移動履歴データから順伝播型ニューラルネットワークで抽出した地域のベクトル表現の意味解釈
―ロジットモデルとの対応関係を通した検討―

白勢拓*・井上亮**

Interpretation of regional embeddings extracted from travel history data using a feedforward neural network
- Examination Through Correspondence with Logit Models -

Taku Shirose* and Ryo Inoue**

Abstract: Previous studies have explored applying Word2Vec, a natural language processing technique, to people flow data to extract embeddings representing regional characteristics. One study has demonstrated that these embeddings reflect inter-regional accessibility. However, analyses considering features of the input data, such as individual attributes and sequence relationships, have not been performed. This study aims to interpret the characteristics of embeddings extracted from stay location sequence data using a feed-forward neural network model that learns embeddings considering individual attributes and sequence relationships. Experimental results suggest that the embeddings extracted by the examined model represent accessibility between regions.

Keywords: 近接性（Accessibility），人流データ（people flow data），ベクトル表現（embeddings）

１．はじめに

近年，情報通信技術や測位技術の発展により，人々の移動の実態を詳細に記録した，人流データの蓄積が進んでいるが，その活用は，地域の滞在人口や地域間の移動量などの把握にとどまっている．人流データを人の移動履歴データとして扱う分析には，移動先の予測を目指したものは数多くみられるが，往復や周遊などの，地域間の人の移動パターンに着目して，地域間の関係性を捉えようとする分析は少ない．

人の移動履歴データの分析アプローチの一つは，各個人の滞在地の順序を表す滞在地系列に対して，深層学習に基づく自然言語処理手法を適用した地域の特徴を表す「ベクトル表現」の抽出である．Crivellari and Beinat (2019) は，自然言語の文章から単語のベクトル表現を抽出する Word2Vec (Mikolov et al., 2013) を滞在地系列に適用し，地域の特徴を表すベクトル表現を抽出できることを示した．さらに，Murray et al. (2023) は，Word2Vec による滞在地系列の分析は，特定条件下にて重力モデルと数学的に等価で，学習されたベクトル表現の内積が地域間の近接性を表すことを理論的に示し，実証的にも検証した．

しかし，Word2Vec が集計モデルである重力モデルと等価であるとの関係は，Word2Vec の入力データが個人単位の滞在地系列であるのに，分析にその「個人単位」「系列」という情報を反映できていないことを意味する．

人流データが有する情報を活かして地域間の近接関係を抽出するには，個人属性を考慮すること，および，滞在地が系列で記録されていることを考慮した分析が必要である．本稿はこの検討の第一段階として，特定の属性を有する個人が，ある滞在地から次の滞在地への１回の移動を１レコードとした移動データから，深層学習モデルによる分析を通して地域間の関係を抽出する分析を検討する．まず，得られるベクトル表現の解釈可能性を担保するため，Word2Vec と同様に，単純な Feed-Forward 型ニューラルネットワーク (FNN) のモデルを前提とする．個人属性と現在の滞在地の入力に対して，次の滞在地を予測するタスクの学習で得られる地域のベクトル表現が獲得する性質を，目的地選択行動に従来から用いられるロジットモデルとの比較を通して解釈し，人流データの深層学習モデルによる地域間関係分析への応用可能性を検討することを目的とする．

２．既往研究

Word2Vec のモデルのうち，Skip-gram モデルは，文章中の一単語を入力し，その前後に表れる単語を予測するタスクを通じて，モデルを学習する．全文章データ中の全単語の数をm，埋め込み次元をnとする．ある単語が入力されると，その単語に対応する行にのみ 1 を与え，それ以外に 0 を与えた $1\times m$の One-Hot ベクトルを入力層

* 学生会員 東北大学 大学院情報科学研究科（Tohoku University）
〒980-8579 宮城県仙台市青葉区荒巻字青葉 6-6-06 E-mail: shirose.taku.r6@dc.tohoku.ac.jp
** 正会員 東北大学 大学院情報科学研究科（Tohoku University）

