

顧客の店舗利用系列を考慮した将来的な優良顧客の 早期予測手法の提案

○谷地陽奈 後藤裕介 (岩手県立大学)

Prediction of Customers Growth Short-term Customer Sequential Purchase History

* H. Yachi and Y. Goto (Iwate Prefectural University)

概要— 本論文では顧客の店舗利用系列を考慮し、商業施設における複数店舗の利用につながる店舗成長度合いの早期予測における分析手法の提案を行う。顧客の店舗利用成長度基準を作成・クラスターを設定し、決定木分析を行う。この際、クラスターごとに早期段階の利用履歴から部分系列を抽出し特徴量として設定する。分析の結果、店舗利用系列は予測精度向上につながる事が確認でき、より詳細な特徴を明らかにすることができた。

キーワード: データマイニング, 成長予測, 利用系列分析

1 はじめに

従来、流行やトレンドがあり、最新機種やブランド品、トレンドのものを手に入れたいという価値観があったため、商業施設ではトレンドを押さえた多種多様なジャンルのテナントが存在していることが大きな強みであった。また、駅ビルでは立地優位により来店客数も担保されていた。しかし近年、立地や扱う商品数の制限を持たないオンライン市場の拡大もあり、以前より来店客数が減少している傾向にある。加えて、生活者の価値観の変化も言われている。近年では、流行・トレンドを追うこと以上に「モノよりコト」を志向していると言われており、商業施設では今まで以上に利用の仕方に多様性があることが十分に考えられ、商業施設ではその対応に苦戦しているようである¹⁾。

以上から、特定のテナントのブランド力に左右されない顧客、つまり複数の店舗を利用し買い回りをしている顧客は、将来的な継続利用の観点から施設自体へのロイヤルティがあり、優良な顧客であることが考えられる。また、このような優良な顧客になる見込みがある顧客を施設の利用開始の早期段階で発見することで施設側からのアプローチが可能になり、将来的な継続利用を考慮した顧客の育成が可能になる。

優良顧客を分析する代表的な手法としてRFM分析²⁾がある。しかしながら、RFM分析は過去の実績を用いて現時点での分析を行うものであり、将来的な予測につながるような評価をする事はできない。また、顧客の購買行動などの考慮もされておらず、分析に用いられる最新購買日・頻度・購買金額の指標では顧客の人物像の推定がしにくい。そのため、将来性の考慮やより詳細な顧客の推定につながる指標を追加する必要がある。

顧客の購買行動を対象とした先行研究に、顧客の店舗間遷移行動を解析するための手法として確率的潜在意味解析 (pLSA)³⁾が挙げられる。この手法では、商業施設のポイントカードの履歴を対象としており、異なる業種をまたぐ店舗間利用の遷移の分析を行って

る。遷移行動別にクラスタリングを行いそれぞれの特徴を明らかにすることにより、顧客の商業施設内における行動が分析されている。しかし、明らかになった遷移行動の中で、どのような遷移パターンが優良であるかはわかっていない。また、これらは顧客が過去に行った遷移を調べ、一定の人数以上の顧客が経験した店舗間遷移を分析の対象とした頻出パターンを分析したものであり、それぞれの利用時期の考慮はされていない。どのような行動パターンが優良であるかを明らかにし、さらに利用時期の考慮をすることで、優良顧客になりえる見込み顧客発見につながる事が考えられる。以上の関連研究についてまとめたものがTable 1である。

Table 1 : 関連研究まとめ

	RFM 分析	確率的潜在意味解析 (pLSA)
概要	最新購買日, 頻度, 総購買金額の指標を用いて顧客の評価を行う分析手法	顧客の異なる業種をまたぐ店舗間の遷移行動を解析する手法
考慮されていない点	● 顧客の行動 ● 予測	● どの遷移行動が優良であるか ● 利用時期

2 提案手法

2.1 提案手法の概要

提案手法は、顧客の店舗利用系列を考慮し、商業施設における複数店舗の利用につながる店舗利用の成長度合いの高い顧客を想定し、顧客の店舗利用の仕方を調査するものである。このとき提案手法は以下のプロ

