

小売店の内側エリアにおける顧客の行動パターンと購買額に関する分析

An analysis of shopping behavior pattern and purchase amount
in the inner areas of the sales floor in a retail store

大畑善裕[†] 大野麻子[‡] 山崎高弘[‡] 常盤欣一郎[‡]
Yoshihiro Ohata Asako Ohno Takahiro Yamasaki Kin-ichiroh Tokiwa

1. はじめに

バーコードが小売業界で用いられるようになり、店舗単位で容易に顧客のPOS (Point of Sales) データを収集することが可能となった。しかし、商品が購入された過程をPOSデータのみを使用した分析により推定することは難しい。そこで、店内における顧客の動きに注目した分析、すなわち、RFID (Radio Frequency Identification) を用いた顧客動線分析が有効である。RFIDとはID情報を埋め込んだタグとリーダー間で無線通信による認証を行う技術のことである。RFIDを用いた動線データの取得は次のようにして行われる。まず、RFIDタグをショッピングカートに取り付け、受信用のリーダーを店舗内の各棚に設置する。カートがリーダー前を通過する度に認証が行われ、タグのID、時刻、位置情報などがデータベースに蓄積される。このようにして得られた位置情報データを時系列的に参照することで店舗内の顧客動線を観測することが可能となる。

一般に、商品を購入する上で最短となる経路をとる顧客に対し、比較的長い経路をとる顧客の方が購買額が高いという傾向がある [1]。また、店内中央の通路を通過することで購買数が増加する可能性があることが示唆されている [2]。このことから、我々は、店内中央の通路およびその付近の売り場、すなわち、内側エリアへの訪問が顧客の購買を説明する重要な役割を果たしていると考えた。

本研究では、内側エリアにおける顧客の行動を定量的に表現し、高額購買顧客 (High-Volume customer: HV) と低額購買顧客 (Low-Volume customer: LV) の分類を行うモデルの構築を目指している。これまでに“内側回遊数”および“内側回遊長”という二つの定量的指標を提案した [3]。提案した二つの指標は、内側エリアにおける顧客の行動特徴を表す統計量であり、顧客動線のもつ時系列的な特徴を考慮した指標ではない。一方、顧客動線から抽出した頻出訪問パターンや各訪問エリアにおける滞在時間といった時系列性を加味した説明変数を併用することで、モデルの分類精度が5%ほど向上したという報告がある [4]。本稿では、エリア訪問順序や共起に基づく新たな指標を用いて顧客動線の時系列的な特徴を定量化する試みについて報告する。これらの指標を用いて高額・低額顧客の分類を行い、本研究で着目する内側エリアにおける顧客の行動パターンと購買額に関する分析を行う。

以下、第2章では、使用するデータの説明と基礎分析を行う。第3章で、顧客の内側エリアにおける行動を定量化する手法の説明を行い、定量化した指標と購

買額の関係について分析を行う。第4章で、時系列性を加味した新たな指標を提案し、それらの指標を使用したモデルの構築を行う。最後に、第5章をまとめとする。

2. 使用するデータと基礎分析

本研究では、2009年5月11日~6月15日の期間中に、関東地方のスーパーマーケットにおいて取得されたPOSデータ(購買データ)および動線データを使用する。店舗内のレイアウトは図1のようになっており、置かれている商品の種類をもとに売り場全体を17種類のエリアに分割している。さらに本研究では、中央通路付近における顧客の行動に着目していることから、売り場を外側エリアと内側エリアに分けている。外側エリアは左下の出入口から売り場の外周を大きく一周する形で、青果(V)、鮮魚(F)、惣菜(G)と続いている。内側エリアには中央通路(H)を挟んで一般食品(B)や菓子(C)などが配置されている。住関(A)は通路を隔てて上下に位置しているため、通路上側をA(上)、下側をA(下)と区別する。ここで、内側エリアに置かれている商品は主に加工食品であり、外側エリアに置かれている商品は生鮮食品である。

