

IV-5 ドットデータに基づいたアクティビティマッチングアルゴリズムに関する研究

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いいがたい。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。

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。

①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。

②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。

③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、①で得られたドットデータに、②、③で得られたアクティビティ・ダイアリー票と情報利用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利用意向調査は、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. パターンマッチング分析

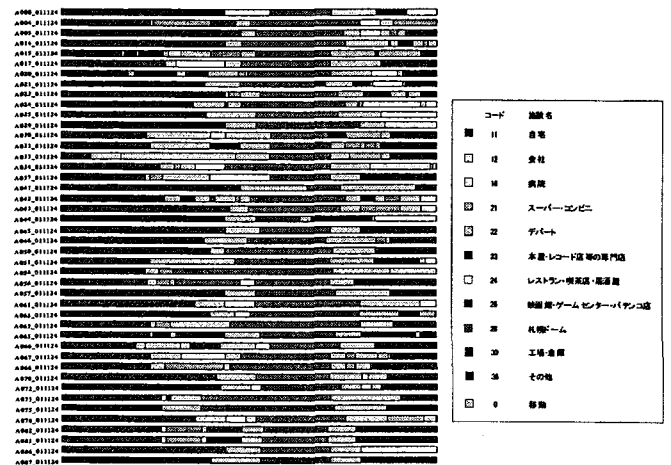


図1 アクティビティパターン図

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。

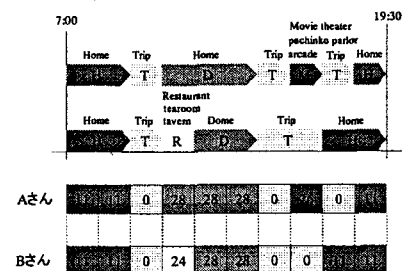


図2 移動・活動配列

2つの配列を比較する方法として、ドットマトリクス法がある。ドットマトリクス法とは、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を考える方法である。この方法では、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、スーパー・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、スーパー・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。

(1)スコアリングモデル

・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} で観察され则认为.

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

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.

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

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.

(2)分析結果

本研究では、調査サンプル44人に対して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3に算出された一致度を示す.

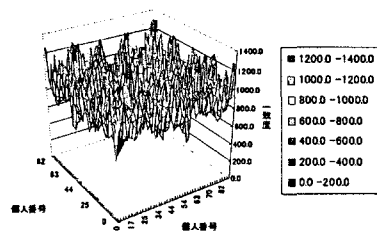


図3 一致度

一致度の高い順に5ペアと最も一致度の低いペアを抜き出し、個人属性、活動配列数、アクティビティパターン図の比較を図4に示す.

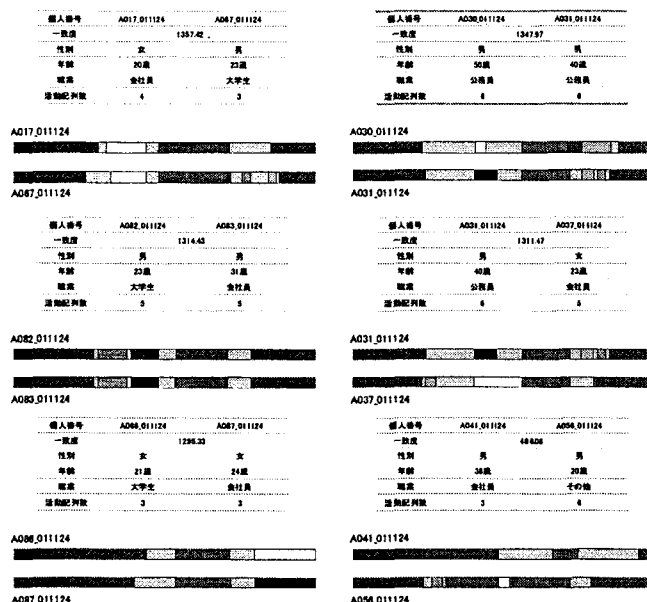


図4 個人間比較

4. まとめ

膨大なドットデータを独自で開発したアルゴリズムに適用して直接操作し分析することに成功した.

スコアリングモデルは、活動配列数が多くなることによって、個々の活動時間が短くなりペア間で一致する確率が低くなる. それにもかかわらず、ペア間で一致するとスコアが大きくなり、一致度は高くなる傾向がある.