

# データマイニング手法を用いた都市公共交通のマーケティングに関する基礎的分析

## Marketing on Urban Public Transport by Data Mining Method

岡村敏之\*\*・藤原章正\*\*・張 峻屹\*\*

By Toshiyuki OKAMURA\*\*・Akimasa FUJIWARA\*\*・Junyi ZHANG\*\*

### 1. 背景および目的

小売業界では、ポイントカード等により獲得した大量の顧客の購買履歴データを用いて個々の顧客の特性を把握することにより、優良顧客(ヘビーユーザー)のさらなる囲い込み、店内陳列レイアウトの設計、販促キャンペーンの検討・実施およびその検証などを行っている。POS(Point of Sales)システム等による「商品管理」(何がどれだけ売れたかを把握し、在庫を最適化し売れ筋商品を絞る)は既に一時代前のものとなり、大量の顧客データを用いた「データベースマーケティング」により、いかに「顧客管理」(誰が何をどれだけ買ったかを把握して、顧客特性別の商品構成と販売戦略を構築する)を行うかが業界競争を勝ち抜く鍵となっている。商品管理に対して顧客管理の考え方の根本的な違いは、「『平均的な消費者』は存在しない」ことを明確に意識していることにある。

翻って交通の世界では、大量顧客データの活用には航空業界が非常に積極的である。しかし都市(公共)交通においては、データは既に存在している(自動改札記録等)がその活用としては小売業での「商品管理」(乗降人員の把握等の定点観測的なデータの利用)の域を出ておらず、個人の行動履歴に着目したマーケティングを行っている事例はまだ少ないと言える。

本研究では、都市公共交通におけるデータベースマーケティングの可能性を明らかにするための基礎的な分析として、広島都市圏で導入されている共通プリペイドカード(磁気式)により蓄積された乗車記録データ(以下「カードデータ」とよぶ)を用いて、記録されたカード発行番号をもとに各利用者(カードホルダー)の一定期間の乗車履歴を抽出することにより、

様々な利用者グループのそれぞれの特徴を把握することを目的とする。この際、データマイニング手法で用いられる決定木(Decision Tree)を構築することにより、マーケティングに有効な利用者類型化のためのルールの記述と類型化を行う。

### 2. 本研究で用いるデータと加工方法

カードデータには、各利用者の各乗車記録(カード発行番号、乗・降した日時/場所、路線系統、運賃等)が蓄積されている。広島都市圏ではJR以外の公共交通機関でカードシステムが共通化されているため、1枚のカードを特定の1人が利用しているとすれば、同一カード番号の利用者(カードホルダー)を一個人とみなし、公共交通利用の全トリップを時系列に追うことが可能である。本研究では、2000年10月の4週間分の全乗車記録から、カード発行番号ベースで2%を抽出(カードホルダー数8735)して分析を行う。

決定木の構築には、Kass(1980)により提案されたBiggs(1991)により改良されたExhaustive CHAID(Chisquare Automatic Interaction Detection)手法を用いる。CHAIDとはカイ二乗検定によって目的変数と関連の強い要因(説明変数)を自動的に選択して決定木を分岐させていく手法であり、Exhaustive CHAIDではそのプロセスを段階的ではなく全数探索を行うもので、任意の分岐停止基準の下で最善の分岐が保証される。変数には質的変数/連続変数とも用いることが可能である。

ここでは基礎分析として、「1ヶ月間の公共交通の総利用頻度(トリップベース)」を目的変数として、表-1に示す変数を用いて、総利用頻度を説明するための決定木の分岐ルールと、決定木により分類されたグループ(ノードとよぶ)を分析する。変数には、各個人について1ヶ月間の履歴を集計したものをを用いる。時間帯はここでは平日/週末別に3つに分割した。

\* キーワーズ: データマイニング, 公共交通計画, 公共交通需要

\*\*正員、博(工) 広島大学大学院国際協力研究科開発科学専攻  
739-8529 東広島市鏡山1-5-1、TEL/ FAX 0824-24-6922  
tokamura@hiroshima-u.ac.jp

表-1 分析に用いた変数

1ヶ月間の総利用頻度(トリップベース, 以下同様)
1ヶ月間の下記の6時間帯/曜日別の利用頻度 3時間帯別(9時までに乗車 9時から18時までに乗車 18時以降乗車)かつ 平日/週末別
1トリップあたり平均乗車時間
1トリップあたり平均運賃(平日/休日別)

