

User centric data を用いたオンライン 情報探索行動研究に関する一考察 —国内の User centric data 構築の現状を踏まえて—

高山 純人（静岡県立大学大学院経営情報イノベーション研究科博士後期課程）

消費者行動の研究では、オンライン上での情報探索行動を明らかにするため、消費者のインターネット上での行動ログである「Click stream data」が使用されてきた。Click stream data は、1つの企業サイトやECサイト内の行動ログである「Site centric data」と、登録されたパネルの全てのオンライン上での行動を記録した「User centric data」の2つに分類できる。本研究では User centric data の研究のさらなる発展への示唆を得るため、まずは先行研究から User centric data の研究における課題を抽出し、次に実際に User centric data を構築しているデータベンダーへのインタビューを行い、User centric data の今後の可能性に関する考察を行った。User centric data の研究では、対象者を絞ると小規模なサンプルサイズになることによる分析結果の頑健性の問題や、データ処理の時間的制約、研究方法の確立の遅れなどの課題があることが先行研究から明らかとなっている。またデータベンダーへのインタビューから、サンプル数の課題は大手事業会社との提携によって解決される可能性があることや、データ処理については User centric data への注目度が高まり、分析実績が蓄積されることで、処理速度が向上する可能性が示唆された。研究方法の確立についても、今後 User centric data のデータ取得の費用が下がることで、研究への活用が増加し、様々な研究方法が確立される可能性があることを指摘した。

キーワード：User centric data, Site centric data, オンライン消費者行動

1. はじめに

スマートフォンやタブレット型端末の普及によって、消費者を取り巻く情報環境は大きく変化した。総務省の通信利用動向調査によると、2020年のインターネット利用率は90%を超え、端末別での通信機器保有率はパソコンが70.1%に対して、スマートフォンが86.8%、タブレット型端末も38.7%となっており、インターネットは人々に生活にとって欠かせないものとなっている [1]。

こうした情報環境変化に伴い、オンライン上での情報探索の機会も増加している。神田らは商品カテゴリ別の購入情報源に関する調査を行い、インターネットが重要な情報源であるという質問に対し、高価格帯商品ではほぼすべてのカテゴリで50%を上回っている [2]。青木らは、消費者は

商品の購入を検討する際に企業、ブランド、製品のWEBサイトから得る製品情報はもちろんのこと、家電カテゴリの価格.comや化粧品カテゴリの@コスメなどをはじめとするクチコミサイトの他、Twitter や Facebook での購買経験や製品に関する使用感に関する書き込みなど、オンライン上の情報を起点に、これまで以上に多くの情報を手に入れながら意思決定を行うようになっている [3]。

このような消費者の情報行動の変化によって、企業のマーケティング活動においても、インターネット上での情報発信や商品販売活動が重要となっているが、それと同時に取得したデータのマーケティング活動への利活用についても関心が高まっている。総務省が2020年3月に行った「デジタルデータの経済的価値の計測と活用の現状に関する

調査研究」によると、企業活動において活用しているデータの種類について、アクセスログデータを活用していると回答した企業は約30%となっており、5年前の「ビッグデータの流通量の推計及びビッグデータの活用実態に関する調査研究」と比較してほぼ倍増となっている。また、ECサイトにおける販売記録データについても約20%の企業が活用していると回答しており、こちらも5年前と比較して3倍以上となっている [4] [5]。

消費者行動研究の分野では、以前はパネルへのアンケート調査の回答データを元に研究が行われていたが、2000年頃から消費者の実際の行動ログを用いた研究が行われるようになった。オンライン上での行動ログデータはClick stream dataと呼ばれており、Randolphらは取得するデータの種類によってClick stream dataを大きく2つに分類して、それぞれの特徴について述べている [6]。1つ目は企業サイトやECサイトなど、1つのWEBサイト内での行動を記録したSite centric dataと呼ばれるもので、前述の企業活動に利用されているデータもほとんどがこのSite centric dataにあたり、研究分野でも多く利用されている。一方で、登録されたパネルの全てのオンライン上での行動を記録したデータはUser centric dataと呼ばれ、消費者がオンライン上でどのようなサイトを閲覧し、どのような情報を得

て購入検討を行っているかについて分析することなどが可能なデータである。特にUser centric dataは、消費者の情報探索行動を知る上で有益なデータであるものの、特に国内においては研究の事例は多くは見られない。

そこで本研究は、User centric dataを用いた研究の有用性と課題について、消費者行動研究の先行研究から明らかにするとともに、User centric dataの現状について、実際に構築や運用を行う企業へのインタビューを実施することで、今後のUser centric dataを用いた研究発展への示唆を得ることを目的とする。

2. 購買意思決定プロセスとオンライン情報探索

消費者の購買意思決定プロセスは、Howardらや、Bettmanなど、いくつかのモデル化が試みられている [7] [8]。その中でもEngelらは、購買意思決定プロセスを問題認識、情報探索、選択肢の評価、購入決定、成果（購買後の評価／将来の行動）の5段階としている [9]。

