

# グラフ研磨手法を用いた 顧客の店舗選択モデルの構築

中原 孝信, 羽室 行信, 宇野 毅明

## 1. はじめに

顧客のライフスタイルや購買行動が多様化し、多業種・業態によってさまざまな商品・ブランドが扱われている現代においては、顧客の店舗選択基準は一律ではなく、さまざまな要因によって顧客は店舗を選択している。小売業に限っても多様な店舗で最寄り品が販売されている状況で、どのような特性を持つ店舗が消費者から選択されるのかを明らかにすることは、店舗の戦略を検討するうえで重要な指針となるであろう。

これまで店舗選択の要因を明らかにするために、定性的な調査を中心とした研究が行われてきた。上田 [1] は、大型スーパーマーケットを構成する各部門が消費者の店舗選択にどのように貢献し、どのような構成要素によって評価されているかを調査しており、消費者が重視する部門とその要因を示している。また、高橋 [2] は、「価格」以外にも、「利便性」「食料品の品揃え」「生鮮食品の鮮度」などの要因がストアイメージを構成していることを示している。同じように Morshett ら [3] は、「価格」や「アクセス性」以外に「店舗のクオリティ」などの要因があることを示している。

本研究では、特定の志向を持った顧客と持っていない顧客で、商品を購入する際に考慮する店舗選択の要因にどのような違いがあるかを明らかにする。どのような小売業がどのような顧客から支持されているのかを把握し、その要因を理解することは、店舗における需要がどのような志向を持つ消費者から構成されてい

るかを理解することとなり、商品ラインの選択や商品開発においても重要な情報になると考えられる。また、これまでの研究の多くはアンケート調査だけが用いられてきたが、本研究は、顧客の購買履歴が記録されたスキャンパネルデータを併せて用いることで、アンケートによる消費者の内面的な属性の理解と、購買履歴による定量的な購買行動の解明を試みる。

具体的には、顧客の志向をアンケートにより特定し、その志向の有無によって選択される商品群と店舗の関係をモデル化し、分類モデルを構築する。そして得られたモデルから、意味解釈の妥当性と分類精度の向上という2つの観点から評価を行う。本研究で利用したスキャンパネルデータとアンケートデータは、経営科学系研究部会連合協議会が主催する平成25年度データ解析コンペティションより提供を受けた。

以降2節で関連研究について述べ、3節では提案手法を説明し、4節ではデータを用いた分類モデルの構築について述べ、5節でまとめと今後の課題について述べる。

## 2. 関連研究

複数店舗の選択要因をアンケート調査から解明した研究は、上述のとおりであるが、ID付きPOSデータを用いた方法としては、渡辺ら [4] の研究がある。ここでは、自己組織化マップを用いて顧客を来店習慣からタイプ分けし、買いまわりタイプによる購買行動の違いを示している。一方で土井ら [5] は、アンケートとID付きPOSデータを用いて、複数店舗の選択とライフスタイルの関係を明らかにするための研究を行っている。ここでは、ライフスタイルを分類するためにアンケートデータを用いた因子分析を行い「こだわり消費派」「節約消費派」「栄養バランス重視派」「新商品消費派」「アクティブ消費派」の5つの因子を抽出している。また、顧客の購買情報をもとに店舗をクラスタリングし、店舗クラスタと購買情報を利用してランダ

なかはら たかのぶ  
専修大学 商学部  
〒214-8580 神奈川県川崎市多摩区東三田 2-1-1  
はむろ ゆきのぶ  
関西学院大学 経営戦略研究科  
〒662-8501 兵庫県西宮市上ヶ原一番町 1-155  
うの たけあき  
国立情報学研究所 情報学プリンシプル研究系  
〒101-8430 東京都千代田区一ツ橋 2-1-2  
受付 14.7.25 採択 14.11.10

ムフォレストを用いたライフスタイルの推定を行っている。土井らの研究は、アンケートデータとID付きPOSデータを利用し、店舗の選択とライフスタイルの関係性を明らかにしようとしていることから、本研究と関連がある。しかし、土井らは、説明変数として店舗と商品を独立に扱っている。一方で本研究では、各店舗で購入された商品を店舗と商品のアイテムペアとして扱い、どのような商品をどのような店舗で選択する傾向にあるかを明らかにしている。