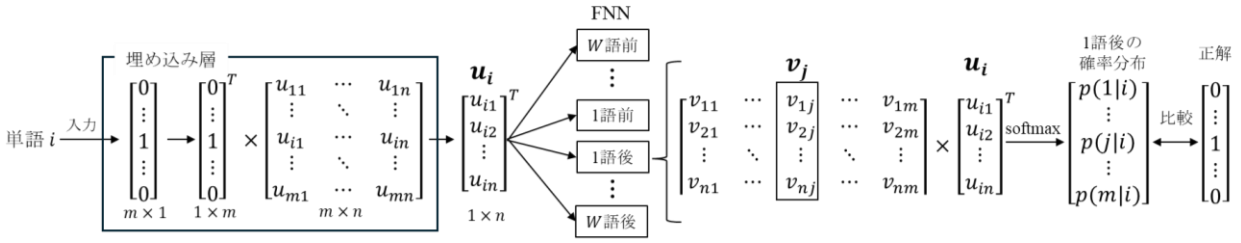


図 1　Skip-gram の概要

に入力する．対応する 1 行にのみ 1 が入ったベクトルのため，$m\times n$の埋め込み層のうち入力語に対応する行のみが演算され，中間層には入力単語に固有の $1\times n$の特徴量ベクトルが伝えられる．

Skip-gram の全結合層は，一般的な FNN である．前後単語の範囲であるウィンドウサイズWについて，$n\times m$の行列が前後それぞれ 2 つずつ，計 $2W$個存在する．出力層では，文章中で入力語の前後W語を予測する確率分布を softmax 関数でモデル化する．

$$p(j|i)=\frac{\exp(\boldsymbol{v_j}\cdot\boldsymbol{u_i})}{\sum_{j'=1}^{m}\exp(\boldsymbol{v_{j'}}\cdot\boldsymbol{u_i})} \tag{1}$$

ここで$p(j|i)$は，単語iが与えられたときに，任意の前後位置において単語jが現れる確率である．また，$\boldsymbol{v_j}$は，全結合層の中で単語jに対応する列ベクトルであり，$\boldsymbol{u_i}$は単語iの特徴量ベクトルである．

Murray et al., (2023) は，滞在系列データに Skip-gram を適用する手法が，モデルの学習時間を短縮する手法である Negative Sampling (NS) (Mikolov et al., 2013) を使用し，ウィンドウサイズ$W=1$の場合にのみ，重力モデルと等価の式で地域間流動を予測すると明らかにした．モデルの学習の際，通常の Skip-gram は学習対象地域全てについてベクトルの積を計算し，softmax 関数でモデル化するが，NS は正例(正解)に対して複数の負例(間違い)をランダムにサンプリングし，その負例を正解もしくは間違いに分類する 2 値分類問題に置き換えることで，モデルの学習時間を短縮する手法である．負例のサンプリングの際には，学習データ中で登場頻度の高い地域が高確率で抽出される．NS の目的関数は，正例のペアの予測確率を 1 に近づけ，且つ負例のペアの予測確率を 0 に近づけることであるため，以下の損失関数の最小化である．

$$\begin{aligned}L^{NS}&=\sum_i\left[\ln\frac{1}{1+\exp\{-\ln f(\boldsymbol{v_j}\cdot\boldsymbol{u_i})+\ln p_0(j)+c\}}\right.\\&\left.+\sum_{k\in V}\ln\frac{1}{1+\exp[-\{-\ln f(\boldsymbol{v_k}\cdot\boldsymbol{u_i})+\ln p_0(k)+c\}]}\right]\\f(\boldsymbol{v_j}\cdot\boldsymbol{u_i})&=(\boldsymbol{v_j}\cdot\boldsymbol{u_i}+\ln p_0(j)+c)\end{aligned} \tag{2}$$

