

食品スーパーにおける 位置情報センサを用いた顧客の移動分析

柳橋 和磨¹ 井上 創造¹

概要：本研究では位置情報センサ・POS データ・FSP データを用いた新たな視点から小売業におけるデータ分析への仮説の中から位置情報センサを用いて顧客の食品スーパー内での移動についての分析を行った。本研究で用いる店舗のデータから各食品エリアごとの滞在時間を元に顧客人気の高い惣菜エリアと各エリア間の相関関係について調べ相関係数は 0.997086 であり、惣菜エリアに立ち寄った顧客の多くは弁当エリアに立ち寄ることが分かった。この分析に用いた特徴量として各エリアごとの滞在時間を選択肢し手法として k-means を用いたが、位置情報センサの前処理によるデータのばらつきが分析する上での誤差に影響していることを鑑み、位置情報センサの前処理を改善し、特徴量を変更することでクラスタリングの指標を元に移動分析を行う際の最適な特徴量を調べることを試みた結果、スーパー内での滞在時間と購入金額に相関が低いエリアがあることが確認でき、滞在時間が長ければ購入金額が高くなるわけではないエリアが存在した。

Analysis of Customer Movement Using Location-Based Sensors in a Food Supermarket

KAZUMA YANAGIBASHI¹ SOZO INOUE¹

1. はじめに

スーパーマーケットなどの小売業の競争は激しく各店舗が様々なサービスを提供することでライバル業者と差を付けようと日々奮闘している。情報技術の進歩が進む中でその技術を活用したデータの収集、分析が多く行われておりスーパーマーケットもまた購買に関するデータを収集、分析を行うことで今後の購買戦略の検討を進めている。しかし、データの収集を行っていてもそのデータをうまく利用している小売業は意外と多くない。研究の分野においても新たなデータ収集やその分析方法について様々な研究が行われているが、分析したデータの活用については触れられていない。これまでの研究では POS システムの導入により収集可能となった購買情報（以下 POS データと呼ぶ）や位置情報センサによる消費者の移動情報、近年ではポイントカードから消費者の情報取得する FSP データなどを分析した研究があるがそのどれもがこれらのデータの中から一つ、あるいは二つのデータのみを使用し分析を行っている。しかし、POS データ、位置情報センサ、FSP データ全ての視点から分析することでこれまでの分析では見えてこなかった知見を追求することができ、データ活用の幅を広げることができる。本研究では位置情報センサ・POS データ・FSP データを用いた新たな視点から小売業における

データ分析を行い、新たな知見を追求しスーパーマーケットにおけるデータの活用について検討、分析を行う。研究結果としてスーパー内での商品エリア毎の滞在時間と平均加速度は顧客の移動分析の目安となることがわかり、滞在時間と購入金額は相関が低いエリアが存在することが確認できた。

2. 関連研究

位置情報を用いた研究はいくつかある。宮崎は取得した RF タグ情報から顧客の動線を可視化し顧客行動の空間的な特徴を把握するため顧客存在確率密度推定を行った [1]。川口らは RF タグを利用し買い物客の位置を推定すると同時に商品の場所を推定しその情報から店舗内での案内を実現するシステム開発に取り組んだ [2]。本研究ではスーパー内のカゴに専用タグを取り付けることでスーパー内のカゴ全ての位置情報を取得することができる。

また POS データを用いた研究には、下田らが行ったスーパーマーケットの購買データ・顧客データ・店舗データ・商品データを用いた購買行動モデル構築を行っており [3]、佐藤はスーパーマーケットの顧客情報を活用し、店舗の売り上げ拡大の施策立案に繋げるための検討を行い、FSP における顧客情報も有効な活用方法についての解説を行った [4]。本研究では POS データや顧客データだけでなく顧客の位置情報も用いて分析することでそれぞれのデータの

¹ 九州工業大学大学院

生田目は FSP データと POS データ、気象データを用いて各顧客の購買行動に天気の変化が影響を及ぼす可能性について焦点を当て天候が売上もしくは消費者行動に与える影響について分析を行った [5].

