

ベイジアンネットワークを用いた消費者行動モデルの構築実験

Consumer behavior modeling using bayesian networks

村上知子 酢山明弘 折原良平
Tomoko Murakami Akihiro Suyama Ryohei Orihara

(株) 東芝 研究開発センター
TOSHIBA Corp. Research & Development Center

In this paper we describe a consumer behavior model using bayesian networks. The model aims to support the marketers. Answer data collected by a questionnaire surveys of the market is often incomplete. It is difficult to collect latent information such as a state of consumers' mind by a questionnaire survey. Considering these facts we apply bayesian networks to construct a consumer behavior model. The merits of the method are to utilize knowledge of domain's specialist, to introduce unobservable variable and the robustness against the lack of data.

We decide the structure of the model utilizing answer data collected by a questionnaire survey to investigate lifestyle of the consumers. Conditional probability for each nodes in the network is calculated based on frequency of the data for observable node and by expectation maximization(EM) algorithm for unobservable nodes.

We perform some experimental studies by probability inference for this model to specify the loyal consumers and to estimate if a brand arouses target consumers' interest.

1. はじめに

POS システムに代表されるデータ獲得技術の進化によって収集された顧客のプロファイルデータや購買履歴データに対して、データマイニング技術を用いた顧客の購買行動の分析が進められてきた [Fayyad 01]。これにより売上に大きく貢献する優良顧客や将来的に優良顧客になる可能性を持つ潜在顧客の特徴などを把握し、ターゲット市場の要求に答える商品の企画・開発を進めることが可能になった。

ところが、消費者の嗜好が多様化し消費者層から個人消費者の嗜好に適応した商品の供給が求められるようになったため、従来の顧客のプロファイルデータや購買履歴データのみでは分析対象データとして十分でなく、消費者の購買に至る心理や内部状態まで深く分析する必要性が生じている。消費者の心理や内部状態が明らかになれば消費者の購買行動を的確に予測したり購買要因を推測したりすることが可能になるため、これらの結果をターゲット顧客や潜在顧客の嗜好にかなう商品開発に生かすことができる。

一方、商品企画・開発に携わる専門家は消費者の購買に至る心理や内部状態まで深く分析し、的確な商品のターゲット市場を定め企業利益を最大化するマーケティング戦略を立てるため、消費者の生活動向を調査するアンケートを実施している。アンケート調査は、比較的容易に大規模な調査が行いやすいものの回答作業の負荷や心理的な回答の困難さによる回答の欠損が多く、心理的・内的状態を聞き出すことが難しいことが知られている [後藤 96]。

本論文では、このようなアンケート調査で収集されるデータの性質を考慮し、ベイジアンネットワークを用いて構築した

消費者行動モデルに関して報告する。消費者行動モデルの構築は、消費者の意識、生活、行動に関するアンケートデータと商品企画の専門家のネットワーク構造知識を用いて決定した。また、このベイジアンネットワーク行動モデルに対して確率推論を行い、購買に至る顧客層の推定や購買要因の特定のための実験に関しても報告する。

2. 消費者行動分析

2.1 消費行動モデル

従来マーケティングリサーチや社会心理学の世界では、消費者の選好に影響を及ぼすと思われる欲求や状況の依存性、購買を決定するメカニズムに関して様々な仮説が導入されてきた [Nelson 70]。消費者の状態を Attention Interest Desire (Memory) Action (Satisfaction Level) と順を追って理解する AID(M)A(S) モデルを始めとして、消費者の製品使用の結果が次回の購買行動に影響を及ぼす側面を重視した Nicosia モデル [Nicosia 66]、外部からの刺激に対する消費者の購買行動を考慮した HowardSheth モデル [Henrion 88]、主体的に情報を探索し意思決定を行う者として消費者を捉えることを特徴とした Bettman モデル [Bettman 79] などが提案された。しかし、これらのモデルは商品毎の意思決定や情報収集の違い、意思決定における消費者の内面の様子など様々な要素を取り込んでモデル化しようとしていることが評価できる反面、直接観測が困難な要素が多いため全顧客を対象としたモデル化やデータによる仮説の検証が困難であった。

