

[特 集]

非補償型意思決定方略を表現するための データマイニング手法の適用に関する分析

山本俊行¹¹正会員 博(工) 名古屋大学大学院助教授 工学研究科社会基盤工学専攻 (〒464-8603 愛知県名古屋市中千種区不老町)

本研究では、非補償型の意思決定方略を表現するためのモデルとして、代表的なデータマイニング手法の1つであるC4.5を適用し、実際のデータに対する選択行動の再現性を通常のロジットモデルと比較することで、データマイニング手法の有効性を検討した。動的パークアンドライドシステムの利用に関するSP調査データを用いた分析結果より、C4.5の的中率はロジットモデルと同程度であった。また、C4.5によって生成されたif-thenルールをダミー変数としたロジットモデルを推定し尤度を算出したところ、通常のロジットモデルより劣る結果となった。さらに、if-thenルールを通常のロジットモデルの説明変数に追加することで統合モデルを構築したところ、選択行動の再現性が向上することを確認した。

Key Words: C4.5, data mining, mode choice, non-compensatory model, random utility model, stated preference data

1. はじめに

(1) 交通行動における意思決定方略

ランダム効用モデルはマイクロ経済理論に基礎を置くモデルであり、これまで多くの交通行動分析に適用されてきた。しかしながら、ランダム効用モデルで用いられる仮定は、行動科学の分野から度々批判されてきた¹⁾。線形効用関数はほとんどの離散選択モデルで適用されてきたが、補償型の意思決定方略を仮定したものである。補償型の意思決定方略では、ある属性の劣位を他の属性の優位で埋め合わせることが可能であることを仮定しているが、意思決定方略は他にも数多くあり、局面によっては属性間の補償のきかない「非補償型」、例えば、属性を一つずつ取り上げて逐次的に絞り込みを行う「逐次選択型」などの意思決定方略が支配的になる場合も考えられる²⁾³⁾。

非補償型の意思決定方略が支配的であると考えられる場合には、想定される非補償型の意思決定方略を表現するモデルの構築を考えるのが自然であると思われるが、これに対しては反論もある。Dawes and Corrigan⁴⁾、Dawes⁵⁾は、たとえ意思決定方略が非補償型であったとしても、線形効用関数はそれらの意思決定方略による選択結果を十分再現できると主張している。一方で、Johnson *et al.*⁶⁾は属性間に負の相関がある場合には線形効用関数による再現力は低くなると主張している。交通行動分析が対象とする選択行動には、高速道路と一般道路の間の選択における所要時間と料金のように、属性間に負の相関を持

つ場合が多く、非補償型の意思決定方略を表現するモデルの構築によって、選択行動の再現力が向上する可能性があると思われる。交通行動分析の分野では、以上のような認識に基づき、非補償型の意思決定方略を表現するための研究が数多く行われている。また、非補償型の意思決定方略を内包した1日の活動パターン生成シミュレータも構築されている⁷⁾⁸⁾。

(2) 非補償型の意思決定方略のモデル化

非補償型の意思決定方略をランダム効用モデルで表現するための方法としては、線形効用関数の代わりに一般化平均概念を用いた分析が行われている⁹⁾。一般化平均を効用関数として用いた場合、通常の補償型の意思決定方略の他、最も優れた値や最も劣った値を持つ属性の値によって決定するような非補償型の意思決定方略についても表現することが可能である。しかしながら、逐次選択型等の方略については十分な表現力を持たない。

より柔軟なモデルとして、ニューラルネットワークの適用が進められてきた。我が国では、秋山¹⁰⁾¹¹⁾、島崎・安田¹²⁾等の研究の蓄積がある。海外での事例についてはDougherty¹³⁾を参照されたい。ニューラルネットワークでは、属性が効用に及ぼす非線形な影響を表現することが可能であり、属性間の交互作用についても表現可能である。ニューラルネットワークは、非線形性や交互作用を組み合わせることによって多くの非補償型の意思決定方略を表現できる可能性を持つ。しかしながら、推定されたモデルは数多くのパラメータを持つため、モデルによ

って表現されている意思決定方略を分析者が解釈するのが困難である場合が多い。ニューラルネットワークモデルから得られたパラメータと選択行動との関係については解明されておらず、今後の研究に期待したい。

一般化平均およびニューラルネットワークは非補償型の意思決定方略を自由度の高い関数形を用いることによって表現しようとするのに対して、より直接的に非補償型の意思決定方略を表現しようとする研究も行われている。これらの研究は、ランダム効用モデルを拡張する形で非補償型の意思決定方略をモデル化したもの^{14), 15)}と、データマイニング手法¹⁶⁾等を用いて、ランダム効用モデルとは独立に非補償型の意思決定方略をモデル化したもの^{17), 18), 19), 20), 21), 22), 23)}に大別される。ランダム効用モデルを拡張したKurauchi and Morikawa¹⁴⁾、森川・倉内¹⁵⁾は、通常のランダム効用モデルに対する選択行動の再現力の優位性を確認しているが、モデルが高度に非線形であるため推定作業が困難であるという指摘もされている。これに対してデータマイニング手法は推定作業の困難性は低く、サンプル数が莫大になっても計算時間が現実的な範囲に収まるという利点をもつ。一方で、ランダム効用理論とは独立に非補償型の意思決定方略をモデル化した研究は、ランダム効用モデルとの比較が不十分である。Wets *et al.*²¹⁾は、前述した活動パターン生成シミュレータに組み込むことを目的として、C4.5²⁴⁾、CHAID²⁵⁾、多項ロジットモデルの的中率の比較を行っており、モデル間での顕著な差は見られないとしている。一方で、Yamamoto *et al.*^{22), 23)}では、C4.5の方が2項ロジットモデルより高い的中率を持つという結果を得ている。

(3) 本研究の目的

本研究では、データマイニングで用いられるモデルの一つであるC4.5を適用し、通常のランダム効用モデル、及び、ランダム効用モデルを拡張したKurauchi and Morikawa¹⁴⁾のモデルとの比較分析を行うことによって、非補償型の意思決定方略の表現のためのデータマイニング手法の有効性に関する知見を得ることを目的とする。Kurauchi and Morikawaは、修正辞書編さん型意思決定方略のモデル化を試みているが、パラメータの推定困難性のため、各個人が最も重要視する変数のみについて取り上げ、その変数について非補償型の意思決定を行うことを仮定している。本研究では、より一般的なモデルの構築が困難であることから、Kurauchi and Morikawaで構築されたモデルをデータマイニング手法の比較対象とする。よって、本研究の分析に用いるデータ及び変数はKurauchi and Morikawaで用いられているものと同一としている。

データマイニング手法には、分類、回帰、クラスタリング、従属性のモデル化、変化や逸脱の検出、決定木、

相関ルールの発見、ニューラルネットワーク等いくつかの手法が存在する¹⁶⁾。このうち、最も注目され、使用頻度が高いのが決定木であり、決定木をコード化したプログラムとしては、前述のC4.5、CHAIDの他、CART²⁶⁾がある。このうち、C4.5は、他のプログラムにない特徴として構築した決定木から新たにif-thenルールを生成するという機能を持つ。本研究では、このような特徴及び交通行動分析での適用実績からC4.5を用いた分析を行う。