本研究で用いるスーパーの商品、売り上げ、顧客のデータは一緒に共同研究を行なっている企業さまから提供していただいたものである。データの収集期間、種類等のデータ構成とデータの詳細について説明をする。

2020 年 2 月 7 日から現在 (2021 年 4 月 28 日) までスーパーでのデータを収集しており 2020 年 2 月 7 日から 2020 年 2 月 13 日までの 1 週間のデータを提供していただいた。本研究では 7 日から 13 日までの 1 週間のデータを用いて分析を行う。本研究で用いたスーパーのデータは 5 つのデータ群に分かれている。会員登録を行った顧客の性別 (sex)、生年月日 (birthday)、会員登録日 (entrydate)、会員番号 (memberid) を記録している個人情報データ。レジで商品の取り扱いが行われた時の商品登録情報 (logtype)、日付 (tradedate)、その日のみのシーケンス番号 (seq)、商品番号 (productcode)、商品の金額 (amount) を記録している POS データ。商品番号 (productcode)、商品名 (productname)、商品の部門コード (category)、大分類コード (class1)、中分類コード (class2)、小分類コード (class3) の商品詳細を記録したデータ。一意なカゴ番号 (tagid)、カゴの x 軸座標 (x2)、カゴの y 軸座標 (y2)、カゴの x 軸方向への加速度 (vx2)、カゴの y 軸方向への加速度 (vy2)、スーパー内でカゴが位置する商品エリア番号 (rezoneid)、時刻 (datetime) を記録しているカゴの移動と位置情報に関するデータ。スーパー内における商品エリア番号 (rezoneid)、商品エリア名 (rezoneiname)、商品エリアの x 座標最小値 (minx)、商品エリアの x 座標最大値 (maxx)、商品エリアの y 座標最小値 (miny)、商品エリアの y 座標最大値 (maxy) を記録している店舗内データ。このスーパーでは屋内での位置測位にも対応している高精度リアルタイム位置測位システム Quuppa を使用してスーパー内のすべてのカゴに取り付けてある専用のタグを読み取っている。これが一意なカゴ番号 (以降 tagid と呼ぶ) であり、本研究ではこの tagid を顧客の位置情報とみなして扱うこととする。tagid 以外での顧客の位置情報を得るデータはないためカゴを利用せず買い物をした顧客の POS データはあるが位置情報はないため、本研究での顧客の移動分析はカゴを使用した場合の顧客にのみ限られる。

- 顧客の移動経路や特徴を把握するためにスーパー内での顧客の移動に規則性の有無を分析する
- スーパーエリアの全体の関係性を知る前に売り上げが高い惣菜エリアとその他エリアの関係を分析する
- 顧客の移動と購買の関係を調べるために各エリアの位置情報と POS データを分析する

tagid の位置情報には顧客の移動とは結びつけられない瞬間のデータが存在する、カゴに専用タグを取り付けることでカゴ利用する顧客の位置情報が分かるが顧客が利用していないカゴは分析には用いられないが収集されているためである。今回データの収集を行ったスーパーでは見取り図 1 の鮮魚・牛肉、加工肉・牛乳 要冷飲料の間には従業員通路の入り口がある、この二箇所の入り口にはカゴがストックされていて買い物客が買い物の途中で利用することができるようになっている。また、入口左と入口右にもカゴがストックしてある。本研究での顧客の移動分析ではスーパー内での顧客の移動と購買商品、金額との分析を視野に入れているため買い物中の一連の経路に重視する。従業員通路入り口のカゴのストック位置周辺の位置情報はあらかじめ省いておく。



tagid は 1 日のうちに複数利用される場合があるために、一度の買い物で取得した tagid の位置情報を区切る必要がある。tagid が持つ位置情報データに状態を示す flag を追加する、2 でスーパー内の位置情報の flag の状態を二つに分けるエリアを示す。入口左、入口右、レジのエリアにある位置情報の flag を 0 としてそれ以外のエリアにある flag はすべて 1 とする。同じ tagid を時系列順に flag の状態を

