情報処理学会論文誌 Vol.58 No.2 298–307 (Feb. 2017)

購入商品レベルでの購買行動に着目したライフスタイルの推定

土井 千章[1,2,a)] 片桐 雅二[1] 太田 賢[1] 重野 寛[2]

受付日 2016年5月19日, 採録日 2016年11月1日

概要：詳細な商品購入情報を活用して，消費者のライフスタイルを高精度に推定する手法を提案する．ライフスタイルと商品購入情報との関連性は，大規模調査パネルによる商品購入情報とアンケート結果から，機械学習手法であるニューラルネットワーク法を用いて獲得し，ライフスタイル推定に利用する．特徴が現れやすい商品種類の購入情報を適応的に選択するとともに購入商品の詳細識別（JAN コード）を用いることによって，高精度にライフスタイルが推定できる．これにより，消費者のライフスタイルに合った商品のレコメンデーション等の高度なサービスをより的確に提供できる．

キーワード：ライフスタイル推定，購買履歴，ニューラルネットワーク

Lifestyle Prediction by Capturing Micro Purchasing Behavior

Chiaki Doi[1,2,a)] Masaji Katagiri[1] Ken Ohta[1] Hiroshi Shigeno[2]

Received: May 19, 2016, Accepted: November 1, 2016

Abstract: This paper proposes a method to estimate consumer's lifestyle by focusing on micro purchasing behavior. The method acquires relevance model for each product category between lifestyles and purchasing histories with detailed product ID (JAN code) beforehand based on consumer panel survey. The model acquisition employs Neural Network, a machine learning method. Appropriate relevance models are adaptively selected for each lifestyle and combined to achieve optimal performance on lifestyle prediction. The method is useful to realize smart recommendation such as suggesting products with reflection of each consumer's lifestyle.

Keywords: lifestyle prediction, purchasing behavior, neural network

1. はじめに

近年，ID-POS（Point of Sales）システムの普及により，消費者の詳細な購買履歴が自動的に収集・蓄積可能となった．これらの ID-POS システムから取得した購買履歴と年齢や性別等の顧客属性データの分析結果に基づいた経営戦略が多くの企業によって取り入れられている．さらに電子マネーやクレジットカード，ポイントカード等の普及により，多業種間での購買履歴情報を取得する試みが行われている．これにより，各店舗やグループ店舗内の局所的な消費行動だけではなく，消費者が訪れた複数店舗での消費行動が網羅的に把握できるようになりつつある．ネット通販事業者の楽天株式会社 [1] では，これらの複数店舗での個人の消費行動に着目し，顧客属性データや店舗での購買履歴等を用いて消費者をいくつかのグループに分類してバナーの出し分け等を行っている [2]．ポイントカードの Ponta [3] や T ポイント [4] は，複数店舗でのポイントカード利用履歴から潜在顧客の可視化や業種を横断した相互送客を行っており [5], [6]，消費者の購買履歴に応じたアプローチが個別に行われるようになった．これらの購買履歴情報を用いた商品や店舗等のレコメンデーションは様々な業態で実施されている．しかし，顧客属性データと購買履歴情報のみを用いて消費者の趣味や価値観等の内面的な属

1 株式会社 NTT ドコモ
NTT DOCOMO, INC., Yokosuka, Kanagawa 239–8536, Japan
2 慶應義塾大学
Keio University, Yokohama, Kanagawa 223–8522, Japan
a) chiaki.doi.tf@nttdocomo.com

性を推定することには限界がある．

本論文では，サイコグラフィック属性の 1 つであるライフスタイルに着目する．サイコグラフィック属性は消費者の心理的な内面性を表しており，消費者のセグメンテーションのための基準として，デモグラフィック属性よりも適していると報告されている [7]．ライフスタイルは，消費行動と相互に関係するといわれており，ライフスタイルによって購買行動が異なることが実証されている [8], [9], [10]．本論文においては日用品の購買行動に関連するライフスタイルに着目し，消費者の性格や購買意識に関するアンケート結果から抽出した 5 つの特徴をライフスタイルと定義する．このライフスタイルを用いることによって，より高度なターゲティングが実現できると考えられる．

消費者のライフスタイルは，アンケートやインタビューを通じて取得して用いるのが一般的である．しかし，すべての消費者に対してアンケートやインタビューを実施することは困難である．そのため，ライフスタイルと購買履歴情報の関係をモデル化することで，消費者が購入した商品からライフスタイルの推定を行うアンケートやインタビューの実施が不要な手法が提案されている [11]．この研究では，購買履歴に含まれる商品数が膨大であることから「野菜」,「魚」等の商品の種類（以降，商品種類と呼ぶ）別購入数や曜日別の来店数等購買行動を特徴づけると思われるいくつかのパラメータを算出し，これによりデータのモデル化を行っており，JAN コード [12] で区別される商品（以降，商品と呼ぶ）レベルの購買情報は扱っていない．しかしながら，消費者のライフスタイルにより購入するものに現れる差は,「牛乳」,「納豆」といった商品の種類の違いや，また同じ牛乳であっても「A 社の牛乳 A」や「A 社の牛乳 B」といった商品の違いに現れること，さらには，個々のライフスタイルによっても特徴が現れる商品種類/商品やその強さは異なることが考えられる．そのため，本論文では商品種類レベルと商品レベルの購買情報を用いてライフスタイルの推定を行う．

本論文では，消費者のライフスタイルが強く現れる商品に関する購買行動を抽出して用いることで，ライフスタイルを高い精度で推定する手法を提案する．具体的には，商品種類ごとに商品単位レベルの購買情報を用いてライフスタイル推定を行う個別モデルを構築し，各ライフスタイルについて推定能力が優れる個別モデル（商品種類）を選択し，それらを組み合わせて結果を得る．本手法は，筆者らの途中経過報告である個々の商品に着目した購買行動からのライフスタイル推定の検討結果 [13], [14] をベースとし，ライフスタイルの推定モデル構築方法および多数ある商品種類からライフスタイルが強く現されている推定モデルの選択方法を拡張したものである．本手法の推定能力を示すため，先行研究，商品種類単位のみの推定と提案手法である商品種類を組み合わせた場合の推定能力を，適合率と再現率の調和平均である F 値および全消費者におけるライフスタイルの推定精度を示す正解率を用いて評価する．

