

行動の類似性に基づく活動分類手法の提案*

A Method of Activity Classification based on Behavioral Similarity*

力石真**・張峻屹***・藤原章正****

By Makoto CHIKARAISHI**・Junyi ZHANG***・Akimasa FUJIWARA****

1. はじめに

交通施策を検討するプロセスにおいては、活動の種類を必須活動、買物活動、自由活動等に分類することで、活動種類間の行動特性の差異を表現することが頻繁に行われている。最も簡便かつ頻繁に行われている活動分類の方法は、データ収集時に設定した活動分類を用いる方法である。すなわち、先見的に決定された活動分類を直接、またはより粗く再分類したものを利用する方法である。しかしながら、このようなア priori な方法は、活動を分類することによって、どの程度行動特性の差異が反映されているのかが明確ではない点が問題である。また、一度活動分類が完了すると、同一セグメント内において行動特性は同質であると仮定されることが多いため、セグメントの与え方によって分析結果の含意は大きく異なると考えられる。

一方で、収集したデータから事後的に活動の分類セグメンテーションを行う方法がある。事後的な分類は、特定の問題に適した分類や、表現したい行動特性に合わせた分類が可能であり、活動分類の意図が明確であるというメリットを持つ。例えばDoherty¹⁾は、Hägerstrand²⁾の研究から示唆されるActivity-basedアプローチの重要な視点「活動/交通行動が持つ時空間上の特性(時空間制約や他者との共同活動等による制約)」は、従来のデータ収集時に設定した活動分類(就業活動、買物活動、自由活動等)では十分に反映されないと主張し、1) 活動遂行の時空間上のフレキシビリティ、2) 他者との共同活動による制約を考慮した新たな活動分類を、主成分分析により導出している。また、調査手法においても、事後的な活動分類が行えるよう工夫がなされているものがある。例えばAmerican Time Use Surveyでは、様々な分析目的への柔軟な対応を可能にするため、3層(大分

類、中分類、小分類)にわたる活動分類システムを構築している³⁾。

以上簡単にみたように、活動の分類(セグメント)方法には、大別して先見的な手法と事後的な手法が存在する。本研究では、不適切な活動分類を採用してしまうと、本来活動が持っている行動特性の情報を失ってしまう恐れがあるとの認識のもと、行動特性の類似性を考慮した事後的な活動分類手法を新たに提案する。提案する活動分類手法を用いることによって、従来の先見的な活動分類と比較して、現象の特性に適した分析が可能になるものと考えられる。特にトリップチェーンや活動間の相互作用等を分析の対象とする場合、分析結果が活動分類の仕方に大きく依存するものと考えられるため、活動分類の方法がより重要な位置を占めるとと思われる。提案方法の有効性を確認するため、活動開始時刻の選択を対象として実証分析を行う。なお本研究では、細分化された活動種類(ここでは61種類)を少数の活動分類に集約する方法を対象とする。

2. 分析方法

マーケティング分野では、事後的なセグメントの方法として、主成分分析、クラスター分析をはじめとする多変量解析や、ニューラルネットワーク、混合モデルなど、様々なセグメントの方法が提案されている⁴⁾。特に消費者のニーズや特徴、行動をベースにした市場志向のセグメントの重要性が認識されており、潜在クラスモデルによるセグメント⁵⁾、Mixed logitモデルによるセグメント⁶⁾等、消費者の行動特性を考慮したセグメントが広く行われている。交通計画においても、活動分類が行動特性の差異を表現する目的で行われているとすると、上述のような行動特性に基づいたセグメント法に基づき分類を行うことが望ましいと考えられる。

マーケティング分野におけるセグメント手法は、主に「個人」を複数のセグメントに分類することを意図している。しかしながら、交通行動や活動の遂行は多くの不確定な変動要因の影響を受けるため⁷⁾、個人のセグメントを主な目的とするマーケティング分野の手法を直接活動分類に応用することは妥当ではないと考えられる。言い換えると、個々の移動・活動単位でのセグメントは、