一方、データマイニング研究においては、顧客のプロファイルデータや購買履歴データ、消費者の生活動向に関するアンケートデータなどに対して、ひとつの目的変数とそれを説明する説明変数を定め、決定木 [Quinlan 93] やニューラルネットワークなどのデータマイニング技術を用いて目的変数を最もよく説明できるモデルの獲得とそれを用いた消費者行動の解釈が行われてきた。このような分析で利用する顧客に関するデータ

連絡先: (株) 東芝 研究開発センター 知識メディアラボトリー
住所: 〒 212-8582 川崎市幸区小向東芝町 1
電話番号: 044-549-2398
メールアドレス: tomoko.murakami@toshiba.co.jp

は一般的にアンケート調査によって収集されるが、アンケート調査には以下のような特徴が挙げられる。

- 回答作業の負荷が大きいため回答の欠損が多い。
- 心理的に回答しづらい設問に対しては信憑性が低い。
- 心理的・内的状態を聞き出すことは難しい。

上記の特徴からアンケートデータの性質を考慮し、消費者行動分析の手法としてベイジアンネットワークを採用した。ベイジアンネットワークには以下のような特長がある。

- 多くの欠損データ、観測が困難な要素を扱うことが可能
- 推測される仮説の確信度を実データに基いて検証可能
- 専門家の持つノウハウをネットワーク構造として導入することが可能

ベイジアンネットワークは前提知識に基づくモデル構築とデータからの学習の両方の側面を持っている。そのため、不完全なデータや観測不可能なデータにも対応可能であり、消費者の購買に至る心理や内部状態などの観測が困難な要素も扱うことができる。また、専門家の持つノウハウをネットワーク構造として導入することが可能なため、推測される仮説を実データに基づいて検証することができるという利点もある。

2.2 ベイジアンネットワーク

ベイジアンネットワークは確率変数間の定性的な依存関係をグラフ構造として記述し、グラフ構造で定義された個々の確率変数間の関係を条件付確率として記述したネットワーク構造を使って問題対象を表現する確率モデルである [本村 03]。ベイジアンネットワークにおける各確率変数間の因果関係を条件付確率表 (CPT) として記述することにより、決定木や回帰モデル、共分散構造分析などと比較して柔軟なモデルを構築することが可能である [本村 03]。また、CPT の各項目に該当する十分なデータが存在すれば、確率変数の各状態の頻度を計算し、

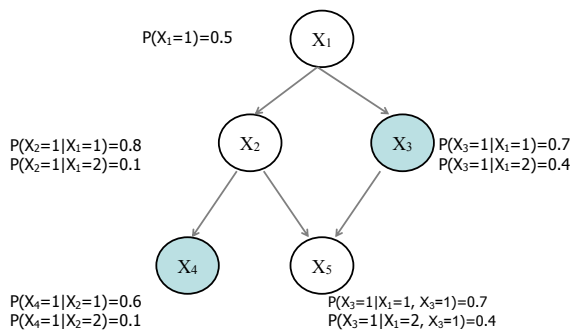


図 1: ベイジアンネットワークの例

表 1: サンプリングの例

サンプル	X_1	X_2	X_3	X_4	X_5	スコア (重み)
1	2	2	1	2	1	0.36
2	1	1	1	2	2	0.28
3	2	1	1	2	2	0.16
4	1	1	1	2	1	0.28
⋮	⋮	⋮	⋮	⋮	⋮	⋮

各項目の値を求めることができる。

ベイジアンネットワークを用いたモデルを構築する際には、

- 何を確率変数として設定するか (グラフ構造)
- 確率変数間の因果関係 (CPT) の値はどのように決定するか (CPT 値)
- どのように確率推論するか (確率推論)