データの加工に際しては,1時間以内に乗り継ぎを行った場合は,前後の乗車をあわせて1回の乗車(「トリップ」)として頻度を算出した。また,カードの保有期間が1ヶ月以下の場合には,カードホルダーごとに対象データ期間内におけるカード保有期間(使用開始日と終了日,または度数が0になった日から把握)を算出することで,1ヶ月当りの利用頻度に換算した。

### 3. 総利用頻度を目的変数とした決定木と分岐ルール

Exhaustive CHAID による決定木を図-4に,決定木の分岐ルール(変数とその範囲)と分類された各ノードの基本特性を表-1に示す(次ページ)。本分析では分岐停止基準を第3層分岐までとした。図-2において,各枝の番号はノード番号であり,表-2のノード番号と対応している。

ここで第一層において「平日9時までの利用頻度」によって4枝に分岐する(ノード1~4)。例えば,ノード52についてみると,まず始めに,目的変数と関連の強い要因として,第1層において「平日9時までの利用頻度」が6.6回以上の利用者がノード4に分類され,さらに第2層において「週末9時から18時の利用頻度」が1.5回以下の利用者がノード23に分類される。そして最終的にノード52に分類されるのは,その中で「平日18時以降の利用頻度」が6.6回以上の利用者である。そして表-2より,ノード52に分類された利用者は全体の3.4%にあたり,1ヶ月あたりの平均利用頻度が33.2回であることがわかる。この分析では,分岐ルールとなった変数は全て「利用頻度」に関するもののみとなった。

図-1に,全サンプルの平均利用頻度10.4(回/月)以上のホルダーを高頻度利用層(それ以外を低頻度利用層),全利用回数のうち午前9時以前に乗車している回数が30%以上のホルダーをピーク層(それ以外

をオフピーク層)として,それぞれが総ホルダー数および総利用回数に占める割合を示す。

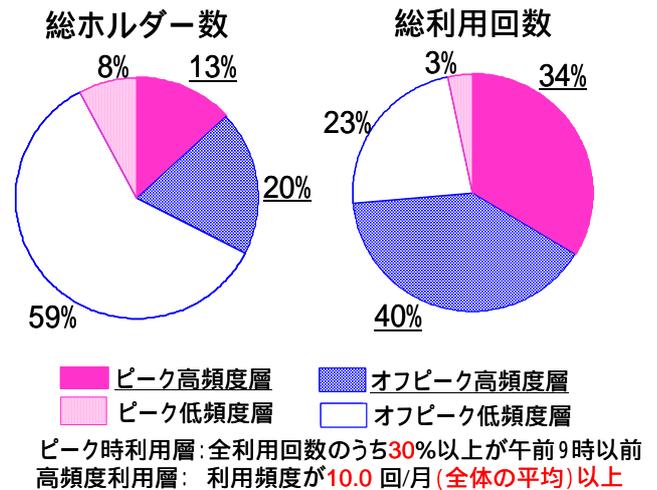


図-1 低頻度利用者,高頻度利用者の全体に占める割合

図-1より,高頻度利用者は総ホルダー数のうち33%しか占めていないにもかかわらず,利用回数では74%を占めていることがわかる。またオフピーク時を主体とする利用者の割合(ホルダー数および利用回数)が少なくないことも分かる(それぞれ20%,40%)。マーケティングの世界で「全体の2割のヘビーユーザーが8割の収入をもたらす」という通説がある。都市公共交通においても(ここでは超ヘビーユーザーである定期券利用者は本分析の対象外であるが)この通説を裏付けるものとなっている。

### 4. 類型化されたノードの特徴

以下では,図-1で示した高頻度利用者について各ノードの特徴を示す。図-3,図-4に各ノードの時間帯別利用割合を総利用頻度の高いものから順に示す。

#### 1) 主にピーク時に利用するノード

図-3に,ピーク時(平日午前9時までの)の利用頻度が総利用頻度の30%以上であるノードを示す。これらのノードはピーク時間帯の利用割合が高く,主に通勤,通学で利用していると考えられる。

この中でも,ノード50では夕方18時以降の利用がないことから通学者と考えられる。ノード52,49においては,昼間の利用が少なく18時以降の利用が多いことから通勤利用者であると考えられる。また,週末利用の割合の高いノード(25)や,平日の通勤・通学のみを利用し週末はほとんど利用しないノード

