

日常購買行動に関する大規模データの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

Customer Behavior Prediction System by Large Scale Data Fusion in a Retail Service

石垣 司
Ishigaki Tsukasa

東北大学大学院経済学研究科
Tohoku University,
isgk@econ.tohoku.ac.jp, <http://www.econ.tohoku.ac.jp/~isgk/>

竹中 毅
Takeshi Takenaka

産業技術総合研究所サービス工学研究センター
National Institute of Advanced Industrial Science and Technology
takenaka-t@aist.go.jp,

本村 陽一
Yoichi Motomura

(同 上)
y.motomura@aist.go.jp, <http://staff.aist.go.jp/y.motomura/>

keywords: service engineering, large scale data modeling, latent class model, ID-POS data, Bayesian network

Summary

This paper describes a computational customer behavior modeling by Bayesian network with an appropriate category. Categories are generated by a heterogeneous data fusion using an ID-POS data and customer's questionnaire responses with respect to their lifestyle. We propose a latent class model that is an extension of PLSI model. In the proposed model, customers and items are classified probabilistically into some latent lifestyle categories and latent item category. We show that the performance of the proposed model is superior to that of the k -means and PLSI in terms of category mining. We produce a Bayesian network model including the customer and item categories, situations and conditions of purchases. Based on that network structure, we can systematically identify useful knowledge for use in sustainable services. In the retail service, knowledge management with point of sales data mining is integral to maintaining and improving productivity. This method provides useful knowledge based on the ID-POS data for efficient customer relationship management and can be applicable for other service industries. This method is applicable for marketing support, service modeling, and decision making in various business fields, including retail services.

1. は じ め に

現状のサービス産業の品質は熟練したサービス提供者の経験と勘への依存度が大きく、その生産性の低さが問題となっている [経済産業省 07, 内藤 09, 本村 09]。その中でも小売業に注目すると、オーバーストア問題によって生じる、クーポン戦略や低価格化戦略などの過当競争による利益率の低下が問題視されている。この状態を継続することで小売業者の提供可能なサービスレベルが低下し、長期的には一般の生活者が享受できる価値も低下する。そのため、低価格化戦略に頼らないサービス品質の向上が課題となっている。商品の価格を下げずに来店率や顧客満足度を向上させることができる顧客との関係構築の方法論と実店舗支援システムが必要とされている。この数十年の間、消費者の価値観やライフスタイルの

多様化が著しい [Ueda 09, Takenaka 10]。そのような社会では、大量生産の方法論が有効ではない商品やサービスも増えている。そのため、個人の特性に合わせたパーソナライゼーションや、あるセグメントに注目するマイクロマーケティングなどの観点からのサービス提供が重要となってきた。その方法論では、顧客を何らかのカテゴリにセグメント化し [中村 08]、各顧客カテゴリに適したサービス提供を実施することで、顧客満足度の向上と持続可能な需要創造を志向する。しかしながら、価値観・ライフスタイルの多様化により、年齢、性別、居住地域などのデモグラフィック属性や個別売上高のみに基づいた従来のカテゴリでは、施策の発案や効果に限界がある。従来の変数に加え、顧客のライフスタイルや価値観に合わせたカテゴリを生成し、そのカテゴリに適したサービス提供を実施できれば、顧客満足度の向上と効

率的な資源の集中が可能となる [中村 09] .

一方、顧客のみではなく、商品に関するカテゴリ化も重要な課題である。大規模な小売業では取り扱う商品の数が数千から数万商品に及び、かつ、その種類も多様である。そのため、適切な商品管理による適正在庫の維持は、環境負荷対策や利益率の低下防止のための重要な業務となる。しかしながら、現状では各商品に対して商品属性のみに基づいた大分類、中分類、小分類などの階層的なラベルを付与し、商品管理に利用している業者が多い。また、それらの商品分類の多くは流通業者の都合で設定されており、必ずしも顧客にとって意味のある分類とはなっていない。カテゴリマネジメントの方法論は顧客のニーズに基づいた商品カテゴリ毎の商品管理の重要性が説いており、かつ米国での成功事例も多数報告されている [Nielsen 05, 麻田 04] . しかし、その成功の鍵となる商品のカテゴリ化は質的調査と属人的な経験に頼っているところが多い。顧客の購買パターンやライフスタイルに基づいた商品カテゴリの生成は、適切な商品管理の支援を可能とすると考えられる。

顧客と商品の双方について、有効なプロモーションが可能なカテゴリを、非熟練者でも過去の日常購買行動に関する大規模データからシステムティックに生成できるようなシステム・方法論を提供することは、業務の高効率化とサービス品質の高付加価値化の観点から有益である。また、そのカテゴリに基づいた、購買状況や購買パターンの関係性をモデル化し、顧客行動の理解を促進することは、提供可能なサービス価値の増大につながると期待できる。そのためには、大規模データから現実的な時間内にカテゴリの自動生成と顧客行動に関する計算モデル構築が可能なシステムを開発し、顧客行動の理解・予測による実問題の解決を図っていく必要がある。

そこで本論では、顧客ライフスタイルに注目した日常購買行動に関する大規模データを融合的に利用した顧客と商品の同時カテゴリ生成と、そのカテゴリをベースとした顧客行動の計算モデル化による顧客行動予測システムについて述べる。図 1 に提案手法の概念図を示す。本研究では、流通量販店の顧客 ID 付きの購買履歴データである 1 年間の ID-POS データと、その会員への 4000 人規模のアンケートデータを用いる。

以下の流れで顧客と商品カテゴリの同時生成を行う。(i) 潜在的な顧客カテゴリと商品カテゴリを仮定し、顧客、商品、潜在顧客カテゴリ、潜在商品カテゴリの関係を潜在クラスモデルとして表現する。(ii) 顧客のライフスタイル特性をアンケートデータから因子分析を用いて抽出する。(iii) 抽出したライフスタイル特性を潜在クラスモデルの制約条件として導入し、顧客ライフスタイル属性と潜在商品カテゴリの関係を潜在商品カテゴリの生成を同時に実行する。その後、生成されたカテゴリ、顧客ライフスタイル特性、商品特徴、購買状況間の関係をベイジアンネットワーク [本村 06, Pearl 09] を用いて計算モ

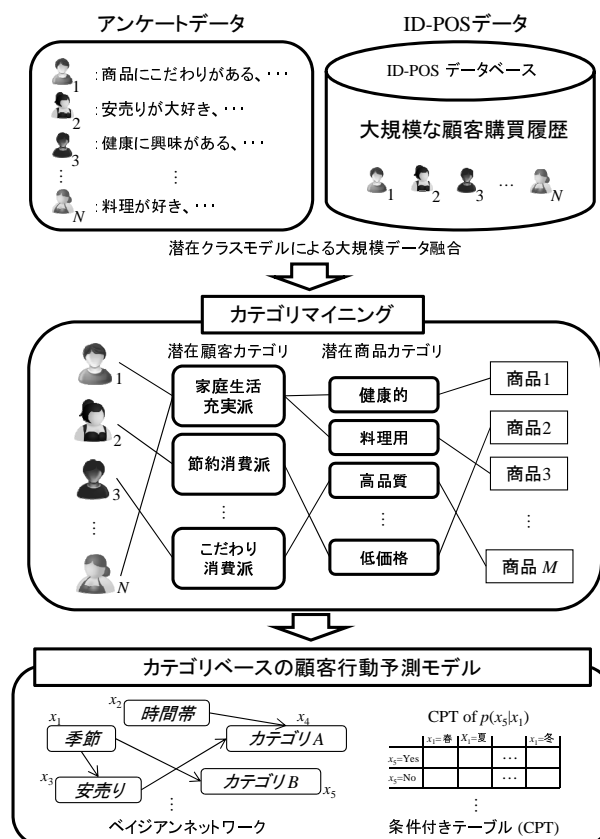


図 1 提案システムの概念図

デル化する。そのモデルから定量的な知識の抽出を行う。それにより、膨大な組み合わせ数が存在する変数間の関係・交互作用を全探索しなくとも、モデル内のリンク構造をたどることで、情報量規準の基で意味のある知見を自動的に抽出可能とする。

従来、顧客ライフスタイルに基づくカテゴリ生成は行われているが、特定の同種商品群の取り扱いを主としている、自由記述による顧客アンケートの結果を属人的経験に基づき解析している、顧客と商品のカテゴリ生成や解析は別々に行われているなど、本論の内容との差異は多い [中村 09] . 提案する潜在クラスモデルでは、多種類の商品群と顧客ライフスタイルを大規模データに基づき客観的に同時に解析することが可能である。本論では、この一連のカテゴリ生成と顧客行動モデリングをカテゴリマイニングと呼ぶこととする [Ishigaki 10, 石垣 10b, 石垣 10c, 石垣 10d] . 大規模データを利用したカテゴリマイニングシステムを開発することで、小売業やその他のサービス業における実問題の解決に貢献可能と考えられる。