セスによって実現される。

- 1) 店舗利用成長度基準の作成: 顧客が各利用日までに1度は利用したことのあるユニークな利用店舗を集計し、ユニークな利用店舗の成長率に基づき k-means 法を用いてクラスタに分割する。
- 2) 店舗利用系列の抽出: 1) で集計したユニークな利用店舗をジャンルに変換し、来店回数5回までの利用履歴をFP-growthアルゴリズムを用いて系列を考慮した店舗利用系列を抽出する。
- 3) 特徴量の選択: 2) で抽出した店舗利用系列から重要度の高い系列を選定するため、アソシエーション分析を行う。
- 4) 特徴量の設定: 3) で選定した店舗利用系列に加え、顧客のデモグラフィック属性、店舗利用頻度、総購買金額を特徴量として決定木分析用データを設定する。
- 5) 決定木分析: 4) で作成したデータを用いて決定木分析を行い、顧客の成長度を予測する。有効性は正解率で各手法を比較し、判断する。
- 6) 顧客の分析: 決定木分析結果の特徴をもとに人物像の推定を行う。

2.2 店舗利用成長度基準の作成

顧客の店舗利用の成長度ごとの特徴理解のためクラスタリングを行い、基準となるクラスタの分割をする。まず、店舗利用の成長度別のクラスタを分割するため、各顧客の各買上日ごとに利用したことのあるユニークな利用店舗の累積値の変動を集計する。次に、その集計データを用いてクラスタを分割し予測・特徴理解のための基準とする。クラスタの分割には、対象が多いことから非階層型クラスタ分析の代表的な手法である k-means法²⁾を用いる。また、k-means法は、事前にクラスタ数の設定を行い要素の集合を集約していく方法であるため、クラスタ数の設定にはエルボー法を用いる。

2.3 店舗利用系列の抽出

顧客がどのような利用の仕方をしているかを指標として特徴を作成するために、店舗のユニークな利用系列を集計する。この際、店舗レベルの系列抽出では件数があまり検出されないことが予測されるため、店舗をジャンルのレベルに変換する。また、早期に予測するため顧客の利用開始日から5回目の利用日までの系列を対象とする。これらの店舗ジャンルのレベルに変換された店舗利用系列データを、データセットの系列を保持した状態で頻出アイテムの集合をデータベースに格納する事ができるFP-growthアルゴリズム⁴⁾を用いて部分系列を抽出する。

2.4 特徴量の選択

FP-growthアルゴリズムで抽出された利用部分系列を全て特徴量として取り込むと、ノイズにつながり精度が落ちることが想定されるため、関連の強い系列のみを抽出し特徴量選択を行う。このフィルターがけとして、support (支持度), confidence (確信度), lift (改善度)の3つの指標を用いて閾値の条件を設定し関

連性を抽出する手法であるアソシエーション分析⁵⁾を用いる。本研究で行うアソシエーション分析ではクラスタと系列の関連の強さを計算するため、以下の計算式を用いる。 X_M を全顧客の集合、 X_i をクラスタ i に所属する顧客の集合、 Y_j を抽出された部分系列 j が履歴に含まれる顧客の集合、 $|X_i \cap Y_j|$ をクラスタ i に所属する顧客のうち部分系列 j が履歴に含まれる顧客の集合、 $|\cdot|$ を集合の要素数とする。このとき系列のconfidence指標、lift指標は以下のように計算する。

- (1) クラスタ i と部分系列 j との confidence 指標

$$confidence(i, j) = \frac{|X_i \cap Y_j|}{|X_i|}$$

- (2) クラスタ i と部分系列 j との lift 指標

$$lift(i, j) = \frac{|X_i \cap Y_j|}{|X_i|} / \frac{|Y_j|}{|X_M|}$$

関連の強い特有のパターンとしては「当てはまればそのクラスタに該当しやすいパターン」と「当てはまればそのクラスタに該当しにくいパターン」とし、それぞれ $lift(i, j) \geq 1.25$, $lift(i, j) \leq 0.75$ とする。また confidence値は抽出数や比率を考慮しながら設定するものとする。