の 3 つが大きなポイントになる。

本研究では、商品企画の専門家の知見を用いて確率変数と変数間の因果関係 (アーク) の存在を決定した。また、変数間の因果関係の度合い (CPT の各項目の値) に関しては、観測変数間の値はアンケートデータにおける頻度分布に従い設定した。 χ^2 乗適合度検定より、CPT の各項目に該当するデータが 5 個以上存在しなければ統計的な正しさが保証できないことが知られているが、今回のアンケートデータでは CPT の各項目に該当するデータ数は上記の条件を満たしている。一方、観測変数と潜在変数間の値はデータが採取できないのでデータの確率分布を仮定し、その分布による期待値計算を行うことが必要になる。本研究では、観測変数と潜在変数間の値は EM アルゴリズム [Dempster 77] を用いて確率分布を推定し、CPT の値を計算した。

確率変数と確率変数間の依存関係を表すグラフ構造、条件付確率によってベイジアンネットワークが定義されれば、これに対する確率推論によって知りたい対象の事後確率を求め、それによって仮説の確信度 (いくつかの確率変数が同時に特定の値を持つ確率) を評価することができる。

ベイジアンネットワークによる確率推論には、確率伝播法をはじめとする様々な手法が提案されている。本研究では厳密計算 [Pearl 88, Jensen 01] に対する近似解法として、logic sampling 法 [Henrion 88] とそれを高速化した likelihood weighting 法 [Fung 89] を採用した。

likelihood weighting 法 [Fung 89] は、例えば図 1 に示すように、5 つの 2 値の確率変数 $X = X_1, X_2, \dots, X_5$ から構成され、そのうち $X_3 = 1, X_4 = 2$ のようにインスタンス化されたベイジアンネットワークが与えられた場合に、ネットワークの確率分布に従い表 1 に示すようにデータ集合をサンプリングする。その際、インスタンス化された確率変数には固定の値が割り当てられ、各サンプルに対して尤度に従った重みを付与する。そして、該当するサンプルの重みの加重平均を計算することにより仮説の確信度を求める。例えば、表 1 に示

すようにサンプリングされた場合に $X_1 = 1$ である確率は

$$P(X_1 = 1 | X_3 = 1, X_4 = 2) = \frac{0.28 + 0.28}{0.36 + 0.28 + 0.16 + 0.28} = 0.52$$

のように求められる。

3. 検証実験

パーソナルコンピュータ (PC) の市場調査のためのアンケートデータと PC のマーケットの知見を導入して消費者行動モデルを構築し、購買に至る顧客層の推定や購買要因の特定のための実験を行った。

利用したアンケートデータの詳細を以下に示す。

- 時期 : 2000 年実施
- 対象 : 3,000 人 (20~59 歳男女)
- 質問項目 : デモグラフィック属性 (年齢, 性別 etc.) , プロファイル属性 (家族構成, 趣味 etc.) , 価値観, ライフスタイル, 好きな番組, タレント, 商品購買時の意識, PC 利用実態

本実験で扱う消費者行動モデルは、6 個の消費者プロフィールに関するノード、3 個の消費者クラスタに関するノード、4 個の購買動機、購買結果に関するノードで構成されている。このうち消費者クラスタに関するノードとして「周囲への追従度」「市場リーダ度」「PC 習熟度」という潜在変数を導入した。これは、PC などの情報機器は市場を引っ張るイノベータ型消費者と周囲に追従して購入するフォロワーズ型消費者 (いち早く反応する消費者とだいたい普及してから追従する消費者を含む)、2 種類の消費者層の間で橋渡し役的な行動をとるブリッジピープル型消費者、周囲の意見に影響を受けないアイソレートッド型消費者が存在する、また、PC リテラシーの習熟度により購買行動に差があるという PC の商品企画の専門家の持つ知識を利用して導入したものである。

本消費者行動モデルに対して、計算における条件式やアンケートデータとして観測可能な変数をインスタンス化して確率推論を実行した結果得られる確率値を利用し、

- 優良顧客の特徴
- 購買要因の推定
- ポジション比較

を実験によって検証した。

3.1 実験 1 : 優良顧客の特定

消費者行動モデルに対して将来的な PC 購入予定確率に関して推論を実施し、消費者全体に対する T 社の PC 購入予定確率 ($P(T \text{ 社製 PC 購入予定})$) より大きい確率を示す消費者の特定を試みた。

