

科学研究費助成事業 研究成果報告書

平成 27 年 5 月 20 日現在

機関番号：11301

研究種目：若手研究(A)

研究期間：2012～2014

課題番号：24683012

研究課題名(和文)顧客セグメンテーションと個別化技術の統合モデルと大規模データ活用基盤の構築

研究課題名(英文)Large-scale personalized marketing modeling for effective use of transaction data

研究代表者

石垣 司 (ISHIGAKI, TSUKASA)

東北大学・経済学研究科(研究院)・准教授

研究者番号：20469597

交付決定額(研究期間全体)：(直接経費) 7,900,000円

研究成果の概要(和文)：購買行動に関する大規模データ活用の一つに個別化技術がある。しかし、店舗を利用する全顧客と店舗で扱う全商品規模でのマーケティングモデルの構築は、購買データの本質的なスパース性による情報不足と大規模モデルのパラメータ推定の計算量の問題から困難であった。本研究では、マーケティングにおける離散選択モデルと機械学習における次元圧縮モデルを階層ベイズモデリングの枠組みで統合した。また、変分ベイズ法による高速な推定アルゴリズムを導出することで、数十万人かつ数千商品種類規模の個別化モデルを実現し、大規模データを活用するための基盤構築を行った。

研究成果の概要(英文)：This research project aims to realize a large-scale personalized marketing model. Large-scale transaction data recorded in supermarkets or convenience stores essentially are sparse with respect to consumers, items, and purchase times. We combine a dimensional reduction model with the hierarchical Bayes binary probit model for overcoming the sparseness of data. For computational feasibility, we employ variational Bayes inference that has computational efficiency compared to the resource-intensive Markov chain Monte Carlo inference in large-scale problem. The result shows that the model is applicable to datasets involving tens of thousands of consumers and hundreds of product items.

研究分野：統計科学、サービス科学

キーワード：マーケティング ビッグデータ 階層ベイズモデル 次元圧縮 パーソナライゼーション

1 . 研究開始当初の背景

情報技術の発展により多くのサービス業で行動履歴を記録した大規模データが蓄積されるようになった。これらの大規模データは、その分析により顧客の行動を理解しマーケティングへつなげるための有力な材料として多くの企業に注目されている。大規模データの具体的な活用として注目されるのは、顧客セグメンテーションや顧客個人を対象とした個別化技術である。セグメンテーションは、顧客を同様の傾向をもつセグメントへ分類し、その俯瞰を可能とする。個別化技術は顧客個人の特性を理解することで個人の購買行動の予測や制御を目指す。個人のマーケティングモデル化は顧客毎の最適な施策を提案することができるため、顧客満足度向上が見込まれる。しかしながら、購買データのスパース性による本質的な情報不足と大規模モデルのパラメータ推定計算量の問題から、店舗を利用する全顧客と店舗で扱う全商品規模でのマーケティングモデル化は困難であった。

2 . 研究の目的

本研究の目的は、店舗を利用する全顧客規模と店舗で扱う全商品規模での大規模な個別型マーケティングモデルの実現である。

3 . 研究の方法

マーケティングモデルと次元圧縮モデルを統合することで、上記の目的を実現する。

マーケティングモデルは個人を対象としたマーケティング刺激への反応のしやすさを推定することで、各個人の特性を理解し、行動を制御するための有効な手段として用いられている。しかし、そのパラメータ推定の数理的性質による計算量の問題から、大規模モデル化には不向きである。一方、機械学習における次元圧縮法は、大規模データからデータの傾向を見出すために有効な手段として用いられている。ID 付き POS データに代表される大規模購買履歴データは本質的にスパースなデータであり、個人化のためには情報の集約や圧縮が必要となる。しかし、マーケティング刺激と行動（説明変数と目的変数）の因果関係を保持したまま情報圧縮する方法の提案は今までなされていない。

ここでは、マーケティングモデルとして、階層ベイズ2項プロビットモデルを、次元圧縮モデルとしてトピックモデルの一種である Voting Blocs Model を採用し、階層ベイズモデリングの枠組みで両者を統合することで、大規模なマーケティングモデルの実現を目指す。また、ベイズモデルにおけるパラメータ推定のための一般的な方法であるマルコフ連鎖モンテカルロ法（MCMC）では、大規模モデルにおいて膨大な計算コストが必要となる。そのため、高速な近似推定アルゴリズムである変分ベイズ法をパラメータ推定アルゴリズムとして採用することで、計算

量の問題を解決する。

4 . 研究成果

ここでは、ID 付き POS データを用いて構築した大規模モデルの検証を行った。そのデータにはマーケティング変数として、価格、エンド陳列の有無、チラシ掲載の有無が付与されている。

