

# 個人の行動パターンのシーケンシャルマイニング\*

## Sequential mining of an individual activity pattern\*

武智環\*\*・羽藤英二\*\*\*・渡部眞幸\*\*\*\*・中山沖彦\*\*\*\*\*

By Tamaki TAKECHI\*\*・Eiji HATO\*\*\*・Masaki WATANABE\*\*\*\*・Okihiko NAKAYAMA\*\*\*\*\*

### 1. はじめに

近年、人の行動の多様化は進み、それぞれが異なった生活スタイルを持つ。結果、個人の日々の行動から得られる目的地選択肢集合は異なり、全体として行動の予測を立てることは困難である。そこで、本研究では、日々の行動の中に存在するであろう個人の周期的な行動に着目した。出発・目的地施設、訪問日時、といった情報のみから、モニター固有の行動パターンの周期性を見出す。そして、モニターにとってその施設はどれくらいの周期・タイミングで訪れるのか、訪問施設の前後関係や時刻・曜日などの、その施設に行く条件は、といったルールを、マイニングを行うことで評価する。そして、そのような条件で訪れる施設というのはどのような施設かということを予測する。出発地目的地、出発到着日時について解答されたアンケートデータを用いて個人の活動の周期性を分析した。

### 2. データ概要

#### (1) 調査方法とサンプルの特徴

本稿で使用するデータは、いつ・どこへ・どのような目的で訪問したかという行動を、自動車トリップに限定して調査した約2ヶ月間の行動調査によるものである。調査期間は、2007年12月28日から2008年2月29日であり、調査項目は走行目的・出発地・経由地・最終目的地、さらに各々について出発時刻・到着時刻・施設名称・所在地を自宅から目的地間の行き帰りについてweb上で記入

\*キーワード：目的地選択、活動分析

\*\*学生員、愛媛大学大学院理工学研究科

(〒790-0826 愛媛県松山市文京町3,

E-mail : takechi.tamaki.04@cee.ehime-u.ac.jp)

\*\*\*正員、工博、東京大学大学院工学系研究科都市工学専攻

(〒113-8656 東京都文京区本郷7-3-1,

E-mail : hato@bin.t.u.tokyo.ac.jp)

\*\*\*\*日産自動車株式会社 電子技術開発本部 IT&ITS開発部

(〒243-0123 神奈川県厚木市森の里青山1-1

E-mail : wata-masaki@mail.nissan.co.jp)

\*\*\*\*\* 同上 (E-mail : nakayama@mail.nissan.co.jp)

してもらった。

こうして集まった128名の回答データのうち、分析に当たり、データクリーニングを行った。本稿では、行動パターンの周期性について検証を行う。そこで、行動に連続性のないトリップを一つでも含むモニターを分析対象から外した。例えば、自宅を出発し勤務地へ向かったが、次のトリップの出発地が勤務地ではなく自宅になっているといったものである。これは、周期性を見る際に、本来周期性をもつ行動であったとしても、一度でもその目的地に訪れたという記録が飛んでいると、周期が途切れると考えたためである。以上を踏まえてデータクリーニングを行った結果、16名分のモニターデータに絞られた。表-1に、抽出されたモニターの属性など、基礎データを示す。

表-1 個人属性

No	moniID	年齢	性別	職業	都道府県	No	moniID	年齢	性別	職業	都道府県
1	44700	52	男性	会社員	愛媛県	9	190129	50	女性	その他	奈良県
2	51629	47	男性	会社員	大阪府	10	200913	25	女性	専業主婦	東京都
3	53864	45	男性	自営業	神奈川県	11	205003	42	男性	会社員	大阪府
4	65724	57	男性	会社員	東京都	12	209546	55	女性	公務員	福岡県
5	100608	54	男性	会社員	北海道	13	215911	64	男性	自営業	神奈川県
6	108299	55	男性	会社員	石川県	14	219295	30	男性	会社員	埼玉県
7	111392	28	女性	パート	茨城県	15	232914	45	女性	会社員	山梨県
8	116138	50	男性	その他	愛媛県	16	233231	29	女性	専業主婦	愛知県

モニターは北海道から九州まで全国各地に広がっており、男女比は5:3で、比較的年齢層が高い。これは、今回の調査の特徴から、自動車を日常で利用している人が調査対象になっているためではないかと考えられる。ゆえに、職業では会社員などの働いている人が目立つ。

