

# 購買履歴と回遊データに基づく小売店舗における商品位置 推定モデルの提案

## Estimating a Goods Position Based on Purchase and Movement Histories in Retail Stores

趙 陽陽<sup>1</sup> 後藤 裕介<sup>2</sup> 森田 裕之<sup>3</sup> 南野 謙一<sup>4</sup> 渡邊 慶和<sup>5</sup>

Yangyang Zhao<sup>1</sup>, Yusuke Goto<sup>2</sup>, Hiroyuki Morita<sup>3</sup>, Ken'ichi Minamino<sup>4</sup>, Yoshikazu Watanabe<sup>5</sup>

<sup>1</sup> 岩手県立大学大学院

<sup>1</sup> Iwate Prefectural University Graduate School

<sup>2,4,5</sup> 岩手県立大学

<sup>2,4,5</sup> Iwate Prefectural University

<sup>3</sup> 大阪府立大学

<sup>3</sup> Osaka Prefecture University

**Abstract:** There are many types of goods in retail stores such as supermarkets. While the goods position may be changed by store staffs, management of goods position may be insufficient. In this study, we proposed a model for estimating goods position in retail stores through association analysis and analyzing the staying time of customers in divided areas in the store, using customer's purchase history and movement history of customers. In this paper, we show some results from our experiments and point out several future works.

## 1 はじめに

スーパーなど小売店舗では商品の種類が多く、商品配置も変更されることがある一方で、商品位置の管理は必ずしも十分に行われていない。店舗内で顧客がどこで商品を購入した(商品をカートに入れた)のかわかると、従来独立に行われてきた購買の分析や、店舗内回遊の分析について、回遊と購買を統合した分析が可能になり、店舗内の顧客行動をより深いレベルで理解することが可能になる。

小売店舗を対象として購買と移動を統合した分析を志向する研究は存在するが、商品位置の推定を行っているものはない。増田ら<sup>1)</sup>はドラッグストアを対象としてエージェントベースシミュレーションにより売場レイアウトの効果分析を行っている。この研究では商品の位置を売場のレベルで定義して、購買履歴の情報から顧客の移動経路(売場間の移動)を逆算的に求めてシミュレーションを行っている。このため、実際に商品が配置されている詳細な位置を考慮しているわけではない。藤野ら<sup>2)</sup>は RFID を用いて顧客の店舗内での回遊データを取得し、購買履歴などを活用した顧客行動モデルを開発している。RFID のデータには精度の問題があり、そのまま動線

を推定することは困難であることから、購買履歴から動線を推測している。この研究では商品位置は商品名称とライン名称から棚のレベルで推定しているが、棚における詳細な位置把握は志向していない。また、実際の店舗で日々行われている商品位置の変更を反映できるものではない。矢田<sup>3)</sup>も RFID を利用して収集した回遊データを用いて、顧客の売場への立寄りに焦点を当て、売場訪問パターンを生成し、回遊データと購買履歴の関係から特徴抽出や分類問題に関する研究を行っている。しかしながら、この研究でも商品の位置は売場のレベルでの把握にとどまり、売場内での具体的な商品位置把握を志向したものではない。このように多くの研究では購買履歴と回遊データを統合した分析を志向しながら、回遊データの精度に問題があることや既存のデータでは商品位置の情報が欠けているために、商品位置を詳細に推定することはなされていない。

そこで本研究では、商品の詳細な位置情報が管理されていない小売店舗を対象として、購買履歴と回遊データを利用した商品位置の推定モデルを提案する。第2章では対象店舗とデータの概説を行う。第3章では商品位置推定モデルの開発に先立つ基礎分析の結果を示す。第4章では現時点で検討している

商品位置推定モデルを紹介する。商品位置推定の着想点とデータの前処理について説明した後、モデルを構成する商品位置推定指標の候補を紹介する。第5章では予備実験として各商品位置推定指標の評価結果を示し、第6章では位置推定モデル改善に向けた検討を行う。第7章はまとめである。