### 3. 店舗選択の要因を抽出するための手法

分類モデルを構築するにあたって、本研究では目的変数として健康志向の顧客群を正例とし、その他の顧客群を負例と定義した。健康志向かどうかは、健康に関する食事関連の4項目のアンケート結果をスコアリングすることにより定義した。具体的には、「1食でより多くの食材が摂れるように料理をする」、「1汁3菜を意識して料理を作る」、「1食あたりのカロリーや塩分・脂質・糖分・食物繊維などを意識しながら食事を作る」、「自分の健康・体調管理よりも、家族の健康・体調管理を意識して料理をする」という4つの質問を対象にして、5件法の回答から平均値を計算し、平均値以上であれば「健康志向」そうでなければ「非健康志向」として定義した。

そして、説明変数として店舗と商品の関係性を表した変数を用いることで、健康志向の顧客と非健康志向の顧客が持つ商品に対する店舗のイメージを表現するモデルを構築する。具体的には、説明変数として、顧客の店舗における商品の購入の有無を表した2値変数を設定する。例えば、「ダイエー」で「牛乳」を買ったことがある顧客は、「ダイエー-牛乳」という変数の値が1となり、購入のない顧客の値は0となる。

店舗数を  $n$ 、商品数を  $m$  とすると、説明変数は  $n \times m$  次元ベクトルとなる。しかし、今回扱うデータでは、カバーするサンプル数があまりにも少ない変数が多数を占めることになり、結果として予測精度の高いモデルが得られないという問題がでてくる。そこで、前処理として変数をクラスタリングすることを考えるが、その方法に本研究の特徴がある。

クラスタリングの方法としては、店舗-商品の変数を節点とし、互いに類似した変数間に枝を張った一般グラフを構築し（「類似度グラフ」と呼ぶ）、そこから密な部分グラフをクラスタとして抽出する（詳細は後述）。変数間の類似度の定義としては、顕在パターン(emerging patterns)における支持度(support)と増

加率(GR: Growth Ratio)を用い、それらの値がユーザの設定した最小支持度および最小増加率以上の場合に枝が張られる。

支持度と増加率の定義は次のとおりである。顧客をトランザクション、そして店舗と商品のペアをアイテムと考え、正例、負例のトランザクション集合をそれぞれ  $D_p$ 、 $D_n$  で、また  $D_p$  においてアイテム  $a, b$  が共起する部分集合を  $Occ_p(a, b)$  で表す。 $D_p$  における2つのアイテム  $a, b$  の支持度  $support_{D_p}(a, b)$  は、式(1)に示されるように、アイテム  $a, b$  の共起確率として定義される。

$$support_{D_p}(a, b) = \frac{|Occ_p(a, b)|}{|D_p|} \quad (1)$$

また、 $D_p$  における2つのアイテム  $a, b$  の  $D_n$  に対する増加率  $GR_{D_n \rightarrow D_p}(a, b)$  は、式(2)のとおり定義される。これは、負例での共起確率に対する正例での共起確率の比であり、1.0より大きければ、アイテム  $a, b$  は正例に特徴的な共起パターンと言える。

$$GR_{D_n \rightarrow D_p}(a, b) = \frac{support_{D_p}(a, b)}{support_{D_n}(a, b)} \quad (2)$$

そして、最小支持度と最小増加率を閾値として与え、変数ペア（顕在パターン）を全列挙し、それらの変数間に枝を張る。このように得られた類似度グラフでは、お互いに類似した変数群の枝密度は濃くなり、逆に類似していない変数群の枝密度は薄くなる。そこで、類似度グラフから、ある程度密度の濃い部分グラフをクラスタとして抽出することで、正例に特徴的な変数クラスタを構成することができる。同様の考えを負例にも適用し、正例と負例それぞれに特徴的な類似度グラフを構成しておく。

一般グラフのクラスタリングには、ニューマンクラスタリング[6]やグラフ分割[7]などがあり、また、クリークをクラスタとして扱う方法としては、極大クリーク列挙や、クリークパーコレーション[8]などがある。このようにこれまでもさまざまな手法が提案されてきたが、どの手法も問題点を抱えており、決定打になっていないというのが現状である。

