

ビッグデータ・アドテク・AI

消費者理解とマーケティング施策での課題



星野 崇宏 慶應義塾大学経済学部教授／理化学研究所AIPセンターチームリーダー

ほしの たかひろ●2004年東京大学大学院博士課程修了。博士(学術、経済学)。04年統計数理研究所助手、08年名古屋大学大学院経済学研究科准教授などを経て15年より現職。第13回日本学術振興会賞受賞。行動経済学会常任理事。日本マーケティング・サイエンス学会理事。日本行動計量学会理事。応用統計学会理事。社会調査協会理事。主な著書に『調査観察データの統計科学』(岩波書店/2009年)。最近のマーケティング分野の主な論文に「ポイント付与と値引きはどちらが効果的か? : マグニチュード効果を導入したプロモーション効果の推定」(共著『流通研究』/2017年)、「Online-Offline チャネルにおける消費者の購買間隔と購買金額の同時モデリング」(共著『オペレーションズ・リサーチ』/2016年)など。

テクノロジーとマーケティングという特集に当たって、まず思い浮かぶキーワードを幾つか並べてみたい。テクノロジー側では「ビッグデータ」「アドテク」「AI」であり、またマーケティング側では「最適化」「施策」「消費者理解」である。これらについて少し議論を行ってみよう。

アドテクの“最適化”

広告分野においてテクノロジーという言葉で最初に思い浮かぶのは、大きな分野に育ってきたデジタル広告(特にインターネット広告)の世界であろう。以前のインターネット広告はマス・メディア同様のマス広告や広告掲載ページのコンテンツ内容に親和的な広告を掲載させるコンテンツターゲティング広告、検索連動型広告などが中心であったが、2010年ごろからはインプレッション(ページ訪問閲覧)単位で広告枠を取引するアドエクスチェンジが行われるようになり、これがインターネット広告の業界を大きく変えたと言っても過言ではない。

EC サイトでの商品ページの閲覧履歴情報などを利用して他サイトでの出稿がなされるアドエクスチェンジでは、そこでの入札を瞬時に実施するリアルタイム入札(RTB)が行わ

れている。またCookieなどを利用して取得できる Web 上での閲覧履歴から、広告枠が表示されるサイトを閲覧した人がどのような属性や関心を有しているかを推測したオーディエンスデータも利用されるようになった。関連してコンバージョン(購入や資料請求など Web マーケティングにおける成果)にどの広告がどの程度貢献したかを分析するアトリビューション分析なども盛んに行われている。

これらはアドテクノロジー(アドテク)と称され、続々と新しい概念や分析方法、取引形態、成果配分の仕組みなどが提案されている分野であり、従来の広告の世界よりも“最適化”を非常に重視しているという点が特徴的である。コンバージョン当たりのコストを最小化するために、より適切な人(実際には広告が表示されることを示す“インプレッション”)に対してより適切なタイミングで広告配信を行うための方法論、データ取得や共有などが志向されているという点での方向性は良い。

しかし、問題は現状のアドテクでの“最適化”が近視眼的な局所最適化となっているのではないか、そこから導出される施策は適切なのか、という点である。近年では TVCM など既存マス・メディアでの出稿情報や実店舗への来店

履歴などの情報も一部利用されるようになってきたが、基本的にはまだインターネット上で得られるデータに閉じた形での評価が行われているにすぎず、他メディアの影響を除去していない以上、デジタル広告の効果が過大推定されているといった問題がある。

また、特に海外のデジタル広告分野でこのところ問題視されているのは、(1)自動化によって自社のブランド価値を毀損させる可能性がある出稿を排除することが難しいという“Brand Safety”の問題、(2) デジタル広告分野の一部企業がボットなどを利用して閲覧回数を水増ししていると考えられる“Ad Fraud”の問題、(3) 本当に出稿された広告が見られているのか、という“Viewability”の問題、である（詳しくは全米広告主協会の報告書などを参照されたい）。