## 2 対象店舗とデータの概説

本研究で使用するデータは衣・食・住にわたる商品を一括して取り扱う総合スーパーマーケット3店舗から取得されている。提供された情報は(1)対象店舗の商品配置図、(2)対象店舗の購買履歴、(3)対象店舗の顧客の回遊データである。

対象店舗の商品配置図を図1に示す。提供された商品配置図は店舗を縦軸A~P、横軸1~28の2次元座標で表現している。その上で商品分類に基づき20種類のエリアが定義されており、商品棚はいずれかのエリア名が記載されている。本研究では、分析のために商品配置図における2次元座標上の点をブロックとして定義する。ブロック $B_{jk}$  ( $j=A\sim P, k=1\sim 28$ )は店舗内の縦軸の位置 $j$ と横軸の位置 $k$ にある位置を指す。このブロックは回遊データで記録されている位置情報と対応しており、本研究では商品の位置をブロックレベルで推定することを試みる。

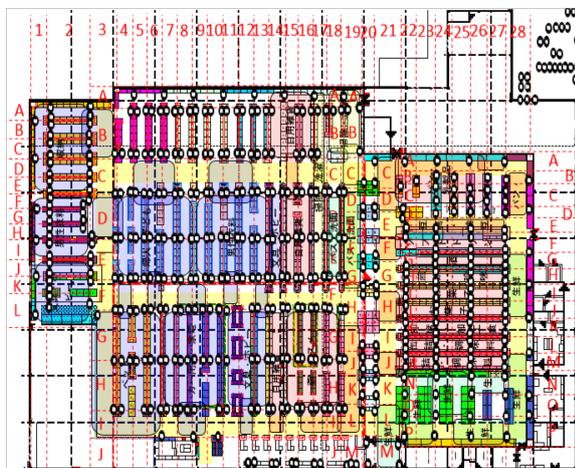


図1: 店舗配置図

購買履歴は2017年10月1日から2018年10月31日までの13ヶ月の期間に記録された3店舗合計で139,561名の会員顧客に関する47,454,867行のデータである。記録されている項目は、購買日、時間帯、購買した商品のJANコード、購買した顧客の会員ID、売上単価、売上数量、値引数量、値引金額の8項目である。なお時間帯とは、一日24時間を1時間ごとに分けたもので、24の時間帯(0, 1, 2, ...)が定義さ

れている。

回遊データは2017年10月1日から2018年10月31日までの13ヶ月の期間に記録されたもので、今回分析対象とした店舗に関しては、4,235名の会員顧客に関する16,804回の来店時回遊履歴であり、2,012,397行のデータである。記録されている項目は、店舗番号、端末ID、日時、会員ID、ブロック、エリアである。

## 3 基礎分析

商品位置推定モデルの開発に先立ち、店舗における顧客の購買傾向や来店傾向を把握するために、提供された3店舗の購買履歴を対象として基礎分析を行った。はじめに、顧客の来店1回あたりの購買金額と購入点数を集計した。集計結果から1回あたりの平均購買金額は約3,000円(中央値では約2,500円)であり、平均して約16点(中央値では約13点)の商品が購入されていることがわかった。図2・図3は店舗ごとに集計した購買金額・購入点数のヒストグラムである。図中の凡例は店舗番号を示している。分布から見ると、来店1回あたり購買金額は6,000円以内の場合が全体の90%を占め、1回あたり購入点数は33点以内の場合が全体の90%を占めることが分かる。また、3店舗はほぼ同じ傾向である。

