

時空間トランスフォーマーと人流予測

東京理科大学 経営学部 ビジネスエコノミクス学科 教授 あんどう しん 安藤 晋

はじめに

深層学習は、2024年のノーベル物理学賞を受賞した Hinton と Hopfield の研究¹⁾をはじめとするいくつかのブレークスルーや、その他多くの発見・改良の積み重ねにより、幅広い分野で利用される人工知能 (AI) の基盤的技術となっている。近年でもトランスフォーマー、生成的敵対的ネットワーク、ノイズ除去拡散モデルなどの革新的なモデルの開発は続いており、その速度は驚異的とも言える。

本稿ではトランスフォーマーを用いて経済活動と関わるの深い人流データの分析を行った内容を紹介する。2, 3 節では人流データとトランスフォーマーの自己注意メカニズムについて解説し、4 節では地理時空間データへのトランスフォーマーの適用方法を紹介する。5 節ではタクシー流量を用いた数値実験の結果を示す。

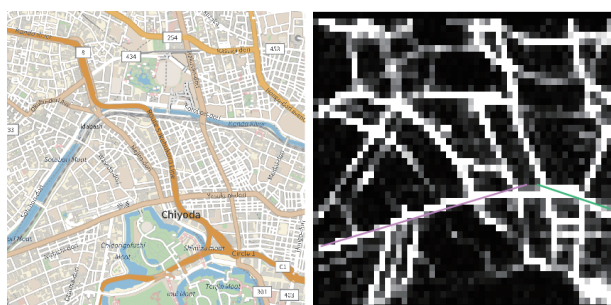
人流データ

人流 (Crowd Flow) データはモバイルデバイスや IoT 等の技術的発展と普及により収集基盤が整い、配車アプリなど応用面の関心も近年高まっている²⁾。輸送分野における重要な問題である交通量予測 (Traffic Forecast) では人や車両の数を各地点で発生・観測した量を扱うのに対し、人流データは配車アプリやシェアサイクルのシステムで収集される 1 件毎に出発 (貸出) と到着 (返却) の位置時間を記録したトリップのレコードを集計したものである。トリップレコードは

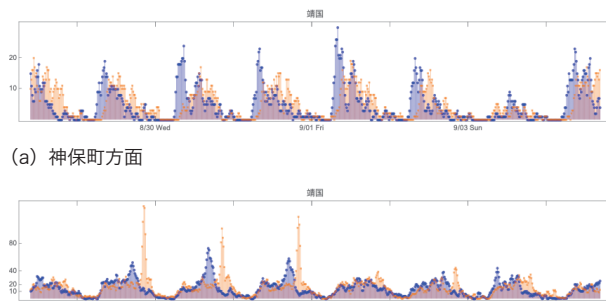
また、IoT デバイスにより収集された GPS 情報と他の属性情報を組み合わせて抽出することもできる。例えばタクシーで記録される GPS 情報をメーターの属性を基に分割すると、タクシーの乗車区間を抽出することができる³⁾。【図 1 (a)】は富士見校舎周辺の地図であるが、この範囲の 3 週間のタクシーの乗車位置をプロットすると【図 1 (b)】に示すように付近の主要道路が浮かび上がるほどの規模がある。

人流データの流量の集計にあたっては空間的・時間的な依存を捉えることを踏まえて分割やスケールを検討する必要がある。関連研究では時間的な分割は時間・日・週といった時間スケールで行い、空間的な分割については一定サイズの格子状に分割するのが標準的である。後者の理由として、集計した流量がテンソル形式となる利便性が大きい点があげられる。ただし、応用面では格子状以外の地理的な特徴を活かした、例えば道路の形状に則した空間的な分割なども有用である。【図 1 (b)】のプロットから線分抽出を用いて道路区間を抽出し、桃色と緑色の線分の一定の距離範囲内で 1 週間分の乗降車数を集計したものがそれぞれ【図 2 (a), (b)】である。このように同じ道路でも区間によって乗降車の大小や時間的關係が異なることから、分割の区切りや粒度は分析に大きく影響する。

