

調査の現場から

ビッグデータとパーソナルデータとマーケティング

長崎貴裕 (株式会社インテージ・マーケティングイノベーション本部・開発ユニットディレクター)

1 パーソナルデータとマーケティング

🔍 データの種類

近年、ビッグデータの活用が企業において大きな話題となっている。その背景には企業側に活用されていない大量のデータが蓄積されているという状況もあるし、企業間の競争、とくに価格競争が厳しいなかでデータを基により合理的な意思決定をしたいという事情もある。企業が活用しているデータとしては様々なものがあるが、マーケティング領域に関連するデータは多くの場合個人に紐づくパーソナルデータである。そのパーソナルデータは自動的に取得されるログデータであることも多い。ここではマーケティングに関連するデータとして「購買（販売）データ」「位置情報」「ネット上のログデータ」「メディア接触データ」を取り上げる。

🔍 購買（販売）データ

購買データの代表的なものはPOSデータである。POSデータは商品の販売データを小売店が記録しているものであり、スーパーなどのレジの記録である。POSデータでは「いつ」「どの店で」「どの商品が」「いくつ」「いくらで」「どんな商品と一緒に」売られた

かがわかる。POSデータは流通の自社内での活用も行われているが、市場調査会社を通じて複数の流通のPOSをまとめたものが小売店パネル（POS）データとしてメーカー等に提供されている。

ID-POSはPOSデータに「誰が買った」という情報が付加されたものである。スーパーなどで会員カードを提示することで割引やポイントがもらえるというサービスがこれにあたる。誰が購入したかがわかり、属性別の分析が可能になる。ID-POSでは購買が複数のチェーンにまたがった場合、同じ人が何を買ったかという分析はできない。チェーンごとに会員カードが分かれているからである。チェーンを超えて分析が可能になるのが「Tポイント」や「ポンタ」に代表される共通ポイントサービスである。共通ポイントの加盟店の範囲内であれば、購入者の分析が可能になり、ID-POSが進歩した形態と理解できる。一方、これらのデータ量は非常に大きく価値あるものではあるが、消費者側の視点でデータを分析しようとしたときにすべての購買が網羅されているわけではないという問題がある。

購買データは消費者（購買）パネル調査でも手に入れられる。消費者（購買）パネル調査は、購入した商品を消費者に記録してもらう調査である。購入した人のプロフィール情

報や意識などの情報も合わせて分析できる。現在、日本で最大の消費者パネル調査は5万人を対象としている。POS データと比べるとデータの量は少ないが、多くの場合、目的とする分析に十分なサンプルサイズである。サンプルサイズの大きさとデータのカバレッジの両方を満たすデータはないので、分析目的に応じて使い分けが行われている。

📍 位置情報

携帯電話の普及により位置情報の活用が進んでいる。代表的なものとしては、NTT ドコモの「モバイル空間統計」がある。モバイル空間統計は、携帯電話の基地局が周期的に各端末のエリアごとの数を把握している。性別・年齢層別・居住エリア別の時間帯ごとの人口の構成（人口構成）などを推定することが可能となり、防災分野や地域開発やマーケティングなどの分野での活用が期待されている。このサービスについては、プライバシー保護の観点から厳格な制限がある。

GPS 等を利用した位置情報は、各端末側でユーザーが取得することができる。スマートフォンの地図上で自分が現在どこにいるかがわかるのはこの機能によるものであり、スマートフォンのアプリによって継続的に移動を記録することもできる。同種のデータとしては交通カードの乗降記録もある。個人の同意があれば位置情報のマーケティング活用も可能になるデータではあるが、プライバシーとのバランスは十分に配慮する必要がある。

📍 ネット上のログデータ

ネット上の行動履歴データは通常、何らかのかたちで記録されている。サイトを訪問した場合には、サーバーにその記録は残る。サーバーから見ると、誰が来たかはわからない

が自社サーバー上の行動は把握できる。多くのサイトが cookie² という識別用のファイルを PC などの端末に送り込むので、同じ人が再びサイトを訪れたことはサーバー側でわかる。とはいえ、cookie はセキュリティソフト等で消されることも多く、確実に把握できるわけではない。もちろん、会員サイトのようにログインを行っている場合には個人別のデータを把握することができる。

ネット行動全体を把握したいというニーズに対応するのが、オンライン視聴パネルである。調査対象者の PC やスマートフォンにトラッキング用のソフトをインストールして、Web のアクセス状況を把握するものである。数万人のアクセス履歴を把握しているパネルもあり、前述の小売店パネルや消費者（購買）パネルと同じような構造がネット上のログデータでも起きている。