ここで，jは正例，Vは負例の集合，$p_o(j)$は学習データ中でのjの登場割合，cは定数項である．$W=1$で，Skip-gram に NS を使用した場合，学習された埋め込み層の重みと全結合層の重みが一致する (Levy and Goldberg. 2014)．そのため，Skip-gram Negative Sampling (SGNS) を用いて予測する確率分布は，以下の式で計算される．

$$\begin{aligned}P^{NS}(j|i)&=\frac{f(\boldsymbol{u_j}\cdot\boldsymbol{u_i})}{\sum_{j'\in V}f(\boldsymbol{u_{j'}}\cdot\boldsymbol{u_i})}\\&=\frac{\exp(\boldsymbol{u_j}\cdot\boldsymbol{u_i}+\ln p_0(j)+c)}{\sum_{j'\in V}\exp(\boldsymbol{u_{j'}}\cdot\boldsymbol{u_i}+\ln p_0(j')+c)}\\&=\frac{p(j)\exp(\boldsymbol{u_j}\cdot\boldsymbol{u_i})}{\sum_{j'\in V}p\,(j')\exp(\boldsymbol{u_{j'}}\cdot\boldsymbol{u_i})}\end{aligned} \tag{3}$$

したがって，SGNS によって予測される，地域iと地域j間の流動量$\widehat{T_{ij}}$は以下の通りである．

$$\widehat{T_{ij}}\propto p(i)P^{NS}(j|i)\propto\frac{p(i)\,p(j)\exp(\boldsymbol{u_j}\cdot\boldsymbol{u_i})}{\sum_{j'}p(j')\exp(\boldsymbol{u_{j'}}\cdot\boldsymbol{u_i})} \tag{4}$$

また，Skip-gram は，滞在地系列の中での滞在地iとjの隣接関係は考慮するが，その前後関係は考慮できないため，地域i, jを発着する相互の流動量が等しいと仮定してモデルを学習することになる．

$$\begin{aligned}&\widehat{T_{ij}}=\widehat{T_{ji}}\Leftrightarrow\frac{p(i)\,p(j)\exp(\boldsymbol{u_j}\cdot\boldsymbol{u_i})}{\sum_{j'\in V}p(j')\exp(\boldsymbol{u_{j'}}\cdot\boldsymbol{u_i})}=\frac{p(j)\,p(i)\exp(\boldsymbol{u_i}\cdot\boldsymbol{u_j})}{\sum_{i'\in V}p(i')\exp(\boldsymbol{u_{i'}}\cdot\boldsymbol{u_j})}\Leftrightarrow\\&\sum_{j'\in V}p(j')\exp(\boldsymbol{u_{j'}}\cdot\boldsymbol{u_i})=\sum_{i'\in V}p(i')\exp(\boldsymbol{u_{i'}}\cdot\boldsymbol{u_j})\\&\Rightarrow\widehat{T_{ij}}=Cp(i)p(j)\exp(\boldsymbol{u_j}\cdot\boldsymbol{u_i})\end{aligned} \tag{5}$$

(5)式より，$W=1$の設定の SGNS は，重力モデル$\widehat{T_{ij}}=Cm_im_jf(r_{ij})$ と等価な方程式で移動を予測すると解釈できる．つまり，学習されたベクトル表現 $\widehat{\boldsymbol{u_i}},\widehat{\boldsymbol{u_j}}$のベクトル空間上での距離が，地域$i,j$間の近接性に対応する．

３．FNN とロジットモデル

ある時点の滞在地を入力し，次の移動後の滞在地を予測して，ベクトル表現を学習する FNN と，ロジットモデルそれぞれが予測する確率分布は，いずれも softmax 関数の形をしており，FNN の全結合層の出力と，ロジットモデルの確定効用は対応関係にある．いずれのモデルも，個人単位の移動行動を扱うため，ある時点の滞在地の入力から次の滞在地を予測する FNN の学習により得られるベクトル表現について，その意味をロジットモデルとの比較を通して明らかにする．

３．１ 検討する FNN モデル

本研究では，以下の式で次の滞在地を予測して，ベクトル表現を学習するモデルを検討する．

$$p^k(j|i) = \frac{\exp\left(\boldsymbol{v_j}\cdot\boldsymbol{u_i} + \boldsymbol{w^a}\cdot\boldsymbol{a_j} + \boldsymbol{w_j^h}\cdot\boldsymbol{h_k}\right)}{\sum_{j'=1}^{m}\exp\left(\boldsymbol{v_{j'}}\cdot\boldsymbol{u_i} + \boldsymbol{w^a}\cdot\boldsymbol{a_{j'}} + \boldsymbol{w_{j'}^h}\cdot\boldsymbol{h_k}\right)} \quad (6)$$