例えば、極大クリーク列挙では、現実データにおいては多くの場合、類似した極大クリークが多数列挙されてしまうという問題がある。列挙された極大クリークの類似関係を用いて、極大クリークをさらにクラスタリングするという方法も提案されているが、列挙される極大クリークの数によっては計算量が問題となる。このような問題の多くは、対象とするグラフにノイズ

が含まれるために起こる問題とも考えられる。

そこで、最近著者らは、対象とするグラフをクリーニングする「グラフ研磨」手法を提案している [9]。これは、グラフをクラスタリングする前に、枝を張り直すことでグラフを再構成し、できる限り構造を明確化しておこうというものである。直感的には、枝密度の濃い部分グラフはより濃く、薄い部分グラフはより薄くするというものである。このような方法を適用することで、列挙されるクリークの数に劇的に少なくなることが期待される。

研磨の方法は至ってシンプルで、すべての節点ペアについて、その類似度がユーザの指定した閾値以上であれば接続し、そうでなければ接続しないというルールに従って、新たなグラフを再構成する。全節点ペアの計算は節点数の 2 乗の計算量が必要となるが、より効率的なアルゴリズムが存在する [9]。

類似度としては Jaccard 係数を用いる。グラフ上の 2 つの節点  $u, v$  の Jaccard 係数  $sim(u, v)$  は、以下のとおり定義される。

$$sim(u, v) = \frac{|N(u) \cap N(v)|}{|N(u) \cup N(v)|} \quad (3)$$

ここで  $N(u)$  は節点  $u$  に直接接続のある節点集合を表している。そしてユーザが与えた最小類似度  $\delta$  以上の類似度を持つ変数ペアに枝を張ることでグラフを再構成する。このようにグラフを再構成すると、大雑把に言えば、共通する隣接節点の多い節点間に枝が張られ、少ない節点間の枝は切断される。これは、SNS における友達紹介のアルゴリズム（すなわち共通する友達が多ければ友達である可能性が高い）と同様なもので、グラフ構造のプリミティブな変化予測（リンク予測）を行っているとも解釈できる。

そして新たに構成されたグラフを入力として同様の研磨手法を繰り返し適用し、グラフの構成に変化がなくなるか、もしくはユーザの指定した最大繰り返し回数に達すれば終了する。最終的に得られたグラフが研磨グラフである。この研磨グラフから列挙された極大クリークを我々はマイクロクラスタと呼んでいる。

以上により得られたマイクロクラスタを説明変数として分類モデルを構築する。マイクロクラスタとしての変数は、マイクロクラスタを構成するアイテム数の 30% 以上のアイテムが顧客のトランザクションに含まれている場合に 1 をとる 2 値の変数と定義した。

分類モデルにはロジスティック回帰モデルを用いる。分類モデルにおける目的変数を  $y \in \{0, 1\}$ （0：負例、

1：正例）、 $p$  個の説明変数（マイクロクラスタ）ベクトルを  $\mathbf{x} = (x_1, x_2, \dots, x_p)$  とすると、ロジスティック回帰モデルは式 (4) で表される。

$$\Pr(y = 1|\mathbf{x}) = f(\boldsymbol{\beta}^\top \mathbf{x} + \beta_0) \quad (4)$$

$f(\cdot)$  はロジスティック関数であり、 $f(a) = 1/(1 + \exp(-a))$  で定義される。 $\boldsymbol{\beta} \in \mathbb{R}^p, \beta_0 \in \mathbb{R}$  は、それぞれ回帰係数ベクトルと定数項であり、これらは訓練サンプルから推定する。

$\boldsymbol{\beta}$  の推定には最尤推定法を利用するのが一般的であるが、説明変数の数  $p$  がサンプル数に比べて多いとき、解の不定性が問題となり異なる推定法が必要となる。また変数増減法などによる変数選択も、 $p$  が大きくなると変数の組合せ数が指数関数的に大きくなり求解が困難となる。この問題に対してさまざまな推定法が提案されてきたが、罰則項付きの対数尤度を最大化する問題（式 (5)）を解く罰則付き最尤推定法が有効であることがわかってきた [10]。

$$\operatorname{argmax}_{\boldsymbol{\beta}} \left[ \frac{1}{N} \sum_{i=1}^N \{y_i \log \Pr(y_i = 1|x_i) + (1 - y_i) \log(1 - \Pr(y_i = 1|x_i))\} - \lambda P(\boldsymbol{\beta}) \right] \quad (5)$$