*キーワード: 活動分類, マルチレベル分析, 活動開始時刻選択

**学生会員, 修(工), 広島大学大学院国際協力研究科
(東広島市鏡山1丁目5番1号, Tel&Fax: 082-424-5971,
E-mail: chikaraishi@hiroshima-u.ac.jp)

***正会員, 博(工), 広島大学大学院国際協力研究科
(東広島市鏡山1丁目5番1号, Tel&Fax: 082-424-6919,
E-mail: zjy@hiroshima-u.ac.jp)

****正会員, 博(工), 広島大学大学院国際協力研究科
(東広島市鏡山1丁目5番1号, Tel&Fax: 082-424-6921,
E-mail: afujiw@hiroshima-u.ac.jp)

行動が持つ変動特性に大きく影響されてしまうことが予想される．そこで本研究では，個人が遂行する一つの活動が「活動種類」と階層的な関係にあることを利用し，個人の上位レベルに存在する「活動種類」の影響を，ランダム変数を用いて扱うマルチレベル分析手法を応用することを考える．具体的には，1) 活動種類の影響をランダム変数として扱うマルチレベルモデルを構築・推定し，2) 推定されたパラメータを用いて，各活動に対する縮約推定量⁸⁾を算出し，3) 縮約推定値に対してクラスター分析を実行し，活動分類を決定する，という流れで活動種類の分類方法を提案する．

1) モデルの定式化

時刻の選択を表現するにあたって，生存時間モデルや離散選択モデルが頻繁に用いられるが，本研究では線形モデルを用いて活動開始時刻選択を表現する．線形モデルを採用する理由は，パラメータと説明変数の積が時刻の単位を持つため，クラスター分析における活動間の距離の定義が明確となるためである．ここで，個人*i*が活動*a* ($=1, \dots, A$)を行う時の活動開始時刻選択を以下の式で定式化する．

$$\mathbf{y}_a = \mathbf{x}_a (\boldsymbol{\beta} + \boldsymbol{\gamma}_a) + \boldsymbol{\varepsilon}_a \quad (1)$$

ただし，

$$\mathbf{y}_a = \begin{bmatrix} y_{a1} \\ y_{a2} \\ \vdots \\ y_{ai} \\ \vdots \\ y_{an_a} \end{bmatrix}, \quad \mathbf{x}_a = \begin{bmatrix} x_{a11} & x_{a12} & \cdots & x_{a1K} \\ x_{a21} & x_{a22} & \cdots & x_{a2K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{ai1} & x_{ai2} & \cdots & x_{aiK} \\ \vdots & \vdots & \ddots & \vdots \\ x_{an_a1} & x_{an_a2} & \cdots & x_{an_aK} \end{bmatrix}, \quad \boldsymbol{\varepsilon}_a = \begin{bmatrix} \varepsilon_{a1} \\ \varepsilon_{a2} \\ \vdots \\ \varepsilon_{ai} \\ \vdots \\ \varepsilon_{an_a} \end{bmatrix},$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}, \quad \boldsymbol{\gamma}_a = \begin{bmatrix} \gamma_{a1} \\ \gamma_{a2} \\ \vdots \\ \gamma_{aK} \end{bmatrix}$$

ここで， y_{ai} は個人*i*が活動*a*を開始する時刻， x_{aik} は活動開始時刻を規定する*k* ($=1, \dots, K$)番目の説明変数であり，それぞれ観測される値である．また， n_a は活動種類*a*を遂行したサンプル数である． β_k は活動種類間で共通の未知パラメータである． ε_{ai} は $\varepsilon_{ai} \sim N(0, \sigma_0^2)$ に従う誤差項である． γ_{ak} は*k*番目の説明変数の影響が活動種類によって異なることを表現するランダム変数であり， $\gamma_{ak} \sim N(0, \sigma_k^2)$ に従うとする．また， γ_{ak} と $\gamma_{ak'}$ の間に相関はないと仮定する ($k \neq k'$)．モデルを同定するための尤度関数は，以下の式で定義される⁹⁾．

$$L(\boldsymbol{\beta}, \boldsymbol{\sigma} | \mathbf{y}) = \int \prod_i [p(\mathbf{y} | \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\sigma}_0)] \Phi(\boldsymbol{\gamma} | \boldsymbol{\Omega}_k) d\boldsymbol{\gamma} \quad (2)$$

