

ビッグデータが社会を変える

第4の経営資源をどう活かすか

アルタミラの洞窟壁画から始まり、象形文字、印刷機、コンピュータと文明は世界の現実をデータとして記録し、伝達、加工することにより発展してきた。

近年、膨大なデジタルデータがコンピュータ・ネットワーク上で生成され、流通、蓄積されている。

こうした、ビッグデータをどう活かすか、企業では、ビッグデータを活用して商品やサービスの付加価値を創造する取り組みが始まっている。



この夏の参議院選挙では与党が76議席を獲得し圧勝したが、ビッグデータが選挙結果を正確に予測していた。検索エンジンや音楽動画の配信などインターネット関連の情報サービスを提供しているヤフー・ジャパンではビッグデータを基にした選挙結果の予測を発表していたが、政党ごとの獲得議席数は、与野党別でみると与党76議席、野党45議席と完全に一致し、政党別の予測も9割がた一致した。ヤフー・ジャパンのビッグデータレポートチームでは、昨年12月の衆議院議員選挙で、インターネットの「ヤフー・検索」の政党名の検索数と比例区の得票数、議席の獲得数に高い相関関係があることを見出した。また、ツイッターやフェイスブックなどのソーシャルメディアの投稿数は小選挙区の得票数に高い相関性があることがわかった。

通常、選挙の予測はアンケート調査や選挙区ごとの情勢分析によって行われているが、ビッグデータを基に驚くほど高い精度の予測結果が出されたわけである。

一方、米国では大統領選挙でビッグデータが大きな役割を果たした。オバマ大統領は2008年の選挙戦ではソーシャルメディアを

活用して効果的にメッセージを伝えたことが勝利につながったが、2012年の選挙ではビッグデータを活用した緻密な選挙手法が勝利につながったと伝えられている。伝えられるところによると、オバマ陣営では支持者のデータベースやインターネット上のデータを活用して政治的な傾向や投票行動を割り出した。そして、「どのような人にとどのようなメッセージで投票を呼びかければよいか」シミュレーションを繰り返し、効果的な選挙運動を展開したという。

ビッグデータの活用の巧拙が政治にも影響を及ぼすようになってきた。そこで、ビッグデータとは何か、ビッグデータがどのように社会やビジネスを変えるのか少し考えてみよう。

「ビッグデータとは何か」

ビッグデータとは文字どおり巨大なデータという意味であるが、我々を取り巻くデータはこのところ爆発的に増加している。人類は先史時代のアルタミラの洞窟壁画から始まり、グーテンベルクの印刷機、1950年代のコンピュータの出現など現実の世界をデータ化し記録し続けてきた。文明は世界の現実をデータとして記録し、

伝達、加工することにより発展してきたといえるだろう。そして、コンピュータからインターネットの時代に入りデータは幾何級数的に増加している。

【情報の爆発】

コンピュータが扱う情報の最小単位は1ビットで「0」か「1」か2とおりの状態を表している。ビットを8つまとめた単位が1バイトで、アルファベットなどの半角

1文字のデータに相当する。データ量の増大に合わせて情報量を示す単位も拡大していった。およそアルファベット1千文字分の1キロバイトから1・4メガバイトのフロッピーディスク、4・7ギガバイトのDVD、「ギガ」から「テラ」「ペタ」「エクサ」「ゼタ」そして「ヨタ」へとデータ量を示す単位は際限なく拡大を続けてきた(表)。