調べ 0 から 1 に変わるときに顧客がカゴを取り買い物を入店したと考え買い物のスタート地点とみなす、flag が 1 から 0 に変わるときにレジで買い物をしているとみなし買い物が終了とする。この間の位置座標を顧客の一度の買い物での移動経路とし経路番号（以後セクション番号とする）を付与する。また、レジでの POS データで買い物をした tagid と直前にレジを通過した tagid が同じセクション番号をを紐付けすることで顧客の位置情報と POS 情報を確認できるようにする。

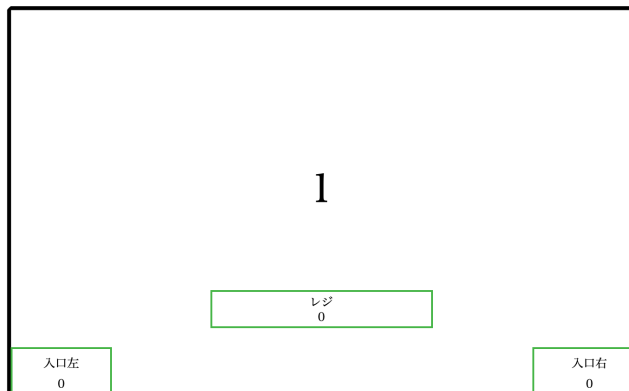


図 2 セクション分け見取り図

4.3 分析手順

セクションではスーパー各エリアごとに位置情報の取得した時間の最上と最大の差から滞在時間を算出している、この滞在時間とは顧客がエリアを通過することで取得される位置情報も影響しているため通過しただけの時間も含まれている可能性もある。また各エリアの顧客の移動情報の目安として x 軸と y 軸の加速度の二乗和平方根から平均加速度の算出を行った。

4.4 結果

図 3 は 1 人の顧客のスーパー内での時系列位置情報データから RNN を用いてエリアの移動順序を予測したものである。横軸がデータ数で縦軸は出力を安定させるためにスーパーのエリア ID を 01 に正規化したグラフとなっている。青色でプロットされているのは正解データであり、オレンジ色が訓練データの予測、緑色がテストデータの予測値となっている。予測精度が高いように思えるが、スーパーでの位置情報は同じエリアにいる場合エリア ID は同じとなるため予測するのは容易である。顧客のエリアが変化を表しているグラフの箇所は予測が外れていることがわかる。

顧客の買い物における一連の移動経路から規則性を見つけて出すことが難しいことがわかり、分析するエリアを絞っ

て分析を試みた。分析するエリアに惣菜エリアを選んだ理由はスーパー側が惣菜エリアの売り上げが良いために顧客の特徴を気にしていたからである。惣菜エリアの滞在時間を分析する上で滞在時間は図 4、図 5 から 3600 秒以下にデータが集中していることが確認できた。各エリアの滞在時間は 3600 秒以下で惣菜エリア (diff_87) と各エリアとの相関関係を表したグラフとは図 6 である、青色が濃い方が相関関係は高くなる。図 6 の相関図から惣菜エリアと相関関係が最も高いのは弁当エリアであることがわかった。弁当エリアと惣菜エリアの顧客の滞在時間の長さは確認できたが図 1 でスーパー内のエリアを確認すると惣菜エリアと弁当エリアは隣接しており、惣菜エリアに立ち寄った顧客が弁当エリアを通る可能性や、惣菜と弁当は食品のジャンルとして近い部類にはいることも考えられ惣菜エリアに立ち寄った顧客が弁当エリアに立ち寄るのは必然的である可能性もある。