商品に対する不満足や在庫切れなど、消費者が現在使用している商品に対して何かしらの問題を認識した状態が、1つ目のプロセスである「問題認識」にあたる。消費者は問題を認識し、商品へのニーズが高まった後に、「情報探索」を行う。

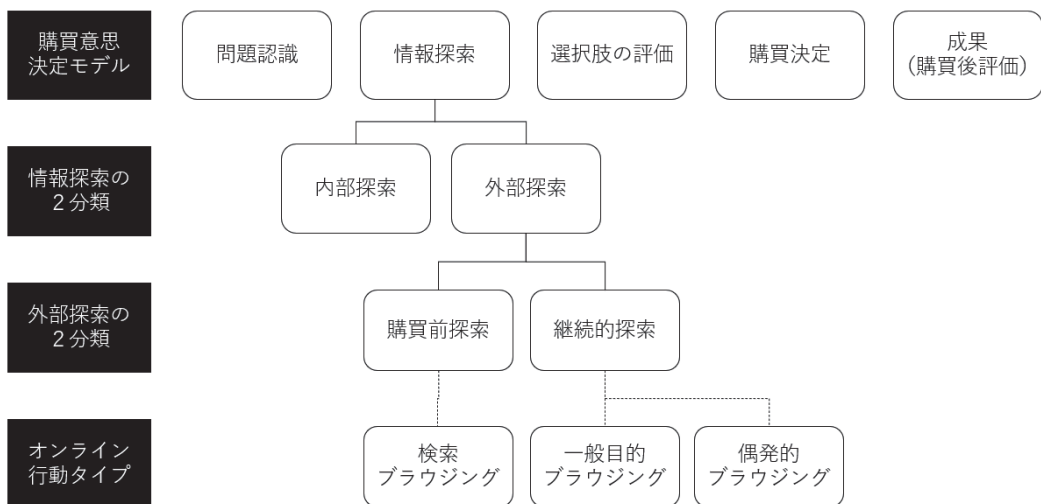


図1 購買意思決定モデルにおける情報探索の分類 出典：Engelら（1968）[9]を参考に筆者作成

情報探索には、大きく分けて2つのプロセスがあると言われている。1つ目は内部探索 (internal search) である。購買意思決定プロセスにおいて、問題を認識した消費者は、自分自身の記憶内から関連のある情報を探し出すことから始める。ここでの記憶には、ブランドに関する記憶や自身の経験に関する記憶が含まれるが、記憶に留まっている情報は非常に限定的である。内部探索の結果、記憶があいまいであったり、製品評価が不十分であったりした場合、例えばカタログや専門誌の記事など、商品に関連する情報の収集を行う。これが2つ目の外部探索 (external search) である [10]。

外部探索については、Block らがさらに購買前探索と、継続的探索の2種類に分類し、その目的について論じている [11]。購買前探索 (Prepurchase Search) は、特定の商品の購入をするための情報収集であり、結果として製品や市場に関する知識の向上、また購買結果への高い満足度をもたらすとされている。もう一方の継続的探索 (Ongoing Search) は、日頃から関心のある製品カテゴリに関してニュースや広告などを見聞きするといった情報探索行動を指し、結果として他者への口コミなどを引き起こす。また、衝動買いの確率を高めたり、探索自体から得られる満足感や問題解決スキルの上昇などをもたらしたりするとされる。

オンライン上での情報探索行動については、Catledge らが、インターネットの歴史の早い段階で、今日でも通用するウェブ上の行動タイプを3種類に分類して論じている [12]。1つ目は、検索ブラウジングで、明確な目的を持ったブラウジングである。検索ブラウジングの特徴は、指向性があり、具体的な検索語で、目的がはっきりしていることである。2つ目の行動タイプは、一般目的ブラウジングと呼ばれるもので、ブラウジングの目的に対するアイデアはあるが、ゴールへの道筋は特定されていないものである。一般目的ブラウジングでは、興味のあるアイテムが含まれている可能性が高いと思われる複数のページを閲覧することが典型的である。最後に、本当にランダ

ムなブラウジング行動と表現される偶発的ブラウジングがある。ここまでに述べた購買意思決定プロセスと、情報探索や外部探索、オンライン行動タイプの関係を図1に整理した。

購買意思決定プロセスの中で特に情報探索は、2000年代のインターネットの普及によって影響を受けた。Ratchford らは、インターネット上での情報探索行動に着目し、1990年と2000年に、自動車購入者を対象として調査を行った [13]。インターネットは他の情報源と同程度の重要な情報源となっていること、また若い世代で、且つ教育水準の高い消費者で、情報探索にインターネットを用いる傾向が強いことが明らかになった。Ratchford らは、インターネット検索とその他の情報源との関係性について調査を行っている [14]。対象は自動車購入者とされた。インターネットによる情報探索量の増加に伴い、ディーラーや印刷媒体を用いての情報探索量が減少していることが明らかになった。特にディーラーにおいての情報探索量は、1990年から2000年、2002年にかけて平均1時間減少しており、消費者の情報探索がオフラインからオンラインへ移行していることが明らかとなった。

