

修正辞書編纂型意思決定モデルの再現性に関する分析*

A study on identifiability of models for lexicographic decision making strategy *

大橋聡子**・山本俊行***・倉内慎也****・森川高行*****

By Satoko OHASHI**・Toshiyuki YAMAMOTO***・Shinya KURAUCHI****・Takayuki MORIKAWA*****

1. はじめに

交通行動分析における離散選択モデルでは、「補償型」の意思決定ルールを仮定したモデルが広く適用されている。補償型意思決定ルールとは、ある属性（例えば所要時間）の劣位を他の属性（例えば費用）の優位によって埋め合わせ（補償）することが可能なことから呼ばれており、線形効用関数を用いたモデルはその代表例である。補償型モデルは、その操作性の高さなどを理由に支配的に適用されているが、意思決定に多大な認知資源を要するものであり¹⁾、人間の情報収集・処理能力を過大に評価していることが指摘されている²⁾。また、実世界の意思決定は必ずしも補償型ではなく、特に選択肢や属性の数が多い選択状況では、認知資源を節約するために、属性間で補償の利かない「非補償型」意思決定ルールのような、より簡略化されたルールが支配的に用いられることが報告されている³⁾。しかし、操作性の高い補償型モデルに対し、非補償型モデルは高度に非線形であるため推定が非常に困難であり、非補償型意思決定ルールを再現するためには、選択結果以外に意思決定過程に関する情報を援用してモデル化を行う必要があると考えられる⁴⁾。

山本⁵⁾は、交通手段選択問題にデータマイニング手法を適用して生成された決定木が、まさに非補償

型意思決定ルールを表現しているとしているが、その再現性については評価されていない。仮に決定木が非補償型意思決定ルールを再現しているのであれば、これを意思決定過程を表す情報として非補償型意思決定ルールのモデル化に用いることができる。

ただし、通常の分析で対象とする実データでは、意思決定者の真の意思決定構造が不明であり、再現性を評価することは困難である。そこで本研究では、意思決定者が非補償型意思決定ルールに従っていると仮定し、人工的に作成した選択データを分析に用いることにより、データマイニング手法の非補償型意思決定ルールの再現性を評価し、それを援用した非補償型意思決定モデルの推定可能性、再現性について検証する。

2. 修正辞書編纂型意思決定ルール

本研究では非補償型意思決定ルールとして修正辞書編纂型を用いる。意思決定者が最も重要とする属性のみに着目して各選択肢を評価し、そこで選択肢が1つに定まらない場合は、次に重要な属性に着目して残った選択肢を評価する。これを選択肢が1つに定まるまで、重要度がより下位の属性について繰り返すのが辞書編纂型の意思決定ルールである。また、属性値が必ずしも最良でなくとも、最良の属性値との差が閾値を超えなければ同等とみなし、次の属性の評価をする意思決定ルールを修正辞書編纂型という。例えば、費用を最も重要な属性としている意思決定者が、費用の最も安いバスと次に安い鉄道との差が10円しかなければ、そこでバスは選択せず、次に重要な属性である所要時間等を評価する、という意思決定ルールである。

修正辞書編纂型は、選択肢や属性の数が増加した場合でも認知負荷はほとんど変化せず、最適な選択に対する正確性もある程度保持していることが明

*キーワード：交通手段選択，交通行動分析，データマイニング

**正員，修(工)，株式会社 長大
(名古屋市千種区名駅南 1-18-24，
TEL:052-586-0734，E-mail:oohashi-s@chodai.co.jp)

***正員，博(工)，名古屋大学大学院工学研究科
(名古屋市千種区不老町，TEL:052-789-4636，
E-mail:yamamoto@civil.nagoya-u.ac.jp)

****正員，修(工)，名古屋大学大学院工学研究科
(名古屋市千種区不老町，TEL:052-789-3565，
E-mail:kurauchi@civil.nagoya-u.ac.jp)

*****正員，Ph.D.，名古屋大学大学院環境学研究科
(名古屋市千種区不老町，TEL:052-789-3564，
E-mail:morikawa@civil.nagoya-u.ac.jp)

らかにされており¹⁾、実際的意思決定において頻繁に用いられることが実証されている³⁾。

3. シミュレーションを用いた推定特性と再現性の比較

(1) 分析に用いるモデル

a) 修正辞書編纂型モデル⁶⁾ (LEX)