表 情報量の単位

単位	情報量	情報量
ヨタバイト	YB	10 ²⁴ バイト 今のところ、最大の単位
ゼタバイト	ZB	10 ²¹ バイト 0.7ゼタバイト:地球上のすべての海岸の砂の数
エクサバイト	EB	10 ¹⁸ バイト 5エクサバイト:今まで人間によって話されたすべての言葉
ペタバイト	PB	10 ¹⁵ バイト 200ペタバイト:すべての印刷物
テラバイト	TB	10 ¹² バイト 10テラバイト:米国連邦議会図書館の収蔵図書
ギガバイト	GB	10 ⁹ バイト 4.7ギガバイト: DVD-R
メガバイト	MB	10 ⁶ バイト 1.4メガバイト: 3.5インチフロッピーディスク
キロバイト	KB	10 ³ バイト 0.8キロバイト: 原稿用紙1枚
バイト	B	1バイト 8ビット、アルファベット1文字
ビット	b	1ビット "0"か"1"か、情報の最小単位

昨年12月に公表されたIDCのデジタル・ユニバースに関する最新の調査(EMC協賛)によると2012年のデジタル・ユニバース(全世界で生成あるいは複製されたデジタル情報の総量)は2・8ゼタバイト、これまで人間によって話されたすべての言葉は5エクサバイトといわれているから、2・8ゼタバイトはその560倍にあたる。また、デジタル・ユニバースは2020年まで2年ごとに倍増し2020年には40ゼタバイトに達すると予想している。40ゼタバイトとは、地球上のすべての海岸にある砂の量の合計(0・7ゼタバイト)の57倍に相当するデータ量である。

国内においても、情報通信白書(平成25年版)によると、ビッグデータ流通量(国内の企業が受信

するデジタルデータ)は2005年の0・4エクサバイトから12年には2・2エクサバイトへ7年間で5・5倍に拡大、蓄積量は9・7エクサバイトに達している。

【ビッグデータの源泉】

では、なぜ最近データが爆発的に増加しているのだろうか。まず、データがデジタル化されたことにより、データの処理と保管、複写や伝達が容易になった。加えて、それに要するコストも加速度的に低下している。コンピュータの計算能力は「半導体の集積密度は18か月〜24か月で倍増する」というムーアの法則が続いており、通信については「通信網の性能は1年で2倍に向上する」というビル・ジョイの法則が働いている。また、データを収納する装置も大容量化と低価格化が急速に進んでいることはメモリーカードを思い浮かべれば容易に想像できるだろう。

そして、最近データを急増させているのが、スマートフォンとセ

ンサ、そして企業業務の高度化である。まず、一つ目がスマートフォン

の増加である。多くの人がスマートフォンを持ち歩き、情報の検索やメールのやりとりをし、チャットとした情報や写真をツイッターなどのソーシャルメディアに投

稿するようになってきた。情報通信白書(平成25年版)によると、昨年の日本のスマートフォン

の普及率は95%と、ほぼ全

ての国民がスマートフォンを所有している。スマートフォンは、製造業のサプライチェーンマネジメント(部品調達から生産、販売の流れを企業横断的に管理する手法)など企業が事業を行ううえで発生する業務関連のデータである。この分野も業務の自動化と高度化を受けて着実に拡大している。

朝起きて、スマートフォンでメールをチェックし、ツイッターで「今日は調子がイマイチだ」とつぶやいたこと、通勤の乗降駅、利用日時や時間帯も交通系のICカードで記録される。昼食のサンドイッチをコンビニで買ったことはPOSデータで記録され、サンド

イッチにするかおにぎりにするか迷った様子もカメラに撮られていた。また、営業で回った車の経路もGPSで追跡される。現実には、我々自身の生活がライフログ（行動の記録）としてデータ化されるようになってきた。

【データの多様化と分析の高度化】

そして、今起きているのは単なるデータの増加ではない。増加しているデータの多くがブログやソーシャルメディアに書き込まれた文字データや画像、センサで検知されたデータなどこれまで分析の対象とされていなかった多種、多様なデータである。こうしたデータはこれまでコンピュータで扱われてきた経理データや販売データなどの構造化データに対して非構造化データといわれている。ビッグデータが注目された背景には、非構造化データを扱うソフトウェアと分析手法が充実してきたこともある。特に、巨大なデータから有用な情報を抽出する技術体系はデータマイニングといわれており、データの中から特徴的なパターンをみつける「相関ルール学習」、相関関係の傾向を数式化する「回帰分析」、似たものどうしを集める「クラス分析」などがある。私の考えでは、ビックデータの

