

# 決定木分析による都市型アミューズメント施設の来訪者特性評価†

梶井昌邦\*, 斎藤参郎\*\*

## 1. 研究のねらいと目的

1996年のキャナルシティ博多のオープンにはじまり、岩田屋Zサイド、マリノアシティ、マリナタウン等、福岡の都心部、及び、近郊では、数多くの新規出店や既存店舗の増床が相次いだ。近年進出している商業施設の大きな特色は、いずれも物販や飲食といった商業機能を核としつつも映画館・観覧車・ゲームセンターなどのアミューズメント施設の併設や、噴水や緑、オープンスペース等の設置、そして、大道芸・ミュージックパフォーマンスなどのイベントによる演出等、エンタテインメントやアメニティといった機能に重点をおいた構成となっている施設が多くみられることにある。このような傾向は、全国的にみられ、これらの機能がどれほどの集客効果をもたらすかを把握することは、これからの店舗戦略を考えていく上で、重要な課題であるといえよう。また、最近では、都市計画のマスタープランにもオープンカフェやオープンスペースの設置が盛り込まれはじめており、エンタテインメントやアメニティは今後のまちづくりを考えていく上でも不可欠な要素となっている。

このような動向の背景には、消費者の意識や嗜好が多様化してきていることが挙げられる。市場には数多くの店舗が存在する。このようななかで自店舗が消費者に選択されるために、店舗は、多種多様な消費者の意識や嗜好を把握した上で、他の店舗にない特色や独自性を打ち立てていかなければならない。物販や飲食という従来の商業機能に、エンタテインメントやアメニティといった機能を付け加え、バラエティにとんだサービスを提供しようという店舗行動は、まさに、このような状況にいち早く対応しようという店舗の反応に他ならない。また、エンタテインメントやアメニティの機能を取り入れたまちづくりが考察され始めていることも、消費者の意識や嗜好の多様化に対応する必然的な流れであるといえよう。

以上のような状況のなか、マーケティング・サイエンスの分野では、one to oneマーケティングに関する手法・モデル開発や実務への応用がさかんに議論されている。one to oneマーケティングとは、これまでのone to manyのマスマーケティングに対応する概念であり、店舗や企業と個々の顧客との間に1対1の関係を築こうとするマーケティングのことである。one to oneマーケティングを実施していくためには、自社や自店舗の顧客が誰であるのかを掴むことが重要となる。その上で、対象とする顧客層を定め、顧客の意識や嗜好に沿ったきめ細かい戦略をとっていく必要がある。すなわち、市場における顧客のセグメントを行った上で、ターゲティングを行い、顧客のニーズに沿った自店舗の独自性や特色を打ち立てていくことが重要となってきているのである。このようにマーケティング・サイエンスの分野では、顧客を軸とした戦略が問題とされている。

本研究は、このような視点・考え方が、これからのまちづくりや都心商業政策を考察していくうえで

† 本稿は、第40回日本地域学会年次大会(於、琉球大学)において発表したものを加筆修正したものである。討論者の労をとって頂いた郵政総合研究所の実積寿也先生(現、九州大学大学院)、早稲田大学の樋口清秀先生をはじめ、多数の方から有益なコメントを頂戴した。ここに記して厚く謝意を表す次第である。

\* 福岡大学経済学部 助教授

\*\* 福岡大学都市空間情報行動研究所 所長、福岡大学経済学部 教授

不可欠であると考え。その理由は次である。まちや商業地にも、多様な嗜好や意識をもった来街者が集まり、これに応じて多種多様な購買パターン、行動パターンが観察されるようになってきている。中心市街地の衰退が問題とされている昨今であるが、まちや商業地は、いかに他のところとの差別化をはかり、激化する商業地間競争を勝ち抜いていくかという課題を抱えている。この問題は、店舗や企業もつ問題と同じ構造をもっており、その問題解決の鍵を握るのは来街者を軸としたまちづくりや都心商業政策であると考えられるからである。

以上の問題意識から、本研究は、都市型アミューズメント施設の1つを取り上げ、その集客力を決める要因となる来訪者特性の評価の問題を考察する。都市型アミューズメント施設とは、エンタテインメント機能の提供をサービスの中心とし、非日常的空間における体験を提供する新しい形態の商業施設である。分析対象の施設として、キャナルシティ博多のラーメンスタジアムをとりあげた。ラーメンスタジアムは、全国の有名なラーメン店を集めた施設であり、2001年12月の開業から2003年2月までの約1年間に200万人の来店者数を記録した集客力のある施設である<sup>1</sup>。

本研究の目的は、マーケティングの分野等で、近年注目されているデータマイニング手法の1つである決定木分析を用い、都市型アミューズメント施設の来訪者特性ルールの抽出による来訪者セグメンテーションを行うことである。具体的には、(1) ラーメンスタジアムへの出向頻度データに、決定木分析の代表的なアルゴリズムであるC4.5を適用し、来訪者特性ルールの抽出を行うとともに、(2) 1Rアルゴリズムを、複合商業施設、百貨店、スーパー等、ラーメンスタジアム以外の業態の出向頻度データにも適用し、得られた来訪者・来店者特性ルールの比較を行っていく。本研究は、このように抽出された複数の商業施設の来訪者特性ルールの比較検討から、どのような商業機能がどのような層の集客に寄与しているかを評価する初めての試みである。

このような分析をとおり、標準的な統計的手法では得ることが不可能であった来訪者特性のルールを抽出する1つの事例を提示することで、都市・地域研究にとって決定木分析が有効なツールとなることを示すことも本研究のねらいのひとつである。

## 2. 分析枠組

### 2.1 決定木分析とは

決定木 (decision tree) とは、機械学習の分野で発展してきた代表的な命題学習器であり、これにより、論理命題によって表現される知識を生成し、大量のデータ群を分類する枠組みを与える。2.1.1で、決定木分析によって抽出される知識の表現形態、2.1.2で、本研究で用いるアルゴリズムについて説明していくこととする。

#### 2.1.1 決定木分析の知識表現

決定木分析によるデータの学習結果は、次の2つの方法で表現される。

##### (1) ツリー構造による表現

決定木分析の知識を表現する1つの方法は、データ群全体をツリー構造の図によって表すものである。条件判断の結果は部分木として表現される。図2.1は、属性  $A, B, C, Y$  をもつリレーションから生成された (属性  $Y$  に関する) 決定木を表現した例である。この図において各矢印を枝 (branch) という。丸であらわされた節点 (node) は、その中にある属性に対応し、その属性値に応じたデータ分類のための分割テストを表している。終端の四角で表された節点を、葉 (leaf) と呼ぶ。中の文字は、属性  $Y$  のカ