2.5 特徴量の設定

目的変数を成長度基準クラスタとし、説明変数として特徴量選択を行った店舗利用系列に加え、精度向上と人物像の推定につなげるため、顧客の年齢や性別、居住地といったデモグラフィック変数や5回目来店までの総購買金額や来店日数などのFMスコアを取り入れ特徴量を設定する。なお、5回目来店時点でのユニーク利用店舗数は精度向上への寄与が大きいが、本分析においては精度向上のみを目的としておらず、顧客の人物像推定につながりやすいことから、今回は特徴量として利用しないこととした。

2.6 決定木分析

作成したデータを用いて決定木分析⁶⁾を行い、顧客の成長度を予測する。この際、提案手法の有効性を評価するため、比較データを作成し判断する。比較データには、それぞれ説明変数として①デモグラフィック変数のみのデータ、②デモグラフィック変数とFMスコアのデータ、③デモグラフィック変数とFMスコアとFP-growthアルゴリズムで抽出された利用系列の全パターンのデータを用いて、本研究での提案手法である④デモグラフィック変数とFMスコアと特徴量選択を行った利用系列のデータを比較した。

2.7 顧客の分析

決定木分析結果をもとに各クラスタの顧客の人物像、店舗の利用のされ方の推定を行う。まず、主要となりそうな決定木結果ノードの調査を行う。主要ノードとしては分類人数が多いボリューム層と、ジニ不純度が

低いノードが考えられる。これらのノードから成長度がより高いクラスタに分類されるにはどのような条件が必要なのか調査を行う。次に分析において重要度の高い説明変数を抽出し、各クラスタの該当比率など詳細を確認し、店舗の利用のされ方などの推定を行う。

3 分析事例

3.1. 対象商業施設の概要

本研究で使用するデータは、複数かつ多様なテナントを抱える商業施設 A から取得された購買履歴である。

- 期間：2017年02月16日～2020年02月29日
- 件数：8,743,763件
- テナント数：279店舗(一部閉店した店舗を含む)
- 顧客数：234,687人
- テナントの構成ジャンル:お土産、カフェ・軽食、サービス、ファッション雑貨、メンズ、ライフスタイルグッズ、レストラン、レディース、催事場・その他、食品関連の10種類

基礎集計により以下のことがわかった。

- 顧客の1回の来店での利用店舗数は、1店舗のみの利用が54.8%、2店舗の利用は24.8%、3店舗の利用は11.6%、4店舗の利用は5.2%、5店舗以上の利用は3.6%であった。
- 各顧客の1日の店舗利用数について、大きくばらつきがある事がわかった。

3.2. 分析結果

3.2.1. 店舗利用成長度基準の作成

提供データをもとに、顧客のポイントID、買上日、利用店舗を示すショッピングコードを用いて各顧客(ポイントID)が各買上日までに1度は利用したことのあるユニークな利用店舗を集計した。

それらのデータから、十分継続的な利用をしていると考えられる来店日数が50日以上、かつ会員登録日が提供データの対象期間内である顧客を抽出し、分析の対象顧客として設定する(対象顧客：14,868人)。

店舗利用成長度基準の作成に、対象の各顧客の50日までの利用店舗数の推移の変動を集計し、成長度と定義する。これらの集計データを基にk-means法を用いて成長度別のクラスタを設定する。なお、クラスタ数の設定にはFig.1よりエルボー法を用いた結果、3に設定した。また、Fig.2より各クラスタはそれぞれ成長度が高い順から「高成長」、「一般」、「低成長」とする。それぞれのクラスタの人数は、「高成長」：3,024人、「一般」：6,879人、「低成長」：4,965人である。