以降 2 章で関連研究，3 章では本論文で使用するデータについて述べる．4 章ではライフスタイルの定義，5 章ではライフスタイルを推定する手法について説明する．6 章で評価，7 章でまとめについて述べる．

2. 関連研究

ライフスタイルは，消費行動と相互に関係するといわれている．アンケート等から消費者のライフスタイルを分類し，ライフスタイルの違いによる消費者の行動の違いを明らかにして活用することを目的とした研究がすでに報告されている [8], [9], [10], [15]．一例として Buckley ら [10] は，購買行動に関連した質問項目に対する回答結果を用いて消費者をインスタント食品に特化したライフスタイルへ分類し，食や購買に関する意識の違いやインスタント食品を購入する動機や行動の違いを明らかにした．

また，ライフスタイルと購買行動をモデル化する研究がすでに行われている [11], [16], [17], [18]．Ishigaki ら [17] は，アンケート結果の因子分析によりライフスタイルを定義し，購買履歴情報と紐づけて顧客モデルの構築を行い，購買予測を行う手法を提案した．石垣ら [18] は，アンケート結果から得たライフスタイルと購買履歴情報を紐づけて，顧客のライフスタイルごとの来店状況（曜日，雨量，気温等）に着目して各日の流通量販店への来店人数を予測する来店人数予測モデルの提案を行った．

小柴ら [11] は，消費者の購買履歴のみから，アンケート結果の得られない消費者についてライフスタイルを推定する手法を提案し，その推定精度を評価している．ライフスタイルと購買履歴情報の関係をモデル化することで，ライフスタイルごとの購買行動が明らかになり，購買行動からライフスタイルを推定することができる．しかしながら，購買履歴を商品種類単位で要約してライフスタイルの推定を試みており，商品単位の購買履歴を活用したライフスタイルの推定は行われていない．

3. データ概要

本論文では，マーケティング活用のために商用提供する目的で調査会社が調査モニタを募って取得しているデータを使用する．これらのデータには，商品購入情報のほかに各モニタの属性情報としてデモグラフィック情報，性格や購買意識に関するアンケート情報が含まれている．本論文では，このデータから 7,023 人のデモグラフィック情報，アンケート情報，2012 年 1 月 1 日から 2012 年 12 月 31 日までの 1 年間の商品購入情報を用いる．対象者の居住地は，一部地域ではなく，日本国内の全都道府県である．

デモグラフィック情報としては，性別，年代，職業，居住地等 38 種類の属性が用意されているが，本論文では性

表 2　各ライフスタイルの属性分布（人）
Table 2　Age distribution of lifestyle types.

	10 代	20 代	30 代	40 代	50 代	60 代	因子別合計
第 1 因子　こだわり消費派	72	368	695	749	542	356	2782
第 2 因子　節約消費派	19	321	526	578	433	203	2080
第 3 因子　栄養バランス重視派	18	154	318	427	294	195	1406
第 4 因子　新商品消費派	1	47	119	147	79	25	418
第 5 因子　アクティブ消費派	5	57	109	103	45	18	337
合計	115	947	1767	2004	1393	797	7023

表 1　対象者の属性分布（人）
Table 1　Characteristics of participants.

	男性	女性	年代別合計
10 代	59	56	115
20 代	450	497	947
30 代	837	930	1767
40 代	1007	997	2004
50 代	733	660	1393
60 代	509	288	797
合計	3595	3428	7023

別と年代のみを用いる．対象者の属性分布（性別，年代）を**表 1** に示す．

アンケート情報は，約 1,600 問の考え方や行動に関する質問の回答結果である．質問項目には，“カロリーに気を使った食生活を心掛けている”，“栄養バランスに気を使った食事をしている” 等の生活についての意識や “新しいことにどんどんチャレンジしていきたい”，“ガヤガヤと騒がしいところにいるのは嫌い” 等の性格についての質問項目が含まれている．

商品購入情報は，消費者 ID，購入を行った日時，商品種類，JAN コード，個数，値段，店舗の組によって構成されるレコードの集合である．1 レコードは，1 商品の購入 1 回に対応している．本論文で扱う商品は，生鮮食品を除いた日用消費財（Fast Moving Consumer Goods）とし，286 種類の商品種類に分類されており，商品数（JAN コード数）は，193,601 である．この商品購入情報には，対象の商品については実店舗，通販等を含めすべての購買が網羅的に記録されている．

また，ここで店舗とは，商号（ブランド名）のみであり，どの店舗かを表す店舗名を含んでいないことに注意する．たとえば，コンビニエンスストア A 横須賀店とコンビニエンスストア A 溜池山王店は，双方ともコンビニエンスストア A として扱われる．店舗の業態は，スーパー，コンビニエンスストア，ホームセンター・ディスカウントストア，薬局・ドラッグストア，酒のディスカウントストア，百貨店・デパート，自販機，宅配・通販，100 円・99 円ショップ，家電量販店，パン・菓子店，販売店（KIOSK 等），酒屋，大学生協であり，日用消費財を扱うおおよその店舗をカバーしている．

4.　ライフスタイル

本論文においては，石垣ら [16] の提案と同様に，3 章で述べた消費者の性格や購買意識に関するアンケート結果から因子分析を用いて特徴軸を導出し，これをライフスタイルの構成要素と考えることにする．本来は，利用価値が高いライフスタイルを得られることが望ましい．しかし，本論文で扱うライフスタイルは，アンケートの設問と回答から数理的に得られるものであり，設問の内容や回答によって得られるライフスタイルは変化し，利用価値の高いライフスタイルにつねに一致するものではないことに注意する．

具体的には，3 章で説明したデータに含まれるアンケート情報（約 1,600 問）から，文献 [16] で使用された 20 問の質問項目と類似した 17 問を抽出し，これらに対する回答結果を用いて因子分析を行った．回答は “あてはまる”，“まああてはまる”，“どちらでもない”，“あまりあてはまらない”，“あてはまらない” の多肢選択式で，各選択肢に対して 5，4，3，2，1 の得点を順序尺度として与えている．文献 [16] では，“家計簿をつけている”，“スーパーでの買い物は出来るだけ早く済ませたい”，“本店でしか買わない商品がある” の質問項目があるが，本論文では類似した質問項目がなかったため，該当する要素を加味していない．