具体的には、例えば2017年1月のInteractive Advertising Bureau's Annual Leadership Meetingで、米プロクター・アンド・ギャンブル(P&G)のMarc Pritchard氏がインターネット広告について、出稿方法の透明性に疑問を呈した講演を行い、その同時期には、同社やユニリーバなど世界最大規模の広告宣伝費を拠出する企業がデジタル広告の削減を行った、というニュースが報じられたことは記憶に新しい。さらに2月にはトムソンロイターなどが、3月にはBBCやThe Guardianなどのメディア、英国政府が、YouTube広告がテロ関連の不適切な動画コンテンツに対して出稿されたことを理由にYouTube広告への出稿を停止した。

これらの問題は今後日本でも非常に重要になってくると考えられるが、ここでは特に上記(1)にも関連する、現状のアドテクの近視眼的な局所最適化の問題点と消費者行動の理解の重要性について少し議論したい。

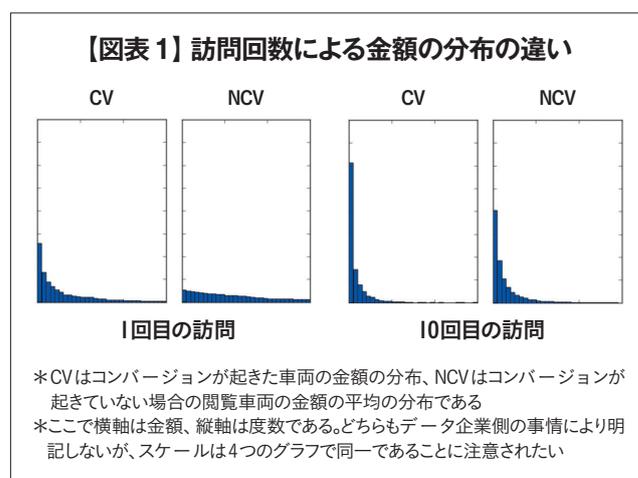
消費者行動の理解の重要性の例：中古車

ここで一つ“近視眼的”なレコメンデーションや広告配信が問題となる例を挙げてみよう。

我々の研究室ではクルマ・ポータルサイト「グーネット」を運営しているプロトコーポレーションと共同研究を行っており、月間500万ユーザーを超える膨大な閲覧履歴情報について詳細な解析を行っている（高畑ら2017、人工知能学会）。中古車は本や日用雑貨とは異なり、高額でかつ維

持費用も掛かる買い物である以上、試し買いや衝動買いをするといった商材ではない。したがってレコメンデーションや広告配信が即時的に購買につながることは多くないが、単価が高いことから購入の手前の見積もりや販売店への来店予約であってもコンバージョンとして重要な成果となり、販売店からは収入を得ることができる。

さて、このような商材である中古車について、実際にコンバージョン(ここでは見積もりである)がなされた車両の金額(【図表1】のCV)の分布を1回目と10回目で比較してみよう。さらに、1回目でコンバージョンに至らなかったユーザーの1回目のサイト訪問時に閲覧した車両の平均購買金額(【図表1】のNCV)、同様にまた10回目でもまだコンバージョンに至らないユーザーの10回目のそれを比較してみよう。



【図表1】を見てわかることは、確かに1回目の閲覧でコンバージョンに至る消費者もいて、彼らは比較的高い価格の車両の見積もりをしている。一方、ここでは記載しないが、より多くの消費者は複数回閲覧して見積もりを行っており、彼らの見積もりを行う車両は1回目の閲覧で購買に至る顧客よりも低めである。さらに、例えば10回目で見積もりを行った消費者が過去どんな車両を閲覧したかを調べると、平均して回を追うごとに価格が下がる傾向にある、ということである。

このことをよりわかりやすく説明するために、実際に存在した某顧客の閲覧履歴を挙げよう。【図表2】は3カ月ほどの間に行われた閲覧の例であり、顧客は最初BMWの幾つ

かの車両を閲覧し、その後サイトを再訪してスバルインプレッサスポーツを中心に閲覧、最後にスズキSWIFTへの見積もりを行った例である。

顧客は最初から「将来的に自分が購入するであろうスペックや価格帯」の車両のみに限定して閲覧して情報探索をする場合も確かに存在するが、多くの消費者は「実際に購入するまで時間がある場合には漠然と高級車に憧れて閲覧をし、いざ（車検やさまざまな理由で）購買意思決定に迫られると現実的に手の届く車種を見て決める」という【図表2】のような閲覧をしばしば行っている、ということである。