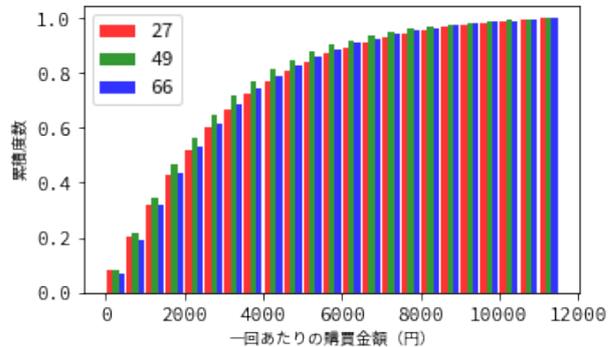


図2: 一回あたりの購買金額の累積度数分布

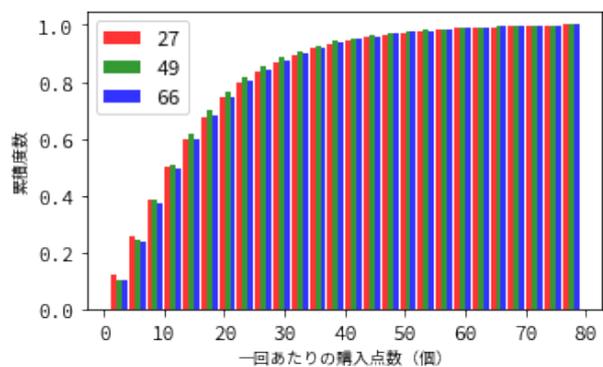


図3: 一回あたりの購入点数の累積度数分布

次に、顧客の来店頻度を把握するため、来店の平均間隔を計算した。計算結果から顧客の平均来店間隔は約 40 日（中央値では約 22 日）であることがわかった。図 4 は顧客平均来店間隔のヒストグラムである。平均して 2 週間以内で来店する顧客は全体の 40%に過ぎず、1 か月に 1 度以上来店する顧客が 6 割以上存在していることが分かった。3 店舗で同じ傾向となっている。

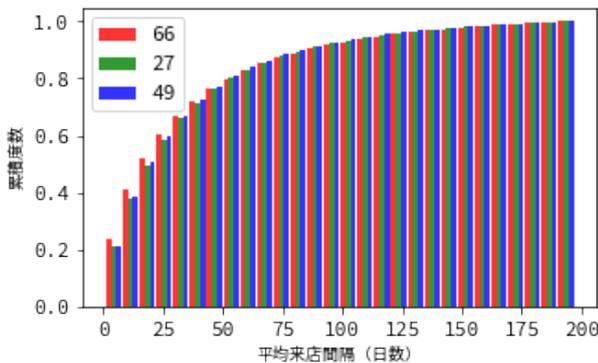


図 4: 顧客平均来店間隔期間の累積度数分布

最後に、店舗内での購買行動パターンを把握するために顧客をクラスタリングした。店舗番号 27 の店舗を対象として、来店回数は大きくばらついていることから、分析に先立って週 1 回程度来店する顧客と週 1 回以上来店する顧客の 2 グループを抽出した。この 2 グループを対象として、k-means 法を用いて来店 1 回あたりの各商品カテゴリの購買頻度を説明変数としてクラスタリングを行った。ここで購買頻度とはデータ期間内における当該商品カテゴリの購買回数を来店回数で除したもので、その顧客の来店 1 回あたりの当該商品カテゴリの期待購買点数を意味する。説明変数に用いる商品カテゴリは、購入された回数上位の商品カテゴリの中から、商品の種類が偏らないようにして 10 カテゴリを選択した。エルボー法によりクラスタ内誤差平方和の減少傾向を把握した上でクラスタ数を 4 とした。

クラスタリング結果を図 5 および図 6 に示す。縦軸はクラスタ ID であり、横軸は選択した 10 商品カテゴリである。各セルの数値はクラスタごとの各商品カテゴリの購買頻度の平均値を表す。セルの背景色は右の軸が示すように、購買頻度の平均値が高ければ高いほど色が濃い。図 5 と図 6 を比較すると、対応するクラスタ ID に関して、購買の頻度は図 5 のものが図 6 のものに比べて相対的に高いが、各商品カテゴリの購買の仕方は同じ傾向であることがわかった。両図において同じクラスタ ID のものに注目すると、ほぼ全ての商品カテゴリについて図 5 のほうが購買頻度が高い。また、両図においてクラスタ ID2