ここで， $\Phi(\cdot)$ は平均0，分散共分散行列 $\boldsymbol{\Omega}_k$ に従う*K*次元正規分布の密度関数である．ただし， $\boldsymbol{\Omega}_k$ は対角行列

であり，対角成分は $\sigma_1^2, \sigma_2^2, \dots, \sigma_K^2$ である． $p(\cdot)$ は平均 $\boldsymbol{\beta}\mathbf{x} + \boldsymbol{\gamma}$ ，分散 σ_0^2 に従う正規分布の密度関数である．

2) 縮約推定量(shrinkage estimator⁷⁾)の算出

マルチレベルモデルでは，得られたパラメータセットをもとに， γ_{ak} の値を事後的に求めることが頻繁になされる．本研究では，その中の一つである縮約推定量 $\hat{\gamma}_{ak}$ を求める．縮約推定量とは， γ_{ak} の信頼性（活動ごとのサンプル数，説明変数のばらつきに依存）が低い場合，0に近づくような特性を持つ推定量である．この特性によって，不確定な変動要因の影響により時刻の選好を記述することが困難な活動種類を同一のセグメントに分類することができる．以下にその算出式を示す⁸⁾．

$$\hat{\boldsymbol{\gamma}}_a = \hat{\boldsymbol{\Omega}}_k \mathbf{x}_a^T (\hat{\sigma}_0^2 \mathbf{I}_a + \mathbf{x}_a \hat{\boldsymbol{\Omega}}_k \mathbf{x}_a^T)^{-1} (\mathbf{y}_a - \mathbf{x}_a \hat{\boldsymbol{\beta}}) \quad (4)$$

3) クラスター分析

式(4)から得られた $\hat{\gamma}_{ak}$ (活動数×パラメータ数)を用いて階層的クラスター分析を行う．クラスター（活動分類）はウォード法を用いて計算する．なお，活動間の距離（行動特性の非類似度） $D_{aa'}$ は，以下の式で計算する．

$$D_{aa'} = \sqrt{\sum_{k=1}^K (\hat{\gamma}_{ak} \bar{x}_{aik} - \hat{\gamma}_{a'k} \bar{x}_{a'ik})^2} \quad (5)$$

ここで， \bar{x}_{aik} 値は，*k*番目の説明変数の活動種類ごとの x_{aik} の平均値である．これより式(5)は，2つの活動種類間の，目的変数に対する説明変数の寄与構造の違いを表す．なお，活動分類数は分析者が事前に与える．

以上，本研究で提案する活動分類手法の手順を述べた．提案手法の特徴として，1) 目的変数に対する各種説明変数の反応（行動特性）に基づいて活動分類を行う点，2) 自由記述式アンケート等，活動種類が膨大な数にのぼる場合においても，比較的システマティックに活動種類を分類できる点（分析者に依存するのは，式(1)の説明変数の選択と，活動分類数の選択の2点）が挙げられる．

3. 実証分析に使用するデータ

実証分析にあたって，61種類の活動分類がなされている社会生活基本調査（調査票B：総務省が平成13年10月13日～10月21日に実施）を用いる．なお実証分析においては，平成13年10月15（月）～10月18日（木）のデータのみピックアップした．また，3:00～24:00までの行動データのみを選別し，さらに，各活動に対して該当日，該時間帯にはじめて行った活動開始時刻のみを分析に用いることとする．言い換えると，同じ日に同じ活動を2回以上行った場合，2回目以降の活動開始時刻はサンプルに含まない．以上の条件のもとで抽出されたサンプル