<sup>1</sup> 参考文献 [7] を参照されたい。

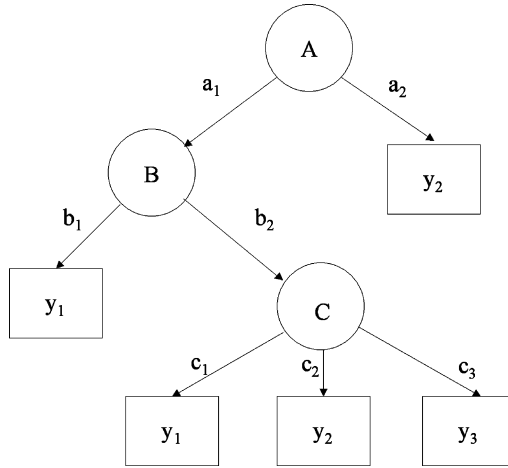


図 2.1 ツリー構造による決定木の表現

テゴリーを表しており、 $Y$  の属性値  $y_1, y_2, y_3$  のいずれかが書かれている。頂点の節点を根 (root) とよぶ。この決定木は、属性  $Y$  のカテゴリーを決定するためにどの属性 (属性どうしの組合せ) で分類するのが有効であるかを表すもので、根から葉に枝を辿って順に読む。ここで、属性  $Y$  を目的属性、分割につかわれる属性 ( $A, B, C$ ) を条件属性と呼ぶ。ルートに近い分岐を生じさせている条件属性が、目的属性に対し強い影響を与えている、すなわち、この図では、ターゲット属性  $Y$  に影響を与えている第 1 の要因は属性  $A$ 、第 2 の要因は属性  $B$ 、第 3 の要因は  $C$  となっている、と解釈することが出来る。

(2) IF-THEN ルールによる表現

上で説明した、根から葉にいたるまでの分割テストは、IF-THEN ルールによって表現することができる。図 2.1 を IF-THEN ルールで、表現すると次のようになる。

- If  $A=a_1$  and  $B=b_1$  then  $Y=y_1$
- If  $A=a_1$  and  $B=b_2$  and  $C=c_1$  then  $Y=y_1$
- If  $A=a_1$  and  $B=b_2$  and  $C=c_2$  then  $Y=y_2$
- If  $A=a_1$  and  $B=b_2$  and  $C=c_3$  then  $Y=y_3$
- If  $A=a_2$  then  $Y=y_2$

2.1.2 1R アルゴリズム

1R アルゴリズム (Holte (1993)) とは、1 段階の決定木であり、結果を単一の属性の分割テストによるルールの集合として表現する。1R アルゴリズムの分割テストの分割基準は、誤り率 (error rate) であり、誤り率が最小となるルールを求め、そのなかでもっとも誤り率の低くなる属性に対応するルールを選択する。1R は、与えられたクラス情報をもとに学習を行う、教師付き学習アルゴリズムである。1R アルゴリズムの擬似コードを以下に示す<sup>2</sup>。

```

FOR EACH 属性 Do
  FOR EACH 属性値 Do
    1) クラスの出現頻度を計算
  
```

<sup>2</sup> 1R アルゴリズムの詳細は Witten, I. H. and E. Frank (2000) を参照されたい。

- 2) 最大頻度のクラスを見つける
- 3) 属性値に 2) で見つけた最大頻度クラスを割り当てるルールを生成
- 4) ルールの誤り率を計算

END

生成されたルールセットの属性別の誤り率を計算

END

総誤り率が最小の属性を選択

この属性によるルールセットを 1R (1-rule) として選択

### 2.1.3 C4.5 アルゴリズム

C4.5 (Quinlan (1995)) は, ID3 (Quinlan (1986)) をもとに開発された, 分割統治法 (Divide and Conquer) によるアルゴリズムである<sup>3</sup>。C4.5 も 1R 同様, 分類を行うクラスをあらかじめ決めておく。C4.5 アルゴリズムを以下に示す。

1. ルートに対応するノードを作成
  2. リレーション  $S$  に含まれる全ての事例が同一クラスに属するとき
    - 1) ノードに最大頻度クラス (クラス分布) のラベルをつける
    - 2) 手続き停止
  3. それ以外のとき
    - (1) 分割テスト
      - 1) 分割テストを行い 1 つの属性  $A$  を選択
      - 2) ノードに属性  $A$  のラベルをつける
    - (2) 部分木の構築
      - 1) 属性  $A$  の属性値  $S$  によりを部分集合  $S_1, S_2, \dots, S_n$  に分割
      - 2) ルートノードの下に  $S_1, S_2, \dots, S_n$  に対応する枝をはる
      - 3) 枝に属性値のラベルをつける
      - 4) 枝の下に  $S$  の部分集合に対応するノードを作成
      - 5)  $S$  の部分集合に含まれる全ての事例が同一クラスに属するとき  
ノードに最大頻度クラス (クラス分布) のラベルをつける
- それ以外のとき, それに対応する各々のノードについて 3. を再帰的に繰り返す

次に, 属性を選択する分割テストを説明する。データセット (リレーション) を  $S$ ,  $S$  を分類するクラスを  $Y = \{y_1, y_2, \dots, y_n\}$  としよう。  $Y$  を確率関数  $P(Y = y_i)$  をもつ確率変数であるとみなすと,  $S$  の事象  $Y$  に対するエントロピーは,

$$H(Y) = - \sum_{i=1}^n P(Y = y_i) \log P(Y = y_i)$$

となる。ここで,  $P(Y = y_i)$  は,  $y_1, y_2, \dots, y_n$  のなかでの事象  $y_i$  ( $i = 1, 2, \dots, n$ ) の出現率を表す確率関数である。また,  $X$  のもとでの  $Y$  の条件付エントロピーは,

<sup>3</sup> 解くべき問題をいくつかの部分問題に分割し, それを統合して問題の解を得る方法を分割統治法 (Divide and Conquer) という。

$$H(Y | X) = \sum_{j=1}^k P(X=x_j)H(Y | X=x_j)$$

となる。

ここで、 $P(X=x_j)$  は、属性  $X$  の属性値  $x_1, x_2, \dots, x_k$  のなかでの事象  $x_j$  ( $j=1, 2, \dots, k$ ) の出現率を表す確率関数であり、 $H(Y | X=x_j)$  は、

$$H(Y | X=x_j) = \sum_{i=1}^n P(Y=y_i | X=x_j) \log P(Y=y_i | X=x_j)$$

で定義される。また、 $H(Y)$  と  $H(Y | X)$  との差をとった、

$$\text{Infor. gain} = H(Y) - H(Y | X)$$

を情報ゲイン (information gain) いう。情報ゲインは、属性  $X$  によってあいまいさがどれだけ減ったかの度合いを表すものと解釈でき、ID3 の分割テストの分割基準として採用されている。本研究で用いる C4.5 では、これを、分割情報量 (split information)

$$\text{Split.info} = - \sum_x p(x) \log p(x)$$

で割ったゲイン率 (gain ratio)、

$$\text{Gain.ratio} = \frac{\text{Infor. gain}}{\text{Split.info}}$$

が分割テストの判定基準として使われている。分割テストでは、ゲイン率を最大にする属性が選ばれる。

## 2.2 使用するデータ

本研究では、2002年6月に福岡大学都市空間情報行動研究所が実施した、第7回福岡都心部回遊行動調査のデータを使用した<sup>4</sup>。回遊行動調査とは、都心部の主要な施設に調査地点を設置し、調査地点に立ち寄った来街者を対象に調査当日立ち寄った店舗、目的、予算等を行動の生起順に尋ねる調査である。調査では、被験者の個人属性や主要商業施設や主要商業集積地区への出向頻度等も同時に尋ねている。第7回福岡都心部回遊行動調査では、出向頻度項目にキャナルシティラーメンスタジアムを新たに追加し、これとともに、ラーメンスタジアムの認知度やそれを知った情報媒体等に関する質問項目を設けている。第7回福岡都心部回遊行動調査の概要は次の通りである。