繰り返しになるが、もちろん1回目とほとんど変わらず同じような価格帯や車種の車両を閲覧し続けてコンバージョンに至るといふ顧客はいる。しかしそうではない顧客も多い。【図表2】のような顧客に通常のアドテックの最適化を行うとどのようなことになるであろうか？ 最初にBMWを見た顧客に同様の車種のレコメンデーションや広告配信を行っても、もちろんコンバージョンには至らない。その後、顧客がインプレッサを見始めたことがわかったとしよう。この場合、では次に何を提示すべきであろうか？ 同様にインプレッサを提示するというのが通常のアドテック最適化である。しかし、顧客がここまで大きく閲覧する価格帯を下げた以上、さらに下がる可能性は十分ある。このような場合に有用なのは、顧客が現在どのようなモードにあり（まだ情報収集を楽しんで非現実的な車種を見ているのか、決定に迫られているのか）、最後までどこで下げ止まるかという消費者の内的状態の理解を行い、競合他社に先んじてコンバージョンの可能性のある車両を提示するということである。

読者の方々も経験があるかもしれないが、Amazonで購

入した、あるいは購入を検討したが買わなかった商品の広告がFacebookや他のSNS、Web サイト、アプリ上で提示されることはしばしばある。現状のアドテックにおける“最適”な自動化をされたレコメンデーションや広告配信とは、結局、別サイトや前回訪問時の閲覧情報を利用したものでしかない。

Amazonで購入した商品と比較検討して買わなかった商品がFacebookでレコメンドされるという“間抜け”な、かつ情報が売買されているという、いわゆる“スカー広告”の気持ち悪さは、マーケティング上の効率が悪いだけではなく、広告配信先のサイトやSNSへの不信感、場合によって利用減少にもつながる。特に上記に取り上げたようなある程度高額な商材でこそ広告の成果が重要であるが、そこでの消費者の行動は本や課金アプリなどの比較的 low 額な商材での購買間隔や購入に至る、いわゆる“カスタマージャーニー”のあり方とは大きく異なるため、おのずとその広告出稿のあり方も異なる。

我々の研究室では幾つかの企業のデータ解析や施策立案の手伝いなどを行っているが、多くの場合“近視眼的”“短期的”なレコメンデーションや広告配信の自動化しか実際には行われていないのは残念である。

ビッグデータ“だけ”を利用することの問題点

近年では企業はその業態に応じて、さまざまなデータを取得するようになってきている。以前から小売業では得られているPOS データやポイントプログラムによるID-POSはもちろん、EC サイトでの閲覧と購買履歴、スマートフォンやPCの閲覧履歴、位置情報、ビーコンや監視カメラを利用した顧客動線など、業務上これらの“集まるデータ”は、企業によってはほぼすべての自社既存顧客から得られている場合もある。またPontaやTポイントなど共通ポイントプログラムの場合であれば会員数は国民の半数近くに上っているため、これらの“集まるデータ”だけから消費者の行動の理解や施策立案が可能であると考えがちである。

