

# エリア毎の滞在利用情報を用いた移動遷移モデルの分析

庄子 和之, 米澤 拓郎, 河口 信夫 (名古屋大学)

Analysis of Trajectory Modeling using Location Usage Information for Each Area

Kazuyuki Shoji, Takuro Yonezawa, Nobuo Kawaguchi (Nagoya University)

## 1. はじめに

近年, GPS (Global Positioning System) 機能を備えたスマートフォンやウェアラブル端末の普及により, 位置情報履歴の収集が容易になった. 位置情報履歴は, 個々のユーザの日々の行動を反映したものであり, ユーザの特徴抽出から都市計画まで様々なことに応用できる.

実際に位置情報履歴を活用するには, その目的に沿ったモデリングが必要になる. 例えば, 滞在遷移をモデリングした場合, 滞在所を座標 (緯度経度) で表現する「座標遷移モデル」やラベル (住宅街, 飲食店など) で表現する「ラベル遷移モデル」がある. モデリングを行うことでデータを抽象化し, ユーザ毎に意味がまとまりやすくなるというメリットがある. しかし, ラベルを用いることには限界がある. 例えば, ラベルが POI (Point of Interest) の場合, ラベルが「一意に決められない」または「振れない」という問題に直面する. 「一意に決められない」とは, 衣料品や日用品, 食料品といった様々な領域をカバーした店舗がこれに該当する. 「振れない」とは, そもそもデータセットに登録されていない POI がこれに該当する.

そこで本研究では, まず初めに, 大量の位置情報履歴を活用し滞在所を分散表現 (ベクトル) で表現する手法を示す. そして, それによって作られたベクトル遷移モデルを用いてユーザを移動遷移の類似性からグループ分けする方法を提案する.

## 2. 関連研究

[1] は, 様々な要素 (滞在目的や天気, 移動手段など) が含まれたデータセット使用し, ユーザの分散表現を作成している. この分散表現を基に, ユーザ同士の行動の類似性を測定している. [2] は, 滞在目的を事前に収集した POI(Point of Interest) 情報を基に類推し, 滞在目的による滞在遷移モデルを作成している. この滞在遷移モデルを基に, ユーザの属性の推定を行っている.

上述した関連研究では, 滞在目的がラベルとして表現 (自宅, 職場, 飲食など) されていて, その種類はせいぜい 10 種類である. ユーザ間の類似度推定やユーザの属性推定といっ

表 1 Time info for staying

曜日	平日, 休日
到着時間	0:00~1:59, 2:00~3:59, ..., 20:00~21:59, 22:00~23:59 (2 時間区切り)
滞在時間	~29 分, 30~59 分, 60~119 分, 120~239 分, 240~359 分, 360 分~

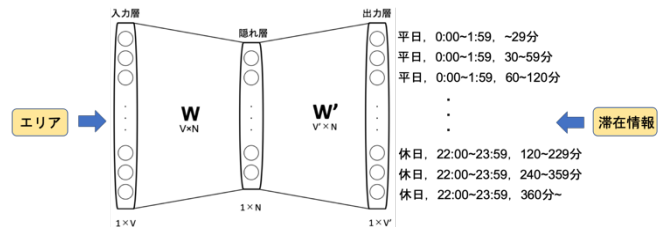


図 1 Remodeled Word2Vec

たことを詳細に行いたい時, このラベルの種類だけでは, ユーザをモデリングするのに不十分であると思われる. 本研究では, ラベルではなくベクトルでユーザの移動遷移を表現する. そのため, 様々なユーザの特徴を遷移モデルに反映できると考えている.

## 3. 提案手法

### 3.1. エリアの分散表現と LU 遷移モデルの作成

エリアの分散表現の作成では, Word2vec の Skip-gram モデルに改良を加えたものを使用する[3]. 入力,  $N$  次元ベクトル( $N$ :エリア数)であり, あるエリアに対して対応する箇所にフラグが立った one-hot ベクトルとする. 出力は, 滞在に関する時間情報(曜日, 訪問時間帯, 滞在時間)を考慮したベクトルで, 対応する箇所にフラグが立った one-hot ベクトルとする. 滞在に関する時間情報には, 表 1 に示した情報を考慮する. Word2vec の Skip-gram の改良モデルを図 1 に示す. これにより, この改良したモデルで学習させることで, 時間情報のみからエリアの分散表現が作成可能となり, 時間的に同じような使われ方をしているエリア同士は近くに, 全く異なる時間的使われ方をしているエリア同士は遠くに, ベクトル空間上で配置できる.