ネット上のデータとしては SNS やブログなどのソーシャルデータをサービスしている会社もあり、ネット上での話題の盛り上がりについてのデータ活用は進んでる。

📍 メディア接触データ

メディア接触データの代表が、テレビ視聴率である。いわゆるテレビ視聴率は世帯でのテレビ視聴率であり、サンプリングによる調査データが一般的であった。今後スマートテレビが普及すれば、スマートテレビの視聴データはメーカー側でも把握可能であり、視聴者の許諾が前提とはなるものの、今までと比べ詳細かつ大規模なデータが活用されていく可能性も考えられる。テレビ以外の新聞、ラジオ、雑誌、交通広告などのメディア接触データはアンケートで把握されることが多い。

2 パーソナルデータをめぐる問題

✿トラブル事例

パーソナルデータの活用がトラブルに発展するケースが増えている。2013年6月、日立はJR東日本のSuicaの駅利用情報を基にした「交通系ICカード分析情報提供サービス」の提供開始を発表したが、利用者の反発を受けJR東日本から社外へのデータ提供は休止となっている。このケースについては、サービスの開始にあたり個人情報保護という観点では注意を払っているように感じられたものの、事業サービスの発表の仕方とその後の対応に失敗したことで、利用者の理解を得られない状況となった。

2012年には、Tポイントツールバーからのデータ収集が情報セキュリティ上の不備から休止となったケースが生じている。ネットでの情報収集については、セキュリティ面でも配慮しないと問題が生じる。同じく2012年には、無料無線LAN利用者からの情報収集について総務省からの指導が入ったケースもある。このケースでは、電気通信事業法第4条に規定する「通信の秘密」を侵害したとされている。

パーソナル情報の収集にあたっては、個人情報保護法などの法令順守およびセキュリティの担保は当然のことではある。情報の収集については利用規約に記述していれば許されるということではなく、利用者への説明をわかりやすく行うことが大事である。利用者の理解が得られないような情報収集は慎まなければならないし、パーソナルデータの利活用については高い企業倫理をもつことが求められている。

✿パーソナルデータの活用検討

このような状況のなかでパーソナルデータの活用についての政府による検討も始まっている。2013年6月に閣議決定として出された「世界最先端IT国家創造宣言」のなかで「『ビッグデータ』のうち、特に利用価値が高いと期待されている、個人の行動・状態等に関するデータである『パーソナルデータ』の取扱いについては、その利活用を円滑に進めるため、個人情報及びプライバシーの保護との両立を可能とする事業環境整備を進める」と記されている。

高度情報通信ネットワーク社会推進戦略本部（IT総合戦略本部）のパーソナルデータに関する検討会も同年9月に開始されている。それに先立って、総務省からは同年6月「パーソナルデータの利用・流通に関する研究会」報告書が公表されている。この報告書のなかでは「パーソナルデータの利用・流通による可能性とその課題」と「パーソナルデータの適正な利用・流通の促進に向けた方策」が記述されている。その方向性としてプライバシー・コミッショナー制度やマルチステークホルダープロセスなどの枠組みも同時に提言されている。プライバシー・コミッショナー制度は「パーソナルデータの利活用に関わる様々な問題について、専門的な知見を有する人材が、パーソナルデータの利活用の基本理念及び原則を実質的に判断して、分野横断的に迅速かつ適切に処理していくことを可能」にするためのものである。マルチステークホルダープロセスは「企業等が自主的に宣言したポリシー・ルール等への遵守を確保する」ための制度である。

これらから読み取れるのは、国としてパーソナルデータの利用促進を進めるという方向

Web 広告とテレビ CM のリーチパターン

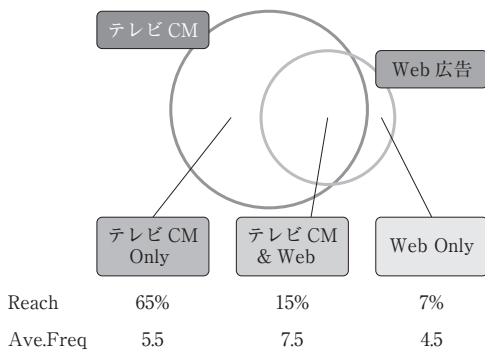


図 1 広告到達状況 (例)

