

アンケートデータを用いた商品認知構造のモデリングによる 継続使用意向の推定

Analysis of Continuous Use by a Bayesian Network Modeling Based on a Questionnaire Data

石垣司*1 本村陽一*1*2 土肥麻佐子*2 持丸正明*1*2
Tsukasa Ishigaki Yoichi Motomura Masako Dohi Masaaki Mochimaru

*1産業技術総合研究所 サービス工学研究センター

Center for Service Research, National Institute of Advanced Industrial Science and Technology

*2産業技術総合研究所 デジタルヒューマン研究センター

Digital Human Research Center, National Institute of Advanced Industrial Science and Technology

The present paper describes an estimation method of continuous use of purchased product with customer's cognitive model in order to know the behavior and satisfaction of customers. The cognitive model is constructed by an integration of questionnaire data and Bayesian network modeling. We can obtain some knowledge with respect to customer's behavior by analysis of the model.

1. はじめに

消費者の価値観やライフスタイルも多様化し、消費行動に関しても大量生産大量消費時代の十人一色の時代から十人十色の時代を経て、現在では時間・状況・気分などの各個人の動的な変数によりその選択行動が変化する一人十色の時代に入っている [井関 91]。同様に企業のマーケティング活動においても、マスマーケティングからマイクロマーケティングへの変遷が生じ、さらには one-to-one マーケティングの必要性が叫ばれて久しい。このような“個”に特化した時代への対応が求められている [吉川 08, 本村 08]。

サービス産業では CRM や顧客満足度の測定を目指したアンケートデータやインタビューデータの蓄積は盛んにおこなわれている。しかしながら、そのようなデータが新たなサービスの設計に上手く活用され、サービス生産性の向上につながっているかどうかは疑問が残る状況である。また、CRM を行うための媒体として情報システムや顧客データベースと共にアンケートデータが注目されているが、従来のクロス集計表による解析や各質問項目間の線形性・独立性を仮定した統計解析手法では、顧客の行動や意思決定のような非線形性や変数間の交互作用を含む事象に対しては取り扱いが困難であった。

そこで本論では、ある商品の購買者に対して行ったアンケートデータを用いた商品の認知構造モデリングを行い、商品の継続使用を促すための顧客が商品に求めている機能の分析を行う。ここでは、各変数の非線形性や変数間の交互作用を明示的に表現可能な統計的モデリングの手法であるベイジアンネットワーク [Jordan98, 繁樹 06, 本村 06] の枠組みと、商品の購入理由や商品継続使用意向を質問項目として含むアンケートデータを機械学習的に統合することで顧客の商品の認知構造のモデリングを行う。作成するモデルの妥当性は継続使用意向をもつ顧客のクロスバリデーションによる判別率で評価する。そのモデルを用いて継続使用意向をもつ顧客に影響を与えている商品購入理由に対する評価と考察を行う。

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 [小西 05] 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。1 にベイジアンネットワークのグラフ構造と条件付き確率表の一例を示す。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 1 内に示すような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 [Murphy99, ピシヨップ 07] 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。

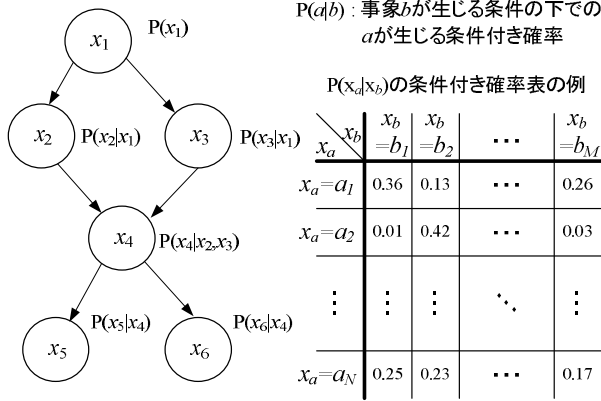
このように柔軟な確率モデルでは、消費者の状態を計算論的に変数として取り扱うことが可能である。かつ、効率的な確率推論のアルゴリズムも開発されているため、ベイジアンネットワークの生活者行動の理解への研究は広く行われるようになってきている [本村 06]。行動履歴データからシステム利用者の行動モデルを作成し、その行動・内的状態を予測することで利用者が欲している情報を提供するリコメンデーションシステムを、携帯電話による映画推奨コンテンツ、カーナビゲーションによる立ち寄り先推奨コンテンツなどに応用されている [本村 06, Motomura05, Ono07]。