本研究では、プログウォッチャー社のプロファイルパス ポートのデータを用い、中部地方の 10km 四方のエリアを 対象として、滞在情報が取得可能な 2 か月間約 2160 万レコードの位置情報履歴を利用した。このエリアを 100m 四方のメッシュで 10000 個に分割した。しかし、実際に使用するのは滞在が 10 回以上行われたメッシュに限定した。その結果、約 1800 個のメッシュにまで絞れ、分散表現の作成にかかる計算量も減らすことができる。

ここで LU (Location Usage) というものを定義する。LU とは、作成した約 1800 個のメッシュの個々の分散表現のことである。各メッシュの分散表現は、多くのユーザのエリアへの滞在に関する情報が反映されたものであるため、このように名付けた。

次に各ユーザの滞在遷移を LU 遷移モデルで表現する。今回は簡単のため、約 1800 個の LU を k-means++ を用いてクラスタリングし、ある程度 LU を塊で捉えることで、移動遷移が類似しているユーザを見つけやすくした。クラスタ数は 6 に設定し、そのクラスタリング結果は図 2 に示す。これは、各クラスタの滞在時間別の人数の分布である。左側が平日について、右側が休日についてである。横軸は時間を表している。縦軸は人数を表している。縦軸に関して、各クラスタに所属しているメッシュの数には違いがあるため、メッシュの数で割ること正規化を行っている。また、平日と休日は日数が違うため、それぞれの収集期間内の日数で割ること正規化を行っている。この処理により、クラスタ間でのグラフの比較と平日休日間でグラフの比較が可能になる。また、長時間滞在を行っている人は、その前後の時間帯にも現れていることに注意してほしい。例えば、10:00 から 12:00 の滞在进行した人がいた場合、10 から 12 の間の目盛り全てに同一人物がカウントされているということである。この図を用いることで、各クラスタに含まれる LU の特徴が解釈可能になる。例えば、クラスタ 1 に関して、8 時くらいから 2 時間以上の滞在を開始した人が多くなっている。さらに、夜中と休日を見るとあまり人は訪れていないことがわかる。そのため、このクラスタは「オフィス街」のような時間的使われ方がされている、という特徴を持った LU が集まっていると解釈できる。クラスタ 2 に関しては、日中どの時間帯でも短時間の滞在をする人が多く、平日に比べて休日の方が人数が多い。そのため、このクラスタは買い物で利用される「店舗」のような時間的使われ方がされている、という特徴を持った LU が集まっていると解釈できる。クラスタ 3 と 4 に関しては、夜中と早朝の長時間の滞在が目立ち、日中はその数が減少している。さらに、平日と休日に変化が小さい。このことから、このクラスタは「住宅街」のような時間的使われ方がされている、という特徴を持った LU が集まっていると解釈できる。クラスタ 5 に関しては、昼食と夕食の時間帯に人数が増え、短時間から中時間の滞在が目立つことから、「飲食店」のような時間的使われ方がされている、という特徴を持った LU が集まっていると解釈できる。このように LU をある程度まとまりを持たせて、そのまとまり間の遷移を LU 遷移とすることで、移動遷移が類似しているユーザを見つけやすくなると考えた。また、LU 遷移には「-1」という

