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

庄子 和之[1] 廣井 慧[2] 米澤 拓郎[1] 酒田 理人[3] 河口 信夫[1,4]

概要：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．この時空間データは，個々のユーザの日々の行動を反映したものである．そのため，これらを分析し深く理解することは，様々なビジネス（混雑予測，都市計画，マーケティングなど）のチャンスを与えてくれる．近年，SNS（Social Networking Service）（Twitter，Facebook） の位置情報を付与した投稿や LBSNS（Location-Based SNS）（Foursquare） の存在により，位置情報履歴をラベル遷移で表現できるようになった．結果として，抽象化されたラベルを用いて，ユーザ同士の移動パターンの比較が行えるようになった．しかし，たかだか数百種類のラベルによって移動遷移がモデル化されることによる情報の損失が懸念される．本論文では，ユーザの移動遷移を，可能な限り多様な情報を保持するように高い次元数の分散表現を使った遷移に変換する手法を提案する．

A Study of Distributed Expression for Each Area using Time Series Information on Staying

KAZUYUKI SHOJI[1] KEI HIROI[2] TAKURO YONEZAWA[1] MASATO SAKATA[3]
NOBUO KAWAGUCHI[1,4]

1. はじめに

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．この時空間データは，個々のユーザの日々の行動を反映したものである．そのため，これらを分析し深く理解することは，様々なビジネス（混雑予測，都市計画，マーケティングなど）のチャンスを与えてくれる．

近年，SNS（Social Networking Service）（Twitter，Facebook） の位置情報を気軽に付与した投稿や LBSNS（Location-Based SNS）（Foursquare） の存在により，移動遷移をラベル遷移で表現できるようになった．これにより，座標（緯度，経度）とタイムスタンプのみからなるデータとは違い，移動遷移をユーザ間で共通したラベルに置き換えることで，行動パターンをユーザ同士で比較ができるようになった [1,2]．

しかし，位置情報履歴をラベル遷移モデルに変換してしまうことに対する懸念も存在する．それは，たかだか数百種類のラベルによって移動遷移がモデル化されることによる情報の損失である．ユーザの行動をラベル遷移に落とし込むことで，日本全国どこにいたとしても，様々な属性（年代，性別，年収など）を持つユーザがいたとしても，移動遷移は同じラベルを用いて表現されてしまう．例えば，あるレストランについて考えてみる．この POI（Point of Interest）のラベルは，「飲食店」である．あるユーザがこのレストランを訪れた場合，ラベル遷移には「飲食店」への滞在が記録される．しかし，このレストランは人気店であるという事実が存在した場合，ラベル遷移の中ではこの情報は消されてしまう．これは 1 つの例に過ぎないが，ユーザの移動遷移をこのような情報を含むような表現が可能になれば，ユーザの行動パターンの推測やユーザ同士の類似度推定がより精度良く行える．

本論文では，座標（緯度と経度）とタイムスタンプのみからなるモデル（座標遷移モデル）を用い，エリアを様々

[1] 名古屋大学大学院工学研究科 Graduate School of Engineering, Nagoya University
[2] 京都大学防災研究所 Disaster Prevention Research Institute, Kyoto University
[3] 株式会社ブログウォッチャー
[4] 名古屋大学未来社会創造機構 Institutes of Innovation for Future Society, Nagoya University

な情報を含んだ分散表現の形にする手法を提案する．これにより，ユーザの移動遷移をこの分散表現の遷移に変換することで，ラベル遷移よりもよりユーザの特徴が現れたモデリングが可能になる．

2. 関連研究

SNS などといったツールを用いて，座標遷移モデルをラベル遷移モデルに変換し，ユーザの特徴を推定したり，ユーザ同士の類似度を算出する研究は多く存在する．[2] は，滞在情報に「アクティビティ」の情報を追加し，ユーザの「行動のモチベーション」まで考慮した行動遷移モデリング手法 「semantics-aware mobility model」を提案している．[3] は，Word2Vec の Skip-gram モデルのアルゴリズム [4] を応用し，滞在目的や移動手段，天気など，あらゆる要素を同じベクトル空間上で扱うための手法 Traj2Vec を提案している．それを用いて，ユーザの分散表現を作成し，ユーザ間の類似度の算出を行っている．[1] は，従来までの移動遷移の類似度推定手法における，様々な要素（移動手段，滞在場所など）において，ユーザ間の類似度を算出する際のそれらの一致度の極端さを指摘し，その問題を改善し柔軟に対応できる手法 MUITAS を提案している．これらの遷移モデルは，ユーザ毎に全く異なる値（緯度，経度）となってしまう座標遷移モデルとは異なり，ユーザ間で共通するラベル（自宅，車，晴など）で表されるので，ユーザ同士の移動遷移の比較といった処理が可能である．これらの研究で扱っているデータは，座標遷移モデルのものとは情報量が異なり，より正確にユーザなモデリングができるという性質を持っている．しかし，これらの手法は，地理座標（緯度と経度）とタイムスタンプのみからなるデータを用いた場合，適用できないという問題が存在する．

