzlauerberg-Friedrichshain-Center データについて実験を行った。本データはベルリン市の中心部のおよそ 6km 四方の情報を含む。この範囲内にある 98 個の地点 (Zone) 間について交通需要量 (OD) の数値が付されている。なお、この数値は自動車での移動の需要量であり、公共交通機関での移動は含まれていない。

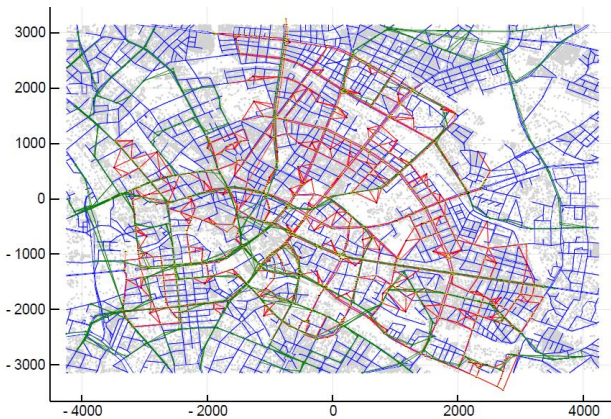


図-1 構築した交通ネットワークグラフ

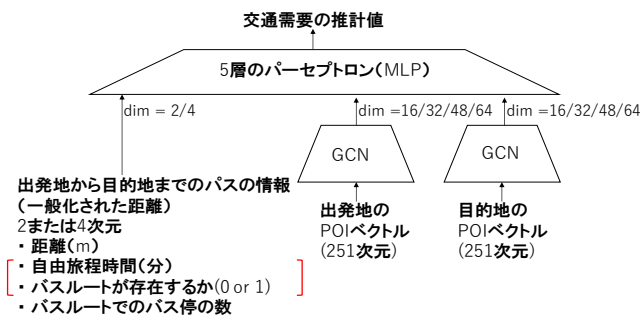


図-2 モデル構成

(2) 交通ネットワークグラフの構築

Berlin-Mitte-Prenzlauerberg-Friedrichshain-Center データに対応する範囲の OSM データを取得し、交通ネットワークグラフを構築した。図-1に構築したグラフを示す。図中の青線が道路ネットワーク、緑線がバスネットワークを表す。なお、道路網が、あまりに煩雑にならないように OSM 上の重要度がクラス 5 以上の道路のみを抽出した。赤線は、TNTP データセットに含まれているネットワークであるが、今回の学習には実験ではしていない。

図中の灰色の点は、OSMから抽出した 40,582 個の POI の位置を表す。各 POI は、OSM 上のタグをもとに、住宅 (building.residential)、レストラン (amenity.restaurant)、衣服店 (shop.clothes) 等の 251 種類に分類した。前述のように、各 POI を付近の交差点に対応させることで、道路網グラフのノード (交差点) 毎に 251 次元のベクトル x_i が付加されることになる。

(3) モデル構成

図-2は全体のモデル構成を示す。まず、出発地、到着地、それぞれに 251 次元の POI ベクトルを入力として Relational-GCN を適用して、各地点の周辺の「雰囲気」を表すベクトル (16次元、または32次元、または64次元) を得る。この2つのベクトルと、出発地から目的地までの一般化された「近さ」を表す2次元または