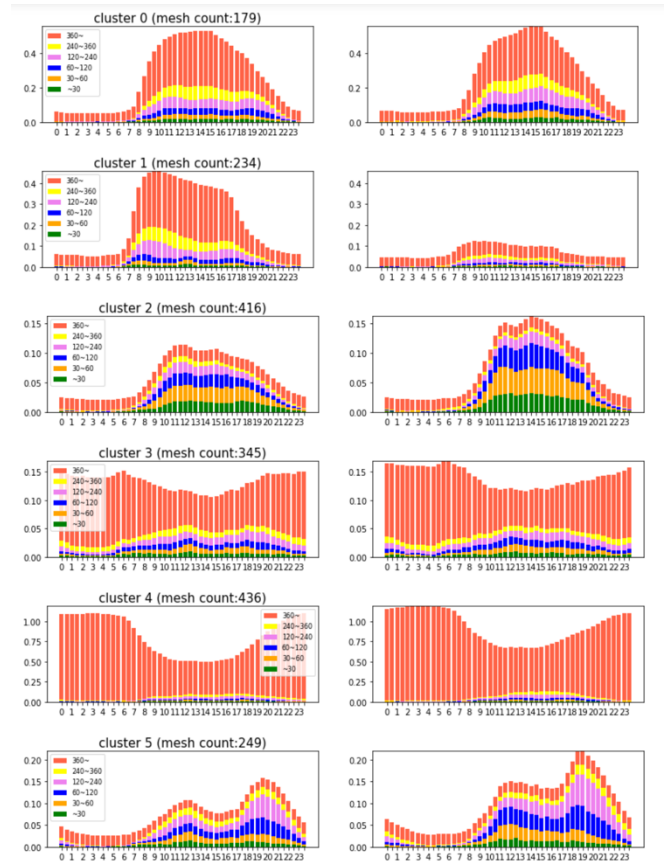


図 3 Result of clustering for LU

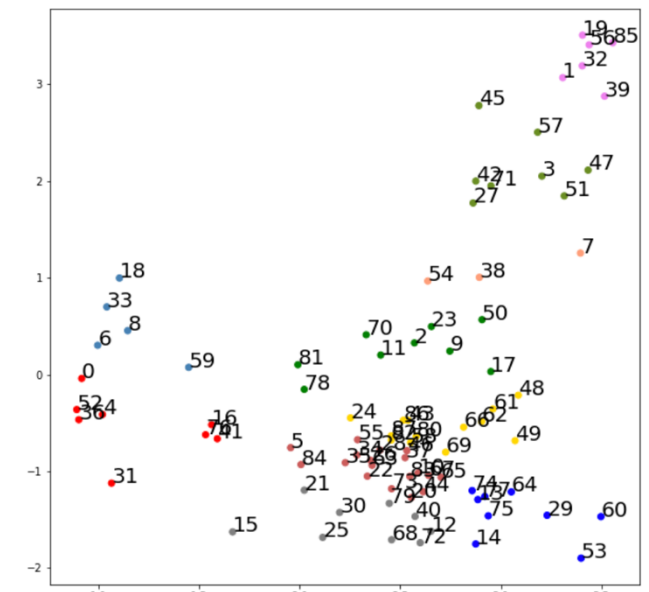


図 2 Result of clustering for users based on LU trajectory

値を持つユーザもいるが、これは滞在回数が 10 回以上という絞り込みをかけた時に弾き出されたメッシュへの滞在を示している。そのため、LU 遷移はクラスタ数 + 1 の 7 つの数値 (-1 ~ 6) の遷移となっている。例えば、あるユーザの LU 遷移は「1→2→1→2→1→2→1→3→...」となっていて、また別のユーザの LU 遷移は「-1→4→5→4→4→0→5→4→...」となった。

### 3.2. LU 遷移モデルからユーザをクラスタリング

今回は、LU 遷移モデルからユーザのクラスタリングが行えるかどうかを確認することが目的であるため、処理時間などを考慮し、使用するユーザ数と期間を少なくした。具体的には、ユーザは 1 日の行動がある程度追えるだけのデータ量がある 88 人に絞り、期間は 2 週間とした。

LU 遷移モデルからユーザの特徴を抽出し、その特徴を基にクラスタリングを行うために、各ユーザのベクトルで表現する必要がある。そのために、Okapi BM25[4] を活用することを考えた。これは、TF-IDF の特徴である、文書に含まれる単語数が多いほど TF 値が小さくなり、複数の文書から重要度の高い単語を抽出して絶対評価で TF-IDF 値を比較する場合、文書ごとの単語数の差による影響が大きくなってしまい、という欠点を克服したものである。各ユーザの LU 遷移を文書に、各ユーザの LU 遷移から作成できる部分遷移を単語に見立てることで、これを可能にした。そして、この部分遷移の作成には、N-gram を用いた。N-gram の N の値を 1 から 4 まで変化させた。この時にできる LU の部分遷移は、1126 個であった。つまり、ユーザ数  $\times$  1126 の行列を Okapi BM25 と N-gram を組み合わせることで作成した。