そこで，座標遷移データから新たに情報を抽出するための研究も存在する．[5] は，滞在目的の推定を行っている．生活の中で比重が大きい自宅と職場の推定とその他（買い物や観光，食事など）を分離して，推定精度の向上を試みている．分類先としては，自宅や通勤先・通学先，買い物，食事・社交・娯楽，観光・行楽・レジャー，その他私用，その他業務の 7 つである．[6] は，滞在場所の推定手法と事前に収集した POI 情報を基にした滞在場所の属性（7 種類：自宅，職場，ショッピングエリア，交通，娯楽，学習，飲食）を推定するための確率的モデルの作成手法を提案している．そして，それを基にラベル遷移を作成し，移動パターンのマイニングを行い，ユーザの属性を推定しようとしている．[7] は，滞在目的の推定を行っている．滞在に関する時間的情報（滞在時間，頻度など）から作成したベイジアンネットワークと事前に収集した空間的情報（POI 情報）を組み合わせるための手法を提案している．これら上述した手法には，2 つの課題がある．1 つ目は，事前に地域の POI 情報を収集しなければいけないことである．これにより，別の地域で実験を試みたいと思った場合，POI 情報の収集から始める必要があり手間となってしまう．2 つ目は，滞在に関して，その目的を推定する際，分類先を固定する点である．この方法では，固定した分類先に当てはまらない特徴的なユーザを見つけることができない．

3. 提案手法

本論文では，ユーザの行動履歴の「滞在」に焦点を当て，各ユーザの滞在遷移をラベルを使用した遷移ではなく，滞在場所の特徴が表されている分散表現を使った遷移で表現することを目指す．そして，我々はこれを以下の考えに基づいて実現しようと検討している．それは，大量の位置情報履歴があれば，エリアに対して，そこへの滞在に関する時間情報のみから，エリアの特徴を示す分散表現が作成可能なのではないか，という考えである．滞在に関して時間情報のみを使用する理由としては，位置よりも時間の方が抽象度が高く，ユーザ間で見た時に意味がまとまりやすいためである．エリア毎の分散表現には，そのエリアに滞在する人が，どの時間帯にどの程度滞在するのか，といった情報が埋め込まれることが望ましい．これにより，クラスタリングを行うと，各クラスタには同じような使われ方をしているエリアがまとめられる．結果として，各クラスタから得られる情報の分析で，そのクラスタの特徴をある程度解釈可能になることが期待される．クラスタの特徴とは，クラスタ内のエリアが，どのような時間的使われ方をしているかを表したものである．これを分散表現としたものを LU（Location Usage）と名付ける．各クラスタ毎に LU が作成される．そして，ユーザの各滞在エリアがどのクラスタに属しているかを同定すれば，LU 遷移モデルを作成可能となる．この遷移モデルは，ラベル遷移モデルと異なり，エリアの特徴が反映されたものとなっている．

ここで，滞在遷移を表現する 3 つのモデルを整理する（図 1）．

- **座標遷移モデル**：滞在場所を地理座標（緯度，経度）を使用して表現したもの．
- **ラベル遷移モデル**：滞在場所をラベルを使用して表現したもの．
- **LU 遷移モデル**：滞在場所をその場所の時間的使われ方の特徴が考慮された分散表現で表現したもの．

ここからは，以下の 2 つのパートに分けて説明する．

- エリアの分散表現の作成
- 分散表現のクラスタリングと LU の算出

3.1 エリアの分散表現の作成

エリアの分散表現の作成には，Word2Vec [4] のアルゴリズムを使用する．Word2vec とは，自然言語処理の分野で開発された技術であり，テキスト処理を行うためのニューラルネットワーク（NN）である．この手法は，類似した