ここで，$\boldsymbol{u_i}$はある時点の滞在地iの特徴量ベクトル，$\boldsymbol{v_j}$は次の移動後の滞在地jに対応する全結合層の列ベクトル，$\boldsymbol{a_j}$は着地の魅力度，$\boldsymbol{h_k}$は移動者kの個人属性，wは重みパラメータである．このモデルは， Word2Vec と同様の，1 層の Feed-Forward 型の全結合層での出力に，着地の魅力度と個人属性の線形結合を加算して，softmax 関数で確率化する．全結合層の数は，ベクトル表現の解釈可能性の担保のために，本研究では 1 層のモデルを考える．

３．２ ロジットモデル

ロジットモデルは，ランダム効用理論に基づく離散選択モデルである．個人は選択肢の望ましさを意味する効用が最も大きい選択肢を選ぶ．ランダム効用理論において，選択肢の効用は確率的に表され，選択肢jの効用U_jは以下の式により計算される．

$$U_j = V_j + \varepsilon_j \quad (7)$$

ここで，V_jは観測可能な説明変数によって表される確定効用，ε_jは観測不能な要因を説明する確率的な誤差項である．ある個人が選択肢jを選択する確率p_jは，

$$p_j = Pr\left[U_j > \max_{j'(\neq j)}\left(U_{j'}\right)\right] \quad (8)$$

ある．ここで，$Pr[*]$は*の生起する確率である．誤差項が独立なガンベル分布に従う場合の，ランダム効用理論に基づく離散選択モデルがロジットモデルと呼ばれ，p_jは以下のように表される．

$$p_j = \int Pr\left[U_j = x\right]Pr[U_1 < x]\cdots Pr[U_m < x]dx$$

$$= \exp\left(V_j\right)\int_{-\infty}^{\infty}\exp(-x)\exp\left[-\left(\sum_{j'=1}^{m}\exp\left(V_{j'}\right)\right)\exp(-x)\right]dx$$

$$= \frac{\exp\left(V_j\right)}{\sum_{j'=1}^{m}\exp\left(V_{j'}\right)} \quad (9)$$

目的地選択問題において，確定効用の説明変数には発地i着地j間の距離d_{ij}，着地の魅力度$\boldsymbol{a_j}$，移動者kの個人属性$\boldsymbol{h_k}$などが一般に用いられ，確定効用項はそれらの線形結合によって表される．地域iにいる個人kにとって，次の滞在地の地域jの確定効用は以下の通りである．

$$V_j^k = w^d\cdot d_{ij} + \boldsymbol{w^a}\cdot\boldsymbol{a_j} + \boldsymbol{w_j^h}\cdot\boldsymbol{h_k} \quad (10)$$

ここでwは重みパラメータである．2 地域間距離，着地の魅力度などの選択肢に固有の変数には，全地域に共通の重みパラメータを用いる．一方，個人属性などの選択肢に固有でない変数は，その線形結合を選択肢固有にするために，選択肢ごとに異なる重みパラメータ$\boldsymbol{w_j^h}$を設定することが一般的である．なお，全ての選択肢に固有の重みパラメータを設定すると多重共線性が生じるために，一つの基準選択肢はその重みパラメータを 0 に固定する．

３．３ 検討モデルとロジットモデルの対応関係

検討モデルは，全結合層での出力に，(10)式のうち，距離以外の説明変数にあたる項を加算して，softmax 関数で確率化する．つまり，ロジットモデルで地域間距離にあたる項を$\boldsymbol{v_j}\cdot\boldsymbol{u_i}$で置き換えた式で次の滞在地を予測するため，$\boldsymbol{v_j}\cdot\boldsymbol{u_i}$は$w^d\cdot d_{ij}$と対応する．したがって，検討モデルにより学習されたベクトル表現と全結合層の重みの内積 $\widehat{\boldsymbol{v_j}}\cdot\widehat{\boldsymbol{u_i}}$は，移動データに潜む地域間の近接性を表す．

