

移動体位置データを基にした移動 活動マッチングモデル*

Activity-travel matching model based on mobile positioning data*

寺谷寛紀** 羽藤英二***

By Hiroki Teratani** Eiji Hato***

1. はじめに

近年、RF-IDタグや携帯電話などのIT技術によって、交通-経済データの量と質に劇的な変化が生まれようとしている。しかし、移動体通信システムで得られるドットデータを用いた研究は、データ精度に対応した分析方法が開発されているとは言いがたい。本研究では、時空間ネットワーク上の人の行動を移動-活動配列に置き換えて、膨大なドットデータを直接操作・分析可能な対象へ変換し、意味ある行動文脈を抽出し、分析するアルゴリズムを開発する。

2. データ概要

2001年11月24日(土)に行われた札幌ドームでのJリーグ戦(試合開始時刻 14:00)において以下の調査を実施した。

トラッキング調査:人の動態情報を集積するため、端末が各基地局から受信する電波の電界強度を一定の時間間隔で計測することにより位置座標を特定する移動体通信(PHS, PEAMON)を用いて、位置情報を把握する。

交通状況調査:トラッキング調査結果との比較を目的として、モニターに事前に配布した当日の移動履歴に関する調査票を後日回収・整理する。

アンケート調査: イベント開催時におけるアクセス交通情報の取得状況や要請、携帯端末を活用した新たな交通調査に対する意見等を把握するためにアンケート調査を実施した。

本研究では、得られたドットデータに、得られたアクティビティ・ダイアリー票と情報

* キーワーズ: パターンマッチング, ドットマトリクス法
スコアリングモデル,

** 学生員, 学士, 愛媛大学大学院理工学研究科

*** 正員, 工博, 愛媛大学工学部環境建設工学科

(愛媛県松山市文京町3番, TEL 089(927)9825, FAX 089(927)9843)

意向調査を総合したデータを使用する。但し、アクティビティ・ダイアリー票と情報意向調査は、モデル化するためデータ統合しコードに変換した。そのうち44人の7:00~19:30の間のデータを使用する。

3. アクティビティパターン分析

得られたドットデータは膨大であり分析するにはそのパターンを判断し分析してやる必要がある。ドットデータの文字列から何らかの特徴的なパターンを見出すため、本章では、配列化したドットデータを使用し、1日の移動活動パターンにどのような特徴があるのかについて基礎的な分析を行う。人がどのような行動をしているかを表したものを、アクティビティパターン(Activity-pattern)として定義し、移動活動配列を用いて、個々の人のアクティビティパターンを抽出する。さらに、その抽出したアクティビティパターンから、類似度の高いパターンを抽出するため、アクティビティ分析の1つとしてホモロジー検索を行う。

(1) アクティビティパターンの抽出

活動場所のドットデータからアクティビティパターンを抽出する。抽出方法は、活動場所のドットデータの配列を見ていき、t番目の要素がt-1番目の要素と異なっていたら、アクティビティパターンの配列に新たにラベリングを行ない追加していく。図1に操作イメージ図を示す。



図1 アクティビティパターン抽出方法

(2) ホモロジー検索

ホモロジー検索とは、対象となる配列とホモロジーのある配列が配列データベースに存在するかどうかを検索する手法で、進化・系統分析の解析、たばく質の機能解析などを目的とした配列解析の最も基本的な手法の1つである(遺伝子学において、ホモロジー(homology; 相同性)とは、「共通の祖先遺伝子から由来していること」を意味する.)。本研究では、配列データベースを、(1)で得られたアクティビティパターンとし、ホモロジーのある配列をホモロジー配列と定義づけ、検索を行う。

ホモロジー検索では、一致度を全個人アクティビティパターンに対するホモロジー配列と一致した個人アクティビティパターンの割合として定義する。

検索する配列を「ホモロジー配列として抽出し得る配列」に変化させて、Jリーグ戦観戦客における一致度の高いアクティビティパターンを分析する。ホモロジー配列の候補としては以下の3つの型を考える。

1. 基本型 [自宅] [札幌ドーム] [自宅]
2. [自宅] [札幌ドーム] [] [自宅]型
3. [自宅] [] [札幌ドーム] [自宅]型

ホモロジー検索の方法は図2に示す。この場合、ホモロジー配列のはじめの要素から順番にアクティビティパターンと1文字ずつ一致するかどうか調べていくという方法を用いる。