その中でも  $P(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_1$  とした lasso、および  $P(\boldsymbol{\beta}) = \|\boldsymbol{\beta}\|_2$  とした ridge 回帰がよく利用される。ここで、 $\|\boldsymbol{\beta}\|_q$  は  $q$ -ノルムで  $\|\boldsymbol{\beta}\|_q = (\sum_{i=1}^p \beta_i^q)^{1/q}$  である。式 (5) の  $\lambda \in [0, \infty)$  は正の定数であり、lasso においては  $\boldsymbol{\beta}$  をどの程度疎に選択するかかのトレードオフパラメータである。 $\lambda$  が大きい場合には、 $\boldsymbol{\beta}$  の多くの値が 0 となる。逆に  $\lambda$  が 0 の場合は通常最尤推定法となる。ridge 回帰においては  $\lambda$  を大きく設定しても回帰係数が 0 と推定されることはないが、推定値が全体的に小さく推定されることになる。

ridge 回帰は共線性への対処法として用いられるが、変数選択としては機能しない。一方で lasso は  $\lambda$  の値によっては多くの回帰係数が 0 となることから変数選択の有効な手法として注目されている。しかしながら一方で共線性のある変数が選ばれにくいといった問題も指摘される。そこで、両者の罰則を結合し、 $P(\boldsymbol{\beta}) = (1 - \alpha) \frac{1}{2} \|\boldsymbol{\beta}\|_2 + \alpha \|\boldsymbol{\beta}\|_1$  とした elastic net がある [10]。 $\alpha$  は 0 以上 1 以下の値で、 $\alpha$  を 0 に近づければ ridge 回帰の罰則が強くなり、逆に 1.0 に近づければ lasso の罰則が優先される。本研究では、ridge 回帰と lasso よりもモデルの予測精度が高かった elastic

net を使うことにした<sup>1</sup>。その際に指定した  $\alpha$  は試行錯誤の実験から 0.001 とした。また、 $\lambda$  は 10 回の交差検証によりモデルの予測誤差を最小化する値に定めた。

#### 4. 計算実験

本研究で利用するデータは、2012 年の 1 年間で約 6,500 人のモニターによるスキャンパネルデータである。

##### 4.1 マイクロクラスタの生成

類似度グラフを生成するために、店舗と商品のペアをアイテムとして利用するが、商品としては大分類、中分類、小分類、細分類の 4 つのアイテム分類をそれぞれ利用した。また顕在パターンの最小支持度は 0.1 で、最小増加率は 1.0 とした。グラフ研磨は、 $\delta$  の値によってさまざまなグラフ構造が得られるため、最適な  $\delta$  を一意に定めることは困難である。そこで、本研究では  $\delta$  を 0.1 から 0.9 までの 0.1 刻みで動かし、各  $\delta$  でマイクロクラスタを列挙してそれらすべてを説明変数に利用した。

図 1 に類似度グラフ、図 2 に研磨後のグラフを示す。これらのグラフは、健康志向を対象に店舗と細分類のペアをアイテムとして扱い生成されたものである。そして図 2 は類似度グラフからグラフ研磨を  $\delta = 0.6$  で行った場合を示している。研磨後のグラフは、ノイズが除去され共起関係の強い節点同士が結びついていることが確認できる。

研磨後のグラフからマイクロクラスタを列挙する際に節点数が 10 個以下のマイクロクラスタのみを列挙した。これはあまりにも要素数の大きいマイクロクラスタが列挙されると意味解釈が困難になるからである。各  $\delta$  で列挙されたマイクロクラスタは重複することはないが、 $\delta$  をまたいで列挙されたマイクロクラスタは重複する可能性があるため、同一の節点集合で構成されるマイクロクラスタがあれば単一化した。このように  $\delta$  を変えてマイクロクラスタを列挙することで、多様なマイクロクラスタを生成した。