因子分析は，バリマックス法を用いて共通因子を抽出した．また，因子数の決定には，スクリー基準を用いて 5 つに決定した（因子分析の結果については付録 A.1 を参照のこと）．消費者は多様な特徴を複合的に持っていると考えられるが，ここでは単純化し理解を容易にするために，文献 [16] 同様，特徴が最も現れた 1 つの軸でその消費者のライフスタイルを代表することとする．各ライフスタイルの特徴を，因子負荷量を用いて主観的に命名したものを所属する消費者の属性分布（年代）とともに**表 2** に示す．

5.　購買情報によるライフスタイルの推定

本章では，消費者のライフスタイルが強く現れる商品に関する購買行動を抽出して用いることで，ライフスタイルを高い精度で推定する手法を提案する．

5.1 提案概要

ライフスタイルの違いが商品の購入に現れる程度は商品種類や商品によって異なるため，すべての商品購入情報を用いると種々雑多な商品購入情報が混在し，ライフスタイルによる違いが埋没してしまうのではないかと筆者らは考えた．また，個々の商品購入情報に含まれる購入された時間帯や曜日，店舗，価格にもライフスタイルが現れる可能性がある．そこで，消費者の 1 来店における商品種類ごとの商品購入情報に着目してライフスタイルの推定に有効な特徴を抽出して用いることができれば，消費者のライフスタイルを高精度に推定できる可能性があると考えた．

本論文では商品種類ごとに商品単位の商品購入情報をデータセットの 1 レコードとして扱い，各商品種類が各ライフスタイルを現す程度を明らかにする．そのうえで各ライフスタイルについて推定能力が優れている商品種類の商品購入情報のみを抽出してライフスタイルを推定する手法を提案する．ここで，同時に購買される商品との関連性よりも，商品個々にライフスタイルは現されやすいと考え，単一の商品単位を 1 レコードとして扱うことにする．**図 1** に提案手法の概要を示す．本提案手法は，ライフスタイルを推定するモデルを商品種類ごとに構築し，推定精度が高いモデルを自動選択する事前分析フェーズと選択されたモデルを用いて推定を行う推定フェーズによって構成される．各フェーズの詳細については，それぞれ 5.4 節，5.5 節に示す．

なお，本論文で扱ったライフスタイルは，4 章で述べたように消費者の性格や購買意識に関するアンケートの結果を用いて導出されたものである．このライフスタイルは普遍的な指標ではなく，現実的にはレコメンドを行う商品やサービスに応じて所望の利用価値が高いライフスタイルが得られるようにアンケートの設計を行う必要がある．本提案手法は，目的のライフスタイルに応じて推定に有効な商品種類を適応的に抽出して用いるため，任意のライフスタイルに対応可能であり，様々なマーケティング用途に応用できることが期待される．

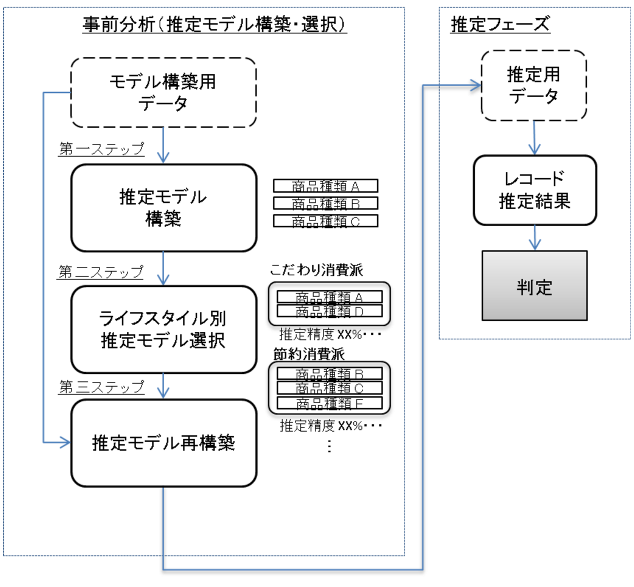

図 1 ライフスタイル推定手法
Fig. 1 Lifestyle prediction method.

5.2 データフォーマット

ライフスタイルの推定に用いるモデル構築用データおよび推定用データは，3 章で述べた商品購入情報から，必要に応じて購買状況レコードを抽出して用いる．購買状況レコードは，消費者の ID，購入日時，商品の JAN コード，購入個数，単価，店舗 ID の情報を持つ．**表 3** に本論文で使用するデータ項目，次元数と値を示す．抽出した購買状況レコードは，下記の変数の形に変換してモデル構築用データまたは推定用データとして用いる．1 購買状況レコードは，1 回の購買行動における 1 商品に関する情報であることに注意する．また，表 3 中の JAN コード数と店舗 ID 数は，抽出した購買状況レコードに現れる数とする．

5.3 評価尺度

本論文では，推定モデルの推定能力を評価する尺度として適合率，再現率，F 値，正解率を用いる．商品種類 p のモデルによるライフスタイル l の適合率 $P(l,p)$ は式 (1) を用いて算出する．ここでライフスタイル l は，$l \in${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} とする．$T(l,p)$ は商品種類 p の商品購入情報を保持している消費者のうち，ライフスタイル l と推定された人数，$S(l,p)$ は商品種類 p の商品購入情報を保持している消費者のうち，ライフスタイル l と推定され，かつ実際にライフスタイル l であった人数とする．再現率 $R(l,p)$ は，式 (2) を用いて算出する．商品種類によっては，全消費者ではなく一部の消費者にのみ商品が購入されている．本論文では，再現率は購入者のみではなく全消費者数を考慮して算出する．すなわち，$M(l)$ は推定用データに含まれるライフスタイル l の人数である．適合率 $P(l,p)$ と再現率 $R(l,p)$ の調和平均である F 値 $F(l,p)$ は，式 (3) を用いて算出する．全消費者における商品種類 p の

表 3 ライフスタイルの推定に用いるデータ
Table 3 Data items for predicting the life style.