3. Click stream data の解析

2000年前後からのインターネットの爆発的な普及に伴って、消費者の情報探索行動は大きく変化した。その変化に応じて、消費者行動研究も Click stream data を使用した研究が徐々に増えた。Click stream data とは、インターネット上でのユーザーの行動を電子的に記録したもので、ユーザーがウェブ上を移動する際にたどる経路を追跡したものである。Click stream data の解析では、どのページにどのくらいのアクセスがあったかというボリュームに関わる項目はもちろんのこと、「パス」と呼ばれるウェブサイト上でユーザーが実際に閲覧したページに順序について明らかにすることで、適切な情報発信への示唆を得ることが可能となる。

最も早く発表された Click stream data 分析

のひとつが、サイトの閲覧行動に関するHubermanらの研究がある [15]。それ以降に Click stream data の分析は頻繁に行われるようになった。

初期の Click stream data を用いた消費者行動研究について Bucklin らは、主な領域は、(1) サイトナビゲーションの選択、(2) EC サイトでの購買決定、(3) ウェブサイト間の選択、(4) 電子エージェント (特にショッピングボット) が関与する購買選択、の 4 領域だと述べている [16]。

サイトナビゲーションの選択とは、閲覧するページ数、ページ (またはサイト全体) の閲覧に費やす時間、特定のページでサイトに留まるか終了するかを選択などが含まれる。EC サイト上での購買決定は、テーマを明確にオンライン上の購入に絞ったもので、購買への転換の確率モデルの開発などが行われている。ウェブサイト間の選択では、ユーザーがあるサイトと別のサイトのどちらを訪れるかを決定するモデルの提案が行われており、ショッピングボットが関与する購買選択では、ショッピングボットの購入への消費者の効用をモデル化する研究などが見られた。

Bucklin らは、オンライン上での消費者行動研究をサイトへの誘導に関する研究、Web サイト利用行動の理解に関する研究、複数サイト・複数チャネル間の行動の研究の 3 つに分類している [6]。サイトへの誘導に関する研究とは、具体的にはバナー広告や検索連動型広告によるサイトへの誘導をテーマとしたものである。Web サイト利用行動の理解としては、訪問時間やサイト内ページ閲覧行動、コンバージョン行動など様々な研究が含まれる。複数サイト・複数チャネル間の行動を捉えた研究では、Web サイト間の行動を見る際にサイトの属性 (娯楽サイト、企業サイトなどのサイトの特徴) の分析などが行われている。

このような研究で扱われる Click stream data が、前述の Site centric data と User centric data の 2 種類である。Site centric data は 1 つのウェブサイトで収集されたデータであるため、訪問者が特定のサイトをナビゲートし、インタラクションする際の行動を非常に詳細に記録することができる。

User centric data は、ウェブ使用中に要求されたすべてのページのユニバーサル・リソース・ロケータ (URL) を、ユーザーのコンピュータからパネル・データの提供者に送信することで取得し、複数のウェブサイトにもたがる活動を記録したデータである。ここからは、それぞれのデータの特徴とその代表的な研究について述べていく。

3.1 Site centric data の特徴と代表的研究

Site centric data は、商品の購入者に絞った行動解析や、サイトへどこから流入してきたのか、その流入元に応じた分析など粒度の細かい分析まで可能であることがその特徴である。そのため Site centric data を用いた代表的な研究も、サイト上での購入に関する研究と、サイトへの誘導に関する研究が多い。

サイト上での購入に関する研究では、Moe は、あるオンライン小売業者の Site centric data を用いて、実際の閲覧行動に基づくサイト来訪者の分類を試みている [17]。まず Web サイト上のページをタイプ別 (ホームページ、商品情報の種類、購入など) に分類し、さらにサイトの閲覧行動を基にセッション尺度、多様性尺度などの一般的なカテゴリの変数を開発した。セッション尺度は製品ページやブランドページなどのページを閲覧したボリュームの尺度で、多様性尺度は閲覧しているカテゴリや製品の多様性を表す尺度と規定している。

これらの指標を基にクラスター分析を行った結果、訪問者は「購買目的」「検索・検討目的」「快楽的閲覧型」「知識構築目的」の 4 つの購買戦略のいずれかに分類されることが判明している。また、快楽的閲覧型、購買目的型、検索・検討目的型を判別するにあたって、これらの尺度は重要であることも明らかになっている。たとえば、快楽的閲覧型ではカテゴリーページの閲覧比率が高く、幅広くカテゴリーページを閲覧している。購買目的型では閲覧するカテゴリが限られており、製品ページの繰り返し閲覧回数が多い。検索・検討目的型も似た特徴を持つものの、製品ページの繰り返し

返し閲覧回数は購買目的に比べると少ない。また、購買目的クラスターのユーザーは最も購入する可能性が高く、知識構築クラスターのユーザーは最も購入する可能性が低いという結果も実証されている。