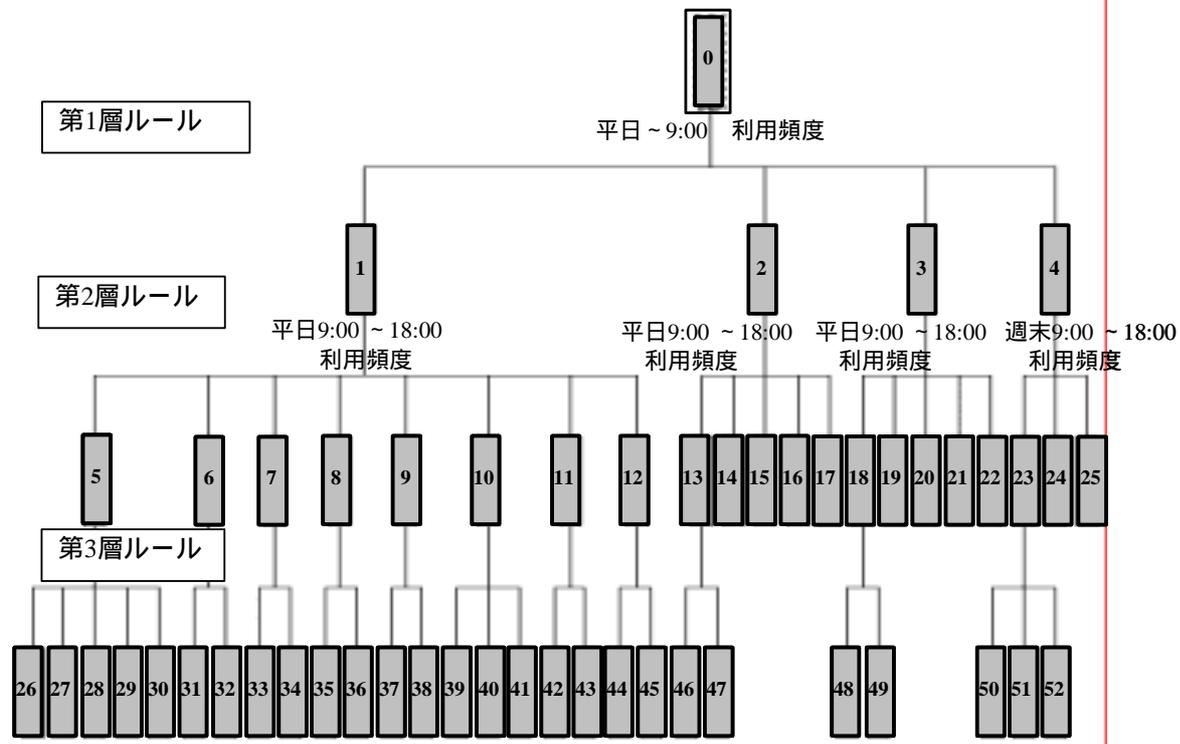


図-2 Exhaustive CHAID によるカード利用者の決定木 (各番号はノード番号)