#### (2) 基礎集計結果

この調査では、目的地施設の登録をユーザー自身に行ってもらうため、ユーザーごとに登録される施設の数、種類が異なる。さらに、車を利用して出かけた日数も異なるため、調査参加日数も異なる。そこで、調査終了時点で登録されていた目的地施設登録数を調査参加日数で割ったものを、新規目的地数とし、図-1に平均トリップ数と合わせて表示した。

目的地数・トリップ数ともに平均値にラインを引いた。これを基準として、目的地数・トリップ数ともに多いタイプは活動的なユーザーといえる。反対に両方少ないタイプはあまり移動を行わないユーザー、また、目的地が

少なくトリップが多いタイプはきまった場所に頻繁に訪れるユーザー、目的地が多くトリップ数が少ないタイプは一度の移動で立ち寄りの多いユーザーといえる。

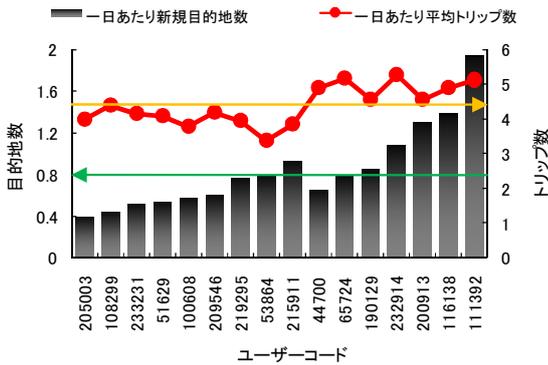


図-1 トリップ数と施設数

### 3. 分析結果

図-2 に示すのは、出現パターンと出現回数の分布である。パターンの詳細は省き、出現頻度の分布をみることとした。期間中1度しか出現しないパターンが圧倒的に多く、75%を占めていた。さらに、以降の分析のために制限を与えた出現回数3以上のところに実線を引き区別したところ、全体の約12%が分析対象となった。3回以上出現したパターンの平均出現回数は8.2回、標準偏差は8.5回であった。こうして得られたパターンを分析の対象とした。

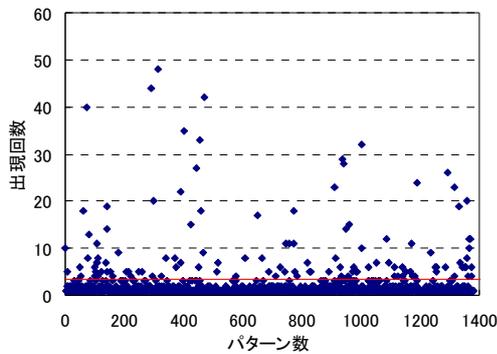


図-2 パターンの頻度分布

#### (1) マイニング分析

ここでは、実際に移動があった2施設間のデータにバスケット分析を適用することによって『Aに行った後、Bへ行く』等の行動文脈の抽出を行う。ルール抽出にはサポート値・確信度・リフト値の3つの指標を用いる。それぞれの算出方法は以下の通りである。

$$\text{サポート値} : \text{Support} = \frac{A \cap B}{S} = P(A \cap B) \quad (3.1)$$

$$\text{確信度} : \text{Confidence} = \frac{A \cap B / S}{A / S} = \frac{P(A \cap B)}{P(A)} \quad (3.2)$$

$$\text{リフト値} : \text{Lift} = \frac{A \cap B / S}{A / S \cdot B / S} = \frac{P(A \cap B)}{P(A) \cdot P(B)} \quad (3.3)$$

A : 条件部を満たす個数

B : 結論部を満たす個数

S : 全個数

サポート値は、全体を1とした時、どの程度条件Aと結論Bが同時に発生しているかという割合を表す。この値が小さいと、その事象は単なる偶然であり、有用な知見ではない。確信度は、条件Aが発生しているうち、どのくらい結論Bも発生しているかを示し、この数値が大きいと、条件Aと結論Bの関連性は高い。リフト値は結論Bの独立性をみる指標となり、1以下であれば、条件Aとは無関係に結論Bが生じ、リフト値が高いと、結論Bのためには条件Aが必要であるという指標になる。

今回は、サポート値0.01以上、確信度0.5以上、リフト値1以上と条件を設定し、抽出した。

移動のあった2施設間においてバスケット分析を行い、同時生起確率の高い組み合わせを抽出する。対象となる組み合わせパターンデータは、期間中3回以上出現しているパターンとする。ここでは、確信度が0.8以上のものを抽出した。確信度が高いということは、そのルールは確かなものであり、施設でみれば、その組み合わせで訪れられることが多いということになる。その後、そのパターンの出現頻度を検証し、周期性について考察を行う。また、確信度が0.8以下のものについては、パターンの発生間隔に着目し、周期性の評価を行う。

