

広告接触による消費者行動変化 デモグラフィック分割による分析

高橋 柊† 岡田 洋明† 丹羽 悠人† 川崎 隆史† 杉山 歩† DAM Hieu Chi†

†北陸先端科学技術大学院大学 知識科学研究科

1 はじめに

多様化・複雑化する情報社会において、消費者に対する効果的な広告手法の発見は、企業にとってますます重要度が増している。2012年度、日本の総広告費は5兆8,913億円となり、前年比103.2%となっている¹。また、インターネット広告費とテレビ広告費が共に前年度に比べ増加していることから、広告が多様化していることが分かる。このような状況の中、企業はターゲットとする消費者に対し効果的な広告出稿先を選択する必要があり、消費者行動変化と広告接触状況の関係性を見つけ出すことは有益である。

これまでのマーケティング研究は、商品群や消費者属性を仮定し、消費者の商品購買プロセスを分析することに主眼を置いている。知識科学的観点から商品購買プロセスを考えた際、商品や消費者属性を仮定することなく、真に重要な属性の特定および知識発見は重要な研究課題である。

そこで本研究では、マーケティングデータでは重要視されないスパースな購買データを対象とし、消費者行動変化を広告接触状況により説明可能な消費者群を発見することで、広告により商品を認知および購入する消費者群を同定する。また、同定した消費者群がどのような広告に反応しているか分析することで、マーケティング分野における知識発見を行うと共に、効果的な広告戦略を行うことが可能となることを目指す。

2 対象データ

株式会社野村総合研究所が提供する時系列シングルソースの顧客データ²(以下、顧客データ)を対象データとする。顧客データは3000人の消費者に関するデータであり、消費者の103商品に関する2区間での「購入実態」、「購入意向」と、その商品に関するテレビCMおよび雑誌広告との接触状況を持つ。

顧客データにおける「購入実態」、「購入意向」を消費者の「商品を購入する」という心理的プロセスに適合するため変換する。商品の購入プロセス分析モデルとして、AIDAモデル[1]が古くから知られている。本研究ではAIDAモデルを踏まえ、顧客データを「認知」、「意欲」、「購入」に変換する。顧客データの変数は該当商品により異なる。データ変換例を図1に示す。

2値化したデータの中で利用可能なデータは表1の4パターンである。その他のデータは、認知していないにも関わらず購入を行っていたり、認知と意欲がないにも関わらず購入しているなど矛盾しているので除外する。

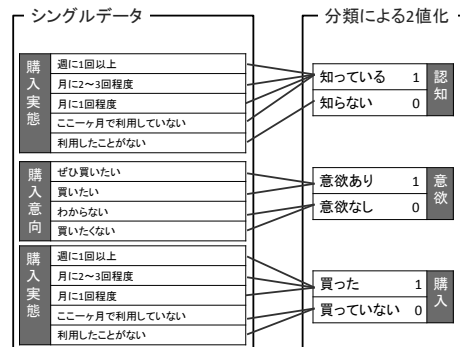


図1: 顧客データ変換例

表1: 利用可能顧客データ

認知	意欲	購入	状態
0	0	0	認知なし
1	0	0	認知あり
1	1	0	意欲あり
1	1	1	購入あり

3 分析手法

顧客データから抽出した103商品について、特定のデモグラフィック属性をもつ消費者行動変化と広告接触状況の関係を調べたい。消費者全体で識別器を構築し、次に特定のデモグラフィック属性によって抽出された消費者群で識別器を構築する。目的変数は、計測区間において状態が「購入あり」に近づくよう遷移していれば1、変わらなければ0の2値とする。説明変数は

Demographic Split Study on the Effect of Advertising Contact for the Consumer Behavior

