

情報爆発時代の広告研究

矢田 勝俊[†]関西大学商学部[†]鷲尾 隆[‡]大阪大学産業科学研究所[‡]

1.はじめに

広告は企業にとって自社の商品やサービスを認知させ、売上を維持する重要なマーケティングツールの一つである。従来、広告計画では、必ずしも顧客に関連付けられたものではなく、集計された指標やデータを頻繁に利用してきた。情報爆発時代と呼ばれる昨今、消費者に関するあらゆるデータが蓄積されるようになっている。このような詳細な顧客データを利用し、効率的な広告計画を策定することは、多くの企業にとって重要な意義を持つ。

本研究の目的は、数値属性やカテゴリ属性を含んだ正負事例を高い精度で分類する CAQEP を用いて、広告計画の中核を成す消費者の購買行動モデルを構築し、実データによる有効性の検証を行うことである。従来の研究でモデル構築に頻繁に用いられるロジットモデルは、正負事例に偏りのあるデータやデータ属性間に相乗・抑制効果がある場合、予測精度が低く抑えられる傾向を持つ。我々は、CAQEP を用いることによって高精度の購買行動モデルを構築することができ、可読性の高い数値領域のルールを獲得することができる。本稿では多様なメディア接触情報を含んだ実データに提案手法を適用し、有用な知識発見を目指す。

2.研究の目的

2.1 データの説明

本研究では、株野村総合研究所、株宣伝会議が開催した「マーケティング分析コンテスト 2007」で提供されたシングルソースデータを利用した。シングルソースデータとは、顧客 ID に関連付けられているメディアへの接触頻度、購買行動、個人属性などの集合を意味する。個人単位で広告などの刺激への接触頻度とその反応（購買行為）を測定できることから、重要な知見が得られると期待される。

データは 2007 年 5 月からの約 1 ヶ月間に、関東エリアの約 2000 人からウェブや携帯でのアンケートによって収集された。データには対象期間の約 7800 番組の視聴の有無、ポータルや情報・コミュニティサイトなどの約 200 サイトの閲覧、120 誌を超える新聞・雑誌の購読、60 種

類のキャンペーン認知などに関して記録されている。購買行動に関しては、食品、耐久消費財、金融商品など約 300 商品の購入意図や購買経験などが含まれている。

2.2 研究の目的

広告研究において、広告効果測定に関する消費者行動モデルは最も基本的な問題の 1 つである。購買行動モデルを基礎にしたメディアプランニングは広告業界で理論と実践を直接結びつける領域であり、多くの研究蓄積が存在する。しかし、それらの多くはシングルソースデータではなかったり、一部のメディア情報に限定されたデータを利用していた。本研究の目的は多様な広告メディアの接触情報を含んだシングルソースデータを用いて、予測精度の高い消費者の購買行動モデルを構築することである。

本研究では、メディアごとの接触情報を媒体や類型ごとに集計した「接触頻度」をモデル構築に利用している。例えば TV 視聴ならば、週末の夜間に視聴した番組数、ウェブサイトならば、訪問したポータルサイト数、新聞ならば一般紙の購読数などである。このようにメディアの接触情報は接触頻度として、数値属性に加工された。ただしキャンペーンの認知はカテゴリ属性としてそのまま利用した。メディアごとの接触頻度を採用した理由は、約 10000 の刺激属性を扱う難しさと、得られるルールの実用性を考慮したためである。

これらのデータを利用して消費者行動モデルを構築する際に、我々は二つの問題を解決しなければならない。第一に予測精度を落とさず、適切な数値区間へ離散化すること。第二に予測精度の高い購買モデルを構築するために、事例数の少ないクラスについても高い精度を得られることである。我々はこれらの問題を解決するため CAQEP を適用し、消費者行動モデルの構築を試みた。

3. CAQEP

データマイニングにおける分類手法の中で、相関規則集合に基づいて分類を行う手法は、

- (1) 多属性を含むデータから、分類に有用な属性組合せを自動的に完全導出する。
- (2) 少数例のクラスに対する分類規則を網羅でき分類精度を落とさない。
- (3) 過学習を起こし難くできる。

という利点があり、CAEP[1]をはじめ多くの手法が開発されてきた。特に、あるクラス事例についてのみ多頻度共

[†]Advertising Research in Information Explosion Age

[†]Katsutoshi Yada, Faculty of Commerce, Kansai University

[‡]Takashi Washio, ISIR, Osaka University

起する属性集合 Emerging Pattern を分類に用いる CAEP は、高精度であることが知られている。

しかし、これらを数値属性を含むデータに適用するには、数値属性間依存性を考慮した離散化前処理が必要であり、離散化区間が不適切だと属性やその値の共起性をうまく捉えられない。そのため、SUBCIU[2]の考え方を各属性軸射影密度に適用して、高速な密度ベース属性部分空間クラスタリングを行う QFMiner[4]を用いて上記問題を解決し、