発生間隔は、その平均・標準偏差・発生回数から8つのパターンに分類し評価する。

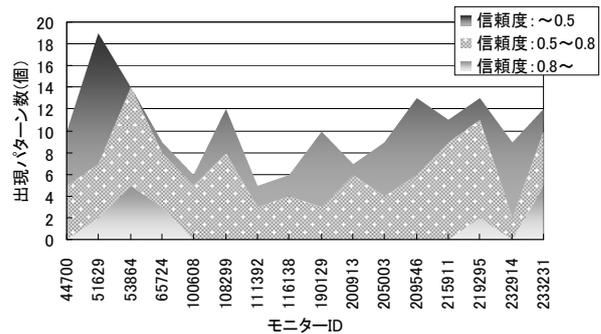


図-3 確信度別パターン数

図-3は、バスケット分析によって得られた確信度を0.5以下、0.5~0.8、0.8以上としてモニターごとにその域に含まれるパターン数を示したものである。確信度0.8以上のパターンを持たないモニターもいることがわかる。さらに、約6割のパターンが確信度0.5~0.8の間にある。すべてのパターンに対して確信度の高いパターンが少ないモニターは、一日の行動がパターン化されにくく、日々の行動がまちまちであると考えられる。また、

確信度の高いパターンの割合が高いモニター（モニターID233231, 53864 など）は、決まった行動パターンをとりやすいと推測される。

表-2 に、バスケット分析の結果、確信度が 0.8 以上として得られたパターンを詳細に挙げる。表中の count はそのパターンの出現回数を示す。サポート値の高いパターンは、その施設間を連続で訪れることが多いことを示しており、確信度の高いパターンは、その施設は必ず連続で訪れているということである。確信度 0.8 以上のものを対象としているため、多様な組み合わせをもつ施設は抽出されず、限定された組み合わせのパターンしか出てこない。サポート値が低いものは、出現回数が 3~5 回というもので、全体に占める割合が少ないにもかかわらず、A に行くときには B にも必ず行くといった強い関連性を持っている施設だと言える。

表-2 確信度 0.8 以上のパターン

monitorCD	条件部A	結論部B	count	support	confidence	lift
51629	HC.1	自宅	5	0.05	0.83	1.31
51629	自宅	HC.1	5	0.05	0.83	1.31
53864	客先	自宅1	3	0.04	1.00	1.20
53864	客先	自宅2	3	0.03	0.83	1.00
53864	自宅	客先	3	0.04	1.00	1.20
65724	会社	取引先	4	0.03	0.80	1.11
65724	自宅	会社	48	0.39	0.80	1.11
65724	取引先	会社	4	0.03	0.80	1.11
219295	テニスコート	自宅	4	0.05	0.80	1.20
219295	自宅	テニスコート	4	0.05	0.80	1.20
233231	SC.1	自宅	19	0.23	0.98	1.26
233231	SC.2	自宅	7	0.09	0.94	1.21
233231	自宅	SC.1	20	0.23	0.98	1.26
233231	自宅	LE	12	0.14	0.86	1.11
233231	LE	自宅	12	0.14	0.86	1.11

(SC : ショッピングセンター, HC : ホームセンター)

## (2) 周期性に関する基礎分析

ここからは、個人に着目して周期性の分析を行う。パターンの出現回数、パターンの出現間隔、出現間隔の標準偏差の大小の組み合わせを用いてパターンの出現タイプを 8 項目のカテゴリーに分類した。分類結果とそれに対応する評価方法を表-3 に示す。例えば、①は、出現回数 7 回以下 and 出現間隔 7 日以上 and 出現間隔の標準偏差 7 日以上といったように、7 回、7 日を基準として分類した。これは、データ取得期間の一番短いモニターが約 7 週間であったためである。7 回以上であれば、平均的に考えて毎週出現しているパターンだと言える。また、曜日による周期性の出現可能性から、間隔は 7 日とした。7 日以下であれば、1 週間を待たずに出現しているパターンだと言える。図-4 にモニターごとの各分類項目の割合を示す。②と⑧の割合が多く、次いで①が多い。⑧の数と①・②では反比例の傾向がみえる。⑦は全く見られなかった。さらに、③、⑤は、この分類に当てはまらないパターンが多く、出現しても少数であった。モニターID108299, 233231 といった②、④、⑥、⑧のパターンが多く見られるモニターは定期的・周期的な行動が多く見られるとわかる。一方、モニターID116138 のような①のパターンが多いモニターからは、周期的な