当該商品について

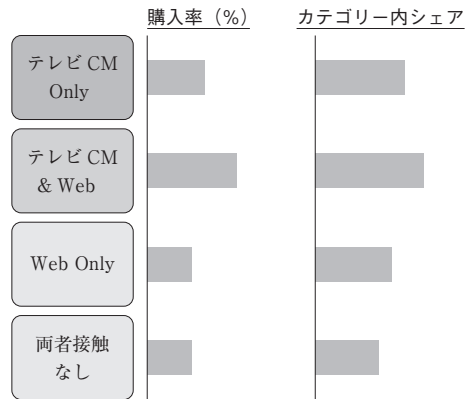


図 2 広告接触と購買の関係 (例)

性である。今後、様々な取組みが実現に向けて動き出すものと思われるが、生活者にとってもデータ活用側にとっても、納得できるかたちになるべきである。

3 シングルソースデータと DMP

シングルソースデータ

パーソナルデータのなかでも最近注目されているのがシングルソースデータである。

シングルソースデータとは、同一個人から購買・広告接触・意識項目・属性などの様々な情報を収集したデータである。元来、マーケティングリサーチのデータは別々の人からそれらを取得していた。「購買」と「メディア接触」という情報も、それぞれ別の人から収集していたのである。別の人から収集していた大きな理由は、調査対象者の負荷を軽減するためである。近年はログでのデータ収集を行うことで、調査対象者に負担をかけることなくシングルソースでのデータ収集が可能になった。これまでは、複数ソースから収集したデータをそれぞれ分析、もしくはデータ

フュージョンによる疑似結合を行うことで分析するしかなく、複数の事象の因果関係を明確に把握することは困難であった。

現在日本では、購買、テレビ視聴、ネット行動 (PC)、ネット行動 (スマートフォン)、属性などのプロフィールデータやブランドに対する意識等をシングルソース化した消費者パネルが実現している。この消費者パネル調査により、図 1 のように Web 広告とテレビ CM がどのように消費者に到達しているかを分析することが可能になる。

また、広告接触と購買の直接的な関係についても、図 2 のように把握することができる。

クロスメディアでの広告接触をした場合の広告効果が測定できるということになる。

購買に至るまでの動線 (カスタマージャーニー) の可視化もできる。カスタマージャーニーを見ていると、その商品についてテレビ CM を見て、店で購入して、その商品について検索して、またその商品を購入してといった行動全体を把握することができる。図 3 は、ある人の A ラーメンについての情報接触と購買行動とネット上の行動を抜き出したものである。

日付	行動
X月20日	CM視聴①
X月23日	しょうゆ味5パック入り購入
X月25日	しょうゆ味5パック入&みそ味5パック入購入
X月26日～月14日	CM視聴②～⑥
△月15日	検索 A ラーメン ⇒ 検索 A ラーメン みそをアレンジ ⇒ 閲覧レシピサイトB ⇒ 検索 ⇒ レシピサイトC
△月16日	CM視聴⑦
□月13日	検索 A ラーメン ⇒ 検索 A ラーメン みそ味に一工夫 ⇒ 閲覧レシピサイトB ⇒ 閲覧 関連ブログ ⇒ 検索 ⇒ レシピサイトD
□月15日	CM接触⑧⑨
□月26日	みそ味5パック入購入
◇月22日	みそ味5パック入購入

図3 A ラーメンのカスタマージャーニー

このように、カスタマージャーニーには分析上の魅力は大きい。しかし生活の様子がダイレクトにわかるので、そのデータの取扱いについては慎重に行う必要がある。シングルソースデータはセンシティブなパーソナルデータともいえる。

🔗 DMP (Data Management Platform)

上記のシングルソースデータは、あくまでもリサーチに協力する調査パネルで構成されているものである。現在、調査パネルではないシングルソースデータが構築されつつある。³それがDMPである。

狭い意味でのDMPは、ネット広告をターゲットに表示するためのデータベースである。ネット上の行動履歴は個々のブラウザごと(たとえばあなたのPCのIE)に、cookieをベースにある程度把握されている。個々のブラウザはUB(ユニークブラウザ)と呼ばれ、UBに対してWeb広告の出し分けを指定する仕組みが⁴DSPである。話がややこしくなるので、狭義のDMPはDSPと一体であると理解してもらえばよい。⁵

狭義のDMP(DSP)は広告を配信する人

からみると、以下のような項目を指定してWeb広告を行うツールということになる。

・性×年代, エリア, 興味領域, 最大広告表示回数, 期間

たとえば「東京に住む30, 40歳代男性で車に興味をもつ人に最大10回広告を出してください」というようなことが行われる。年齢や性別は推定に基づくことが多い。興味領域はそのUBが訪問したサイト履歴に基づくものである。