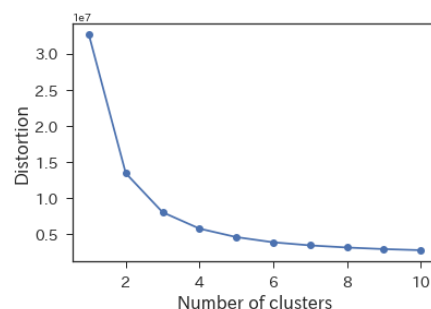


Fig.1：エルボー法による分割数の評価

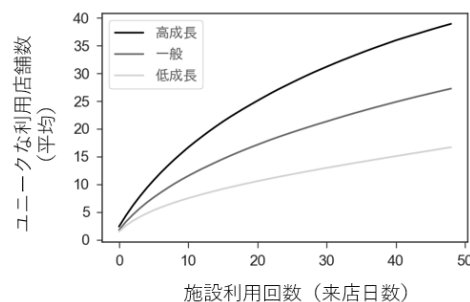


Fig.2：各クラスタの平均利用店舗数

3.2.2. 分析に使用する特徴量の作成

初回来店日から5回来店までに利用したユニークな店舗利用履歴から、それぞれ店舗ジャンルに変換し、各クラスタごとにFP-growthアルゴリズムを用いて部分系列を抽出する。次に抽出された利用系列から実際に各顧客の利用履歴に当てはまる部分系列をダミー変数化し、特徴量として用いる。Table 2より抽出されたパターンは高成長で404件、一般で388件、低成長で229件であり、各クラスタで抽出された全パターンを統合し、重複を省いたユニークなパターン件数は631件であった。

Table 2：FP-growthアルゴリズムにより抽出された部分系列

	抽出パターン数
高成長	404
一般	388
低成長	229
全体	1,021
全体 (ユニーク)	631

3.2.3. 特徴量の選択

Table 3より、 $lift(i, j) \geq 1.25$ を該当しやすいルール、 $lift(i, j) \leq 0.75$ を該当しにくいルールとし抽出する。全体(ユニーク)は、各クラスタで抽出された全パターンを統合し、重複を省いたユニークなパターン件数である。confidenceの値は高いほど信頼度が高く望まし

いが、抽出される系列の件数が少ないと人物像の推定が難しくなるといったジレンマから、今回はエルボー法のような考えのもと、0.005を境に該当パターン数の変化の仕方が大きく変わっており、減少幅が緩やかになる地点であるため、全体のユニークな抽出パターン数631件の約10%にあたる66件のルールを選定した。

Table 3 : confidence 値ごとの該当パターン数

	confidence ≥0.001	confidence ≥0.005	confidence ≥0.01
高成長	135	61	38
一般	108	46	32
低成長	78	33	21
全体	321	140	91
全体 (ユニーク)	163	66	44

3.2.4. 特徴量の作成

次に、Table 4 で示すような、店舗利用成長度基準クラスタを目的変数とし、「店舗利用系列」に加え、「デモグラフィック属性」、「顧客ごとの5回来店までの総購買金額・購買頻度 (FM スコア)」を説明変数としたデータを作成し、決定木分析を行う。

Table 4 : 目的変数と説明変数

目的変数	店舗利用成長度	0 : 高成長, 1 : 一般, 2 : 低成長
説明変数	デモグラフィック属性	性別 男性, 女性, 不明 年代 10代, 20代, 30代, 40代, 50代, 60代, 70代, 80代, 90代, 不明 居住 A 県 B 市, A 県, A 県外
	FM スコア	5回来店までの総購買金額 1~5000, 5001~10000, 10001~15000, 15001~20000, 20001~25000, 25001~30000, 30001~35000, 35001~40000, 40001~45000, 45001~50000, 50001~
	5回来店までの日数	4~9, 10~19, 20~29, 30~39, 40~49, 50~59, 60~69, 70~79, 80~89, 90~99, 100~
	店舗利用系列	全抽出 FP-growth アルゴリズムを用いて抽出された全店舗利用系列 特徴量選択 アソシエーション分析を用いて特徴量選択を行った店舗利用系列

3.2.5. 決定木分析

決定木分析にはそれぞれクラスタを2つ用いて判別を行った。ハイパーパラメータの値は criterion : [gini, entropy], splitter : [best, random], max_depth : [8, 12, 16, 20], min_samples_leaf : [5, 10, 15, 20]の範囲でグリッドサーチによる最適化を行った。

