

Twitter 位置情報を用いた地域間移動傾向の要因分析とそのモデル化

Modeling and Factor Analysis of Human Mobility Tendency between Areas Using Spatio-temporal Data of Twitter

前田 高志ニコラス *¹

Takashi Nicholas MAEDA

鳥海 不二夫 *¹

Fujio TORIUMI

大橋 弘忠 *¹

Hirotsuda OHASHI

*¹ 東京大学大学院工学系研究科

Graduate School of Engineering, The University of Tokyo

It is necessary to analyze spatial interactions between cities for evaluating cities' sustainability. On the other hand, today we can more easily collect fine-grained spatio-temporal data generated through social media and mobile phones than before. In this research, we propose a method that enables us to see cities' attractiveness and how much distance between two cities affect the number of mobility. Using Twitter data with geotag, we tried to extract people's infrequent traveling from Twitter data, and we calculate cities' attractiveness and the influence of distance by multiple regression analysis.

1. はじめに

近年、地方経済の疲弊や高齢化、また大規模災害などにより、地方都市の存立条件について注目が集まっている。とりわけ、都市間のヒト・モノ・カネの流れがどのような空間的なネットワークを形成しているかを把握することが地方都市の分析にとって大きな重要性を持つ。他方、スマートフォンの普及によって人の位置情報について、空間的にも時間的にもきめ細かい精度の情報を得ることが可能になった。そのため、今後これらのデータをうまく利用することによって、都市間の移動傾向とその要因を深く分析することが期待される。

2都市間の移動件数は、目的地の魅力と2都市間の移動コスト(距離)、出発地の人口や経済状態などによって決まると考えられる。本研究は、位置情報付きのTwitter投稿記事を利用して、目的地が持つ魅力を推定し、2地点間の距離が移動件数に与える影響を明らかにする。

これまで、人文地理学の領域では、人や物や情報の空間的フローを説明する空間的相互作用モデルの構築がなされてきた[石川 88]。特に歴史が深いものはニュートンの万有引力を用いた重力モデルである。これは、2都市間の流量が両都市の規模の積に比例し、距離に反比例するとしたモデルである。

空間情報学の領域では、Phithakkitnuookoonaら[Phithakkitnuookoon 14]が、大規模な携帯電話の位置情報をもとにして、旅行者の行動を詳細に分析することに成功している。この研究では、旅行者の旅行頻度、移動距離、目的地、出発地、移動手段、現地での滞在時間の関係性を個人レベルで抽出し、その傾向の分析を行った。

若宮ら[若宮 13]は、位置情報付きTwitterデータとパーソントリップ調査のデータをもとに、群衆の移動傾向を分析している。この研究では、地域間の移動について、移動距離・移動時間・移動量の3つの値をもとに各地域間の直感的な近接性を多次元尺度構成法により示すことに成功している。

本研究では、日常の定常的な移動(通勤・通学・帰宅など)ではない、非定常的な移動(旅行・出張・帰省など)をTwitterから抽出し、2地域間の移動件数と距離のみから、到着地の魅

力を推定することを目的とする。さらに距離が移動件数に与える影響を算出する。その際、移動件数の算出に際し、隣接地域の魅力の競合を考慮する。

2. 非定常移動の抽出と出発地・到着地のクラスタリング

2.1 非定常移動の抽出

位置情報付きTwitterデータからいかに非定常な移動を抽出するかを述べる。非定常移動とは、旅行・出張・帰省など、頻度の少ない移動を指す。これに対して、定常的な移動とは、家や職場、学校のような地点の間の日常的な移動を指す。本研究では、位置情報付きTwitterデータから非定常移動を以下の手順で抽出する。

まず、各ユーザのTweet(Twitter投稿記事)について、平日朝・平日昼・平日夜に投稿されたものを別々に抽出する。それらのTweetについて、平日朝・平日昼・平日夜それぞれの時間帯ごとに、Tweetに付与された緯度経度の情報をもとに、空間的な中心点と標準偏差を求める。これらの3つの中心点から各々の標準偏差の2倍の距離を半径とした円を描く。このユーザのすべて時間帯のTweetから、この3つの円に含まれていないTweetを抽出し、その位置と平日昼の円の中心点と結び、平日昼の中心点を出発地、もう一方の点を目的地とする。これを非定常移動とする。