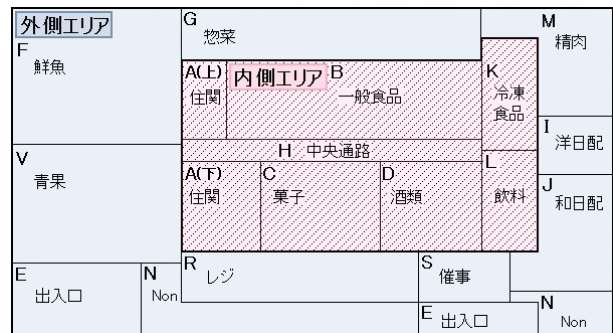


図1: ストアマップ

表1に示すように、RFIDを使用して取得された動線データには、カートのID、リーダー設置個所の通過時刻、店舗内における顧客の位置などが含まれる。各時刻における顧客の位置は2次元座標上のxy座標として表される。本研究では、座標をV、Fなどのエリアを示す英記号に置き換え、記号列として表現した動線データを使用する。購買データはレジで精算を行う際に記録されているため、すべての顧客に存在するが、動線データは買い物カートにRFIDタグを取り付けて観測しているため、カートを使用して購買を行った顧客のみ存在する。本研究では、期間中に記録された6,997件の購買データのうち、動線データと紐づけすること

[†]大阪産業大学大学院工学研究科[‡]大阪産業大学工学部

ができた 5,900 件のデータを使用する。データの集計結果を表 2 に示す。顧客一人あたりの購買額は最大で 21,621 円、最小で 84 円、平均が 3,509 円である。

表 1: 動線データの一部

カート ID	通過時刻	x	y	エリア
5110201	120570	14	10	E
5110201	120584	40	12	V
5110201	120624	28	22	F
5110201	120652	4	28	G
5110201	120656	7	28	B
⋮	⋮	⋮	⋮	⋮

表 2: 購買額 (円)

	全体	HV	LV
最大値	21,621	21,621	904
最小値	84	7,718	84
平均値	3,509	10,197	224
標準偏差	2,281	2,625	692
中央値	3,037	9,360	692
最頻値	3,208	12,411	766

基礎分析として、5,900 件のデータを購買額順に 295 件ずつの 20 グループに分割し、グループ毎にエリア訪問数などを集計することで顧客全体の傾向を調査する。このうち、購買額上位 295 件のグループを HV、下位 295 件のグループを LV とする。図 2 は各グループにおける全エリアと内側エリアの平均訪問数を表したものである。横軸の 1 から 20 は顧客のグループ番号であり、グループ 1 は購買額最下位の LV である。グループ番号が 2, 3 と増加するに従いグループの平均購買額も増加していき、グループ 20 は購買額最上位の HV となる。グラフ上にプロットされている黒丸 (●) と十字 (+) 記号はそれぞれ各グループにおける全エリアまたは内側エリアの平均訪問数を表している。図 2 から、購買額の高い顧客グループの方が全エリア・内側エリアの両方において訪問数が多い傾向にあることが分かる。

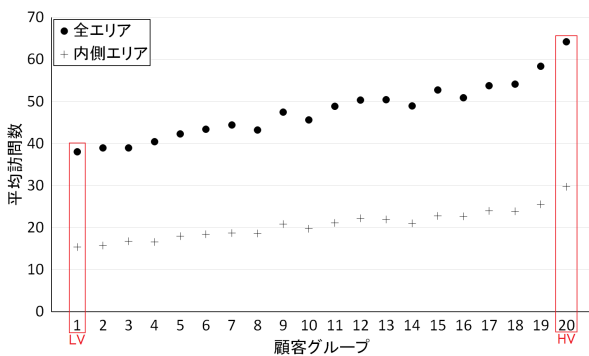


図 2: 各グループの平均エリア訪問数

3. 提案手法

本研究では顧客の内側エリアにおける行動をもとに、高額購買顧客と低額購買顧客の分類を行うモデルの構築を目指している。本章では内側エリアにおける顧客の行動を“内側回遊数”および“内側回遊長”という指標を用いて定量的に表現し、それらの指標を説明変数とする決定木により顧客の分類を行う手法について説明を行う。

