

ベイジアンネットワークを利用した ユーザ適応システム

本村 陽一

ベイジアンネットワークの統計的学習によって大規模データから利用者のモデルを構築し、それを活用することでユーザに適応して動作するシステムの研究がある。本稿では情報推薦システムや顧客分析システム、需要予測システムなどの例を紹介する。また、こうした日常生活空間における大規模データから人に関する計算モデルを構築し活用する「実サービスを通じた調査・研究」と人の相互作用であるサービスを対象にするサービス工学についても触れる。

キーワード：ベイジアンネットワーク、ユーザモデル、サービス工学

1. はじめに

「『モノ』から『コト』へ」、というフレーズなどにも表れているように、これまでのモノ中心のパラダイムから人や経験を重視する世界観への転換が進んでいる。社会における人の相互作用であるサービスに対する工学を考えるとというサービス工学[1][2]もそうした流れの一つである。サービスの特性はサービス提供と消費が同時に行われ、その品質はサービス利用者や状況に依存し、保存できないという同時性、異質性、消滅性に起因している。また、提供されるサービスの評価は利用者（ユーザ）と状況に依存していることから、現場の熟練者の属人的な経験と勘への依存性が強く、サービスが複数の人の相互作用（コト）として伝搬しており、人の心理や行動、状況といった情報の技術的な取り扱いが本質的な課題になっている。

こうしたサービス特有の課題に対して、熟練者の経験と勘を脱却し、サービスの現場で起こる事象を客観的に「観測」し、それを「分析」して得られる計算論的モデルに基づいて、あるべきサービスを再「設計」し、それを現場に「適用」という「最適設計ループ」によって、サービスを連続的に改良するアプローチが提案されている[3]。この最適設計ループは制御工学的なアプローチでもあるが、サービス特有の課題に対処するための独自のモデル化が必要になる。具体的には、利用者の集団を対象にすることから来る不確実性や利用者の心理的要因や状況依存性に基づく不確

実性などに対処するために、大規模データに基づいてユーザに適応するシステムと、それを実現するための利用者のモデル構築・活用技術が必要となる[4]。本稿では、大規模データから自動的に構造化したモデルを構築し、さらに構築した計算モデルの上で確率的推論を実行できるベイジアンネットワーク技術[5]とそれを活用したさまざまな実用システムを紹介する。これはサービスの現場でリアルタイムに集積されるデータを分析し、最適化を実行するシステムである。さらに、こうした大規模データを活用する技術を社会実装するために重要となるサービスや利用者との相互循環的な枠組みについても議論する。

2. ベイジアンネットワーク

日常生活中における現象をモデル化するためには人の行動や心理などからくるさまざまな不確実性に対処することが必要であるため確率モデルを使って対象を

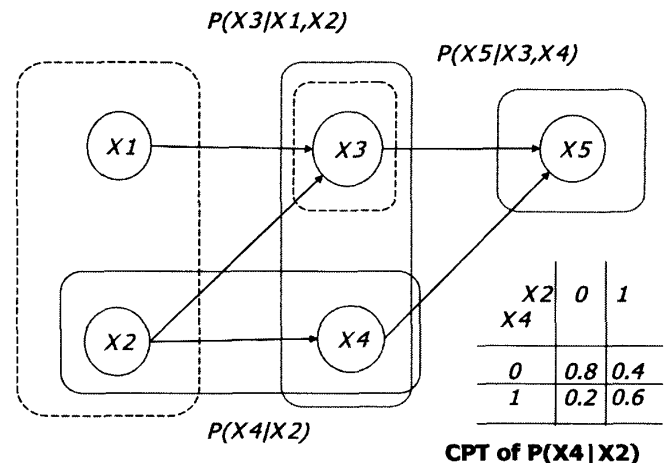


図1 ベイジアンネットワーク

もとむら よういち

(独)産業技術総合研究所 サービス工学研究センター
〒135-0064 江東区青海 2-3-26

モデル化することで、知りたい変数の確率分布を推定し、おこりえる各状態の確率（確信度）を評価する枠組みが有効である。最近のITの普及により大量の統計データから経験的にノンパラメトリックな確率分布を構成することが容易になり、経験ベイズ的といわれるアプローチが実用的な方法として有望視されている。また確率分布間の関係として変数間の条件付分布を考えこれを構造化したモデルにベイジアンネットワークがある。