C4.5では決定木とif-thenルールが生成されるが、いずれも非補償型の意思決定方略を表現するものとして捉えることが出来る。ただし、生成された決定木やif-thenルールは実際に用いられた意思決定方略そのものを表すとは限らない。このような実際の意思決定方略とモデル構造の乖離は、ロジットモデル等の補償型効用関数が実際の意思決定方略をそのまま反映しているとは限らず、あくまでデータから推定される関数に過ぎないのと同様である。しかしながら、C4.5ではモデルの構造がロジットモデル等の効用関数と比べて柔軟であるため、選択行動をより正確に再現することが期待される。

本研究では、さらに、生成されたルールを通常の離散選択モデルの説明変数に追加することによって、非補償型の意思決定方略的な要素を組み込んだ補償型意思決定モデルの選択行動の再現性の向上を目指す。

(4) 検証項目

本研究では、手法の有効性を表す指標のひとつとしてデータの再現性についての分析を中心に行う。C4.5での出力結果と従来のロジットモデル、及びランダム効用を拡張した非補償型意思決定モデルの的中率、及び、尤度を比べることにより、選択行動の再現性の比較を行う。ここで、的中率はC4.5等のデータマイニング手法で主に用いられるモデルの精度を表す指標であり、尤度はランダム効用モデルで主に用いられる指標である。本研究では、より中立な比較を行うために両者の指標を用いる。

これらの指標のうち、的中率には、確率的情報を0-1の離散的な情報に置き換えていることによる情報の損失があるという問題点がある。よって、的中率は、確率分布を仮定しているランダム効用モデルの精度を十分表現できない。一方で、データマイニング手法はモデルの推定に最尤推定法を用いないため、通常は尤度を算出しない。そのため、本研究では、Yamamoto *et al.*^{22), 23)}と同様に、尤度を算出するために、C4.5で生成されたif-thenルールを説明変数としたロジットモデルを構築し、通常のロジットモデル、及び、非補償型意思決定モデルと尤度の比較を行う。

また、手法の有効性を表す指標として、再現性の他に、モデルの安定性、及び、得られた結果の解釈可能性についても検討を行う。モデルの安定性については、通常の

ランダム効用モデルと同様に、パラメータの値について検証する。ただし、C4.5は直接パラメータの値を算出しないため、上述のif-thenルールを説明変数としたロジットモデルの推定結果を用いる。また、誤差項の分布の仮定に対するモデルの安定性について検証する。ここでは、C4.5の推定結果と通常のロジットモデルの推定結果をmaximum score推定量²⁷⁾によって再推定し、パラメータの安定性の比較を行う。クロスバリデーション等による他のサンプルに対するモデルの安定性の検証は今後の課題とする。また、解釈可能性については分析者の主観によるところが大きい。C4.5の推定結果が非補償型の意思決定方略として解釈できるかについて検討する。

2. 決定木構築手法 : C4.5

(1) 概要

C4.5では、はじめに、サンプルを再帰的にセグメントに分割することで、最終的に同一セグメントに属するケースの全てが同じ選択肢を選択するようにセグメント分割を行う。この時、1回の分割は1つの属性に基づいて行われる。こうして生成された決定木は、しばしばサンプルの選択行動を過剰に反映し、モデルの構造が複雑となり、サンプル内の誤差変動までモデルで表現するため予測モデルとして有効ではなくなる。そこで、一旦生成された決定木に対し、一定の基準に従っていくつかのセグメントを統合することで単純化し、最終的な決定木を得る。決定木では、この単純化を枝刈りと呼ぶ。

if-then ルールの生成に当たっては、まず、枝刈り前の決定木で分割された各セグメントを規定する属性の組み合わせを条件 (if) , 当該セグメントの全ケースが選択した選択肢を選択結果 (then) としてルール集合を生成する。このようにして生成されたルール集合は、最初に生成された決定木と同様に、予測モデルとしての有効性が低い可能性がある。よって、一定の基準に従ってルール集合を単純化し、最終的な if-then ルールを得る。

(2) 決定木の生成

上述したように、サンプルは再帰的に分割される。各分割時には、情報理論に基づく利得比 (gain ratio) と呼ばれる指標が最も大きくなるような説明変数が分割の基準として用いられる。説明変数がカテゴリ変数の場合には、カテゴリの全ての組み合わせ (例えば、変数値がA, B, Cの3つの場合には、{A, B}と{C}, {B, C}と{A}, {C, A}と{B}, {A}と{B}と{C}の4通り) のうち最も利得比が大きい組み合わせを用いて他の説明変数との比較を行う。一方、説明変数が序列表変数や連続変数の場合には、変数値によってケースを並び替え、変数値の大小関係に

よって2分した時に最も利得比が大きくなる閾値による2分割を用いて他の説明変数との比較を行う。よって、序列表変数や連続変数では必ず2分割されるのに対して、カテゴリ変数の場合には3つ以上に分割される場合もある。

利得比の考え方の基礎となっている情報理論的な考え方は、メッセージによって伝えられる情報量はメッセージの生起確率で決まる、というものである。すなわち、情報量は、生起確率の値の2を底とする対数を取り、-1倍したビット単位で表される。セグメント、 S からランダムに1ケースを選び、そのケースの選択結果が C_j であると伝えたとする、このメッセージの情報量は以下の式で表される。

$$-\log_2 \left(\frac{\text{freq}(C_j, S)}{|S|} \right) \quad (1)$$

ただし、 $\text{freq}(C_j, S)$ はセグメント、 S に含まれるケースのうち、選択肢、 C_j を選択したケース数を表し、 $|S|$ はセグメント、 S に含まれるケース数を表す。このようなメッセージによってセグメント、 S に含まれる1ケースの選択結果を伝える場合の情報量の期待値、 $\text{info}(S)$ は以下の式で与えられる。

$$\text{info}(S) = -\sum_{j=1}^k \left\{ \frac{\text{freq}(C_j, S)}{|S|} \times \log_2 \left(\frac{\text{freq}(C_j, S)}{|S|} \right) \right\} \quad (2)$$

ただし、 k は選択肢数を表す。 $\text{info}(S)$ はエントロピーとして知られている。

ここで、一つの説明変数が分割の基準として選ばれ、セグメント S が (S_1, S_2, \dots, S_n) の n 個のセグメントに再分割されたとしよう。この分割を X とすると、分割後に同様のメッセージを伝える場合の情報量の期待値、 $\text{info}_X(S)$ は以下の式で表される。

$$\text{info}_X(S) = \sum_{i=1}^n \left\{ \frac{|S_i|}{|S|} \times \text{info}(S_i) \right\}. \quad (3)$$