４．東京都を対象とした分析

本研究では，実データに検討モデルを適用し，得られた地域のベクトル表現が獲得した性質を考察する．

４．１ 使用データとモデルの設定

東京大学空間情報科学研究センター「人の流れプロジェクト」提供の「擬似人流データ」から，東京都内で移動が完結した約 870 万人の移動を対象とする．東京都市圏パーソントリップ調査の計画基本ゾーン 164 地域で個人の滞在地を表した系列を作成した．移動者の個人属性は年齢，性別，職業の 3 種類を使用し，年齢は 0~19 歳，20~39 歳，40~59 歳，60~89 歳の 4 区分，性別は男性，女性の 2 区分，職業は就業者，無職，学生の 3 区分に分類した．「0~19 歳」「女性」「学生」を基準のカテゴリーとし，ダミー変数を設定した．着地の魅力度を表す指標として，平成 28 年経済センサス-活動調査 従業者数，平成 27 年度国勢調査 人口総数を用いる．従業者数は繁華街やオフィス街などの性質による魅力度を説明し，人口総数は住宅街の性質による魅力度を説明する変数として用いる．

検討モデルは埋め込み次元を 25 次元とし，滞在地系列データを学習用と検証用に 8:2 の割合で分割し，Early stopping を用いて，validation loss の改善が見られなくなってから 2 回まで学習を行い，最良のモデルを選択した．

検討モデルの比較対象として，SGNS の学習とロジットモデルの推定を行う．SGNS は，Murray et al. (2023)と同じ設定のモデルを用いて，埋め込み次元を検討モデルと同じ 25 次元と設定する．ロジットモデルの説明変数は，着地の魅力度と個人属性は検討モデルと同じ変数，地域間距離は計画基本ゾーンの重心間距離(km)を用い，(10)式で確定効用を計算する．なお，この推定では，全データを用いた推定は計算量の観点から困難なため，対象データから 3%をサンプリングして使用した．

４．２ ベクトル表現の特徴

検討モデルにより学習されたベクトル表現と全結合層の重みの内積 $\widehat{\boldsymbol{v_j}}\cdot\widehat{\boldsymbol{u_i}}$を縦軸，推定されたロジットモデルでの距離の重みパラメータと距離の積 $\widehat{w^d}\cdot d_{ij}$を横軸とした散布図を**図 2** に示す．推定された距離の重みの値は，-0.368 であった．対応する両者の間には，比例に近い関係性が見てとれ，類似した性質を持つと考えられる．

図 2 の中で，都心 5 区(千代田，中央，港，新宿，渋谷)

を発地とする移動に注目したものを**図 3**(a)に，都心 5 区を着地とする移動に注目したものを**図 3**(b)に示す．**図 3**(a)は，**図 3**(b)と比較して，$\widehat{w^a} \cdot d_{ij}$の値が小さい範囲に，$\hat{\boldsymbol{v}}_j \cdot \hat{\boldsymbol{u}}_i$の値が特に小さい点が多く見られる．これは，距離が離れた地域間の移動において，都心を発地とする場合と，都心を着地とする場合では，ベクトル表現が表す近接性に大きな違いがあることを示している．オフィス街や繁華街が集中する都心の人を集める範囲の広さに起因して，都心を着地とする方向の近接性が高くなったと考えられ，ベクトル表現は移動の向きを区別した近接性を表すと解釈できる．

次に，$\hat{\boldsymbol{v}}_j \cdot \hat{\boldsymbol{u}}_i$について，新宿駅西口地域，八王子駅地域を発地とする移動に注目した分布を**図 4** に，新宿駅西口地域，八王子駅地域を着地とする移動に注目した分布を**図 5** に示す．八王子駅地域は両図の間に違いがみられない一方，新宿駅西口地域は，**図 4** と比較して，**図 5** では地理的距離が近い地域だけでなく，距離の離れた地域にも，その値の大きい地域が確かに多く分布している．比較対象の SGNS によるベクトル表現の内積について，新宿駅西口地域，八王子駅地域に注目した分布を**図 6** に示す．移動の向きを区別できない SGNS の分布は，**図 4**, **図 5** と異なる傾向の分布をしている．