例えば不確定さを含む対象を確率変数として X で表し、その変数がとりうる具体値を x_1, x_2, \dots, x_n と表すことにする。次に変数間の依存関係を考える。例えば変数 X_i が x という値を取るならば、 X_j は y となる、という関係が成立しているとき、 X_j が X_i に依存していると考える（if $X_i=x$ then $X_j=y$ ）。現実には起きている複雑な事象を考えると、複数の変数間の依存関係は複雑になり、「if $X_1=x_1, \dots, X_i=x_i, \dots$, then $X_j=y$ 」のように明示的に全ての関係を列挙することはあまり現実的でない。また、たとえこのようなIF-THENルールを膨大に挙げたとしても実際には例外などがあり、必ずしも完全に状況を記述することは難しいだろう。そこで厳密な表現をあきらめ、主要な変数のみに注目し、ルールが成立する確信の度合いを定量的に表すために「 $X_i=x_i$ であるとき $X_j=y$ である確率は $P(X_j=y|X_i=x_i)$ 」という確率的な表現を導入する。二つの量 x, y の間の一意的な依存関係は、例えば関数 $y=f(x)$ によって表せるが、これと同様に、確率変数 X_i, X_j の依存関係は条件付確率分布 $P(X_j|X_i)$ によって表すことができる。これは X_i の取る値に応じて、 X_j の分布が影響をうけ、その依存関係の定量的関係が条件付確率分布 $P(X_j|X_i)$ で定められることを示している。

複数の確率変数の間の定性的な依存関係をグラフ構造によって表し、個々の変数の間の定量的な関係を条件付確率表で表したモデルがベイジアンネットワークである。この確率モデルは数学的には確率変数をノードで表し、これらを有向リンクで結合して依存関係を表現した確率分布として定義される。有向リンクの元にあるノードを親ノード、有向リンクの先にあるノードを子ノードと呼ぶ。有向リンクは親から子の向きに条件付の依存関係があることを示し、子ノード X にリンクを張る親ノード（集合）を U とすると、この子ノードの確率変数は条件付確率分布 $P(X|U)$ に従う。確率変数が k 通りの状態を持つ確率変数の場合、

子ノードは $X=x_1, \dots, x_k$ のそれぞれの値を取る可能性のあるものと考え、それぞれの値を取る確率が $P(x_1), \dots, P(x_k)$ であれば、これにより X の確率分布を与えることができる。離散的な確率変数ならば親ノード（集合）についても取る値の全ての組み合わせを列挙することができるので、 X についての確率分布が親ノード U に依存していれば、その条件付確率分布を考えて、全ての U の取り得る値の組み合わせについての確率値 $P(x_1|U), \dots, P(x_k|U)$ を並べた表、条件付確率表（CPT）としてこれを定義することができる。つぎに変数間の依存関係、つまり各子ノードについての親ノードが結合しているかという親ノードの集合を定義するとベイジアンネットワークのグラフ構造が決定する。つまりベイジアンネットワークのモデルは、ノード集合とグラフ構造と、各子ノードにそれぞれ一つ割り当てる条件付確率表（CPT）の集合によって規定できる（図1）。