3.1. 内側エリアにおける顧客行動の定量化

本研究では RFID により取得された動線データに前処理を施し、顧客動線をエリア訪問順を示す記号列として表している。図 3 に動線データの例を示す。ここで、連続した同エリアへの訪問はマージして 1 回の訪問とみなす。本研究の提案手法では、内側エリアへの 1 回以上の連続した訪問を“内側回遊”と定義している [3]。図中の動線データ上で外側エリアを示す記号列に挟まれた内側エリアの記号列部分が内側回遊である。このときの内側回遊の数を“内側回遊数”，内側回遊の長さを“内側回遊長”と定義する。一動線データ上に複数の内側回遊が存在する場合、最も大きな内側回遊長をその動線の内側回遊長とする。図 3 における動線の内側回遊数は 2、内側回遊長は 3 である。内側回遊数は 1 回の買い物の中でどれだけ内側エリアに出入りしているかの頻度を表し、内側回遊長は一旦内側エリアに入った顧客が外側エリアに出るまでにどれだけ回遊を行っているかという特徴を表す。

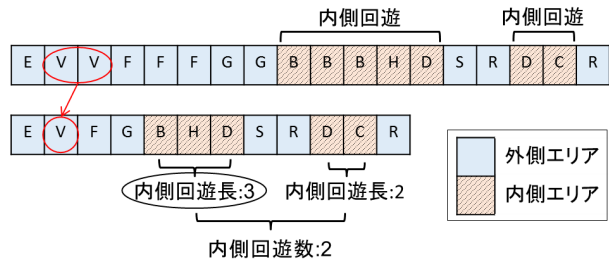


図 3: 内側回遊の定義

内側回遊数と内側回遊長についても前述した 20 グループ毎に平均回数を算出し、全体の傾向を調べる。図 4 に各グループの平均内側回遊数、平均内側回遊長を示す。内側回遊数と内側回遊長についても購買額の高いグループほど値が大きくなっていることがわかる。

3.2. 内側回遊を使用したモデルの構築

ここでは、3.1 で定義した“内側回遊数”および“内側回遊長”を説明変数として、目的変数である顧客 (HV, LV) の分類を行うモデルを構築する。ここで、分類モデルには決定木を使用する。決定木はナイーブベイズやサポートベクタマシンといった他の分類モデルに比べ精度は劣るものの分類結果の解釈が容易であるという利点をもつ。決定木の出力には Weka (Waikato Environment for Knowledge Analysis)[5] を使用する。ここでは、情報エントロピーの概念を用いて決定木の構築や枝刈りを行う C4.5 アルゴリズムの Weka 実装である J48 アル

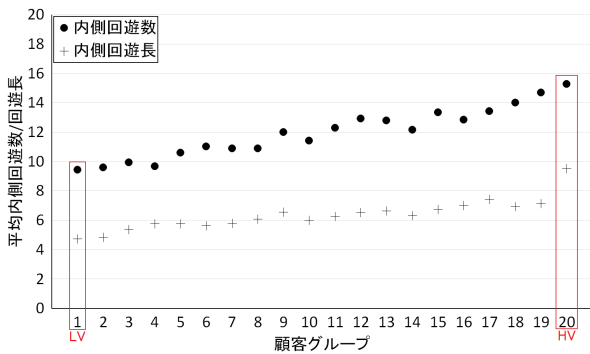


図 4: 平均内側回遊数・回遊長

ゴリズムを使用する。出力した決定木を図 5 に示す。楕円で囲まれた箇所は決定木の分岐を、四角で囲まれた箇所は決定木の葉 (分類の結果) を表している。各葉の括弧内に書かれている左側の数字は全データ中で分岐の条件を満たしたデータの件数を、右側の数字は条件を満たしているが誤って分類されたデータの件数を意味する。この決定木による顧客の分類精度は 78.4% であった。