では顧客はスナック\_洋菓子や小型 PET を他のクラスタに比べてより購買する傾向があるなど、クラスタ ID が同じであれば各商品カテゴリに対する購買頻度の傾向も同様である。以上から、来店頻度が異なっても店内での購買のパターンはある程度同じものとして捉えることができ、このことから店舗内での移動経路についても同様のことが期待できる。

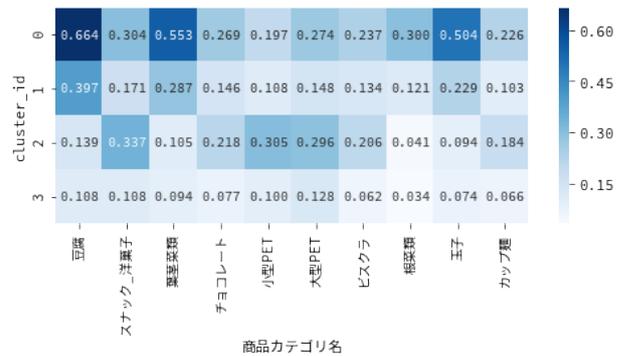


図 5: 週 1 回程度来店する顧客の各クラスタの購買傾向

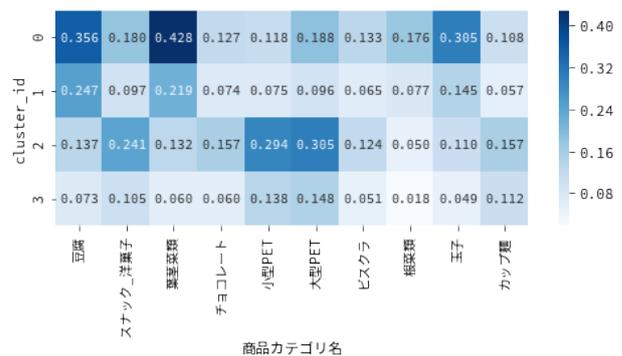


図 6: 週 2 回以上来店する顧客の各クラスタの購買傾向

## 4 商品位置推定方法の検討

### 4.1 商品位置推定の着想点

本研究における商品位置推定とは、商品位置を地図上のエリア配置のレベルで把握することではなく、商品カテゴリのレベルで店舗内のどのブロックに配置されているかを推定することである。本研究では購買履歴と店舗内の回遊データを利用して、商品位置の推定を試みる。

本研究では、マーケティングで利用される代表的なデータ分析手法であるアソシエーション分析を応用する。アソシエーション分析は顧客が商品を購入する際の購買パターンを分析することで、ある商品 A と商品 B の売れ行きについてその関連性を発見す

るものである。

アソシエーション分析では、異なる事象間の共起性を発見することができるため、店舗における商品の購買と、その商品が配置されている場所の通過の関係性を発見することができる。この考えを応用して、アソシエーション分析において購買履歴と回遊データを活用して、購買と回遊の共起性に基づいて、商品位置の推定を行うことを試みる。

## 4.2 データの前処理

店舗 66 の購買履歴と回遊データを用いて商品位置推定を試みる。購買履歴における日・会員 ID でのユニーク数は 1,084,220 件であり、回遊データの日・会員 ID でのユニーク数は 16,804 件である。アソシエーション分析では購買と回遊の共起性を計算するが、ここでは簡単のため時系列での回遊は考慮せず、購買履歴の日・会員 ID ごとに通過したブロックを追加して、購買履歴と回遊データを統合した。

さらに、一定量以上のデータがないと正確な推定は困難であることを考え、期間内に当該商品カテゴリが購買された回数が 50 より小さいものやブロックを通過した回数が 500 回より少ないデータを除いた。統合したデータの日・会員 ID ユニーク数は 16,804 件である。具体的な項目は日、時、会員 ID、商品カテゴリ名、通過ブロックである。