よって、分割、 X によって得られた利得、 $\text{gain}(X)$ は以下の式で与えられる。

$$\text{gain}(X) = \text{info}(S) - \text{info}_X(S) \quad (4)$$

利得は分割数を多くするほど大きくなるという性質もっているため、例えば全ケースが別のセグメントに分割

されれば利得が最大となるが、そのような分割には意味はない。よって、分割自体を伝えるのに必要なメッセージの情報量で基準化する。分割情報量, $split\ info(X)$ は以下の式で与えられる。

$$split\ info(X) = - \sum_{i=1}^n \left\{ \frac{|S_i|}{|S|} \times \log_2 \left(\frac{|S_i|}{|S|} \right) \right\} \quad (5)$$

利得比, $gain\ ratio(X)$ は、利得を分割情報量で基準化した以下の式で与えられる。

$$gain\ ratio(X) = \frac{gain(X)}{split\ info(X)} \quad (6)$$

式(6)は、利得比が分割によって得られる情報量のうち、選択枝の特定に寄与する部分の割合を示しており、利得比は分割の効率性を表すものである。しかしながら、セグメントに属するケースのうち非常に少数のものとそれ以外のものに分割するような小さな分割は分割情報量が小さくなり、利得比の値が不安定になる。よって、説明変数の選択の際には、全ての変数のうち、利得が少なくとも平均以上であるという条件の下で利得比が最大となる説明変数が選ばれる。

(3) 枝刈り

枝刈りは、期待誤差に基づいて行われる。(2)で生成された決定木は期待誤差が減少しなくなるまで再帰的にセグメントが統合される。誤差は、セグメントに属するケースのうち、分割時に当該セグメントに割り当てられた選択枝とは異なる選択枝を持つケースの数として定義される。しかしながら、サンプルに含まれる誤差は統合によって常に増加するため枝刈りに用いることが出来ない。C4.5では、サンプルに含まれる誤差はある真の誤差確率にしたがって生成された標本であるとの認識に基づき、サンプルの出現確率が一定の信頼区間に含まれるという条件のもとで誤差確率の上限値を用いて期待誤差を計算する。誤差確率の上限値は以下の式で表される。

$$U_{CF}(E, |S|) = \max \{ p |_{\beta} C_E \cdot p^E \cdot (1-p)^{|S|-E} \geq CF \} \quad (7)$$

ただし、 $U_{CF}(E, |S|)$ はケース数が $|S|$ のセグメントに E ケースの誤差が存在する場合の誤差確率の上限値を表し、 $|_{\beta} C_E$ は $|S|$ から E を取り出す場合の組み合わせの数、 CF は信頼度を表す。

期待誤差は、誤差確率の上限値、 $U_{CF}(E, |S|)$ とセグメントに含まれるケース数の積として定義される。ある分割が統合される条件は以下の式で表される。

$$|S| \times U_{CF}(E, |S|) < \sum_{j=1}^k \left\{ |S_j| \times U_{CF}(E_j, |S_j|) \right\} \quad (8)$$

ただし、 E_j はセグメント S_j に含まれる誤差を表す。

(4) ルール集合の生成

ルール集合の生成にあたっては、はじめに、枝刈り前の決定木で分割された各セグメントを規定する属性の組み合わせを条件、当該セグメントの全ケースが選択した選択枝を選択結果とするルール集合に変換する。枝刈り前の決定木は全てのケースをいずれか1つのセグメントに分割しており、かつ、各セグメントに含まれるケースは全て同じ選択枝を持つ。よって、変換されたルール集合についても、各ケースは必ずただ1つのルールの条件を満たすことに注意されたい。

各ルールは条件に含まれる属性のうち、いくつかの属性を取り除くことによって一般化される。属性の削除は決定木の枝刈りと同様に誤差確率に基づいて行われる。あるルールの条件に X という属性が含まれており、 X を削除することでルールを一般化する時、残った属性で構成される条件によって規定されるセグメント、 S_X は一般化する前のセグメント、 S を含む。ここで、属性 X は、削除した場合に誤差確率が減少する場合に削除される。削除される条件は以下の式で与えられる。

$$U_{CF}(E_X, |S_X|) < U_{CF}(E, |S|) \quad (9)$$

ここで、 E_X はセグメント S_X の誤差数を表す。誤差確率が最小となるまで1つずつ属性が削除され条件が一般化される。

このようにして一般化された条件は、もはや背反ではなく、1つのケースが2つ以上のルールの条件を満たす場合もある。よって、ルールは順番に並べられ、最初に条件を満たすルールが適用される。ルールの順番は以下のように決定される。まず、同じ選択枝を選択結果とするルール同士がまとめられる。こうしてまとめられたグループのなかでは、どのルールが適用されても選択結果が同じであるため順序を気にする必要がない。次に、まとめられたルール群のなかで、最小記述長 (minimum description length) 原理²⁸⁾とよばれる基準によって最も効率がよくなるように不要なルールが削除される。ここで、最小記述長原理は情報理論に基づくものであり、ルールの単純さと例外の少なさのバランスを取るものである。一般的に、ルールが単純であれば例外が多くなる可能性が高く、例外を伝達するための情報量は増える。一方で、ルールを複雑にすれば例外は減るもののルール自体を伝達する情報量は増える。ここでは、ルールを記述するの

に必要な情報量と、ルールに合わないケース（誤差）を記述するのに必要な情報量をコーディングに必要なビット数で計算し、その和が最も少なくなるようなルールの組み合わせが選択される。ただし、総当たりでルールの組み合わせを検討するのは現実的ではないため、焼き鈍し法が用いられる。

各ルール群に含まれるルールを最小記述長原理で選択した後、誤差が最小となるようにルール群同士の順番が決定される。最後に、いずれのルールの条件にも当てはまらないケースのために、サンプルの中でいずれのルールにも当てはまらないケースのうちで最も多くのケースが選択した選択肢をデフォルトの選択肢として設定する。

3. データ

本研究の事例分析で用いるデータは 1997 年に名古屋市北東部、およびそれに近接する郊外地域で収集された動的パークアンドライド（DP&R）システムの利用に関する SP 調査データである。調査では、情報提供時の仮想的な状況として、都心部へ向かって自動車を運転している途中で DP&R に関する交通情報板を見た時に、DP&R を利用して地下鉄に乗り換えるかそのまま自動車で都心部へ向かうかという選択を尋ねている。また、トリップ目的による選択行動の違いを考慮するため、通勤・通学目的と買い物目的の 2 つの目的でのトリップについてそれぞれ質問している。

考慮されている要因は、都心部まで自動車を利用した場合の所要時間と費用（都心の駐車場料金）、都心まで DP&R を利用して地下鉄を利用した場合の所要時間と費用（DP&R 駐車場の料金と地下鉄往復運賃）、DP&R 駅での乗り換えのしやすさ、都心駐車場満空情報である。ただし、通勤・通学目的トリップについては、目的地での駐車場は確保されているという仮定により、都心の駐車場料金及び都心駐車場満空情報は考慮されていない。本研究では、通勤・通学を行っていない被験者は買い物目的トリップに関する項目のみ回答しているため買い物目的トリップに関するデータの方がサンプル数が多いこと、及び、買い物目的トリップに関する設定のみに都心の駐車場料金と都心駐車場満空情報が提供されているため買い物目的トリップの設定の方が変数が多く、より非補償型の意思決定方略が用いられやすい^{29),30)}こと、の 2 つの理由により、買い物目的トリップに関するデータを用いた分析を行う。