ベイジアンネットワークの中の一つの子ノードに注目した依存関係、つまり一つの目的変数（従属変数： Y ）と、それに対する説明変数（独立変数： X ）の間の依存関係について着目すると X - Y 空間を条件付確率表にしたがって量子化し、個々の確率値を割り当てたものになっており、これにより任意の非線形性、非正規性を表現できる。また複数の親ノードによる交互作用を表せること、つまり複数の親を持つ場合、非線形を持つ交互作用も表すことができることが現実の社会で起こる多様な事象をモデル化するために有用な特長になっている。日常生活場面では個人差、状況依存性などを反映する必要がある、この点で交互作用や非線形性、非正規性を含むモデル化が果たす役割が非常に大きい。また実データからモデルを自動的に構築し、さらにそのモデル上で確率推論を連続して実行できるソフトウェアが提供されていることも、さまざまな情報サービスを開発する上では重要な特長である。

3. ユーザ適応型情報推薦システム

ベイジアンネットワークモデリングと、その上で確率推論を実行することで、消費者の嗜好性や意図などを推定することができる。そこでこれを用いて情報推薦に応用できる。特に新規のコンテンツに対応できるようにコンテンツ属性を変数として用い、さらにユーザ属性や状況を表す変数もベイジアンネットワークのノードとしてモデルに組み込むことで、状況やユーザの傾向に応じた推薦が可能になる。筆者らとKDDI

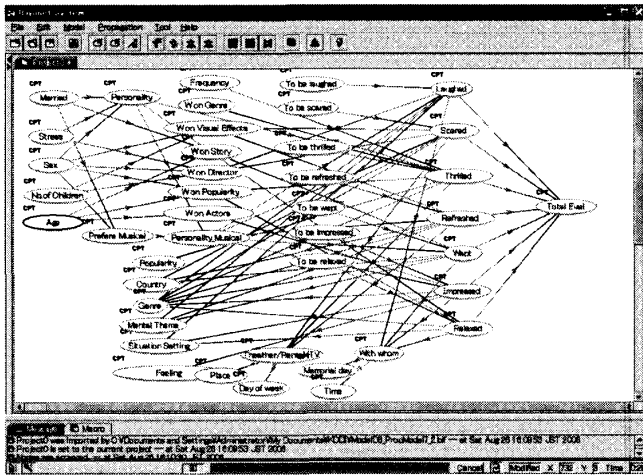


図2 映画推薦のために構築したベイジアンネットワーク

研究所のグループによる、携帯電話サービスのためにベイジアンネットワークを用いた映画コンテンツを推薦する事例[6][7]では約1,600名の被験者に対して映画コンテンツを提示するアンケート調査により収集したユーザ属性、コンテンツ属性、コンテンツ評価履歴からベイジアンネットワークモデルを構築した(図2)。アンケートでは年齢・性別・職業などのデモグラフィック属性の他にライフスタイルなどに関する質問項目、さらに映画視聴に関する態度属性として鑑賞頻度、映画選択時の重視項目、映画を見る主要目的(感動したい等7項目)、コンテンツに対する評価(良い・悪い)、そのときの気分(感動した等7項目)などを収集し、さらに約1,000人について別途、各映画コンテンツについて、どんな気持ちや状況で、どこで(映画館、DVDで家)、誰と何人で、どんなときに、鑑賞するか、を自由記述文により収集した。

このような実際に映画を選ぶ際の状況まで含んだデータを収集したことで、状況とユーザの嗜好性に応じて映画を推薦する携帯情報システムのプロトタイプを開発できた。ユーザが携帯電話からサービスへの要求を状況に関する情報とともに送ると、システムはデータベースから登録済みのユーザ属性情報と状況情報を使って確率推論を実行する。その結果選択される確率が高いと判断されたコンテンツを上位から推薦する。