---

### 第7回福岡都心部回遊行動調査

調査日時：2002年6月28日(金)、29日(土)、30日(日)

回収票数：1155票

調査方法：聞き取りアンケート調査

調査場所：キャナルシティ博多、岩田屋 Z-side、大丸・エルガーラ、ショッピングダイエー福岡、博多リバレイン、博多駅コンコース、ソラリアプラザ、ソラリアステージ、福岡三越

調査主体：福岡大学都市空間情報行動研究所

主要質問項目：個人属性、都心部回遊行動(立ち寄り地点、目的、支出金額)、主要商業集積地への

---

<sup>4</sup> 本調査における実査は、福岡大学都市空間情報行動研究所の指導のもと、福岡大学経済学部産業経済学科講義フィールド調査、福岡大学斎藤ゼミ、及び、梅井ゼミの受講生が行っている。

## 出向頻度，主要商業施設への出向頻度

## 3. キャナルシティラーメンスタジアムの利用形態

## 3.1 キャナルシティラーメンスタジアムの認知度とその情報源

図 3.1.1 はラーメンスタジアムの認知度，表 3.1.1 はラーメンスタジアムを知っていると答えたサンプルがどのような情報源で施設を知ったかを集計したものである。図 3.1.1 より福岡都心部の来街者の 87.0% がラーメンスタジアムを知っており，かなり高い認知度であることが分かる。情報源に関しては，口コミが 32.0% と圧倒的に高く，テレビ (24.6%)，タウン情報誌の九州ウォーカー (13.2%) と続いている。

## 3.2 ラーメンスタジアムの来店者特性

図 3.2.1 は，開業から調査日までの 6 ヶ月間に，何回ラーメンスタジアムに訪問したかの出向頻度の分布を示したものであり，図 3.2.2 は，出向頻度のデータを，一回以上ラーメンスタジアムに訪れたことがあるとそうでないというカテゴリをもつ変数に変換し，それを集計したものである。この変換データを，次章以降の決定木分析で目的属性として用いる。

図 3.2.3 から図 3.2.8 は，図 3.2.2 の分析を個人属性や施設までの時間距離といった属性別に，クロス集

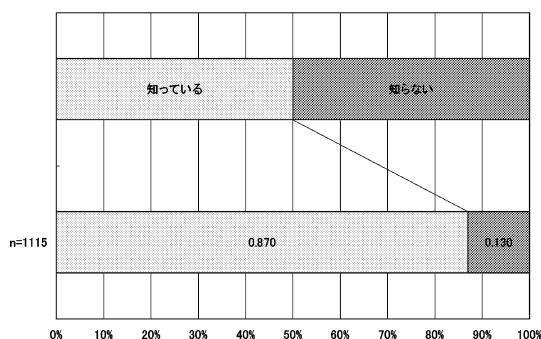


図 3.1.1 ラーメンスタジアムの認知度

表 3.1.1 ラーメンスタジアムの情報源 (複数回答)

|                    | 度数   | 構成比   | 実回答者数に対する比率 |
|--------------------|------|-------|-------------|
| シティ情報FUKUOKA       | 86   | 0.063 | 0.088       |
| 九州ウォーカー            | 179  | 0.132 | 0.183       |
| その他情報誌             | 31   | 0.023 | 0.032       |
| 週刊誌・月刊誌            | 16   | 0.012 | 0.016       |
| 旅行ガイド・タウンガイド       | 4    | 0.003 | 0.004       |
| 新聞                 | 30   | 0.022 | 0.031       |
| テレビ                | 333  | 0.246 | 0.340       |
| ラジオ                | 21   | 0.015 | 0.021       |
| インターネット            | 4    | 0.003 | 0.004       |
| ダイレクトメール           | 1    | 0.001 | 0.001       |
| チラシ・看板・車内広告        | 72   | 0.053 | 0.074       |
| 友人どうしの話(口コミ)       | 434  | 0.320 | 0.443       |
| キャナルシティ博多に訪れたとき知った | 145  | 0.107 | 0.148       |
|                    | 1356 | 1.000 |             |

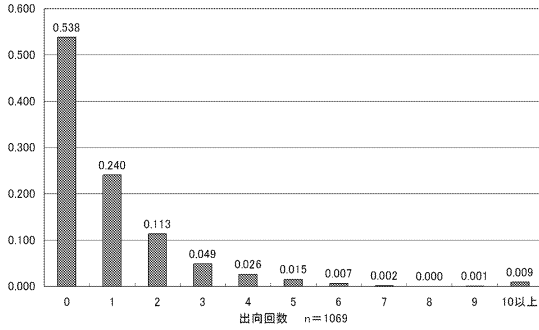


図 3.2.1 ラーメンスタジアムへの出向回数

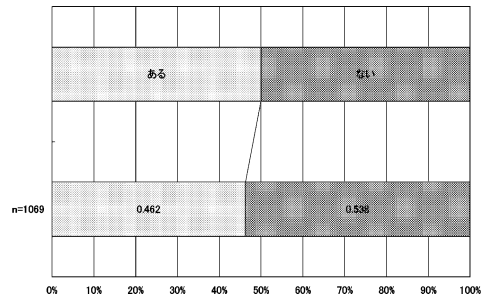
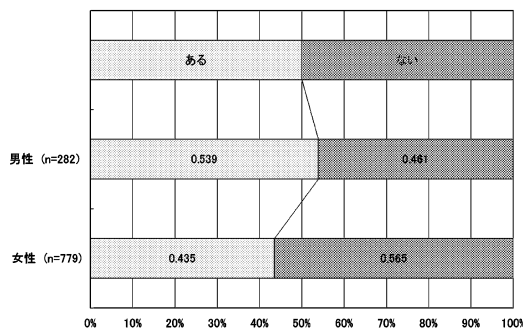