本研究で用いる買い物目的の場合の仮想的な設定条件は表-1 に示す通りである。要因の組み合わせは実験計画法により決定されており、 $L_{18}(2^1 \times 3^3)$ 型直交表により割り付けられている。被験者の負担を減らすため、上記の

表-1 SP 調査の設定条件

設定	所要時間(分)		費用(円)		乗り換え 易さ*	都心 駐車場
	自動車	DP&R	自動車	DP&R		
A-1	40	20	500	500	遠い	満車
A-2	40	30	500	1500	近い	空き少
A-3	40	10	500	900	近い	空車
A-4	30	20	500	900	近い	空き少
A-5	30	10	500	1500	近い	満車
A-6	30	10	500	500	遠い	空き少
A-7	30	20	500	1500	近い	空車
B-1	40	30	500	1500	遠い	満車
B-2	40	20	500	900	近い	空車
B-3	40	10	500	500	近い	空車
B-4	30	20	500	1500	遠い	空き少
B-5	30	10	500	900	遠い	満車
B-6	30	10	500	1500	遠い	空車
B-7	30	20	500	500	近い	満車

*DP&R 駐車場と乗り換え駅の距離

組み合わせのうち、個々の被験者に A-1~7 あるいは B-1~7 の 7 つの設定条件のそれぞれについて自動車を選擇するか DP&R を選擇するかを尋ねている。回答には繰り返し観測に伴う系列相関が含まれることが危惧されるものの、本研究ではそれに対する修正は行っていない。

本研究で用いるデータは自動車を運転中に行われる意思決定を対象としたものである。一般に、意思決定に費やすことの出来る時間が少ないほど非補償型の意思決定方略が用いられやすいことが確認されており³¹⁾、自動車を運転中に行われる意思決定は時間的な制約を多く受けると考えられるため、非補償型の意思決定方略の適用に対する妥当性は高いと考えられる。しかしながら、本研究で用いるデータは SP データであり、実際にはこのような時間的な制約を受けない。一方で、今回用いる SP データは同一の被験者から 7 つの設定条件のもとでの回答を求めており、多くの意思決定を行う必要がある場合には、事件的な制約が大きい場合と同様に非補償型の意思決定方略が用いられやすいとの知見³¹⁾もある。以上より、本研究で用いるデータは確実に非補償型の意思決定方略を観測したものとは言えないものの、非補償型の意思決定方略が用いられやすい状況で得られたデータであると言える。

調査は訪問配布・回収により実施され、604世帯、1102個人からのデータが得られている。世帯回収率は90.4%と高く、母集団代表性の高いデータが収集できたものと考えられる。調査の詳細については中村ら³²⁾、平田³³⁾を参照されたい。本研究では、1102人から得られたデータのうち、個人属性等に不備のない4778ケースの選択データを用いた分析を行う。分析に用いた説明変数をまとめたものを表-2に示す。自動車保有台数に関しては、自動車依存率の高い中京都市圏の特徴から保有自動車が3台以上の世帯についてダミー変数を設定している。変数の

表-2 説明変数の定義

変数	定義
費用	DP&R-自動車 (1000 円)
所要時間	DP&R-自動車 (時間)
乗換便利	1, DP&R 駐車場と乗り換え駅が近い; 0, それ以外
空き少し	1, 都心駐車場の情報が「空き少し」の時; 0, それ以外
満車	1, 都心駐車場の情報が「満車」の時; 0, それ以外
子供有り	1, 子供のいる世帯の場合; 0, それ以外
男性	1, 男性; 0, それ以外
若者	1, 30 歳未満; 0, それ以外
老人	1, 60 歳以上; 0, それ以外
学生	1, 学生; 0, それ以外
高収入	1, 年収 1,000 万円以上; 0, それ以外
保有車多	1, 自動車保有台数 3 台以上; 0, それ以外

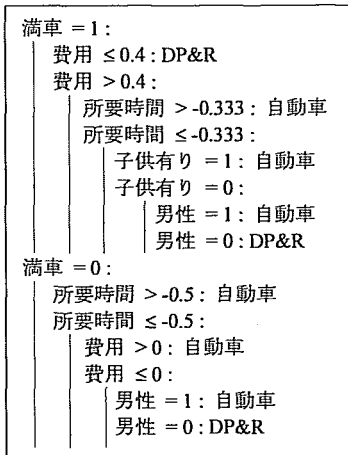


図-1 生成された決定木

設定は、比較可能性を確保するため Kurauchi and Morikawa と同一のものを用いている。

4. 推定結果

(1) 決定木

はじめに、得られた決定木を図-1に示す。図では、コロン (:) の左側がセグメントを規定する属性を表し、コロン (:) の右側は予測される選択肢を表す。コロン (:) の右側が空白の場合には、そのセグメントはさらに分割されることを意味し、その分割はすぐ下の行に示されている。図では、最初に満車が否かでセグメントが分割され、満車の場合には、次に費用が 0.4 以下か否かによってセグメントが分割されることを示しており、もし費用が 0.4 以下の場合には DP&R が選択され、そうでない場合には、次に所要時間によってさらにセグメントが分割されることを示している。

各セグメントはいくつかの条件式によって規定されるが、条件式が全ての変数を含んでいる場合には、そのセグメントは全ての変数を考慮していることを意味し、補

償型の意思決定方略を取っていることを意味する。一方、少数の変数のみを含む場合には、それら少数の変数のみを考慮した非補償型の意思決定方略を取っていることを示唆するものと捉えることが可能である。

図-1 より、最初に都心部の駐車場の混雑状況により意思決定構造が異なることが示されている。都心部の駐車場が満車の場合には、次に費用、所要時間の順で考慮されている。一方、都心部の駐車場が満車でない場合には、所要時間、費用の順で意思決定が行われている。また、都心部の駐車場が満車の場合には、費用差の閾値は DP&R が自動車より 400 円以上高いか否か、所要時間の閾値は DP&R が自動車より 20 分以上早いかなどとなっているのに対し、都心部の駐車場が満車でない場合には、それぞれ 0 円、30 分となっている。これらはいずれも DP&R に対して厳しい閾値となっており、都心部の駐車場の状況を見て選択基準を変更していることを示すものと考えられる。特に、都心部の駐車場が満車の場合には、DP&R の費用が自動車より 400 円以上高くなければ所要時間に関わらず DP&R を選択し、都心部の駐車場が満車でない場合には、DP&R の所要時間が自動車より 30 分以上早くないと費用に関わらず自動車を選択するという非補償型の意思決定が行われていることを示唆する結果が得られている。