Table 5 より以下の4つの分析結果を用いて本研究手法の有効性を比較し、判断した。「①: デモグラフィック属性のみ」、「②: ①と FM スコア」、「③: ②と店舗利用系列」、「④: ②と特徴量選択を行った店舗利用系列」の正解率を比較したところ、④が1番高いことがわかった。しかし、高成長_低成長の極端なテストデータでは0.78であったのに比べ、高成長_一般や一般_低成長では正解率が0.67, 0.64と低下していた。また、①のデモグラフィック属性のみのテストデータでの正解率と比較した際にも、高成長_一般や一般_低成長では5ポイントの改善にとどまっていた。

次に Table 6 より、「④: ②と特徴量選択を行った店舗利用系列」の決定木分析のコンフュージョンマトリクスと正解率について確認を行う。高成長_低成長では、テストデータの中で1,200人中560人が高成長、540人が低成長に分類された。そのうち、TPに該当する高成長顧客は437人で、全顧客の36.4%であり、TNに該当する低成長顧客は497人で、41.4%であった。

これらのことから、隣り合うクラスタでは FM スコアや利用系列の指標による正解率の向上は難しいことが考えられ、成長度の程度の差によって正解率の向上への寄与の程度が異なることがわかった。

Table 5 : 決定木分析正解率比較

	①	②	③	④
高成長_低成長	0.65	0.74	0.75	0.78
高成長_一般	0.62	0.64	0.65	0.67
一般_低成長	0.59	0.60	0.63	0.64

Table 6 : 正解率評価

	高成長_低成長	高成長_一般	一般_低成長
TP	437	301	376
FN	143	200	217
FP	123	196	211
TN	497	403	396
正解率	0.78	0.67	0.64

3.2.6. 顧客の分析

作成した決定木分析用のデータは説明変数をダミー変数としていたため、結果の見方は各ノードで各要素が該当するなら1、該当しないなら0となる。各決定木分析の結果より、まずクラスタを分類する指標を主要ノードにより調査した。本研究における主要ノードとしては、分類人数が多いボリューム層ノードとジニ不純度が0または低い層ノードを指す。Table 7の例を含め利用系列は(お土産, 食品関連), (カフェ・軽食, 食品関連), (食品関連, 食品関連)の利用が重要度の高い条件であることがわかり、かつ成長度が高いクラスタほどそれらの系列を利用している傾向が確認できた。このように従来得られる情報に加えて、重要となる店舗利用系列の発見が可能となる。

Table 7 : 一般と比較した際の各種用分類ノードの例

		高成長	低成長
特徴	該当	<ul style="list-style-type: none"> ・(お土産, 食品関連) ・(食品関連, 食品関連) ・女性 ・利用額 2.5万~5万円 	<ul style="list-style-type: none"> ・女性 ・利用額 5千円以下
	該当なし	<ul style="list-style-type: none"> ・(お土産, お土産, 食品関連) ・(ライフスタイルグッズ, 食品関連) ・80代 	<ul style="list-style-type: none"> ・(お土産, 食品関連) ・(食品関連, 食品関連) ・(カフェ・軽食, 食品関連) ・(催事場・その他, 食品関連) ・(ライフスタイルグッズ, 食品関連) ・20代, 50代, 70代
該当者		44/4800人	66/4800人

次に重要度の高い説明変数の中から、顧客の推定につながりやすいと考えられる変数の詳細を確認していく。Fig. 3より商業施設Aでは顧客の約8割が女性で、大多数を占めているが、クラスタごとにその比率を見ると高成長のクラスタになるほどさらに女性の比率が高くなっていることがわかる。