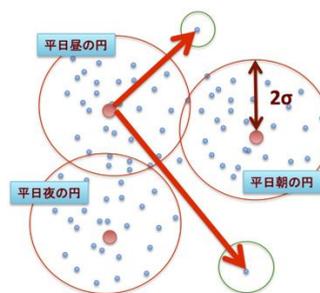


図 1: 非定常移動の抽出方法

連絡先: 東京大学大学院工学系研究科システム創成学専攻

〒113-8656 東京都文京区本郷 7-3-1 工学部 8 号館 526

TEL: 03-5841-6991

E-mail: maeda@crimson.q.t.u-tokyo.ac.jp

2.2 出発地・到着地のクラスタリング

前述の手順で数多くの出発地と到着地が得られるが、空間的に近いものをひとつにまとめたい。そこで Mean Shift Clustering[Fukunaga 75]を用いる。

Mean Shift Clustering とは以下のような、漸次的な手順によって近い点同士をひとつにまとめる手法である。各ステップごとに各点が次に移る先の点を、自身を含めた近傍半径 r_n の円内に含まれるすべての点の重心とする。すべての点について、近傍半径 r_n の円内の点がそれよりさらに小さい収束半径 r_c の円内に収まれば、そこでこの処理を終了する。同じ収束半径に収まった点同士を同じクラスタとしてまとめ、各クラスタの重心をそのクラスタの代表点とする。

ここで、出発地と目的地が同一のクラスタとなった移動は、移動データと見なさないことにする。

また、毎ステップですべての点同士の距離を求めると計算時間が膨大になるため、不要な計算を省く工夫をする。緯度・経度を等間隔で区切ったメッシュを張り、それらの各メッシュの縦・横の長さが r_n よりも長くなるようにする。このようにすれば、各点について、その点が含まれるメッシュと近傍の 8 メッシュ内の点以外は r_n 以上の距離にあるため、距離の計算をせずにすむ。これによって計算時間の短縮をはかる。

本研究では、 r_n を 25km とし、 r_c を 10km とした。

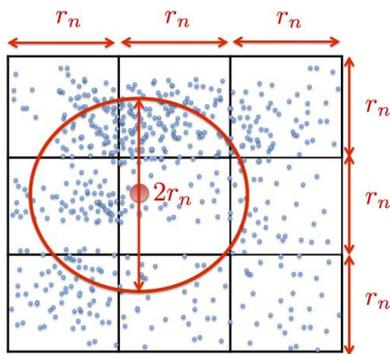


図 2: Mean Shift Clustering の効率化

3. モデル化と指標値の定義

3.1 概要

本研究では、2 地点間の移動件数は下記 4 つの値によって決まると考え、到着地の魅力を推定するために、これらの指標値が互い持つ関係性をモデル化する。

- 出発地の放出力
出発地の放出力が高ければ、その地点からの移動件数と移動距離が大きくなる。
- 到着地の魅力
到着地の魅力が高ければ、その地点への移動件数と移動距離が大きくなる。
- 移動コスト
2 地点間の移動コストが高ければ、その区間の移動件数が減る。本研究では、移動距離をコストの指標値として用いる。