図では、いくつかの社会経済属性がセグメントを規定する属性として推定されている。これらの変数は選択肢の状況を表すものではないが、意思決定構造が個人によって異なることを表現するものである。今回の推定結果では、性別と子供の有無が最終的なモデルに含まれている。一方で、年齢や職業、収入、自動車保有台数等の変数は選択に及ぼす影響が統計的に有意な値を示さなかったため、最終的なモデルに含まれていない。モデルの推定結果より、性別は 2 つの部分でセグメントを分割する属性として用いられているが、いずれの場合も、男性はその上のセグメント分割で他方のセグメントに割り当てられた選択肢と同じ選択肢を持つことが示されている。満車 = 0, 所要時間 ≤ -0.5 の時を見ると、男性は費用に関係なく自動車を選択することを示している一方で、女性は、そのような場合に次に費用を考慮することを示している。この結果は、男性の方がより単純な意思決定構造を持っていることを示すものである。すなわち、非補償型の意思決定方略を取っている傾向が強いと考えられる。ただし、満車 = 1, 費用 > 0.4 の時を見ると、次に所要時間を考慮するのは子供のいない女性のみであるが、男性がより単純な意思決定構造を持っていることに変わりはない。

(2) if-then ルール

次に、生成されたルールを表-3に示す。表ではルール

表3 生成されたルール

ルール	条件	選択肢
1	子供有り = 0, 男性 = 0, 満車 = 1, 費用 ≤ 0.4, 所要時間 ≤ -0.333	DP&R
2	男性 = 0, 費用 ≤ 0, 所要時間 ≤ -0.5	DP&R
3	費用 ≤ 0.4, 満車 = 1	DP&R
4	費用 > 0.4, 所要時間 > -0.333	自動車
5	子供有り = 1, 費用 > 0.4	自動車
6	所要時間 > -0.5, 満車 = 0	自動車
	それ以外	自動車

表4 各ルールの的中率

ルール	ケース数	的中率
1	105	0.667
2	170	0.647
3	1011	0.548
4	1357	0.819
5	297	0.741
6	1225	0.739
それ以外	613	0.558
合計	4778	0.693

は優先順位に従って並べられており、ルール1から順に条件に当てはまるルールの選択肢が当該ケースの選択肢として採用される。いずれの条件にも当てはまらない場合はデフォルトの選択肢として自動車がそのケースの選択肢として採用される。

推定結果より、if-then ルールも決定木と同様の意思決定構造を示している。ただし、ルール数は6つであり、決定木では9つのセグメントに分割されているのに比べて、ルールが単純化（一般化）されていることが分かる。個々のルールを見ると、ルール1および2は女性のみ適用されることが分かる。すなわち、男性がDP&Rを選択するのはルール3の条件が満たされる時に限られる。また、DP&Rを選択するルールのうち、ルール1とルール3は都心部の駐車場が満車の場合のみ適用されることが分かる。よって、都心部の駐車場が満車でない場合には、DP&Rが選択されるのは、ルール2の条件が満たされる場合に限られることが分かる。この結果からも、DP&Rの選択には都心部の駐車場の混雑状況が大きな影響を与えていることが分かる。

(3) 再現性の検証

if-then ルールについて各ルールに当てはまるケース数、および、それらのケースのうち、実際の選択結果がルールで示された選択結果と一致しているか否かを表す的中率を計算した結果を表4に示す。各ケースは最初に当てはまるルールが適用されるため、各ケース数の合計はサンプル数に一致する。決定木についても各セグメントに属するケース数とそのセグメントでの的中率の算出を行ったが、合計的中率は全く同一であり、個々のケースが的中したか否かも全てのケースについて一致したため、より単純化されているif-then ルールの結果のみを示す。

表より、ルール1のケース数は105と最も少なくなっていることが分かる。これはルール1の条件が5つもの変数の積からなることによるものであると考えられる。しかしながら、同様にDP&Rを選択肢に持つルールであるルール2やルール3よりも高い的中率を持っているために、有意なルールとして最終的なif-then ルールに含まれていることが分かる。また、DP&Rを選択肢に持つルール1からルール3に比べて自動車を選択肢に持つルール4からルール6の方が的中率が高くなっている。これは、自動車を選択する場合の条件は表2に含まれる説明変数でよく表現されているのに対して、DP&Rを選択する条件はそれほど明確に表現されていないことを示すものである。この原因として、DP&Rの選択に今回のモデル化で考慮したような一般的な変数の他に、非観測異質性のようなものがより大きく影響していることが考えられる。もう1つの可能性としては、自動車を選択する、言い換えれば、DP&Rを選択しないという意味決定はいくつかの条件式によって表現が容易であるのに対して、DP&Rを選択するという意思決定は、そのような条件式では表現が困難であるということが考えられる。後者が原因である場合、DP&Rの利用意向を分析するモデルとしてのC4.5の限界を表しており、今後の適用に際して注意が必要である。

5. ランダム効用モデル及びランダム効用モデルを拡張した非補償型意思決定モデルとの比較

(1) ロジットモデルの推定

4.で構築した決定木、if-then ルールと通常のロジットモデルの選択行動の再現性を比較するために、表2の説明変数を用いてロジットモデルを推定し、5%の有意水準で統計的に有意でない変数を順に取り除く変数減少法によって最終的な推定結果を得た。ランダム効用モデルを拡張した非補償型意思決定モデルについては、同一データを対象とした推定がKurauchi and Morikawaで示されており、本研究での比較にはその結果を用いる。Kurauchi and Morikawaでも通常のロジットモデルが推定されているが、統計的に有意でない変数も導入されているため今回の結果とは異なっていることに注意されたい。通常のロジットモデルの推定結果を表5の左列に示す。表では、推定値が正の場合、変数値が増加するほどDP&Rが選択される確率が増加することを示す。

推定結果より、通常のロジットモデルでは、決定木やif-then ルールより多くの個人属性が選択に影響を及ぼす変数として導入されていることが分かる。ただし、このような個人属性は、効用関数の定数項にのみ影響を及ぼすものであり、費用や所要時間といった選択肢属性が選

表-5 ロジットモデルの推定結果

説明変数	通常のモデル		if-then ルール		統合モデル	
	推定値	t値	推定値	t値	推定値	t値
定数項	-1.828	-8.49			0.267	2.30
費用	-0.816	-9.20			-1.022	-6.82
所要時間	-3.373	-8.46				
乗換便利	0.382	5.17				
空き少し	0.304	2.58				
満車	1.378	13.75			1.192	10.64
子供有り	-0.281	-4.11			-0.227	-3.12
男性	-0.342	-5.18			-0.324	-4.77
老人	0.198	2.27			0.205	2.33
ルール1			0.693	3.35		
ルール2			0.606	3.78	0.398	2.05
ルール3			0.192	3.05	-0.905	-6.13
ルール4			-1.508	-21.40	-0.938	-7.19
ルール5			-1.050	-7.93	-0.624	-3.37
ルール6			-1.040	-15.99	-0.757	-7.19
デフォルト			-0.233	-2.86		
サンプル数	4778		4778		4778	
L(0)	-3312		-3312		-3312	
L(C)	-3078		-3078		-3078	
L(β)	-2749		-2810		-2737	
χ^2	1125		1004		1150	
AIC	5517		5634		5496	
的中率	0.691		0.693		0.690	