データ項目	次元数	値
JAN コード	(JAN コード数)	指示関数
購入個数	1	スカラー
単価	1	スカラー
旬 （上旬，中旬，下旬)	3	指示関数
曜日	7	指示関数
時間帯 (1 時間単位)	24	指示関数
店舗 ID	(店舗 ID 数)	指示関数

正解率 $A(p)$ は，正しくライフスタイルが推定された消費者数 N_p と全消費者数 O から式 (4) を用いて算出する．

$$P(l,p) = S(l,p)/T(l,p) \quad (1)$$

$$R(l,p) = S(l,p)/M(l) \quad (2)$$

$$F(l,p) = 2 * P(l,p) * R(l,p)/(P(l,p) + R(l,p)) \quad (3)$$

$$A(p) = N_p/O \quad (4)$$

$$N_p = \sum_l S(l,p) \quad (5)$$

$$O = \sum_l M(l) \quad (6)$$

5.4 事前分析

図 1 に示すように，事前分析は，商品種類別推定モデル構築，ライフスタイル別商品種類選択，ライフスタイル別推定モデル再構築の 3 つのステップで構成される．

第 1 ステップの商品種類別推定モデル構築では，機械学習の一手法である階層型ニューラルネットワーク [19]（以降，NN）を用いて購買状況レコードを入力としライフスタイルを出力する推定モデルを商品種類ごとに構築する．推定モデルの構築に NN を採用した理由は，筆者らの途中経過報告であるランダムフォレスト法を用いたライフスタイルの推定結果 [14] と比較して推定精度が高かったためである．

式 (8) で示される購買状況レコードごとに NN から推定される結果 B は，ライフスタイル l への所属確率 b_l を要素とするベクトルである．ここで，式 (7) を用いて消費者 u の商品種類 p におけるライフスタイル l の所属確率 b_l の平均値を求め，A_{up} を算出する．rn_{up} は，消費者 u の商品種類 p におけるレコード数とする．L は，ライフスタイルの総数である．また B_{upi} は消費者 u の商品種類 p に関する i 番目の購買状況レコードに対する NN からの出力結果である．そして A_{up} が要素として持つ各ライフスタイルの所属確率の平均値のうち，最も所属確率が高かったライフスタイルをその消費者のライフスタイルとする．

$$A_{up} = \frac{1}{rn_{up}} \sum_{i=1}^{rn_{up}} B_{upi} \quad (7)$$

$$B = (b_1, \cdots, b_L) \quad (8)$$

NN の学習方法はバック・プロパゲーション法を用い，隠れ層は 1 層とし，隠れ層のユニット数は入力層のユニット数と同じとする．各入力信号の次元数（入力層のユニット数）は，表 3 に示すとおりである．学習時のデータの重みはすべて同じ重みとした．結合荷重の増大を抑制するため，目的関数に付加する weight decay 項の結合荷重の初期値には乱数を設定した．隠れ層の活性化関数にはシグモイド関数を用い，出力層の活性化関数としてソフトマックス関数を用いる．推定モデルの性能評価には，モデル構築用データに対して 9 割の消費者のデータを学習データ，1 割の消費者のデータを評価データとした 10 分割交差検証を行う．各推定モデルの評価尺度として，5.2 節で述べた F 値 $F(l,p)$ を算出する．

第 2 ステップのライフスタイル別商品種類選択では，第 1 ステップで構築された商品種類ごとの推定モデルからライフスタイルごとに推定精度が高い複数の推定モデルを選択し，組み合わせることで能力の高い推定モデルを構築する．ここでは，推定モデルの各ライフスタイルにおける F 値 $F(l,p)$ と推定モデルが出力した消費者ごとの平均所属確率 A_{up} を用いる．また，複数の商品種類のモデルを組み合わせるとは各商品種類で得られた A_{up} の平均を用いることを意味する．

ライフスタイル別商品種類選択では，貪欲前向き選択を用いる．今，ライフスタイル l ごとに F 値 $F(l,p)$ 上位 n 商品種類のモデルを組み合わせ，ライフスタイルの推定を行い，その F 値を得ることを考える．n を 1 から始め，F 値を観測しておく．次に n を 1 増加させる．このとき F 値が向上または等しかった場合はさらに n を増加させる．もし F 値が低下したなら，増加させる前の n を用いた上位 n 番目までの商品種類を組み合わせた推定モデルを，そのライフスタイル推定向けモデルとして採用し，結果の商品種類の集合を P_l と表すことにする．上記の手順をすべてのライフスタイルにおいて行う．

第 1 ステップでは，全データに対して 9 割のデータをモデルの構築に使用している．第 3 ステップのライフスタイル別推定モデル再構築では，第 2 ステップで採用されたライフスタイルごとの商品種類においてすべてのモデル構築用データを学習データとし，購買状況レコードを入力としてライフスタイルを出力する推定モデルを構築する．

5.5 推定フェーズ

図 1 に示した提案手法の推定フェーズについて述べる．推定フェーズでは，ライフスタイルの推定を行いたい消費者の商品購入情報を推定用データとし，5.3 節の事前分析で構築されたライフスタイル別の推定モデルを用いて推定を行う．すなわち，消費者 u についてライフスタイル l への所属確率の推定は次のように行う．消費者 u による商品種類 $p \in P_l$ のレコードを抽出し，各 p について式 (7) により A_{up} を得て，式 (9) を用いてそれらの平均を得る．これを C_{ul} と表す．消費者ごとに，C_{ul} において所属確率が最も高かったライフスタイルと，l が一致した場合に，l を消費者のライフスタイルとする．上記の手順をすべての消費者 u とライフスタイル l の組合せに対して行う．このとき，ライフスタイルの推定は，ライフスタイルごとに行われるため，1 人の消費者が複数のライフスタイルに所属する結果となることがある．その場合は，該当する C_{ul} を用いて所属確率が最も高いライフスタイルに所属することとする．所属確率が同値であった場合には，推定されたライ