そして最後に、k-means++ を使ってクラスタリングを行い、ユーザの分類を行う。しかし、ここで作成した行列は疎行列であり、ユーザ間でまとまりを作るのは困難である。そこで、SVD を使って次元圧縮を行い、ユーザ数  $\times$  2 の行列まで圧縮し、これをクラスタリングした。クラスタ数は 10 に設定した (図 3)。あるクラスタ (ユーザ 12, 15, 21, 25, 30, 40, 68, 72, 79 が所属) は、部分遷移とし「1」と「4」の組み合わせが多いという特徴を持つユーザの集合であることを見つけた。LU のクラスタリング結果 (図 2) より、「1」と「4」はそれぞれ「オフィス街」と「住宅街」のような時間的使われ方がされている、という特徴を持った LU が集まっていると解釈したため、このクラスタに所属しているユーザは、「住宅街」と「オフィス街」の往復が多いという特徴を持っていることが推定できる。また、別のクラスタ (ユーザ 0, 4, 16, 31, 36, 41, 52, 76 が所属) は、部分遷移として「2」と「3」と「4」の組み合わせが多いという特徴を持つユーザの集合であることを見つけた。同様に LU のクラスタリング結果より、「2」と「3, 4」はそれぞれ「店舗」と「住宅街」のような時間的使われ方がされている、という特徴を持った LU が集まっていると解釈したため、このクラスタに所属しているユーザは、「住宅街」と「店舗」の往復が多いという特徴を持っていることが推定できる。以上より、自然言語処理の技術である Okapi BM25 と N-gram を用いて、各ユーザの LU 遷移を文章に、各ユーザの LU 遷移の部分遷移を単語に見立てることで、ユーザをベクトルで表現し、クラスタリングできることが判明した。

### 4. まとめ

本論文では、まず初めに大量の位置情報履歴を用いて作成したエリアの分散表現を使って、ユーザの滞在遷移をラベル

を使わず、ベクトルでモデリングする手法を提案した。その後、ユーザ毎の滞在遷移の類似度を、自然言語処理の技術である Okapi BM25 と N-gram を用いて分析する手法を提案し、この手法がユーザの滞在遷移の特徴を捉えられていることを確認した。

今後の課題としては、まずは時間を考慮するということが挙げられる。何時にそこへ滞在したかという情報は重要なものであり、これを考慮することで、より詳細なクラスタリングが可能になるだけでなく、LU 遷移からユーザの 1 週間の生活パターンを推定するなど、様々なことが可能になると考えられる。また他にも、クラスタリング手法に関しても改善が挙げられる。適切なクラスタリング手法を選択することは、マーケティングで本研究を活用する際、ピンポイントで精度良くユーザにレコメンドできるということが期待できるはずである。

### 謝辞

本研究は、JST CREST JPMJCR1882 および総務省 SCOPE に支援を受けたものです。また、データ提供にご協力頂きましたプログウォッチャー社に感謝します。

### 文 献

- [1] Andrea Esuli, Lucas May Petry, Chiara Renso, and Vania Bogorny. Traj2user: exploiting embeddings for computing similarity of users mobile behavior, 2018.
- [2] Wanlong Zhang, Xiang Wang, and Zhitao Huang. A system of mining semantic trajectory patterns from gps data of real users. Symmetry, Vol. 11, No. 7, p. 889, Jul 2019.
- [3] 庄子 和之, 廣井 慧, 米澤 拓郎, 酒田 理人, 河口 信夫. 滞在に関する時系列情報を用いたエリア毎の分散表現の検討. マルチメディア, 分散, 協調とモバイル(DICOMO2020)シンポジウム. pp.1008-1013
- [4] S.E.Robertson, S.Walker, S.Jones, M.M.HancockBeaulieu and M.Gatford. "Okapi at TREC-3", Proc. of TREC-3, pp.109-126 (1995)