2. ID-POS 活用に関する現状と関連研究

現在では ICT の発達により日常の購買行動の現場で観測されるデータ量は爆発的に増大している。小売業においても、1990 年代よりポイントカードや会員カードを利用した ID-POS データと呼ばれる顧客 ID 付きの購買履歴

データも大量に蓄積されている。ID-POS データには「いつ、どこで、誰が、何を、いくつ」購買したのかという履歴が数百万から数億件規模で蓄積されているため、顧客と商品の関係を知るための豊かな情報源となることが期待される。しかしながら小売業の現場の多くでは、そのデータの大規模性からデシル分析や RFM 分析などの基本特徴量の算出のみが主な解析となっており、その解析結果が十分に活用されていない現状がある。

小売業においてサービス品質の向上と適切な商品管理のためには、どのような顧客がどのような商品を求めているのかを定量的に把握することが望ましい。従来の研究では、顧客アンケートデータ、購買調査データ、スキャンパネルデータ [岡太 01] などを用いて階層的クラスタリング [Saunders 08, Xu 09]、潜在クラス分析 [Goodman 78, 佐藤 03]、ジョイントセグメンテーション [Ramaswamy 96]、自己組織化マップ [Weng 03, 徳高 02] などの手法により顧客や商品のセグメント化が行われている。これらの手法は顧客側、もしくは商品側に注目した単方向のセグメンテーションである。また、コレスポネンス分析 [Greenacre 06, 君山 05] などの顧客と商品の関係を線形的に低次元空間に次元圧縮し、両者の関係性を把握する手法もある。しかしながら、データ収集の方法や適用手法の限界から数百～数万サンプルのデータに対する研究がほとんどである。加えて、ID-POS データの解析として、例えば顧客来店行動の解析 [佐藤 08]、消費者の異質性を考慮した消費者モデル [照井 09]、新製品の動向解析 [Yada 07]、時系列解析手法の応用 [阿部 05]、潜在クラスモデルを利用した顧客行動の動的変化の解析 [Iwata 09] などの研究成果があるが、いずれも本論が目的としている顧客と商品のカテゴリをベースとした顧客行動予測システムの構築とは目的が異なる研究である。

自然言語処理などの分野で使用されている確率的潜在意味解析 (PLSI) [Hofmann 99, Hofmann 01] により顧客と商品を同時にクラスタリングする研究もある。[石垣 11a] では百貨店の ID-POS データに対して PLSI により作成されたカテゴリに対する顧客行動のモデル化を行っている。また、[石垣 10a] では流通量販店の ID-POS データに対して PLSI を利用した同時分類が行われているが、このモデルでは顧客カテゴリと商品カテゴリを分離せず購買行動データのみに基づき同一の潜在カテゴリへ分類している。PLSI では分類された結果から各カテゴリの意味内容を把握する必要がある。ID-POS データを分類の対象とした場合、商品の分類結果から各カテゴリの意味内容を把握するためには大量の商品の特徴を把握している必要があり、非熟練作業員やデータ分析者にその条件を求めることは現実的ではない。提案モデルでは、カテゴリ生成には顧客ライフスタイルアンケートの分析結果を反映させるため、顧客ライフスタイルを軸として潜在商品カテゴリの意味内容が直感的に理解しやすいという利点もある。

3. 大規模データ融合によるカテゴリ自動生成

3.1 潜在顧客・潜在商品カテゴリのモデリング

本章では、ID-POS データと顧客アンケートデータを用いたカテゴリ生成について述べる。ここでは、2 重の潜在変数を仮定した顧客、商品、潜在顧客カテゴリ、潜在商品カテゴリの関係をモデル化する。そのグラフィカルモデルを図 2 に示す。このモデルは

- (1) 顧客はいくつかの顧客カテゴリに分類できる
- (2) 商品は顧客の購買傾向から、いくつかのカテゴリに分類できる
- (3) 顧客カテゴリは特定の商品カテゴリを購買する傾向にある

という 3 つの期待をモデル化したものである。

ここでは X 人の顧客と Y 個の商品を対象とし、顧客 i と商品 j を表す変数をそれぞれ $x_i (i = 1, \dots, X)$ と $y_j (j = 1, \dots, Y)$ とする。また、潜在顧客カテゴリ数を U 、潜在商品カテゴリ数を V とし潜在顧客カテゴリ k と潜在商品カテゴリ l を表す変数をそれぞれ $u_k (k = 1, \dots, U)$ と $v_l (l = 1, \dots, V)$ とする。 x, y, u, v は上記のそれぞれの変数を含むベクトルとする。また、確率分布 p は多項分布と仮定する。ここでは顧客、商品、潜在カテゴリ間の関係を

$$p(x_i, y_j, u_k, v_l) = p(u_k)p(x_i|u_k)p(v_l|u_k)p(y_j|v_l) \quad (1)$$

としてモデル化する。このモデルは顧客 i の商品 j の購買数を N_{ij} とすると、その対数尤度は

$$\begin{aligned} l &= \sum_i^X \sum_j^Y N_{ij} \log p(x_i, y_j) \\ &= \sum_i^X \sum_j^Y N_{ij} \log \left\{ \sum_k^U \sum_l^V p(u_k)p(x_i|u_k)p(v_l|u_k)p(y_j|v_l) \right\} \end{aligned} \quad (2)$$

となる。

この潜在クラスモデルの対数尤度は EM アルゴリズムにより最大化できる。ここでは推定すべき条件付き確率をパラメータと呼ぶ。このモデルにおいて推定すべきパラメータ数は $X \times U$ 個の $p(x|u)$ 、 $U \times V$ 個の $p(v|u)$ 、 $Y \times V$ 個の $p(y|v)$ 、 U 個の $p(u)$ である。各パラメータに対して初期値を乱数で与えると、式 (2) の変形から潜在変数の条件付き確率は以下のように計算できる (E ステップ)

$$\begin{aligned} p(u_k, v_l | x_i, y_j) &= \frac{p(x_i, y_j, u_k, v_l)}{p(x_i, y_j)} \\ &= \frac{p(u_k)p(x_i|u_k)p(v_l|u_k)p(y_j|v_l)}{\sum_k^U \sum_l^V p(u_k)p(x_i|u_k)p(v_l|u_k)p(y_j|v_l)} \end{aligned} \quad (3)$$

また、ラグランジュの未定乗数法から各反復計算ステップの式 (3) の条件付き確率を最大化するパラメータは以

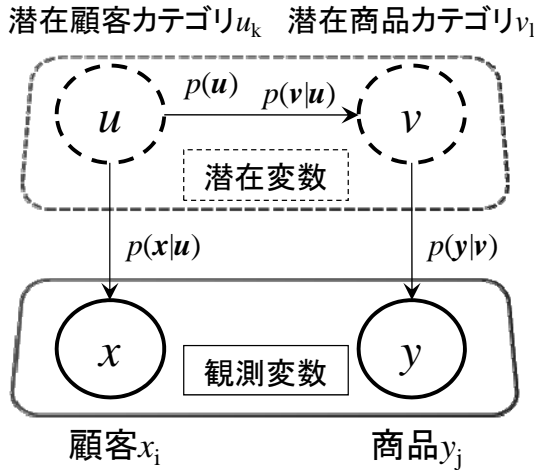


図2 提案モデル

下のように求めることができる (M ステップ)

$$p(x_i|u_k) = \frac{\sum_j^Y \sum_l^V N_{ij} p(u_k, v_l | x_i, y_j)}{\sum_i^X \sum_j^Y \sum_l^V N_{ij} p(u_k, v_l | x_i, y_j)} \quad (4)$$

$$p(v_l|u_k) = \frac{\sum_i^X \sum_j^Y N_{ij} p(u_k, v_l | x_i, y_j)}{\sum_i^X \sum_j^Y \sum_l^V N_{ij} p(u_k, v_l | x_i, y_j)} \quad (5)$$

$$p(y_j|v_l) = \frac{\sum_i^X \sum_k^U N_{ij} p(u_k, v_l | x_i, y_j)}{\sum_i^X \sum_j^Y \sum_k^U N_{ij} p(u_k, v_l | x_i, y_j)} \quad (6)$$

$$p(u_k) = \frac{\sum_i^X \sum_j^Y \sum_l^V N_{ij} p(u_k, v_l | x_i, y_j)}{\sum_i^X \sum_j^Y \sum_k^U \sum_l^V N_{ij} p(u_k, v_l | x_i, y_j)} \quad (7)$$