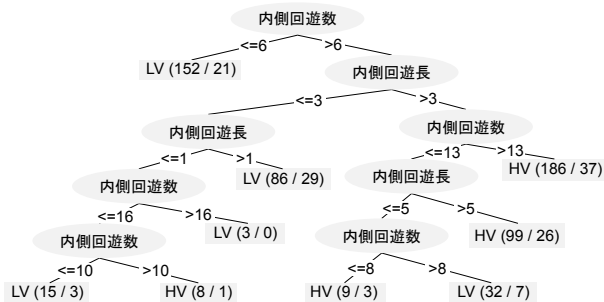


図 5: 内側回遊を使用した決定木

3.3. 内側回遊と購買額の関係

図 6 は HV295 件の内側回遊数と購買額を散布図にプロットしたものである。グラフの横軸は内側回遊数を、縦軸は購買額を示し、両要素には、平均が 0、分散が 1 になるよう正規化を施している。内側回遊数の低いグループに比べ、高いグループの購買額のばらつきが大きいことが確認された。また、図 7 は HV295 件の内側回遊長と購買額を散布図にプロットしたものである。内側回遊長においても同様の傾向が見られた。このため、内側回遊数・内側回遊長といった二つの指標では分類できなかったグループを正しく分類することができるような説明変数を新たに定義する必要がある。

4. 顧客の訪問パターン抽出

これまでに提案した内側回遊数・内側回遊長という二つの指標は、内側エリアにおける顧客の行動特徴を表す統計量であり、顧客動線のもつ時系列的な特徴である訪問順序や、訪問したエリアの種類について考慮した指標ではない。そこで、顧客動線から時系列的な

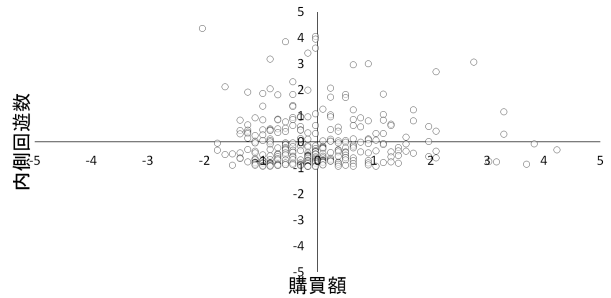


図 6: 内側回遊数と購買額の関係 (HV)

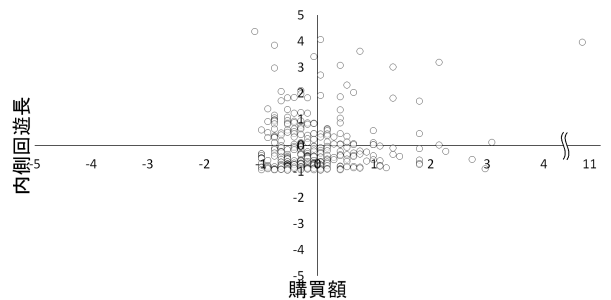


図 7: 内側回遊長と購買額の関係 (HV)

特徴やエリアの種類に関わる特徴を抽出し、顧客分類モデルの精度を高めるための新たな指標を定義するための検討を行う。具体的には、エリア訪問順序や共起に基づく新たな指標を定義し、HV・LV の分類や特徴的な行動パターンの同定を試みる。

4.1. 2 エリア間の移動

動線データは前処理により記号列として表現されており、連続した同エリアの訪問は一つの記号にマージされている。そのため、動線データ上において隣接する二つのエリアは必ず異なるエリアとなっている。図 8 に示すように、動線データの内側回遊部分から 2 エリア間の移動を抽出することで、HV・LV の分類に繋がる特徴的なエリア訪問パターンの発見を目指す。ここでの 2 エリア間の移動は、図 1 の内側エリアにおける A(上) と B、K と L などの隣接した売り場間のみを考える。