前章で述べた修正辞書編纂型意思決定ルールを仮定した非補償型モデルである。ただし、全ての属性を評価した後に複数の選択肢が残った場合、意思決定者は残った選択肢の中から1つをランダムに選択すると仮定する。例として、選択肢が*i, j*の2種類、属性が*a, b*の2種類の選択状況を想定すると、属性の重要度の順が{*a, b*}である場合の選択肢*i*の選択確率 $P(i|a, b)$ は以下のように定式化される。

$$P(i|a, b) = Q(a) + (1 - Q(a)) \times Q(b) + (1 - Q(a)) \times (1 - Q(b)) \times \frac{1}{2} \quad (1)$$

ここに、 $Q(a) : X_{ia} - X_{ja} > \tau_a$ である確率、 X_{ia} : 選択肢*i*に関する属性*a*の値、 τ_a : 属性*a*に関する閾値。閾値 τ_a に適切な分布形を仮定し、全サンプルにおける平均値と標準偏差を未知パラメータとする。

b) データマイニング手法 (DM)

データマイニング手法とは大規模なデータを、与えられた変数からいくつかのセグメントに分割することによって、そのデータの持つ何らかのパターンを発見する手法である。本研究ではデータマイニング手法のモデルの1つとしてC4.5⁷⁾を用いる。

(2) 分析方法

本研究では、仮想的な意思決定者の選択行動が修正辞書編纂型意思決定ルールに従うものと仮定し、そのルールに従って意思決定者の行動をシミュレートすることにより、人工的に選択データを作成する。これにより、通常分析ではわからない意思決定者の真の意思決定構造が既知である状態で分析を行うことができる。具体的には、データマイニング手法の再現性(属性の順位や閾値)、正しいモデル(修正辞書編纂型モデル)の推定可能性、データマイニング手法の修正辞書編纂型モデルへ援用可能性、各モデルの適合度や予測精度などを比較・検

証することが可能となる。

a) 想定する選択状況

選択肢を自動車、バス、鉄道の3肢、属性を所要時間、費用、端末時間の3属性とし、サンプル数を1,000とした交通手段選択問題とする。

b) 属性の望ましさの相関

ある選択状況下で属性の望ましさに正の相関があるとき、例えば、所要時間と費用のどちらについても自動車より鉄道のほうが優れている場合、辞書編纂型意思決定ルールで評価する属性の順序が変わっても選択結果は変わらない。一方、所要時間は鉄道のほうが優れているが費用は自動車のほうが優れているなど、属性の望ましさに負の相関がある場合、評価する属性の順序が変わることにより、選択結果が反転することがある。そこで本研究では、属性の望ましさに正の相関を持つ意思決定者の割合が大きい場合と負の相関を持つ意思決定者の割合が大きい場合の2種類のデータについて分析を行う。

c) 選択結果のシミュレーション

本研究で仮定する修正辞書編纂型意思決定ルールでは、属性の重要度の順位には個人間の異質性を考慮せず、全ての意思決定者が1.所要時間、2.費用、3.端末時間の順で属性を評価して選択するものとする。また、各属性の閾値には正規分布、または対数正規分布を仮定し、その平均値を所要時間：5分、費用：20円、端末時間：1分とし、標準偏差を所要時間：2.5分、費用：10円、端末時間：0.5分とした乱数を生成する。生成された乱数を閾値として修正辞書編纂型に従った選択行動のシミュレーションを行い、各意思決定者について選択結果のデータを得る。乱数の値によっては異なる選択結果が得られるため、同一条件に対して乱数生成を10回を行い、各々のデータに対してモデル推定を行う。

4. 分析結果

(1) データマイニング手法の再現性

真の意思決定構造が前章で述べた修正辞書編纂型であるデータについてデータマイニング手法を適用した場合、その推定結果は以下のような点で修正辞書編纂型意思決定ルールを再現していた。

・属性間に負の相関がある場合には、決定木上の

評価属性順位が正しく推定される。

- ・ 閾値に正規分布を仮定した場合，決定木上の閾値は真の閾値の平均値に近い値が推定される。

これは，属性の望ましさに正の相関がある場合，属性の評価順位が変わっても選択結果が同じであるため，属性順位を正確に再現することが困難となり，また，データマイニング手法は閾値の最頻値を推定する傾向にあり，正規分布のように平均値と最頻値が一致するような分布形に対しては良好な再現性が得られるものと考えられる。ただし，属性順位，閾値が真の構造を再現しているのは，決定木の上位のみであり，下位の構造は誤差部分までも過度に的中させるために，意味を持たない構造となっている。