この結果、将来的な T 社製 PC 購入予定確率は 3.59 (%) であり、全体の消費者の 3.59% が T 社製 PC の購入を検討していることが分かった。自営業者、学生、専門職・自由業の T 社製 PC 購入予定確率はそれぞれ (3.865%)、(3.745%)、(3.455%)

表 2: T 社の PC 購入予定者の特徴

購入者特徴	購入確率 (%)
流行や人の意見に追従して購入を決める流行追従型消費者	24.1
経済力があり新しいモノはいち早く体験する既婚で子供のいない男性会社員	4.57
経済力があり新しいモノはいち早く体験する 20 歳代の女性	4.36

となり、職業により購入予定率に若干の変化が見られたが、その他のデモグラフィック属性で購入予定率に差はなかった。

しかしながら、生活や生活態度状況に関する属性で大きく変化が見られた。将来的な T 社製 PC 購入予定確率が高い消費者特徴として表 2 のような結果が得られた。表 2 から、「周囲への追従度」すなわち流行や人の意見に従って購入を決めるか否かという軸 (隠れ変数) に関して有意な差が得られたことが分かる。その他の隠れ変数として導入した「市場リーダ度」すなわち周囲より先んじて購入し市場を牽引する消費者か否かという軸や「PC 習熟度」すなわち PC リテラシーや技術に熟練しているか否かという軸によって特徴付けられる顧客層に有意な顧客層は発見できなかった。

3.2 実験 2 : 購買要因の推定

消費行動モデルに対して確率推論を実施し、購買予定の PC のメーカー別に購買動機の確率を計算し比較した。例えば、T 社製 PC の購入予定者が過去の購入経験をもとに購入する確率、 $P(\text{過去の購入経験} | T \text{ 社製 PC 購入予定})$ を求めた。メーカーは T 社、N 社、F 社、S 社、I 社、D 社、A 社の 7 社、購買動機として「最高の品質性能」「顧客サービスの良さ」「納得できる価格」「過去の購入経験」「技術先端性」「多様な機能・楽しめる工夫」「耐久性・丈夫さ」「個性的なデザイン」の 8 種類を選択した。

結果を図 2 に示す。図 2 から、T 社製 PC は他社と比較し

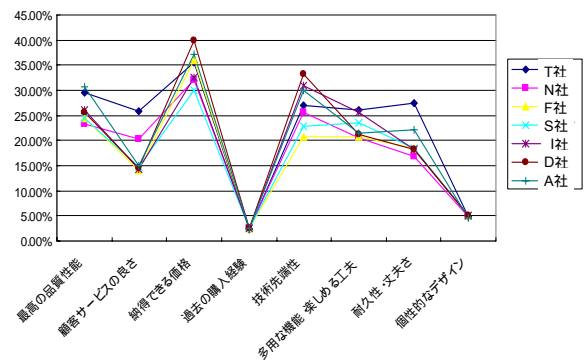


図 2: メーカー別購買要因の推定

て「顧客サービスの良さ」「耐久性丈夫さ」で圧倒的に支持が高く、「多様な機能楽しめる工夫」「最高の品質性能」で高い支持を得ていることが明らかになった。

3.3 実験3：ポジション比較

消費行動モデルに対して確率推論を実施し、ある消費者層におけるメーカーのシェアを確率として計算し比較する実験を行った。例えば、20歳代で流行をいち早く取入れたいと思っている人がT社製PCを購入予定である確率は、 $P(T社製PC購入予定 | 年齢=20歳代 \text{ and } 流行をいち早く取入れる)$ として計算する。実験2と同様に7社のメーカーで比較し、消費者層として「年齢」「経済力」「新しいモノに対する感度」「周囲への追従度」「市場リーダ度」などから2種類の軸を選択して比較を行った。

その結果、ほとんどの市場においてN社、F社、S社が3強であることが分かった。例外として、「周囲への追従度」別の消費者に関して大きく差が見られ、「周囲への追従度」が高い消費者はN社、F社、S社、T社製PCを購入予定であるが、「周囲への追従度」が低い消費者はA社やD社の製品を購入予定であることが分かった。また、「性別」と「年齢」の2軸を選択してシェアを比較した場合には、若い女性、特に20歳代の女性がA社の製品を購入していることも明らかになった。図3は「年齢」(横軸)と「新しいモノに対する感度」(縦軸)を選択した場合の各消費者層におけるメーカーのポジション比較を可視化した結果である。