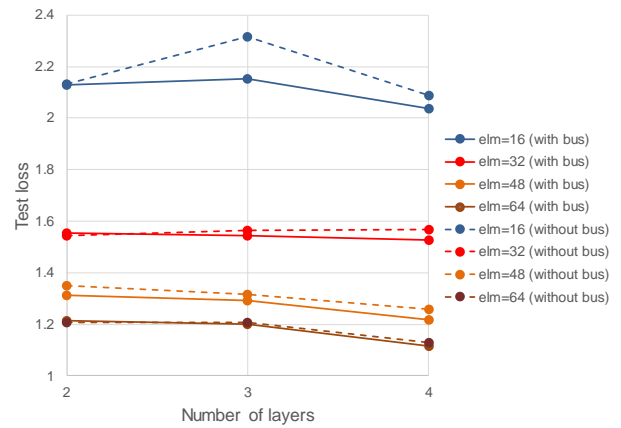


図-3 テストセットでのロス

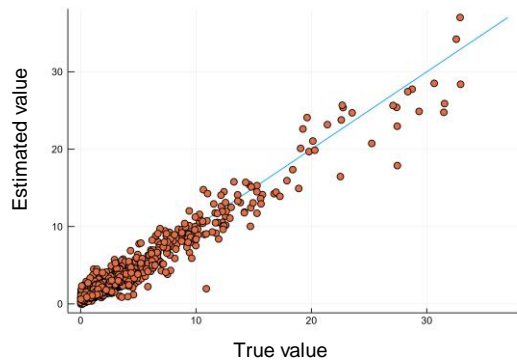


図-4 テストセットでの予測精度

4次元のベクトルを5層の多層パーセプトロン (Multi-Layer Perceptron, MLP) に入力して、最終的にスカラー値で交通需要の推定値を得る。

なお、出発地から目的地までの一般化された「近さ」は、交通ネットワークグラフ上の最短距離、および、TNTP データセットのに含まれる 2 点間の自由旅程時間 (Free flow time)、すなわち、渋滞が一切発生していない場合に出発地から目的地までの移動にかかる時間、の 2 つの変数と、さらに追加で、出発地から目的地までにバスルートが存在するかどうかを 0 または 1 で表すダミー変数、および、バスルートが存在する場合にはルート上のバス停の数の 2 つの変数とで構成される 2 次元または 4 次元の変数である。これは、バスルート (公共交通機関) の存在が、自動車の交通需要に影響を及ぼすかどうかの検証のために、バスルートの報を用いない場合 (2次元) と、バスルート情報を用いる場合とで、予測精度を比較することを意図している。

(3) 実験結果

実験は、TNTP データセットに含まれる全 9,501 個の OD ペアを無作為に 8 割 (7,604 個) のトレーニングセットと 2 割 (1,901 個) のテストセットに分割し、トレーニングセットを教師データとしてモデルの学習を行った。学習は、2乗誤差をロス関数として、Adam法によってパ

ラメータを最適化することで行った。学習期間は 5 万 epoch とした。

図-3 は学習後のパラメータを用いてテストセットでのロスを評価した結果を表す。図中の横軸は、Rational-GCN の層数を、縦軸がテストセットでの平均ロスを表す。また、グラフの色は、Rational-GCN の出力（各地点の「雰囲気」ベクトル）の次元数を表す。なお、GCN の各層の次元は全て共通とした。また、図中の破線はバスルート情報を使わなかった場合を、実線はバスルート情報を使った場合を表す。

図-3 より、Rational-GCN の出力層数を増すことで予測精度が向上すること、また、バスルート情報を使うことで、若干、予測精度が向上することがわかる。

図-4 は、最も精度が高かった層数 4、Rational-GCN の出力次元数 64、バスルート情報を使用した場合について、テストセットの 1901 個の OD 値について、横軸に真の値、縦軸の予測値をプロットしたものである。図中の 45 度線近くに点が集まっており、高い精度で予測できていることがわかる。

3. まとめと今後の展望

オープンな地理情報データベースである OpenStreetMap (OSM) から交通ネットワークグラフを構築し、グラフ畳み込みネットワーク (GCN) を適用することで、交通需要分布の推計を行った。

なお、本手法は、交通需要分布の予測に限らず、注目地点の周辺環境に依存すると考えられる指標全般に適用できる可能性がある。今後、様々な指標の予測を試みたい。

参考文献

- 1) OpenStreetMap, <https://www.openstreetmap.org/>.
- 2) Y. LeCun, et. al., Gradient-based learning applied to document recognition, Proceedings of the IEEE, 86 (11), pp. 2278–2324, 1988.
- 3) T.N. Kipf and M. Welling, Semi-Supervised Classification with Graph Convolutional Networks, ICLR, 2017.
- 4) M. Schlichtkrull, et. al., Modeling Relational Data with Graph Convolutional Networks, 2017.
- 5) Transportation Networks for Research Core Team, Transportation Networks for Research, <https://github.com/bstabler/TransportationNetworks>. Accessed March 2019.
- 6) OpenStreetMapX.jl, <https://github.com/pszufo/OpenStreetMapX.jl>