DSPは単に広告配信だけではなく、広告の効果(クリック, サイトでの申込み)を把握する機能も併せもつことが多い。そのため、広告主からみるとネット広告の出稿の指定と結果が把握できるツールでもある。

ネット広告を多用する広告主にとってDMP, DSPは普通に利用するものである。DSP業者は約2億UB(5,000万~4億UB)を把握していると称しているが、cookieはアンチウイルスソフト等で消されるケースも多く、消された後に再度cookieが入るなど二重カウントされていることも多く完璧なものではない。データとしては完璧ではなくとも、広告効率を上げるという面では機能している。

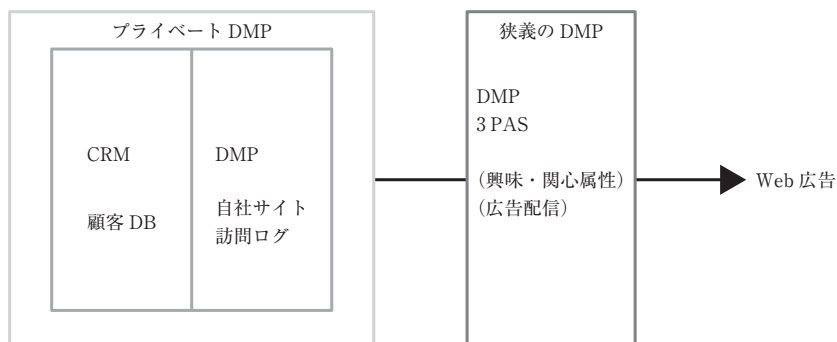


図4 D M P

このように狭義の DMP はすでにネット広告業界においては整備されている状態である。

広い意味での DMP は、顧客（見込み顧客含む）データベース＝CRM に狭義の DMP が結合したものだと考えて構わない。この CRM 部分と自社サイトの訪問履歴を結合したものをプライベート DMP と呼ぶことが多い（図4）。

前述のデータベース（狭義の DMP）に自社のサイトや自社会員情報が加わることになる。自社のサイトに訪問したときも、前述の DMP と同様に cookie によりブラウザは識別される。すなわち、狭義の DMP に自社サイトに来たことがある人、自社サイトに訪問し会員登録している人の情報がマッチングされることになる。

その結果、「自社会員で前にピザを注文したけど、最近1 ヶ月は注文していない人に 50% オフのクーポン付のバナー広告を今日の夕方を出したい」というような広告出稿が可能になる。このとき大事なのは、広告枠が空いていれば、自社のサイト以外のところにもバナー広告が出せるということである。自社のサイトに来なくても自社の会員や見込み客にアプローチ可能というのが大きい。

もう少し頑張ると、その他の情報と組み合

わせて「夕方、雨が降ったら当社のデリバリー可能な地域の人にピザの広告を出してください。今まで注文したことのない人には全商品 30% オフのクーポンを出し、注文したことがある人には新商品のクーポンを表示、サイトに来たけど注文していない人には 50% オフのクーポンを出す」というような芸当も可能になる。

もう一段広い意味での DMP は、マーケティングデータベースと呼んだほうがふさわしいものである。この段階となると外部のデータとも結合し、マーケティング全体をコントロールするデータベースとなる（図5）。

この領域の整備をするための仕組みに様々な企業が参入しており、激しい競争になっている。顧客データベースを管理し、いわゆる CRM を得意にしてきた IT ベンダーやネットとの連動を得意とするネット系企業、周辺データを保有する企業、DMP 構築をコンサルティングする企業など百花繚乱となっている。

注目されている DMP だが、課題も多い。1 つは、cookie の問題である。ネット広告のターゲティングは cookie に依存している部分が大きく、今後、cookie の利用範囲が限定的になった場合には肝心の Web 広告の表