(2) 修正辞書編纂型モデルの推定可能性

修正辞書編纂型モデルは，選択結果のシミュレーションと同じ意思決定ルールを仮定した，正しいモデルである。ただし，属性の重要度の順位については，分析者には正しい順位がわからないものとし，全ての属性順位のパターンをそれぞれ仮定したモデルを推定した。その結果，誤った属性順位を仮定したモデルは収束することがほとんどなく，正しい属性順位を仮定したモデルでも，10回の推定のうち収束した回数は，属性間に正の相関：4回，負の相関：1回，であり，頻度こそ高くなるものの完全に収束することはなかった。つまり，属性順位に異質性がない場合でも収束可能性から属性順位を特定することは困難である。修正辞書編纂型モデルでは，選択結果がどの属性評価段階で決定され，残りの選択肢がどの段階で棄却されるかの全事象の出現確率の周辺和を尤度としており，選択肢や属性が増えるにつれ尤度関数の複雑さは急激に上昇し，初期値の与え方によってはパラメータが収束しない，または局所最適に陥ることがあることが確認された。

表1 修正辞書編纂型モデルの収束回数

初期値	援用なし	DM 援用	真値援用
収束回数	1	2	7

(3) データマイニング手法の援用

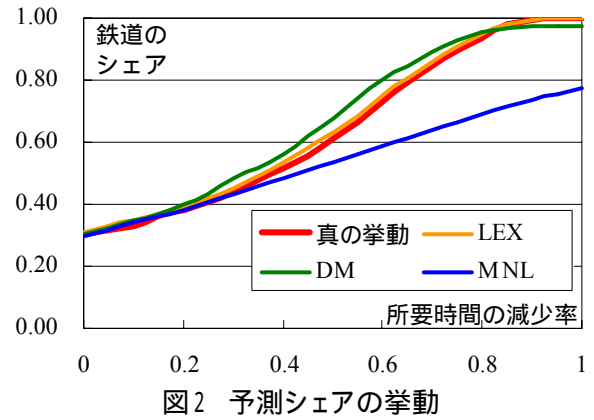
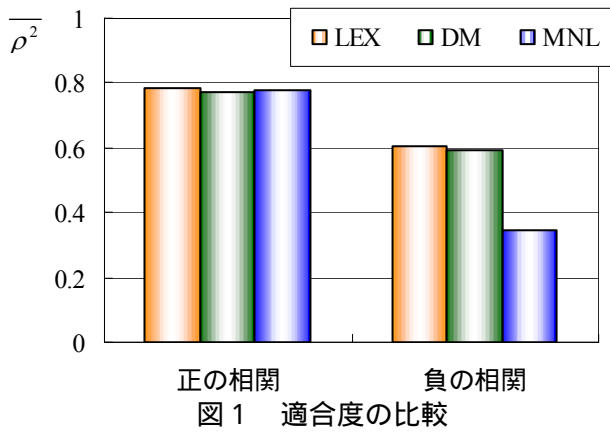
前節で述べた修正辞書編纂型モデルの問題点を解消する手段として，(1)節の決定木を援用する。具体的には，決定木で表される属性の評価順位から，修正辞書編纂型モデルの属性順位を特定し，決定木で表される各属性の閾値をパラメータ推定時の初期値として援用することができる。

表1は属性間に負の相関のある場合に，修正辞書編纂型モデルを10回推定したうちの収束回数を示しており，初期値を0とした場合（援用なし），データマイニング手法を援用した場合，さらに閾値の分布に関して事前情報がある場合の極端な例として真値を初期値として用いた場合，の3パターンを試みた。データマイニング手法の援用では，収束回数はほとんど変わらないのに対し，真値を援用した場合には収束回数が増加している。つまり，収束可能性を向上させるためには，データマイニング手法から得られる閾値の平均値に関する情報だけでは不十分であり，標準偏差に関する情報も必要であると考えられる。

表2は初期値の3パターンについて，収束する，しないに関係なく10回の推定パラメータ値の平均値と標準偏差を示している。評価順位の高い属性（所要時間，費用，端末時間の順）の閾値の平均値に関するパラメータは，初期値の援用がなくとも推定値は真値をほぼ再現しているといえるが，評価順位の低い属性の閾値や，各属性の閾値の標準偏差については，初期値に真値を援用しても再現されない。ただし，初期値に真値を援用した場合，端末時間に関するパラメータのばらつきが小さくなっていることから，初期値の援用がない場合に収束回数が少ないのは，端末時間のパラメータの変化が尤度関数に