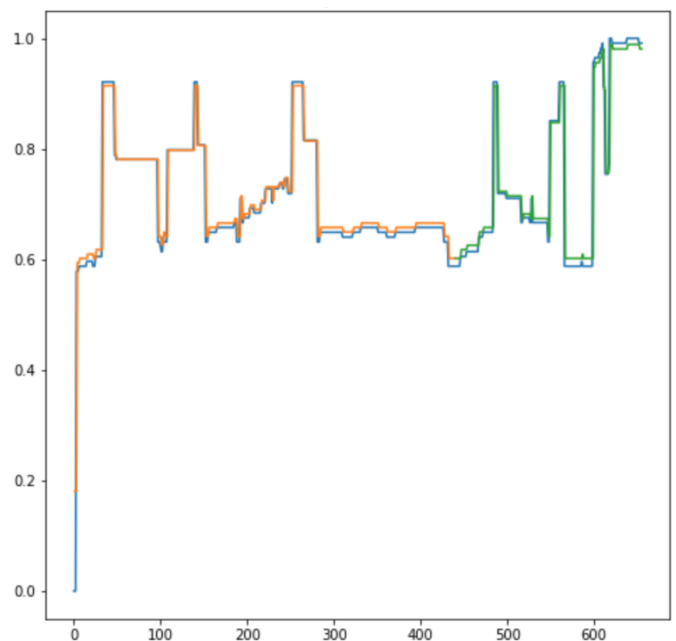


図 3 RNN による顧客のエリア移動予測

図 8 は 2020 年 2 月 7 日から 2020 年 2 月 13 日の間における、顧客のセクションをでエリア毎の滞在時間 (time) ・購入金額 (amount) ・平均加速度 (avg_ac) で示したものである。各エリアの滞在時間は 1 時間以内に限定している、左から購入金額の高いエリア順となっている。購入金額が高い野菜、惣菜は滞在時間も長いことがわかった。しかし、滞在時間が一番長い加工肉や次に滞在時間が長い牛乳要冷飲料は購入金額はさほど高くないことが分かり、滞在時間と購入金額の順位が同じということではないことが確認できた。購入金額だけでは商品単価の差が順位に影響する可能性を考慮して図 9 では購入金額ではなく各エリアでの購入商品数 (purchases) と滞在時間 (time) ・平均加速度