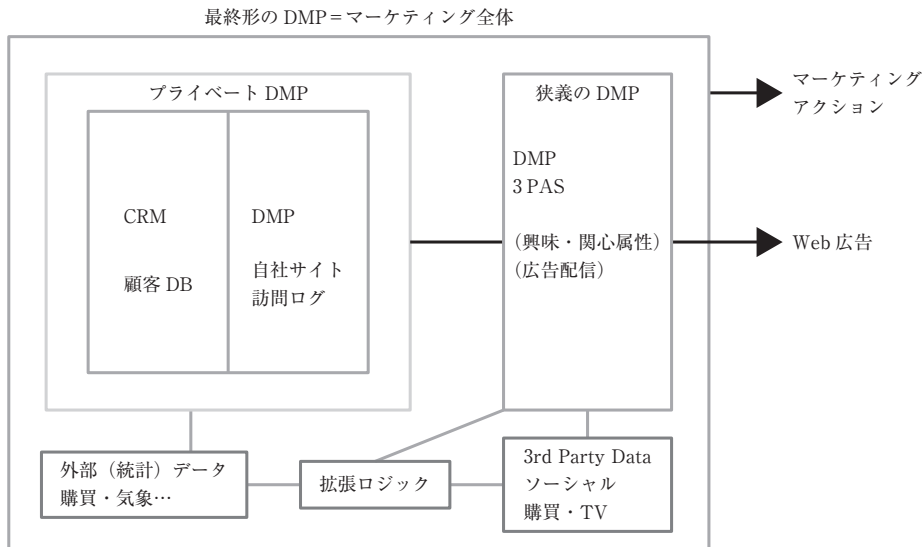


図5 マーケティングデータベースとしての DMP

示が難しくなるかもしれない。もう1つは、DMPによる広告の効率化の限界である。当たり前だが、ターゲティングをすればターゲットが小さくなる。ターゲットが小さくなりすぎると、広告効率がよくなったとしても多くの人に売りたいというマーケティング活動と矛盾してしまう場合が出てくる。さらに、モバイル広告の広告フォーマットの問題がある。現状スマートフォンへの広告はスマホユーザーにとって邪魔な広告としか受け止められていない。そのためか、スマートフォン広告市場は伸び悩んでいる。DMPが本当に機能するのは時間やエリアによる広告の出し分けができるスマートフォンである。スマートフォン広告が伸びればDMPの必要性も増すはずである。

🔗 データ活用の可能性と今後

現在、様々なデータが各企業に蓄積されている。今後、測定可能なログデータもますます増えていくと考えられる。官庁のデータのオープンデータ化も進むであろう。数千万規

模の人数のデータベースに各企業のもつデータがID（もしくはそれに準ずるもの）を基に結合される。付帯されるデータが増えるため、そのデータベースを解析するだけで様々なことがわかるようになる可能性はきわめて高い。そのデータベースはDMPのように直接的なマーケティングも行うプラットフォームになる。プライバシーの問題は残るものの、分析や施策上は魅力的なデータベースになることは間違いない。とはいえ、たとえば購買データについてすべてをカバーするとは考えにくいので、10万規模の消費者パネルデータがそれを補完するような役割となると思うし、ブランドに対する意識を測定するような調査や定性的なアプローチはなくなることはない。

データベース内の情報の解析で様々なことがわかるようになることで、「世界最先端IT国家創造宣言」に書かれているような「世界最高水準のIT利活用社会を実現」できるかもしれない。

注

- ・1 販売記録データ。POSはpoint of salesの頭文字をとったもの。
- ・2 サイトを訪問してきた端末に特定の情報を保存する仕組み。
- ・3 Data Management Platformの略。データを管理するプラットフォーム。
- ・4 あるサイトを訪れたブラウザの数。
- ・5 Demand-Side Platformの略。ネット広告配信の統合プラットフォーム。

文献

横山隆治・菅原健一・草野隆史, 2013, 『顧客を知るためのデータマネジメントプラットフォーム DMP 入門』インプレス R & D。

参照 URL

NTTドコモ「モバイル空間統計に関する情報」(https://www.nttdocomo.co.jp/corporate/disclosure/mobile_spatial_statistics/)。

日立「交通系ICカード分析情報提供サービス」(<http://www.hitachi.co.jp/products/it/bigdata/field/statica/>)。

JR東日本「Suicaに関するデータの社外への提供についてよくいただくお問い合わせ」(http://www.jreast.co.jp/pdf/suica_data.pdf, 2013年9月29日取得)。

ITメディアニュース「『Tポイントツールバー』公開中止」Web閲覧履歴を平文で収集 (<http://www.it-media.co.jp/news/articles/1208/20/news045.html>, 2013年9月29日取得)。

インターネットウォッチ「総務省がNTT-BPとコネクトフリーに行政指導、公衆無線LANで通信の秘密を侵害」(http://internet.watch.impress.co.jp/docs/news/20120404_523848.html, 2013年9月29日取得)。

閣議決定「世界最先端IT国家創造宣言」2013。 (<http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryou1.pdf>)。

「第1回パーソナルデータに関する検討会 議事次第」2013 (<http://www.kantei.go.jp/jp/singi/it2/pd/dai1/gijisidai.html>, 2013年9月29日取得)。

総務省「パーソナルデータの利用・流通に関する研究会」報告書 (http://www.soumu.go.jp/main_content/000231357.pdf, 2013年9月29日取得)。