この反復を尤度が収束するまで実行することで各パラメータを推定することができる。

ここでは、顧客はアンケート回答者を対象とする。また、商品は1年間の売上個数の上位1,000商品を対象とする。つまり、 $X = 3965$, $Y = 1000$ である。また潜在クラス数は赤池情報量規準 (AIC) 等の情報量規準 [小西 04] により決定することができる。本モデルの自由パラメータ数を $|Para| = U(X-1) + V(U-1) + V(Y-1) + (U-1)$ とするとその AIC の値は

$$AIC = -2l + 2|Para| \quad (8)$$

と計算できる。

3.2 日常購買行動に関するデータ

§1 大規模 ID-POS データとデータ抽出

本節では、カテゴリマイニングの実現のために使用する顧客の日常購買行動に関する大規模データについて説明する。

兵庫県を事業エリアに約150店舗を展開する流通量販店で記録された2008年10月1日から2009年9月30日の期間におけるID-POSデータを利用する。この期間の全データのトランザクション数は約6.7億件 (669,511,467件) である。このID-POSデータから前節の N_{ij} が観測

可能となる。

本ID-POSデータは、ポイントカードの提示によりポイントカードのIDと購買記録が関連付けて記録されている。また、各トランザクションには、ポイントカードID、店舗、利用日時、各商品に対して1対1対応の商品コード、購買数、購買価格の情報が含まれている。商品コードと関連付けることができる流通量販店独自の商品大分類・中分類・小分類・商品名のマスタが利用可能である。ここではMySQLを用いてID-POS用のデータベースを作成し、データ抽出環境を構築した。

§2 顧客アンケートデータとその解析

顧客の購買行動データのみでは、提案モデルに対して顧客ライフスタイルを反映させることができない。そのため、ここでは顧客のライフスタイルに関するアンケートデータを用いる。

同流通量販店において、顧客のライフスタイルやパーソナリティを把握するためのアンケート調査を実施した。2009年12月に同流通量販店の会員約17,000人に対しダイレクトメールによりアンケートを送付し、その内3,965名から回答を得た。

質問項目は消費者心理学、社会心理学、マーケティング、個人差研究における消費者のライフスタイルや価値観に対する知見に着目し、様々な先行研究を元に設計した [杉本 97, 竹村 00, 山本 01, 吉田 01, 上里 01]。デモグラフィック特性 (年齢, 性別, 家族構成, 家族人数, 職業), 日常行動・生活時間 (家での朝食夕食の頻度, 家事時間, 交通手段), 健康状態 (健康に関する心配, ダイエット志向), 食に対する意識 (料理, 国産品への意識), 消費傾向 (堅実志向, 冒険志向, 論理的, 感覚的), パーソナリティ・社会参加 (外向性, 情緒安定性, 社会性) 等の質問項目も設定した。また、ビッグ5法 [Goldberg 90] を用いたパーソナリティを把握するための質問項目も設定した。デモグラフィック特性と日常行動・生活時間に関する質問を除いた20項目の回答は質問項目への当てはまりの強度順に“良く当てはまる”から“全く当てはまらない”までの4カテゴリの選択式である。4つの選択肢に対し強度順に4, 3, 2, 1の得点を間隔尺度として与えている (各質問項目については付録Aを参照のこと)。アンケートの項目数は全35問である。

§3 ライフスタイルカテゴリの抽出

因子分析によりアンケートデータから顧客の消費に関するライフスタイル因子の抽出を行った。ここではKaiserの正規化を伴うバリマックス法を用いた。また、抽出する因子の数は固有値法により決定した。その結果、付録Bに示す特徴的な6つの軸が抽出された。それぞれの因子は表1に示すように6つのライフスタイルカテゴリとして特徴づけることができる。

各顧客に対して、6つのライフスタイルカテゴリのスコアを付与することができる。ここでは、顧客 i の質問項目 q に対する回答の点数を $Q_q^{(i)} \in \{1, 2, 3, 4\}$ とす

表 1 アンケートから抽出されたライフスタイルカテゴリ

カテゴリ名	特徴	所属人数
こだわり消費派	高くても健康に良いものを選ぶ．産地への関心，こだわりのブランドがある．	1670
家庭生活充実派	料理が好きで食事生活も充実している．気分も安定している．	1385
アクティブ消費派	外向的で，新商品や話題の商品は試しに買ってみる．ただ無駄遣いは多い．	384
節約消費派	チラシを見てお得な商品を買う．安ければ少々遠い店にも行く．高い商品は買わない．	707
堅実生活派	几帳面で家計簿をつけ，無駄遣いはしない．毎日の献立は店に行く前に決める．	364
パパッと消費派	スーパーでの買い物はできるだけ早くすませたい．お弁当を作ることがある．	869

る．そこで，ライフスタイルカテゴリ k に属する質問項目の集合 $\mathcal{L}_k^{(i)}$ を $\mathcal{L}_1^{(i)} = \{Q_5^{(i)}, Q_{13}^{(i)}, Q_{14}^{(i)}, Q_{15}^{(i)}\}$, $\mathcal{L}_2^{(i)} = \{Q_1^{(i)}, Q_2^{(i)}, Q_3^{(i)}, Q_{18}^{(i)}\}$, $\mathcal{L}_3^{(i)} = \{Q_{12}^{(i)}, Q_{16}^{(i)}, Q_{19}^{(i)}, Q_{20}^{(i)}\}$, $\mathcal{L}_4^{(i)} = \{Q_7^{(i)}, Q_8^{(i)}\}$, $\mathcal{L}_5^{(i)} = \{Q_6^{(i)}, Q_{10}^{(i)}, Q_{11}^{(i)}, Q_{17}^{(i)}\}$, $\mathcal{L}_6^{(i)} = \{Q_4^{(i)}, Q_9^{(i)}\}$ とする． Q_{10} , Q_{11} , Q_{18} に関しては因子の値が負であるため，得点を逆点させている．ここで， $|\mathcal{L}_k|$ をライフスタイルカテゴリ k の要素数とし，各顧客 i のライフスタイルカテゴリ k のスコアを以下のように定義する．

$$L_k^{(i)} = \frac{\sum_{Q_q^{(i)} \in \mathcal{L}_k^{(i)}} Q_q^{(i)}}{|\mathcal{L}_k|}. \quad (9)$$

表 1 の所属人数は各アンケート回答顧客に対して $L_k^{(i)}$ を計算し，最も高い得点をもつライフスタイルカテゴリへの所属を割り振った数値である．また各 k に対して最大同得点の場合は（例えば，こだわり消費派 4 点かつ堅実生活派も 4 点などの場合），2 重に数値をカウントしている．

本論ではアンケート回答者 3,965 人の購買記録を対象とし議論を進める．その ID-POS データのトランザクション数は 4,175,441 件である．

3.3 ライフスタイルスコアに基づく制約条件の導入

提案する潜在クラスモデルを購買履歴データのみに基づいて尤度最大化を行うと，顧客ライフスタイルを反映できない．そのため，前節で述べたライフスタイルスコアを提案モデルのパラメータ推定における制約条件として用いることで，顧客ライフスタイルを反映した商品カテゴリの生成を行う．

各ライフスタイルスコアを制約条件として以下のように利用する．

$$p(u_k | x_i) = \frac{L_k^{(i)}}{\sum_k L_k^{(i)}}, \quad (10)$$

$$p(u_k) = \frac{\sum_i L_k^{(i)}}{\sum_i \sum_k L_k^{(i)}}, \quad (11)$$

$$p(x_i) = \frac{\sum_j N_{ij}}{\sum_i \sum_j N_{ij}}. \quad (12)$$

そのモデルの概念図を図 3 に示す．パラメータ $p(u_k | x_i)$ は各顧客 i のライフスタイルスコアの正規化された値を

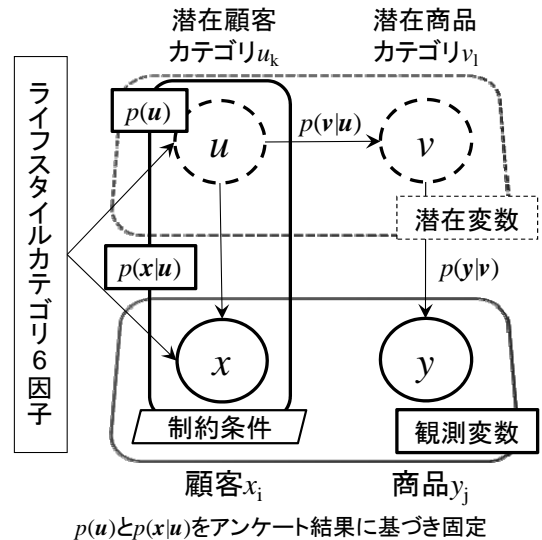


図 3 制約条件を用いた提案モデル

代入することを式 (10) は表現している．式 (11) はパラメータ $p(u_k)$ は各カテゴリの得点を全顧客に関して正規化した値である．式 (12) は $p(x_i)$ は各顧客に関して正規化された商品購買数である．また，ベイズの定理からパラメータ $p(x_i | u_k)$ が計算できるため，この制約条件を式 (4) に代入することができる．