この映画推薦システムはインターネットサービスにも発展し、auoneラボ(<http://labs.auone.jp>)において2007年から一般に公開されるべ約7,000件の推薦を実行した。その推薦履歴からさらにモデルの再学習を行うことで推薦精度が向上する。またここで構築したモデルは映画の属性やユーザ属性、状況属性が異なる用途でも再利用可能な変数として表現されているため、同じモデルを用いて、あるコンテンツに対してもっとも視聴する確率の高いユーザ層や、その際の状況などを推論することができる。つまり映画として公開が終わった後に、同じ映画のDVD販売戦略の最適化などに再利用することができる[8]。KDDI研究所と産総研、松竹の3者はこの仕組みを用いて、実際にある映画のDVD販売施策支援を行った。これは情報推薦でよく用いられる協調フィルタリングでは実現することが難しい、大規模データから構築した計算モデルの再利用性を活用した事例といえる。

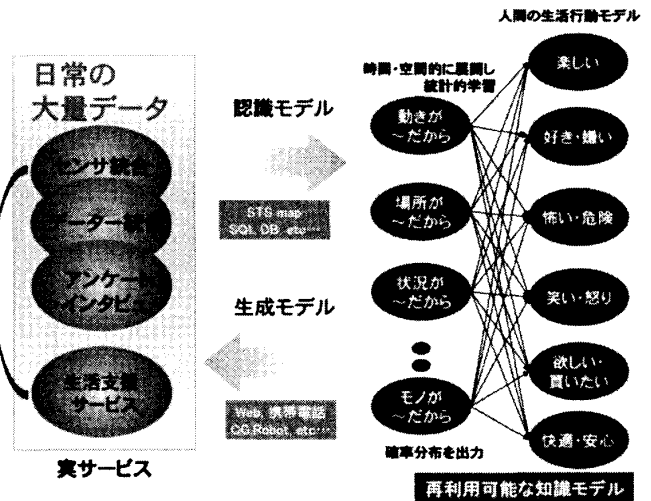


図3 日常の大量データからの生活行動モデルの構築

る用途でも再利用可能な変数として表現されているため、同じモデルを用いて、あるコンテンツに対してもっとも視聴する確率の高いユーザ層や、その際の状況などを推論することができる。つまり映画として公開が終わった後に、同じ映画のDVD販売戦略の最適化などに再利用することができる[8]。KDDI研究所と産総研、松竹の3者はこの仕組みを用いて、実際にある映画のDVD販売施策支援を行った。これは情報推薦でよく用いられる協調フィルタリングでは実現することが難しい、大規模データから構築した計算モデルの再利用性を活用した事例といえる。

4. サービスにおける利用者モデル

POSシステムを採用しているスーパーなどの小売店では、店舗営業終了後にID付POSデータなどとして数十ギガバイトにも及ぶ1日間の全ての販売記録がデータセンターへと送られる。そのため、データの収集は常時行われており、分析対象のデータ量は大量となる。このような大規模データはまずはデータベースへの保存と格納、次に平均や分散といった古典的な統計量や会計的指標の計算、さらにはクラスタ分析や線形解析などに利用されている。この大規模データを分析し、顧客セグメントを可視化することでサービスの生産性向上に寄与することもできる。

従来のID付POSデータを用いた顧客セグメンテーションの代表的なものはRFM分析と呼ばれる、直近の来店日、来店回数、購入金額などの数値指標で顧客の特性をモデル化するものであり、そこから直接ライフスタイルや商品ニーズなど真の生活理解を行うことは難しい。さらに、顧客のセグメント化のみなら

ず、商品のセグメント化も重要な課題である。大規模な小売サービス業では商品数が数千から数万点に及び、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のが現状である。またカテゴリマネジメントに代表される顧客視点での商品管理の重要性も説かれているが、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大きい。

そこで、顧客と商品の双方について、有効な施策の実施が可能な顧客セグメントや商品カテゴリを、大規模データから自動的に抽出する技術が重要になる。そこで得られたカテゴリに基づいた購買状況や購買パターンとの関係性をモデル化し顧客行動を理解することで、提供するサービスの生産性と価値の増大につなげることができる。そこで、サービス現場で観測される大規模データから実務に利用可能な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技術として多層の潜在クラス分析とベイジアンネットワーク、定量的ライフスタイル調査を組み合わせた顧客モデル化技術が開発されている。ここでは、ID付POSデータから顧客の購買行動をモデル化した事例[9][10]を紹介する。