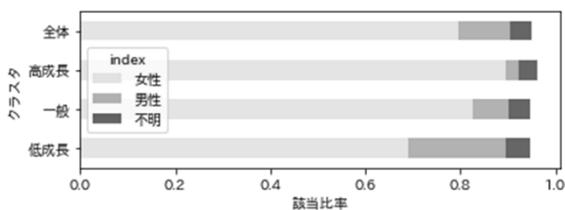


Fig. 3 : 各クラスタの性別比率

次に重要度の高い利用系列の該当比率を確認する。Fig.4は重要度の高い系列の各クラスタの利用割合を集計したものである。各決定木分析における重要度の高い系列は食品関連の店舗が含まれることがわかった。特に(a)の系列は高成長のクラスタにおいて約半数の顧客が利用している系列であった。一番少ない割合である低成長のクラスタにおいても3割を越えており、食品関連の店舗は複数の同ジャンルの店舗と利用され

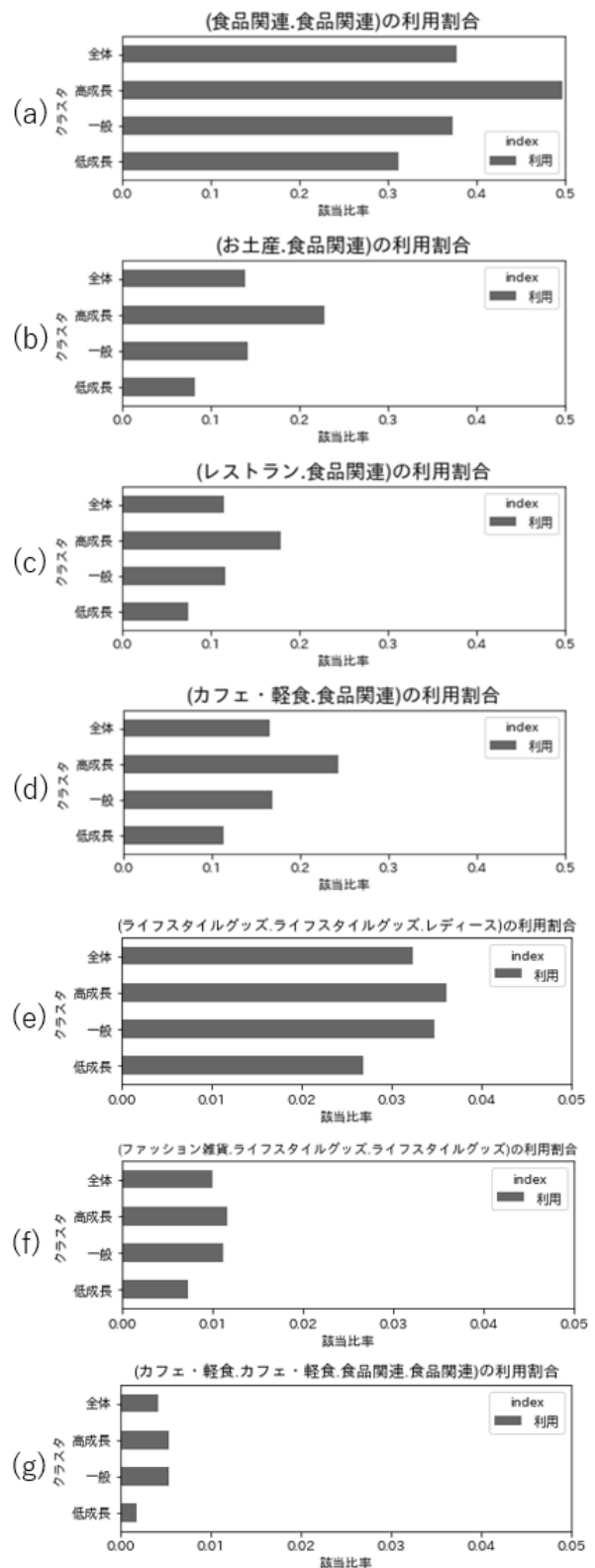


Fig. 4 : 各クラスタの利用割合

やすく、食品関連の店舗を買い回っている可能性が高いことがわかった。

また、(b)より、低成長の利用割合が0.08であったのに対し、高成長の利用割合が0.23であり3倍近く差が出た。(c)では、高成長の利用割合が0.180、一般が0.116、