内側エリアにおける 2 エリア間の訪問パターンを抽出し、動線に含まれる訪問パターンの出現率 A_i を式 (1) より導出する。

$$A_i = \frac{N_i}{N_{all}} \times 100(\%) \quad (1)$$

ここで、 N_i はある一つの動線に含まれる特定の 2 エリア間の訪問パターンの出現数であり、 N_{all} はその動線に含まれる 2 エリア間の訪問の総数である。図 8 の顧客動線における訪問パターン A → B の出現率 $A_{A \rightarrow B}$ は $(1/7) \times 100 = 14.3(\%)$ となる。訪問パターンの出現率は顧客毎に算出し、HV・LV それぞれのグループに

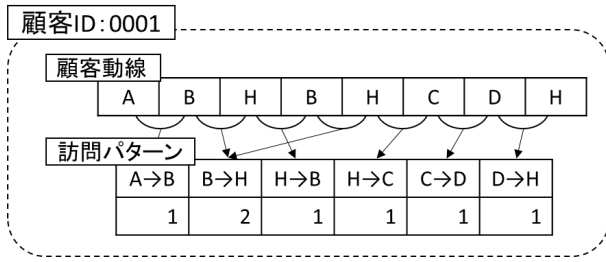


図 8: 2 エリア間の訪問パターン抽出

おける平均をとっている。HV・LV それぞれについて多く見られた訪問パターン上位5件を表3に示す。集計の結果からHV・LV共に中央通路(H)を含む2エリア間の訪問が多く見られた。

表 3: 2 エリア間の訪問パターンと出現率 (%)

順位	HV		LV	
	パターン	出現率	パターン	出現率
1	B → H	4.7	B → H	3.0
2	H → B	4.5	H → B	2.4
3	L → H	1.5	H → C	1.3
4	H → C	1.4	L → H	1.2
5	H → L	1.1	H → D	1.1

また、表3からも分かるように、同一訪問パターンの割合を比較すると多くのパターンにおいてHVの方が出現率が高くなっている。しかし、図9に示すように内側エリア左端のA(下)→C、H→A(下)、内側エリア右端のK→Lの訪問パターンについてはLVの方が出現率が高い。また、図10に示すように、(特定のエリア訪問数/動線に含まれる全訪問エリア数)により算出したエリアCの訪問割合は、HVとLVの差が0.23%であったが、C→HのようにエリアHを絡めた2エリア間のパターンで比較した場合0.41%と上昇している。

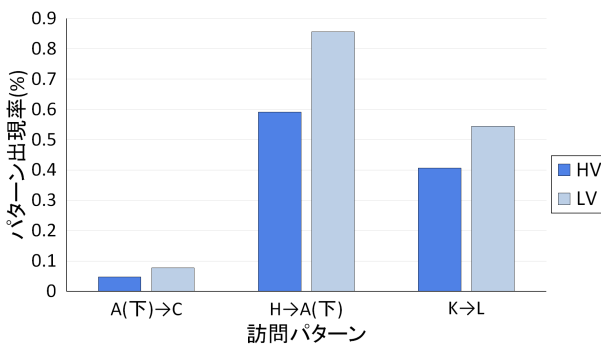


図 9: LVの方が出現率の高い訪問パターン

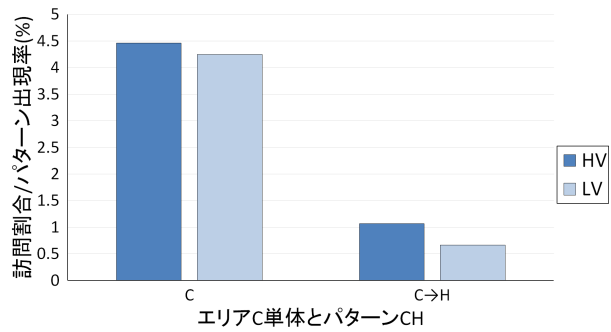


図 10: エリアCの訪問割合とパターンCHの出現率

エリアC単体における訪問割合はHV・LVにおいて同程度であるが、エリアCから他のエリアへの移動を調査すると、C→Hの出現率はHVの方が大きいことが確認された。一方、図11に示すように、エリアCから外側エリアへの移動はLVに多く見られた。ここで、記号oは外側エリアを表している。内側エリアから隣接する外側エリアへ移動する顧客の割合として、LVについてはA(下)やC、Dなどの内側エリア内でも下の方に位置するエリアから外側エリアへ移動しており、HVについてはA(上)やB、Kなどの内側エリア内上部のエリアから外側エリアへ多く移動している。