Sismeiro らは、サイト上において購入に至るまでのプロセスを、ユーザーが購入するために完了しなければならない一連のタスクに分解している [18]。この研究では、インターネット上の自動車販売店の Site centric data を用いて、3つのユーザータスクのモデル化を試みた。最初のステップは商品設定の完了、次に個人情報を入力、最後がクレジットカードによる注文確認の3ステップである。より一般的に言い換えれば、商品をショッピングカートに入れ、配送先情報を入力し、クレジットカードで注文するという、EC サイト上での一連の流れに沿ったステップにあたる。この研究ではサイト訪問者のうち、注文手続きを完了したのは約2%であったが、タスクごとに分解すると、購入プロセスはサイト訪問者の30%が1つ目の商品設定のタスクを完了し、そのうち20%が次の配送先情報入力のタスクを完了し、双方を完了した人の34%が最後のクレジット情報入力確認のタスクを完了したことが明らかになった。

サイト誘導に関する研究では、Chatterjee らは、セッション中に広告がまだクリックされていないユーザーが、バナーをクリックする確率をモデル化した [19]。その結果、消費者の属性、回線速度、インターネットアクセス料金、オンライン経験などが、広告のクリック行動に影響を与えることが明らかとなり、平均的にあまり広告をクリックしない消費者の場合には、バナーを繰り返し表示することでクリック率が上昇することも判明している。また、消費者はナビゲーションパスの早い段階でも遅い段階でも同じようにバナー広告をクリックすることが明らかになっている。

Rutz らは、自動車関連のウェブサイトでバナー広告の露出が、その後の訪問者のページビュー選択にどのような影響を与えるかについて研究を行っている [20]。彼らは Site centric data を用いて、バナー広告の露出後にユーザーが行う製品関連情

報のページ固有の選択をモデル化した。その結果、現在のセッションで配信されたバナー広告は、その後の閲覧行動に対して、ユーザーセグメントごとに異なるレスポンス効果を持つことがわかった。1つのセグメント（訪問者の37%）は、宣伝されたブランドに関する詳細情報を求めて応答し、2番目のセグメント（23%）は否定的に反応し、3番目のセグメント（40%）は、バナー広告の露出に対して大きな反応ないことが述べられている。

この他にも、Site centric data を用いた研究は様々なテーマで行われている。

3.2 User centric data の特徴と代表的研究

Site centric data では、EC サイト内での購入に焦点を当てた研究や、サイト流入に焦点を当てた研究が多く見られる。一方、User centric data を用いた研究に関しては、サイトナビゲーションの選択や、サイト間比較等を用いた研究については見られるものの、サイト外での情報探索行動や、商品購入に焦点を当てた研究については Site centric data と比較して多くは見られず、特に日本国内では、その研究事例は少ない。その中でも代表的な研究について取り上げる。

Johnson らは、Media Metrix 社の User centric data を使用し、複数のサイトを対象としてユーザーの同一サイトへの訪問回数と1セッションあたりの滞在時間について調査した [21]。この研究では、ウェブサイトへの訪問回数が2倍になると、1回の訪問でサイトに滞在した時間は平均19%減少すると明らかとなった。これは、ユーザーがサイトを再訪する際に、サイトの操作方法やコンテンツに慣れることで、より効率的な時間配分が可能になることを示唆している。

Johnson らは、同じく Media Metrix 社の User centric data を使用し、書籍、コンパクトディスク、航空券などの商品を購入した1万世帯以上のインターネットユーザーの分析を行った [22]。その結果、CD と書籍の購入者の約70%は、1つのウェブサイトしか利用していないことを明らかにしている。これは旅行カテゴリにおいても

同様の結果となっている。

Montgomery らが行ったモデリング研究では、Media Metrix 社のデータを用いて、Barnes and Noble 社のウェブサイトでの行動に焦点を当てている [23]。Moe と同様に、彼らはサイトのページをいくつかの異なるページタイプのカテゴリ（ホームページ、情報、製品、ショッピングカート、注文など）に分類し、次にユーザーがサイトをナビゲートする際にどのページタイプからどのページタイプへと移行するかについてのモデリングを試みている。

日本国内での User centric data の研究を見ると、新美らはビデオリサーチ社の WEB PAC パネルのデータを使用し、(1) 自社サイトへ顧客を誘導する際の Web 広告・コミュニケーション活動を事前に策案するためのモデルと、(2) 実際の自社サイトへの訪問者の経路情報の分析の同時モデルを提案している。具体的には、サンプルセレクションモデルを用いて、(1) 個人が特定の Web サイトを閲覧するかどうか、(2) 訪問した場合の閲覧時間の 2 段階のモデル構築を行っている [24]。

