

百貨店 ID 付き POS データからのカテゴリ別 状況依存的変数間関係の自動抽出法

石垣 司, 竹中 毅, 本村 陽一

1. はじめに

現在、百貨店業界は業績の不振が続き、さらなる業務の高効率化とサービス品質の高付加価値化が求められている。この現象は百貨店業界のみならず、日本のサービス産業全体に共通する課題である。その課題を解決するため、サービスについての科学的・工学的な研究が盛んになってきている[11]~[13]。現状の小売サービス業では、そのサービス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視されている。例えば、店舗ごとの来客や顧客ごとの商品購買数などの予測が最適ではないことから生じる欠品や過剰在庫などによるコストの増大や利益の減少が生産性を低下させていると考えられる。

百貨店や流通量販店などの大規模な小売業では、提供している商品数が多く、その内容も多種多様であることが特徴の一つとしてあげられる。そこでは一般的に、顧客、商品、購買状況に関する情報の組み合わせ数は膨大である。それにもかかわらず、熟練者は“どのような商品”が“いつ”、“どのくらい”売れるのかを経験的に知っている。また、どのような顧客カテゴリが存在し、どのような商品カテゴリを購買する傾向にあるのかも経験的に知っている場合もある。それらの知識を非熟練者でも過去のデータからシステムティックに抽出できるような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、価値観やライフスタイルの多様化により、年齢・性別・職業などの顧客のデモグラフィックスに基づいた従来のセグメンテーションでは、顧客の多様性

に対応することが困難になってきている[13][19]。このような現代において、その多様化に対応するためには、顧客の購買状況や購買パターンを考慮した生活者としての顧客理解が必要となってきている。そのためにも大規模な顧客行動履歴データから適切な顧客と商品のカテゴリを自動的に発見し、その状況や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、有効なマーケティング支援、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。

そこで本論では、顧客の購買パターンに基づいた自動的なカテゴリ生成と、そのカテゴリ別の状況依存的な顧客の購買行動に関する変数間関係の抽出方法について述べる。ここでは、潜在クラスモデルの一種である確率的潜在意味解析 (Probabilistic Latent Semantic Indexing: PLSI) [5]により ID 付き POS データから顧客と商品 (基本アイテム) を同時分類し、カテゴリ生成を行う。その後、そのカテゴリに加え、天気、購買時間帯、バーゲンなどの状況変数を ID 付き POS データに付与し、非線形・非ガウス性かつ変数間の相互作用を表現可能なベイジアンネットワーク[10][16]により、各変数間の関係を確率的構造モデルで表現する。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・相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状況依存性を伴う顧客の購買行動に関する知見を自動的に抽出可能となることを示す。また、本論で述べる方法論は、百貨店 ID 付き POS データのみならず、一般の ID 付き POS データに対して適用可能な枠組みを提供する。

2. 百貨店 ID 付き POS データと既存研究

2.1 百貨店 ID 付き POS データ

本論では、「経営科学系研究部会連合協議会主催、平成 21 年度データ解析コンペティション」で提供さ

いしがき つかさ, たけなか たけし, もとむら よういち
産業技術総合研究所 サービス工学研究センター
〒135-0064 江東区青梅 2-3-26
受付 10.7.22 採択 10.11.8

れた百貨店 ID 付き POS データを用いる。

本論で使用する ID 付き POS データは百貨店マルイの3店舗(有楽町店, 池袋店, 渋谷店)における2008年4月から2009年3月までの1年間の顧客ID付き購買履歴データである。本データ内には商品の被購買年月日, 時刻, 個数, バーゲン品であるか否か, 購買店舗, 購買した顧客のデモグラフィックス(年齢, 性別, 居住地)が同時に記録されている。また, 商品に関しては個々の商品名ではなく, 基本アイテムと呼ばれる小分類コードのみが記録されている。全シートデータ数は3,125,394件, 全基本アイテムは513種である。

2.2 既存研究

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などを用いて階層的クラスタリング, 自己組織化マップなどの手法により顧客ごと, 商品ごとのセグメント化が行われている[1][14]。また, コレスポネン分析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[7]。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。また, 潜在クラス分析による顧客購買行動に関する分析は数多くの報告例[3][4][15][17]があるが, その目的はアンケート質問項目などの低次元のカテゴリカル変数のセグメンテーションや数百人規模の顧客セグメンテーションであり, 本研究のような1年間のID付きPOSデータを用いた数万人規模の顧客セグメンテーションの例は報告されていない。加えて, 例えば顧客来店行動の解析[21], 消費者の異質性を考慮した消費者モデル[18], 新製品の動向解析[23], などのID付きPOSデータを利用した研究報告も多数あるが, いずれも本研究の目的とは異なるものである。