図 3.2.2 ラーメンスタジアム来店の有無

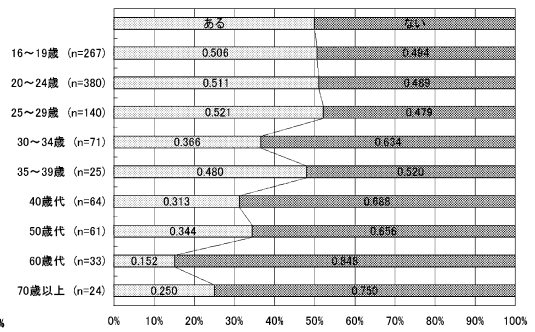
計をした結果である。図 3.2.3 は男女別の集計結果である。出向経験があると答えた割合が、男性の 53.9% に対し、女性が 43.5% となっており、男性が女性よりも出向経験があると答えた比率が高くなっている。これは、百貨店、専門店等の通常の商業施設と逆の傾向であると思われる。図 3.2.4 は、年齢別の集計結果である。来店経験があると答えた比率は、老年層に対し、若年層、中年層が高く、全体として若い層ほど出向経験があると答えた比率が高い傾向がみられた。図 3.2.5 は、職業別の集計結果である。出向経験があると答えた比率は、その他を除くと、上から順に、短大・大学生が 54.3%、販売・サービス系勤め人が 51.7%、高校生が 45.7% となっている。図 3.2.6 は、自宅からキャナルシティまでの時間距離別の集計結果である。ここでは、時間距離が大きいほど来店経験があると答えた比率が小さくなるといったきれいな傾向がみられたが、 $\chi^2$  検定の結果は、有意差なしであった。図 3.2.7 は、結婚の有無別の集計結果である。来店経験があると答えた比率は、既婚者の 34.8% に対し、未婚者は 48.8% であった。図 3.2.8 は、自家用車の所有形態別の集計結果であるが、来店経験の比率に所有形態別に大きな違いは見られず、 $\chi^2$  検定を行った結果は、有意差なしであった。

次章以降の決定木分析では、これらの変数を条件属性として用いる。



$\chi^2(1) = 8.979, p < 0.01$

図 3.2.3 ラーメンスタジアム来店の有無 (男女別)



$\chi^2(8) = 36.583, p < 0.01$

図 3.2.4 ラーメンスタジアム来店の有無 (年齢別)

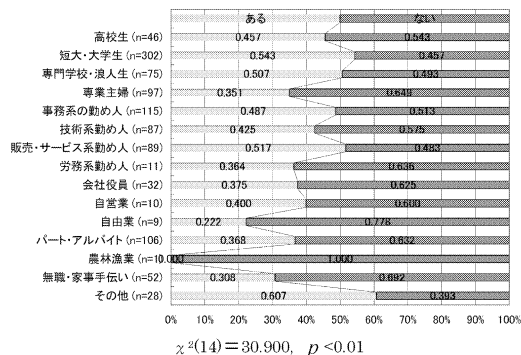


図 3.2.5 ラーメンスタジアム来店の有無 (職業別)

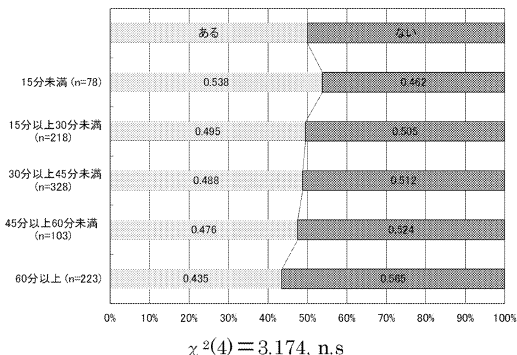


図 3.2.6 ラーメンスタジアム来店の有無 (時間距離別)

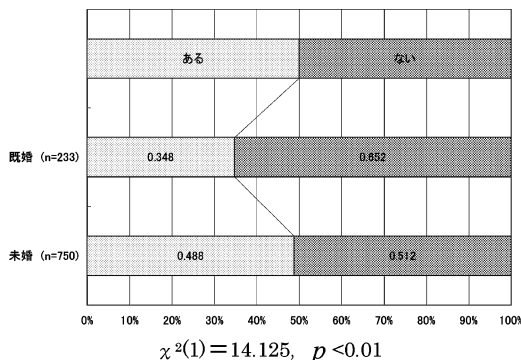


図 3.2.7 ラーメンスタジアム来店の有無 (結婚の有無別)

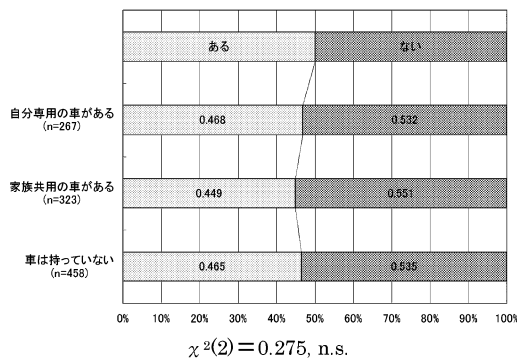


図 3.2.8 ラーメンスタジアム来店の有無 (自家用車の所有形態別)

#### 4. 決定木分析による都市型アミューズメント施設の来訪者特性の評価

##### 4.1 C4.5 アルゴリズムによる都市型アミューズメント施設の来訪者特性ルール抽出

###### 4.1.1 分析の入力変数情報

分析に用いた、目的属性、及び、条件属性は次である。

目的属性：ラーメンスタジアムへの出向の有無(開業から調査日までの6ヶ月間)

条件属性：(1) 性別 (男性, 女性)

(2) 年齢(16~19歳, 20~24歳, 25~29歳, 30~34歳, 35~39歳, 40歳代, 50歳代, 60歳代, 70歳以上)

(3) 職業 (高校生, 短大・大学生, 専門学校・浪人生, 専業主婦, 事務系勤め人, 技術系勤め人, 販売・サービス勤め人, 労務系勤め人, 会社役員, 自営業, 自由業, パートアルバイト, 農林漁業, 無職・家事手伝い, その他)

(4) 自家用車の所有形態(自分専用の車がある, 家族共用の車がある, 車は持っていない)

(5) 自宅からの時間距離 (数値データ)