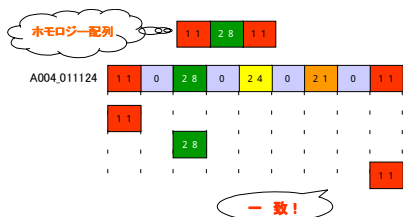


図2 ホモロジー検索

ホモロジー検索を実行し、3つの型のホモロジー配列に対する一致度を表1に示す。

表1 ホモロジー検索結果

1) 基本型								
配列名	1	2	3	4	5	一致人数	全人数	一致度(%)
h1	11	28	11			30	44	68.18

2) [自宅][札幌ドーム][] [自宅]型								
配列名	1	2	3	4	5	一致人数	全人数	一致度(%)
h2	11	28	12	11		0	44	0.00
h3	11	28	16	11		0	44	0.00
h4	11	28	21	11		5	44	11.36
h5	11	28	22	11		4	44	9.09
h6	11	28	11			1	44	2.27
h7	11	28	24	11		6	44	13.64
h8	11	28	26	11		1	44	2.27
h9	11	28	30	11		1	44	2.27
h10	11	28	36	11		2	44	4.55

3) [自宅][札幌ドーム][自宅]型								
配列名	1	2	3	4	5	一致人数	全人数	一致度(%)
h11	11	12	28	11		1	44	2.27
h12	11	16	28	11		1	44	2.27
h13	11	21	28	11		9	44	20.45
h14	11	22	28	11		0	44	0.00
h15	11	23	28	11		3	44	6.82
h16	11	24	28	11		9	44	20.45
h17	11	26	28	11		1	44	2.27
h18	11	30	28	11		0	44	0.00
h19	11	36	28	11		3	44	6.82

(3) 分析結果

・特徴1 : h1の基本形である[自宅] [札幌ドーム] [自宅]のアクティビティパターンが68.18%と最も高い。基本型([自宅] [札幌ドーム] [自宅])は、札幌ドームでのJリーグ戦観戦客が対象であったため、すべての人のアクティビティパターンに一致し、一致度が100%になると考えられる。但し、ホモロジー検索結果は約7割の人としか一致しなかった。この原因は、データの取得範囲の設定にあると考えられる。本研究では、調査時間を7:00から19:30までとした。よって、19:30までに帰宅しなかった人は基本型と一致しなかったためと考えられる。

・特徴2 : Jリーグ戦観戦客は、家を出て札幌ドームに行く途中、何らかを買ったり、食べたりした後向かう...2番目に一致度が高いのは、h13([自宅] [スーパー・コンビニ] [札幌ドーム] [自宅])とh16([自宅] [レストラン・喫茶店・居酒屋] [札幌ドーム] [自宅])の20.45%である。理由として、Jリーグ戦の開始時間が14:00であったため、札幌ドームに行く前にスーパー・コンビニで昼食を買ったり、レストラン・喫茶店・居酒屋で昼食を食べてから向かう人が多かったことが挙げられる。

・特徴3 : Jリーグ戦観戦客は、自宅に帰る途中、夕食を済まして帰る...3番目に一致度が高いのは、h7([自宅] [札幌ドーム] [レストラン・喫茶店・居酒屋] [自宅])で13.64%である。これは、札幌ドームでサッカー観戦をして、熱狂冷めやらぬうちに居酒屋で飲んだり、夕食を外で済ましてから自宅に帰る人が多かったと考えられる。

4. パターンマッチング分析

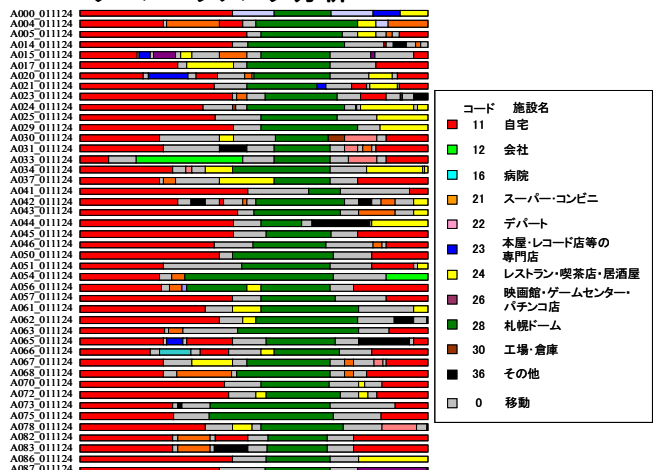


図3 アクティビティパターン図

図3にドットデータを視覚化した図を示す。色は施設、長さは活動時間を表している。ドットデータを使用し、2人のある一定の長さをもつ移動-活動配列を比較した際に異なる個人間の移動-活動パターンにどのような特徴があるか分析する。図4に例として2人のアクティビティ配列を示す。2つの配列を比較する方法として、ドットマトリクス法とスコアリングモデルとSequence Alignment Method(SAM)がある。