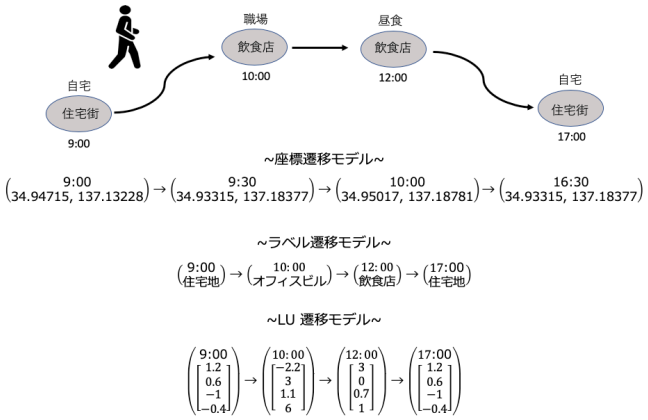

図 1　座標遷移モデルとラベル遷移モデルと LU 遷移モデルの違い

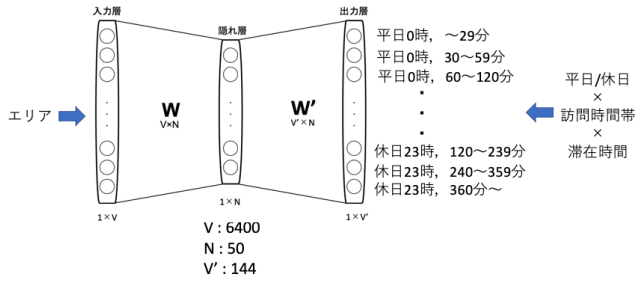

図 2　Skip-gram モデルの改良版

文脈で用いられる単語は類似した意味を持つという分布仮説に基づいたものである．Word2vec には学習手法として，Skip-gram と Continuous Bag of Words（CBoW）の二つのモデルがある．Skip-gram は，ある単語を入力にその周辺の単語を予測するタスクを NN で学習する．逆に CBoW は，周辺単語を入力にその中心の単語を予測するタスクを NN で学習する．こうして得られた中間層の重みが各単語の分散表現を表すようになる．この単語のベクトル空間において，ある単語の周辺によく表れる単語同士は近くに配置され，逆に文章中に同時に出現しない単語同士は遠くに配置されるようになる．本研究では，Skip-gram モデルを用いて，エリアへの滞在に関する時間情報のみを学習させて，エリアの分散表現を作成する．

Word2Vec の Skip-gram モデルは，本来であれば，ある文に注目し，その中のある単語を入力にその周辺の単語を予測するタスクを NN で学習させる．しかし，このままでは本研究が扱うデータを適用できないので，このモデルを改良する必要がある．具体的な改良点は，あるエリアを入力にそこへの滞在情報を予測する，というタスクに変更することである．図 2 が改良後のモデルである．

入力は，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 ベクトルとなる．出力は，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 ベクトルである．本研究での曜日，訪問時間帯，滞在時間の定義を表 1 にまとめた．

表 1　曜日，訪問時間帯，滞在時間の定義

曜日	平日，休日
訪問時間帯	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～

ここで出力側に用いるデータにおいて工夫した点を紹介する．それは，図 2 にある通り，曜日×訪問時間帯×滞在時間の全組み合わせを用意したことである．既存研究に Word2Vec の Skip-gram ベースのアルゴリズムを用いて，曜日や滞在時間，滞在目的といった様々な側面を考慮してユーザの分散表現を作成し，ユーザ間の類似度を算出するというものがある [3]．入力は，N 次元ベクトル（N：ユーザ数）であり，あるユーザに対して対応する箇所にフラグが立った one-hot ベクトルである．出力は，用いる要素毎のベクトルを足し合わせた 88 次元ベクトルとなっている（例えば，曜日は 1～2 次元目に対応し，滞在時間は 3～6 次元目，滞在目的は 7～20 次元目）．しかし，この方法では Word2vec の性質上，共起関係が考慮されないという問題がある．つまり，「土曜日にレストランで 1 時間食事」と「月曜日に 8 時間の仕事」という情報が，単語毎にばらばらに扱われてしまい，「土曜日」，「月曜日」，「1 時間の滞在」，「8 時間の滞在」，「レストラン」，「職場」という別々の情報になってしまう．この問題を解決し「何曜日の何時に何時間滞在」という共起を考慮可能にするために，曜日 × 訪問時間帯 × 滞在時間 の全組み合わせを用意して，144 次元の one-hot ベクトルとした．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するエリア同士は遠くに，ベクトル空間上で配置されることが期待できる．