(6) 結婚の有無 (既婚, 未婚)



```

年齢 = 16～19 歳
| 性別 = 男性: 出向経験有り (50.44/17.07)
| 性別 = 女性
|   職業 = 高校生
|     自家用車の所有 = 自分専用の車がある: 出向経験無し (2.05)
|     自家用車の所有 = 家族共用の車がある: 出向経験有り (14.37/4.37)
|     自家用車の所有 = 車は持っていない
|       時間距離 (チャンネル) <= 21: 出向経験有り (3.33/0.15)
|       時間距離 (チャンネル) > 21: 出向経験無し (19.43/5.0)
|     自家用車の所有 = 無回答: 出向経験無し (0.0)
|   職業 = 短大・大学生: 出向経験無し (107.93/52.12)
|   職業 = 専門学校・浪人生
|     自家用車の所有 = 自分専用の車がある: 出向経験無し (3.25/1.25)
|     自家用車の所有 = 家族共用の車がある: 出向経験有り (6.81/2.81)
|     自家用車の所有 = 車は持っていない
|       時間距離 (チャンネル) <= 30: 出向経験有り (19.02/6.0)
|       時間距離 (チャンネル) > 30: 出向経験無し (12.17/2.17)
|     自家用車の所有 = 無回答: 出向経験無し (0.0)
|   職業 = 専業主婦: 出向経験無し (1.0/0.0)
|   職業 = 事務系勤め人: 出向経験無し (4.02/1.02)
|   職業 = 技術系勤め人: 出向経験有り (1.0)
|   職業 = 販売・サービス系勤め人: 出向経験有り (3.01/1.0)
|   職業 = 労務系勤め人: 出向経験無し (0.0)
|   職業 = 会社役員: 出向経験無し (0.0)
|   職業 = 自営業: 出向経験無し (0.0)
|   職業 = 自由業: 出向経験無し (0.0)
|   職業 = パート・アルバイト: 出向経験無し (15.88/4.07)
|   職業 = 農林漁業: 出向経験無し (0.0)
|   職業 = 無職・家事手伝い: 出向経験有り (3.01/1.0)
|   職業 = その他: 出向経験有り (1.26)
|   職業 = 無回答: 出向経験無し (0.0)
| 性別 = 無回答: 出向経験有り (0.0)
年齢 = 20～24 歳
| 性別 = 男性: 出向経験有り (127.05/46.05)
| 性別 = 女性: 出向経験無し (254.38/113.71)
| 性別 = 無回答: 出向経験有り (0.0)
年齢 = 25～29 歳
| 職業 = 高校生: 出向経験無し (0.13/0.0)
| 職業 = 短大・大学生: 出向経験無し (1.01/0.01)
| 職業 = 専門学校・浪人生: 出向経験無し (3.29/1.15)
| 職業 = 専業主婦
|   時間距離 (チャンネル) <= 35: 出向経験無し (5.69/2.06)
|   時間距離 (チャンネル) > 35: 出向経験有り (3.37/0.37)
| 職業 = 事務系勤め人: 出向経験有り (41.29/17.0)
| 職業 = 技術系勤め人: 出向経験無し (31.22/12.22)
| 職業 = 販売・サービス系勤め人
|   結婚 = 既婚: 出向経験無し (3.31/0.31)
|   結婚 = 未婚: 出向経験有り (20.86/5.0)
|   結婚 = 無回答: 出向経験有り (0.0)
| 職業 = 労務系勤め人: 出向経験有り (0.0)
| 職業 = 会社役員: 出向経験有り (2.01/1.0)
| 職業 = 自営業: 出向経験無し (1.01/0.01)
| 職業 = 自由業: 出向経験有り (2.01/1.0)
| 職業 = パート・アルバイト: 出向経験有り (14.1/5.0)
| 職業 = 農林漁業: 出向経験有り (0.0)
| 職業 = 無職・家事手伝い: 出向経験無し (8.06/1.06)
| 職業 = その他: 出向経験有り (3.15/1.0)
| 職業 = 無回答: 出向経験有り (0.0)
年齢 = 30～34 歳: 出向経験無し (71.27/26.13)
年齢 = 35～39 歳
| 時間距離 (チャンネル) <= 30: 出向経験無し (9.8/2.39)
| 時間距離 (チャンネル) > 30: 出向経験有り (15.29/5.64)
    
```

図 4.1.1 C4.5 アルゴリズムの適用結果—枝刈りされた決定木 (チャンネルシティラームスタジアム)

|  |
|--|
| 年齢 = 40 歳代: 出向経験無し (64.24/20.12)         |
| 年齢 = 50 歳代                               |
| 時間距離 (チャンネル) <= 70                       |
| 時間距離 (チャンネル) <= 55: 出向経験無し (38.91/13.39) |
| 時間距離 (チャンネル) > 55: 出向経験有り (5.87/0.68)    |
| 時間距離 (チャンネル) > 70: 出向経験無し (16.45/2.54)   |
| 年齢 = 60 歳代: 出向経験無し (33.12/5.06)          |
| 年齢 = 70 歳代以上                             |
| 時間距離 (チャンネル) <= 16: 出向経験有り (2.67/0.45)   |
| 時間距離 (チャンネル) > 16: 出向経験無し (21.42/3.82)   |
| 年齢 = 無回答: 出向経験無し (0.0)                   |
| 予測的中率 63.985 %                           |
| 誤り率 36.015 %                             |

図 4.1.1 つづき

|             |   |                   |   |                  |        |                     |       |                   |       |        |      |      |      |
|-------------|---|-------------------|---|------------------|--------|---------------------|-------|-------------------|-------|--------|------|------|------|
| 年齢 = 16~19歳 | & | 性別 = 女性           | & | 職業 = 技術系勤め人      | →      | 出向経験あり              | 1     | 0.00              |       |        |      |      |      |
| 年齢 = 16~19歳 | & | 性別 = 女性           | & | 職業 = その他         | →      | 出向経験あり              | 126   | 0.00              |       |        |      |      |      |
| 年齢 = 16~19歳 | & | 性別 = 女性           | & | 職業 = 高校生         | &      | 自家用車の所有 = 車は持っていない  | &     | 時間距離(チャンネル) <= 21 | →     | 出向経験あり | 3.33 | 0.15 | 0.05 |
| 年齢 = 25~29歳 | & | 職業 = 専業主婦         | & | 時間距離(チャンネル) > 35 | →      | 出向経験あり              | 3.37  | 0.37              | 0.11  |        |      |      |      |
| 年齢 = 50歳代   | & | 時間距離(チャンネル) <= 70 | & | 時間距離(チャンネル) > 55 | →      | 出向経験あり              | 5.87  | 0.68              | 0.12  |        |      |      |      |
| 年齢 = 70歳代以上 | & | 時間距離(チャンネル) <= 16 | → | 出向経験あり           | 2.67   | 0.45                | 0.17  |                   |       |        |      |      |      |
| 年齢 = 25~29歳 | & | 職業 = 販売・サービス系勤め人  | & | 結婚 = 未婚          | →      | 出向経験あり              | 20.86 | 5                 | 0.24  |        |      |      |      |
| 年齢 = 16~19歳 | & | 性別 = 女性           | & | 職業 = 高校生         | &      | 自家用車の所有 = 家族共有の車がある | →     | 出向経験あり            | 14.37 | 4.37   | 0.30 |      |      |
| 年齢 = 16~19歳 | & | 性別 = 女性           | & | 職業 = 専門学校・浪人生    | &      | 時間距離(チャンネル) <= 30   | →     | 出向経験あり            | 19.02 | 6      | 0.32 |      |      |
| 年齢 = 25~29歳 | & | 職業 = その他          | → | 出向経験あり           | 3.15   | 1                   | 0.32  |                   |       |        |      |      |      |
| 年齢 = 16~19歳 | & | 性別 = 女性           | & | 職業 = 販売・サービス系勤め人 | →      | 出向経験あり              | 3.01  | 1                 | 0.33  |        |      |      |      |
| 年齢 = 16~19歳 | & | 性別 = 女性           | & | 職業 = 無職・家事手伝い    | →      | 出向経験あり              | 3.01  | 1                 | 0.33  |        |      |      |      |
| 年齢 = 16~19歳 | & | 性別 = 男性           | → | 出向経験あり           | 50.44  | 17.07               | 0.34  |                   |       |        |      |      |      |
| 年齢 = 25~29歳 | & | 職業 = パート・アルバイト    | → | 出向経験あり           | 14.1   | 5                   | 0.35  |                   |       |        |      |      |      |
| 年齢 = 20~24歳 | & | 性別 = 男性           | → | 出向経験あり           | 127.05 | 46.05               | 0.36  |                   |       |        |      |      |      |