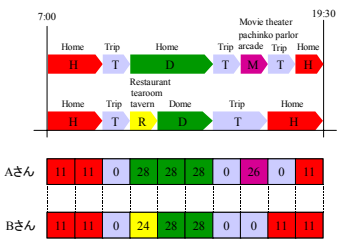


図4 移動-活動配列

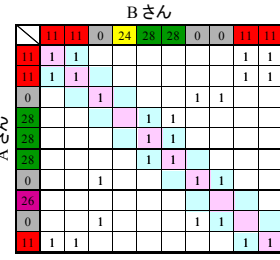


図5 ドットマトリクス法

(1) ドットマトリクス法

ドットマトリクス法とは、図4に示す比較する配列を列方向、行方向に並べ、対応する要素が一致したとき1、一致しないとき0を値として持つマトリクス(図5)を考える方法である。移動-活動配列間で相同性のある領域を図5において、ハッチングして示す。この領域では、時間軸上の意思決定の不確実性を考慮できないので、隣接したk要素(斜線部)を加えた領域を相同領域とする。2つの配列の相同性の尺度、一致度M(%)は相同領域における一致した要素数の割合とする。

(2) スコアリングモデル

ドットマトリクス法では、対応する要素が一致したときの重みを表すことができない。例えば、一日の行動の中で、自宅にいる時間より、コンビニにいる時間の方が短い。よってある時間、自宅にいて一致するより、コンビニにいて一致する方が確率的に低いことが分かる。この重みを考慮するために、本研究ではスコアリングモデルの概念を導入する。

・表記法の定義
比較する2つの配列を配列X, Yとし、それぞれの配列の長さをm, nとする。x_iをXのi番目コードとし、y_jをYのj番目コードとする。ここでは、m=nの場合を考える。

与えられた2つの配列について、(配列間に何らかの関連性がある)/(配列間に何の関連性もない)で表される対数尤度の尺度に基づいたスコアを割り当てる。この場合、2つの配列から配列間に何らかの関連性がある確率と配列間に何の関連性もない確率を推定し、その比を考える。

まず、配列間に何の関連性もない場合のモデル、ランダムモデル(random model)Rを考える。コードaが独立に頻度q_aで観察されると仮定すると、与えられた2つの配列が偶然観察される確率は、2つの配列の各位置におけるコードの観察頻度を掛け合わせたものである。

$$P(X, Y | R) = \prod_{i=0}^m q_{x_i} \prod_{j=0}^n q_{y_j} \quad (1)$$

次に、配列間に何らかの関連性がある場合のモデル、一致モデル(match model)Mを考える。一致モデルでは、Xのt番目のコードをa, Yのt番目のコードをbとすると2つの配列のt番目のコードペアは同時確率p_{ab}で観察されると考える。

$$P(X, Y | M) = \prod_{i=0}^n p_{x_i, y_i} \quad (2)$$

これら2式の尤度比を、オッズ比として以下の式(2.3)で定義する。

$$\frac{P(X, Y | M)}{P(X, Y | R)} = \frac{\prod_{i=0}^n p_{x_i, y_i}}{\prod_{i=0}^n q_{x_i} \prod_{i=0}^n q_{y_i}} = \prod_{i=0}^n \frac{p_{x_i, y_i}}{q_{x_i} q_{y_i}} \quad (3)$$

ここで、オッズ比の対数をとることによって、対数オッズ比Sと呼ばれる加算的なスコアリングシステムを導出することができる。今回、この対数オッズ比Sを2つの配列の一致度と定義する。

$$S = \sum_{i=0}^n s(x_i, y_i) \quad (4)$$

ここで、

$$s(a, b) = \log \left(\frac{p_{ab}}{q_a q_b} \right) \quad (5)$$

は、配列間に何の関連性もないコードペア確率に対する配列間になんら関連性のあるコードペア確率の対数尤度比である。この対数尤度比s(a,b)を2つの配列XのコードaとYのコードbの相同性スコアとする。

(3) Sequence Alignment Method(SAM)

Willson (1998) が導入した方法で、分子生物学、クロマトグラフィー及び音声認識のような学問の用いられているシーケンス比較法の一つである .SAM によって、移動-活動パターンの比較において、指標に順序情報が入れ込まれる .SAMの配列間の一致尺度(類似尺度)としては、2つの配列を等しくさせるための最小操作とする .等しくさせるための操作として『削除(deletion)』、『挿入(insertion)』、『置換(substitution)』、『一致(identity)』がある .