3. 顧客—基本アイテムカテゴリ生成

3.1 確率的潜在意味解析

本節では, ID-POS データから, どのような顧客がどのような基本アイテムを購買しているのかを把握するための顧客—基本アイテムの同時カテゴリ生成について述べる。

ここでは, 顧客が商品に対して求める意味カテゴリを潜在クラスとして仮定し, 顧客と商品を同時分類する。PLSIは自然言語処理の分野で文書と単語行列から文書の意味を推定するために提唱された潜在クラス

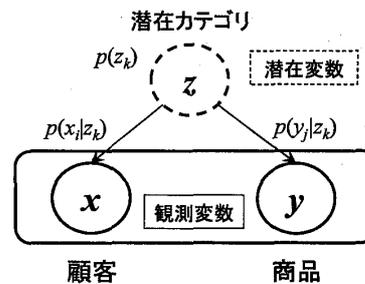


図1 PLSIのグラフィカルモデル

分析の一種である。そのモデルは, 文書がある潜在意味を持ち各文書内の単語がある潜在意味から確率的に発生するという確率モデルであり, Bayesの定理を介することで, マーケティングの分野で研究されている局所独立性を仮定した潜在クラス分析と等価となる。

ここでは N 人の顧客と M 個の基本アイテムを対象とし, 顧客 i と基本アイテム j を表す変数をそれぞれ $x_i (i=1, \dots, N)$ と $y_j (j=1, \dots, M)$ とする。また, 潜在クラスの数を K と仮定し, その変数を z_k と表す。ここでは顧客, 意味カテゴリ, 商品を多項分布として表現し, その関係を以下の同時確率としてモデル化する。

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

その概念図を図1に示す。このモデルは潜在クラス z_1 から z_K が与えられた下で, 各 x_i, y_j 間の関連が独立となることを表現している。また, 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

$$L = \sum_i \sum_j N_{ij} \log \sum_k p(z_k) p(x_i|z_k) p(y_j|z_k) \quad (2)$$

となる。

この潜在クラスモデルはEMアルゴリズムによる反復計算で対数尤度を最大化する条件付き確率値を推定することが可能となる。ここでは, 各々の条件付き確率値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は $K-1$ 個の $p(z_k)$, $(N-1) \times K$ 個の $p(x_i|z_k)$, $(M-1) \times K$ 個の $p(y_j|z_k)$ 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(1)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。

$$p(z_k|x_i, y_j) = \frac{p(z_k) p(x_i|z_k) p(y_j|z_k)}{\sum_k p(z_k) p(x_i|z_k) p(y_j|z_k)} \quad (3)$$

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各パラメータは以下のように求めることができる。

$$p(x_i|z_k) = \frac{\sum_j^M N_{ij} p(z_k|x_i, y_j)}{\sum_i^N \sum_j^M N_{ij} p(z_k|x_i, y_j)} \quad (4)$$

$$p(y_j|z_k) = \frac{\sum_i^N N_{ij} p(z_k|x_i, y_j)}{\sum_i^N \sum_j^M N_{ij} p(z_k|x_i, y_j)} \quad (5)$$

$$p(z_k) = \frac{\sum_i^N \sum_j^M N_{ij} p(z_k|x_i, y_j)}{\sum_i^N \sum_j^M \sum_k^K N_{ij} p(z_k|x_i, y_j)} \quad (6)$$

上記の反復を尤度が収束するまで実行することで尤度を最大化するパラメータを推定することができる。また、潜在クラス数は情報量規準により決定することが可能である [8]。

3.2 自動カテゴリ生成

ここでは、1年間に100件以上の被購買があった126の基本アイテム、1年間に23件以上の購買があった20,158人分の顧客IDを対象として自動カテゴリ分類を行う。利用する顧客IDと基本アイテム数を購買件数により選別する理由は、両者の共起頻度のスパー