低成長が 0.075 であり、比較的各クラスタによって利用の割合に差が出た。(d)でも、高成長の利用割合が 0.242、一般が 0.169、低成長が 0.114 で同様に比較的各クラスタによって利用の割合に差が出た。これらのことから、高成長の顧客はレストランやカフェ・軽食の食事関連の店舗やお土産の店舗の利用の後に食品関連の店舗を利用している傾向もあり、食品関連のみなど特定の利用用途に留まっていない顧客が多いことも予想される。

(e)より、高成長の利用割合が 0.036、一般が 0.035 でありあまり差違が見られなかった。(f)より、高成長の利用割合が 0.012、一般が 0.011 であり(e)と同様にあまり差違が見られなかった。これらから、「ファッション雑貨」や「レディース」、「ライフスタイルグッズ」の店舗は成長度合いへの寄与は少ないことが考えられる。

さらに、(g)では高成長の利用割合が 0.0053、一般が 0.0054 とわずかながら一般が高成長の割合を上回った。(カフェ・軽食・食品関連)や(食品関連・食品関連)では割合に差が出ていたことを考慮すると、(カフェ・軽食・カフェ・軽食)のユニークな利用の増え方は少なく、ほとんど成長度合いへの寄与が無いことが考えられる。

一方で低成長のクラスタが高い割合の利用系列は確認されなかった。Fig.5 より、低成長のクラスタはユニークな利用店舗数が少なく 5 回来店の間ではまだ 1 店舗しか利用していない顧客が一定数存在することも十分想定されるため、系列の抽出が少なく、低成長のみ高い割合の系列がなかったと考えられる。

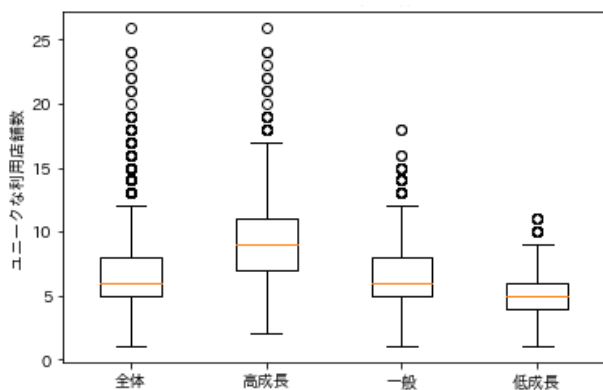


Fig. 5 : 5 回来店時点での利用店舗数

以上より、各クラスタの顧客について以下のことが考えられる。

- 高成長：買い回りがしやすい食品関連の店舗利用を中心に、ユニークな店舗の利用が増えていく可能性が高い。加えて女性の利用者である可能性が高い。
- 一般：買い回りはするが、多少好みの系統の店舗の利用に落ち着く可能性が高い。
- 低成長：限定的な目的の店舗のみに利用が留まっている可能性が高い。加えて男性の利用者である可能性が高い。

利用店舗数の成長における観点から店舗の利用のさ

れ方に関して以下のようなことが考えられる。

- 「食品関連」は同ジャンルの中でも魚屋、八百屋などカテゴリで店舗が分かれているため、買い回りが多いことが予想され、抽出された利用系列の中でも多く組み込まれていた。
- 「カフェ・軽食」や「レストラン」などの食事関連の店舗は、多少好みに偏ることもあるが、比較的多様な店舗を利用しやすい可能性が考えられる。
- 「ファッション雑貨」や「レディース」の店舗は比較的買い回りをせず、好みの系統の店舗の利用に留まることが考えられることから、系列での抽出が食品関連などに比べ少ないことが予想される。
- 「ライフスタイルグッズ」は全体の利用割合が高いとされる主要店舗が含まれているが、成長度合いへの寄与は少ないことがわかった。また、目的購買による利用のされ方が強いと考えられる。

このように、従来得られる顧客のデモグラフィック属性に加えて、介入の際に店舗の利用系列を見つけることが可能となり、顧客の特徴理解や複数店舗を対象とした施策に検討に活用できる可能性がある。