この制約条件によりパラメータ推定に顧客のライフスタイルが反映される．PLSI 法においては行動履歴のみに基づき顧客と商品を同時分類する．また，各カテゴリに分類された商品の傾向から，その分類結果の意味を考える．小売という業態上，数千から数万以上の顧客や商品の種類を扱う必要があり，分類結果の可視化や現場への知識の反映に大きな困難が生じる．また提案モデルでも同様に自由度が高く，さらに潜在クラスが多層となっているため，ライフスタイルのような軸を考慮せず独立に最尤推定を行うと，より分類結果の意味の把握が困難になる．そのため，本節のような制約条件を用いる．また，推定される潜在商品カテゴリに対しても流通量販店のサービス従事者が理解しやすいカテゴリが生成されることが期待できる．

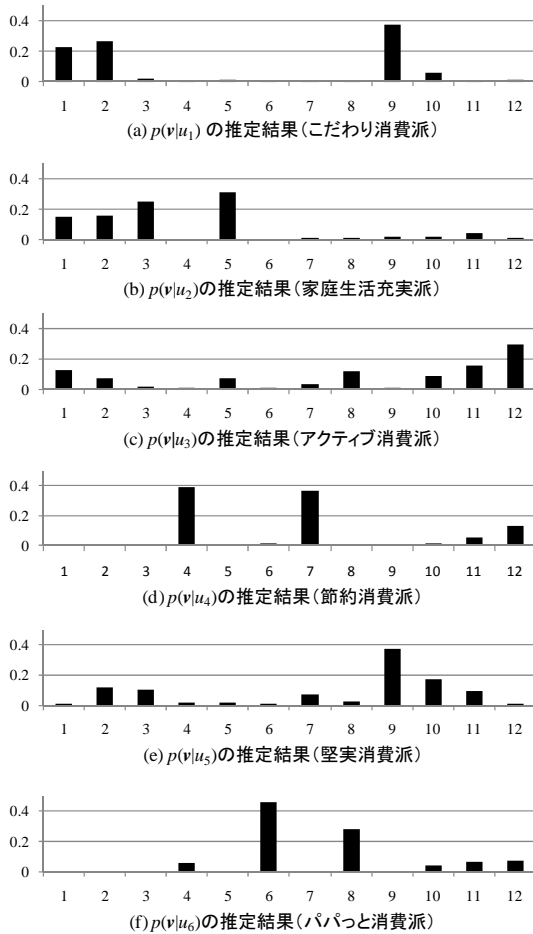


図4 $p(v|u_k)$ の条件付き分布

3.4 制約条件の導入による計算量の削減

式 (3) 内の $p(u_k)p(x_i|u_k)$ は制約条件を導入することで定数として扱うことができる．ここで式 (10) ~ (12) とベイズの定理を用いて事前に計算した制約条件を定数として $p(u_k)p(x_i|u_k) = \gamma_{ik}$ とおくと，制約条件を導入したときの式 (3) は以下のように変形できる．

$$p(u_k, v_l | x_i, y_j) = \frac{\gamma_{ik} p(v_l | u_k) p(y_j | v_l)}{\sum_k^U \sum_l^V \gamma_{ik} p(v_l | u_k) p(y_j | v_l)} \quad (13)$$

これにより，EM アルゴリズムの M ステップにおいて式 (5), (6) を計算するのみで，反復に必要な $p(u_k, v_l | x_i, y_j)$ の計算が可能となり，計算量が削減できる．また，上式を用いた場合でも，ラグランジュの未定乗数法により M ステップにおける更新式は式 (5), (6) と同様となる．

制約条件の導入により式 (13) には x_i に関する変数が含まれておらず， u_k と y_j を観測変数， v_l を潜在変数， $p(u_k)$ を所与とした潜在クラスモデルに変形することも可能である．本論で各顧客 x_i を明示的にモデル化している理由は，個人に訴求可能なサービスを記述するためである．潜在顧客カテゴリはアンケート回答顧客に関しては事前知識を観測可能な制約条件として導入できるが，非アンケート回答顧客に関しては本質的に潜在変数である．しかし，推定したパラメータを用いることで非アン

ケート回答顧客の購買履歴からライフスタイルの傾向を推定することが可能となる．[石垣 11b]，[本村 11] では，推定したパラメータを用いて推定された顧客ライフスタイルカテゴリごとに各日の来店人数の予測を行うことで予測精度が向上する事例を示しており，パラメータ推定の結果と非アンケート回答顧客に対する潜在顧客カテゴリ導入の実応用上の有効性を示している．また，[村上 09] では，レコメンデーションにおいて，ユーザの習慣から外れており，かつユーザの嗜好に適合したコンテンツの推薦が顧客満足度向上のために有効であることが示されている．ここでは，推定したパラメータを用いて $p(v_l | x_i)$ を計算することができる．この値の高い潜在商品カテゴリは，顧客 i の嗜好性が高い潜在商品カテゴリとみなすことができる．その商品カテゴリの中から普段は顧客 i が購買していない商品を推薦することで顧客の満足度を向上できる可能性がある．

3.5 カテゴリ生成実験

カテゴリ生成実験は Mac OS X，プロセッサ $2 \times 2.93\text{GHz}$ Quad-Core Intel Xeon，メモリ 32GB の PC により動作させた．また， $U = 6$ とし $V = [2, 5, 10, 15, 20, 30, 50]$ について各 3 回異なる初期値で AIC の値を計算した．その結果として平均的に $V = 10$ が最適な値であると決定された．そこで $V = [6, 7, 8, 9, 10, 11, 12, 13, 14]$ について各 5 回異なる初期値で AIC の値を計算した結果， $V = 12$ が平均的に最適であると判断されたため，ここではその値を用いる．また，以下で用いるカテゴリは $U = 6, V = 12$ に対して異なる初期値で最適化計算を 10 回実行し，最も尤度が高かった結果を採用している． $U = 6, V = 12$ とした時の 1 回の試行にかかる計算時間は平均 53 分であった．

3.6 カテゴリ生成の結果

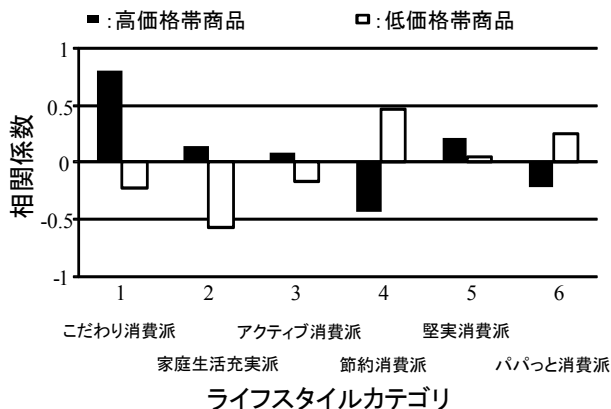
§1 パラメータの推定結果

図 4 に推定された $p(v|u_k)$ の値を示す．図中の横軸は各カテゴリ番号，縦軸は条件付き確率を示す． u_1, u_2, u_4, u_6 は数個の大きな条件付き確率をとるカテゴリが存在する． u_3, u_5 に関しては比較的複数のカテゴリで小さな条件付き確率値をとっていることがわかる．また以下では，各顧客 i と商品 j はそれぞれ $p(u_k | x_i)$ と $p(v_l | y_j)$ の値が最も高い潜在顧客カテゴリ k と潜在商品カテゴリ l に分類する．表 2 に各潜在カテゴリへ分類された商品の傾向を示す．表中の PB とはプライベートブランドを意味する．また，表中の商品傾向は，筆者らと流通量販店の担当者が，各潜在カテゴリ内の商品群を定性的に評価した内容を記している．

商品の中に 2 種類の見切り品が存在し，この 2 商品は節約消費派が最も高い条件付き確率を示している潜在商品カテゴリ 4 に属している．また，全 19 商品が存在する 10 個詰めたまごの内，平均単価が高い 5 商品がこだわり消費派が高い条件付き確率を示している潜在商品カ

表 2 潜在商品カテゴリへ分類された商品の傾向

カテゴリ番号	特徴的な商品
1	高品質 PB 商品
2	生野菜・生鮮品
3	日配品
4	低価格帯商品
5	鮮魚・肉類
6	肉類・パン・飲料
7	セールへの頻出商品
8	多種混合
9	高価格帯野菜
10	小サイズの野菜・日配品
11	飲料
12	惣菜・飲料

図 5 $p(v|u)$ と高・低価格帯商品の所属確率の相関係数

テゴリ 1, 2, 9 に属している。また、平均単価が一番安い商品は節約消費派が高い条件付き確率を示している潜在商品カテゴリ 7 に属している。料理をすると回答している家庭生活充実派が高い条件付き確率を示している潜在商品カテゴリ 1, 2, 3, 5 には調理済みの惣菜品がほとんど分類されていない。以上のように、商品の分類結果から、カテゴリ生成のある程度の妥当性を見ることができる。