数は、44,728であった。

4. 分析結果

4.1 モデルの推定結果と考察

活動開始時刻選択モデル(式(1))の推定結果を表1に示す。モデルの推定は、61種類の活動を合わせたデータを用いているため、パラメータ β の推定値は、活動に依存せず、全体として活動開始時刻が前後にシフトすることを意味する。また、パラメータ σ_a の値は、当該説明変数が活動開始時刻選択に与える影響の活動間の差の程度を示している。

パラメータ β の推定値を見ると、高齢者の生活のリズムは、その他の個人と比べて70分程度早い時間帯にシフトしていることが分かる。また、既婚者は37分、世帯収入が200万円未満の個人は15分程度、活動開始時刻が早い時間帯にシフトしていることが分かる。しかしながら、これらの説明変数に対応するパラメータ σ_a の値は大きく、活動種類によってその出発時刻は大きく変動する。一方で、天気ダミーに対応する σ_a の値は比較的小さく、雨の日は、活動全体を通して平均的に10分程度早い時間帯に活動開始時刻がシフトすることが分かる。

次に、ランダム変数 y_{lak} の分散 σ_a^2 の値について考察する。なお、ランダム変数を考慮する必要があるかどうかは、当該ランダム変数を除いた場合の最終尤度と、含んだ場合の最終尤度との間で χ^2 検定を行うことで評価している⁹⁾。まず、天気ダミー、自動車保有ダミーにかかる σ_a の値が有意でないことから、雨かどうか、自動車を保有しているかどうかは活動開始時刻選択の活動間のばらつきに影響を及ぼさないと考えられる。すなわち、どのような活動分類を行ってもこれらの変数が活動開始時刻選択に与える影響は一定であると見なすことができる。一方、その他の説明変数にかかる σ_a の値はすべて統計的に有意であり、これらの説明変数の影響は活動種類によって異なると考えられる。特に、定数項に対応する σ_a の値が大きく推定されていることから、平均活動開始時刻は活動種類によって大きく異なることが分かる。続いて年齢、性別、既婚ダミーにかかる σ_a の値が大きく、これらの変数に対応するパラメータの値は、活動分類の仕方によって大きく変化すると考えられる。

4.2 活動分類結果と考察

式(5)に基づき算出した非類似度を用いてクラスター分析を行った結果を表2に示す。例として、活動分類数が6のときの各クラスターの活動開始時刻分布を図1に示す。以下、活動分類数が6の場合について、考察を行う。

表2、図1より、6A、6Bは、早い時間帯に行われる活動であり、主に必須活動で構成されていることがわかる。

両者の違いは、6Aが決められた時間帯に繰り返し行われる活動種類であるのに対し、6Bは必須活動ではあるものの、時刻の選好が個人によって異なる活動であると考えられる。ただし6Aと6Bは、クラスター数(活動種類数)を5に変更した場合に統合される活動分類であり、両者の行動特性は類似していると予想される。6C、6D、6Eに分類される活動種類は、自由裁量型の活動が比較的多く、同じ自由裁量型の活動であっても活動種類によって活動開始時刻の選好は大きく異なることが分かる。特に、活動分類数2の場合においても、6Cに含まれる活動種類と、6D、6Eに含まれる活動種類は異なる活動分類にセグメントされることから、自由裁量型の活動でも活動種類によって時間的な選好は大きく異なることが明らか