流量の集計にも 2 つの代表的な形式がある。従来は In-Out (IO) 流量テンソルが中心であったが、近年では Origin-destination (OD) 流量テンソルが多く用いられる²⁾。IO 流量は格子毎の合計入出量であり、OD 流量は出発地と到着地の組み合わせ毎の流量であ



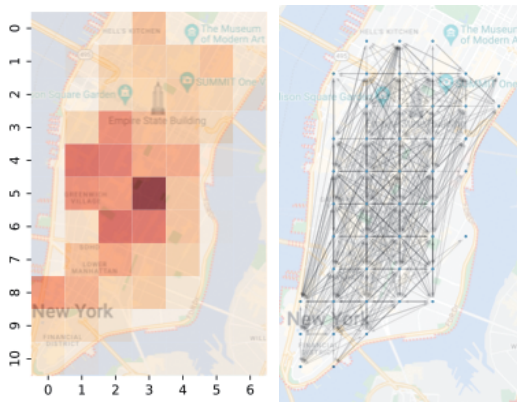
【図 1】校舎周辺地図 (a) とタクシー乗車プロット (b)



(a) 神保町方面

(b) 九段・市ヶ谷方面

【図 2】靖国通りタクシー乗降車数時系列



【図3】IO 流量 (a) と OD 流量 (b) の例

る。【図3 (a)】はニューヨークのバイクシェアサービスのIO 流量テンソルを表すヒートマップを地図に重ね表示し、【図3 (b)】はOD 流量テンソルを同様に表す。【図3 (a)】では格子の色の濃さが流入量を、【図3 (b)】では矢印の太さが移動量を表す。

空間計算量のオーダは格子数 n と時間長 T に対して IO 流量のオーダが $O(nT)$ である一方、OD 流量では $O(n^2T)$ となる。関連研究は情報量が多い OD 流量を用いるモデルがより高い予測性能を示すことを報告しているが、この計算量はトランスフォーマーを含む大規模モデルに対する制約となっている。

トランスフォーマーと自己注意

トランスフォーマーは ChatGPT に代表される大規模言語モデルの枠組みとして知られているが、時系列予測においても有効であり、電力需要や交通量など経済活動に関わるデータを用いた予測にも応用されている。トランスフォーマーには数多くの技法が実装されているが、中でも自己注意 (Self-Attention) メカニズムはその特徴的な仕組みといえる。自己注意は画像処理で用いられる畳み込みの仕組みと比較すると、近傍の要約だけでなくデータ中広い範囲の依存性を学習できる点を特徴とする。そのような仕組みは多くのパラメータを要し、訓練に要するデータと時間も増大する。

自己注意の基本的な構造は以下の式で表される。本稿では、これを数多い拡張と区別するため Vanilla 自己注意 (VSA) と呼ぶ。

$$\text{VSA}(X) = \sigma \left(\frac{QK^T}{\sqrt{D}} \right) V \quad (1)$$

自己注意に関する解説は多く存在するが、以下では本稿で扱う人流予測の文脈での (1) の解釈を示す。

X は入力であり、この問題では各地点の OD 流量をベクトル上に並べた行列となる。【図3 (b)】のような OD 流量がある時、各行が【図3 (b)】中の1つの格子点の流量の時系列を表すベクトルとなる。 Q, K, V はそれぞれクエリ・キー・バリューと呼ばれる、 X をさらに線形変換した行列である。それぞれの行ベクトルもまた、1つのマスや格子の表現となっている。自己注意はこれら3つの生成表現を用いて相互に関係する部位を発見し、利用する仕組みといえる。

富士見校舎の近くを走る靖国通りの九段方面へのタクシーの流入量を予測する問題で考えると、 Q と K の行列積は九段方面と他の地点の関連の強さを対応する行ベクトルの内積によって測っている。例えば神保町と市ヶ谷方面の流量がターゲットの九段方面と関連するならば、九段方面のクエリ行と神保町、市ヶ谷方面のキー行の内積が大きくなるよう、 Q, K を生成するパラメータは更新される。