一方、 $U = 6, V = 12$ の同条件で顧客ライフスタイルに関する制約条件を加えないパラメータ推定も行った。その結果、低価格帯商品カテゴリ・高価格帯商品カテゴリが生成されるなど一部の妥当性は見られたが、その他の潜在商品カテゴリには多様な商品が所属してしまい、特徴的な商品例の発見が困難であった。各潜在顧客カテゴリの分類基準がライフスタイルやデモグラフィック属性などの人間にとって理解しやすい軸を反映していないため、各潜在商品カテゴリの意味の把握・追求が困難となったことが理由の一つに挙げられる。

§2 価格帯による分類傾向

また、高価格帯、低価格帯商品の潜在商品カテゴリへの出現確率と $p(v|u_k)$ との相関をみる。ここでは、たま

表 3 クロスバリデーションの平均絶対誤差

ランダム	k -means	PLSI	提案法	
			制約無し	制約あり
8.63	2.15	1.05	1.03	0.89

ご、牛乳、見切り品、高価格帯ブランドの全 71 商品（高価格帯 50 商品，低価格帯 21 商品）を対象とした。たまごと牛乳を採用した理由は、その価格が生鮮品などと比べ季節的に変動しにくいためである。たまごと牛乳に関しては同量のたまご商品と同量の牛乳商品を平均価格順に 3 等分し、その上位と下位をそれぞれ高価格帯、低価格帯商品と定義した。ここでは潜在商品カテゴリ u_k に対する高価格帯商品の出現確率を $p(high) = \{p(high_1), \dots, p(high_l), \dots, p(high_V)\}$ ，低価格帯商品の出現確率を $p(low) = \{p(low_1), \dots, p(low_l), \dots, p(low_V)\}$ とし、各 k に対し $p(v|u_k)$ と $p(high)$ ， $p(v|u_k)$ と $p(low)$ の相関係数を計算した。その結果を図 5 に示す。その結果、

- こだわり消費派は高価格帯商品を買う傾向がある
- 家庭生活充実派は低価格帯商品を買わない傾向がある
- 節約消費派は高価格帯商品を買わず、低価格帯商品を買う傾向がある

などの傾向が読み取ることができ、潜在商品カテゴリの生成結果は価格帯の購買行動に関して、ある程度の妥当性を持っていることが分かる。

3.7 カテゴリ生成の結果検証

ここではカテゴリ生成結果の検証を行う。通常、PLSI 等を用いた文書クラスタリングでは Perplexity や F-measure 等の指標がよく用いられている [Hofmann 01]。しかしながら、Perplexity は単語列の予測精度の指標であり本問題には適さない。また、F-measure を計算するためにはテストデータに正解ラベルが必要となるが、本問題では顧客と商品の両カテゴリ分類について正解が存在しない。そのため、提案法と既存のカテゴリ生成法の未知データに対する説明力をクロスバリデーションにより検証する。

2008 年 10 月から 2009 年 9 月までの全 ID-POS データを 1 カ月毎の 12 データセットに分類し、11 カ月分のデータを訓練データ、残り 1 カ月分のデータをテストデータとし、12-fold クロスバリデーションによりその未知データに対する説明力を検証する。

潜在顧客カテゴリ k に所属する顧客の潜在商品カテゴリ l に属する商品の購買数を \tilde{N}_{kl} とする。ここで、訓練データで構成された \tilde{N}_{kl} を \tilde{N}_{kl}^{Train} ，テストデータで構成された \tilde{N}_{kl} を \tilde{N}_{kl}^{Test} とする。ここでは、潜在顧客カテゴリは 3.2.3 節での顧客アンケートの解析結果に基づき決定する。また、テストデータにおける各商品の潜在商品カテゴリへの所属は、訓練データにより作成された \tilde{N}_{kl}^{Train} の所属と同様とする。

ここでの目的は、 \tilde{N}_{kl}^{Train} と \tilde{N}_{kl}^{Test} の各潜在顧客カテゴリ毎の各潜在商品カテゴリに所属する商品の購買割合

がどの程度類似しているのかを示すこと、すなわち顧客ライフスタイルカテゴリの訓練データの購買傾向からどの程度 \tilde{N}_{kl}^{Test} を説明可能であるかを検証することである。

ここでは、 \tilde{N}_{kl}^{Train} には 11 カ月分の購買が、 \tilde{N}_{kl}^{Test} には 1 カ月分の購買のみが反映されているため、その直接の残差を比較することはできない。そのため、正規化した \tilde{N}_{kl}^{Train} と \tilde{N}_{kl}^{Test} の誤差 E_{kl} により分類の良さを評価する。本論では、 E_{kl} は以下のように正規化絶対誤差として定義する。

$$E_{kl} = \left| \frac{\tilde{N}_{kl}^{Train}}{\sum_k \sum_l \tilde{N}_{kl}^{Train}} - \frac{\tilde{N}_{kl}^{Test}}{\sum_k \sum_l \tilde{N}_{kl}^{Test}} \right| \quad (14)$$

ここでは、ランダムな分類、代表的なクラスタリング手法である k -means 法 [?], PLSI, 制約条件無しの提案法, 制約条件ありの提案法により作成されたカテゴリに対し \tilde{N}_{kl} をそれぞれ作成し、12-fold クロスバリデーションによる評価を行った。ランダムな分類とは、全ての商品がランダムに分類された場合についての評価である。また、本指標ではカテゴリ数が同一のときのみ、そのカテゴリ間の一致度を比較することができる。そのため、ここでは全てのカテゴリ生成法において $U = 6, V = 12$ と固定した。また、 k -means 法は提案モデルや PLSI 法で実現される共クラスタリングの方法ではないため、ここでは商品の被購買個数に関するクラスタリングを行い、その後、3.2 節の顧客アンケート結果に基づき顧客の分類を行った。[石垣 11a] では通常の PLSI 法を用いているため、本検証は先行論文との比較となっている。

表 3 に 3 つの分類における正規化絶対誤差の全潜在顧客カテゴリと全潜在商品カテゴリに対する合計値の平均 ($\frac{1}{UV} \sum_k \sum_l E_{kl}$) をそれぞれ示す。また、表中に示す値は 12-fold クロスバリデーションの平均値である。その結果、提案法による分類の誤差が最も少なくなっている。そのため提案法を用いたカテゴリ化は、顧客ライフスタイルカテゴリ毎の購買傾向の説明力の意味で良い性能を示すことが示された。

4. カテゴリベース顧客行動モデリング

4.1 ベイジアンネットワーク

本章では、前章で生成したカテゴリを単位とした顧客行動予測のための計算モデル化について述べる。ここでは、ベイジアンネットワークにより、商品の特徴や購買された状況とカテゴリの関係をネットワーク構造で図示することにより、情報量規準の基で意味のある変数間関係を自動抽出することができる。また、その変数間の確率構造から、定量的な関係理解を実行できる。

ベイジアンネットワークは対象とする確率変数のノードと変数同士の依存関係を確率的なネットワークとしてモデル化したものである。その確率ネットワークはグラフ構造として表現することが可能で、視覚的に表現・理解し易く、グラフィカルモデルによる確率推論の手法を

直接応用することができる。また、そのグラフ構造は情報量規準などによりデータから自動的に探索・構築することも可能であるし、設計者の経験や物理的・社会的な法則をモデル内に取り込み柔軟に決定することも可能である。グラフ構造が決定すると、そのモデルの同時分布はリンクが張られている変数間の条件付き確率の積として表現することが可能である。その条件付き確率は条件付き確率表としてデータから学習することができる。確率伝搬法、LoopyBP などの確率推論のアルゴリズムを用いることで、ある変数にエビデンスを与えたときの事後確率の計算や感度分析などが実行可能となる。このように柔軟な確率モデルでは、顧客の購買状況を計算論的に変数として取り扱うことが可能である。そのため、ベイジアンネットワークを用いた生活者行動理解の研究は広く行われるようになってきている [本村 06]。

4.2 確率構造の学習法

ここでは AIC の基で最適なモデル構造の探索を行う。AIC を用いることでデータに対しての過学習を避ける目的がある。説明変数の次元を N_v 、 z を各変数を要素としてもつベクトル ($z = \{z_1, z_2, \dots, z_{N_v}\}$)、 $P(z)$ を確率モデルの同時分布とする。また、変数 z_i の親ノードの集合を $pa(z_i)$ 、ベイジアンネットワークの確率構造を S とすると、確率構造 S が与えられた時のベイジアンネットワークの同時分布は

$$P(z|S) = \prod_{i=1}^{N_v} p(z_i | pa(z_i), S) \quad (15)$$

として表現できる。さらに、変数 z_i の親ノードの集合 $pa(z_i)$ がとるパターン数を $N_{pa(i)}$ 、親ノードの集合 $pa(z_i)$ がとるパターンが決まった時の z_i がとりうる状態の数を $N_{K(i)}$ とし、 $\Theta_S = \{\theta_{ijk}\}$ を構造 S が与えられた時の親ノードの集合 $pa(z_i)$ が、 j 番目のパターンをとった時に $z_i = k$ となる条件付き確率に対応したパラメータ集合とする。また、データセット D において、 $pa(z_i)$ が j 番目のパターンをとり変数 z_i に対して k 番目の値をとったデータの数 N_{ijk} と表記する。このとき、データセット D が与えられた時の対数尤度 $l_{BN}(\Theta_S|D)$ は

$$l_{BN}(\Theta_S|D) \propto \sum_{i=1}^{N_v} \sum_{j=1}^{N_{pa(i)}} \sum_{k=1}^{N_{K(i)}} (N_{ijk} \log \theta_{ijk}) \quad (16)$$