統合モデルではトピックモデルの性質に基づいて顧客の商品購買傾向より各顧客を潜在セグメントへ分類する。その潜在圧縮空間においてプロビットモデルの性質であるマーケティング刺激に対する反応係数の推定を行う。上記のとを階層ベイズモデリングによって同時分布として表現し、変分ベイズ法でパラメータ推定を行う。その後、その推定結果を元の大規模空間に還元する。ここでは、購買データの存在しない部分の効用を、マーケティング変数を利用し情報補完する。そして、その効用から各顧客、各商品に対する反応係数を推定する。

図1に推定した各商品に対する反応係数のヒストグラムを示す。反応係数は顧客ごと商品ごとに推定できるため、図中の各値は顧客に関しては平均値として算出している。図中の Price は価格に対する反応係数を示している。具体的には、反応係数が負の値をとる商品は価格が下がると購買されやすいことを示している。図中ではほとんどの商品が負の係数を示しており、推定結果が合理的であることの証左となるであろう。一方、Display はエンド陳列の有無、Feature はチラシ掲載の有無に対する反応係数を示している。多くの商品は正の値を示しているが、Price と比べると、ゼロ付近に値をとる商品が多い。これは、エンド陳列やチラシ掲載が有効である商品と有効とは言えない商品が存在することをしめしており、実務上の経験則と一致する結果となっている。また、どの商品が有効・非有効の判別を行うことができるため、実務において有益な示唆を与えることが可能である。

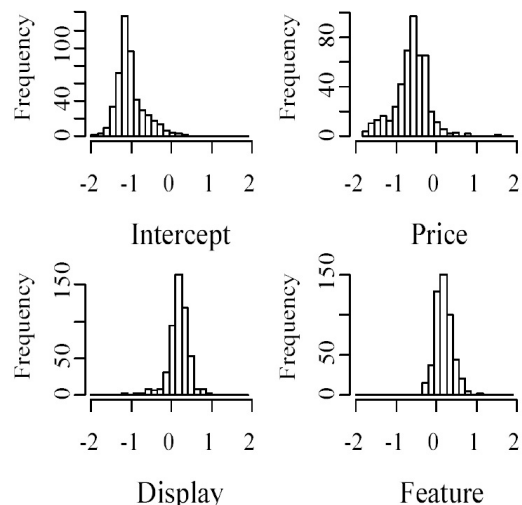


図1 500 商品の反応係数のヒストグラム

計算量の検証を下記のように行った．3.5 GHz processor (Quad-Core Xeon; Intel Corp.)と 64 GB memory の計算機を利用し，MCMC と変分ベイズ法の計算時間の検証を行った．ここでは，顧客 1 万人，商品千種類のシミュレーションデータを作成し計算時間を調べた．また，トピック数は 20 とした．その結果，MCMC では約 580 時間の計算時間が必要と推定されたが，変分ベイズ法では約 17 時間であった．また，複数の設定で検証を行った結果，顧客数，商品種類数，トピック数の増加に対して計算時間もほぼ線形に増加した．

図 2 に大規模マーケティングモデルによる大規模な顧客・商品種に対するパーソナライゼーションの結果を示す．図 2 中の各図の黒い点は，各マーケティング刺激に対する反応係数が 5%水準で有意であることを示している．図より推定結果には，顧客毎・商品毎に有効なマーケティング変数が異なることが分かった．この結果を用いることで，個人毎・商品毎の最適なマーケティング刺激を推定できる可能性があり，全店舗規模でのマーケティング活動の最適化を図ることができるといえる．

本研究で構築した大規模マーケティングモデルは，ID 付き POS データのみではなく様々な説明変数を伴う分散行動データに対して有効に作用することが期待できる．

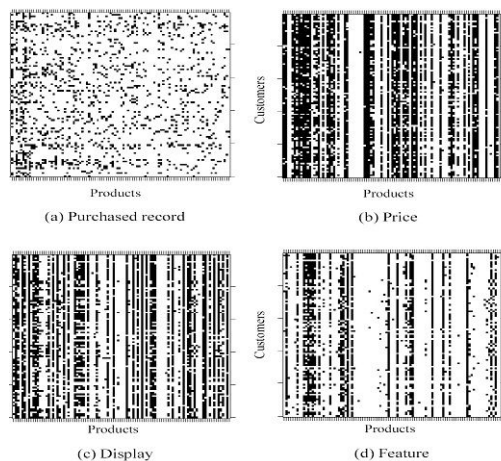


図 2 個人毎・商品毎の有効なマーケティング刺激

5. 主な発表論文等

(研究代表者，研究分担者及び連携研究者には下線)