ス性を抑制し、安定した推定結果を得るためである。上記の値は顧客IDと基本アイテムの購買を共起頻度行列 $\{N_{ij}\}_{N \times M}$ として考えたとき、その行列密度充填率 (N_{ij} がゼロではない要素の数) が10%となるように設定した。このとき利用したレシートデータ数は689,103件である。

加えて、自動分類のためには潜在クラス数 K を与える必要がある。ここでは、潜在クラス数 K を2から30まで変化させ、その赤池情報量規準 (AIC) [8] を計算した。また、EM アルゴリズムの推定結果は初期値により影響を受けるため、各 K の値に対し各10回の異なる初期値でAICの値を計算した。そのAICの平均値を表1に示す。その結果として平均的に $K=8$ がAICの意味で最適な潜在クラス数であると決定された。そのため、本論では潜在クラス数は8として取り扱う。また、本事例では $N=20,158$, $M=126$, $K=8$ であるため、推定すべきパラメータの数は162,263個である。

本論における以降の顧客-基本アイテム分類結果は $K=8$ において30回の分類実験を実行し、最も尤度が高かった推定結果を採用したものである。また、各顧客 i と基本アイテム j はそれぞれEMアルゴリズム収束後の条件付き確率 $P(z_k|x_i)$ と $P(z_k|y_j)$ の値が最も大きい値をもつ潜在カテゴリ k へ分類している。

各分類実験は Mac OS X, プロセッサ 2×2.93 GHz

表1 AICの値

潜在クラス数	2	3	4	5	6	7
AIC	9365073	9273126	9225698	9205238	9163411	9148998
潜在クラス数	8	9	10	15	20	30
AIC	9143042	9149057	9156032	9189711	9198151	9247460

表2 分類結果

カテゴリ名	分類された基本アイテム
婦人物カテゴリ (8334人)	婦人スーツ, ワンピース, 婦人ジャケット, 婦人スカート, 婦人パンツ, 婦人コート, 婦人ブルゾン, 婦人ブラウス, 婦人シャツ, 婦人ベスト, 婦人セーター, 婦人カーディガン, 婦人カットソー, 婦人トレーナー, 婦人レザー, 婦人ドレス, 婦人ベルト, 婦人マフラー, 婦人スカーフ, パンプス, 婦人サンダル, 婦人ブーツ, カジュアルヘアアクセサリ, 補修婦人, 婦人リフォーム
着物&アクセサリ カテゴリ (1256人)	婦人水着, アクセリ, 女浴衣, 女帯, 巾着, 女下駄, 着物雑貨, 福袋, 婦人シューズ, 婦人バッグ, Jリング, リング, ネックレス, ブレスレット, ピアスイヤリング,
紳士物カテゴリ (2859人)	紳士スーツ, 紳士ジャケット, 紳士ショートパンツ, 紳士スラックス, 紳士コート, 紳士ブルゾン, 紳士シャツ, 紳士ベスト, 紳士セーター, 紳士カーディガン, 紳士ポロシャツ, 紳士Tシャツ, 紳士トレーナー, 紳士レザー, 紳士水着, 紳士セットアップジャケット, 紳士セットアップスラックス, オフショーン, タイピン, 紳士アンダー, 紳士ソックス, 紳士ベルト, 紳士マフラー, ネクタイ, 紳士手袋, 紳士帽子, アウトドア用品, ビーチサンダル, 紳士ウインドブレーカー, 紳士ジャージ, 紳士グッズ, 紳士シューズ, 紳士バッグ, 紳士革小物, 喫煙具, 補修紳士, 紳士リフォーム,
化粧品カテゴリ (1883人)	化粧品, 化粧雑貨, タオル, スクール
ハンカチカテゴリ (1268人)	ハンカチ
婦人小物カテゴリ (1237人)	婦人ストッキング, 婦人ソックス, 婦人手袋, 婦人アパレル229, 婦人グッズ, Jピアス, 傘, ヘアウィッグ, レンタル
下着&その他 (796人)	ランジェリー, 婦人革小物, メガネフレーム, レンズ, ナイティ, 花
バラエティグッズ &お出かけ品カテ ゴリ (1280人)	婦人ポロシャツ, 婦人帽子, 婦人アパレル220, マリン用品, スポーツ用品, ゴルフ用品, スキー用品, レオタード, 婦人ウインドブレーカー, 婦人ジャージ, 紳士アパレル451, ケアグッズ, 婦人スリッポン, 婦人スニーカー, Jネックレス, Jアクセサリ類, 時計, サングラス, ヘアアクセサリ, 美容料金, リラクゼーション, 書籍, シーズングッズ, バラエティグッズ, ステーションナリー, 日用雑貨, インテリアグッズ, ホビー860, CD