括に及ぼす影響の個人間異質性を示すものではない。これは、変数の導入の際に、各変数の主効果のみを考慮し、変数間の交互作用を考慮していないことによる当然の結果である。

変数間の交互作用を考慮するためには、予め複数の変数の積を説明変数に導入する必要がある。しかしながら、McClelland and Judd²⁴⁾は、説明変数としてモデルに導入した複数の要因が行動に及ぼす影響に交互作用がある場合にも、モデルの推定結果に基づいて経験的に判断するといった方法でこの交互作用を発見することは困難であることを示している。実際、if-then ルールのルール1の条件では5つの変数の積を取っているため、各変数の直接効果に加えて、2つの変数の積から順に5つの変数の積まで考慮すると、12の変数があるため、全部で、 $1_2C_1 + 1_2C_2 + \dots + 1_2C_5 = 1585$ の説明変数を考慮する必要がある。3つの変数の積まで限定したとしても、その数は298となり現実的ではない。もちろん、全ての変数組を一度に推定せずに、いくつかの変数組をランダムに取り上げ推定するという作業を繰り返すことによって、有意な変数組を探索することは可能である。ただし、ランダム効用モデルは分析者の立てた仮説を検証するという形でモデルの推定を行うのが一般的であり、繰り返しモデルを推定することで有意な変数組の探索を行うという方法はデータマイニング手法に属する考え方である。すなわち、このような交互作用が実際の意思決定に存在する場合には、C4.5等のデータマイニング手法が有効であると考え

られる。

(2) 的中率の比較

通常のロジットモデルの個々の説明変数のパラメータ推定値は予想通りの符号を持っており、決定木や if-then ルールと共通に用いられている変数についても同一の傾向を示しているものと解釈できる。例えば、子供有りダミー、男性ダミーの推定値はいずれも負の値を示しているが、いずれの変数も決定木では、子供有り = 1 の場合には自動車を選択する、男性 = 1 の場合には自動車を選択する、という分割に用いられている。これらの結果は、通常のロジットモデルと決定木、if-then ルールのいずれのモデルであっても、ある程度の意思決定方略の再現が可能であることを示唆するものと考えられる。的中率については、C4.5が0.693に対して、通常のロジットモデルの的中率は0.691となっており、C4.5と通常のロジットモデルでは的中率に大きな違いはないことを示している。また、Kurauchi and Morikawa では的中率を示していないため直接の比較は出来ないが、通常のロジットモデルより高い的中率を持つことが予想されるため、C4.5よりも高い的中率を持つ可能性が高い。

(3) 尤度の比較

ロジットモデルをはじめ、ランダム効用モデルではモデルの適合度を表すのに尤度が用いられるのが一般的であるが、通常、決定木や if-then ルールは尤度を算出せず、そのままでは尤度の比較が出来ない。そこで、if-then ルールに含まれる各ルールに当てはまるか否かを示すダミー変数を作成し、それらのダミー変数のみを用いてロジットモデルを推定した。表-4の集計と同様に、各ケースは最初に条件に当てはまったルールのダミー変数のみ1を取り、他のルールについては0を取る。また、デフォルトはどのルールの条件にも当てはまらないケースが1、いずれかのルールに当てはまったケースは0をとるダミー変数を表す。よって、全てのケースが必ずただ1つのダミー変数の値として1を取り、線形従属関係が成り立つため定数項は含まれていない。推定結果を表-5の中列に示す。

推定結果より、全てのダミー変数が統計的に有意であり、推定値の符号もif-thenルールでの各ルールの選択肢と整合している。また、各推定値の絶対値は各ルールの的中率と比例している。最終尤度、L(β)は-2810であり、通常のロジットモデルよりも劣る結果となった。また、Kurauchi and Morikawaのモデルは通常のロジットモデルを特殊ケースとして内包するため通常のロジットモデルよりも尤度が高く、if-thenルールは3つのモデルの中で最も尤度が低い結果となった。モデル間の比較に際し、片方のモデルが他方のモデルを特殊ケースとして含むよう

な関係にない場合、 χ^2 検定を用いることは不適切であるため³⁵⁾、本研究では以下の式で計算されるAIC(赤池の情報量基準)³⁶⁾を用いる。

$$AIC = -2 * \{L(\beta) - K\} \quad (10)$$

ここで、Kは推定されたパラメータ数を表す。AICは値が小さいモデルほどデータへの適合度が高いことを示す指標であり、AICで見てif-thenルールによるロジットモデルがKurauchi and Morikawaのモデルや通常のロジットモデルより適合度が劣る結果となった。

(4) 安定性の比較

前節で述べたように、if-thenルールによるロジットモデルの推定結果より、ルールを表すダミー変数の全てが統計的に有意な値をとっており、C4.5の安定性を確認することが出来る。一方、通常のロジットモデルについても、推定の際に有意水準を用いて変数選択を行っているため、全ての変数が統計的に有意な値をとっており安定性をもつモデルと考えられる。

前節において、if-thenルールと通常のロジットモデルの比較において、的中率では大きな差は見られないのに対してAICでは差が見られたことは、両モデルが非補償型意思決定方略と補償型意思決定方略を表現しているという違いよりも、パラメータの推定方法の違いを反映したものと考えられる。すなわち、if-thenルールでは各々のケースの実際の選択肢とモデルの選択結果が一致するようにパラメータが推定されるのに対して、通常のロジットモデルでは各々のケースの実際の選択肢をモデルが選択する確率を最大化するようにパラメータが推定されている。ここで、ランダム効用理論で確率(尤度)が最大化されるのは効用関数の誤差項に特定の分布、ロジットモデルではガンベル分布を仮定しているためである。効用関数の誤差項に特定の分布を仮定しないパラメータの推定方法としてはmaximum score推定量がある。本研究では、通常のロジットモデルと同様に線形効用関数を仮定しつつ的中率の最大化を図るため、maximum score推定を試みた。推定には統計ソフト、LIMDEP³⁷⁾を用いた。maximum score推定量そのものはパラメータの標準誤差の推定値を算出しないため、bootstrap法³⁸⁾によって推定したものをを用いたが、多くの説明変数が統計的に有意でなく、「費用」と「乗換便利」の2つの説明変数のみが有意となった。それら2つのみを導入したモデルの的中率も0.675に留まり、誤差項にガンベル分布を仮定した通常のロジットモデルよりも的中率が低くなった。一方で、if-thenルールによるモデルをmaximum score推定量を用いて推定したところ、全てのダミー変数が統計的に有意となり、的中率も誤差項にガンベル分布を仮定した場合