表1. モデルの推定結果

説明変数	パラメータ	t値
既婚ダミー(既婚: 1, その他: 0)	β	-36.89 -3.94 **
	σ_a^2	3021 [55] (294.5) **
天気ダミー(雨: 1, その他: 0)	β	-10.17 -3.56 **
	σ_a^2	26 [5.1] (0.2)
自動車保有ダミー(1台以上保有: 1, その他: 0)	β	-2.59 -0.78
	σ_a^2	0 [0] (0.0)
性別ダミー(男性: 1, 女性: 0)	β	4.60 0.40
	σ_a^2	6008 [77.5] (462.0) **
世帯年間収入ダミー1(1000万以上: 1, その他: 0)	β	13.50 2.63 **
	σ_a^2	316 [17.8] (5.2) *
世帯年間収入ダミー2(200万未満: 1, その他: 0)	β	-14.62 -2.21 *
	σ_a^2	704 [26.5] (19.9) **
年齢ダミー1(18歳以下: 1, その他: 0)	β	45.57 2.03 *
	σ_a^2	17076 [130.7] (369.0) **
年齢ダミー2(65歳以上: 1, その他: 0)	β	-73.07 -6.92 **
	σ_a^2	3932 [62.7] (416.2) **
10歳以上の世帯構成員数(人)	β	2.80 1.42
	σ_a^2	97 [9.8] (25.1) **
10歳未満の世帯構成員数(人)	β	-1.38 -0.36
	σ_a^2	331 [18.2] (39.0) **
定数項	β	846.06 37.07 **
	σ_a^2	27802 [166.7] (2713.7) **
残差	σ_0^2	58482 [241.8]
LL(C)		-324728
LL(β, σ)		-309348
サンプル数		44,728

()内は χ^2 統計量, []内は標準偏差 σ_a **1%有意, *5%有意

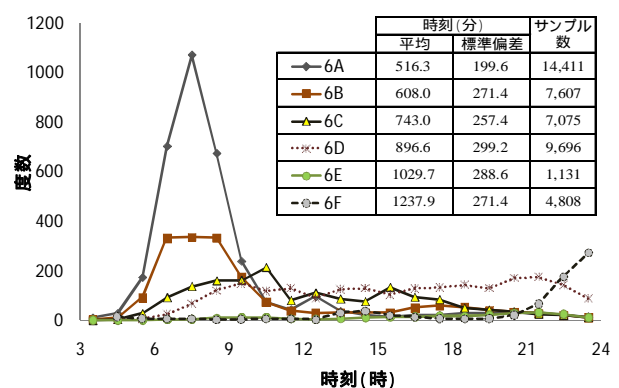


図1. クラスターごとの活動開始時刻分布(活動分類数6)

かとなった。この結果は、安易に「自由活動」等の分類を行うことによって、本来現象が持っている行動特性の情報を失ってしまう恐れがあることを示している。最後に、6Fは睡眠と社会生活基本調査に関連する行動で構成されるセグメントである。これは、就寝前に調査票を記入する個人が多いことに由来している。

5. おわりに

本研究では、本来現象が持っている行動特性の情報を失ってしまう恐れがあるとの認識のもと、行動の類似性を反映した新たな活動分類手法を提案した。具体的には、1) 活動種類の影響をランダム変数として扱うマルチレベルモデルを構築・推定し、2) 推定されたパラメータを用いて、各活動に対する縮約推定量⁸⁾を算出し、3) 縮約推定値に対してクラスター分析を実行する、という流れで活動種類の分類を行った。細分化された活動種類を持つ平成13年度の社会生活基本調査データを用いた実証分析の結果、特に自由裁量型活動の分類を安易に行ってしまうと、本来現象が持っている行動特性の情報を失ってしまう可能性が高いことを明らかにした。

活動分類の方法は、トリップチェーンや活動間の相互作用等を分析の対象とする場合だけではなく、他の多くの場面においてその重要性は高まっていると考えられる。例えば、活動種類別に時間価値が異なることをどの程度考慮する必要があるのかという問題¹⁰⁾についても、どのような活動分類を採用するかによって、設定される時間価値は大きく異なるものと考えられる。今後、活動分類の違いによってもたらされるモデルからのインプリケーションの違いを定量的に評価していく必要がある。