3.2　分散表現のクラスタリングと LU の算出

次は，得られた分散表現をクラスタリングし，エリアの特徴が時間情報からまとめられているかを確認する．クラスタリングでは，k-means++ を使用した．このクラスタリング手法は，クラスタ数がハイパーパラメータとなり，値を自由に変更できる．そのため，滞在目的を推定するものとは違い，思いもしなかったことを発見できる可能性を秘めている．

そして次は，このクラスタリング結果から得られた各クラスタの時間的使われ方の特徴が反映された分散表現 LU を求める．LU の算出方法は，各クラスタに属しているエリアの分散表現を全て足し合わせ，エリア数で割ったものとする．つまり，各クラスタに属しているエリアの分散表現の平均である．

以上より，LU 遷移モデルを作成するための準備は完了した．ユーザの滞在エリアがどのクラスタに属しているか判定できれば，そのクラスタの LU を割り当てることで，座標遷移モデルを LU 遷移モデルに変換可能になる．

4. エリア毎の分散表現の計算実験

以下では，ブ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．実際に WordVvec の Skip-gram モデルをベースに改良を行ったモデルで，この 約 1800 個のメッシュの分散表現を作成し，k-means++ を使ってクラスタリングを行った．図 3 は，各メッシュを所属しているクラスタに応じ，マップ上で色分けしたものである（クラスタ数は 10）．色がないメッシュは，滞在が 10 回未満であるために分散表現が作成されなかったメッシュである．図 4 には，クラスタ数を 3 に設定した時の各クラスタの時間的特徴をグラフ化したものを示す．この図は，各クラスタの 1 メッシュあたり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．縦軸は人数を表している．横軸は時間を表している．ただし，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 の滞在を行った人がいた場合，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．そして，クラスタ番号の横には，そのクラスタに属しているメッシュ数を記している．クラスタ 0 を見てみると，逆に長時間の滞在を夜から開始する人が多いのが分かる．これのクラスタは，「住宅街」のような特徴を持つ分散表現であると解釈できる．また，クラスタ 1 を見てみると，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」のような特徴を持っている分散表現であると解釈できる．さらに，全エリアの滞在時間分布（クラスタ数 1）（図 5）とクラスタ数 10（図 6）の場合のクラスタリング結果も出した．これらの図から，クラスタ数 1 からクラスタ数 3 に分割する際，またはクラスタ数 3 からクラスタ数 10 に分割した際について，どのような時間的要因が分割要因なのかが判明するかもしれない．また，クラスタ数 10 のクラスタリング結果についても分析を行う．クラスタ 0 は，夜中から早朝にかけて（0〜6 時）には，滞在はほとんど行われていない．そして，昼間から夕方にかけて（11〜19 時）には，短時間の滞在を行う人の割合が多く，休日も人数を