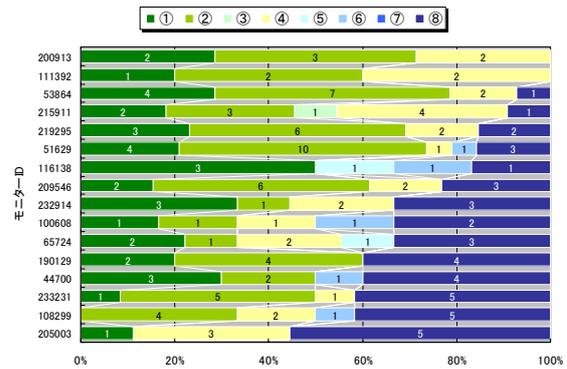


図-4 モニター別パターン分類結果

表-3 パターン分類

	間隔大		間隔小	
	標準偏差大	標準偏差小	標準偏差大	標準偏差小
回数少	①訪れる間隔がまちまち、時々しか行かない。	②間隔を空けて定期的に行く、実家など。	③突発的に行く	④短期間に日を空けずに少し行く。
回数多	⑤期間をあけてよく行くが間隔はばらばら。	⑥間隔を空けて、定期的によく行く。	⑦短期間によく行くが、その間隔はばらばら。	⑧頻繁に定期的に行く。

行動は観測されにくい。社員がモニターに多いことから、⑧のような頻繁に定期的に行くといった施設は観測されやすいのだと考えられる。必然と毎朝会社に行かねばならないため、周期的な行動となっておかしくないはずである。パターンの出現回数、パターンの出現間隔、その標準偏差の組み合わせで、周期的行動を行いやすいモニターと、そうでないモニターを判別することが可能ではないかと考える。詳細な行動パターンの予測は、日々行動が変化するために難しいが、カテゴリー単位なら、そこに含まれるパターンのいずれかを選択するといった予測ができるのではないだろうか。

表-4 32 日以降の選択パターン（モニターID : 233231）

	②	④	⑧	new	total
32日目	2	0	2		
33日目		5	5		
34日目	2	0	2		
35日目	2	0	2		
37日目		3	3		
38日目	2	2	4		
39日目	2	0	2		
40日目	2	0	2		
41日目	1	5	5		
42日目		2	2		
43日目	1	4	6	7	
44日目	2	0	2		
45日目	2	2	4		
46日目	2	2	4		
47日目		2	2		
48日目	1	3	3		
49日目	2	2	4		
50日目	2	2	4		
51日目	2	2	4		
52日目	2	2	4		
53日目	1	1	2		

表-4 は、調査 1 か月終了後からのパターンの出現の様子を示している。このモニターは、この期間②、④、⑧しか出現せず、また、1 か月たった後でも新たなパターンが出現していることがわかる（表中の new）。45 日までは、頻繁に行っていた行動パターンが、48 日以降変化し、新たなパターンを形成している。⑧から④へ変わったということから、今までとは異なった行動をとるようになり、さらに新たな繰り返しの行動がとられるようになったと言える。このカテゴリー選択の周期性を

知ること、行動パターンの予測がより容易になると考えられる。そこで、実際に観測されたパターンと分類を比較検証した。(図-5)

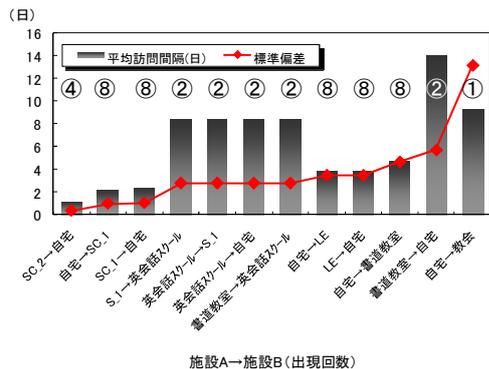


図-5 訪問間隔と標準偏差 (モニターID : 233231)