User centric data を活用した研究の課題として、Bucklin らは対象者を限定するとサンプルサイズの問題が発生してしまうという分析結果の頑健性の課題を指摘している。また、Click stream data 研究全体の課題として、データ処理に時間がかかることや、マーケティングへの活用については方法論の確立がコンピュータサイエンスの分野に比べて遅れていることなどを挙げている [6]。先行研究で明らかとなっている User centric data の課題について、その原因は大きく 3 つあると考えられる。1 つ目は、ユーザーの確保が難しいことである。例えば、1 万人のデータ提供者を獲得できたとしても、日本国内のインターネット人口からの割合で考えると決して多いとは言えず、分析対象のサンプルサイズの問題が生じる可能性が高い。また EC サイトの購入者などに対象を絞ろうとする場合は、更に対象者の数が少なくなるため、User centric data は、分析に耐えうるサンプル数の確保が大きな課題となることが考

えられる。

2 つ目は、集計の煩雑性が挙げられる。特に User centric data の場合は Site centric data と比べて複数のサイトが分析の対象となるため、データの抽出や集約の工数が大きい可能性が考えられる。

3 つ目の要因は、データ利用にかかる費用面である。クロス集計などに耐えうるサンプル数を確保しようとすれば、その費用はより高額になる。User centric data を構築しているデータベンダー数は少なく、研究者側の選択肢も少ないため。価格を指標としてデータベンダーを選択することは難しい。

これらの課題やその要因仮説を解決していく方法はあるのか、実際に User centric data を構築、運用を行っている企業にインタビューを行い User centric data の研究における課題や要因仮説の確認を行い、その解決へのヒントを探った。

4. User centric data 運用者へのインタビュー

4.1 インタビュー対象企業について

本研究では、日本最大級の WEB パネルを保有する株式会社ヴァリューズに対してインタビュー調査を行った。

株式会社ヴァリューズは2009年に日本で設立され、インターネット行動ログ分析事業を中心にビッグデータ/AI/クリエイティブに関するコンサルティング事業などを展開している。現在の社員数は約100名。インターネット行動ログ分析事業は、会員数1,000万人を超えるクレジットカード会社と提携した共同事業として展開。クレジットカード会員のうち、約20万人程度の許諾を得て、アクセスした全 WEB サイトの URL を取得し分析用のデータベースを構築。2018年からは200万人の会員を保有する事業者とも提携し、日本最大級の User centric data を保有している。今回はヴァリューズ社に所属するビジネスプロデューサーとカスタマー担当部署のリーダーの 2 名の社員の方に協力いただき、インタビュー調査を実施し

た。

4.2 運用者から見た User centric data の課題

テクノロジーが進化した現在においても、User centric data を構築、維持する難易度はいまだに高い。それには大きく3つの要因があると考えられる。1つ目はユーザー数の確保の難しさ、2つ目はデータ取得や集計処理などの難しさ、3つ目は人材確保の難しさがある。2つ目のデータ取得や集計処理の難しさは創業当時に比べて現在は解消されつつあるが、ユーザーの確保と人材確保の難しさについては、現在も大きな課題となっている。

4.2.1 ユーザー数確保の課題

3つの要因の中でも最も大きな要因となっているのが、ユーザー数の確保の難しさである。現在大手クレジットカード会社と提携した User centric data のパネルは約20万人の登録があるが、日本のインターネット人口の0.2%程度となっているため、まだサンプル数が十分とは言えない。例えば、Site centric data で10万ユーザーが確認できているWEBサイトを対象とする場合、ヴァリューズ社のパネルでは約200人の分析用サンプルしか確保することができない計算となる。そこからさらにターゲットを絞り込んでクロス集計を行う場合には、さらにサンプルが難しくなる。User centric data は10万人以上のパネル確保できたとしても、分析に耐えうるとは言い難い。

また、オンライン上の行動データの提供を許可したパネルを1度に大量に確保する方法は無い。パネルをインターネット広告のプロモーションで獲得する場合、その費用は少なくとも1人のパネル獲得あたり1,000円以上となる。また、データベースの管理やモニター維持にも最低限1人以上のエンジニアが必要となるため、その人件費も含めると、そのコストは膨大なものとなる。

ヴァリューズ社では前述の通り、2018年に新たな事業会社と提携した。多くの会員を抱える事業

会社との提携は、低コストでサンプル確保が可能な1つの方法ではあるが、会員のデータを提供することは個人情報保護なども障壁となるため、そもそも外部の事業者と提携しデータ事業を行うことに消極的な企業も少なくないものの、データビジネスへの関心度は年々高まっている。

4.2.2 集計処理などの技術の課題

クレジットカード会社の会員からのデータ取得方法は、専用の検索ツールバーをブラウザにインストールし、ツールをインストールされたブラウザに打ち込まれたURLを、サーバに記録していくという手法である。2009年の創業当時は、このデータ取得の技術を開発できるエンジニアが社内にはいなかったため、海外の企業と技術提携を行った。

また、Site centric data の集計の場合には、URLの構造に共通のルールがある場合が多く、分析に際してもそのルールに則ってデータの集計を行うことが多いが、User centric data の場合、ドメインやURLの構造も各サイトによって異なるため、集計処理にかかる工数も大きい。これについては、一部自動処理を行っているものの、今後データの活用パターンや集計パターンが蓄積されていけば、データ処理の自動化範囲を広げる開発も行われる可能性がある。