Quad-Core Intel Xeon, メモリ 32 GB 1066 Hz DDR3 の PC 内で Python2.6.4 の 64 bit 版により動作させた。また, EM アルゴリズムの反復回数は, 予備実験から 100 回の反復で十分に尤度が収束することが確認できたため, すべての試行で 100 回に設定した。ここでの収束とは, 尤度の上昇率が尤度の絶対値に対して 0.01% 以下の値になることとした。また, $K=8$ のとき, 100 回の反復に要する時間は 10 試行の平均で 51.2 秒であった。

3.3 分類結果と検証

表 2 に PLSI により生成されたカテゴリと分類された基本アイテムを示す。各カテゴリ名は分類された基本アイテムの傾向から付与した。また, 各カテゴリに対して分類された顧客の数も表中に示す。婦人物と紳士物が異なるカテゴリに分類されている, 化粧品・ハンカチ・小物などの特徴のあるカテゴリが生成されているなど, 定性的にある程度の妥当性をみることができる。

PLSI による分類の妥当性を定量的に検証する。PLSI の枠組みでは一般的に F-measure, Perplexity 等の指標が分類精度評価に用いられるが, 本事例の場合はカテゴリ分類の正解が存在しないため, それらの指標は使用できない。そのため, クロスバリデーションによる未知データの予測力によって分類結果の妥当性の検証を行う。ここでは 1 カ月ごとにデータセットを分け, その中から 11 カ月分のデータを訓練データ, 残りの 1 カ月分をテストデータとして使用するアウトサンプルの 12-fold クロスバリデーションを実行し, その妥当性を検証する。

各顧客 i の各潜在カテゴリ k に属する基本アイテムの購買個数をそれぞれ訓練データとテストデータにおいて N_{ik}^{Train} , N_{ik}^{Test} とすると, その訓練・テストデータにおける正規化された購買個数の絶対誤差 (AE) の合計は

$$AE = \sum_i \sum_k \left| \frac{N_{ik}^{Train}}{\sum_i \sum_k N_{ik}^{Train}} - \frac{N_{ik}^{Test}}{\sum_i \sum_k N_{ik}^{Test}} \right|$$

により計算できる。ここでは, その絶対誤差の値が小さいほど, 未知データに対する予測力が高い分類であ

表 3 クロスバリデーションによる平均絶対誤差

	ランダム	k-means	PLSI
平均絶対誤差	2.04	1.54	1.37

るとみなすことができる。

表 3 に 12-fold クロスバリデーションの平均絶対誤差を示す。ここでは, 代表的なクラスタリングの手法である k -means 法[2]とその性能を比べた。 k -means 法では, 基本アイテムを説明変数として 8 個のクラスタに分類し, 上記の方法で平均絶対誤差を算出した。また, 表中のランダムとは, 仮に対象顧客全員が同じ割合ですべての基本アイテムを購入した場合の平均絶対誤差の値である。表 3 からは PLSI による分類の平均絶対誤差が最も小さくなっていることが分かる。よって, PLSI による顧客と基本アイテムのカテゴリ化は未知データに対する説明力の意味において妥当性を持っていることが示された。