## 4.3 商品位置推定指標

アソシエーション分析の観点から購買履歴と回遊データを使用し、商品位置推定指標を定義する。対象とする商品カテゴリ数を  $C_n$  としたとき（本分析では  $C_n = 538$ ）、商品カテゴリ  $i$  が購買されている回数を  $X_i (i = 1, 2, \dots, C_n)$ 、ブロック  $B_{jk}$  を通過した回数

$Y(B_{jk})(j=A, B, \dots, P)(k=1, 2, \dots, 28)$  とすると、商品

カテゴリ  $i$  の商品を購入した顧客がブロック  $B_{jk}$  を通過した回数は  $X_i \cap Y(B_{jk})$  となる。また、顧客の履歴データは 1 日 1 回だけのものを対象として、その場合のトランザクション総数を  $A (A=16,222)$  とする。

アソシエーション分析で代表的に用いられる Dice 係数、Jaccard 係数、Simpson 係数は、本研究では以下のように定義される。

$$Jaccard(i, j, k) = \frac{X_i \cap Y(B_{jk})}{X_i + Y(B_{jk}) - X_i \cap Y(B_{jk})} \quad (1)$$

$$Dice(i, j, k) = \frac{(X_i \cap Y(B_{jk})) \times 2}{X_i + Y(B_{jk})} \quad (2)$$

$$Simpson(i, j, k) = \frac{X_i \cap Y(B_{jk})}{\min(X_i, Y(B_{jk}))} \quad (3)$$

また、レコメンデーションにも用いられる 2 つの指標 Confidence と Lift 値を考えると、本研究では以下のように定義される。

$$Confidence1(i, j, k) = \frac{X_i \cap Y(B_{jk})}{X_i} \quad (4)$$

$$Confidence2(i, j, k) = \frac{X_i \cap Y(B_{jk})}{Y(B_{jk})} \quad (5)$$

$$Lift(i, j, k) = \frac{(X_i \cap Y(B_{jk})) / X_i}{\frac{Y(B_{jk})}{A}} \quad (6)$$

これらは方向性があるため、Confidence は二つの計算方法で定義した。Confidence1 は商品カテゴリ  $i$  を購買した顧客がブロック  $B_{jk}$  を通る確率であり、Confidence2 はブロック  $B_{jk}$  を通った顧客が商品カテゴリ  $i$  を購買する確率を表している、分母から見ると Simpson 係数とはほぼ同じ計算式である。Confidence 指標は商品カテゴリの購買回数やブロック通過回数に大きく依存していることが分かるため、Lift 値の指標を定義した。Lift 値では店舗全体においてそもそもブロック  $B_{jk}$  を通る確率と比較して、商品カテゴリ  $i$  を購買した顧客がブロック  $B_{jk}$  を通る確率がどの程度上昇しているかを求めている。

## 5 予備実験

### 5.1 実験の準備

まず、分析の対象とする商品レベルについてデータをまとめた。購買履歴データの項目の中で、商品については JAN コードで記録されており、229,932 種類の商品がある。このレベルで商品が数多く購買される場合はごく少数であるため、JAN コードのレベルの商品を 538 種類の商品カテゴリのレベルでまとめて分析を行った。商品カテゴリは単品の商品を酒、生鮮、家電などの商品の種類や特性ごとのグループで分けしたものである。例えば、「いちご」、「りんご」などは生鮮の商品カテゴリに属している。また、店内のレイアウトは 20 個のエリアに分割されており、それぞれのエリアはいくつかのブロックを含まれている。このとき、エリア名から、そのエリアに含まれる商品カテゴリ群はある程度推測が可能である。

指標の推定結果を評価するために、正解データが必要になるが、実際の商品位置を知ることはできない。このため、商品カテゴリ名からそれが含まれるエリア名の対応関係を定義して、エリアに相当するブロック集合に商品が配置されていると仮定する。商品カテゴリ名からエリアの対応づけが可能であると考えられた商品カテゴリ数は 538 種類の内 486 種類であった。さらに、商品カテゴリの購買回数が 50 より小さいデータやブロック通過回数が 500 回より少ない場合のデータを除き、評価対象とする商品カテゴリ数は 305 とした。