５．おわりに

本研究は，個人属性と現在の滞在地の入力に対して，次の滞在地を予測する FNN で得られる地域のベクトル表現の性質を，ロジットモデルとの比較を通して解釈を行った．実データを用いた分析により，そのベクトル表現は，移動の向きを区別した地域間の近接関係を表すと評価でき，人流データの深層学習モデルによる地域間関係分析への応用可能性が示唆された．今後，過去の滞在地と現在の滞在地の関連性を学習し次の滞在地を予測する，LSTM や Transformer などの深層学習モデルによるベクトル表現について，同様の議論へと発展させたい．

謝辞

本研究は，東京大学 CSIS 共同研究（No. 1184）による成果である（利用データ：擬似人流・トリップデータ 東京都データセット CSIS 人の流れプロジェクト事務局提供）．

参考文献

Crivellari, A. and Beinat, E. (2019) From motion activity to geo-embeddings: Generating and exploring vector representations of locations, traces and visitors through large-scale mobility data. *ISPRS International Journal of Geo-Information*, **8**(3), 134.

Levy, O., & Goldberg, Y. (2014). Neural word embedding as implicit matrix factorization. *Advances in neural information processing systems*, 27.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013) Efficient estimation of word representations in vector space. arXiv:1301.3781.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.

Murray, D., Yoon, J., Kojaku, S., Costas, R., Jung, W. S., Milojević, S., & Ahn, Y. Y. (2023). Unsupervised embedding of trajectories captures the latent structure of scientific migration. *Proceedings of the National Academy of Sciences*, 120(52), e2305414120.

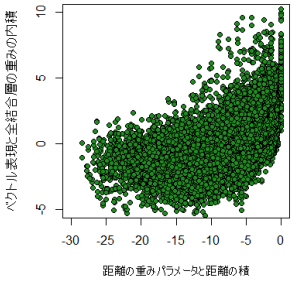
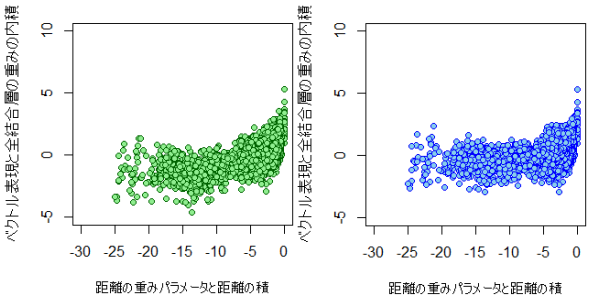


図 2　$\hat{\boldsymbol{v}}_j \cdot \hat{\boldsymbol{u}}_i$と$\widehat{w^a} \cdot d_{ij}$の散布図



(a) 発地　　(b) 着地

図 3　都心 5 区に着目した$\hat{\boldsymbol{v}}_j \cdot \hat{\boldsymbol{u}}_i$と$\widehat{w^a} \cdot d_{ij}$の散布図

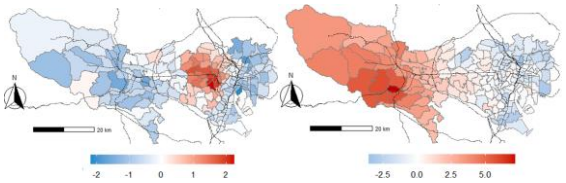


図 4　注目地域を発地とする移動の$\hat{\boldsymbol{v}}_j \cdot \hat{\boldsymbol{u}}_i$の分布

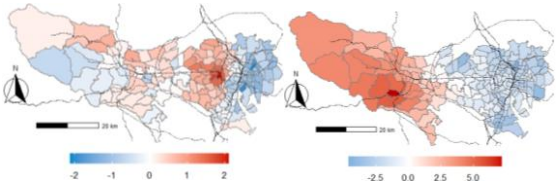


図 5　注目地域を着地とする移動の$\hat{\boldsymbol{v}}_j \cdot \hat{\boldsymbol{u}}_i$の分布

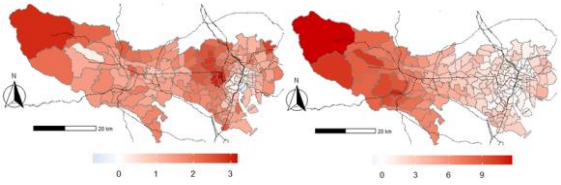


図 6　SGNS の$\hat{\boldsymbol{u}}_j \cdot \hat{\boldsymbol{u}}_i$の分布