4.2.3 人材確保の課題

User centric data の構築には多くの費用が必要となるため、データベンダーとしてはデータベース構築のコストを捻出することも同時に必要となる。それを人材という視点で見ると、データベースの構築や集計、分析を行うエンジニアの他に、ビジネス視点やマーケティング視点でデータの価値を捉えることができる人材も欠かすことができない。データベース構築やデータ分析のスキルを保有し、かつビジネスやマーケティング視点をもっている人材は非常に稀有な存在ではあるが、User centric data の構築には、その双方を保有

する人材が必要となる。ヴァリューズ社が User centric data を構築できた背景には、創業時のメンバーにデータベースの構築とデータ分析に長けたエンジニアと、データを理解しビジネス視点を持ったが創業者が含まれていたという要因もあった。

現在は領域を超えたチーム編成や人材育成が盛んに行われているため、今後はビジネス人材とエンジニア人材を保有する企業が増える可能性も考えられる。

4.2.4 User centric data の新たな機会

ここまで述べたように、User centric data の運用や構築にはいくつかの課題があるため、新規で立ち上げることは難しく、データを提供するベンダーが急増することは考えにくかった。一方で、昨今の環境変化によって、そのデータへの注目度は更に高まっている。

その要因の1つ目は、コロナ禍による人々の消費行動の変化である。新型コロナウイルス感染症の拡大により、人々の消費行動の中心がオンラインへと急速に移行した。そのため、消費者の全てのオンライン上の行動を捕捉することができる User centric data の分析は、消費者の購買行動を明らかにする上で、さらに重要度が増加している。

2つ目の要因は Site centric data の取得に対する懸念がある。多くの場合、Site centric data を取得しているツールは、cookie の技術を活用している。しかし、今後は cookie の利用制限などが行われるため、自社のサイト来訪者からのデータ取得にも許諾が必要となる可能性がある。Site centric data の取得の難易度が高まると、企業はその分析を User centric data に頼らざるを得なくなる。このような要因で、User centric data への注目は、昨今更に高まっている。

近年見られる User centric data への新たなニーズは、スマートフォンアプリのアクセスログの取得である。消費者は購買行動プロセスにおいて、通常のスマートフォンサイトで情報探索と同様に、

スマートフォン上のアプリを使用しているケースがある。

現在のデータ取得は safari などの一般ブラウザからの取得が主となっているため、消費者のオンライン行動を全て取得できているとはいえない。今後はアプリの立ち上げや閲覧についてもより詳細な情報を取得可能となれば、購買に関連するデータ分析についても、より幅広く発展すると考えられる。

4.2.5 インタビュー内容まとめ

User centric data を実際に構築し、データ提供を行っているデータベンダーの視点から挙げられた User centric data 構築の課題は、ユーザー数確保の課題、集計処理などの技術の課題、人材確保の課題の3点であった。この3つの課題を解決できなければ、分析に耐えうる User centric data を新たに立ち上げることは難しい。ただしそれぞれの課題に関して、その解決の兆しも少しずつ見えてくる。

また昨今の環境変化によって、Site centric data の取得が難しくなるかもしれないという懸念が発生していることや、コロナ禍によって消費者の情報探索や商品購買のオンライン化が加速しているという状況もあり、User centric data への期待が一気に高まる可能性があると言及した。

5. 考察

消費者行動研究における先行研究では、購買意思決定プロセスの中でも情報探索のステップがインターネットの普及によって変化が起きやすいということが示されており、その中でも外部探索、特に購買前探索においてオンライン上での情報探索は重要な意味を持つことが明らかになった。

また、オンライン上での情報探索の中でも、実際の消費者の行動ログである Click stream data を活用することは、アンケートによる調査では得られない詳細な情報探索行動の分析を行うことができる。

Click stream data の1つである Site centric data の分析では、WEB サイトでの閲覧コンテンツの選択や、EC サイト内での商品購入の確率予測などの研究があり、企業のマーケティングにも活用可能な示唆がある。ただし消費者のオンライン情報探索は複数の WEB サイトをまたいで行われる可能性が高いため、消費者の全ての情報探索を捕捉することは難しい。

User centric data は、その情報探索を全て補足したデータである。そのため、消費者行動研究においても有益なデータである。しかし、その研究には大きく3つの課題があることが先行研究で明らかになっている。第1に対象者を購入者のみに限定するなどターゲットを絞った場合、サンプル数が不足するケースが多いため、分析結果の頑健性を確保できないという点。第2に、URL の構造などがサイトにより異なるため、集計に大きな手間がかかるという点。第3に研究方法の確立が遅れているという点である。

この3つの課題については、その要因が解決される可能性があることをデータベンダーのインタビューは示唆している。

1つ目は分析結果の頑健性については、データ

ベンダーが多くの会員を抱える事業会社との提携を増やすことで、現在もサンプル拡大に努めていると述べている。データベンダー側でサンプル数が拡大すれば、研究においても詳細なターゲティングを行った状態の分析も可能となるため、研究の幅の広がりを見せる可能性がある。