表 1 は、 $\delta$  を変えたときに得られたマイクロクラスタ数と節点平均数を示している。またこれは健康志向を対象に店舗と細分類のペアをアイテムとして扱った場合の結果である。ORG は研磨を行わずに極大クリークを列挙した場合の結果であり、節点数が 10 個以下の極大クリークのみを列挙した。項目名「MC 数」は得られたマイクロクラスタの数、「節点平均数」は 1 つのマイクロクラスタに属する平均節点数を示している。

<sup>1</sup> 統計解析ツール R のパッケージ glmnet1.9-5 を用いている。

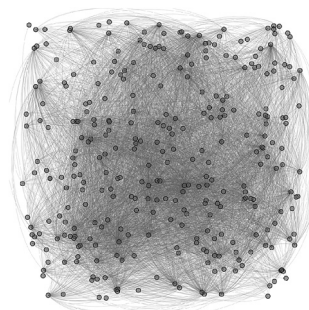


図 1 類似度グラフ

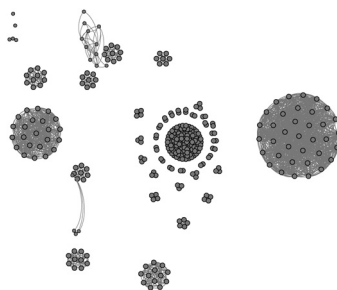


図 2 研磨後のグラフ ( $\delta = 0.6$ )

表 1  $\delta$  別のマイクロクラスタ (MC) 数と節点平均数

$\delta$	MC 数	節点平均数
0.1	3	4.667
0.2	8	5.875
0.3	22	4.773
0.4	19	3.105
0.5	21	4.143
0.6	38	3.421
0.7	41	3.634
0.8	37	3.027
0.9	17	2.176
ORG	108	3.4537

ORG は研磨前の類似度グラフに対する結果である。

$\delta$  が 0.1 や 0.2 と小さい場合は、共起関係が弱くても枝が追加されるため研磨後のグラフは密なグラフになる。したがって、 $\delta$  が小さいと節点数の多い巨大なマイクロクラスタが少数列挙されることになり、またそれにともないマイクロクラスタの節点平均数も大きくなる。しかし本研究では節点数が 10 個以下のマイクロクラスタを省いているため、マイクロクラスタ数と節点平均数も少なくなっている。

また  $\delta$  が 0.4 から 0.7 まではマイクロクラスタ数が増加している。これは、 $\delta$  を増加させると、間接的に共起関係の弱い枝は削除されるため、小さいサイズのマイクロクラスタが生成されるからである。また、それ



表2 マイクロクラスタの抜粋

<b>セブン&amp;i系クラスタ</b>
{ セブン&i系-その他水物, セブン&i系-蒲鉾, セブン&i系-その他畜産, セブン&i系-コンニャク, セブン&i系-冷凍農産素材, セブン&i系-キャンディ・キャラメル, セブン&i系-炭酸フレーバー, セブン&i系-油揚げ, セブン&i系-その他加工水産, セブン&i系-インスタントカレー }
<b>ダイエー系クラスタ</b>
{ ダイエー系-半生菓子, ダイエー系-生麺・ゆで麺, ダイエー系-食パン, ダイエー系-その他畜産, ダイエー系-牛乳, ダイエー系-菓子パン, ダイエー系-ヨーグルト, ダイエー系-豆腐 }
<b>生協クラスタ</b>
{ 生協の個人宅配-その他農産, 生協の個人宅配-和惣菜, 生協の個人宅配-冷凍水産素材, 生協の個人宅配-冷凍調理 }

以上の $\delta$ になるとマイクロクラスタ数が減っているが、これは接続がなくなり、2 節点以上のマイクロクラスタではなく、単一の節点が増えていることが理由である。マイクロクラスタ数は、各 $\delta$ で ORG に比べて大幅に少なくなっている。前述したように、極大クリーク列挙では、現実データにおいては多くの場合、類似した極大クリークが多数列挙されてしまうという問題があるが、グラフ研磨によりその問題を解決できていることがわかる。

このようにして提案手法では、健康志向と非健康志向を対象に店舗と4つのアイテム分類で、それぞれ列挙されたマイクロクラスタを合計すると、最終的に1,488個のマイクロクラスタが得られた。