(4) 数値属性を含むデータについても高い分類精度を達成できる。

という利点を持つ CAQEP[5]を開発した。(1)～(4)は本研究の対象データに適合するものであり、CAQEP によって適切な消費者行動モデルが導出されると期待できる。

4. 実験結果

本研究では提案手法 CAQEP を上記のデータに適用し、その予測性能、抽出されたルールについて考察する。目的変数として、シャンプー、ビール、野菜ジュース、液晶テレビ、コーヒーの 20 ブランドの購入の有無を使用した。

4.1 性能評価

マーケティング研究において、数値属性を持つ正負事例の分類モデルにはロジットモデルが利用されることが多い。ここでは、提案手法とロジットモデルの予測精度を比較した。説明属性はメディアの大分類を用い、新聞、雑誌、ウェブサイト、TV 視聴、キャンペーン認知の 5 媒体の接触頻度を採用した。

モデルの性能評価には Matthews Correlation Coefficient(C) [3]を用いた。分類後、正しく分類された正負事例をそれぞれ P_T, N_T 、誤って分類された事例を P_F, N_F としたとき、 C は以下のように定義される。

$$C = \frac{P_T \times N_T - P_F \times N_F}{\sqrt{(P_T + N_T)(N_T + P_F)(P_T + P_F)(N_T + N_F)}}$$

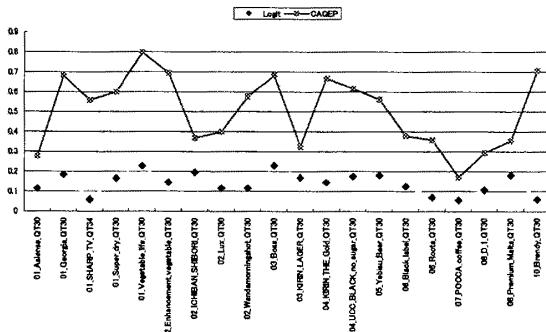


図 1 CAQEP とロジットモデルの性能比較

図 1 は 10 回の交差検証を実行し得られた 20 商品の購

買モデルの C を示している。すべての商品の購買モデルにおいて、本提案手法はロジットモデルよりも高い予測精度を達成している。実験結果を詳細に検討すると、ロジットモデルは正事例数（購入者）の少ない商品について、購入者をほとんど識別することができなかった。

4.2 ビジネスへのインプリケーション

次により詳細なメディア中分類の説明属性を用いて購買モデルを構築し、抽出されたルールに考察を加えた。我々は実験結果から従来の統計手法では見つけることが難しい、局所的なルールを網羅的に抽出することができた。例えばシャンプー購入者は一般的にウェブサイトの利用頻度が低いことが知られているが、花王㈱のアジェンスに関しては、月に 7-1 回とコミュニティサイトの利用頻度が高い顧客グループが存在していることが分かった。このような局所的な顧客グループは専門家に新規性の高い、興味深い示唆を提供することができた。

5. むすび

本研究では数値アイテムを含むデータに対する分類手法 CAQEP を広告計画における購買モデル構築に適用し、実データによる実験を行った。提案手法は従来手法よりも高い精度の購買モデルを構築することができ、重要なビジネスへの示唆も得ることができた。しかしながら本研究に残された課題は多い。本稿では購買モデルの構築に焦点を当てたため、それを用いた広告予算配分問題にまで拡張することはできなかった。また広告の長期にわたる影響などは考慮されておらず、理論的な検討も含め今後の課題である。

参考文献

- 1) G. Dong, X. Zhang, L. Wong, J. Li: CAEP: Classification by Aggregating Emerging Patterns, *Int'l Conf. on Discovery Science*, pp. 30-42 (1999).
- 2) K. Kailing, H.-P. Kriegel, and P. Kroger.: Density-connected subspace clustering for high-dimensional data, *Fourth SIAM Int'l Conf. on Data Mining (SDM'04)*, pp. 246-257 (2004).
- 3) N. Matsatsinis, and Y. Siskos: *Intelligent Support Systems for Marketing Decisions*, Springer (2002).
- 4) T. Washio, Y. Mitsunaga and H. Motoda: Mining Quantitative Frequent Itemsets Using Adaptive Density-based Subspace Clustering, *ICDM'05: The Fifth IEEE Int'l Conf. on Data Mining* pp. 793-796 (2005).
- 5) T. Washio, K. Nakanishi and Hi. Motoda: Deriving Class Association Rules Based on Levelwise Subspace Clustering, *PKDD2005: 9th European Conf. on Principles and Practice of Knowledge Discovery in Databases*, LNBI 3721, Springer, pp. 692-700 (2005).