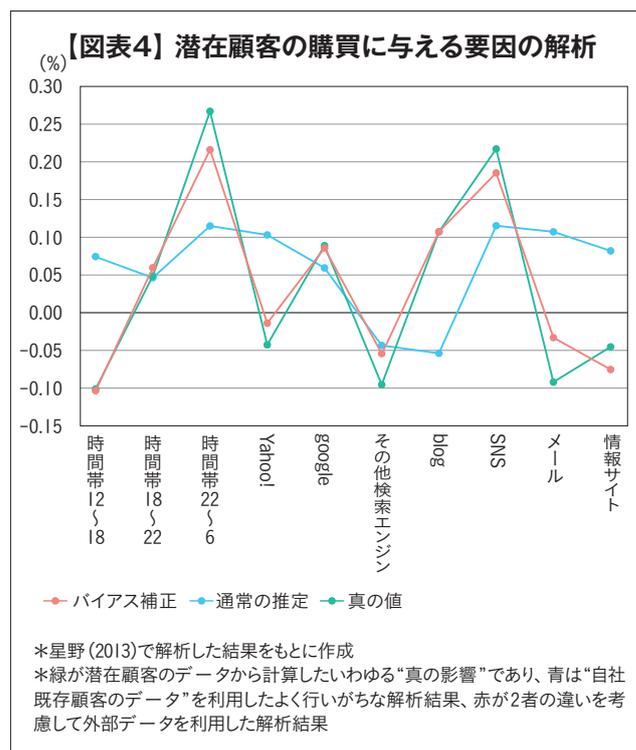
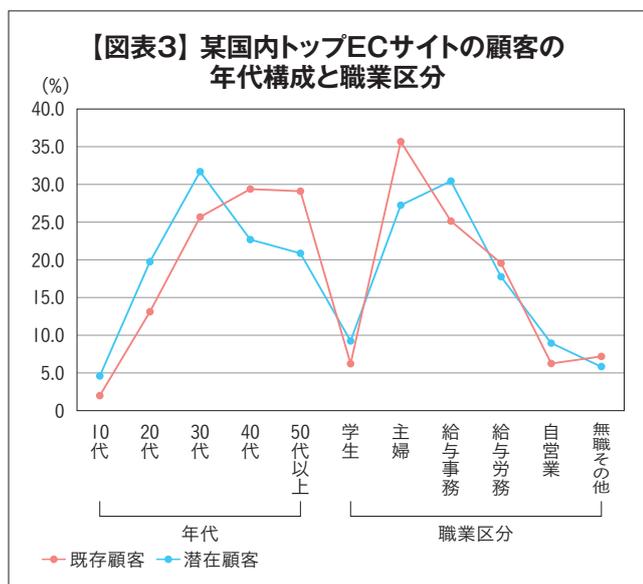
しかし、“集まる”ビッグデータだけでは測定できない決定的に重要な情報があることがこれまでも指摘されており、

特に(1)消費者の態度や認知の理解、(2)潜在顧客の理解、(3)競合他社での行動の理解、の3点は企業にとって非常に重要である。特に(1)については、たとえ同じ顧客が繰り返し同一ブランドを購入している、それは“他に良い選択肢がないから”“どうせあまり質が変わらないであろう別のブランドを試すのも面倒だから”などという消極的な理由で同一商品を購入する、いわゆる“偽のロイヤルティ”によるものかもしれない、このような状態の顧客は競合企業の革新的な商品やサービスが登場すれば容易に離反する可能性がある。したがってやはり行動だけではなく顧客の態度や状態を理解することが非常に重要である(例えば西本、2014;コラー、2013参照)。ただし誌面の都合上ここでは(2)や(3)の問題について議論したい。

筆者は以前、国内の2大ECサイトの顧客の購買行動について、ビデオリサーチ・インタラクティブによって提供されたインターネット閲覧データのパネル調査データを利用して解析を行った。このデータは1万3,000人ほどの調査協力者からPCの閲覧履歴を得ているものであり、通常の“集まるデータ”ではわからないこと、つまり競合企業での閲覧や購買がわかる。さて、このデータを解析してわかったこととして、2者の顧客はだいぶ異なるということであり、例えばA社の既存顧客はその潜在顧客(B社のみで購買する顧客)

客)に比べて年齢層が高く、主婦が多いなどといった偏りがある(【図表3】参照)。A社が自社顧客を増やすために、B社から顧客を奪うための広告販促を含めたマーケティング施策を立案するとして。しかし“集まるデータ”である既存顧客の閲覧・購買履歴から示唆されるのは、既存顧客にとっては魅力的であるが、潜在顧客にとってはそうでもないような施策のみである。実際、ECサイトにどこから流入したか(リファラー)の情報を利用して潜在顧客が購入に至る要因の解析を行うと、使用しているサーチエンジンや時間帯、SNSからかブログからかなど、さまざまな点で異なっており、いくらビッグデータがあるからといって、それを“ブラックボックスのツールで分析する”のでは潜在顧客を取り込むための施策が失敗することは明白である(【図表4】)。

上に記載したことは、すでに触れたアドテクの局所最適化の問題にも直結する。自社ECサイトやアプリ・Webサービスでコンバージョンを取得できるようになって、広告出稿と広告成果が直接結び付くようになったことが、ターゲティングを必要以上に狭めているという問題である。広告の