2つ目の課題である分析に時間がかかるという点については、分析の実績が蓄積されていけば、研究に用いられることが多いデータ加工のパターンや、データ集計の視点などが蓄積されていくと、繰り返し用いられるデータ加工や集計のパターンについては、データベンダー側でも自動化ツールの開発の可能性があることから、データ処理の工数が今後削減される可能性がある。

3つ目の分析方法の確立が遅れているという課題については、データ利用の費用が下がることで、研究事例が増えれば、分析方法が確立されていく可能性がある。User centric data 構築にもいくつか課題があることが確認されたが、本稿で取り上げたヴァリューズ社のように、エンジニア人材とビジネス人材の掛け合わせによって User centric data の開発を行うプレーヤーが増えれば、ビジネスの競争原理によってデータ利用の費用が

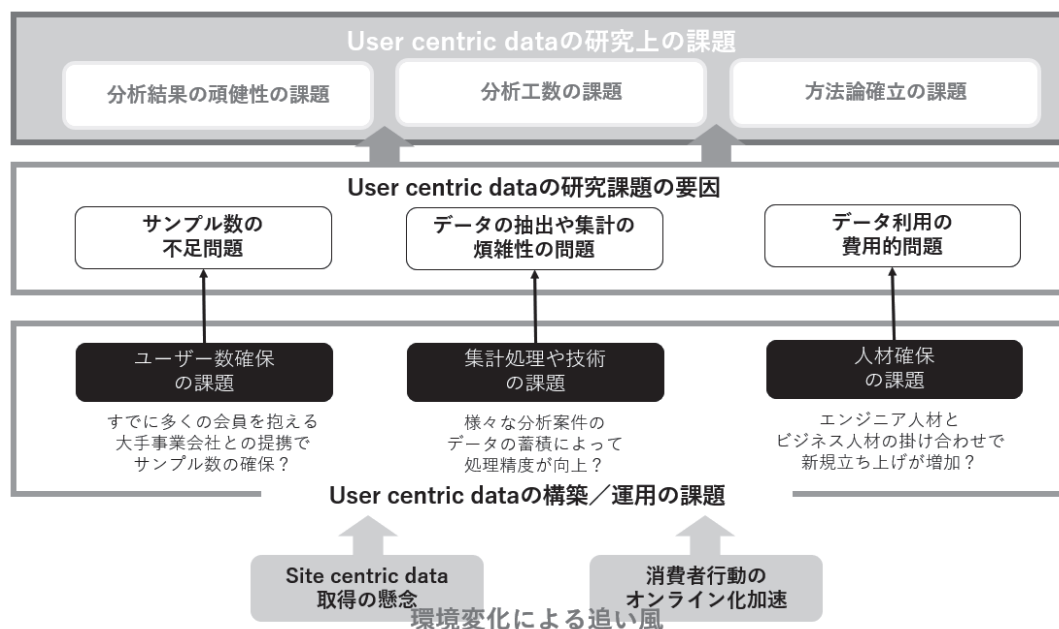


図2 User centric data の研究課題と、構築／運用の課題の関係性（インタビューから筆者作成）

下がる可能性があり、市場の拡大によって、副次的に研究拡大に貢献する可能性を示唆している。

6. 結論

本研究では、User centric data の研究に関する先行研究と、実際に User centric data の構築と運用を行っているデータベンダーへのインタビューから、User centric data の研究における課題とその解決の可能性について調査した。その結果を図2に整理した。

環境変化の追い風も契機にして構築や運用での課題を解決することにより、研究そのものの課題についても解決の可能性があることが示唆された。しかし、消費者にスマートフォンアプリなどが普及したことにより、オンライン上での情報探索はブラウザ上のみではなく、スマートフォンアプリを通して行われている可能性が高いが、現在の User centric data では、スマートフォンアプリでの情報探索行動まで取得することができない。今後スマートフォンアプリ上での行動ログデータとブラウザ上での行動ログデータを保有した User centric data の開発が進めば、User centric data を用いた消費者行動研究はさらなる発展を遂げる可能性がある。

