

# 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

愛媛大学大学院 学生員 寺谷寛紀  
愛媛大学工学部 正会員 羽藤英二

## 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．データとしては，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)における調査で得られた44人の7:00~19:30のデータを使用する．

## 2. パターンマッチング分析

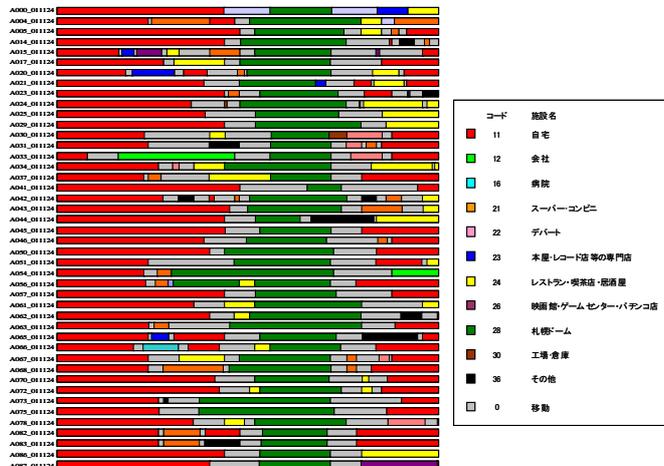


図1 アクティビティパターン図

図1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1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2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がある．

### (1)ドットマトリクス法

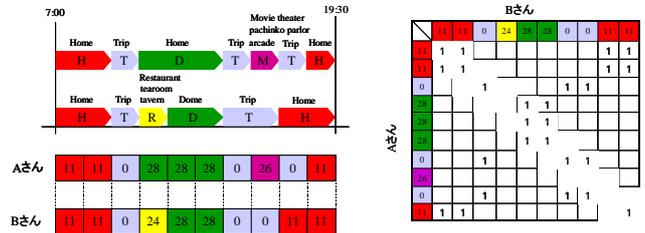


図2 移動-活動配列

図3 ドットマトリクス法

ドットマトリクス法とは，図2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3)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3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番目のコードとし， $y_j$ 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

キーワード: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  
連絡先:〒790-8577 愛媛県松山市文京町3番、089(927)9825、FAX089(927)9843

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．

$$P(X, Y | R) = \prod_{i=0}^m q_{x_i} \prod_{j=0}^n q_{y_j} \quad (2.1)$$

次に，配列間に何ら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．

$$P(X, Y | M) = \prod_{i=0}^n p_{x_i y_i} \quad (2.2)$$

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．

$$\frac{P(X, Y | M)}{P(X, Y | R)} = \frac{\prod_{i=0}^n p_{x_i y_i}}{\prod_{i=0}^n q_{x_i} \prod_{i=0}^n q_{y_i}} = \prod_{i=0}^n \frac{p_{x_i y_i}}{q_{x_i} q_{y_i}} \quad (2.3)$$

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．

$$S = \sum_{i=0}^n s(x_i, y_i) \quad (2.4)$$

ここで，

$$s(a, b) = \log \left( \frac{p_{ab}}{q_a q_b} \right) \quad (2.5)$$

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何らかの関連性のあるコードペア確率の対数尤度比である．今回，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)分析結果

本研究では，調査サンプル 44 人に対してドットマトリクス法とスコアリングモデルを適用し，異なる個人間の一致度を算出する．図 4 に算出されたそ

れぞれの一致度を示す．

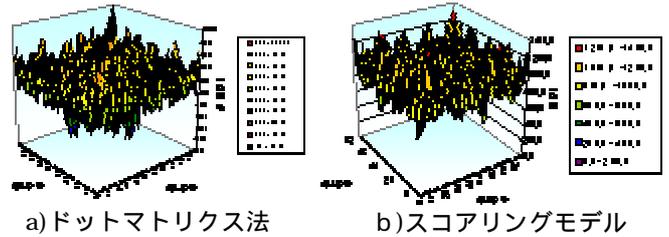


図 4 一致度

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 5 に示す．

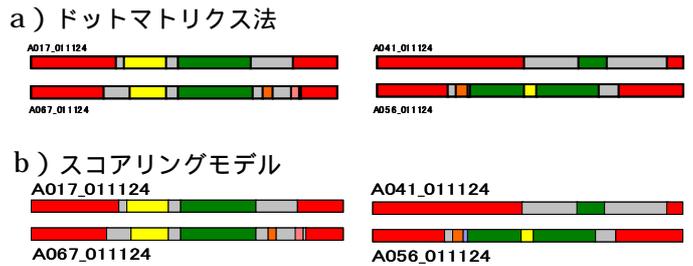


図 5 個人間比較

(4)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 6 に示す．図 6 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．

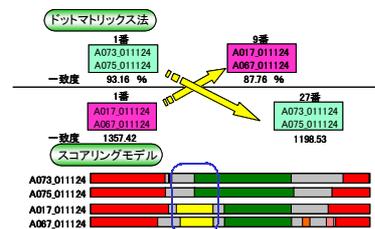


図 6 アルゴリズムの指標比較

4.まとめ

ドットデータを独自開発したアルゴリズムで直接操作し分析することに成功した.ドットマトリクス法は，移動-活動パターンが単純になる程，各活動時間が長くなり一致度が高くなる.スコアリングモデルでは,1日の活動で確率的に低い文字ペアが一致するとスコアが大きくなり,一致度が高くなる．