機能として、本来は見込み客（潜在顧客）に対して製品やサービスを知ってもらい、可能ならば利用してもらおうということが求められているが、上記の結果からもわかるように、コンバージョンのみを最大化する広告出稿は長期的には潜在顧客を取りこぼすということにつながる。同様に、前記の(3)に記載した自社顧客の競合他社での購買や利用については、当然ながら外部の調査データなども活用した上で初めてわかるものであるが、その情報をいかに計画的に取得し競合からシェアを奪う施策を実施するかが、少なくとも成熟化し縮小傾向すらある、日本を含めた先進国でのマーケティング課題となりつつある。

AIですべて解決できるのか？

上記の話に対して、「これからはAI技術が解決してくれるのでは」と考える方もおられるかもしれない。そこでAI技術の課題についても議論しておこう。筆者の専門分野の一つは統計学や統計的機械学習であり、国が省庁横断で研究費を投じた、日本のAI研究の総本山ともいえる理化学研究所AIPセンターのグループリーダーとして、ディープラーニング等の機械学習手法の理論的研究および経済経営分野での応用研究を行っている（例えば新美・星野、2017）。

このところ“AI”や“人工知能”といった用語を見ない日はないほど、AI技術に対して国民や産業界の注目と期待が集まっているようである。囲碁ではGoogleによるAlphaGoが世界のトップ囲碁棋士に圧倒的な差で勝利し、また自動運転についても完全自動運転が可能となるのは遠い先ではないと考えられている。

しかし、今のAIが成功している分野は、方法論的にはほぼすべて統計的機械学習というAI研究全体のごく一部の分野の成果であり、また応用としては画像認識や自然言語処理の研究成果が利用できる対象に限定される。具体的にはネット上にある膨大な画像データやテキストデータを事前に学習することで、画像に何が映っているか（人や動物、物体）を十分な精度で判定する、自動運転であれば道路上の物体が何で（例えば自転車）、その位置はどこであり、今後数秒以内にどのような位置にいる可能性があるか（自転

車なので2秒で動ける範囲はせいぜい10m程度であるなど）、カーナビシステムに登録されていない通行止めの標識などを把握する、といったものは画像認識技術である。また裁判の際、膨大な判例や関連法令の中から該当の案件に関連するものを探し出す、膨大なSNSデータを分類するなど、大量の画像や文書を人が見なくてもある程度の精度でまとめたり関連付けたりするという前処理は自然言語処理の成果である。

画像認識や自然言語処理がなぜ非常に有効かといえば、「ある画像と、それが道路標識であり何を意味しているか」といった組み合わせの情報（教師データと呼ばれるもの）を一度に大量に取得しそれを機械学習にかけることで、ある程度長期にわたり有用な最適化ルールを構成できるという点にある。例えば猫、自転車や道路標識の形が短期間に変化していく、交通ルールが日々変わるということは想像できない。SNSの書き込みであっても、短縮語や流行語は一般の言葉遣いのほかに例外処理を行えばよい。またAlphaGoの成功は、囲碁という評価関数を含めた“ルールが不変”な世界において最適化が行われているにすぎない。

一方、テロ関連動画など、不適切あるいは出稿によってブランド毀損を起こしえる動画はどうだろうか？ Googleという世界最高峰のAI技術を持つ企業をもってしても、冒頭に述べたYouTube動画に対して網羅的にそのコンテンツを「広告出稿予定の企業・ブランドにとって適切かどうか」を自動的に判定することは容易ではない。これはなぜか？

YouTuberの動画などを見ればわかるように、今後我々が想像しなかったような新しいタイプの動画が次々と出てくることは想像に難くない。すでに述べたように現状のAI技術は（画像や文字情報のように）教師データが大量に存在する、あるいは（囲碁のように）最適化の評価関数が明確な場合の統計的機械学習によって成立している。したがってAIが不得意とするのは“状況が変わる”“前提が変わる”“ルールが変わる”不確実な世界における予測や説明である。AIで自動化が可能だといっても、それは過去のデータからルールを導き出し、そこで得られる評価関数に対して最適化を行っているにすぎない。

動画の自動評定の問題には確かに画像認識が応用がで

きるとはいえ、何が不適切でブランドを毀損するかは事前にルールを決めておくことは難しく、教師データがない、あるいは非常に少ない、といった機械学習がうまく利用できない問題となる。