フスタイル数で1を除算した数を消費者数としてカウントし評価することにする．

$$C_{ul} = \frac{1}{|P_l|} \sum_{p \in P_l} A_{up} \quad (9)$$

6. 評価

本章では，先行研究と5章で述べた手法の推定能力の比較を行い，提案手法の有効性について評価を行う．

6.1 先行研究の推定精度

本論文の目的は，JANコードで区別されるような個々の商品レベルに着目し，消費者のライフスタイルを推定することである．文献[11]では，商品数が膨大であることから購買行動を特徴付けると思われるいくつかのパラメータを算出してデータのモデル化を行い，ライフスタイルの推定を行っている．ライフスタイルの推定を行う手法として，文献[11]の手法が適用できることから，3章で説明したデータを用いて評価し，提案手法の有用性を確認するための比較対象とする．推定精度は扱うデータによって異なると考えられるため，本論文で使用したデータを用いて先行研究の提案手法を評価する．ただし，先行研究[11]では，商品購入情報を人単位で集計して扱っており，個々の購買履歴は使用していないことに注意する．また，先行研究[11]では単一店舗における購買履歴の利用が前提であるが，本論文で扱うデータには複数店舗の購買履歴が含まれているため，購入した店舗を示す店舗IDを用いて店舗ID別来店回数を算出し，複数店舗の購買を反映させるように拡張を行う．

先行研究の評価に用いるデータを**表4**に示す．時間帯ごとの来店回数は，文献[11]では15分単位であるが，この購買情報には24時間営業の店舗も含まれるため，この評価では，1時間単位で集計するものとした．

ライフスタイルの推定は，文献[11]と同様に，ランダムフォレスト法[20]を用いて行う．評価は，10分割交差検証法を用いて行い，F値と正解率を確認する．交差検証においては，9割の消費者のデータをモデル構築，1割の消費者のデータを評価に使用した．図2のPriorに上記で説明した方法でライフスタイルを推定した結果を示す．37.6%の正解率でライフスタイルを推定できることを確認した．

表4　先行研究の評価に用いるデータ

Table 4　Data items for predicting the life style.

データ項目	次元数	値
店舗ID別来店回数	485	スカラー
来店回数（曜日別）	7	スカラー
来店回数（1時間ごと）	24	スカラー
商品購入金額（総額）	1	スカラー
商品購入数（総数）	1	スカラー
食品購入金額（総額）	1	スカラー
食品購入数（総数）	1	スカラー
商品種類別商品購入数（総数）	286	スカラー

6.2 提案手法の推定精度

本節では，まず提案手法の動作を確認するため，事前分析の第1ステップにおける商品種類単位のみでの推定能力について示す．次に提案手法の推定能力を明らかにし，その有効性を示す．

6.2.1 商品種類単位での推定能力

本項では，商品種類個々の商品購入情報のみでライフスタイルを推定した際の推定精度と推定に使用したNNによる推定モデルの学習精度を示し，商品種類単位のみでの推定能力を明らかにする．

商品種類個々の商品購入情報のみで（式(7)に示した平均所属確率 A_{up} を用いて）推定した際のライフスタイルの推定精度を示す．図2では，各商品購入情報のうち，正解率が高い上位5商品種類を用いた場合の結果を示している．商品種類を単体で用いた場合の正解率は「液体茶」は41.1%，「コーヒードリンク」は40.3%であり，先行研究「Prior」より高いことを確認した．また，ライフスタイルごとに単一の商品種類を選択することで「こだわり消費派」，「節約消費派」，「新商品消費派」や「アクティブ消費派」において先行研究を上回る能力で推定できる商品種類が明らかになった．ただし，「栄養バランス重視派」においてはF値が先行研究よりも低くなることがあわせて確認された．

表5にライフスタイルごとのF値が高い商品種類を示す．こだわり消費派であれば「コーヒードリンク」や「液体茶」，節約消費派であれば「豆腐類」，「食パン」等ライフスタイルごとに異なる特徴を確認した．また，ライフスタ

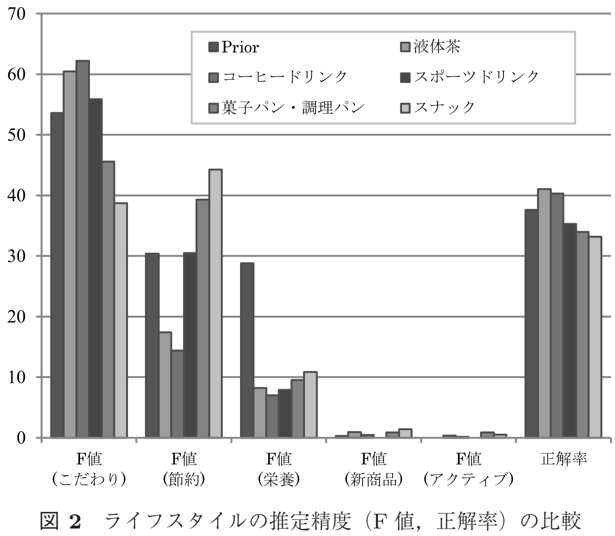

図2　ライフスタイルの推定精度（F値，正解率）の比較

Fig. 2　Comparison of lifestyle prediction accuracy (F-measure, Accuracy rate).

表 5　各ライフスタイルの推定に適した商品種類と推定精度（F 値）
Table 5　Optimal product types and accuracy rate of each life style (Accuracy of learning, Accuracy, F-measure).