4. おわりに

本論文では、アンケート調査で収集されるデータの性質を考慮し、ベイジアンネットワークを用いて構築した消費者行動モデルに関して報告した。消費者行動モデルは消費者の意識、生活、行動に関するアンケートデータと商品企画の専門家のネットワーク構造知識を用いて決定し、モデルに対して確率推論を実施することにより購買に至る顧客層の推定や購買要因の特定のための実験を行った。

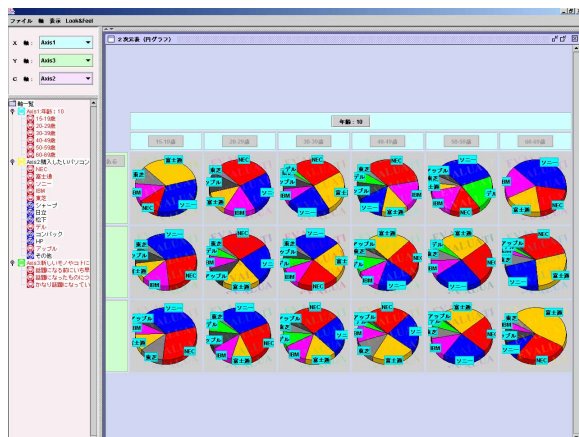


図3: 「年齢」と「新しいモノに対する感度」別消費者層におけるポジション比較

今後はさらにモデルを改良して実験を継続していく。それとともに、継続した消費者行動調査のデータ解析のためのアルゴリズム、結果に強く結びつく要因の効率的な発見手法、EMアルゴリズムの高速化を検討する予定である。

参考文献

[Fayyad 01] Fayyad, M. Usama et al.: "Information Visualization in Data Mining and Knowledge Discovery", Morgan Kaufmann Series in Data Management Systems, (2001).

[後藤 96] 後藤秀夫: "市場調査ケーススタディ 改訂新版", 日本マーケティング教育センター, (1996).

[Quinlan 93] Quinlan, J.R. *C4.5: Programs for Machine Learning*, Morgan Kaufmann, San Mateo, CA, 1993.

[Nelson 70] Nelson, P.: "Information and Consumer Behavior", *Journal of Political Economy*, Vol.78, pp.311-329, (1970).

[Nicosia 66] Nicosia, F. M.: "Consumer Decision Processes", Englewood Cliffs, NJ: Prentice Hall, (1966).

[Howard 69] Howard, J. A. and Sheth, J.N.: "The Theory of Buyer Behavior", John Wiley & Sons, Inc., (1969).

[Bettman 79] Bettman, J. R.: "An Information Processing Theory of Consumer Choice", Addison-Wesley, (1979).

[本村 03] 本村 陽一: "特集「ベイジアンネット」ベイジアンネットソフトウェア", *人工知能学会誌*, Vol.17, No.5, pp.559-565, (2002).

[本村 03] 本村 陽一: "ベイジアンネットによるヒューマンモデリング", *ベイジアンネットセミナー資料*, pp.79-84, (2003).

[Pearl 88] Pearl, J.: "Probabilistic Reasoning in Intelligent Systems", Morgan Kaufmann, CA, (1988).

[Jensen 01] Jensen, F. V.: "Bayesian Networks and Decision Graphs", Springer Verlag, (2001).

[Henrion 88] Henrion, M.: "Propagation of uncertainty by probabilistic logic sampling in Bayes' networks", In J. F. Lemmer & L. N. Kanal (Eds.), *Uncertainty in Artificial Intelligence 2*, pp.149-163, (1988).

[Fung 89] Fung, R. and Chang, K. C.: "Weighting and integrating evidence for stochastic simulation in Bayesian networks", In *Proc. of Conf. UAI-89*, (1989).

[Dempster 77] Dempster, A. et al.: "Maximum likelihood from incomplete data via the EM algorithm", *Journal of the Royal Statistical Society*, B39, pp.1-38, (1977).