意義は単にデータの量が増えたことではなく、今まで分析の対象でなかった非構造化データを分析し、新しい知見を見出し商品やサービスの付加価値を創造するところにある。

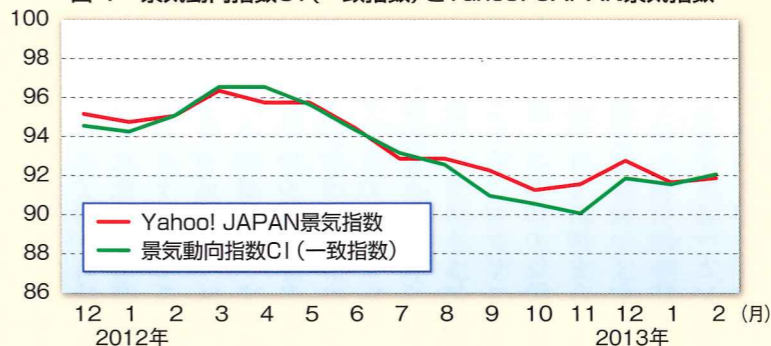
【ビッグデータをどう使うか】

では、ビッグデータからどのような新しい知見や価値が生まれているのか、そのいくつかを見てみよう。

【景気の今をみる】

景気判断の基になるのは経済活動を数値化した経済指標であるが、公表されるまでには1か月から2か月かかる。実際、我々は数か月前の情報を基に景気の現状を推測しているわけである。しかし、足元の景気を知る方法はないのだろうか。ヤフー・ジャパンでは、「ヤフー・検索」で使われたキーワードをもとに景気の「今」を示す「ヤフー・ジャパン景気指数」を開発し公表している。「ヤフー・ジャパン景気指数」の基準になっているのは内閣府が毎月公表している景気動向指数C-Iの一致指数である。景気動向指数C-Iとは、鉱工業生産指数や商業販売額な

図1 景気動向指数C-I(一致指数)とYahoo! JAPAN景気指数



出所: Yahoo! JAPAN HP「景気の「今」を把握することは可能か?」(2013.4.16)をもとに作成
URL <http://event.yahoo.co.jp/bigdata/keiki/>

図2 消費者物価指数と東大日次物価指数(月次指数)前年同月比



(注) 消費者物価指数は東大日次物価指数のカバーする商品だけを取り出したもの
出所: 東大日次物価指数プロジェクト(月次指数)により作成
URL http://www.cmdlab.co.jp/price_u-tokyo/

ど景気の動きに敏感な経済指標を合成した総合的な景気指標で、そのうち景気の動きに一致するのが一致指数である。一方、「ヤフー・ジャパン景気指数」は、「ヤフー・検索」のキーワード75億種類のなかから分析に適したキーワードを抽出し景気動向指数C-Iとの相関関係を調べ、相関関係の高いキーワードの動きをもとに計算されている(図1)。経済の因果関係から結論を導く

のではなく、巨大なデータ群からパターンを読み取るのがビッグデータの分析手法である。また、物価指数については東京大学が「東大日次物価指数」を全国のスーパー300店舗のPOSデータをもとに公表している。日次で集計され3日後までに指数を作成、月次指数も早い段階で公表されている(図2)。総務省の公表している消費者物価指数は、全国の調査員が調査対象となる店舗

に出向き聞き取り調査するという方法をとっており、公表されるのは翌月下旬となる。「東大日次物価指数」はスーパーで販売されている商品が対象で、すべての商品やサービスを網羅しているわけではないが、2%の物価上昇率をターゲットとした「異次元の金融緩和」のなかでリアルタイムの物価指標として政策効果を示してくれる。