その上で、指標値を計算し最上位の値となるブロックを予測結果として、そのブロックが当該商品カテゴリの含まれるエリアのブロック集合の要素であれば、正解したと考える。表 1 において、商品カテゴリ「ワイン」に関する指標値の最上位のブロックが G22, G23, G24, G25, G26, G27 に含まれていれば正解であると判定する。

表 1: 正解データの定義例

商品カテゴリ名	block	エリア名
ワイン	G22	酒
ワイン	G23	酒
ワイン	G24	酒
ワイン	G25	酒
ワイン	G26	酒
ワイン	G27	酒

## 5.2 評価実験結果

各指標を用いて推定した結果が正解かどうか判定し、正解率を表 2 にまとめた。すべての商品カテゴリを対象として評価したとき、Jaccard 係数と Dice 係数が 61.3%と最も高い正解率であった。61.3%の正解率は 1/3 程度は不正解になることを意味し、改善が期待される。このとき、生鮮売場は店舗内で大きなエリアを占めており、また生鮮商品の位置と種類は季節により変わりうることから、推定が難しい商品のタイプであると考えられる。そのため、生鮮売場に含まれる 95 の商品カテゴリを除き、生鮮以外の 210 の商品カテゴリに対する評価も行った。このとき、Confidence1 を除くすべての指標において正解率が最大 10%程度改善することがわかった。

この結果より、正解率を考慮して有用であると考えられる指標としては Jaccard 係数、Dice 係数、Confidence2、Lift 値であることがわかった。また、生鮮売場の商品に関しては、推定方法を別途で考える必要があることもわかった。

表 2: 各指標の正解率

範囲 指標	全て (305 種類)	生鮮以外 (210 種類)
Jaccard	61.3%	68.6%
Dice	61.3%	68.6%
Simpson	28.5%	30.5%
Confidence1	28.5%	27.1%
Confidence2	47.9%	59.5%
Lift	58.7%	66.2%

## 6 推定方法改善に向けた検討

### 6.1 推定指標の組合せ利用

商品位置推定の正解率を改善するために、相対的に正解率が高い指標である Jaccard 係数、Dice 係数、Confidence2、Lift 値を組み合わせて利用することを考える。Jaccard 係数と Dice 係数の推定結果は全く同じであるため、ここでは Dice 係数を採用する。全て 305 種類商品カテゴリを対象として三つの指標の正解件数のお互いの関係を表 3, 4, 5 に示す。表 3 を見ると、Lift と Confidence2 両方とも正解である件数は 146 件であり、両方とも正解ではない件数は 126 件が存在する。Lift は正解であるが、Confidence2 は正解ではない件数は 33 件であり、逆に Confidence2 は正解であるが、Lift は正解ではないケースがない。

表 3: Lift と Confidence2 推定結果の関係 (件数)

		Confidence2	
		正解	不正解
Lift	正解	146	33
	不正解	0	126

表 4: Dice と Confidence2 推定結果の関係 (件数)

		Confidence2	
		正解	不正解
Dice	正解	141	46
	不正解	7	118

表 5: Lift と Dice 推定結果の関係 (件数)

		Dice	
		正解	不正解
Lift	正解	174	5
	不正解	13	126

表3~5の結果から、組合せることで一番多くの件数が改善できそうなのはDiceとLiftである。表より、DiceとLiftがともに正解である174件に加えて、Diceが不正解でLiftが正解の5件、Liftが不正解でDiceが正解の13件を合計すると192件となる。全305であるから正解率は $192/305=62.9\%$ と改善できる可能性がある。今後は推定が難しいと思われる生鮮を除いた210カテゴリについての組合せ分析結果も確認していきたい。