まず、潜在クラスモデルの一種である確率的潜在意味解析[11]により、顧客ごとの商品の買い方、商品から見ると買われ方が似ているものを同一クラスに併合する操作を繰り返すことによってID付POSデータから顧客と商品を同時分類しカテゴリ生成を行う。最適なカテゴリの数は事前に決めることは難しいため、情報量規準(AIC)に基づいて最適なクラス数を探査する。このアルゴリズムによって、数千から数万の膨大な数の顧客と商品が比較的少数のカテゴリへ所属する確率ベクトルが得られる。

この確率ベクトルは商品や顧客属性のように使うことができ、大量の商品群、顧客群をベイジアンネットワークで扱える変数表現に変換できる。さらに、ID付POSデータに加え、コーザルデータと呼ばれる、そのときの天気、購買時間帯、バーゲンなどの状況変数を付与した大規模データからベイジアンネットワークを構築することで、ある顧客がその状況である商品を購入する条件

付確率を表すユーザモデルが作れる(図4)。

神戸地域の流通量販店の会員のID付POSデータと、会員アンケートのデータを統合することで、心理的属性もユーザモデルに反映させることができる。潜在クラスを顧客セグメントと商品カテゴリの2つに分けた多層潜在クラスモデルを考え、顧客セグメントでは会員アンケートから得られた顧客ライフスタイル(因子)を反映させ、これとID付POSデータによる潜在意味解析の結果得られた商品カテゴリを関連づける(図5)。このモデルは、i)顧客はいくつかの顧客セグメントに分類できる、ii)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、iii)顧客セグメントは特定の商品カテゴリを購入する傾向にある、iv)顧客セグメントは顧客のライフスタイルに影響を受ける、という全体構造に基づいている。どの顧客セグメントがどの商品カテゴリを購入する傾向にあるのかが計算できるだけでなく、ベイズの定理を用いることで、ある商品を買う確率が高い顧客セグメントの確率 $P(\text{顧客} | \text{商品})$ や、ある特定の顧客が購入する確率の高い商品セグメントの確率 $P(\text{商品} | \text{顧客})$ も計算できる。これは、実際に訪れる顧客の心理や行動を反映した店内の売り場や施策の最適化や、ユーザモデルを導入したデジタルサイネージによる最適な情