と近似できることを利用すると、構造とデータセットが与えられた時のベイジアンネットワークの AIC は以下のように表現できる。

$$AIC_{BN} = -2l(\Theta_S|D) + 2 \sum_{i=1}^{N_v} N_{pa(z_i)} \quad (17)$$

この値を最適にするようにモデル構造の学習を行う。

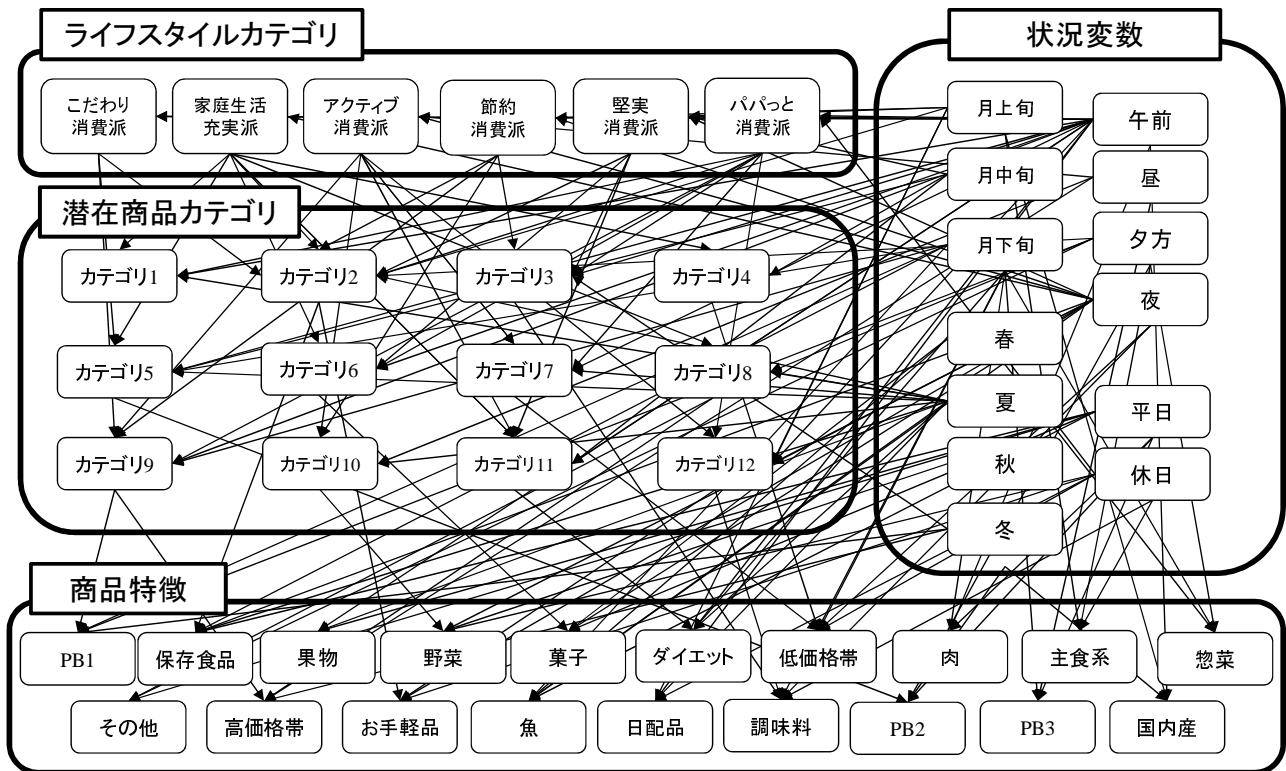


図 6 構築したベイジアンネットワークモデル

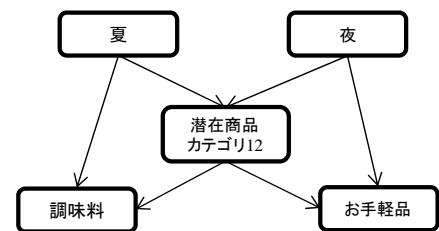
4.3 確率構造モデルの構築

ここでは対象となっている約 420 万件のトランザクションデータを用いてベイジアンネットワークモデルを構築した。各商品に対して購入した顧客に対し、合計購入数、合計金額、購入平均単価、特定保健用食品（トクホ）購入回数、プライベートブランド購入回数、国産品購入回数、健康食品購入回数、お手軽品購入回数、高級品購入回数、ダイエットの商品購入回数、お買い得商品購入回数に関しては対象店舗の顧客に対して ABC 分析を行い、その結果をラベルとして付与した。また、6 種類のライフスタイルカテゴリ属性と 12 種類の潜在商品カテゴリ属性のラベル、状況変数（購買の時間帯、平日か休日、季節、月旬）を付与した。そのデータに対してベイジアンネットワーク構築ソフトウェアである Bayonet[本村 03]を用いて、Greedy search により AIC の意味で最適になるような確率構造の探索を行った。その結果、図 6 に示すようなベイジアンネットワークモデルが構築された。カテゴリ生成と同じ計算機環境で、モデル構築の計算時間は 45 時間 58 分であった。

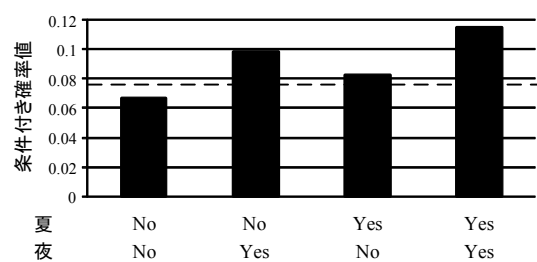
4.4 カテゴリに対する知識抽出

§1 潜在商品カテゴリに対する顧客行動理解の一例

ここでは、潜在商品カテゴリ 12 に着目する。図 7 に構築したベイジアンネットワークの部分グラフとカテゴリ 12 に関する購買の条件付き確率を示す。カテゴリ 12 はお手軽品と調味料、夏と夜からリンクが張られている。また、図中の破線はカテゴリ 12 の 1 年を通した被購買



(a) 構築されたベイジアンネットワークの部分グラフ



(b) 潜在商品カテゴリ12に関する確率値

図 7 潜在商品カテゴリに対する条件付き確率の一例

確率を示している。図よりカテゴリ 12 の商品は夜に多く購買され、特に夏の夜には平均被購買確率よりも約 3.5 ポイントも高い被購買確率を示している。このように確率構造モデルを利用することで、生成された潜在商品カテゴリに対して、付加的な情報を抽出することが可能となり、状況依存的な商品被購買の予測が可能となる。

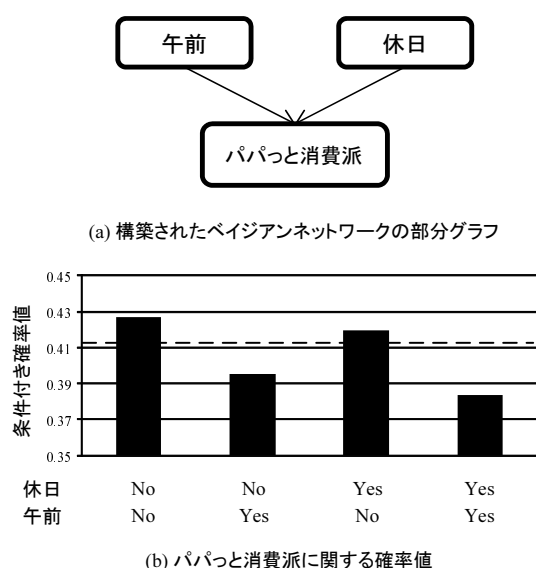


図8 顧客ライフスタイルカテゴリに対する条件付き確率の一例

§2 顧客ライフスタイルカテゴリに対する顧客行動理解の一例

顧客のライフスタイルカテゴリに対しても同様の解釈を行うことができる。ここでは、パパッと消費派に着目し、図8に構築したベイジアンネットワークの部分グラフとパパッと消費派に関する購買の条件付き確率を示す。図より、パパッと消費派は休日、午前からリンクが張られている。また、図中の条件付き確率より、パパッと消費派の顧客は午前中の購買確率が他の時間帯よりも総じて低いことが読み取れる。また、休日の午前中の購買確率はさらに低くなっていることが読み取れる。このように確率構造モデルを利用することで、顧客ライフスタイルカテゴリに対しても付加的な情報を抽出することが可能となる。また、そのカテゴリに属する顧客の状況依存的な購買行動の予測が可能となる。