図 4.1.2 C4.5 によるラーメンスタジアムの来訪者特性ルール抽出結果 (誤り率 ≤ 36%)

#### 4.1.2 ラーメンスタジアムへの C4.5 アルゴリズムの適用結果

図 4.1.1 は、チャンネルシティラーメンスタジアムへの出向の有無に関する、C4.5 アルゴリズムの適用結果を示したものである。決定木では、根に近いノードに対応する条件属性ほど、目的属性のクラス分類に大きな影響を与えていると解釈できる。図 4.1.1 をみると、第 1 階層の属性は年齢となっており、ラーメンスタジアムへの出向の有無を決める 1 番目の要因は、年齢であることが分かる。第 2 階層は、16 歳から 19 歳、20 歳から 24 歳の層は性別、25 歳から 29 歳の層は職業、35 歳から 39 歳、50 歳代、70 歳代以上では、自宅からの時間距離となっており、若年層では、性別や職業、中高年層は時間距離が出向の有無の決定要因になっている傾向が把握できる。その他の層は、第 1 階層の年齢が葉ノードとなっている。

ツリーは、全部で 59 の葉からなり、59 の来訪者特性ルールが抽出されている。図 4.1.2 は、これらのルールの中で、分類クラスが出向経験有りとなっているルールを取り出し IF-THEN 形式に変換したものである。但し、葉ノードの事例数がゼロ、推定誤り率が 36% より大きいルールを取り除いている<sup>5</sup>。15 のルールが抽出されている。これらのルールを見ると、30 代や 40 代の属性はまったく抽出されておらず、若年層もしくは、高齢層をひきつけている施設であることが分かる。その他の条件属性もあわせてみると、若年層の男性、若年層の女子学生、若年層の無職・家事手伝いの女性、近郊に住む老年層といった層がラーメンスタジアムの主要な来訪者層であることが分かる。

## 4.2 1R による主要商業施設における来訪者・来店者特性ルールの比較分析

### 4.2.1 分析対象

本節では、都心部に立地する他の商業施設のルールとの比較から、都市型アミューズメント施設の来

<sup>5</sup> 総誤り率の水準を下回るルールを採用した。

訪者特性ルールの特徴を検討していく。分析アルゴリズムには、1Rを採用する。比較する商業施設として、(1) 複合商業施設、(2) スーパーマーケット、(3) 百貨店、(4) 専門店の4つの業態から、福岡に立地する以下の7つの施設を取り上げた。

- (1) 複合商業施設 : キャナルシティ博多
- (2) スーパーマーケット : ショッピングダイエー福岡
- (3) 百貨店 : 岩田屋, 岩田屋Zサイド, 博多大丸, 福岡三越
- (4) 専門店 : ソラリアプラザ

キャナルシティ博多は、1996年に博多地区に開業した、九州を代表する複合商業施設である。キャナルシティ博多は、物販系のほかに、AMC映画館や劇団四季のミュージカル専用の福岡シティ劇場、ゲーム等のアミューズメント施設、ホテル等、様々な機能をもつ施設である。ショッピングダイエーは、全国に展開しているスーパーマーケットである。岩田屋は、福岡資本の地場の老舗百貨店であり、岩田屋Zサイドは、若者に特化した商品構成で岩田屋が1996年に新たに天神地区にオープンさせた百貨店である。ソラリアプラザは、物販系、飲食系中心の地場の専門店である。上にあげた店舗は、キャナルシティ博多をのぞき全て、九州最大の商業集積地である天神地区に立地している。

都心部回遊行動調査では、これらの施設の出向頻度を、平均出向頻度回答法<sup>6</sup>で尋ねている。4.1.1の分析にあわせ、平均して6ヶ月に1回以上施設に訪れるときにアウト頻度、それ以外のときアウト頻度となの変数を作成し、これらを目的属性に採用している。

#### 4.2.2 都心部主要商業施設への1Rアルゴリズムの適用結果

上記8商業施設への1Rアルゴリズムの適用結果を、図4.2.1から図4.2.8に示す。選択された属性を順にみていこう。都市型アミューズメント施設のラーメンスタジアムでは職業、複合商業施設のキャナルシティ博多、スーパーマーケットのショッピングダイエー福岡では時間距離が選択されている。百貨店では、福岡三越は性別が選択されているが、それ以外の全ての店舗で年齢が選ばれている。専門店のソラリアプラザもまた年齢が選択された。次に抽出された、来訪者・来店者特性ルールを見ていこう。

図4.2.1はキャナルシティラーメンスタジアムへの1Rアルゴリズムの適用結果である。属性には、職

|                  |   |        |
|------------------|---|--------|
| 職業:              |   |        |
| 職業 = 高校生         | ⇒ | 出向経験無し |
| 職業 = 短大・大学生      | ⇒ | 出向経験有り |
| 職業 = 専門学校・浪人生    | ⇒ | 出向経験有り |
| 職業 = 専業主婦        | ⇒ | 出向経験無し |
| 職業 = 事務系勤め人      | ⇒ | 出向経験無し |
| 職業 = 技術系勤め人      | ⇒ | 出向経験無し |
| 職業 = 販売・サービス系勤め人 | ⇒ | 出向経験有り |
| 職業 = 労務系勤め人      | ⇒ | 出向経験無し |
| 職業 = 会社役員        | ⇒ | 出向経験無し |
| 職業 = 自営業         | ⇒ | 出向経験無し |
| 職業 = 自由業         | ⇒ | 出向経験無し |
| 職業 = パート・アルバイト   | ⇒ | 出向経験無し |
| 職業 = 農林漁業        | ⇒ | 出向経験無し |
| 職業 = 無職・家事手伝い    | ⇒ | 出向経験無し |
| 職業 = その他         | ⇒ | 出向経験有り |
| 職業 = 無回答         | ⇒ | 出向経験無し |
| 予測的中率 57.1562 %  |   |        |

図 4.2.1 1Rアルゴリズムの適用結果 (キャナルシティラーメンスタジアム)