また、先ほど取り上げた“潜在顧客”と“既存顧客”の違いや、企業にとって自社での購買や利用しかわからないといった議論からも、“集まる”ビッグデータだけをブラックボックスのAI・機械学習モデルに投入すればよいというのは明らかな誤解であることがわかるだろう。実際に、新美・星野(2017)ではディープラーニングの一種のDNN(Deep Neural Network)を利用したデータ融合の研究を行っている。データ融合とは別々のデータセットをあたかも同一データセットであるかのようにみなして解析を行う方法論であるが、マーケティング分野では「TV視聴率調査の対象者」に「ブランド認知の調査をする」ことはTVやCM視聴行動に影響を与えることが問題視されているものの、ニールセンなどヨーロッパの市場調査会社によって(顧客のマッチングという古典的な方法論として)古くから利用されてきた。我々の研究では「自社でのサイト閲覧や購買履歴データ」と「他社を含めた購買等の調査データ」の融合を行うことで「自社のどのような顧客が他社でどの程度購買しているか」を理解するための方法論の開発と実データへの応用を行っている(【図表5】)。

ここで、ディープラーニングなど近年利用されている機

械学習の方法論がこれまでの方法と異なるとされるのは、データを分析するに当たってどのような変数や情報の組み合わせ(専門的な用語では“特徴量”と呼ばれる)が予測に有用かということが事前にわからずとも、その自動抽出(事前学習)がある程度可能である、とされる点である。実際に、AI分野で成功している画像認識や自然言語処理であれば、構造が不変かつ網羅的にデータを取得することが可能である場合が多いため、特徴量の自動抽出とその後の予測の精度は極めて高い。

一方、マーケティングデータや経済時系列の予測、ファイナンスデータなどでは、AIシステムを利用した種々の予測の精度がそれほど高くないのは(1)時間によって市場構造が変化する、(2)他社での購買・利用など必要なデータが欠測している、(3)(特にファイナンスなどの場合)同様のブラックボックスモデルが世界中で利用されており、そこでの予測をもとに投資などが行われると需給構造が変化するという影響がある、などの画像認識や自然言語処理とは全く異なる状況が存在するためである。

実際に新美・星野(2017)の研究においても、自社顧客のデータだけから自動的に抽出される特徴量による予測の精度は高くなく、事前に消費者行動のメカニズムを踏まえて設定された顧客行動の多様性指標というものを組み込んだ場合のほうが予測精度は大きく上昇している。

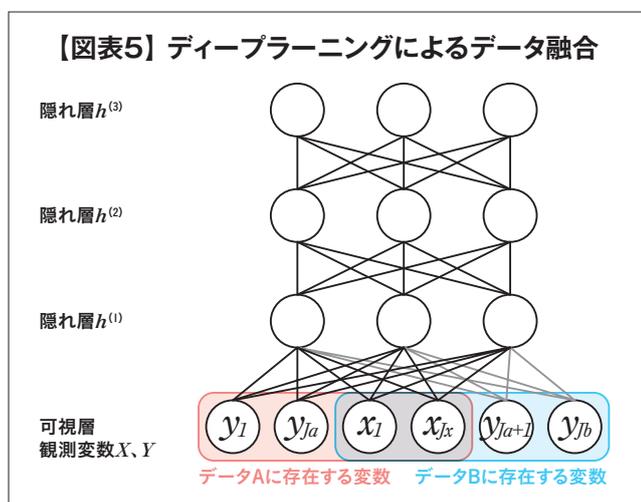
まとめ： ストーリーのある顧客理解の必要性

現在のAI技術が特定の課題にせよ成功しているのは、以下の4つの組み合わせによる。

(1)ディープラーニングなど機械学習の新しい解析法が登場した。(2)計算機の高速度と記憶容量の低コスト化が進んだことで実課題で集められるような大量のデータを処理できるようになった。(3)ネット時代になり大量のデータが収集されるようになった。(4)ECサイトや実店舗などにおいてフィールド実験が行われるようになった。