と変わらなかった。これらの結果は、C4.5の方が線形効用関数を仮定したモデルよりも誤差項の分布の仮定に対して頑健であることを示している。

(5) 統合モデルの再現性

最後に、非補償型意思決定方略と補償型意思決定方略のいずれも表現できるモデルとして、通常のロジットモデルの説明変数とif-thenルールを表すダミー変数を統合してロジットモデルの推定を行った。ただし、定数項を用いる関係で、if-thenルールを表すダミー変数のうち、デフォルトダミーについてはあらかじめ説明変数から除いた。通常のロジットモデルを推定した時と同様に変数減少法を用いた結果を表-4の右列に示す。通常のロジットモデルで用いられていた説明変数のうちいくつかの変数が統計的に有意でなくなり最終的なモデルからは除かれている。所要時間についてはif-thenルールでもルールを構成する属性として用いられており、従来より選択に大きな影響を及ぼす要因として考えられている。しかしながら、推定結果からは、いつでも所要時間が影響を及ぼすというわけではなく、影響を及ぼすのは、ルール2, 4, 6, のように、他の属性による条件が成り立つ時に限られるということが分かる。一方で、費用については統計的に有意であり、いつでも選択に影響を及ぼすことが確認された。これは、if-thenルールでも6つのルールのうち、ルール6を除く5つのルールで費用が条件に含まれていることとも整合的である。

一方、if-thenルールを表すダミー変数については、ルール1以外は統計的に有意となった。ルール1については、もともとルールにあてはまるケース数が最も少なかったこともあり、他の説明変数と相関を持つことによってパラメータ推定値の標準偏差が大きくなったものと考えられる。あるいは、条件に含まれる属性が5つと他のルールよりも多く、実際は補償型の意思決定方略が取られているのにも関わらず、無理やりif-thenルールで非補償型意思決定方略としてサンプルを再現しようとしていたのが原因かも知れない。後者の場合には、統合モデルの構築によって非補償型意思決定方略と補償型意思決定方略の両者を表現することが可能となり、そのようなモデルの過誤が修正されたと捉えることが出来る。

統計的に有意となったルールのうち、ほとんどのダミー変数の符号はif-thenルールのみを用いた場合と変化しなかったものの、ルール3については正負が逆転している。これは、ルール3の条件に含まれる属性である「費用」と「満車」のいずれも単独の説明変数としてもモデルに含まれており、ルール3とそれらの変数間の相関が高くなったことによることが考えられる。本研究で用いたデータはSP調査によって得られたものであり、直交表により要因が割り付けられているため、「費用」や「満空情報」

といった変数間には相関がない。しかしながら、ルール3の条件式はそれらの変数を含んでいるため、変数間の相関が生じてしまう。ただし、ルール5についても「子供有り」と「費用」はいずれも単独の説明変数としてモデルに含まれているが、符号条件は変化していない。

統合モデルのAICは5496となり、通常のロジットモデル、if-thenルールによるモデルのいずれのモデルよりも優れた値を示しており、非補償型意思決定方略と補償型意思決定方略を同時に考慮することでモデルの適合度が向上したものと考えられる。ただし、的中率は0.690と3つのモデルの中で最低の値を示した。なお、統合モデルのAICはKurauchi and Morikawaで構築された非補償型意思決定方略を取り込んだモデルのAICの値よりも小さく、本研究での統合モデルの有効性が示されたと考えられる。しかしながら、Kurauchi and Morikawaでは、推定作業の困難さから統計的に有意でない変数も多くモデルに含まれており、厳密な比較のためにはKurauchi and Morikawaでのモデルから統計的に有意でない変数を除去した上で再推定することが必要である。

6. おわりに

データマイニング手法は大規模なデータから何らかのパターンを発見するための手法であり、本研究では離散選択行動における非補償型意思決定方略を表現するために、代表的な手法の1つであるC4.5を適用してDP&Rの選択意向に関する分析を行った。データ分析の結果より、C4.5によって生成されたif-thenルールは通常のロジットモデルとほぼ同程度の的中率であった。また、if-thenルールをダミー変数として推定したロジットモデルのAICは通常のロジットモデルより劣っているという結果となった。この結果はWets *et al.*²¹⁾、Yamamoto *et al.*^{22), 23)}と整合的であり、C4.5が誤差項の分布を仮定せず選択結果の再現性を最大化しようとするのに対して、ランダム効用モデルでは誤差項に特定の分布を仮定し選択確率を最大化するという、両モデルの推定方法の違いを反映したものである。Maximum score 推定量を用いた推定結果では、if-thenルールが頑健である一方で、通常のロジットモデルでは誤差項の分布形を仮定しないことで多くの説明変数の統計的有意性を失った。この結果は、C4.5の方が誤差項の分布の仮定に対して頑強であることを意味する。C4.5によって生成された決定木及びif-thenルールは非補償型意思決定方略の存在を示唆するものであったが、実際の意思決定方略が非補償型であるという保証はない。今回の結果からはif-thenルールと通常のロジットモデルのいずれが優れているかについて、断定的な判断は不可能であり、今後も引き続いて比較分析を進める必

要があるものと考えられる。その際、シミュレーションデータにより真の意思決定方略が既知として比較分析を行うことが有効であると考えられる。

C4.5によって生成されたif-thenルールはランダム効用モデルを拡張した非補償型意思決定モデルよりAICが大きく、再現性に劣るが、if-thenルールと通常のロジットモデルを統合したモデルを推定した結果、AICが最も小さいモデルが得られた。これは、統合モデルで非補償型意思決定方略と補償型意思決定方略の両者を表現することが可能であることによるものであると考えられる。

また、今回の分析では、if-thenルールの生成にあたり、選択肢属性に加えて個人属性も用いた。これによって、ルール集合が個人毎に異なるという形で個人間異質性を表現することが可能であった。今後は、非補償型意思決定方略を用いている個人と補償型意思決定方略を用いている個人が混在しているという個人間異質性に関しても、潜在クラスモデル²⁹⁾等を適用したモデルの拡張を行うことで、さらなる適合度の向上が期待される。

謝辞：本稿の分析で用いたデータは名古屋大学・中村英樹助教授が中心となり実施されたアンケート調査に基づくものである。また、3名の査読者には有益なコメントを多数頂いた。ここに記して感謝の意を表します。

参考文献

- 1) Gärling, T., Kwan, M.-P. and Golledge, R.G.: Computational-process modeling of household travel activity scheduling, *Transportation Research B*, Vol. 25B, pp. 355-364, 1994.
- 2) Tversky, A.: Elimination by aspects: A theory of choice, *Psychological Review*, Vol. 79(4), pp. 281-291, 1972.
- 3) Wright, P.: Consumer Choice Strategies: Simplifying vs. Optimizing, *Journal of Marketing Research*, vol. 12, pp. 60-67, 1975.
- 4) Dawes, R.M. and Corrigan, B.: Linear models in decision making, *Psychological Bulletin*, Vol. 81, pp. 95-106, 1974.
- 5) Dawes, R.M.: The robust beauty of improper linear models in decision making, *American Psychologist*, Vol. 34, pp. 571-582, 1979.
- 6) Johnson, E.J., Meyer, R.J. and Ghose, S.: When choice models fail: Compensatory representations in negatively correlated environments, *Journal of Marketing Research*, Vol. 26, pp. 255-270, 1989.
- 7) Vause, M.: A rule-based model of activity scheduling behavior, *Activity-based Approaches to Travel Analysis*, Ettema, D.F. and Timmermans, H.J.P. eds., Elsevier Science, Oxford, UK, pp. 73-88, 1997.
- 8) Arentze, T.A., Hofman, F., Van Mourik, H. and Timmermans, H.J.P.: Albatross: a multi-agent rule-based model of activity pattern decisions, Paper presented at the 79th Annual meeting of the Transportation Research Board, Washington D.C., pp. 9-13, 2000.
- 9) 森地茂, 目黒浩一郎, 小川圭一: 一般化平均概念を用いた交通情報提供の影響分析手法に関する研究, 土木学会論文