3. アンケートデータを用いた商品認知構造モデリング

3.1 実験に用いたアンケートデータ

本論では、総合通信販売大手企業である株式会社千趣会によるオリジナルシューズブランドカタログ「BENEBIS (ベネビス)」に掲載されたカスタマイズインソール (カスタマイズ可能な女性用の靴の中敷き) を対象とする。カスタマイズインソールを購入した 1236 人にアンケート調査の回答を依頼し、

連絡先: 産業技術総合研究所 サービス工学研究センター、〒135-0064 東京都江東区青海 2-41-6、ishigaki-tsukasa@aist.go.jp



上図で表現される確率モデルの同時分布

$$P(x_1, x_2, x_3, x_4, x_5, x_6) = P(x_1)P(x_2|x_1)P(x_3|x_1)P(x_4|x_2, x_3)P(x_5|x_4)P(x_6|x_4)$$

図 1: ベイジアンネットワークのグラフィカルモデル

返答があった 311 人分 (返答率 25.2%) のアンケートデータを使用する。

3.2 商品認知構造のモデリング

本実験ではアンケートデータの質問項目よりデモグラフィック属性 5 変数 (子供の有無、既婚・未婚、職業、年齢、地域)、商品購入理由 12 変数 (価格、機能、サイズ、疲れ軽減、痛み軽減、はき心地、カスタマイズ可能、靴を調整したい、新商品、商品コピー、色、ビジュアル)、商品の使用後の評価 3 変数 (価格の妥当性、はき心地の満足度、総合満足度)、使用前の期待感 1 変数 (はき心地への期待感)、商品の継続使用意向 1 変数の計 22 変数を商品認知構造のモデリングのための変数として使用した。

それらの変数を使用し、赤池情報量規準 (AIC) の基で最適なモデル構造の探索を行う。ここでは、デモグラフィック属性の変数群が他の変数から影響を受けるという状況は自然ではないため、デモグラフィック属性の変数群に関しては他の変数から有向リンクを張られないという制約のみを与え、それ以外のモデル構造に関しては Greedy Search により探索的に AIC の意味で最適なモデル構造の推定を行った。そのグラフィカルモデルを 2 に示す。

作成したモデルを商品の認知構造モデルとして捉えると、各変数はデモグラフィック属性、商品により与えられる機能的ベネフィット、商品により与えられる情緒的ベネフィット、商品使用前の期待感、商品使用後の評価、商品継続使用意向の 6 つの変数群に分類することができる。継続使用意向に関しては価格の妥当性と総合満足度から有向リンクが張られており、直感的にも自然な構造となっている。デモグラフィック属性に関しては、このモデルでは継続使用意向や総合満足度などの満足度評価へつながるリンクは張られていない。

3.3 妥当性の評価に用いる統計的判別器

作成したモデルの妥当性について評価するため、そのモデルを商品の再購買意向を目的変数とするベイジアンネットワーク判別器とみなし、その判別率でモデルの妥当性を評価する。

ここで N 個の入出力のペア $(x^{(i)}, c_i) \in X \times C$, $C = \{c_1, c_2, \dots, c_{N_c}\}$ ($i = 1, \dots, N$) が存在し、 C を判別対象のクラス、 $x^{(i)}$ をサンプル i の特徴ベクトルとする。また、 $P(x^{(i)}|C)$ を特徴ベクトル $x^{(i)}$ に対する各クラスの尤度ベクトル、 $P(C)$ を各クラスの事前分布とする。そのとき、事後分布 $P(C|x^{(i)})$

はベイズの定理より、

$$P(C|x^{(i)}) = \alpha P(x^{(i)}|C)P(C) \quad (1)$$

$$\alpha = \left\{ \sum_{j=1}^{N_c} P(x^{(i)}|c_j) \right\}^{-1}$$

として表現できる。ここで α は正規化項であり事後確率に対して本質的な影響は与えない。本論で用いるベイジアンネットワーク判別器では最大事後確率をもつクラスを判別結果の出力として取り扱う。また、実際の事後分布の計算には確率伝搬法を用いた。

