

決定木分析法の交通行動分析への適用性の検討

岐阜大学工学部 学生員 ○豊田 将隆
 岐阜大学工学部 正会員 奥嶋 政嗣
 岐阜大学工学部 正会員 秋山 孝正

1. はじめに

本研究では、機械学習のなかでも原始的な手法である「木構造」による学習を用いる。「木構造」とは人工知能の分野で考え出された膨大な人間の知識データを帰納的に分類するための方法である。交通行動分析への適用性の検討については、初めに、各トリップメーカーがそれぞれの交通機関を選択する際の決定条件となる属性を求める。次に、求めた属性を木構造によって表現する。最後にその木構造からの的中率を求め、考察を行う。

2. 決定木分析法の手法

ここでは、「ID3」「C4.5」の2つの手法の説明を行う。一般的に木の形状は、ID3では広く浅い決定木、C4.5では狭く深い決定木を生成する特徴がある。ID3とC4.5の関係を図-1に示す。この図は、テスト属性の選択基準、数値属性に対する分類方法の相関関係を表している。

2.1 ID3 アルゴリズムとファジィ決定木への拡張

ID3は情報量を用いることにより「データを分類する時のテスト回数の期待値が最小になる決定木」を作成することを目的とする方法である。獲得情報量 $G(A_i, D)$ の計算方法を以下に示す。

$$G(A_i, D) = M(D) - B(A_i, D)$$

$$M(D) = - \sum P_k \log_2 P_k \cdots D \text{ の情報量}$$

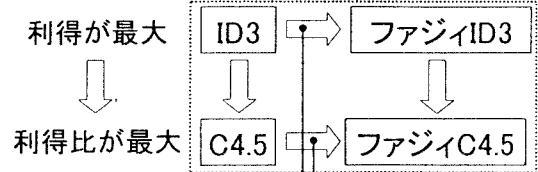
$$B(A_i, D) = \sum P_{ij} I(D_{ij}) \cdots \text{平均情報量}$$

$$P_k = |D_{ck}| / |D| \cdots \text{クラス } C_k \text{ の出現確率}$$

$$P_k = |D_{aij}| / |D| \cdots \text{属性値 } a_{ij} \text{ の出現確率}$$

獲得情報量が最大となるものをテスト節点として選択する。しかし、ID3では属性値がシンボルの属性で表現されたデータしか取り扱うことができない。本研究では連続属性に対して、確信度、メンバーシップ関数を定義する。分類後のデータの確信度は、各データの分類前の確信度の値と属性値がメンバーシップ関数に属する度合いとの積とする。

テスト属性選択基準



数値属性に対する分類方法

図-1 ID3とC4.5の相関図

2.2 C4.5 アルゴリズムとファジィ決定木への拡張

ID3では属性値の数が多い属性が選ばれる傾向にあるので、C4.5ではデータ自体の分割の度合いを表す分割情報量を計算し、獲得情報量と分割情報量との比(情報利得比)が最大となる属性を選択する。情報利得比の計算は以下に示す。

$$S(A_i, D) = - \sum_{j=1}^m (P_{ij} \log_2 P_{ij}) \cdots \text{分割情報比}$$

$$GR(A_i, D) = G(A_i, D) / S(A_i, D) \cdots \text{情報利得比}$$

やはり、C4.5においてもシンボルの属性で表現されたデータしか取り扱うことができないため、ファジィへの拡張方法はID3と同様に連続属性に対して、確信度、メンバーシップ関数を定義する。

3. 決定木分析法の交通行動分析への適用性の検討

決定木分析法を用いて交通機関を決定するモデルを構築する。決定木を作成し的中率を算出する。その値をロジットモデルの的中率と比較することで決定木分析法の交通行動分析への適用性を検討する。

3.1 モデルの概要

モデルの対象地域となる岐阜市を図-2に示す。岐阜市在住者の市街地(1.2ゾーン)への出勤トリップ(1,183トリップ)の一部である215トリップ抽出し決定木分析法を用いて各トリップで利用された交通機関(自動車、マストラ、その他)を決定する。説明変数は、世帯人数、自動車の保有台数、性別、免許の有無、出発時間、各交通手段別の所要時間、自動車の所要費用、距離を用いる。