表2は、マイクロクラスタの例を示している。各マイクロクラスタの要素を見ると、購入商品は異なるが、それぞれ同一の大型店舗から構成されるマイクロクラスタを形成しており、店舗選択の要因として、セブン&i系や、ダイエー系、そして生協を対象にそれぞれバラエティーに富んだ商品が購入されている。

4.2 健康志向予測モデルの構築

得られた1488個のマイクロクラスタを説明変数に利用して、罰則付きロジスティック回帰によって「健康志向」と「非健康志向」の分類モデルを構築した。

表3は、10回の交差検証で評価したモデル全体の予測精度を示している。予測精度は、式(4)のロジスティック回帰モデルにおける確率が0.5以上なら健康志向、0.5より小さい場合に非健康志向と予測し、予測クラスと実クラスとの一致から正答率を計算し予測精度

表3 全体の精度比較

グラフ研磨	説明変数の数	選択された変数の数	予測精度
あり	1488	160	70.69%
なし	1820	328	65.35%

表4 クラス別の精度比較

	グラフ研磨	適合率	再現率	F 値
健康志向	あり	0.800	0.438	0.566
健康志向	なし	0.761	0.301	0.431
非健康志向	あり	0.677	0.915	0.779
非健康志向	なし	0.631	0.927	0.751

とした。「説明変数の数」は、罰則付きロジスティック回帰モデルで利用した説明変数の数を表している。また「選択された変数の数」は、罰則付きロジスティック回帰モデルの結果、選択された変数の数を表している。最終的に選択された変数の数は、グラフ研磨ありの場合は160個、なしの場合は328個で、グラフ研磨の効果によってより少数の変数が選択された。提案手法であるグラフ研磨を用いたマイクロクラスタによる予測精度は高く70.69%であった。一方で、グラフ研磨を用いずに、極大クリークを列挙した場合の予測精度は65.35%であった。したがって、グラフ研磨を行うことにより少ない説明変数で予測精度を約5%改善することができた。これは、グラフ研磨によりノイズを除去することで予測精度を高めるマイクロクラスタが生成できたためと推察する。

また表4は、モデルのクラス別予測精度を評価するための適合率、再現率、F 値を示している。グラフ研磨ありのF 値は、グラフ研磨なしに比べていずれも高くなっており、特に健康志向におけるF 値は顕著に異なる。

次に、提案手法によって得られたモデルの内容について考察を行う。興味深い内容を持つ変数を表5、6に示す。表5は回帰係数がマイナス、すなわち非健康志向に特徴的な変数である。これらの変数は非健康志向を分類するために寄与している変数である。つまり、本来食品を購入するスーパーマーケットでは、コーラやインスタント袋麺のような不健康の代名詞となるものを購入し、その一方で、コンビニエンスストアや自動販売機では、食品や日用品を購入するという行動をとっている。これらのことから、非健康志向の顧客が店舗を選択する要因は、鮮度や品質よりも利便性を重視していると考えられる。

一方で表6は回帰係数がプラス、すなわち健康志向

表5 非健康志向のマイクロクラスタ (抜粋)

マイクロクラスタ	係数
その他スーパー-コーラ	-0.194018
その他スーパー-インスタント袋麺	-0.198015
NEWDAYS-食品	
その他屋外の自販機-食品	
サンクス-食品	
セブンイレブン-日用品	
デイリーヤマザキ-食品	-0.262028
ミニストップ-食品	
住宅街の道路沿いの自販機-食品	
家電量販店-文化用品	
楽天市場-文化用品	
職場(オフィス)の自販機-食品	
ファミリーマート-飲料・酒類	-0.26732
ローソン-飲料・酒類	-0.26732

表の横罫線は1つのマイクロクラスタを表している。

表6 健康志向のマイクロクラスタ (抜粋)

マイクロクラスタ	係数
ローソン-生菓子	0.169906
その他スーパー-スープ	0.181586
その他スーパー-ホームメイキング材料	
その他スーパー-ラッピングフィルム	
その他スーパー-水	0.194462
その他一般小売店-その他農産	
その他スーパー-ビール	0.218102
その他スーパー-マカロニ	
サンドラッグ-衣料用洗剤類	0.226391
クリエイト-菓子	0.357944
その他ホームセンター-食品	
セイジョー-日用品	
ダイエー系-日用品	0.378628
L-楽天市場-日用品	
生協の個人宅配-食品	
その他スーパー-乳製品	0.520824
その他スーパー-加工肉類	