ここでは判別率を比較するため、ナイーブベイズ判別器、数量化 類 [林 74]、カーネル法を用いた SVM [Cristianini00] の 3 つの判別器と判別率を比較する。ナイーブベイズ判別器は一つの目的変数を親ノードとし、説明変数間の局所的な独立性を仮定してそれらの子ノードとする簡単な構造をもったベイジアンネットワークとみなすことができる判別器である。数量化 類はカテゴリーデータのための判別分析であり、カテゴリーデータからダミー変数を用いて 2 値の特徴量を作成し判別分析を行う手法である。ここでは、判別面の構成には線形関数を用いる。SVM は 2 値クラス判別のための学習器であり、カーネル法と組み合わせて使用することで非線形判別が可能となる。現在では、様々な分野でその成功例が報告されている [Muller01]。ここでは、カーネル関数として Gaussian カーネルを用いる。また、SVM のチューニングパラメータ (正則化パラメータとカーネルパラメータ) は cross validation により最適な値を選択する。

3.4 モデルの評価方法

本実験では $N = 311$ 、判別対象のクラスを商品の継続使用意向 $C = \{c_1 = \text{継続使用意向あり}, c_2 = \text{継続使用意向なし}\}$ とする。また、判別率は各クラス c_1, c_2 からそれぞれのサンプルの約 3 分の 2 を学習データとして使用し、残りのサンプルをテストデータとして使用する 3-fold Cross Validation の平均により評価する。

ここでは、全ての説明変数 (21 変数)、購買理由のみの説明変数 (12 変数) の 2 種類の説明変数の組み合わせを用いて、それぞれ判別実験を行った。ただし、3.2 節で作成したモデルにおいては、その変数の組み合わせのみをエビデンスとして与えたときの判別率を評価したことを意味する。他の判別器においては、その変数を用いて判別器を構成し、判別率の評価を行ったことを意味する。つまり、全ての説明変数を使用した実験においては、同じデータにおける条件下での判別率とみなすことができる。

判別実験の結果を 1 に示す。表中の記号 NBC、QT2、BNC はそれぞれナイーブベイズ判別器、数量化 類、作成したベイジアンネットワーク判別器を指す。ここでは 2 つの実験の両者においてベイジアンネットワーク判別器が高い判別率を示している。また、ナイーブベイズ判別器は低い正判別率にとどまっている。これはナイーブベイズ判別器は変数間に独立性を仮定した最尤推定と等しいため、学習データに対して過学習を引き起こしている可能性がある。

3.5 判別結果と判別モデルとしての考察

ナイーブベイズ判別器はそれぞれの変数が独立であると仮定した予測モデルを構築することと等しい。本論で構築したモデルはナイーブベイズ判別器と比べて高い判別結果を示している。そのため、本アンケートデータに対しては変数間に独立性