(1) の右辺では、行列積 QK^T をスケーリング、正規化した上で V との行列積をとっている。 V は各地点の流量の予測に用いる要素を含む行ベクトルで構成されるので、この行列積は関連の強い地点に対応する V の要素を混合 (線形結合) する処理といえる。

九段方面の流量の予測は $\text{VSA}(X)$ の対応する行を取り出し、さらに線形変換や活性化などの処理を加え求める。したがって、自己注意の学習は $Q \cdot K$ で関連する場所を抽出するのに加え、 V を関連する場所のベクトルを結合した時に予測が合うよう更新する。

なお、時間的な自己注意においては、 $Q \cdot K$ の行が t 時刻前等の時点を表すベクトル表現に置き換わり、同様の仕組みで関連する時刻を抽出する。

地理的時空間トランスフォーマー

画像処理分野では動画のような平面と時間の次元を持つデータを扱う時空間 (Spatio-temporal) トランスフォーマーが、補完や予測などの問題に用いられている。人流データを【図3 (a) (b)】のような画像が時間的に変化する動画データとみなすと、時空間トランスフォーマーに直接的に入力として与えて予測モデルとして用いることも可能である。このような用い方はオフザシェルフと呼ばれる。

オフザシェルフのモデルは画像処理や自然言語処理など汎用的なデータを用いた事前学習が可能な分野では非常に効率的である一方、本来とは異なるデータへの応用では問題を生じやすい。例えば2節に示した

通り、OD 流量の空間計算量は格子数 n の 2 次のオーダーである。自己注意の行列積 QK^T を求める計算量は、そのさらに 2 乗のオーダー $O(n^4)$ となるため、オフザシェルフの時空間トランスフォーマーではスケール性が低く、非常に小規模のデータしか扱えない。

加えて、人流データの地理的な特性についても注意が必要である。【図 2 (a)】の格子間の依存性は動画を構成する画像の画素同士よりも安定しており、時間的な変化も小さい。筆者らは、上記の考察を踏まえ、自己注意の複雑性を抑えてスケール性の課題に対処し、OD 流量の地理的特性を考慮した予測モデル地理的時空間トランスフォーマー (GTFormer) を開発した²⁾。

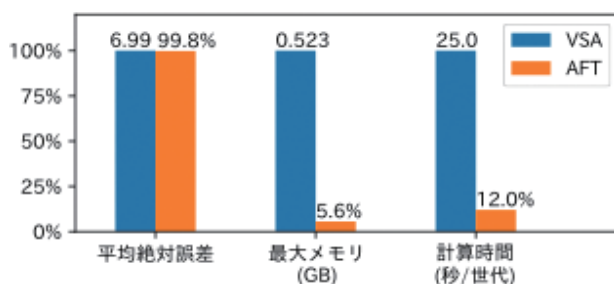
効率的な注意機構

GTFormer は自己注意に替えて Attention-Free Transformer (AFT) と呼ばれるメカニズムを採用している。その仕組みは (2) で表されるように、行列積を用いず、要素毎の積のみ用いることで計算量を行数の線形のオーダーに抑えている。なお、 $\odot$ は行列同士の要素毎の積、 $\circ$ は行列とベクトルの行毎の積を表す。

$$\begin{aligned} \text{AFT}(X) &= \sigma(Q) \circ \text{Softmax}(K, V) \\ \text{Softmax}(K, V) &= \frac{\sum_t (\exp(K_t) \odot V_t)}{\sum_t \exp(K_t)} \quad (2) \end{aligned}$$

(2) では自己注意で Q と K の一致により捉えている依存性を Q のみで捉える。これは、地点間の相関が固定されたものとみなしたものと見える。 K はバリューを調整するための重みの役割のみを持つ。AFT は計算量削減のために提案された手法だが、画像処理タスクにおいては性能のトレードオフも大きい手法である。一方、地点間の関係の大きく変化しない人流データではより有力と考えられる。