## 6.2 滞在時間を用いた行動理解

顧客が商品を購入する時と購入しない時ではブロックの滞在時間が異なることが考えられる。本節では滞在時間に注目して、位置推定への活用を検討する。

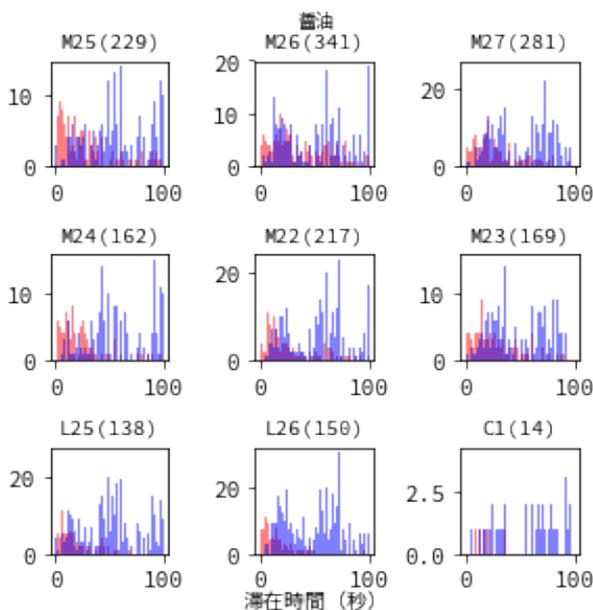


図7: 醤油の購買時・非購買時の滞在時間分布

図7は商品カテゴリ「醤油」を対象として各ブロックでの購買時・非購買時の滞在時間の分布を示したものである。各ヒストグラムの横軸は滞在時間(秒)であり、縦軸は滞在時間に対応する件数である。また各ブロックの合計の通過回数を括弧の中に示している。各ブロックはLift値の上位順に配置している。なお、Lift値が最上位であるのはブロックM25であり、このときM25も調味料売場であるため、推定結果が正解である状況である。

図7から、商品を購入したときは購入していないときより分布が左側に偏り、平均滞在時間が短い傾向にあることがわかった。この結果は、顧客の商品購買は多くの場合で計画的であり、滞在時間が短く非購買時には商品の選別や探索のために、逆に滞在

時間が長くなる可能性があることを示唆している。この結果をふまえて、前述のアソシエーション分析による推定結果と組合せることで正解率を向上できないか今後検討していきたい。

## 7 おわりに

本稿では小売店舗における商品位置推定する手法を提案した。提案に先立って行った基礎分析結果をふまえて、アソシエーション分析を応用してDice係数、Jaccard係数、Simpson係数、Confidence1、Confidence2、Lift値という6指標を利用して商品位置の推定を行った。また、商品位置の正解データを作成した上で、各提案指標の正解率を算出した。正解率の改善のための指標の組合せ利用と滞在時間を考慮する案を検討した。指標は類似しているものの、組み合わせを検討することでより強力な説明力を持つ可能性がある。また、滞在時間の状況についても図7の状況をいくつかの統計量で表現して説明することによって、説明力を持つ変数を作成することができると考える。最終的にはこれらを組み合わせたモデルを作ることによって、全体としてより強力な推定方法となることを期待している。

## 謝辞

本研究はJSPS科研費JP25350343の助成を受けたものです。使用しているデータは、株式会社トライアルホールディング様より提供いただきました。改めてお礼申し上げます。

## 参考文献

- [1] 増田浩通, 菊池晋矢, 新井健: エージェントベースシミュレーションによる小売店舗レイアウトの効果分析, 日本経営工学会論文誌, Vol.60, No.3, pp.128-144 (2009)
- [2] 藤野俊樹, 北澤正樹, 山田隆志, 高橋雅和, 山本学, 吉川厚, 寺野隆雄: スーパーマーケットで客はどう動く? -顧客動線分析とエージェントシミュレーションからわかること, 第5回社会システム部会研究会, pp.57-68 (2014)
- [3] 矢田勝俊: スーパーマーケットにおける顧客動線分析と文字列解析, 統計数理, Vol.56, No.2, pp.199-213 (2008)