4. 状況依存的確率構造モデリング

4.1 モデル化される変数

ここでは, 顧客変数, 状況変数, 潜在カテゴリ間の関係をベイジアンネットワークによりモデル化する。モデル化される変数を表 4 に示す。

顧客変数として顧客のデモグラフィクス (性別, 年齢, 住所), 購買価格・頻度・個数の ABC 分析の結

表 4 ベイジアンネットワークでモデル化する変数

顧客の属性	
年齢	10代, 20代, 30代, 40代, 50代, 60代以上の6分類
性別	男性, 女性の2分類
住所	各都道府県の47分類
購買個数	合計購買個数に対し上位20%=A, 上位20~50%=B, 下位50%=C
購買金額	合計購買金額に対し上位20%=A, 上位20~50%=B, 下位50%=C
来店頻度	来店頻度(日毎)に対し上位20%=A, 上位20~50%=B, 下位50%=C
購買状況と条件	
月	購買された月の1か月毎の12分類
曜日	購買された曜日(月曜日から日曜日)の7分類
時間帯	購買された時間帯(午前, 昼, 夕方, 夜)の4分類
平日休日	購買された日(平日, 休日)の2分類
降水量	無し・極少量・少量・雨・大雨の5分類
日照時間	無・小・中・多の4分類
店	購買された店舗(有楽町・池袋・渋谷)の3分類
バーゲン	バーゲン品か否かの2分類
顧客-基本アイテム潜在カテゴリ	
潜在カテゴリ	購買された商品が属するクラスタに対してダミー変数としてラベルを付与(2分類×8クラスタ)

果を付与した。状況変数として、商品購買に関する月、時間帯、店舗、曜日、平日休日、降水量、日照時間、バーゲンを取り入れた。また、潜在カテゴリは購買された基本アイテムが所属するカテゴリを1、それ以外を0とするダミー変数として付与した。

4.2 ベイジアンネットワーク

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・理解しやすく、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[2][10][16]。

変数間関係のモデル化に関して、例えばベイジアンネットワークの代わりに構造方程式モデリング[20]等の線形モデリングの手法も利用することが可能である。しかしながら、顧客の購買行動という本質的に非線形

性を含む事象の解析には変数間の非線形性を取り扱うことが可能なベイジアンネットワークの適用が適切であると考えられるため、本方法を採用した。

4.3 確率構造モデリング

ベイジアンネットワークモデル構築ソフトウェアである Bayonet[9]を用いて、確率的構造モデルを構築した(図2)。対象としたレシートデータはカテゴリ分類時に使用した689,103件のデータである。

ここでのモデル構築に関する条件は、以下の4つを指定した。

①潜在カテゴリ変数間にはリンクを張らない。②デモグラフィック属性に向けてリンクを張らない。③顧客購買履歴の変数間にはリンクを張らない。④条件付き確率表の頻度が100以下の事例がある場合はリンクを張らない。これらの①から③の条件は過度にリンク構造が複雑になるのを避けるため、実務上不必要もしくは現実に沿わないと判断した制約を導入した。④に関しては、少数事例を伴うリンク構造を排除するために導入した。

また、モデル選択基準はAICを採用し、Greedy Search[9]によるモデル構築を行った。その結果を図2に示す。PLSIによる分類と同じ計算機環境でモデル構築にかかった時間は4分32秒であった。

5. 状況依存的な変数間関係の抽出

本節では構築したベイジアンネットワークの構造からID付きPOSデータに含まれている状況依存的な変数間関係の抽出事例について述べる。それらは、構築されたベイジアンネットワークモデルのリンクが張られた変数間の関係を読み取ることで、仮説無しに抽出された事例である。