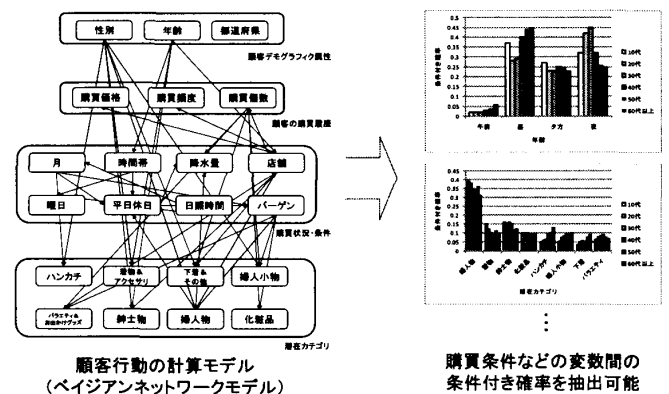


図4 顧客行動計算モデルと顧客購買行動予測

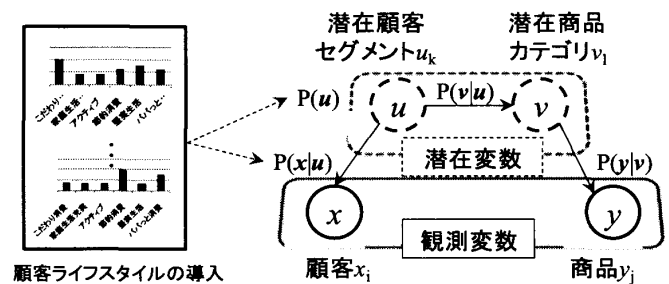


図5 ライスタイルを導入した多層潜在クラスモデル

報推薦を行うことでサービスの生産性を向上することが期待され、現在、いくつかの企業が実際に導入を進めている。

5. 需要予測システム

これまでに述べたような大規模データから構築した利用者モデルを用いて需要予測の精度改善を行うことができる。適切な在庫管理による廃棄率の低下や過剰在庫の抑制は小売サービス業の生産性にとって重要であるが、いまだに流通量販店における需要予測は、前年度の同月同曜日における売上との比較や、各店舗の店長の経験と勘により行われている現状もある。

そこで各日の状況（曜日、祝日、雨量、気温、イベント）などのコーザルデータと呼ばれる説明変数をデータに加えて、数量化Ⅰ類と先に説明した利用者モデルに基づいた顧客セグメントを使った来店人数予測の実験評価を行った[12][13]。ある一店舗において行った需要予測の結果を図6に示す。図中の表における①は前年同月同日の来店人数から今年度の来店人数を予測した結果である。②は全来店人数に対し、数量化Ⅰ類を適用し、その予測残差を計算した結果である。③は顧客セグメントごとに数量化Ⅰ類を実施し、それぞれの予測値の合計をその日の全体の来店人数の予測値とし予測残差を計算した結果である。この店舗では来店人数の曜日周期性が弱く、また顧客セグメントごとの来店の独立性が高い店舗であった。そのため、顧客セグメントごとの需要予測結果は前年同月同日、全体の来店予測と比べてその精度を改善できている。

6. おわりに

本稿では、日々収集される大規模データを用いたユーザ適応システムの事例を紹介した。こうした情報サービスをリアルタイムに日常生活環境で実行し続ける

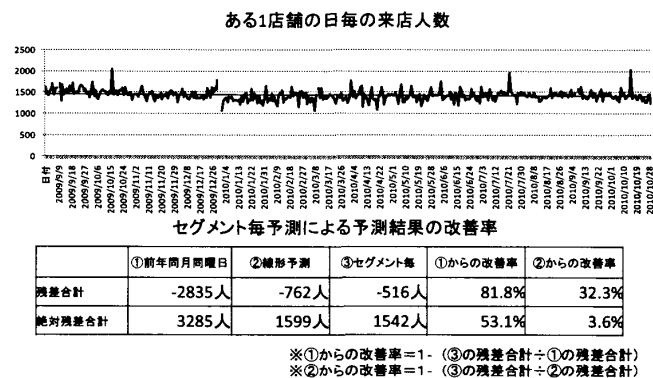


図6 顧客セグメントを導入した来店人数予測

ことで、これまでの実験室環境では得られなかった大量で意味のあるデータを分析できるようになる。日常環境でのデータ収集は技術的には容易に実行できるようになっているが、データを取得する上で、プライバシーの問題や、研究目的のためだけに協力を得ることは社会的に受容されにくくなっている。またたとえ観測容易な事象だとしても、実際に使う場面において、状況依存性の高い説明変数を十分網羅的に収集するためには、データを観測する環境が日常的な利用環境と合致していなければならない。

そこで、こうした問題に対して実サービスと調査・研究を一体化すべきであるとする「サービスとしての調査・研究 (Research as a service)」と呼ぶアプローチが考えられている[14]。ここでは調査・モデル化の段階とそのモデルを用いたサービスを切り離すことなく、情報サービスを日常環境で実行しながら、観測や利用者のフィードバック（心理的調査）の結果を網羅的に収集する。これは古くはサイバネティクス、また信頼性工学ではデミングサイクルとして知られるPDCA (Plan, Do, Check, Action) サイクルを実社会の中で回し続けることで、現実のモデルを常に修正していき、さらにそのモデル上であるべき理想的な状態の確率が高まるようなシミュレーションを行い、その状態を社会に反映するというものである(図7)。この枠組みは単に実データの収集だけにとどまらず、研究を実フィールドの中で実践することで新しい価値を生み出すというアクションリサーチにも通じるものである。アクションリサーチとは、社会心理学の祖である K. Lewin によって提唱された理論研究と現実のフィールドでの適用を相互循環的に推進して社会問題の解決にあたる介入研究のことである。こうした社会の中でのアクションリサーチと位置付けた大規模デー