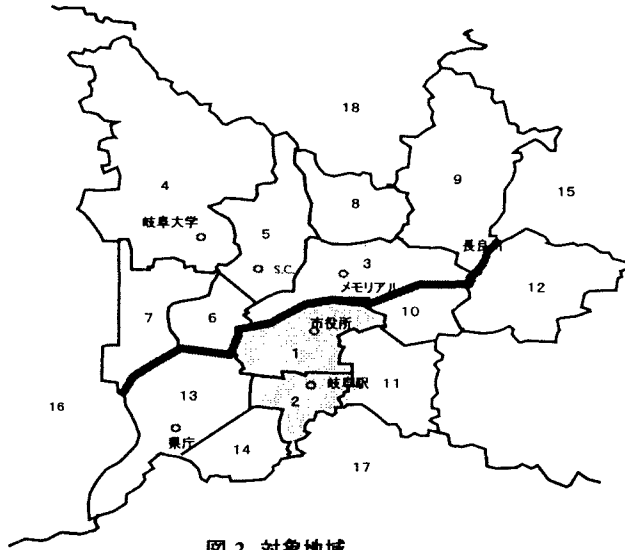


図-2 対象地域

3.2 決定木の作成

ID3とファジィID3を用いて決定木を作成してみる。獲得情報量を求めることでテスト節点を決定することができる。2つの手法ともに第1階層のテスト節点として免許の有無が選択された。この計算を繰り返し行い、テスト節点を決定していくことで決定木を作成することができる。ID3で作成された決定木の上位層を図-3、ファジィID3で作成された決定木の上位層を図-4に示す。また、各説明変数が最初に出現した階層の数を表-1に示す。作成された二つの木を比較してみると、ファジィID3では第2階層で枝を多く持つ変数である世帯人数を選択している。ここでは、ファジィID3で作成された決定木はID3の決定木に比べ、より広い木が作成されている。

3.3 的中率の比較と適用性に関する考察

作成された決定木を上から辿り、正解数を数えることでそのデータ集合の的中率を求めることができる。的中率の結果を表-2に示す。今回のモデルでの的中率は全ての手法においてロジットモデルの的中率よりも非常に良い結果を得ることができた。

これはロジットモデルでは記述不可能であった各属性間の非線形な関係がモデル構造に含まれていることが考えられる。このことより交通行動分析において決定木分析法は適用性があると考えられる。それぞれの決定木を用いたモデルの適中率はほぼどのような値となっている。しかしながら、本来ならファジィ決定木への拡張を行えば、精度は向上するはずであるが、ここではそのような結果を得ることが出来なかった。原因として定義した確信度の値やメンバーシップ関数の形状に問題があると考えられる。

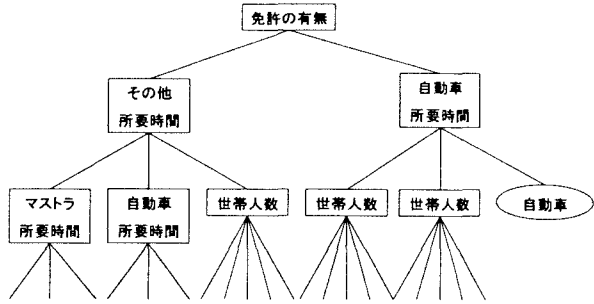


図-3 ID3で作成される決定木の上位層

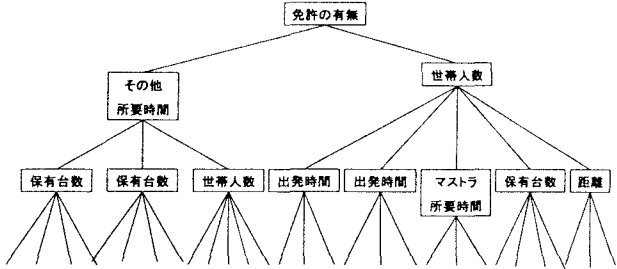


図-4 ファジィID3で作成される決定木の上位層

表-1 説明変数の第一出現階層

説明変数	ID3	ファジィID3
世帯人数	3	2
保有台数	4	3
性別	5	4
年齢	4	4
免許	1	1
出発時間	4	3
所要時間	2	4
マストラ	3	2
その他	2	2
所要費用	-	5
距離	4	3

表-2 的中率の比較

ロジットモデル	64.70%		
ID3	94.80%	ファジィID3	91.40%
C4.5	94.80%	ファジィC4.5	91.40%

4. おわりに

本研究では決定木分析法により各トリップメーカーの交通機関を決定するモデルを構築した。またこのモデルの的中率の結果から決定木分析法は交通行動分析への適用性について確認した。

今後の課題としては、2点挙げられる。

- ①ファジィ決定木における確信度やメンバーシップ関数の見直し。
- ②対象地域の全出勤トリップにモデルを適用しその安定性の確認を行う。

【参考文献】

1) 馬野, 岡本, 鳩野, 田村, 河内, 梅津, 木下: ID3に基づくファジィ・ルールの抽出とその推論, 9th Fuzzy System Symposium (1993)