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	
商品種類	F 値(%)	商品種類	F 値(%)	商品種類	F 値(%)	商品種類	F 値(%)	商品種類	F 値(%)
コーヒードリンク	62.2	豆腐類	48.9	チーズ	27.9	リップクリーム	5.8	その他雑貨	6.1
液体茶	60.5	食パン	48.3	サラダ油・天ぷら油	24.0	皮膚用薬	5.0	味噌汁・吸物類	4.5
スポーツドリンク	55.9	生麺・ゆで麺	47.3	漬物	24.0	たわし・スポンジ	5.0	殺虫剤	4.1
果汁飲料	52.0	ヨーグルト	45.5	柔軟剤	23.9	口紅	4.8	合わせ調味料	3.9
炭酸飲料	51.1	牛乳	45.4	ラッピングフィルム	23.4	ハンド&スキンケア	4.7	中国茶	3.9
コーラ	49.6	冷凍調理	44.4	その他汎用調味料	23.2	その他雑貨	4.7	検査薬	3.7
ミネラルウォーター	49.2	スナック	44.3	納豆	23.0	眉目料	4.5	その他缶詰	3.5
紅茶ドリンク	48.9	中華風食品	43.9	香辛料	23.0	シロップ類	4.4	麦茶	3.2
栄養ドリンク	47.2	納豆	43.8	畜肉ハム	23.0	住居用クリーナー	4.4	エチケット品	2.9
機能性ドリンク	45.8	ビスケット・クラッカー	43.6	豆腐類	22.6	ファンデーション	4.3	キャンディ	2.8

イルによって F 値の最大値に差が見られた．これらの結果は，ライフスタイルによって推定に適した商品種類とそうではない商品種類があることを表している．

ライフスタイルによる推定精度の違いは，各ライフスタイルに所属する消費者数の偏りが影響を与えていることも考えられる．そのため，推定モデル構築に使用した学習データの消費者数を全ライフスタイル同数にして推定モデルの構築と評価を行った．その結果，推定精度が低いアクティブ消費派や新商品消費派の F 値が向上することを確認した．これから，消費者数の偏りが推定精度に影響を与えている可能性があると考えられる．また，消費者数を同数にしなかった場合と比較してライフスタイルの推定能力の高い商品種類がほぼ同じ商品種類であったことから，各ライフスタイルの特徴が現れやすい商品種類が消費者の偏りが存在しても，おおむね選択できていることを確認した．

次に NN によって商品種類別推定モデルがどの程度学習できているかを明らかにする．NN 単体の評価として，図 2 の結果を得るために用いたデータを使用し，学習終了時の学習正解率と評価データの評価正解率を比較する．正しいライフスタイルが推定されたレコード数を評価に使用したレコードの総数で除算した値を正解率とする．学習正解率は，NN 単体の評価に使用した学習データを評価に使用して算出した正解率とする．評価正解率は，評価データを評価に使用して算出した正解率とする．**表 6** に表 5 で示した商品種類の推定モデルの学習正解率と評価正解率を示す．こだわり消費派，節約消費派，栄養バランス重視派は，学習正解率が 40～60%であり，評価正解率は 30～50%であった．こだわり消費派，節約消費派，栄養バランス重視派と比較して新商品消費派とアクティブ消費派の学習正解率は高く，評価正解率は低い傾向を確認した．学習正解率と評価正解率の関連性を考察するため，学習正解率を評価正解率で除算した値（正解比率）を指標として採用する．**図 3** に商品種類別に購入者数と正解比率をプロットした図を示す．正解比率が 1.0 に近いほど NN がライフスタイルの特徴を学習できていることを示し，正解比率が低い場合は過学習が疑われる．F 値の低い商品種類は正解比率も低い傾向が見られ，推定精度と正解比率の関連性を確認した．また，これらの正解比率は購入者が多い商品種類ほど 1.0 に近い傾向が見られた．正解比率の値が低い商品種類については，ライフスタイルの特徴を学習するのに十分なデータ量が得られていないかライフスタイルの特徴が現れにくい商品種類であることが考えられる．今後データ量が増加することによって，正解比率が向上する可能性がある．

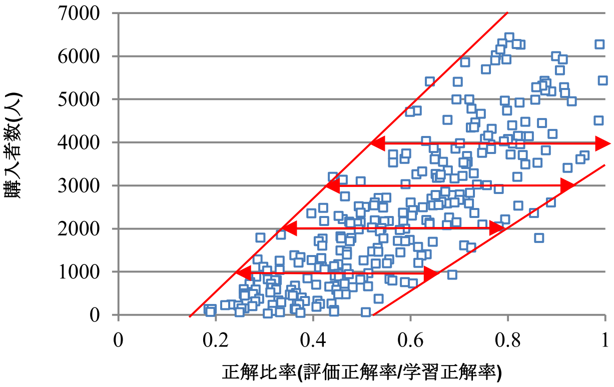

図 3　購入者数と正解比率
Fig. 3　Number of purchaser and accuracy ratio.

6.2.2　提案手法の推定能力

図 4 にライフスタイルの推定精度が高い商品種類のみを用いてライフスタイルを推定する提案手法の推定精度を示す．事前分析のライフスタイル別推定モデル選択を行った結果，「こだわり消費派」は 5 商品種類，「節約消費派」は 5 商品種類，「栄養バランス重視派」は 4 商品種類，「新商品消費派」は 6 商品種類，「アクティブ消費派」は 5 商品種類が採用され，これらを組み合わせたモデルを使用した．提案手法の正解率は 44.0%であり，各ライフスタイルに対して適切な商品種類を選択することによって正解率が向上することを確認した．

また，先行研究の手法や単一商品種類を用いた場合に，

表 6　推定モデルの推定結果（学習正解率，評価正解率）
Table 6　Result of the prediction model (Learning accuracy rate, Evaluation accuracy rate).

こだわり消費派

商品種類	学習正解率(%)	評価正解率(%)
コーヒードリンク	59.6	52.2
液体茶	46.5	46.0
スポーツドリンク	44.8	46.2
果汁飲料	47.3	43.3
炭酸飲料	43.7	39.7
コーラ	43.0	42.4
ミネラルウォーター	43.9	39.1
紅茶ドリンク	43.3	40.4
栄養ドリンク	48.0	46.0
機能性ドリンク	46.5	44.2

節約消費派

商品種類	学習正解率(%)	評価正解率(%)
豆腐類	38.3	38.2
食パン	39.7	36.3
生麺・ゆで麺	40.8	35.9
ヨーグルト	40.6	36.5
牛乳	44.0	36.0
冷凍調理	40.7	35.7
スナック	45.0	37.2
中華風食品	42.9	38.3
納豆	38.5	35.4
ビスケット・クラッカ	42.9	33.7

栄養バランス重視派

商品種類	学習正解率(%)	評価正解率(%)
チーズ	38.5	33.1
サラダ油・天ぷら油	45.1	31.7
漬物	37.7	32.3
柔軟剤	42.5	31.1
ラッピングフィルム	43.2	32.3
その他汎用調味料	46.7	33.5
納豆	38.5	35.4
香辛料	43.4	31.3
畜肉ハム	41.2	32.7
豆腐類	38.3	38.2