	ml	11	11	0	28	28	28	0	0	11	11
ml	0	1	2	3	4	5	6	7	8	9	10
11	0	1	2	3	4	5	6	7	8	9	
11	2	1	0	1	2	3	4	5	6	7	8
0	3	2	1	0	1	2	3	4	5	6	7
28	4	3	2	1	2	1	2	3	4	5	6
28	5	4	3	2	3	2	1	2	3	4	5
28	6	5	4	3	4	3	2	3	4	5	6
0	7	6	5	4	5	4	3	2	3	4	5
26	8	7	6	5	6	5	4	3	4	5	6
0	9	8	7	6	7	6	5	4	3	4	5
11	10	9	8	7	8	7	6	5	4	3	4

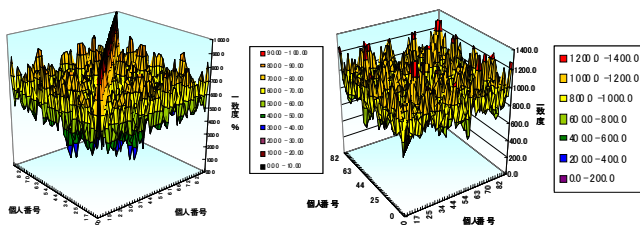
矢印	操作	コスト
→	挿入	1
↔	置換	2
↓	削除	1
↔	一致	0

図6 comparison table

図6に図4に示す比較する配列に対してSAMを適用した例を示す .図6の各セルの値は、そのセルに到着するために必要とされる最小コストとする .それ故、図6のハッチングしてあるセルの値は、比較する配列を等しくさせるための最小コストを示している .

(4) 分析結果

SAMは、移動-配列パターンの順序は考慮されているが、時刻による一致が考慮されない .本研究では、短時間の需要の集散に関する位置データであるため、時間の取扱いが厳密であることが望ましい .このため先に述べた3つの方法のうち、時刻による一致を考慮したドットマトリクス法とスコアリングモデルを調査サンプル44人に対して適用し、異なる個人間の一致度を算出する .図7に算出されたそれぞれの一一致度を示す .

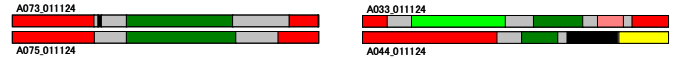


a) ドットマトリクス法 b) スコアリングモデル

図7 一一致度

さらに、一致度の最も高いペアと低いペアを抜き出し、アクティビティパターン図の比較をそれぞれ図8に示す .

a) ドットマトリクス法



b) スコアリングモデル

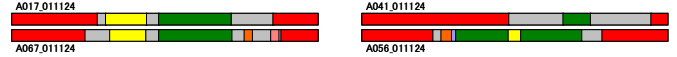


図8 個人間比較

(5) アルゴリズム比較

2つのアルゴリズムの指標において最も一致度の高いペアについての比較を図9に示す .図9に示す通り、指標による評価の逆転が生じている .これは一致度の定義の違いに起因する .ドットマトリクス法の指標はある時刻にある活動場所で一致すれば、一致度が高くなり、文字間の一致による指標として識される .一方、スコアリングモデルでは、文字ペア間において1日の活動におけるそのペアの一致する確率を考慮している .つまり、文字が一致しても文字ペア間の一致する確率が高ければ、スコアリングモデルでは一致度は高くない .

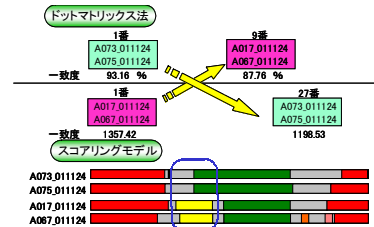


図9 アルゴリズムの指標比較

5 .まとめと課題

位置データを独自に開発したアルゴリズムを適用し、直接操作する移動-活動パターン分析の方法を提案した .今後の課題として、膨大なday-to-dayの位置データへの適用することで行動パターンの摂動メカニズムの解明を考えている .

<参考文献>

- 1) Arentze, T., Hofman, F.C., Joh, H., Timmermans, H. Activity pattern similarity: a multidimensional sequence alignment method . *Transportation Research B* 36, pp.385-403 . 2001 .
- 2) 羽藤英二, 朝倉康夫 . 時空間アクティビティデータ収集のための移動体通信システムの有効性に関する基礎的研究, *交通工学*, vol.35, No.4, pp.19 - 28 . 2000 .