参考文献

- 1) Doherty, S. T.: Should we abandon activity type analysis? Redefining activities by their salient attributes, *Transportation*, 33, pp. 517-536, 2006.
- 2) Hägerstrand, T.: What About People in Regional Science?, *Papers and Proceedings of the Regional Science Association*, Vol. 24, pp. 7-21, 1970.
- 3) Shelley, H. J.: Developing the American Time Use Survey Activity Classification System, *Monthly Labor Review*, Vol. 128, No. 6, pp. 3-15, 2005.
- 4) Wedel, M., Kamakura, W. A.: *Market Segmentation: Conceptual and Methodological Foundations*, Kluwer Academic Publishers, 1998.
- 5) 櫻井尚子: 潜在クラス分析を用いたマーケット・セグメンテーション 顧客の購買パターンによるマーケットセグメント創出のモデル, *計算機統計学*, 第17巻, 第1号, pp. 21-30, 2004.
- 6) Scarpa, R., Giudice, T. D.: Market Segmentation via Mixed Logit: Extra-Virgin Olive Oil in Urban Italy, *Journal of Agricultural & Food Industrial Organization*, Vol. 2, pp. 1-18, 2004.
- 7) 力石真, 藤原章正, 張峻屹, K.W. Axhausen: 6週間の交通行動データを用いた活動パターンの変動に関するマルチレベル分析, *土木計画学研究・講演集*, Vol.37, 2008 (CD-ROM).

- 8) Harville, D.: Extension of the Gauss-Markov Theorem to Include the Estimation of Random Effects, *The Annals of Statistics*, Vol. 2, No. 2, pp. 384-395, 1976.
- 9) Goldstein, H.: *Multilevel Statistical Models*, 3rd Edition, London: Arnold, 2003.
- 10) 上田孝行: 道路事業の評価に関する研究課題と今後の展望, *交通工学*, Vol. 43, No. 1, pp. 6-13, 2008.

表2. クラスター分析による活動分類結果

活動種類数	活動種類数	活動種類数	活動種類数	活動種類数	活動種類数	活動種類数	社会生活基本調査の活動分類
2	3	4	5	6	7		
				6A	7A		身の回りの用事(自分自身や家族等が行うもの) 食事 通勤・通学 主な仕事 礼拝・読経
	3A	4A	5A	6B	7B		主な仕事中の移動 副業中の移動 学校での授業・その他学校での行動 食事の管理 住まいの手入れ・整理 衣類等の手入れ その他の家事 乳幼児の身体の世話と監督 その他の趣味・娯楽 療養
	2A				7C		家庭教師による勉強・学習塾・予備校での勉強等
						2A	身の回りの用事(個人サービスの利用) 乗り物の手入れ 公的サービスの利用 商業的サービスの利用 家族の身の回りの世話 乳幼児の介護・看護 乳幼児以外の家族の介護・看護 乳幼児の付き添い等
		3B	4B	5B	6C	7D	移動 新聞・雑誌 軽飲食 仕事中・学校での学習(学業)中の休憩 教養・娯楽 ドライブ エアロビクス系スポーツ ウォーター系スポーツ ボランティア活動 冠婚葬祭 受診 求職活動
				5C	6D	7E	副業 建築・修繕 乳幼児と遊ぶ 買物 テレビ・ラジオ 家族とのコミュニケーション その他の教養・くつろぎ 学習・研究(学業以外)
		2B	3C			4C	創作 ゲーム 読書 CD・カセットテープ・ビデオ 球技 その他のスポーツ 社会参加活動 人と会って行う交際・付き合い 電子メールや手紙による交際・付き合い
				5D	6E	7F	学校の宿題 世帯管理 子供の教育 うたたね 趣味 電話による交際・付き合い
			4D	5E	6F	7G	睡眠 社会生活基本調査に関連する行動