- 到着地の競合

出発地の周りに多くの魅力ある到着地があれば、それぞれの到着地への移動件数が分散する。

3.2 目的地選択のモデル

出発地点 s に存在する人間が、数ある目的地から目的地 d を選択する確率 $P(s \rightarrow d|s)$ を以下の式で表す。

$$P(s \rightarrow d|s) = \frac{A_d}{D_{sd}^\alpha} / E_s \quad (1)$$

- A_d : 目的地 d の持つ絶対的な魅力 (未知変数)
- D_{sd} : 出発地 s と到着地 d の間の距離 (既知変数)
- α : 距離が目的地の魅力に与える影響を決定する係数 (未知変数)
- E_s : 出発地 s の周囲の魅力の総和 (未知変数)

$$E_s = \sum_k \frac{A_k}{D_{sk}^\alpha} \text{ によって与える。}$$

出発地点 s に存在する人間が、数ある目的地から目的地 d を選択する確率 P は、実データによる観測値を用いると下記のように表すことができる。

$$P_{obs}(s \rightarrow d|s) = \frac{T_{s \rightarrow d}}{\sum_k T_{s \rightarrow k}} \quad (2)$$

- $T_{s \rightarrow d}$: データで得られた、出発地 s から目的地 d へ移動した移動件数 (既知変数)

4. 重回帰分析による指標値の算出

4.1 データセット

今回用いたデータセットは位置情報付きの Twitter 投稿記事データのうち、2013 年 7 月～2014 年 6 月に日本で投稿されたものである。本研究では、この期間を 2013 年 7 月～2013 年 9 月と 2013 年 10 月～2013 年 12 月と 2014 年 1 月～2014 年 3 月と 2014 年 4 月～2014 年 6 月の 4 期間に分けて同じ計算処理を行った。

4.2 ユーザーの選定方法

日常的に位置情報付きの投稿をしているユーザのデータでなければ、Tweet がなされた位置が日常的に行く場所か否かを判断することができない。そのため、各ユーザに対して、Tweet の頻度に基準を設けて、その基準を超えたユーザの Tweet のみを分析の対象とする必要がある。本研究では、各 4 期間のそれぞれ 3 ヶ月間の中で、平日朝に投稿された日が 15 日以上あり、かつ、平日昼に投稿された日が 20 日以上あり、かつ、平日夜に投稿された日が 15 日以上あるユーザのみを対象とした。ただし、ここで朝の時間帯は午前 4 時から午前 10 時まで、昼の時間帯は午前 10 時から午後 6 時まで、夜の時間帯は午後 6 時から午前 4 時までとした。

4.3 重回帰分析の方法

式 (1) と式 (2) を等号で表し、両辺の対数をとる.

$$\log\left(\frac{T_{s \rightarrow d}}{\sum_k T_{s \rightarrow k}}\right) = \log A_d + \alpha \cdot \log D_{sd} - \log E_s \quad (3)$$

ここで左辺および D_{sd} はデータから求まる. ただし, 緯度経度からの距離の計算では, GRS80 楕円体をもとにした計算によって求める.

Mean Shift Clustering をもとに出発地・到着地をクラスタリングしたのちに, 各 2 地点間の移動件数を算出し, 式 (3) に代入して, 重回帰分析により, 各目的地の魅力 (A_d), 距離が魅力に与える係数 (α), 各出発地の周囲の魅力 (E_s) を求める.

4.4 各指標値の算出結果

4.4.1 距離が目的地の魅力に与える影響を決定する係数

距離が目的地の魅力に与える影響を決定する係数 (α) は重回帰分析の結果, 各期間ごとに表 1 の通りとなった.

2013/7~ 2013/9	2013/10~ 2013/12	2014/1~ 2014/3	2014/4~ 2014/6
0.7316	0.8299	0.8432	0.9330

表 1: 距離が目的地の魅力に与える影響を決定する係数

4.4.2 到着地の絶対的な魅力

到着地の絶対的な魅力については, 図 3 にて, 2013/7~2013/9 の期間分を緑色の円の半径の長さで表している (なお, 図内の赤い矢印線は 2 地点間の移動を表し, その線の濃さは $P_{obs}(s \rightarrow d|s)$ に応じた濃さとなっている).