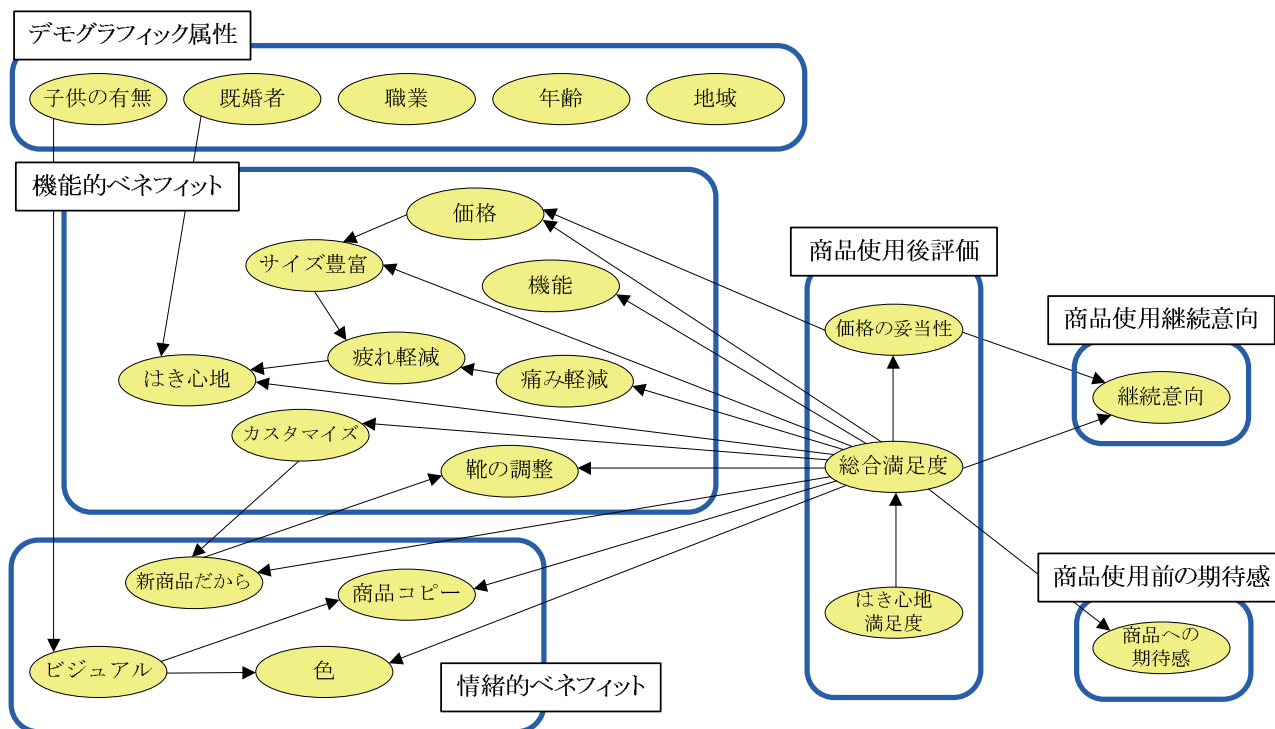


図 2: グラフィカルモデルで表現した商品の認知構造モデル

表 1: 判別実験の結果 (%)

	NBC	QT2	SVM	BNC
全ての説明変数	66.8%	71.1%	74.1%	74.9%
理由のみ	67.0%	71.2%	72.5%	73.5%

表 2: 総合・価格満足度から継続使用意向への感度分析結果

	継続使用意向 の事後確率		
	大	中	小
エビデンスなし	0.685		
総合満足度 (3段階評価)	0.861	0.768	0.340
価格満足度 (3段階評価)	0.821	0.735	0.509

を仮定するよりも 2 のような交互作用をもつ構造をあてはめたほうが説明力が増している。ページアンネットワーク判別器の正答率が高いことが認知モデルとしての正統性とはつながらないが、少なくとも予測モデルとしては説明力が優れていることが実証されている。

4. 認知構造モデルからの知識発見

4.1 認知構造モデルを利用した感度分析

本節では、顧客の継続使用意向に影響を与える変数を探るため、作成したモデルに対して感度分析を行う。作成したモデルではデモグラフィック属性に含まれる変数群は継続使用意向に対してほとんど影響を与えないことが 2 の構造から読み取れる。そのため、ここでは継続使用意向に大きく影響を与える

総合満足度・価格満足度と、商品の特性を表わす顧客の購入理由に注目する。なぜなら、2 の構造から総合満足度と価格満足度に対してエビデンスを与えると、継続使用意向変数とその他の変数は条件付き独立 [宮川 04] となるためである。それらを操作変数とした感度分析の結果を??と??にそれぞれ示す。

本モデルでは継続使用意向以外の変数を周辺化したときに、本商品の継続使用意向をもつ割合が 0.685 である。これらの表中の値は、それぞれの 1 つの操作変数に対してエビデンスを与え、他の変数にはエビデンスを与えない状態での継続使用意向変数の事後確率を表わしている。その結果、総合満足度、価格満足度が高いほど継続使用意向が高いという、当然の結果が再現されている。また、その 2 つの変数に比べ、購入理由変数が継続使用意向に与える影響は小さくなっている。

4.2 商品機能の分析

2 の構造と感度分析の結果から、商品の使用継続意向のある人は商品使用後の総合満足度と価格妥当性が高いと回答した顧客であり、価格妥当性は総合満足度から影響を受けるということがわかる。そのため、継続意向が高い顧客を見つけるためには総合満足度が高くなるような商品が与えるベネフィットを考えればよいことになる。しかしながら、商品購買理由変数から総合満足度へ与える影響の感度分析を行った結果、顕著に説明力の高い変数を見つけることは出来なかった。また、「機能」「疲れ軽減」「痛み軽減」「カスタマイズ可能」を商品購入理由として挙げている顧客は全体の 9 割以上を占めるが、3 の結果からはそれらの変数が継続使用意向の向上に与える影響は大きくない。また、「痛み軽減」「疲れ軽減」を購入理由として挙げていない顧客の継続使用意向は全体と比較し 10% 以上も低減している。そのため、商品を購入した顧客は、商品の機能的なベネフィットをインソール機能に対する付加価値としてではなく、購入の条件として考えていたことが読み取れる。