筆者らは AFT の効果を測定するための切除実験 (Ablation Study) を NY のタクシー流量データを用いて行った。切除実験は機械学習分野で一般的な新しい仕組みの効果を測る手続きで、対象を実装した場合と除いた場合の予測を比較評価する。【図 4】は平均絶対誤差 (MAE) と最大使用メモリ、1 世代の学習に要する時間の比較を示す。GTFormer の AFT を VSA で置き換えた場合を VSA、本来の GTFormer を AFT と表示している。GTFormer は MAE に差がなくメモリと計算時間をそれぞれ 94.4%、88% 削減し、性能を維持しつつ、計算量を大きく削減したことが分かる。



【図 4】AFT 切除実験の評価

反復的時空間エンコーダ

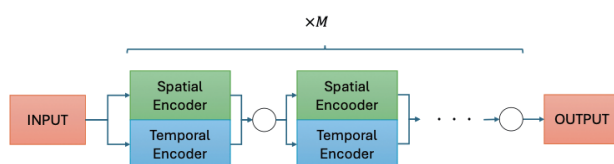
多くの時空間トランスフォーマーはまず、特徴抽出器により【図 3 (a) (b)】に見られるような空間的なパターンをベクトル表現に変換し、その表現における【図 2 (a) (b)】のような時間依存を時間エンコーダにより捉え予測を生成する。このような処理は各モデルが直列的に接続する構造で実装される。直列的構造は大規模な画像データセットにより抽出器を事前学習しやすいことを前提とした、多くの画像処理モデルで一般的である。

しかし、人流データのように事前学習のためのデータが十分でなく、直列構造のモデルをランダムな初期値から学習する場合、出力からより遠い抽出器が適合不足に陥りやすい。同時に、特徴抽出器が十分に学習されないまま時間エンコーダのパラメータが早期収束してしまう恐れがある。さらに、時空間の複合的な依存性は直列的な構造では捉えきれない場合もある。

このような問題を回避するため、GTFormer では反復的時空間エンコーダを実装している。反復的エンコーダは【図 5】のような構造で、自己注意による空間的なエンコーディングと時間的なエンコーディングをそれぞれ並行に行った上で線形結合する。この処理を繰り返すことで時空間の複合的な依存性を捉えることも可能である。この構造の効果は、次節に紹介する従来モデルとの比較により評価している。

タクシー需要予測への応用

本節では Yellow Taxi Trip Record データセットを



【図 5】反復的時空間エンコーダの構造

用いた数値実験の結果を示す。このデータセットは NYC Yellow Cab Taxi^(注1)にて公開されているレコードから、2016年の1月から3月までの期間の記録を1時間毎のOD流量に集計したものである。機械学習における成果の一つとして、共通ベンチマーク問題において State-of-the-art (SOTA) を達成することが挙げられる。SOTA は分野内の技術的に最先端の手法を指すが、一般的にベンチマークや指標が複数あるため、単一の手法やモデルが SOTA といえない場合も多い。このため、SOTA を達成するにはできるだけ多く、幅広い種類の既存 SOTA 手法と比較し、優位性を示す必要がある。ここでは4つの既存 SOTA 手法に加え、基本的な LSTM を対象とした比較結果を示す。

【図6】に絶対平均誤差 (MAE) の比較を示す。各手法は MAE の降順に並び、MAE が2番目に小さい手法には MAE 値を、GTFormer には2位の MAE に対するパーセント値を棒の上に表示している。GTFormer は最も誤差が小さく、2位の手法に比べて 63.4% の誤差削減を実現した。この結果は GTFormer の設計の妥当性を示すものと考えられる。