<sup>6</sup> 出向頻度の質問方法については、斎藤参郎・梶井昌邦・中嶋貴昭 (1999)を参照されたい。

|                |           |
|----------------|-----------|
| 時間距離 (チャンネル) : |           |
| < 122.5        | ⇒ 出向多頻度   |
| >= 122.5       | ⇒ 出向少頻度   |
| 無回答            | ⇒ 出向少頻度   |
| 予測的中率          | 71.0952 % |

図 4.2.2 1R アルゴリズムの適用結果 (チャンネルシティ博多)

|             |           |
|-------------|-----------|
| 時間距離 (天神) : |           |
| < 5.5       | ⇒ 出向多頻度   |
| >= 5.5      | ⇒ 出向少頻度   |
| 無回答         | ⇒ 出向少頻度   |
| 予測的中率       | 60.2667 % |

図 4.2.3 1R アルゴリズムの適用結果 (ショッピングダイエー福岡)

|              |           |
|--------------|-----------|
| 年齢 :         |           |
| 年齢 = 16~19 歳 | ⇒ 出向少頻度   |
| 年齢 = 20~24 歳 | ⇒ 出向少頻度   |
| 年齢 = 25~29 歳 | ⇒ 出向少頻度   |
| 年齢 = 30~34 歳 | ⇒ 出向少頻度   |
| 年齢 = 35~39 歳 | ⇒ 出向少頻度   |
| 年齢 = 40 歳代   | ⇒ 出向多頻度   |
| 年齢 = 50 歳代   | ⇒ 出向多頻度   |
| 年齢 = 60 歳代   | ⇒ 出向多頻度   |
| 年齢 = 70 歳以上  | ⇒ 出向多頻度   |
| 年齢 = 無回答     | ⇒ 出向少頻度   |
| 予測的中率        | 70.4728 % |

図 4.2.4 1R アルゴリズムの適用結果 (岩田屋)

|              |           |
|--------------|-----------|
| 年齢 :         |           |
| 年齢 = 16~19 歳 | ⇒ 出向多頻度   |
| 年齢 = 20~24 歳 | ⇒ 出向多頻度   |
| 年齢 = 25~29 歳 | ⇒ 出向多頻度   |
| 年齢 = 30~34 歳 | ⇒ 出向多頻度   |
| 年齢 = 35~39 歳 | ⇒ 出向多頻度   |
| 年齢 = 40 歳代   | ⇒ 出向多頻度   |
| 年齢 = 50 歳代   | ⇒ 出向少頻度   |
| 年齢 = 60 歳代   | ⇒ 出向少頻度   |
| 年齢 = 70 歳以上  | ⇒ 出向少頻度   |
| 年齢 = 無回答     | ⇒ 出向少頻度   |
| 予測的中率        | 76.4337 % |

図 4.2.5 1R アルゴリズムの適用結果 (岩田屋 Z サイド)

|              |          |
|--------------|----------|
| 年齢 :         |          |
| 年齢 = 16~19 歳 | ⇒ 出向少頻度  |
| 年齢 = 20~24 歳 | ⇒ 出向多頻度  |
| 年齢 = 25~29 歳 | ⇒ 出向多頻度  |
| 年齢 = 30~34 歳 | ⇒ 出向多頻度  |
| 年齢 = 35~39 歳 | ⇒ 出向多頻度  |
| 年齢 = 40 歳代   | ⇒ 出向多頻度  |
| 年齢 = 50 歳代   | ⇒ 出向多頻度  |
| 年齢 = 60 歳代   | ⇒ 出向多頻度  |
| 年齢 = 70 歳以上  | ⇒ 出向多頻度  |
| 年齢 = 無回答     | ⇒ 出向少頻度  |
| 予測的中率        | 64.795 % |

図 4.2.6 1R アルゴリズムの適用結果 (博多大丸)

|                 |         |
|-----------------|---------|
| 性別 :            |         |
| 性別 = 男性         | ⇒ 出向少頻度 |
| 性別 = 女性         | ⇒ 出向多頻度 |
| 性別 = 無回答        | ⇒ 出向少頻度 |
| 予測的中率 63.6282 % |         |

図 4.2.7 1R アルゴリズムの適用結果 (福岡三越)

|                 |         |
|-----------------|---------|
| 年齢 :            |         |
| 年齢 = 16~19 歳    | ⇒ 出向多頻度 |
| 年齢 = 20~24 歳    | ⇒ 出向多頻度 |
| 年齢 = 25~29 歳    | ⇒ 出向多頻度 |
| 年齢 = 30~34 歳    | ⇒ 出向多頻度 |
| 年齢 = 35~39 歳    | ⇒ 出向多頻度 |
| 年齢 = 40 歳代      | ⇒ 出向多頻度 |
| 年齢 = 50 歳代      | ⇒ 出向多頻度 |
| 年齢 = 60 歳代      | ⇒ 出向少頻度 |
| 年齢 = 70 歳以上     | ⇒ 出向少頻度 |
| 年齢 = 無回答        | ⇒ 出向少頻度 |
| 予測的中率 75.6684 % |         |

図 4.2.8 1R アルゴリズムの適用結果 (ソラリアプラザ)

業が選ばれている。1R ではルールは、選択された属性の属性値の数だけ抽出される<sup>7</sup>。抽出されたルールは、全部で 16 であった。ラーメンスタジアムに行ったことがあるというルールは、次の 4 つであった。

- IF 短大・大学生 THEN 出向経験あり
- IF 専門学校・浪人生 THEN 出向経験あり
- IF 販売・サービス系勤め人 THEN 出向経験あり
- IF その他 THEN 出向経験あり

キャナルシティ博多とショッピングダイエー福岡は、ともに時間距離が選択されているが、抽出されたルールは、キャナルシティ博多が「時間距離<122.5 ならば出向多頻度」「時間距離>=122.5 ならば出向少頻度」であるのに対し、ショッピングダイエー福岡は、「時間距離<5.5 ならば出向多頻度」「時間距離>=5.5 ならば出向少頻度」と大きく異なっており、分析結果は、大きな商圈をもつ複合商業施設と、近場からの来店者をメインとするスーパーマーケットの違いを明確に示しているといえよう。百貨店は、4 店舗のうち 3 つの店舗で年齢が選択されている。岩田屋が 40 歳代以上、岩田屋 Z サイドが 40 歳代未満の来店者を集め、博多大丸が全ての年齢層の来店者を集客しているというルールがきれいに抽出されている。また、福岡三越は、「男性ならば出向少頻度」「女性ならば出向多頻度」というルールが、ソラリアプラザでは、50 歳代以下の全年齢層の来店者を集客しているというルールが抽出された。

以上より、主力商品が生鮮食品や日用品であるスーパーでは近場の来店者をひきつけているといえるとともに、主力商品がファッションアイテムである百貨店や専門店では年齢層がきいており、商品構成が来店者特性ルールを決めている原因となっていることがうかがえる。エンタテインメント性を売りにする複合商業施設では、遠方の人まで幅広く集客している傾向がわかる。