†TAKAHASHI Shu †OKADA Hiroaki †Niwa Yuto

†KAWASAKI Takafumi †Sugiyama Ayumu †DAM Hieu Chi

†School of Knowledge Science, Japan Advanced Institute of Science and Technology

¹http://www.dentsu.co.jp/books/ad_cost/2012/

²(出所)株式会社野村総合研究所 インサイトシグナル(2010年3-4月期)データ

商品テレビCMおよび雑誌広告とし、それぞれ接触していれば1、接触していなければ0の2値とする。識別器の構築には、多量の説明変数をもつデータに対し有効なアンサンブル学習アルゴリズムである Random Forests[2] を用いる。また、目的変数のパラ付きを抑えるため、ランダムサンプリング時にサンプル数を最小クラスに合わせ補正した。

4 分析結果

4.1 生成識別器の誤り率

生成された識別器の誤り率分布を図2に示す。左は消費者全体で識別器を構築した場合であり、右は特定のデモグラフィック属性によって分割した消費者群で識別器を構築した場合の結果である。縦軸の Error rate of 1 は1(状態遷移あり)の誤り率、横軸の Error rate of 0 は0(状態遷移なし)の誤り率を示す。消費者全体で識別器を構築した際、誤り率が共に0.5以下となる有効な識別器(以下、採択識別器)は2個であるが、デモグラフィック属性によって分割された消費者群で識別器を構築した場合、採択識別器は316個と大幅に増加した。

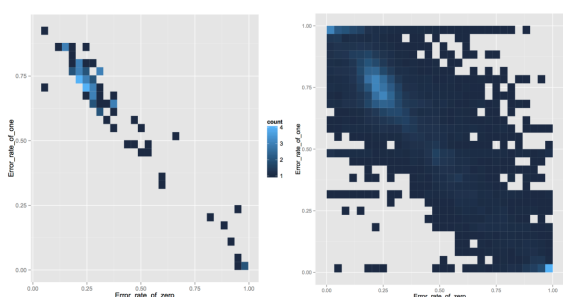


図2: 識別器の誤り率分布

4.2 最良結果識別器

デモグラフィック属性によって分割された消費者群で構築した識別器のうち、一例として最も誤り率が低い識別器の概要を表2に示す。OOB(out of bag) は cross validation における誤り率を示している。

表2から、デモグラフィック属性「職業」に「派遣」という変数をもつ消費者は、ケンタッキーフライドチキンの広告接触状況により、ケンタッキーフライドチキンの消費者行動変化を説明可能である事がわかる。

4.3 採択識別器構築に寄与している属性

識別器における説明変数にはテレビCMおよび雑誌広告の出稿データを用いている。デモグラフィック分

表2: 高精度識別器

デモグラフィック	”職業”=”派遣”
商品	ケンタッキーフライドチキン
サンプル数	83
OOB	0.0370

割後生成された採択識別器の説明変数として採択されたものは、テレビCMのみとなり、雑誌広告との接触状況は消費者行動変化に影響しないことが分かった。

表3にデモグラフィック分割後の採択識別器生成に寄与しているテレビCMの放送時間帯を示す。平均値に対し、分散値が非常に大きいことから、消費者行動変化に影響するテレビCMの放送時間帯は、全体としての特徴を持たず、デモグラフィック属性ごとに異なることが分かる。

表3: テレビCMの放送時間帯

	平均	分散
4:00-7:59	0.0849	0.0515
8:00-11:59	0.2421	0.1210
12:00-15:59	0.1060	0.0613
16:00-19:59	0.1850	0.0623
20:00-23:59	0.2410	0.0853
24:00-27:59	0.1410	0.0613

5 まとめ

本研究では、Random Forest を用いることで、消費者行動変化を広告接触状況により説明する識別器を構築した。構築した識別器より、消費者をデモグラフィック属性により分割することで、広告接触と購買行動の相関が属性群により異なることを示した。また本解析により、消費者の商品認知・購入実態に影響を及ぼす広告の出稿時間帯には明確な差異があることが分かった。

このことから、スパースなマーケティングデータに対し適切な分割を行うことで、説明変数と目的変数の関係が顕在化することが示唆される。

参考文献

- [1] Edward K. Strong. *The psychology of selling and advertising*. McGraw-Hill book company, inc., New York [etc.], 1925.
- [2] Leo Breiman. Random forests. *Machine learning*, Vol. 45, No. 1, pp. 5–32, 2001.