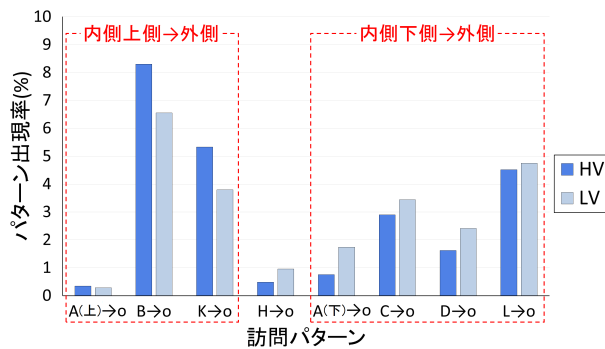


図 11: 内側エリアから外側エリアへの移動

前述したように、基礎分析の結果として、LVに比べHVの方が内側回遊の値が大きかった。図11の結果から読みとれる内側回遊のもう一つの傾向として、内側回遊には大きく分けて“商品を購入するための内側回遊”と“レジへ向かう内側回遊”が存在することが考えられる。すなわち、HVに多く見られたパターンである店内上部の外側エリア(エリアGなど)への移動は引き続き商品を購入するための回遊であり、LVに多く見られたパターンは購買行動を終了し、レジへ向かう内側回遊であると推察される。

4.2. エリア共起に基づく訪問パターンの絞り込み
2エリア間の訪問パターンを調査した結果、HV・LVに共通してB→Hとその逆のH→Bという移動が多

く見られた。そこで、(a) エリア B とエリア H の両方を訪問している場合 (共起あり)、(b) エリア B のみを訪問した場合、(c) エリア H のみを訪問した場合、(d) 両方訪問しなかった場合について考える。表 4 はエリア B とエリア H への訪問の共起について集計した結果である。(c) および (d) の場合は HV と LV を 80% 以上分類することができる。しかし、(a) および (b) の場合には HV と LV が混在している。以降、前述の分類困難な二つの場合において HV と LV を分けることが可能なパターンの発見を試みる。

表 4: エリア B・H に関する訪問傾向の比較

	HV	LV
(a) 共起あり	272	197
(b) エリア B のみ訪問	18	45
(c) エリア H のみ訪問	1	33
(d) 両方なし	4	20

(a) 共起ありの顧客には、 $H \rightarrow B$ を除くと $L \rightarrow H$ や $H \rightarrow C$ などの中央通路を含む移動が多く見られた。図 12 に示すように、エリア H と他の内側エリア間の移動パターンを見たとき、ストアマップ上で中央通路より上に位置するエリアへ移動している割合および、中央通路より下のエリアから中央通路へ移動している割合は HV の方が高く、中央通路より下に位置するエリアへ移動している割合は LV の方が高いことが分かる。

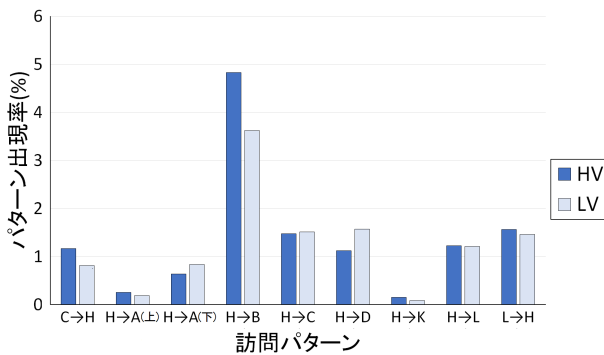


図 12: エリア B・H について共起ありの顧客に見られた訪問パターン

次に、(b) エリア B のみを訪問した顧客について考える。図 13 にエリア B のみを訪問した顧客に多く見られた訪問パターンを示す。エリア B のみに訪問していた顧客は $B \rightarrow K$, $D \rightarrow L$, $K \rightarrow L$ など、内側エリア内の右側に位置するエリアを移動するパターンが多く見られた。中でも $D \rightarrow L$, $L \rightarrow D$ という経路については HV には見られず、LV のみが見られている移動パターンであった。