新商品消費派

商品種類	学習正解率(%)	評価正解率(%)
リップクリーム	56.5	29.5
皮膚用薬	59.4	27.7
たわし・スポンジ	62.9	31.4
口紅	99.0	32.0
ハンド&スキンケア	61.3	29.1
その他雑貨	67.9	28.6
眉目料	87.8	29.4
シロップ類	60.8	27.8
住居用クリーナー	58.5	31.3
ファンデーション	90.6	26.5

アクティブ消費派

商品種類	学習正解率(%)	評価正解率(%)
その他雑貨	67.9	28.6
味噌汁・吸物類	46.2	33.1
殺虫剤	61.0	28.2
その他合わせ調味料	50.9	28.8
中国茶	64.8	33.3
検査薬	98.6	25.6
その他缶詰	59.0	25.0
麦茶	49.3	33.5
エチケット品	81.0	28.6
キャンディ	47.7	34.0

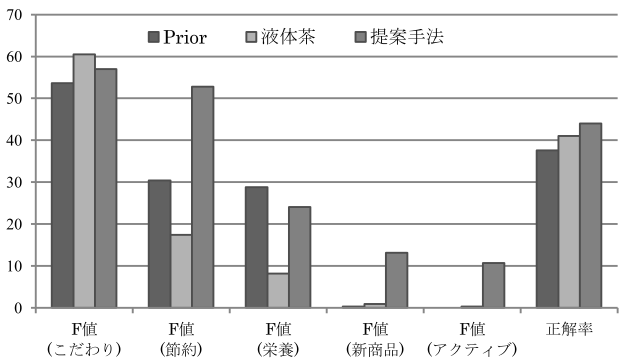

図 4　提案手法の推定精度（F 値，正解率）
Fig. 4　Lifestyle prediction accuracy of proposed method (F-measure, Accuracy rate).

「新商品消費派」と「アクティブ消費派」については著しく F 値が低く（先行研究の手法では 0.3%，最善の単一商品種類を用いても 6.1%以下）推定が困難であったが，提案手法では双方とも F 値が 10%以上と大幅に向上しており，詳細な商品購入情報を適切に利用している効果と考えられる．これらの結果より，提案手法の有効性が確認された．提案手法は，数百万人規模のユーザを対象としたレコメンデーションの実施時に適用することを想定しているため，正しいターゲティングが行えるユーザ数に大きな影響を与える正解率の数ポイントの差は大きいと考えられる．そのため，提案手法を用いることにより先行研究よりも 6.4 ポイント高い精度でターゲティングを行えることは，有用である．

本手法は，消費者のライフスタイルが強く現れる商品に関する購買行動を抽出し，多種多様な消費者のライフスタイルを推定可能にする．しかしながら，ライフスタイルごとの推定能力には差が見られ，推定精度のさらなる向上は今後の課題である．また，目的とするライフスタイルによっては，そもそも購買行動に特徴が現れにくい場合も考えられ，そのような場合には手がかりとなる新たな情報の追加利用が必要であろう．

7. おわりに

本論文では，消費者によって購入される多数ある商品からライフスタイルが強く現れる商品に関する購買行動を抽出して用いることで，ライフスタイルを高い精度で推定する手法を提案した．6 章で示したライフスタイル推定結果

より，対象とする商品種類を限定して商品に着目することで先行研究よりも高い 44.0%の正解率で推定できること，一部のライフスタイルの F 値が大幅に向上することが確認できた．

消費者のライフスタイルを理解することによって，消費者に合った商品やサービスをレコメンドする等，より高度な One to One マーケティングが実現できる．ただし，ライフスタイルは普遍的な指標ではなく，目的に応じて所望のライフスタイルが得られるようにアンケートの設計を行う必要がある．これに対して本手法は，本論文で扱うライフスタイル以外でも目的のライフスタイルに応じて，推定に有効な商品種類を自動的に抽出できるという利点がある．

今後は，各ライフスタイルのさらなる推定精度の向上を目指すため，動径基底関数ネットワークやベイジアンネットワーク等，他の手法を用いた推定モデルの構築と評価を行っていきたい．また，推定精度に影響するパラメータについても寄与度を考慮して自動的に抽出可能にしたい．

さらに推定したライフスタイルを用いて顧客へ情報配信等の施策を実施し，ライフスタイルごとに顧客の単価や購買頻度の向上に有効な特徴を明らかにしていくことも課題である．また，本論文で扱った日用消費財は，購入された数が少ない商品や新商品，生産終了品が存在する．今後，これらの商品の状態変化や商品抽出等への対応も検討していきたい．

謝辞　本研究を進めるにあたり，株式会社インテージと株式会社ドコモ・インサイトマーケティングの協力を得た．本研究で用いたデータ「i-SSP データ（インテージシングルソースパネルデータ）」に関して，株式会社インテージの協力を受けた．また，株式会社ドコモ・インサイトマーケティングから有益なコメントをいただいた．ここに記して感謝する．

参考文献

[1] 楽天株式会社，入手先 ⟨http://corp.rakuten.co.jp/⟩（参照 2016-04-11）．

[2] 森　正弥：E-commerce 企業におけるビッグデータへの挑戦と課題，電子情報通信学会第 15 回情報論的学習理論ワークショップ (2012).

[3] 共通ポイント Ponta，入手先 ⟨http://www.ponta.jp/⟩（参照 2016-04-11）．

[4] T サイト，入手先 ⟨http://tsite.jp/⟩（参照 2016-04-11）．

[5] 中村博之：データサイエンティストの虚像と実像―データ分析をビジネスに結びつけるための人材と体制，野村総合研究所 IT ロードマップセミナー AUTUMN2013 (2013).

[6] 客ごとにクーポン発行，スリーエフ T カードの履歴分析，日経 MJ（流通新聞），入手先 ⟨https://messe.nikkei.co.jp/rt/news/112860.html⟩（参照 2016-04-11）．

[7] Straughan, R. and Roberts, J.: Environmental segmentation alternatives: A look at green consumer behavior in the new millennium, *Journal of Consumer Marketing*, Vol.16, No.6, pp.558–575 (1999).