[例1]状況依存的な非線形効果の抽出

図3に顧客年齢別の購買時間帯ごとの商品販売個数を

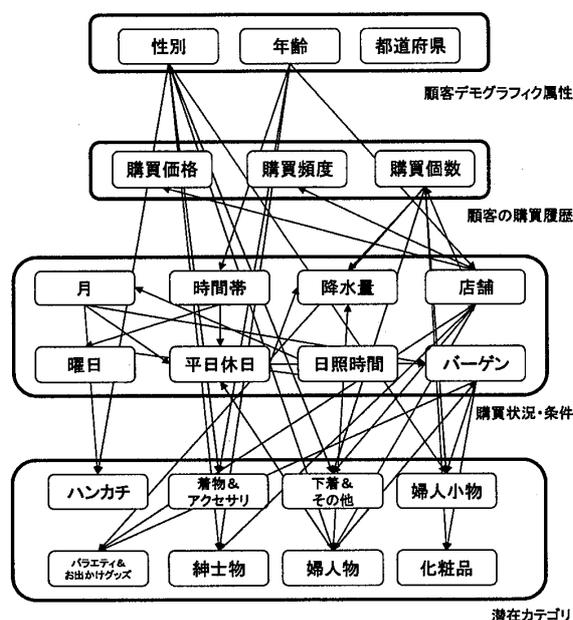


図2 構築されたベイジアンネットワークモデル

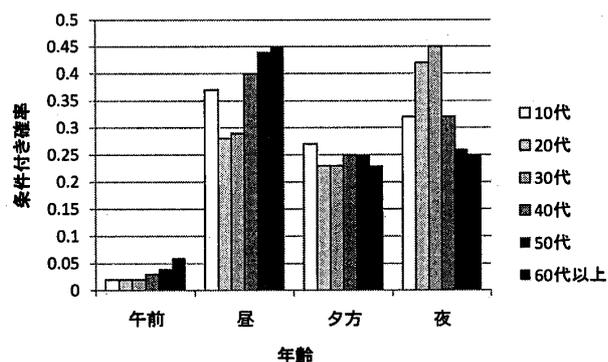


図3 年齢ごとの購買時間帯

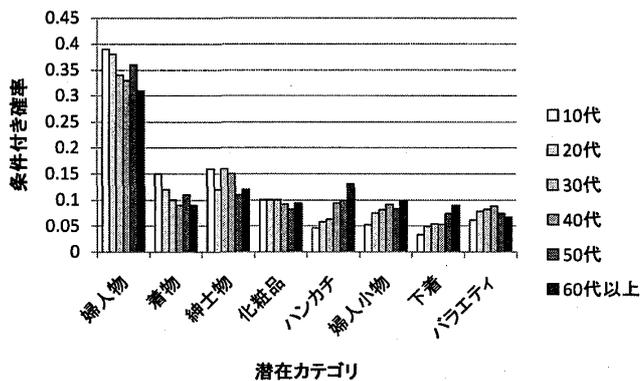


図4 カテゴリごとの購買者年齢

の条件付き確率を示す。本百貨店のメインターゲットの顧客は20代から30代の顧客であり、その年齢による商品購買割合は対象データに関しては87%であった。メインターゲットの年代は他の年代の顧客と比べ“夜”に商品を購入していることが読み取れる。逆に、10代と40代以上の顧客は“昼”に購買している割合が大きいことが分かる。このように、構築したベイジアンネットワークモデルは、非線形的な効果を含む変数間の状況依存関係を自動的に発見できている。

加えて、潜在カテゴリに含まれる基本アイテムの年齢別の購買割合を図4に示す。図4からはハンカチカテゴリ、婦人小物カテゴリ、下着&その他カテゴリに属する基本アイテムが10代から30代に比べて40代以上の顧客によく売れていることが読み取れる。そのため、これらのカテゴリに属する商品を昼の間は販促し、夜には20代30代向けの商品を販促する等の、時間帯ごとの棚割プロモーション戦略に応用できる可能性がある。

[例2] イベント駆動による状況変化の抽出

図2よりハンカチカテゴリには月と性別の変数からリンクが張られている。図5に各月ごとの男女別のハンカチカテゴリの購買条件付き確率を示す。男性のハンカチの購買確率は3ポイント付近で推移しているが、12月と3月のみ7ポイント以上に購買確率が倍増しているのが分かる。これは、12月の年末イベント、3月のホワイトデー等のイベントに駆動されてハンカチを購入する効果が表れていると推測できる。

このように、ベイジアンネットワークモデルに月や曜日などの時間状況を取り込むことで、イベントに駆動される売り上げの変化を自動的に発見し、かつその効果を定量的に把握することができている。

[例3] 常識的知識の定量的把握

図2の構造に着目することで、常識的な知識ではあ

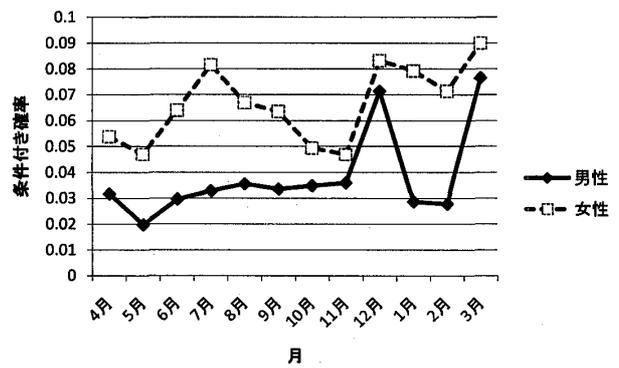


図5 男性女性別の月ごとのハンカチカテゴリの購買

るが、定量的に把握すべき状況依存性の効果や影響を条件付き確率として確認することができる。例えば、大雨の日には購買割合が少ない、紳士物カテゴリの約71%が男性IDにより購買されている、店舗間で年代、男女間に購買傾向の差がある、バーゲンを実施している期間の婦人物カテゴリの購買の70%以上はバーゲン品が占める等、常識的な知識に関しても定量的に把握することができる。