## 5. 結論と今後の課題

本研究では、都市型アミューズメント施設の集客力を決める来訪者の特性評価の問題を取り上げ、これを、C4.5 アルゴリズム、1R アルゴリズムという 2 つのアルゴリズムを用いた施設の来訪者・来店者特

<sup>7</sup> 1R では、欠損値も 1 つのカテゴリーとして取り扱う。ここで職業=無回答がそれに対応する。

性ルールの抽出によって行った。

都市型アミューズメント施設の例として、福岡市のキャナルシティラーメンスタジアムを挙げ、C4.5アルゴリズムを適用した結果、

- (1) 年齢、性別、職業といった特性がラーメンスタジアムの出向に強く影響を与えていることが明らかになるとともに、
- (2) 推定誤り率が低いルールを抽出すると、出向経験有りというルールの条件属性の組合せとして、
  - (a) 若年層の男性(若年層 & 男性)、若年層の女子学生(若年層 & 女性 & 学生)、若年層の独身OL(若年層 & 女性 & 販売・サービス)、若年層の主婦(若年層 & 主婦)
  - (b) 老年層かつ近場に住む人(70代以上 & 時間距離が16分以下)

が選ばれた。また、

- (c) 働き盛りである30代、40代の属性を含むルールは全く得られていない。

(2)(a)(c)は、直観にあうルール、事実であるが、(2)(b)は、分析の前には想定していなかった意外な知見である。事前に立てた仮説にもとづき分析を行う通常のパラメトリックな統計手法と異なり、決定木分析では、目的属性との関係から分割テストをクリアする条件属性をひとつ選び、それによってリレーションを分割するという作業を逐次すすめていく。(2)(a)(c)もそうであるが、特に(b)のような意外な知見は、このように細分割されたリレーションの部分集合を一つ一つ探索的に検索しルールを見出していくという作業をとおしてはじめて得られるものであり、貴重な知見であるといえよう。

また、複合商業施設、スーパーマーケット、百貨店、専門店等、主要商業施設の出向頻度データに、1Rアルゴリズムを適用し業態別来訪者・来店者特性ルール抽出を行った結果、

- (1) 複合商業施設、スーパーの出向には距離が影響を与えており、複合商業施設が広域の商圈から、スーパーが近場から集客していること、
- (2) 百貨店、専門店の来店者特性ルールは年齢が決められていること、
- (3) 都市型アミューズメント施設の来店者特性ルールは職業が決められていること、

が明らかとなった。

また、複数の店舗で分析を行った百貨店では、4店舗中3店舗が年齢に関する来店者特性ルールが抽出されており、そこでの各々のルールも、店舗開発のコンセプトと合致するなど、直観にあったものとなっている。これは1Rアルゴリズムの適用が妥当であることを示唆する結果であるといえよう。

近年、消費者や来街者の意識や嗜好の多様化により、商業地間、都市間競争が激しくなってきた。このような状況のもとでまちづくりを考えていくためには、従来の、どこに、どれだけの規模の施設を整備、開発していくかといった観点からばかりではなく、来街者の意識や嗜好を把握した上で、そのニーズを反映した内容の施設を開発し来店者を魅きつけていくといった、きめ細かい計画を練っていくことが課題となってきた。本研究は、この課題の解決に向けた第1歩として、One to oneマーケティングの考え方を適用し、最近の店舗開発や都心商業政策の鍵となっているエンタテイメント機能に着目した、来訪者・来店者セグメンテーションの事例を示したものである。

今後の課題としては、(1) 今回は、都市型アミューズメント施設等の商業施設を対象としたが、オープンカフェやセレクトショップなどの新しい業態や都心商業地のデータを対象とした分析を行うこと、(2) これらの知見にもとづきエンタテイメント機能と集客メカニズムの関係を記述する基礎的な枠組みを構築すること、(3) ターゲティングに結びつける枠組みの構築、が挙げられる。また、(4) エンタテイメント機能による満足度の測定や(5) リピーターの決定要因分析なども今後の課題としたい。

## 参 考 文 献

- [ 1 ] Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.J., *Classification and Regression Trees*, Chapman & Hall/CRC, 1984.
- [ 2 ] Holte, R.C., "Very simple classification rules perform well on commonly used datasets," *Machine Learning*, 11, 1993, pp. 63-90.
- [ 3 ] Quinlan, J. Ross., "Induction of decision trees," *Machine Learning*, 1, 1986, pp. 81-106.
- [ 4 ] Quinlan, J. Ross., *C4.5 : Programs for Machine Learning*, Morgan Kaufmann, 1995 (古川康一監訳『AIによるデータ解析』トッパン, 1995年)
- [ 5 ] Witten, I.H. and Frank, E., *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufman, 2000.
- [ 6 ] 斎藤参郎・梶井昌邦・中嶋貴昭, "来街地ベースサンプリングによる都心商業地への入込み者数予測モデルの構築と評価", 『地域学研究』 Vol. 29, No. 1, 1999年.
- [ 7 ] キャナルシティ博多公式ホームページ  
[http://www.canalcity.co.jp/ramenstadium/topics\\_back.html](http://www.canalcity.co.jp/ramenstadium/topics_back.html)

## Discovering Rules to Extract Critical Attribute of Frequent Shoppers at Urban Commercial Complexes Using Decision Trees

Masakuni KAKOI\* and Saburo SAITO\*\*

In recent years, many retail establishments in Japan have attached amusement functions along with selling goods such as movie theaters, game centers, and themed food courts. A reason for these trends is a drastic diversification of consumer's attitudes and preferences. In order to acquire new and more loyal customers, shops must grasp diversified consumer demands and provide for their customers the unique characteristics other stores cannot offer such like amusement facilities. Similarly, the functions of open cafes, events, street performers, and music performances in an open urban space also have been reevaluated because of their important roles for attractiveness of city. Exploring what kinds of consumers are attracted by what kinds of amusement functions is worth investigating further in urban studies.

For the purpose, we address the problem to extract critical attributes of frequent visitors for different kinds of urban commercial complexes for investigating the roles of amusement functions to attract customers. First, we apply C4.5 algorithm to visit frequency data for a themed noodle food court set up in a commercial complex for exploring what kinds of customers are attracted by the amusement function of the themed noodle court. The C4.5 is the representative algorithm of decision trees, which are classification algorithms that serve to discover significant relationships between explanatory attributes and a given predicted class in the dataset. We adopt customer's profiles as explanatory attributes and frequent shopper as a predicted class. Thus, using C4.5, we extract the rules to explain critical profiles of frequent shopper for this noodle court. Furthermore, we apply 1R algorithm, which produces one-level decision tree, to visit frequency data for different kinds of 8 urban commercial complexes for exploring what kinds of profiles characterize frequent shoppers for these complexes.

---

\* Associate Professor, Fukuoka University

\*\* Professor, Fukuoka University ; Director, Fukuoka University Institute of Quantitative Behavioral Informatics for City and Space Economy (FQBIC)