- 集, No. 555/IV-34, pp. 15-26, 1997.
- 10) 秋山孝正: 知識利用型の経路選択モデル化手法, 土木計画学研究・論文集, No. 11, pp. 65-72, 1993.
 - 11) 秋山孝正: 知的情報処理を利用した交通行動分析, 土木学会論文集, No. 688/IV-53, pp. 37-47, 2001.
 - 12) 島崎敏一, 安田誠一: ニューラルネットワークによる交通手段選択モデル, 土木学会論文集, No. 494/IV-24, pp. 79-86, 1994.
 - 13) Dougherty, M.: A review of neural networks applied to transport, *Transportation Research C*, Vol. 3, pp. 247-260, 1995.
 - 14) Kurauchi, S. and Morikawa, T.: An exploratory analysis for discrete choice model with latent classes considering heterogeneity of decision making rules, *Travel Behaviour Research: The Leading Edge*, Hensher, D. ed., Pergamon, Oxford, UK, pp. 409-423, 2001.
 - 15) 森川高行, 倉内慎也: 合理的選択の拡張とモデリングへのインプリケーション, 土木学会論文集, No. 702/IV-55, pp. 15-29, 2002.
 - 16) 岩崎学: データマイニングと知識発見, 行動計量, Vol. 26, No. 1, pp. 46-58, 1999.
 - 17) Recker, W.W. and Golob, T.F.: A non-compensatory model of transportation behavior based on sequential consideration of attitudes, *Transportation Research*, Vol. 13B, pp. 269-280, 1979.
 - 18) Murtaugh, M. and Gladwin, H.: A hierarchical decision-process model for forecasting automobile type-choice, *Transportation Research A*, Vol. 14A, pp. 337-348, 1980.
 - 19) Smith, T.R., Clark, W.A.V. and Cotton, J.W.: Deriving and testing production system models of sequential decision-making behavior, *Geographical Analysis*, Vol. 16, No. 3, pp. 191-222, 1984.
 - 20) Clark, W.A.V. and Smith, T.R.: Production system models of residential search behavior: a comparison of behavior in computer-simulated and real-world environments, *Environment and Planning A*, Vol. 17, pp. 555-568, 1985.
 - 21) Wets, G., Vanhoof, K., Arentze, T. and Timmermans, H.: Identifying decision structures underlying activity patterns: an exploration of data mining algorithms, *Transportation Research Record*, No. 1718, pp. 1-9, 2000.
 - 22) Yamamoto, T., Kitamura, R. and Fujii, J.: Driver's route choice behavior: Analysis by data mining algorithms, *Transportation Research Record*, No. 1807, pp. 59-66, 2002.
 - 23) Yamamoto, T., Kurauchi, S. and Morikawa, T.: Comparison of non-compensatory models of driver's choice on dynamic park and ride, *Proceedings of the 9th World Congress on Intelligent Transport Systems*, Chicago, Illinois, CD-ROM, 2002.
 - 24) Quinlan, J.R.: *C4.5 Programs for Machine Learning*, Morgan Kaufmann Publishers, San Mateo, Calif., 1993.
 - 25) Kass, G.: An exploratory technique for investigating large quantities of categorical data, *Applied Statistics*, Vol. 29, No. 2, pp. 119-127, 1980.
 - 26) Breiman, L., Freidman, J.H., Olshen, R.A. and Stone, C.J.: *Classification and regression trees*, Wadsworth, Belmont, Calif., 1984.
 - 27) Manski, C.: Maximum score estimation of the stochastic utility model of choice, *Journal of Econometrics*, Vol. 3, pp. 205-228, 1975.
 - 28) Rissanen, J.: A universal prior for integers and estimation by minimum description length, *Annals of Statistics*, Vol. 11, pp. 416-431, 1983.
 - 29) Biggs, S.F., Bedard, J.C., Gaber, B.G. and Linsmeier, T.J.: The effects of task size and similarity on the decision behavior of bank loan officers, *Management Science*, Vol. 31, pp. 970-987, 1985.
 - 30) Sundstrom, G.A.: Information search and decision making: the effects of information displays, *Acta Psychologica*, Vol. 65, pp. 165-179, 1987.
 - 31) Payne, J.W., Bettman, J.R. and Johnson, E.J.: *The Adaptive Decision Maker*, Cambridge University Press, New York, 1993.
 - 32) 中村英樹, 加藤博和, 内海泰輔, 平田哲: SPモデルを適用した名古屋におけるDynamic Park and Rideの導入効果分析, 土木計画学研究・論文集, No. 16, pp. 949-954, 1999.
 - 33) 平田哲: 名古屋都市圏におけるダイナミック・パーク・アンド・ライドの適用可能性に関する研究, 名古屋大学工学研究科修士論文, 1998.
 - 34) McClelland, G. and Judd, C.: Statistical difficulties of detecting interactions and moderator effects, *Psychological Bulletin*, Vol. 114, pp. 376-390, 1993.
 - 35) Ben-Akiva, M. and Lerman, S.R.: *Discrete Choice Analysis: Theory and Application to Travel Demand*, The MIT Press, Cambridge, 1985.
 - 36) Akaike, H.: Information theory and an extension of the maximum likelihood principle, *2nd International Symposium on Information Theory*, Petrov, B.N. and Csaki, F. eds., Akademiai Kiado, Budapest, pp. 267-281, 1973.
 - 37) Greene, W.H.: *LIMDEP version 8.0: Econometric Modeling Guide*, Econometric Software, New York, 2002.
 - 38) Efron, B. and Tibshirani, R.J.: *An Introduction to the Bootstrap*, Chapman & Hall, London, UK, 1993.
 - 39) Bhat, C.R.: An endogenous segmentation mode choice model with an application to intercity travel, *Transportation Science*, Vol. 31, pp. 34-48, 1997.

(2003. 6. 6 受付)

APPLICATION OF DATA MINING ALGORITHM FOR REPRESENTING NON-COMPENSATORY DECISION STRATEGY

Toshiyuki YAMAMOTO

C4.5, one of the popular data mining algorithms, is applied to represent non-compensatory choice behavior, and the predictability of the choice behavior is compared with the conventional logit model. The empirical analysis of this study is based on the SP data about the use of a hypothetical dynamic park and ride system. The results suggest that C4.5 and the conventional model have a same predictability in terms of the hit ratio. Though, the log-likelihood at convergence of the logit model with the dummy variables representing if-then rules produced by C4.5 is lower than that of the conventional logit model. When the dummy variables are added to the conventional logit model with conventional independent variables, the higher log-likelihood at convergence than that of the conventional logit model is given even after the insignificant variables are excluded.