参考文献

- [1] 総務省：令和2年度通信利用動向調査、
https://www.soumu.go.jp/johotsusintokei/statistics/data/210618_1.pdf (2021/07/20最終閲覧)
- [2] 神田晴彦、鳥山正博、清水聰：購入に影響を及ぼす情報源と情報発信の変化 —39商品カテゴリーの横断分析—、マーケティング・ジャーナル、128、pp.79-91、2013
- [3] 青木幸弘、新倉貴士、佐々木壮太郎、松下光司：消費者行動論、有斐閣、2012。
- [4] 総務省：デジタルデータの経済的価値の計測と活用の現状に関する調査研究、
<https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r02/pdf/n3200000.pdf> (2021/07/20最終閲覧)
- [5] 総務省：ビッグデータの流通量の推計及びビッグデータの活用実態に関する調査研究、
https://www.soumu.go.jp/johotsusintokei/linkdata/h27_03_houkoku.pdf (2021/07/20最終閲覧)
- [6] Bucklin, R. E. and Catarina, S.: Click here for internet insight: Advances in clickstream data analysis in marketing, Journal of Interactive Marketing, 23(1), pp. 35-48, 2009.
- [7] Howard, J. A. and Sheth, J. N.: The Theory of Buyer Behavior, John Wiley & Sons, 1969.
- [8] Bettman, J. R.: An Information Processing Theory of Consumer Choice, Adison-Wesley, 1979.
- [9] Engel, J. F., Kollat, D. T. & Blackwell, R. D.: Consumer Behavior, Holt, Rinehart and Winston, 1968.
- [10] 平久保伸人：消費者行動論、ダイヤモンド社、2005。
- [11] Block, Peter H., Sherrell, Daniel L. and Ridgway, Nancy M.: Consumer Search: An Extended Framework, Journal of Consumer Research, 13 (June), pp.119-126, 1986.
- [12] Catledge, L. D. and Pitkow, J. E.: Characterizing browsing strategies in the world-wide web, Comput. Netw. ISDN Syst., 27(6), pp.1065-1073, 1995.
- [13] Ratchford, B., Lee, M.-S. and Talukdar, D.: The Impact of the Internet on Information Search for Automobiles, Journal of Marketing Research, 40 (2), pp.193-209, 2003.
- [14] Ratchford, B., Talukdar, D. and Lee, M.-S.: The Impact of the Internet on Consumers' Use of Information Sources for Automobiles, Journal of Consumer Research, 34 (1), pp.111-119, 2007.

- [15] Huberman, B., Pirollis, P., Pitkow, J., and Lukose, R.: Strong Regularities in World Wide Web Surfing, *Science*, 280(3), pp.95-97, 1998.
- [16] Bucklin, R.E., Lattin, J.M., Ansari, A., Gupta, A., Bell, D., Coupey, E., Little, J.D.C., Mela, C., Montgomery, A. and Steekel, J.: Choice and the Internet:From Clickstream to Research Stream, *Marketing Letters*, 13(3), pp.245-258, 2002
- [17] Moe, W. W.: Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream, *Journal of Consumer Psychology*, 13(1&2), pp.29-39, 2003.
- [18] Sismeiro C, and Bucklin, R. E.: Modeling Purchase Behavior at an E-Commerce Web Site: A Task-Completion Approach, *Journal of Marketing Research*, 41(August), pp.306-323, 2004.
- [19] Chatterjee, P., D. L. Hoffman, and T. P. Novak.: Modeling the Clickstream: Implications for Web-Based Advertising Efforts, *Marketing Science*, 22(4), pp.520-541, 2003.
- [20] Rutz, Oliver J., and Bucklin, R. E.: Effects of Banner Ad Exposure on Internet Site Browsing Behavior, Working Paper, School of Management, Yale University, 2007.
- [21] Johnson, Eric J., S. Bellman, and G. L. Lohse.: Cognitive Lock-In and the Power Law of Practice, *Journal of Marketing*, 67(April), pp.62-75, 2003.
- [22] Johnson, E. J., Moe, W., Fader, P. S., Bellman, S., & Lohse, J.: On the Depth and Dynamics of Online Search Behavior, *Management Science*, 50(3), pp.299-308, 2004.
- [23] Montgomery, A. L., Li, S., Srinivasan, K., & Liechty, J. C.: Modeling Online Browsing and Path Analysis Using Clickstream Data, *Marketing Science*, 23(4), pp.579-599, 2004.
- [24] 新美潤一郎、星野崇宏：ユーザー別アクセス・パターン情報の多様性を用いた顧客行動の予測とモデリング, *応用統計学*, 44(3), pp. 121-143、2015

**A Study on Online Information Seeking Behavior Using User Centric Data
—Based on the current status of user centric data construction in Japan—**

SUMITO TAKAYAMA

Graduate School of Management and Information of Innovation, University of Shizuoka

Abstract:

In the study of consumer behavior, "Click stream data", which is a log of consumers' behavior on the Internet, has been used to clarify their online information-seeking behavior. Click stream data can be categorized into two types: "Site centric data," which is a log of behavior within a single company site or e-commerce site, and "User centric data," which records all online behavior of a registered panel. In this study, to obtain suggestions for the further development of research on user centric data, we first extracted issues in research on user centric data from previous studies, and then conducted interviews with data vendors who are actually constructing user centric data. Previous studies have shown that user centric data research faces issues such as the robustness of analysis results due to the small sample size when the target population is limited, time constraints on data processing, and delays in establishing research methods. In addition, from the interviews with data vendors, it is clear that there are issues such as the robustness of analysis results due to the small sample size, time constraints of data processing, and delays in establishing research methods. Interviews with data vendors suggest that the issue of the number of samples may be resolved by forming alliances with major business companies and that the processing speed may be improved as more attention is paid to user-centric data and more analysis results are accumulated. As for the establishment of research methods, it was pointed out that the cost of acquiring user centric data will decrease in the future, which will increase its use in research and may lead to the establishment of various research methods.

Key words: User centric data, Site centric data, Consumer online behavior