6. 考察

作成したベイジアンネットワークモデルを用いて、各変数について状況依存性を考慮した確率シミュレーションを実行することが可能である。シミュレーションにより顧客行動の予測を行うことで、カテゴリごと効果的な商品推薦のためのダイレクトメールの送付、カテゴリごとの最適在庫を考慮した商品発注など、その行動に適合したサービス提供や最適化を実施できる可能性がある。

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、PLSI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、PLSI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、PLSI処理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。

本論で述べた手法は一般のID付きPOSデータに対して適用できるため、流通量販店などへの直接的な応用が可能であると考えられる。本論で取り扱った基本アイテム数は126種類と比較的少数であったが、その数が10,000商品に増大しても、同条件での自動カ

カテゴリ生成は実務に対応可能な時間内に計算可能であることが確認されている。

7. おわりに

本論では、百貨店 ID 付き POS データからの自動カテゴリ生成とそのカテゴリ別の状況依存的変数間関係の抽出法について述べた。本論で述べた方法論を、ID 付き POS データを蓄積している実務家が利便性良く使用できるソフトウェアとして提供することが今後の課題である[6]。また、需要予測システムや自動発注システムなどに組み込むことで、マーケティング支援のみならず、店舗・従業員支援へつなげることが可能であると考えられる。それらのシステム開発も今後の課題とする。

参考文献

- [1] 阿部誠, 近藤文代:「マーケティングの科学—POS データの解析」, 朝倉書店, (2005).
- [2] C.M. ビショップ:「パターン認識と機械学習」, シュプリンガー・ジャパン株式会社, (2007).
- [3] Goodman, L.: Analyzing: “Qualitative/Categorical Data: Log-Linear Models and Latent-Structure Analysis,” Abt Books, (1978).
- [4] Hagenaars, J.A. and McCutcheon, A.L.: “Applied Latent Class Analysis,” Cambridge University Press, (2002).
- [5] Hofmann, T. and Puzicha, J.: “Latent class models for collaborative filtering,” *Proc. 16th International Joint Conference on Artificial Intelligence*, (1999).
- [6] Ishigaki, T., Takenaka, T. and Motomura, Y.: “Customer-Item Category Based Knowledge Discovery Support System and Its Application to Department Store Service,” *Proc. IEEE APSCC 2010*, pp. 371-377 (2010).
- [7] 君山由良:「コレスポネンス分析の利用法」, データ分析研究所, (2005).
- [8] 小西貞則, 北川源四郎:「情報量規準」, 朝倉書店, (2004).
- [9] 本村陽一: “ベイジアンネットソフトウェア Bayo-Net”, 計測と制御, Vol. 42, No. 8, pp. 693-694 (2003).
- [10] 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).
- [11] 本村陽一, 石垣司:「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).
- [12] 本村陽一, 西田佳史:「サービス可能知識としての日常生活行動の計算モデル」, 人工知能学会論文誌, Vol. 25, No. 5, pp. 651-661 (2010).
- [13] 内藤耕 (編):「サービス工学入門」, 東京大学出版会, (2009).
- [14] 中村博 (編):「マーケット・セグメンテーション」, 白桃書房, (2008).
- [15] 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).
- [16] Pearl, J.:「統計的因果推論」, 共立出版, (2009).
- [17] 佐藤栄作, 廣松毅, 椿広計: “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).
- [18] 佐藤忠彦, 樋口知之: “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).
- [19] Takenaka, T., Fujita, K., Nishino, N., Ishigaki, T. and Motomura, Y.: “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).
- [20] 豊田秀樹:「共分散構造分析 入門編—構造方程式モデリング」, 朝倉書店, (1998).
- [21] 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).
- [22] Ueda, K., Takenaka, T., Vancza, J. and Monostori, L.: “Value creation and decision-making in sustainable society,” *CIRP Annals Manufacturing Technology*, Vol. 58, No. 2, pp. 681-700 (2009).
- [23] Yada, K., Ip, E. and Katoh, N. “Is this brand ephemeral? A multivariate tree-based decision analysis of new product sustainability,” *Decision Support Systems*, Vol. 44, No. 1, pp. 223-234 (2007).