一般にはこれらがセットとして存在するとき、画像や文書、購買データの処理が容易になるということであって、そこから効果的な施策を生み出せるかといえば、それほど簡単な

【図表5】ディープラーニングによるデータ融合



話ではない。また前記の(1)や(2)は情報システムを整備すればよい話であるとして(3)や(4)はどうだろうか？ 顧客はどのようにして新製品を知るのか？ そのときのイメージは何か？ 棚の前でいかに顧客が迷っているか？ いくら安くなるなら新製品を試してもいいと思うか？ 顧客のマインドを理解し優れたマーケティング施策を生み出すための“リアルなデータ”や“実験”、そういったデータも実験フィールドもないのに、高価なシステムを導入する必要はあるのか？ まずはいかに顧客のリアルなマインドと行動のデータ、さらには実験をする環境を整備するかが重要である。

もちろん機械学習的な手法、AI技術には在庫管理やマスタ管理、データ入力自動化といった経営実務など、マーケティングにおいて有用な用途は存在するが、筆者の知識および友人の研究者などによる情報を総合しても、少なくとも2017年現在のAI技術はマーケティング施策の立案と実行どころか、その前段階であるマーケティングデータの分析を自動化するレベルにすら達していない。このことは、前に取り上げたようなYouTube動画広告によるブランド毀損が検出できないといった問題、出稿自動化による近視眼的な最適化、さらには潜在顧客を見逃してしまう問題などの事例からも明確であろう。ビッグデータ自体の歪みや機械学習自体の限界があることを踏まえて賢くシステムを利用していく必要がある。

テクノロジーの進歩とともに、マーケティングも変化すべきだろうか？ すでに見てきたように、答えはYesでもありNoでもある。ECサイトの興隆、PCからスマートフォンへ、そしてセンサーチップが安価になったことに伴うIoT時代の到来など、急速に変化する環境の中で、データ取得や顧客へのアプローチのあり方は今後とも変化し続けるであろう。一方、マーケティングの対象たるヒトはそれほど容易には進化しない。ヒトの認知や感情など、消費者行動の基礎となる部分については狩猟採集生活をしてきた時代に適したシステムがある程度遺伝子に組み込まれているからである(Buss, 2014)。デジタルネイティブという言葉があるが、農耕の発生から現在まで1万年前ほどしかたつておらず、また携帯電話の普及ですらこの20年ほどのものであり、ヒトが

いかに意思決定を行うかについてのメカニズムは、この短期間では容易に変更されない。したがっていくらテクノロジーが進化しようと、心理学や脳科学の経済やマーケティングへの応用である行動経済学や消費者行動研究(大垣・田中, 2014; 清水, 2006)の知見をもとに顧客を理解し施策を考えることは今後とも有用である。

ブラックボックスモデルの分析から誤った施策を打って立場を悪くする前に、一消費者として納得でき、かつ経営陣に対しても説明できるストーリーのある顧客理解と施策立案をぜひ心がけていきたいものである。

【参考文献】

- The Association of National Advertisers (2017) “The Bot Baseline: Fraud in Digital Advertising 2017 Report”
- Buss, D. (2014) Evolutionary Psychology: The New Science of the Mind 5th ed. Psychology Press
- 大垣昌夫・田中沙織(2014)『行動経済学』 有斐閣
- 清水聰(2006)『戦略的消費者行動論』 千倉書房
- 高畑圭佑・星野崇宏・柳博俊・渋谷友磯子(2017)「行動経済学的な知見を用いた消費者の情報探索行動の予測とレコメンデーション法の開発」人工知能学会第31回全国大会抄録
- 新美潤一郎・星野崇宏(2017)「顧客行動の多様性変数を利用した購買行動の予測」人工知能学会論文誌、32(2)
- 西本章宏(2014)「ビッグデータ時代の消費者行動分析—消費者行動データと消費者選択行動に関する一考察」マーケティングジャーナル、34(2)、47-60
- フィリップ・コトラー(2013)「インタビュー：ビッグデータはマーケティングを変えるのか」、『ダイヤモンド・ハーバード・ビジネス・レビュー October 2013』、54-62、ダイヤモンド社
- 星野崇宏(2009)『調査観察データの統計科学：因果推論・選択バイアス・データ融合』岩波書店
- 星野崇宏(2013)「継続時間と離散選択の同時分析のための変量効果モデルとその選択バイアス補正：Webログデータからの潜在顧客への広告販促戦略立案」日本統計学会誌 43(1)、41-58