[雑誌論文](計 2 件)

T. Ishigaki, N. Terui, T. Sato and G.M. Allenby, Topic Modeling of Market Responses for Large-Scale Transaction Data, Data Science and Service Research Discussion Paper, 査読無,

No.35, pp. 1-41 2015

小柴等，石垣司，竹中毅，櫻井瑛一，本村陽一，行動履歴データとライフスタイル調査にもとづく顧客モデル構築技術，電気学会論文誌 C，査読有，Vol. 133, No. 9, pp. 1787-1795, 2013

[学会発表](計 15 件)

石垣司，統計的モデリングとデータ活用による社会価値創出，文部科学省委託事業「数学共同プログラム」ワークショップ 社会システムデザインのための数理と社会実装へのアプローチ，2 月 11-12 日，九州大学(福岡県福岡市)，2015

T. Ishigaki and N. Terui, A Dynamic Marketing Model Based on Topic Modeling for Large-scale Customer Data, International Conference on Statistical Analysis of Large Scale High Dimensional Socio-Economics Data, November 6-7, 東北大学(宮城県仙台市), Japan, 2014

石垣司，消費者行動データ分析のための構造モデリング：入門からビッグデータ活用まで，日本行動計量学会第 42 回大会チュートリアルセミナー，9 月 2 日，東北大学(宮城県仙台市)，2014

T. Ishigaki, N. Terui, T. Sato and G.M. Allenby, A Large-Scale Marketing Model using Dimension Reduction and Variational Bayes Inference, International Workshop on Data Science and Service Research, July 18, 東北大学(宮城県仙台市), Japan, 2014

石垣司，照井伸彦，佐藤忠彦，変分ベイズ法を用いた大規模マーケティングモデル，日本マーケティング・サイエンス学会第 95 回研究大会，6 月 21~22 日，関西学院大学(兵庫県西宮市)，2014

石垣司，成長を実現するためのビッグデータからの確率的潜在意味解析と構造的モデリング，第 98 回行動計量シンポジウム，3 月 21 日，産業技術総合研究所(東京都江東区)，2014

石垣司，照井伸彦，佐藤忠彦，変分ベイズ法を用いた購買履歴データ分析のための大規模マーケティングモデル，第 8 回日本統計学会春季集会，3 月 8 日，同志社大学(京都府京都市)，2014

石垣司，照井伸彦，佐藤忠彦，ビッグデータ対応型消費者行動分析モデルとその応用，第 5 回横幹連合コンファレンス，12 月 21 日，香川大学(香川県高松市)，2013

石垣司，照井伸彦，佐藤忠彦，ビッグデータ対応型消費者行動分析モデル，2013 年度 統計関連学会連合大会，9 月 8~11 日，大阪大学(大阪府豊中市)，2013

石垣司，サービス科学における大規模

データと統計的モデリング，日本機械学会設計情報学研究会，12月7日，強羅青雲荘（神奈川県足柄郡），2012

小柴等，石垣司，竹中毅，本村陽一，行動履歴データからのライフスタイル推定技術：顧客 ID 付き POS データとアンケート調査による小売りサービスでの実証，第 15 回人工知能学会，社会における AI 研究会，11月16日，東京，2012
石垣司，照井伸彦，佐藤忠彦，潜在クラス型階層ベイズプロビットモデルによる大規模購買行動モデル，第 15 回情報論的学習理論ワークショップ（IBIS2012），11月7～9日，慶應義塾大学（神奈川県横浜市），2012

石垣司，照井伸彦，佐藤忠彦，大規模購買データ活用のためのマーケティングモデル，2012 年度 統計関連学会連合大会，9月4～7日，北海道大学（北海道札幌市），2012

T. Ishigaki, N. Terui, T. Sato and G.M. Allenby, Personalized Marketing Model by Using Massive Customer Data, International Workshop on Marketing Science and Service Research, July 2-3, 統計数理研究所(東京都立川市), Japan, 2012

T. Ishigaki, N. Terui and T. Sato, Toward Large-scale Customer Analysis in the Framework of Random Utility Model, ISBA 2012 World Meeting, June 25-29, 京都テルサ（京都府京都市），Japan, 2012 by poster presentation

〔図書〕(計 1 件)

本村陽一，竹中毅，石垣司（編著），サービス工学の技術 ビッグデータの活用と実践，東京電機大学出版局，2012年11月，204 ページ

〔産業財産権〕

出願状況（計 0 件）

取得状況（計 0 件）

〔その他〕

ホームページ等

6. 研究組織

(1) 研究代表者

石垣 司 (ISHIGAKI Tsukasa)

東北大学・大学院経済学研究科・准教授

研究者番号：20469597