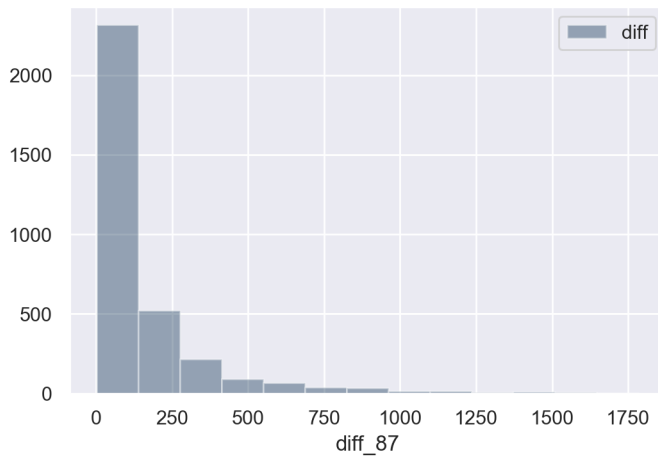


図 4 惣菜エリアの滞在時間 (diff) が 1800 秒以下のデータ数

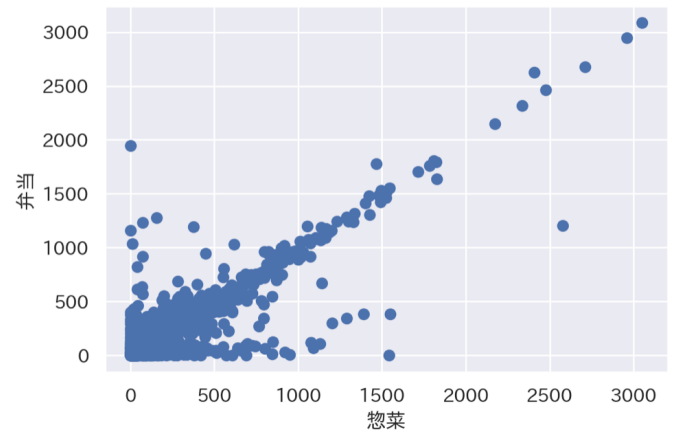


図 7 惣菜エリアと弁当エリアにおける滞在時間の相関

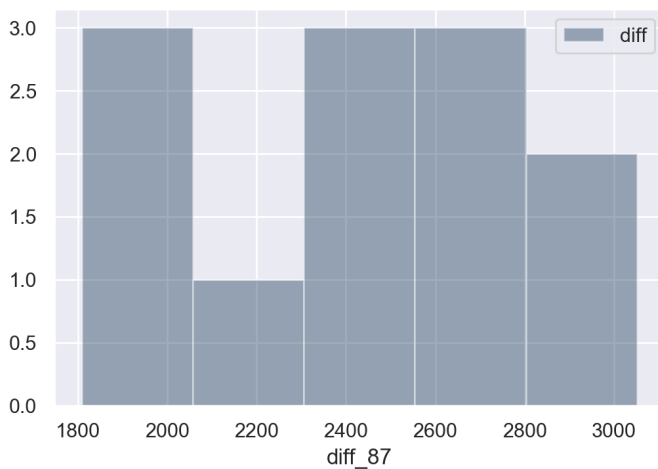


図 5 惣菜エリアの滞在時間 (diff) が 1800 秒を超え 3600 秒以下のデータ数

diff_rezoneid	diff_87
0	route -0.000642319
1	diff_9 0.638246
2	diff_12 0.323602
3	diff_16 -0.000296446
4	diff_17 0.427673
5	diff_54 0.2386
6	diff_55 0.149339
7	diff_56 0.0880492
8	diff_57 0.334832
9	diff_58 -0.000452697
10	diff_71 0.990523
11	diff_72 0.989045
12	diff_73 0.968175
13	diff_74 0.966525
14	diff_75 0.997283
15	diff_76 0.991576
16	diff_77 0.996057
17	diff_78 0.997086
18	diff_79 0.958978
19	diff_80 0.99436
20	diff_81 0.971159
21	diff_82 0.967562
22	diff_83 0.98387
23	diff_84 0.981635
24	diff_85 0.929445

25	diff_86 0.953009
26	diff_87 1
27	diff_88 0.998079
28	diff_90 0.955512
29	diff_91 0.990854
30	diff_92 0.996341
31	diff_93 0.99617
32	diff_94 0.988594
33	diff_95 0.989222
34	diff_96 0.996706
35	diff_97 0.994635
36	diff_98 0.980338
37	diff_99 0.976705
38	diff_100 0.972758
39	diff_101 0.990206
40	diff_102 0.987721
41	diff_103 0.972286
42	diff_104 0.933115
43	diff_105 0.984265
44	diff_106 0.983142
45	diff_107 0.843808
46	diff_108 0.970955
47	diff_109 0.982152
48	diff_110 0.986752
49	diff_111 0.978313
50	diff_112 0.988119
51	diff_113 0.996238
52	diff_114 0.993439
53	diff_115 0.960451
54	diff_116 0.997749

55	diff_117 0.981261
56	diff_118 0.981652
57	diff_119 0.996898
58	diff_120 0.994597
59	diff_121 0.996668
60	diff_122 0.996411
61	diff_123 0.996316
62	diff_124 0.98038
63	diff_125 0.977632
64	diff_126 0.997369
65	diff_127 0.997093
66	diff_128 0.996222
67	diff_129 0.997433
68	diff_130 0.992492
69	diff_131 0.997204
70	amount -0.0119969

図 6 惣菜エリアと各エリアにおける滞在時間 (diff) の相関

(avg_ac) を表した．左から購入商品数の多いものとなっている．購入金額と購入商品数で売り上げ順に多少の変化があるエリアもあるが、さほど影響があるとは思えないため、売上に繋がる購入金額の方に注力して分析を行なった．滞在時間が最も長い加工肉は購入金額が 17 番目となった、次に滞在時間が長い冷凍の購入金額は下から 2 番目ということになったが平均加速度はどれも高く他のエリアに比べて移動が早いことがわかった．