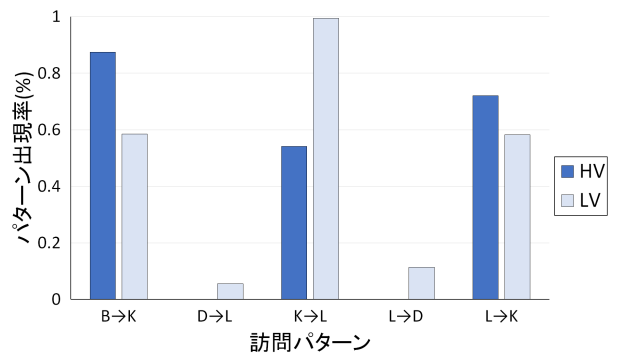


図 13: エリア B のみを訪問した顧客に多く見られた訪問パターン

明変数に 2 エリア間の訪問パターンのみを使用して決定木の構築を行う。ここで、説明変数として使用する 2 エリア間の訪問パターンは、HV と LV で同一のパターン出現率を比較した場合において、その差が大きかった 5 件を使用する。差が大きかった 2 エリア間の訪問パターンは、表 3 や図 9, 10 でも挙げている、 $H \rightarrow B$, $B \rightarrow H$, $C \rightarrow H$, $H \rightarrow A(\text{下})$, $L \rightarrow H$ といったパターンである。構築したモデルを図 14 に示す。この木の分類精度は 69.3% であった。

次に、説明変数に 2 エリア間の訪問パターンと共起を使用して決定木を構築する。ここで使用する共起パターンは 4.2 でも取り上げている、B と H のみを使用する。2 エリア間の訪問パターンは、共起ありという条件下において HV と LV で同一のパターン出現率を比較したとき、その差が大きかった 5 件を使用した。具体的には、 $H \rightarrow B$, $B \rightarrow H$, $H \rightarrow D$, $C \rightarrow H$, $H \rightarrow A(\text{下})$ の 5 つである。前述した 5 件の 2 エリア間の訪問パターンに加え、B と H の共起を説明変数として決定木を構築した。構築したモデルを図 15 に示す。この木の分類精度は 73.7% であり、2 エリア間の訪問パターンのみを使用した決定木に比べて精度が約 4% 向上している。内側回遊数および内側回遊長を使用したモデルには及ばないものの、いずれのモデルについても 70% 以上の精度で分類を行うことができた。