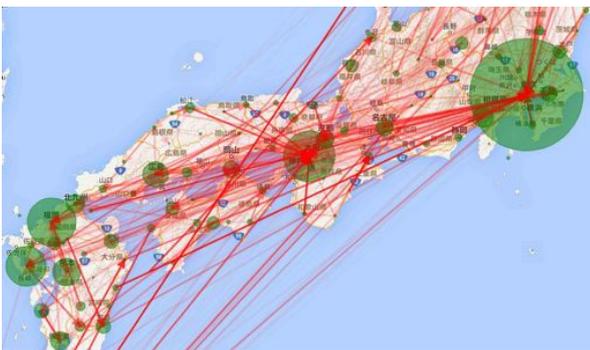


図 3: 到着地の絶対的な魅力 (2013/7~2013/9)

表 2 は, 4 つの期間についてそれぞれ, 到着地の絶対的な魅力を上位 10 地点を並べたものである.

図 4 は 2013/7~2013/9 の期間の到着地の絶対的な魅力について, 横軸にその順位, 縦軸にその魅力をとった両対数グラフである.

4.4.3 出発地の周囲の魅力

同様に出発地の周囲の魅力についても, 図 5・表 3・図 6 に表した.

	2013/7~ 2013/9	2013/10~ 2013/12	2014/1~ 2014/3	2014/4~ 2014/6
1	東京都 目黒区	沖縄県 与那国町	東京都 港区	東京都 渋谷区
2	大阪府 大阪市	東京都 港区	福岡県 福岡市	沖縄県 北谷町
3	福岡県 福岡市	沖縄県 沖縄市	大阪府 大阪市	大阪府 大阪市
4	沖縄県 浦添市	鹿児島県 与論町	沖縄県 宜野湾市	福岡県 福岡市
5	長崎県 諫早市	福岡県 珂川町	北海道 札幌市	愛知県 名古屋市
6	沖縄県 宮古島市	大阪府 大阪市	愛知県 名古屋市	沖縄県 竹富町
7	北海道 札幌市	沖縄県 宮古島市	沖縄県 石垣市	北海道 札幌市
8	沖縄県 与那国町	北海道 札幌市	鹿児島県 奄美市	沖縄県 宮古島市
9	熊本県 熊本市	京都府 京都市	鹿児島県 鹿児島市	京都府 久御山町
10	鹿児島県 屋久島町	長崎県 諫早市	沖縄県 宮古島市	長崎県 長崎市

表 2: 到着地の絶対的な魅力 (各期間について上位 10 地点)

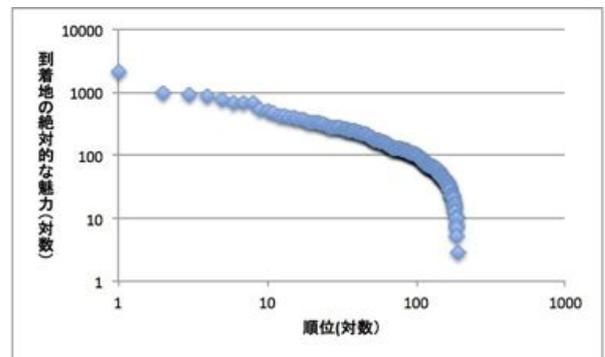


図 4: 到着地の絶対的な魅力 (2013/7~2013/9)



図 5: 出発地の周囲の魅力 (2013/7~2013/9)

	2013/7~ 2013/9	2013/10~ 2013/12	2014/1~ 2014/3	2014/4~ 2014/6
1	京都府 京都市	京都府 京都市	愛知県 名古屋市	京都府 久御山町
2	滋賀県 竜王町	大分県 大分市	埼玉県 深谷市	兵庫県 神戸市
3	東京都 目黒区	群馬県 玉村町	京都府 久御山町	大阪府 大阪市
4	静岡県 御殿場市	大阪府 大阪市	滋賀県 竜王町	愛知県 名古屋市
5	千葉県 四街道市	滋賀県 東近江市	東京都 港区	東京都 渋谷区
6	愛知県 名古屋市	愛知県 名古屋市	熊本県 熊本市	埼玉県 深谷市
7	大阪府 大阪市	大阪府 泉大津市	兵庫県 神戸市	滋賀県 近江八幡市
8	静岡県 浜松市	熊本県 熊本市	千葉県 習志野市	長崎県 長崎市
9	福岡県 北九州市	東京都 港区	福岡県 福岡市	宮城県 仙台市
10	兵庫県 姫路市	兵庫県 姫路市	福岡県 北九州市	千葉県 印西市