【SC\_2→自宅】のパターンは期間中 7 回出現、出現間隔は 1 日で、標準偏差も小さいことから④に分類される。そのほかのパターンについても、図中のような分類になり、このモニターは②、⑧の分類が多いことから、周期的な行動をとりやすいと言える。英会話スクール、書道教室など曜日が固定される行動パターンが多くみられることから、周期的な行動は、曜日を決めて行う行動・日々必然的に発生する行動、といった制約のある行動から生まれるのではないかと考えられる。例えば、通勤・通学、幼稚園の送迎、レンタル、クリーニング、スーパーの特売、教会、習い事等は、社会的制約が生じ、周期性を生み出しやすい行動となる。日々の行動のうち、これらの行動の占める割合が高いと行動に周期性が生まれ、次に行くだろう施設の予測が立てやすくなると考えられる。

これまでの分析では、必ず 3 日間隔で訪れるといった行動を対象としてきた。しかし、実際は、2 日連続で訪れて、1 日空けてまた 2 日連続で訪れるといった変則的ではあるが、周期的な行動をとることもある。このような行動も同様にして出現間隔を平均値・標準偏差として算出してしまうと、周期的でないとして評価されてしまうため、出現間隔をパターンごとに見てみることにした。図-6 は、例えば【書道教室→自宅】は最初に出現してから 7 日後に再び出現し、次に 21 日後に出現、といったことを示している。【自宅→SC\_2】のような直線で表現されるパターンは、常に一定の間隔で出現し、周期的な行動を示す。【英会話スクール→自宅】のように途中から間隔が変わるパターンは、出現間隔を平均値などで評価してしまうと、折角出現した周期性を見落としてしまう。これらをよく見てみると、【自宅→LE】(LE: 幼児教室)のように、0 日・7 日・0 日・7 日と間隔は違いますが、周期のある行動となっており、これらを実評価していく必要がある。このように、出現間隔の推移を時系列で

ならべることにより、これまで見落としてしまっていた活動の周期性の発見につながると考えられる。

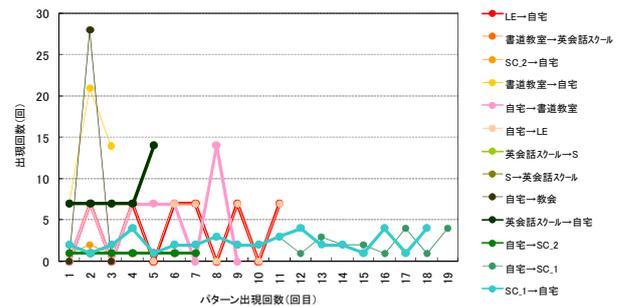


図-6 パターンの出現間隔 (モニターID : 233231)

#### 4. まとめ

今回、まず、マイニングによってどれほど行動パターンを抽出することが出来るかを検証した。その組み合わせの強さを表す確信度の高いものを抽出したが、全モニターから 15 パターンしか抽出されなかった。これは、今回期間中 3 回以上出現したパターンに限定したことから、意思を持ってとった行動に限定されたためだと思われる。例えば、ある日突然訪問して、その後ほとんど訪問しなかったというような施設への行動パターンが除外される。バスケット分析の性質上このような稀なパターンだと確信度が高く評価されてしまうのだが、3 回以上という閾値を設定することで、解消された。

また、行動がパターン化されるということから、そのパターンの出現には周期性が見られるのではないかと推測がたつた。本稿では、周期性の指標として、パターンの出現間隔の平均値・標準偏差・出現回数を用いて 8 つの項目に分類し評価した。結果、周期的な行動をとりやすいモニター、そうでないモニターを把握することができたのではないかと考える。さらに、個人についてさらに詳細にパターンの周期性を検討していくと、変則的ではあるが、周期的といえる行動パターンを見つけることができた。今後これらを定量的に評価していく必要がある。さらには、行動の履歴から、将来の行動を予測していき、その的中率についても分析を行っていく。

#### 参考文献

- 1) 平手勇宇, 山名早人: 時間情報を含むシーケンシャルパターンマイニングの一般化, 第 4 回日本データベース学会年次大会(DEWS2006)
- 2) 平手勇宇, 小松俊介, 山名早人: イベント発生時間間隔を考慮したシーケンシャルパターンマイニング, 情処研報, (DBS), Vol.2005, No.137, pp.321-328, 2005
- 3) 山城貴久, 平野靖, 梶田将司, 間瀬健二: ペイジアンネットによるユーザ行動モデルの体験記録からの学習, 第 4 回情報科学技術フォーラム一般講演論文集, pp.79-80, 2005