表-2 Exhaustive CHAID によるカード利用者の決定木の分岐ルール

ノード番号0 (サンプル全体) 平均利用頻度(終日) 10.4 回/月, 分散 12.3, 総カードホルダー数 8735人											
ノード番号1 第1層ルール: 平日9時までの利用頻度 [0 回/月] 平均利用頻度(終日) 7.4回/月 分散9.6 ホルダー数 5982人 全体に占める割合 68.5%					ノード番号2 第1層ルール: 平日9時までの利用頻度 (0, 1.9] 回/月 平均利用頻度(終日) 7.1回/月 分散 8.1 ホルダー数 925人 全体に占める割合 10.6%						
ノード番号	第2層ルール	平均利用頻度(回/月)	分散	ホルダー数(人)	割合	ノード番号	第3層ルール	平均利用頻度(回/月)	分散	ホルダー数(人)	割合
5	平日9時～18時頻度 [0]回/月	3.6	6.0	1704	19.5%	26	週末9～18頻度 [0]回/月	2.6	4.3	615	7.0%
						27	週末9～18頻度 (0, 1.5]回/月	2.0	2.1	492	5.6%
						28	週末9～18頻度 (1.5, 2.3]回/月	3.2	3.4	275	3.2%
						29	週末9～18頻度 (2.3, 4.0]回/月	4.8	3.2	194	2.2%
						30	週末9～18頻度 (> 4.0]回/月	14.1	14.4	128	1.5%
						31	週末9～18頻度 [0]回/月	1.4	1.1	667	7.6%
6	平日9時～18時頻度 (0, 1]回/月	1.9	1.8	818	9.4%	32	週末9～18頻度 (> 0]回/月	4.0	2.4	151	1.7%
						33	平日18～頻度 [ < = 1.2]回/月	3.1	1.8	167	1.9%
7	平日9時～18時頻度 (1, 1.9]回/月	4.5	3.7	282	3.2%	34	平日18～頻度 [ > 1.2]回/月	6.5	4.7	115	1.3%
						35	平日18～頻度 [ < = 1.2]回/月	3.0	1.8	688	7.9%
8	平日9時～18時頻度 (1.9, 2.8]回/月	3.6	2.6	789	9.0%	36	平日18～頻度 [ > 1.2]回/月	7.2	4.2	101	1.2%
						37	平日18～頻度 [ < = 1.2]回/月	4.7	2.3	591	6.8%
9	平日9時～18時頻度 (2.8, 4.2]回/月	5.5	3.5	699	8.0%	38	平日18～頻度 [ > 1.2]回/月	9.9	5.3	108	1.2%
						39	週末9～18頻度 [0]回/月	6.7	2.9	279	3.2%
10	平日9時～18時頻度 (4.2, 6.6]回/月	8.2	4.0	523	6.0%	40	週末9～18頻度 (0, 2.3]回/月	8.0	2.6	132	1.5%
						41	週末9～18頻度 (> 2.3]回/月	12.4	4.6	112	1.3%
						42	平日18～頻度 [ < = 2.6]回/月	11.7	3.6	476	5.5%
						43	平日18～頻度 [ > 2.6]回/月	21.5	7.3	144	1.7%
11	平日9時～18時頻度 (6.6, 12.3]回/月	14.0	6.3	620	7.1%	44	週末9～18頻度 [ < = 4.0]回/月	23.7	10.0	307	3.5%
						45	週末9～18頻度 [ > 4.0]回/月	35.7	12.5	240	2.8%
12	平日9時～18時頻度 (> 12.3]回/月	29.0	12.7	547	6.3%						
ノード番号3 第1層ルール: 平日9時までの利用頻度 (1.9, 6.6] 回/月 平均利用頻度(終日) 12.7回/月 分散10.6 ホルダー数 948人 全体に占める割合 10.9%											
ノード番号	第2層ルール	平均利用頻度(回/月)	分散	ホルダー数(人)	割合	ノード番号	第3層ルール	平均利用頻度(回/月)	分散	ホルダー数(人)	割合
18	平日9時～18時頻度 [ < = 1]回/月	6.7	4.5	379	4.3%	48	平日18～頻度 [ < = 2.6]回/月	4.7	2.1	259	3.0%
						49	平日18～頻度 [ > 2.6]回/月	11.2	5.1	120	1.4%
19	平日9時～18時頻度 (1, 2.8]回/月	8.3	3.9	138	1.6%						
20	平日9時～18時頻度 (2.8, 4.2]回/月	11.0	5.5	131	1.5%						
21	平日9時～18時頻度 (4.2, 12.3]回/月	18.0	8.4	197	2.3%						
22	平日9時～18時頻度 (> 12.3]回/月	1.2	13.6	103	1.2%						
ノード番号4 第1層ルール: 平日9時までの利用頻度 (> 6.6] 回/月 平均利用頻度(終日) 31.5回/月 分散12.9 ホルダー数 880人 全体に占める割合 10.1%											
ノード番号	第2層ルール	平均利用頻度(回/月)	分散	ホルダー数(人)	割合	番号	第3層ルール	平均	分散	ホルダー数(人)	割合
23	週末9時～18時頻度 [ < = 1.5]回/月	28.1	11.6	569	6.5%	50	平日18～頻度 [0]回/月	24.0	12.3	127	1.5%
						51	平日18～頻度 (0.6, 6.6]回/月	21.7	8.7	149	1.7%
						52	平日18～頻度 (> 6.6]回/月	33.2	10.1	293	3.4%
24	週末9時～18時頻度 (1.5, 4.0]回/月	31.9	9.5	152	1.7%						
25	週末9時～18時頻度 (> 4.0]回/月	42.9	13.7	159	1.8%						

(51) など、週末の利用にも違いが現れている。また、利用頻度では、1ヶ月あたり 43.3 回(約 22 往復つまり週 5 回の割合)利用しているノード(25)や、1ヶ月あたり 11 回(約 5 往復つまり週 1~2 回の割合)利用しているノード(20)など、主に通勤・通学目的でも、利用頻度には様々な形態が存在する。

- ・パートなどで定期的に活動している利用者
- ・営業などで平日昼間のみカードを使用する利用者などとも考えられる。週末の利用割合が非常に高く平日は定期券や車などを利用していると考えられるノード(30)、や平日の昼間(9時から18時)と週末のみ利用するノード(42)のように、買い物目的のみで共通カードを利用していると推測されるノードも確認される。