ライフスタイル属性に対する購買行動の特徴抽出に関する研究は多様に行われているが、それらは定性的なモデルの構築や小サンプル・線形モデルでの解析事例がほとんどである[井関91, Michman 91]。構築した確率構造モデルを用いることで、対象としているライフスタイル属性に対して大規模実データに基づいた特徴を探索的に発見することも本手法の有用性の一つである。

5. 考 察

多くの小売業では、ID-POS データに代表される大規模データの有効活用が望まれている。しかしながら、その利用が進まない背景には、日々の業務に労働時間の多くを費やしていることや、大規模データを取り扱うことができる専門家の人材不足などがある。そのため、本手法で扱うような大規模データを恒常的に実際の現場で有効活用するためには、現場の従業員が容易に利用可能な

GUI やインタフェースを備えたシステムとして提供する必要がある。本論で示した方法論は、そのような実際のサービス現場で使用できるシステムとして構築することが十分に可能であり、そのための作業は現在進行中である。

制約条件の導入は、各パラメータの推定結果に大きな影響を与える。本論で用いた制約条件は心理学的知見に基づいた質問項目を設計したアンケートの分析結果を用いているため、その結果には心理学的な意味が反映される。無計画に収集されたデータでは、例えば制約条件が購買行動と結びつかない、カテゴリ間の差異がデータに表れずカテゴリ化できない、などの現象が生じる可能性がある。そのため、制約条件の決定には、分析対象としての観点からの知見を十分に反映させることが重要である。

大規模データ解析の実務への応用のためには一般的な計算機パワーによる現実的な時間内での計算可能性が一つの課題となる。本論では、潜在クラスモデルを含んだベイジアンネットワークを計算するのではなく、提案モデルによる分類の結果をベイジアンネットワークの変数として取り込み確率モデルを構築している。それにより、提案モデルのパラメータ推定とベイジアンネットワーク構築の計算時間の線形和で目的とするモデルの構築が可能であり、実務への応用が可能な計算機パワーと計算時間内で本論の目的を達成することを可能としている。加えて、実務でベイジアンネットワークモデルを利用するためには、数か月や一年の間隔でモデル構築を実行することで、新しいデータがモデル構築に反映されと考えられる。なぜなら、ID-POS データの蓄積にはある程度の時間が必要なためである。また、各変数の条件付き確率の可視化や確率推論などのモデルの利用はリアルタイムで可能である。そのため、構造学習に必要な約46時間の計算時間は実利用における障害とはならない。

ベイジアンネットワークでは確率伝搬法などによる確率シミュレーションの実行も可能である。サービス提供者が知りたい状況を設定した顧客購買行動の確率シミュレーションを実行することで、その状況下での顧客行動の予測を行い、その行動に適合したサービス提供を実施できる可能性がある。そのような実店舗への介入実験は今後の課題とする。また、提案モデルによるパラメータの最尤推定アルゴリズムは並列実装することができ、PLSIと同様[Das 07]にスケーラブルな処理が可能である。

提案手法では顧客のライフスタイルを軸に潜在商品カテゴリとの関係性と潜在商品カテゴリの生成を行ったが、本モデルを利用することでライフスタイル以外の観点からの確率値の推定も可能である。例えば、行動履歴データと関連付けが可能な別種アンケートや仮説を制約条件として反映させることで、本システムの利用者が見たい観点からの潜在カテゴリ化が可能である。また、ID-POS データから店舗間での各商品カテゴリの売れ方を計量で

きる．また，その売れ方から逆算により店舗間の顧客ライフスタイル構成率の違いも検出することができる．そのため，店舗ごとの顧客構成率に合わせた商品発注量や在庫量の最適化に応用できる可能性がある．

6. む す び

本論では顧客の購買行動に関する大規模データを融合した顧客と商品のカテゴリの発見法とそのカテゴリに関する顧客行動予測システムの開発について述べた．顧客の価値観やライフスタイルの多様化に合わせ，個々の顧客の要求に過不足なく対応することがサービス産業における生産性向上の鍵である．本手法は小売業においてカテゴリを単位とした効率的な店舗業務の支援を可能とする．例えば，顧客に合わせたサービス提供による顧客満足度の向上や，需要予測による品揃え最適化による提供価値の増大など，顧客接点の最適化に利用できると期待できる．また，本システムは流通量販店のみならず，インターネットショップや通信販売などの業態においても適用可能である．ベイジアンネットワークの確率推論を行うことで，各変数に対して様々な条件下での予測が可能となる．その予測結果に基づくプロモーションや店舗支援などの介入実験は現在進行中の課題である．

謝 辞

本論で用いた ID-POS データと顧客アンケートデータは生活協同組合コープこうべから提供を受けた．ここに感謝の意を表する．また，本研究は経済産業省 IT とサービスの融合による新市場創出促進事業（サービス工学研究開発事業）の委託研究費を受けている．

◇ 参 考 文 献 ◇

- [阿部 05] 阿部，近藤，マーケティングの科学-POSデータの解析，朝倉書店，2005
- [麻田 04] 麻田孝治，戦略的カテゴリーマネジメント，日本経済新聞社，2004
- [Das 07] A. Das, M. Datar and A. Garg “Google news personalization: scalable online collaborative filtering”, *Proc. 16th international conference on World Wide Web*, pp.271-280, 2007
- [Goldberg 90] L.R. Goldberg, “An alternative description of personality: The Big-Five factor structure”, *Journal of Personality and Social Psychology*, Vol.59, No.6, pp.1216-1229, 1990.
- [Goodman 78] L. Goodman, *Analyzing Qualitative/Categorical Data: Log-Linear Models and Latent-Structure Analysis*, Abt Books, 1978
- [Greenacre 06] M. Greenacre and J. Blasius (ed.), *Multiple Correspondence Analysis and Related Methods*, Chapman & Hall, 2006
- [Hofmann 99] T. Hofmann and J. Puzicha, “Latent class models for collaborative filtering”, *Proc. 16th international joint conference on Artificial intelligence*, 1999
- [Hofmann 01] T. Hofmann, “Unsupervised Learning by Probabilistic Latent Semantic Analysis”, *Machine Learning*, Vol.42, No.1-2, pp.177-196, 2001
- [井関 91] 井関，室井，生活起点発想とマーケティング革新，国元書房，1991
- [石垣 10a] 石垣，竹中，本村，確率的潜在意味解析を用いた大規模 ID-POS と顧客アンケートの統合利用による顧客 - 商品の同時力
- テゴリ分類，信学技報，Vol.109, No.461, NC2009-160, pp.425-430, 2010
- [石垣 10b] 石垣，竹中，本村，日常購買行動に関する大規模データを融合したベイジアンネットワークモデル，人工知能学会全国大会論文集，3J1-NFC1a-2, 2010
- [石垣 10c] 石垣，竹中，本村，2重潜在クラスモデルとベイジアンネットワークを結合した小売サービスにおける顧客購買行動モデリング，信学技報，vol.110, No.76, IBISML2010-24, pp.167-17, 2010
- [石垣 10d] 石垣，竹中，本村，大規模 ID-POS データ活用のための顧客の計算モデル化～小売サービスにおける生産性向上の試み，第 80 回 SIG-FPAI, pp.15-18, 2010
- [Ishigaki 10] T. Ishigaki, T. Takenaka and Y. Motomura, “Category Mining by Heterogeneous Data Fusion Using PdLSI Model in a Retail Service”, *Proc. IEEE International Conference on Data Mining*, pp. 857-862, 2010
- [石垣 11a] 石垣，竹中，本村，百貨店 ID 付き POS データからのカテゴリ別状況依存の変数間関係の自動抽出法，オペレーションズ・リサーチ，Vol.56, No.2, pp.77-83, 2011
- [石垣 11b] 石垣，竹中，本村，“潜在クラスモデルによる流通量販店の来店人数予測の精度改善の評価”，2011 年度人工知能学会全国大会論文集，1B3-2, 2011
- [Iwata 09] T. Iwata, S. Watanabe, T. Yamada, and N. Ueda, “Topic tracking model for analyzing consumer purchase behavior”. *Proc. 21st International Joint Conference on Artificial Intelligence*, pp. 1427-1432, 2009.
- [経済産業省 07] 経済産業省編：サービス産業におけるイノベーションと生産性向上に向けて，経済産業調査会 (2007)
- [君山 05] 君山由良，コレスポネンズ分析の利用法，データ分析研究所，2005
- [小西 04] 小西，北川，情報量規準，朝倉書店，2004
- [Michman 91] R.D. Michman, *Lifestyle Market Segmentation*, Praeger Pub, 1991
- [本村 03] 本村陽一，ベイジアンネットワークソフトウェア BayoNet，計測と制御，Vol.42, No.8, pp.693-694, 2003
- [本村 06] 本村陽一，岩崎弘利：ベイジアンネットワーク技術，東京電機大学出版局 (2006)
- [本村 09] 本村，石垣，“サービス工学における計算論的モデル”，システム / 制御 / 情報，Vol.53, No.9, pp.374-379, 2009
- [本村 11] 本村，竹中，石垣，“条件付層別差分モデルによる需要予測の高精度化”，2011 年度人工知能学会全国大会論文集，1B3-1, 2011
- [村上 09] 村上，森，折原，“推薦の意外性向上のための手法とその評価”，人工知能学会論文誌，Vol.24, No.5, pp.428-436, 2009
- [内藤 09] 内藤耕 (編)，“サービス工学入門”，東京大学出版会，2009
- [中村 08] 中村博 (編)，“マーケット・セグメンテーション”，白桃書房，2008
- [中村 09] 中村，寺本，矢野：顧客視点の商品マスター (商品 DNA) の可能性，流通情報，No.477, pp.22-33
- [Nielsen 05] A.C. Nielsen and A. Heller, *Consumer-Centric Category Management: How to Increase Profits by Managing Categories Based on Consumer Needs*, Wiley, 2005.
- [岡太 01] 岡太，守口，木島 (編) “マーケティングの数理モデル”，朝倉書店，2001
- [Pearl 09] J. Pearl, *Causality: Models, Reasoning and Inference (2nd ed.)*, Cambridge University Press, 2009.
- [Ramaswamy 96] V. Ramaswamy, R. Chatterjee and S.H. Cohen, “Joint segmentation on distinct interdependent bases with categorical data,” *Journal of Marketing Research*, vol. 33, no. 3, pp. 337-350, 1996
- [佐藤 03] 佐藤，廣松，椿，“潜在クラスモデルを利用した取引データのセグメンテーション”，行動計量学，Vol. 30, No.1, pp.121-133, 2003
- [佐藤 08] 佐藤，樋口，“動的個人モデルによる消費者来店行動の解析”，日本統計学会誌，Vol.38, No.1, pp.1-19, 2008
- [Saunders 08] J.A. Saunders, “Cluster Analysis for Market Segmentation”, *European Journal of Marketing*, Vol.14, no.7, 422-435, 1980
- [杉本 97] 杉本徹雄 (編)，編集消費者理解のための心理学，福村出版，1997
- [竹村 00] 竹村和久 (編)，消費行動の社会心理学 消費する人間のこころと行動，北大路書房，2000