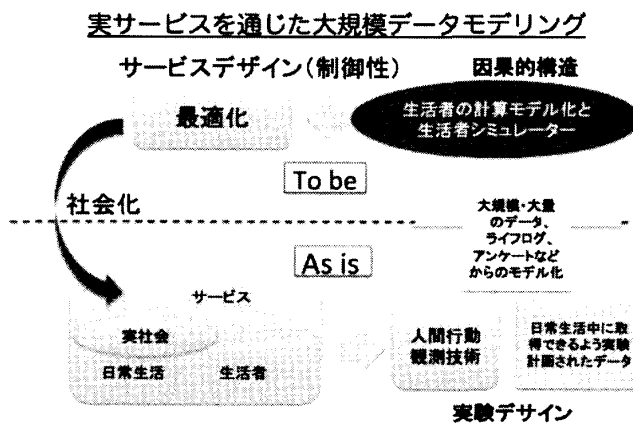


図7 実サービスにおける大規模データモデリング

タに基づく生活者のモデル化やサービス工学は今後さらに発展が期待される研究領域といえよう。

参考文献

- [1] 吉川弘之, “サービス工学序説—サービスを理論的に取り扱うための枠組み—,” シンセシオロジー, Vol. 1, No. 2, pp. 111-122, 2008.
- [2] 内藤耕, サービス工学入門, 東京大学出版会, 2009.
- [3] 本村陽一, 西田佳史, 持丸正明, 赤松幹之, 内藤耕, 橋田浩一, “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.
- [4] 本村陽一, “サービス工学におけるユーザモデリング”, 電子情報通信学会誌, Vol. 94, No. 9, pp. 783-787, 2011.
- [5] 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.
- [6] C. Ono, M. Kurokawa, Y. Motomura and H. Asoh, “A context-aware movie preference model using a Bayesian network for recommendation and promotion,” Proc. User Modeling 2007, LNCS, Vol. 4511, pp. 257-266, Springer, 2007.
- [7] 小野智弘, 黒川茂莉, 本村陽一, 麻生英樹, “ユーザ嗜好の個人差と状況差を考慮した映画推薦システムの実現と評価,” 情報処理学会論文誌, Vol. 49, No. 1, pp. 130-140, 2007.
- [8] 落合香, 下角哲也, 小野智弘, 麻生英樹, 本村陽一: ベイジアンネットワークを用いた映画コンテンツのマーケティング支援, 人工知能学会全国大会, 3D2-NFC1-1, 2009.
- [9] T. Ishigaki, T. Takenaka and Y. Motomura, “Category Mining by Heterogeneous Data Fusion Using PdLSI Model in a Retail Service,” Proc. on IEEE International Conference on Data Mining (ICDM), pp. 857-862, 2010.
- [10] 石垣司, 竹中毅, 本村陽一, “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.
- [11] T. Hofmann and J. Puzicha, “Latent Class models for collaborative filtering,” Proc. 16th Int. Joint Conference on Artificial Intelligence, pp. 688-693, 1999.
- [12] 竹中毅, 石垣司, 本村陽一, “生活者行動に着目したサービス需要予測技術の検討,” 人工知能学会全国大会, 1B1-2, 2011.
- [13] 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, 1B3-2, 2011.
- [14] 本村陽一, “大規模データからの日常生活行動予測モデリング,” シンセシオロジー, Vol. 2, No. 1, pp. 1-11, 2009.