本章で述べた知見は条件付き確率の値として定量的に評価

表 3: 購入理由から継続使用意向への感度分析結果

	継続使用意向 の事後確率	
エビデンスなし	0.685	
購入理由	あり	なし
はき心地	0.685	0.592
サイズ	0.706	0.609
価格	0.720	0.597
機能	0.686	0.620
疲れ軽減	0.694	0.572
痛み軽減	0.700	0.564
カスタマイズ	0.695	0.658
調整	0.675	0.700
新商品	0.685	0.686
ビジュアル	0.696	0.673
色	0.694	0.674
商品コピー	0.673	0.742

可能である。そのため、本手法で商品の認知モデルを構築することで、商品への期待度と継続使用意向で示したような非線形な変数間の関係についても条件付き確率として定量的に把握することが可能である。

5. まとめ

本論ではアンケートデータを用いた商品の認知構造モデリングの方法論とそれを利用した商品の継続使用を促すための顧客が商品に求めている機能の分析を行った。カスタマイズインソールの購買者に対して行ったアンケートデータとベイジアンネットワークを機械学習的に統合することで顧客の商品の認知構造のモデリングを行い、そのモデルの妥当性は継続使用意向をもつ顧客のクロスバリデーションによる判別率で評価した。その後、作成したモデルを用い、商品に求められている機能の分析を行った。本論ではカスタマイズインソールという特定の商品についての分析を行ったが、本論で展開した方法論は他のアンケートデータや商材に対しても親和性が高く、応用展開が可能である。

6. 謝辞

アンケートデータを提供していただいた株式会社千趣会に心より感謝の意を表す。また、本研究は経済産業省サービス研究センター基盤整備事業の委託研究費を受けている。

参考文献

- [井関 91] 井関利明, 室井鐵衛: 生活者起点発想とマーケティング革新, 国元書房 (1991)
- [Ono07] C. Ono, M. Kurokawa, Y. Motomura and H. Asoh: A context-aware movie preference model using a Bayesian network for recommendation and promotion, Proc. of User modeling 2007, LNCS, Vol. 4511, pp. 257-266, Springer (2007)
- [Cristianini00] N. Cristianini and J. S. Taylor, An Introduction to Support Vector Machines and Other Kernel-

based Learning Methods, Cambridge University Press (2000)

- [小西 05] 小西貞則, 北川源四郎: 情報量規準, 朝倉書店 (2005)
- [繁樹 06] 繁樹算男, 植野真臣, 本村陽一: ベイジアンネットワーク概説, 培風館 (2006)
- [Jordan98] M. I. Jordan: Learning in graphical models, MIT Press (1998)
- [林 74] 林知己夫: 数量化の方法, 東洋経済新報社 (1974)
- [ピシヨップ 07] C. M. ピシヨップ: パターン認識と機械学習 (上・下), シュプリンガー・ジャパン株式会社 (2007)
- [Murphy99] K. Murphy, Y. Weiss, M. I. Jordan, Loopy belief propagation for approximate inference: An empirical study, Proceedings of Uncertainty in Artificial Intelligence (1999)
- [Muller01] K-R. Müller, S. Kika, G. Ratsch, K. Tsuda and B. Schölkopf, An introduction to kernel-based learning algorithms, IEEE Trans. Neural Networks, vol. 12, no. 2, pp. 181-202 (2001)
- [宮川 04] 宮川雅巳: 統計的因果推論, 朝倉書店 (2004)
- [Motomura05] Y. Motomura and T. Kanade: Probabilistic human modeling based on personal construct theory, Journal of Robot and Mechatronics, Vol. 17, No.6 pp. 689-696 (2005)
- [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 08] 本村陽一, 西田佳史, 持丸正明, 橋田浩一, 赤松幹之, 内藤耕, サービスイノベーションのための大規模データの観測・モデリング・サービス設計・適用のループ, 人工知能学会誌, Vol. 23, No. 6, pp. 736-742 (2008)
- [吉川 08] 吉川弘之: サービス工学序説 サービスを理論的に扱うための枠組み, 人工知能学会誌, Vol. 23, No. 6, pp. 714-720 (2008)