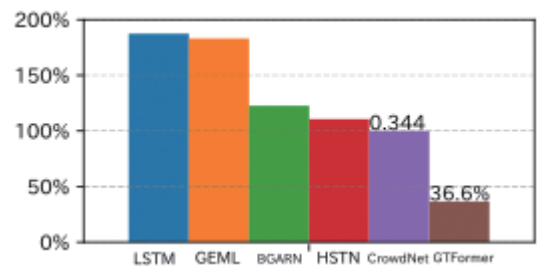
【図7】にはある地点の流入量の実測・予測時系列を示す。緑線の2位の SOTA 手法と比べ、橙線の GTFormer は高いピークにおいて青線の実測値によりよく適合していることが分かる。

最後に IO 流量と OD 流量を地図上に可視化した結果を【図8】に示す。格子の色は IO 流入量の大きさを表す。矢印は一定以上の大きさの OD 流量を表し、またその太さは流量の大きさを反映している。

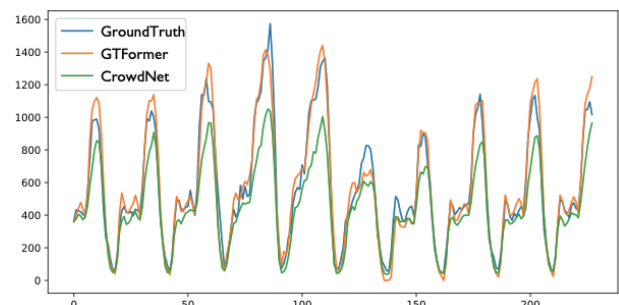
IO 流量に関しては、ミッドタウン、ブルックリンハイツ、イーストハーレム、JFK と LaGuardia 空港を含む格子の流入量が多く、GTFormer がこれらを正確に予測していることが分かる。また、OD 流量は、2つの空港、ブルックリンハイツ、ブロンクスなどへの大きな矢印が再現されている。一方で、実測値に数多く見られる頻度が少なく、目的地が広範囲に分散している流量については予測が難しいことが分かる。

おわりに

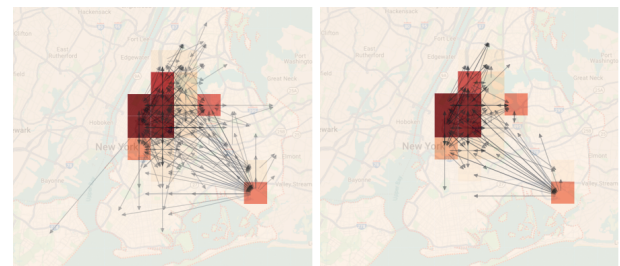
様々な分野に応用されるトランスフォーマーでも、新しい応用において SOTA を達成するには様々な工夫を導入する必要があることを示した。数値実験において SOTA を達成している一方で、小規模流量の予測が課題として残されている。将来の研究課題としてはイベント情報の導入による精度向上やエンコーダ・デコ



【図6】 平均絶対誤差の比較



【図7】 流量時系列の予測値と実測値の比較



(a) 観測値 (b) 予測値

【図8】 Yellow Taxi OD 流量の観測値と予測値

ーダを用いた長期予測などが実用的なテーマとして挙げられる。さらに、予測を人間にとって理解しやすい自然言語形式の説明や提案に変換する大規模言語モデルによるデータ to テキスト生成といったチャレンジングだが興味深い課題もある。

注

(注1) <https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page>

【参考文献】

- 1) John J. Hopfield and Geoffrey E. Hinton. Foundational discoveries and inventions that enable machine learning with artificial neural networks. Nobel Prize in Physics, The Royal Academy of Sciences, 2024.
- 2) Sohei Kodama and Shin Ando. GTFormer: A geospatial temporal transformer for crowd prediction. In 2024 IEEE International Conference on Knowledge Graph (ICKG), pp. 167-176, 2024.
- 3) 山本綾華, 楊博健, 安藤晋. 深層ニューラルネットワークを用いたタクシー乗車数のスパイク予測. 人工知能学会全国大会論文集, pp. 4Rin188-4Rin188. 一般社団法人 人工知能学会, 2020.