3.2.7. エキスパートレビュー

本研究の提案手法の有用性の評価について、2021年2月16日(火)にデータ提供元の商業施設営業部担当者2名からエキスパートレビューをしていただいた。提案手法による分析結果を、クーポン配布・セールの実策検討の材料や、買い回り分析への応用として利用できそうだと評価をいただいた。

食品関連の傾向についてはある程度想定がされており、以前ファッション雑貨・レディースの店舗で買い上げた顧客に対し、食品関連のクーポンなどを発行していたこともあった。一方で、ライフスタイルグッズ関連の利用について、店舗利用成長への寄与が少ないことが想定外のケースとしてあげられた。これについては、ライフスタイルグッズの利用はどのクラスタでもよく利用されており、各クラスタの特徴を分ける材料にはなり得なかったことが考えられる。さらに、来店トリガーなどの調査のための買い回りパターン分析に応用できる可能性など今後の展開につながりそうといった評価をいただいた。

4. 考察

本研究の分析結果から、既存の顧客分析手法では明らかになっていなかったより詳細な顧客の推定につながる店舗利用の特徴である、買い回りがしやすい食品関連の利用を中心に、ユニークな店舗の利用が増えていき優良な顧客になる可能性が高いといったことを明らかにする事ができた。加えて、店舗の利用のされ方についても顧客の趣味嗜好に左右されやすい店舗のジャンルは成長度合いへの寄与が少ないなど、ある程度推定が可能になった。これらのことから、どのような店舗を利用しているかの系列とユニークな店舗利用の成長の度合いにはある程度関係があることが考えられる。また、分析の精度評価結果では提案手法において

一番高い正解率を得ることができた。

今回の分析では複数のジャンルの店舗のデータを対象に分析を行ったため、結果を得ることができたといえる。しかし、ポイントカードなどで複数の店舗で共通する顧客 ID が存在して購買履歴が得られていれば、商業施設のみでなく、デパートや商店街などでも同様の分析が展開可能であるといえる。

今後の課題として、本研究で用いた説明変数は顧客の店舗のユニークな利用系列と FM スコア、デモグラフィック属性の情報に留まっているが、顧客の店舗の利用時間などは考慮していないため、説明変数として取り入れることで、精度を向上させることが期待される。

5. おわりに

本研究では、将来的な継続利用の見込める施設自体へのロイヤルティがある優良な顧客の育成を行うため、店舗利用系列を考慮した商業施設における複数店舗の利用につながる店舗成長度合いの早期予測を行った。具体的に、まず店舗利用の成長度における基準を作成しクラスタを分割、そのクラスタを目的変数とし、店舗利用系列や顧客の人物像の推定につながりそうなデモグラフィック属性と FM スコアを考慮し、決定木分析を用いて顧客の分析を行った。説明変数の組み合わせを変えたデータを用いる事で、正解率の評価の確認もできた。各決定木分析結果による顧客の分析では、ユニークな店舗利用の成長には食品関連の店舗が成長度合いへの寄与が多く、食品関連の店舗利用を中心に買い回りが発生していく可能性が高いなど、顧客の推定や店舗利用のされ方について明らかにすることができた。

参考文献

- 1) 松本阿礼：商業施設における来店動機に関する研究 一駅ビルを対象として一、プロモーショナル・マーケティング研究, **10**, 40/55 (2017)
- 2) 生田目崇：マーケティングのための統計分析, オーム社 (2017)
- 3) 吉田, 本村, 竹中：複数業種店舗間における顧客の店舗間遷移の分析手法の提案, 日本行動計量学会大会抄録集, **41**, 214/217 (2013)
- 4) C. H. Mooney and J. F. Roddick: Sequential Pattern Mining – Approaches and Algorithms, *ACM Computing Survey*, **45-2**, Article 19 (2013)
- 5) C. Zhang and S. Zhang : Association Rule Mining: Models and Algorithms, *Lecture Notes in Artificial Intelligence*, **2307** (2002)
- 6) J. R. Quinlan: Induction of Decision Trees, *Machine Learning*, **1-1**, 81/106 (1986)