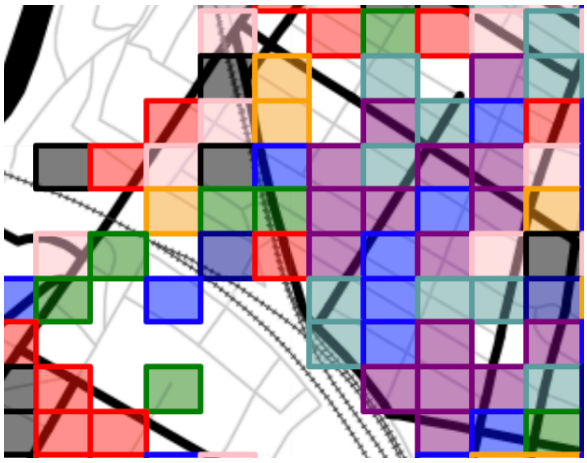

図 3　クラスタリング結果のマップへの描写（一部）

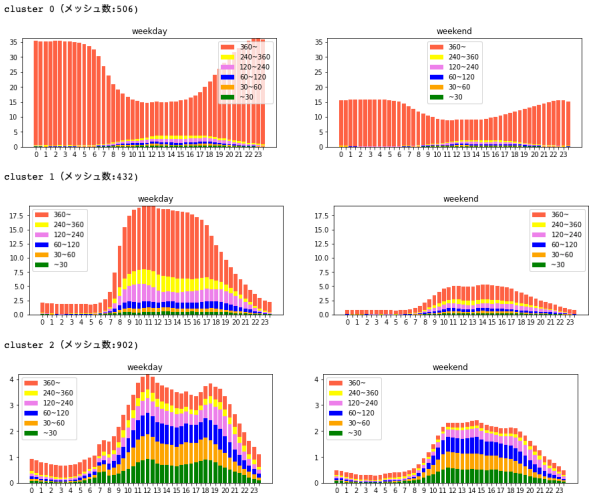

図 4　クラスタリング結果：クラスタ数 3

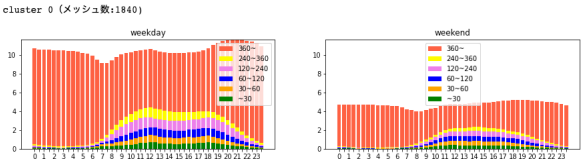

図 5　クラスタリング結果：クラスタ数 1

維持したまま同じような傾向が見える．したがって，このクラスタは，「店舗」（買い物）を表していると解釈できる．クラスタ 6 は，昼頃（11〜13 時）と夕方（18〜21 時）に，滞在が集中していることから，「飲食店」を表していると解釈できる．クラスタ 7 は，朝（7〜8 時）と夕方（18〜19 時）に短時間の滞在の人が目立つことから，通勤・通学に使用される駅やバス停を表していると解釈できる．このように，各クラスタには，そこに属しているメッシュへの滞在に関する情報が含まれている．これを分析することで，そのクラスタをある程度解釈することが可能であることが判明した．このことから，各メッシュの分散表現は，期待通り，滞在に関する時間的特徴からその近さが決定していることが分かった．