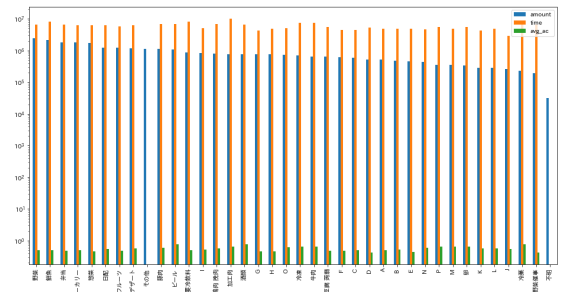


図 8 エリア毎の購入金額・滞在時間・平均加速度

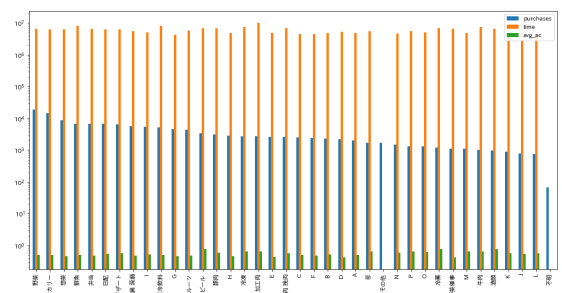


図 9 エリア毎の購入回数・滞在時間・平均加速度

5. 考察

顧客の移動分析から規則性を見つける分析において図3でも分かるように予測精度が良かったのは滞在エリアを予測したためだと思う。顧客があるエリアに入り次の位置情報でそのエリアにいる可能性はとて高くエリア内にいる間に取得される位置情報データは多いため予測精度が高くなりやすい。また、エリアから次のエリアに移動したエリアの変わり目の予測が外れていることから時系列の一つ前に位置座標が大きく影響していると考えられる。

滞在時間に2日のような極端に大きな値が確認されたのはセクション分けの際に入り口からカゴを持って入店された顧客がスーパー内でカゴを放置することがあるという話をスーパー側からお聞きしたので放置されたカゴを利用した顧客がレジを通ることでセクションとして認識されてしまったと考える。図4、図5の滞在時間のデータ数を見ても3600秒以下がほとんどであり更に1200秒以上も外れ値しても良いと考える。惣菜エリアに対する分析で弁当エリアとの相関が高いのは弁当エリアとが隣接していることや食品のジャンルとして同じと言えることが大きいのではないかと考える、顧客が経路として通過しているだけの可能性があったが図8で惣菜エリアも弁当エリアも平均加速度は低い方でありエリアに滞在している可能性が高いと考える。

図8で顧客の各エリアでの滞在時間と購入金額の関係から滞在時間が長いのに購入金額が高くない理由はエリアを通路として利用している可能性が平均加速度から確認できたが、購入金額が高い野菜や鮮魚も図1でエリアを確認すると加工肉や冷菓と配置は似ているので加工肉なども購入金額が上がって良いはずだがエリアの大きさに違いがあるため加工肉や冷菓もエリアを商品数を増やすことで売り上げが上がると考える。

6. おわりに

滞在時間が長ければ購入金額も増えるというわけではないことが確認できた。今回の分析で出た滞在時間が最も長いのに購入金額はさほどでもない加工肉エリアは滞在時間が長い理由が通路として利用している人がいることが平均加速度から推測することができ原因の一つとして考えられるが、実際にエリアで売られている商品の詳細や配置位置にも要因があるかもしれないため、今後の研究として加工肉、冷菓と野菜、鮮魚等の似た配置のエリアのPOSデータの分析を試みる

参考文献

- [1] 宮崎慎也: RFID 技術を用いて取得したスーパーマーケット内の顧客動線の可視化, 日本建築学会技術報告書, Vol.18, No.40, pp.1022-1037, 2012

- [2] 川口正太郎, 山下晃弘, 松林勝志: RFID タグを用いた買い物支援システムにおける買い物客と商品の位置推定法, 第78回全国大会公演論文集, Vol.2016, No.1, pp.403-404, 2016
- [3] 下田稜, 小山竜平, 杉浦孝典: スーパーマーケットのPOSデータに基づく消費者の購買行動モデル構築, 第18回社会システム部会研究会, pp.175-178, 2019
- [4] 佐藤稔, 加藤悦子, 松田芳雄: 食品スーパーにおけるFSP分析の研究, UNISYS TECHNOLOGY REVIEW, No.87, pp. 56-64, 2005
- [5] 生田目 崇, 須山 憲之: 小売店の購買行動における天気の影響-スーパーマーケットのPOSデータを用いた分析, 商学研究所報, Vol.41, No.8, pp. 1-29, 2017 No.8, pp. 1-29, 2017

謝辞

本研究の分析に用いたデータを提供していただきました株式会社GLORY様には心より感謝申し上げます。