表2 パラメータ推定値（10回の平均値，()内は標準偏差）

	真値	援用なし	DM援用	真値援用
所要時間の閾値の平均	5.00	4.95 (0.04)	4.95 (0.04)	4.96 (0.04)
ln(所要時間の閾値の標準偏差)	0.92	-0.38 (0.11)	-0.38 (0.11)	-0.38 (0.11)
費用の閾値の平均	20.00	19.29 (0.52)	19.39 (0.52)	19.47 (0.51)
ln(費用の閾値の標準偏差)	2.30	0.34 (0.27)	0.33 (0.28)	0.33 (0.28)
端末時間の閾値の平均	1.00	12.64 (8.77)	34.96 (47.30)	27.22 (8.13)
ln(端末時間の閾値の標準偏差)	-0.69	8.27 (6.67)	2.72 (6.67)	0.22 (1.35)



及ぼす影響が極めて小さく、結果として非常に不安定になっていることが原因と考えられる。

(4) 適合度と予測精度の比較

前節までの分析で用いた修正辞書編纂型モデル (LEX), データマイニング手法(DM)に加えて、交通行動分析で最も頻繁に適用されている多項ロジットモデル(MNL)についても同じデータに対して推定を行い、真の構造が非補償型である場合に対する、各モデルの適合度や予測精度の比較を行う。

図1は属性の望ましさに正の相関がある場合と負の相関がある場合のそれぞれについて、各モデルの適合度を ρ^2 値で示したものである。正の相関がある場合、どんな意思決定ルールを用いても選択結果はほぼ一致するため、誤った意思決定ルールを仮定したモデルでも高い適合度を得ることができる。

一方、図2は属性の望ましさに正の相関がある場合について、鉄道の所要時間を0~100%減少させたときの鉄道のシェアの挙動を示している。前章で述べた選択結果のシミュレーションと同様の方法で求めた真のシェアの挙動と比較すると、多項ロジットモデルは線形効用関数を仮定しているため、属性値の減少率に対して直線的にシェアを伸ばしているのに対し、真の挙動に表される修正辞書編纂型の意思決定ルールでは選択基準に閾値を設けているため、属性値がある特定の減少率に達するまではシェアの増加率は小さく、その減少率を超えた時点でシェアの増加率が大きくなっている。図2の修正辞書編纂型モデルやデータマイニング手法はその傾向を再現できているといえる。

本研究より、属性の望ましさに負の相関がある場合、データマイニング手法から推定される決定木は修正辞書編纂型意思決定ルールの評価属性順位を再現しており、また決定木の上位部分であれば閾値も再現できることが明らかとなった。しかし、これらの情報を援用しても修正辞書編纂型モデルにおいて不安定なパラメータを再現することはないため、修正辞書編纂型モデルを収束可能とする手段を他の視点から検討する必要がある。

これまで支配的に適用されてきた補償型モデルは、特に属性の望ましさに正の相関がある場合には、誤った意思決定ルールを仮定しても再現性に遜色はないが、これを将来予測に適用すると予測シェアに大きな誤差が生じる。このため、操作性が高いという理由だけで補償型モデルを適用することは非常に危険であると言える。

参考文献

- 1) Payne, J. W., Bettman, J. R. and Johnson, E. J. : *The Adaptive Decision Maker*, Cambridge University Press, 1993.
- 2) Simon, H. A. : A Study of Decision Making Process in Administrative Organization. *Administrative Behavior*, 2nd ed. Macmillan, 1957.
- 3) 例えば, Wright, P. : Consumer choice strategies: Simplifying vs. optimizing. *Journal of Marketing Research*, Vol.12, pp.60-67, 1975.
- 4) 倉内慎也, 三古展弘, 森川高行, 大橋聡子: 意思決定ルールの仮説が個人行動モデルに及ぼす影響の実証的分析, 土木計画学研究・講演集, No.25, (CD-ROM), 2002.
- 5) 山本俊行: 非補償型意思決定ルールを表現するためのデータマイニング手法の適用に関する分析, 土木学会論文集(印刷中), 2004.
- 6) Kurauchi, S. and Morikawa, T. : An Exploratory Analysis for Discrete Choice Model with Latent Classes Considering Heterogeneity of Decision Making Rules, In Hensher, D.A. (ed.) *Travel Behaviour Research: The Leading Edge*, Pergamon, pp.409-423, 2001.
- 7) Quinlan, J. R. : C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, 1993