### 5. まとめ

都市公共交通の1ヶ月間の利用者の乗車履歴データを用いて、データマイニング手法の一つである決定木(Exhaustive CHAID)により、利用者の類型化と類型化のためのルールの記述を試みた。本分析では、基礎的な分類を行うために、目的変数を「総利用頻度」とし、説明変数も「平日休日別/時間帯別の利用頻度」と限られたものであったが、いくつかの特徴的な行動特性を持つグループが存在することを把握し、都市公共交通でもデータベースマーケティングの潜在性が高いことが示された。

- さらには、
- ・説明変数に乗降地点やLOS等を加える。
  - ・特定の地域や路線を対象に分析を行う。
  - ・目的変数をより特定のものとする  
(例えば週末の利用頻度を目的変数として、週末利用の多い or 少ない層のルールを記述でき、週末の利用促進施策を示すことが出来る(かもしれない))
  - ・複数年次のデータを用いて、ルール変遷を分析するなど、都市公共交通におけるデータベースマーケティングには、様々な可能性があるかと筆者は考えている。

### 参考文献

- 1) G. V. Kass: An Exploratory Technique for Investigating Large Quantities of Categorical Data, Applied Statistics, No.2, pp. 119-127, 1980.
- 2) SPSS Inc: AnswerTree 3.0 User's Guide, 2001.
- 3) 岡村敏之, 藤原章正, 神野優, 杉恵頼寧: 共通プリペイドカードによる都市圏内公共交通乗車記録の特性分析, 土木計画学研究・論文集 19, pp29-36, 2002.9

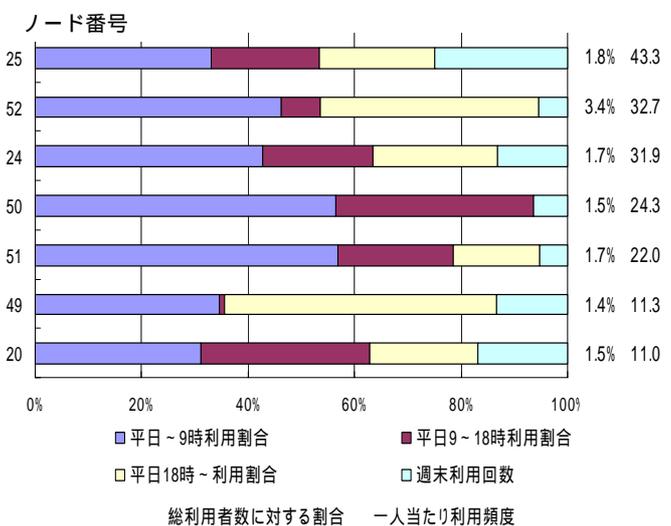


図-3 主にピーク時に利用するノードの時間帯別利用割合

### 2) 主にオフピーク時に利用するノード

図-4 にオフピーク時(平日午前9時以降)の利用頻度が総利用頻度の60%以上を占めるノードを示す。

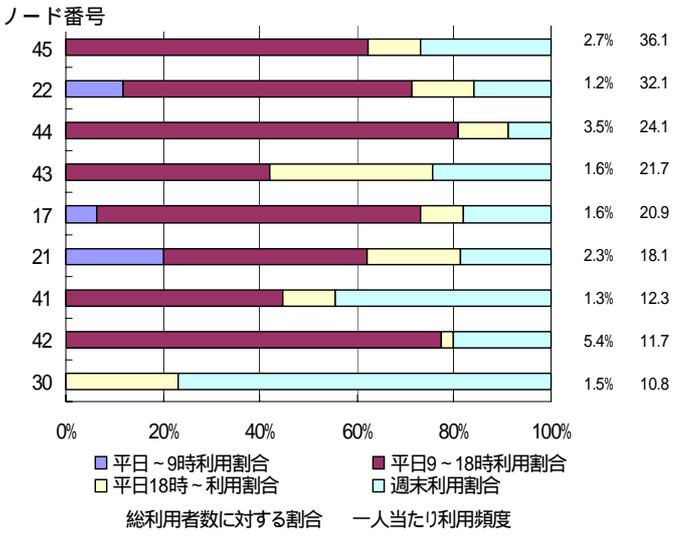


図-4 主にオフピーク時に利用するノードの時間帯別利用割合

ノード 45, 22 では1ヶ月の総利用頻度が30回以上であり、オフピーク時間帯においても非常に利用頻度の高い利用者が存在することが示される。これらは買い物などの目的の他に