【カゼを予防する】

カゼなどの感染症に対しても、感染の実態がリアルタイムで把握できれば予防におおいに役立つだろう。米国では、インターネットの検索データを使ってインフルエンザの流行を予測した話が伝えられているが、日本ではエスエス製薬の「カゼミル+(プラス)」がツイッターのデータを使ってカゼの流行をリアルタイムで教えてくれる。カゼミル+では、東京大学の「知の構造化センター」の先端技術によりツイッターの会話を解析しカゼをひいている可能性の高いユーザーをエリアごとに集計して増減傾向を教えてくれる。しかし、ツイッターからカゼをひいている可能性のあるユーザーを選び出すことは、そう簡単なことではない。「昨日からのどが痛くてイガイガする」人と「カラオケでの

どが痛い」人をシステムの区分する必要があるのである。

また、カゼに関するツイートの増減と天候の変化を分析すると、関連性の高い天気因子があることもわかった。この結果をもとに、週間天気予報と組み合わせることによりカゼツイートの増減を予測して、注意を喚起してくれる。

【顧客の声を聞き取る】

ツイッターなどのソーシャルメディアにはカゼ情報だけでなく、製品やサービスの改善につながる顧客の声も隠れている。自社のデータだけではなく、広くソーシャルメディアの投稿データを分析し経営に活かす手法はソーシャルリスニングといわれている。NTTデータ経営研究所とgooliサーチの「企業によるソーシャルリスニングに関する動向調査」(2013年8月)によると、「現状の顧客の声を把握」するためにソーシャルリスニングを「積極的に実施」している企業は14.2%、「それなりに実施」している企業は28.9%と合わせると43.1%にのぼる。また、ソーシャルリスニングのメリットとしては、「タイムリーに顧客の声を把握」できることが、「大いにそう思う」と「思う思う」を合わせて71.8%、「お

客様窓口に寄せられないようなサイレントカスタマーの声を把握」できることが69.8%と、およそ7割の企業が顧客の声を聞き取れることをメリットと考えている。また、コンビニの店頭においても、ビッグデータを使って顧客の声を聞き取る試みが始まっている。従来のPOSシステムでは「誰が何をいつ買ったか」はわかるが、顧客が本当は何を求めたのかまではわからない。来年の2月パナソニックとローソンでは実験店を設け、店内に設置した高度なカメラで顧客の性別や年齢を判断、立ち寄った陳列棚や滞留時間など店内での行動を分析し、顧客の心をつかむマーケティングにつなげるという。

【第4の経営資源をどう活かすか】

このように、企業においてもビッグデータの活用は広がりをみせはじめているが、情報通信白書(平成25年版)によると、ビッグデータの活用により創出される潜在的な経済効果は7兆7千億円以上にのぼる。最も経済効果が大きいのが製造業の「リモート監視によるメンテナンス人件費の効率化」で4兆7千億円、2番目は「交通状

況をリアルタイムで提供できるプロップ情報による渋滞回避、それに伴う燃費改善」の1兆1千億円であった。流通業では、ビックデータにより消費者ニーズを詳細にとらえ、売り逃しや売り残しを減らすことなどにより全体で1兆1千億円。農業においても、気象、土壌などの環境データ、植物工場の制御、流通・販売情報の活用などによる効果があげられている。この推計は、対象とした業種のヒアリング調査により行われたもので、産業界全体の潜在的なポテンシャルはさらに大きいものと考えられている。

データは21世紀における石油だといわれている。まず、データはデータマイニング (mining・採掘)といわれるように採掘する必要がある。そして、精製、加工することによりさまざまな価値が生まれるからである。情報は、ヒト、モノ、カネに続く第4の経営資源として指摘されてから久しい。データを経営戦略から日常業務にいたるまで経営にどう活かすか。これによって、企業のパフォーマンスに格差が出てくる時代に入ってきたのではないだろうか。

(株)京都総合経済研究所

調査部長 森 秀人