表の横罫線は1つのマイクロクラスタを表している。

に特徴な変数を示している。特徴的な商品は、ホームメイキング材料、ラッピングフィルムなどの手作りを連想させる商品や、健康の代名詞となる乳製品、水などである。そしてこれらの商品をスーパーで購入し、日用品をセイジョー、または楽天のお取り寄せなどで購入している。さらに、食品を生協の個人宅配を利用して購入するなど、利便性よりも品質や素材を重視した購買が確認できる。また、ローソンではスナックではなく、生菓子を購入しているなど、コンビニエンスストアを利用する場合も非健康志向とは異なる内容のマイクロクラスタが出現していた。

## 5. おわりに

本研究では、グラフ研磨を用いたマイクロクラスタを分類問題に利用しその有効性を検証した。そして、健康志向と非健康志向に特有の店舗選択の要因となる商品の違いを明らかにした。そこから、非健康志向の消費者は日用品を購入する際には利便性を重視した店舗選択を行っており、健康志向の消費者は水や乳製品のような健康を意識した商品の購入には、質や素材を重視した店舗選択を行っているという知見が得られた。また、マイクロクラスタを分類モデルの説明変数に利用することで、予測精度が改善できることを示し、グラフ研磨は予測問題に有効であることを明らかにした。このように提案手法を用いることで、特定の志向を持つ顧客によって、商品を購入する際にイメージされる店舗を把握することが可能である。

今後の課題は、他の志向に対するモデル化を実施することで、健康志向や非健康志向以外の異なる志向における本手法の有効性を確認することである。そして獲得した知見を評価し、最終的にはマーケティング施策を展開できるように研究を進めていきたいと考えている。

**謝辞** 大阪大学産業科学研究所の河原吉伸准教授からは有意義な情報と適切なコメントをいただいた。(株)KSK アナリティクスの北島聡氏はデータの可視化について実験していただいた。ここに感謝の意を表します。また、本研究の一部は、科学技術振興機構CREST、及びERATO 湊離散構造処理系プロジェクト、文部科学省の科研費若手研究(B)4730375の研究助成を受けている。

## 参考文献

- [1] 上田隆穂, “地域内複数店舗における店舗選択及び売場等部門別評価要因検討,” 学習院大学経済論集, **25**, 63-92, 1988.
- [2] 高橋郁夫, 『消費者購買行動—小売マーケティングへの画像—』, 千倉書房, 1999.
- [3] D. Morschett, B. Swoboda and T. Foscht, “Perception of store attributes and overall attitude towards grocery retailers: The role of shopping motives,” *The International Review of Retail, Distribution and Consumer Research*, **15**, 423-447, 2005.
- [4] 渡辺亮, 北村裕人, 星野直人, 関庸一, “買回りタイプによる顧客購買行動の理解,” *オペレーションズ・リサーチ*, **50**, 644-653, 2005.
- [5] 土井千章, 小西哲平, 中川智尋, 片桐雅二, 稲村浩, 太田賢, “店舗クラスタを用いた購買行動に関するライフスタイルの推定,” *DEIM Forum 2014 F3-5*, 2014.

- [6] M. Girvan and M. E. J. Newman, “Community structure in social and biological networks,” *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 7821–7826, 2002.
- [7] G. Karypis and V. Kumar, “Multilevel  $k$ -way partitioning scheme for irregular graphs,” *Journal of Parallel and Distributed Computing*, **48**, 96–129, 1998.
- [8] G. Palla, I. Derényi, I. Farkas and T. Vicsek, “Uncovering the overlapping community structure of complex networks in nature and society,” *Nature*, **435**, 814–818, 2005.
- [9] 宇野毅明, 中原孝信, 前川浩基, 羽室行信, “データ研磨によるクリーク列挙クラスタリング,” 情報処理学会アルゴリズム研究会報告書, 2014-AL-146(2), pp. 1–8, 2014.
- [10] J. Friedman, T. Hastie and R. Tibshirani, “Regularization paths for generalized linear models via coordinate descent,” *Journal of Statistical Software*, **33**, 1–22, 2009.