4.3. 2 エリア間の訪問パターンや共起を使用したモデルの構築

4.1, 4.2 で提案した 2 エリア間の訪問パターンと共起を使用して顧客分類の決定木を構築する。まず、説

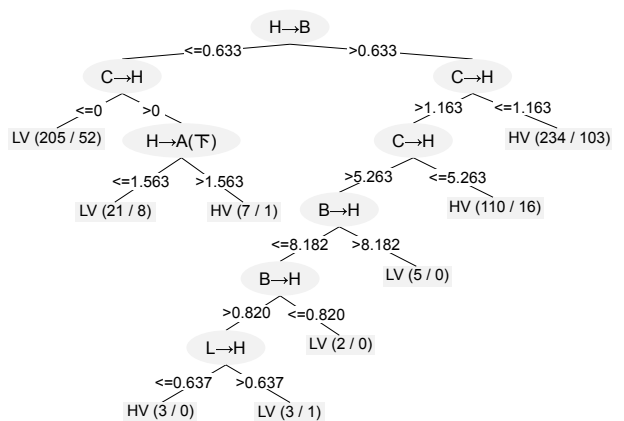


図 14: 2 エリア間の訪問パターンを使用した決定木

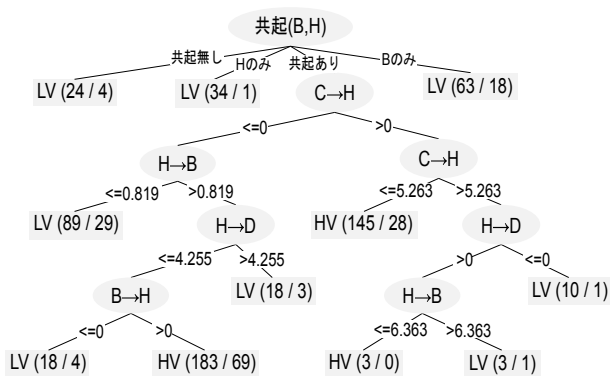


図 15: 2 エリア間の訪問パターンと共起 (B, H) を使用した決定木

内側回遊数および内側回遊長を使用したモデルには及ばないものの、いずれのモデルについても 70%以上の精度で分類を行うことができた。

5. おわりに

本研究では、これまでに内側エリアにおける顧客の行動を内側回遊数や内側回遊長といった指標により定量表現し、これらを説明変数とした決定木による顧客の分類を行ってきた。本稿ではこれらの指標で分類不可能な顧客の行動特徴を定量化するため、時系列性を加味した特徴であるエリア訪問パターンの発見を試みた。結果として、複数のエリア訪問パターンや訪問したエリアの共起情報を組み合わせて使用することで、高額・低額顧客の行動特徴について差異を見出すことができた。また、全体的な傾向として、どの内側エリアにおいても中央通路 (エリア H) へと移動する割合は HV の方が高く、エリア H に移動したあとも HV においては店内上部のエリア (エリア B やエリア G) へと移動する傾向が多く見られた。一方で LV は店内下部のエリア (エリア D やエリア R) へと移動する傾向が多く見られた。これらは HV の買い物を継続する行動と LV のレジへ向かう行動をそれぞれ表していると考えられる。

内側エリアにおける 2 エリア間の訪問パターンや訪問エリアの共起を説明変数として、高額・低額顧客の分類を行う決定木モデルを構築した結果、約 70%の精度で顧客を分類することができた。しかし、2 エリア間の訪問パターンに基づく説明変数としては代表的な 5 件のみ、訪問エリアの共起に基づく説明変数としてはエリア B とエリア H に関する共起情報のみを用いて決定木の構築を行っており、これらの組み合わせが最適であるかという検証は行っていない。今後の課題として、説明変数の組み合わせの最適化を行う必要がある。また、本稿において検討したエリア訪問パターンや共起をもとにした指標とこれまでに提案した内側回遊数・内側回遊長といった指標を説明変数として組み合わせたモデルの構築を行い、分類精度のより高いモデルの構築に向けて引き続き検討を行う必要がある。

謝辞

本研究で使用したデータは関西大学商学部矢田勝俊教授により提供頂いたものであり、ここに記して感謝いたします。

参考文献

- [1] Hui, S. K., Fader, P., Bradlow, E. : The traveling salesman goes shopping: the systematic deviations of grocery paths from TSP-optimality, *Marketing Science*, vol.28, no.3, pp.566–572 (2009).
- [2] 藤野俊樹, 北澤正樹, 高橋雅和, 山田隆志, 山本学, 吉川厚, 寺野隆雄 : 小売店舗内における顧客行動シミュレーションに関する研究, 第 3 回社会システム部会研究会資料, pp.125–128 (2013).
- [3] Ohata, Y., Ohno, A., Yamasaki, T., Tokiwa, K. : An analysis of the effects of customers' migratory behavior in the inner areas of the sales floor in a retail store on their purchase, *Procedia Computer Science*, vol.35, pp.1505–1512 (2014).
- [4] 中原孝信, 矢田勝俊 : 系列パターンマイニングを用いたストーリーミングデータからの特徴抽出, *人工知能学会誌*, 第 27 巻, 第 2 号, pp.146–153 (2012).
- [5] Machine Learning Group at the University of Waikato : Weka 3: Data Mining Software in Java, <http://www.cs.waikato.ac.nz/ml/weka/>, (2014/01/31 参照).