表 3: 出発地の周囲の魅力 (各期間について上位 10 地点)

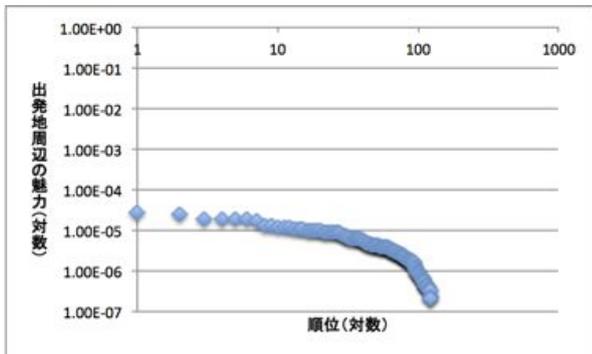


図 6: 出発地の周囲の魅力 (2013/7~2013/9)

4.5 考察

距離が目的地の魅力に与える影響を決定する係数は、表 1 で見るように、0.73~0.93 の間にある。これが指し示すのは、出発地から近距離にある到着地は距離が増えるに伴って大きくその魅力を減じるが、遠距離にある到着地同士を比べる場合、到着地の魅力への距離の影響は比較的緩やかになることを意味する。

到着地の絶対的な魅力については、図 4 から上位約 100 地点に限ればベキ分布に似た分布になっている。到着地の絶対的な魅力は、表 2 より、期間に応じて大きく変動することが分かる。

出発地の周囲の魅力については、表 3 より、京都・大阪・滋賀・兵庫のように、距離の近い範囲に中規模以上の都市がまとまって存在する地域は、どの期間も上位に存在しやすい傾向があることが分かる。

以上のことを総合すると、出発地から一定の近距離にあるか否かが到着地の魅力を大きく左右すること、到着地の絶対的な魅力は期間に応じて変動しやすいこと、近距離に中規模以上の都市が多く存在する場合はどの期間であっても周囲の魅力が大きく変動することはないことが確認できた。

5. おわりに

本研究により、複数の都市が互いに空間的に近接なネットワークを形成すると、各々の都市の周囲の環境が安定的に良好な状態になるという示唆が得られた。したがって、各都市のレジリエンス性を確保するためには、その周囲の都市との間に、互いに良いフィードバック与え合うようなネットワークを維持する必要がある。

今後は以下の 3 点を検討することで、地域間移動傾向をさらに理解するべく研究を行う。第一に、本研究では、Mean Shift Clustering の近傍半径を大きく設定したが、今後はより細かくすることで、ミクロな移動傾向を分析する。第二に、移動の目的が現状では不明であるため、Twitter 投稿記事の内容について自然言語処理を行い、移動目的の分類を行う。最後に、本研究では、移動コストに 2 地点間の距離を用いたが公共交通機関などの移動手段についても注目し、より確からしい移動コストの算出方法を検討する。

参考文献

- [石川 88] 石川 義孝: 空間的相互作用モデル—その系譜と体系, 地人書房, 1988
- [Phithakkitnukoon 14] Phithakkitnukoon, S., Teerayut Horanont, T., Witayangkurn, A., Siri, R., Sekimoto, Y., Shibasaki, R.: Understanding tourist behavior using large-scale mobile sensing approach: A case study of mobile phone users in Japan, Pervasive and Mobile Computing, 2014
- [若宮 13] 若宮 翔子, 李 龍, 角谷 和俊: 位置ベース SNS を通じた群衆の移動経験に基づく都市空間の近接性分析, 情報処理学会論文誌, 2013
- [Fukunaga 75] Fukunaga, K., Hostetler, L.: The estimation of the gradient of a density function, with applications in pattern recognition, IEEE Transactions on Information Theory, 1975