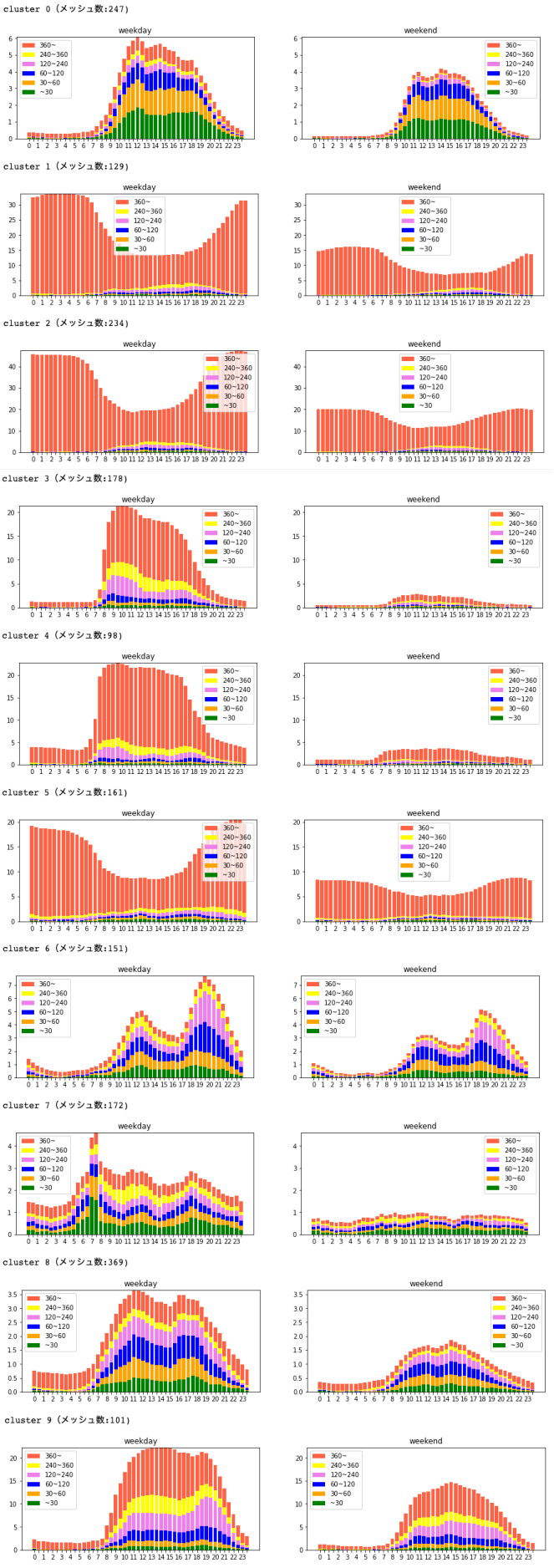

図 6　クラスタリング結果：クラスタ数 10

5.　今後の展望

実際にクラスタ数 10 における，ユーザの LU 遷移モデルを作成した（図 7）．LU 遷移モデルは，ラベル遷移モデルと比べると，滞在場所の特徴がより現れた情報を持っている．そのため今後は，ユーザ同士の移動パターンの類似度推定やユーザの属性推定などを可能にする LU 遷移モデルを用いた手法を検討する．また，様々な地域で LU を作成し，地域性による違いが存在するかも調べる価値があるはずである．

さらに別のアプローチとして，あるメッシュにおけるユーザ個人に依存した意味（PLS：Personal Location Semantics）[8] を反映した滞在遷移モデルの作成があげられる．LU には，そのエリアに訪れたユーザ全員の情報が含まれている．それに対し，PLS とは，ユーザ毎に滞在場所の意味は異なる，という考えが反映されたものである．例えば，レストランについて考えてみる．あるユーザにとっては「食事」をする場所でも，別のあるユーザにとっては「職場」かもしれない．このような情報を考慮した滞在場所の遷移モデルが作成できれば，一般的なエリアの使い方が反映された LU 遷移モデルと比較することで，そのユーザの特徴をより正確に推定可能になると考えられる．

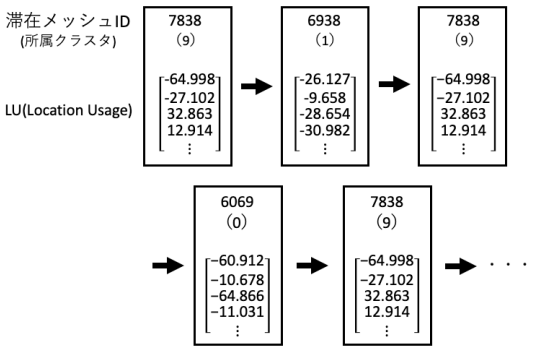

図 7　あるユーザの LU 遷移（一部）

6.　謝辞

本研究は，JST CREST JPMJCR1882，NICT 委託研究，総務省 SCOPE，JST OPERA(JPMJOP1612) の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ブログウォッチャー社に感謝します．

7.　参考文献

参考文献

[1] Lucas May Petry, Carlos Andres Ferrero, Luis Otavio Alvares, Chiara Renso, and Vania Bogorny. Towards semantic-aware multiple-aspect trajectory similarity measuring. *Transactions in GIS*, Vol. 23, No. 5, pp. 960–975, 2019.

[2] H. Shi, Y. Li, H. Cao, X. Zhou, C. Zhang, and V. Kostakos. Semantics-aware hidden markov model for human mobility. *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–1, 2019.

[3] Andrea Esuli, Lucas May Petry, Chiara Renso, and Vania Bogorny. Traj2user: exploiting embeddings for computing similarity of users mobile behavior, 2018.

[4] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pp. 3111–3119. Curran Associates, Inc., 2013.

[5] Mori KUROKAWA, Hiroki ISHIZUKA, Takafumi WATANABE, Shigeki MURAMATSU, Chihiro Ono, Hiroshi KANASUGI, Yoshihide SEKIMOTO, and Ryosuke SHIBASAKI. Estimating semantics of significant places using location information associated with telecommunication histories of mobile phones. *IEICE technical report. MoNA, Mobile network and applications*, Vol. 113, No. 398, pp. 79–84, jan 2014.

[6] 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.

[7] Mingqi Lv, Ling Chen, and Gencai Chen. Discovering personally semantic places from gps trajectories. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*, CIKM ’ 12, p. 1552–1556, New York, NY, USA, 2012. Association for Computing Machinery.

[8] 庄子和之, 米澤拓郎, 廣井慧, 河口信夫. 個人に依存した時空間セマンティクスの分散表現の検討. 情報処理学会研究報告ユビキタスコンピューティングシステム（UBI）2020-UBI-65(30)122020-03-02.