[8] 洪　懐馨：ライフスタイルと購買行動の関連性―シャンプー購買についての台湾と日本の比較，創価大学大学院紀要，Vol.28, pp.55–77 (2006).

[9] Boer, M., McCathy, M., Cowan, C. and Ryan, I.: The influence of lifestyle characteristics and beliefs about convenience food on the demand for convenience foods in the Irish market, *Food Quality and Preference*, Vol.5, pp.155–165 (2004).

[10] Buckley, M., Cowan, C. and McCarthy, M.: The convenience food market in Great Britain: Convenience food lifestyle (CFL) segments, *Appetite*, Vol.49, pp.600–617 (2007).

[11] 小柴　等，石垣　司，竹中　毅，本村陽一：行動履歴データとライフスタイル調査に基づく顧客モデル構築技術，電気学会論文誌 C，Vol.133, No.9, pp.1785–1795 (2013).

[12] 一般財団法人流通システム開発センター，入手先 ⟨http://www.dsri.jp/jan/about_jan.htm⟩（参照 2016-05-14）．

[13] 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，第 6 回データ工学と情報マネジメントに関するフォーラム（DEIM2014），F3-5 (2015).

[14] 土井千章，片桐雅二，小西哲平，中川智尋，稲村　浩，太田賢：複数店舗における購入商品レベルでの購買行動に着目したライフスタイルの推定，マルチメディア・分散・協調とモバイルシンポジウム，pp.672–683 (2015).

[15] 大沼　進：ライフスタイルから見る環境配慮行動，廃棄物資源循環学会論文誌，Vol.22, No.2, pp.101–113 (2011).

[16] 石垣　司，竹中　毅，本村陽一：日常購買行動に関する大規模データ融合による顧客行動予測システム：実サービス支援のためのカテゴリマイニング技術，人工知能学会誌，Vol.26, No.6, pp.670–681 (2011).

[17] Ishigaki, T., Takenaka, T. and Motomura, Y.: Customer-item category based knowledge discovery support system and its application to department store service, *Asia-Pacific Services Computing Conference*, pp.371–377 (2010).

[18] 石垣　司，竹中　毅，本村陽一：潜在クラスモデルによる流通量販店舗の来店人数予測の精度改善の評価，人工知能学会全国大会論文集，Vol.25, pp.1–4 (2011).

[19] Riedmiller, M.: Advanced supervised learning Multi-layer perceptrons from back propagation to adaptive learning algorithm, *Computer Standards and Interfaces Special Issue on Neural Networks*, Vol.16, No.3, pp.265–275 (1994).

[20] Breiman, L.: *Machine Learning*, Vol.45, pp.5–32, Kluwer Academic Publishers (2001).

付　　録

A.1　因子分析の結果

ライフスタイルの特徴軸に関する因子分析の結果を**表 A·1** に示す．

表 A·1　ライフスタイルの特徴軸に関する因子分析の結果
Table A·1　Result of factor analysis.

質問項目	ライフスタイルの特徴軸				
	第一因子	第二因子	第三因子	第四因子	第五因子
カロリーに気を使った食生活を心掛けている		0.12	0.54	0.22	
お弁当の調理を行う	-0.29		0.15		
多少高くても有機・無農薬野菜を買う	0.89			0.15	
料理を作るのが楽しい，好きだ	0.76				
ガヤガヤと騒がしいところにいるのは嫌い		0.14			-0.24
お金が手元にあればあるだけ使ってしまう方だ		-0.51	-0.11	0.15	0.18
産地直送の食べ物に弱い			0.22	0.66	
食品の新商品は率先して試す			0.12	0.51	0.25
今を楽しく過ごすことを大切にしている					0.63
栄養バランスに気を使った食事をしている			0.94	0.14	
新しいことにどんどんチャレンジしていきたい				0.16	0.46
無駄の少ない合理的な生活をしたい	0.8				
特売していた商品が献立を決めるきっかけになる	0.33				
お金を使うことは慎重なほうだ		0.81			
休日は予定を詰め込んで活動的に過ごす方だ				0.11	0.58
モノを買う時は，じっくり比較検討してから決める		0.52			
同じものを買うなら安いところを探して買う		0.29			

土井 千章　（正会員）

平成 21 年慶應義塾大学大学院理工学研究科博士前期課程修了．同年（株）NTT ドコモ入社．現在，株式会社NTT ドコモ先進技術研究所勤務．モバイルコンピューティング，Android アプリケーションのセキュリティ，顧客理解を目的とした行動モデリングや大規模データを用いた行動予測技術の研究に従事．

片桐 雅二　（正会員）

昭和 61 年早稲田大学大学院理工学研究科博士前期課程修了．日本電信電話，UC Berkeley 訪問研究員，NTT ドコモマルチメディア研究所，ドコモUSA 研究所等を経て，現在，NTT ドコモ先進技術研究所主幹研究員．図形処理，マルチメディアアプリケーション，データマイニングの研究開発に従事．大阪大学博士（情報科学）．電子情報通信学会，IEEE 各会員．本会シニア会員．

太田 賢　（正会員）

平成 10 年静岡大学大学院博士課程修了．博士（工学）．平成 11 年 NTT 移動通信網（株）入社．現在，NTT ドコモ先進技術研究所勤務．モバイルコンピューティング，端末セキュリティ，分散システムに関する研究に従事．共著『モバイルネットワーク』，訳書『コンピュータネットワーク第 5 版』等．電子情報通信学会会員．本会シニア会員．

重野 寛　（正会員）

1990 年慶應義塾大学理工学部計測工学科卒業．1997 年同大学大学院理工学研究科博士課程修了．現在，同大学理工学部教授．博士（工学）．情報処理学会論文誌編集委員，電子情報通信学会英文論文誌 B 編集委員等を歴任．現在，情報処理学会マルチメディア通信と分散処理研究会主査，Secretary of IEEE ComSoc APB．ネットワーク・プロトコル，ITS 等の研究に従事．著書『ユビキタスコンピューティング』（オーム社），『情報学基礎第 2 版』（共立出版）等．電子情報通信学会，IEEE，ACM 各会員．