- [Takenaka 10] T. Takenaka, K. Fujita, N. Nishino, T. Ishigaki and Y. Motomura, "Transdisciplinary approach to service design based on consumer's value and decision making," *International Journal of Organizational and Collective Intelligence*, vol. 1, no. 1, pp. 58-75, 2010
- [Ueda 09] K. Ueda, T. Takenaka, J. Vancza and L. Monostori "Value creation and decision-making in sustainable society" *CIRP Annals-Manufacturing Technology*, vol. 58, no. 2, pp. 681-700, 2009
- [上里 01] 上里一郎 (監), 心理アセスメントハンドブック, 西村書店, 2001
- [照井 09] 照井伸彦, ウィラワン・ドニ・ダナハ, 伴正隆, マーケティングの統計分析, 朝倉書店, 2009
- [徳高 02] 徳高, 山川, 藤村, 自己組織化マップ応用事例集, 海文堂出版, 2002
- [Weng 03] S.S. Weng and M.J. Liu, "Feature-based recommendations for one-to-one marketing", *Expert Systems with Applications*, Vol.26, No.4, 2003
- [Xu 09] R. Xu and D. C. Wunsch, *Clustering*, Wiley, 2009
- [Yada 07] K. Yada, E. Ip, N. Katoh, "Is this brand ephemeral? A multivariate tree-based decision analysis of new product sustainability", *Decision Support Systems*, Vol.44, pp.223-234, 2007
- [山本 01] 山本真理子 (編) 心理測定尺度集 1 人間の内面を探る "自己・個人内過程", サイエンス社, 2001
- [吉田 01] 吉田富二雄 (編), 心理測定尺度集 2 人間と社会のつながりをとらえる "対人関係・価値観", サイエンス社, 2001

〔担当委員：山川 宏〕

2011 年 6 月 13 日 受理

◇ 付 録 ◇

A. アンケート項目

本論で使用したアンケートの質問項目を以下に記す。

- Q1 バランスの良い食事がとれていると思う
- Q2 毎日の生活が充実している
- Q3 調理をするのが好きだ
- Q4 お弁当を作ることがある
- Q5 低カロリーの食品を選んでいる
- Q6 家計簿をつけている
- Q7 商品が安いお店があれば, 少々遠くても行く
- Q8 チラシなどを見て, お得な商品を買う
- Q9 スーパーでの買い物はできるだけ早くすませたい
- Q10 毎日の献立はお店で買い物をしながら決める
- Q11 無駄遣いが多い方だと思う
- Q12 新しい商品や話題の商品があると, 試しに買ってみる
- Q13 本店でしか買わない商品がある
- Q14 少々高くても健康に良いものを買うようにしている
- Q15 食料品の産地や, それを使ったレシピなどの情報に関心がある
- Q16 にぎやかなところが好きだ
- Q17 きちょうめんな方だと思う
- Q18 明るくなったり暗くなったり気分が変わりやすい
- Q19 新しいことを体験することが好きだ
- Q20 家族や友人と一緒に買い物に行くことがある

表 B.1 アンケートの因子行列

	F1	F2	F3	F4	F5	F6
Q1	0.23	0.55	0.41	0.01	0.16	0.06
Q2	0.09	0.60	0.17	-0.04	0.03	0.09
Q3	0.20	0.50	0.21	0.08	0.07	-0.03
Q4	-0.03	0.12	0.13	0.11	-0.04	0.14
Q5	0.31	0.20	0.04	0.01	0.07	0.06
Q6	0.10	-0.02	0.04	0.11	0.38	0.07
Q7	0.01	-0.03	0.13	0.59	0.06	0.09
Q8	0.11	0.00	0.05	0.71	0.08	0.05
Q9	0.06	-0.03	-0.01	0.06	0.02	0.49
Q10	0.06	-0.03	0.11	0.04	-0.21	0.13
Q11	0.08	-0.21	0.27	-0.01	-0.42	0.05
Q12	0.24	-0.06	0.43	0.09	-0.18	0.05
Q13	0.38	-0.01	0.05	0.07	0.04	0.09
Q14	0.71	0.10	0.15	-0.13	0.01	0.00
Q15	0.54	0.15	0.23	0.07	0.05	-0.06
Q16	0.06	0.00	0.50	0.07	-0.02	-0.02
Q17	0.20	0.10	0.10	0.05	0.36	0.05
Q18	0.04	-0.28	0.08	0.04	-0.01	0.09
Q19	0.12	0.12	0.56	-0.02	0.03	0.07
Q20	0.07	0.11	0.34	0.08	0.02	0.01

B. アンケートの因子行列

アンケートの因子分析により抽出された因子行列を表 B.1 に示す。表中の F1 から F6 はそれぞれの因子軸である。

著 者 紹 介

石垣 司(正会員)



2007 年総合研究大学院大学複合科学研究科統計科学専攻博士課程修了。博士(学術)。同年科学技術振興機構 CREST 研究員, 2008 年産業技術総合研究所特別研究員を経て, 2011 年より東北大学大学院経済学研究科講師。日本統計学会, 電子情報通信学会, 計測自動制御学会, 日本マーケティング・サイエンス学会, IEEE の会員。

竹中 毅(正会員)



2002 年神戸大学大学院文化学研究科社会文化専攻修了。2002 年より東京大学人工物工学研究センターにおいて PD 研究員, 助手を経て, 2007 年~2009 年まで特任准教授。2009 年 4 月より産業技術総合研究所サービス工学研究センター研究員。専門は認知心理学, 共創工学。工学と心理学の融合によるサービスの創発的理解を目指した研究に従事。博士(学術)。日本心理学会, 計測自動制御学会などの会員。

本村 陽一(正会員)



1993 年電気通信大学大学院博士前期課程修了。同年通産省工技院電子技術総合研究所, 2003 年~産業技術総合研究所デジタルヒューマン研究センター主任研究員, 2008 年~同研究所サービス工学研究センター大規模データモデリング研究チーム長。2011 年同研究所サービス工学研究センター副研究センター長。統計数理研究所客員教授, 東京工業大学連携准教授兼任。博士(工学)。人工知能学会研究奨励賞, 全国大会優秀賞, 人間工学会大島賞, ドコモモバイルサイエンス賞, インタラクション 2011 ベストペーパー賞など受賞。