

## ビッグデータの潮流と課題に関する一考察

## A Consideration of Tide and Problem About Big Data in Business

坂巻 英一

Yoshikazu Sakamaki

## 1. はじめに

近年、メディア等でビッグデータという言葉を目にする機会が多い。企業活動の仮定に於いて蓄積されたテラバイト、ペタバイトといった市販されている分析ツールでは分析することが困難な巨大なデータを指す。近年、こうしたデータを効率的に処理する技術が次第に確立されつつあり、企業の経営効率改善に役立てようとする試みも盛んになってきている。巨大なデータを処理するに当たり必要となるコンピュータの性能向上と共に、RDBを中心としたデータベースだけではなく、並列分散処理を実現するソフトウェアとして Hadoop が無料で使用できるようになったことが、ビッグデータの活用を後押ししていると言えよう。

ところが、企業に蓄積された大規模データを分析し、そこから得られた知見を企業の経営効率改善に役立てようという動きは 1990 年代の前半には既に始まっていたのである。有名な事例として POS データを分析しそこから得られた知見を基に店舗のレイアウトを変更したところ、売上改善を実現することができた、という報告が挙げられる。こうしてみると、ビッグデータやデータサイエンスに関連する技術は今に始まったものではないことに気付かされる。本稿ではビジネスにおけるビッグデータ利用の歴史について概観すると共に、ビッグデータ活用の今後の展望について概観する。

## 2. ビールと紙おむつの事例

第 1 節で既述したように、ビッグデータを分析した結果得られた知見をビジネスで活用した事例は 1990 年代の前半にまで遡る。有名な事例として、アメリカのあるスーパーマーケットチェーンが行った POS データの分析事例が有名であろう。これは店舗に蓄積された POS データを分析した結果、金曜日の夜になると多くの顧客が缶ビールと紙おむつを一緒に購入する傾向があることが分かった、というものである。この店舗では分析結果を基に、缶ビール売り場の横に紙おむつを並べたところ、缶ビールと紙おむつの売上が共に上がった、というのだ。分析結果は 1992 年 12 月に発行された Wall Street Journal で紹介され、ビッグデータ分析の初期の頃の代表的な成功事例として後に広く知られるようになった。

こうした分析手法は一般に併買分析と呼ばれており、現在、流通小売業界において最もよく行われる分析手法のひとつである。POS データを分析することで得られた知見を基に、同時に購入される傾向のある商品を見つけ出し、陳列棚の棚割りを決定する際に利用されている。

併買分析はリアル店舗だけではなく、ネットショップでも利用しているサイトは多い。Amazon.com 等で商品を購入した際に、他の商品を勧められたことはないだろうか。これはレコメンデーションエンジンと呼ばれており、併買分析の結果を活用した典型的な事例であると言える。

## 3. データサイエンティストの仕事

ビッグデータを分析する技術者は、最近、データサイエンティストと呼ばれており、近い将来、人材不足に陥る可能性が高いと言われている。ところが、データサイエンスという言葉自体、既に様々な場所で使われているにも拘らず、データサイエンティストが行う仕事については、未だに定義があいまいなままである。

最近の求人広告を見ていると、データサイエンティストと呼ばれる職種は大きく分けて 2 種類に分かれているように思われる。

一つ目がビッグデータの管理等を効率的に行う、SE タイプの職種である。Hadoop のコーディングや Javascript を利用した帳票を出力させるための WEB 画面の構築もここに含まれる。そして二つ目がアナリストタイプの職種である。営業部門やマーケティング部門と密に連絡をとりながら、データベースから抽出されたデータを解析し、得られた知見を営業やマーケティングの現場へフィードバックする必要な役割を果たすことになる。データサイエンティスト募集、と書かれた求人広告の中でよく目にする SAS や SPSS, R 等の BI ツールのオペレーションスキルを持った人材がここに該当する。

ところで、SAS や SPSS を使用した分析業務はいつ頃からあるのだろうか。少なくとも、90 年代の中ごろには、企業が保有する顧客データを分析し、得られた知識を経営改善に生かそうという取り組みが存在していたように記憶している。こうして見ても、データサイエンティストと呼ばれる職業が今に始まったものではないことに気付かされるのではなかろうか。

## 4. ビッグデータの活用を進めるために

「ビッグデータ」や「データサイエンス」という言葉が消滅しても、企業に蓄積された大規模データを活用しよう、という動きは引き続き残ると考えられる。なぜならば、大規模データのビジネスでの活用は今に始まったものではなく、20 年近くも前から存在するからである。それでは、ビッグデータの活用を成功に導くためにはどうしたらよいのだろうか。図 1 は社内にデータサイエンス専門の部隊を有するか否か、社内に独自の大規模データを保有しているか否か、を基に企業を 4 つのパターンに分類した結果である。

独自の大規模データを保有しており社内にデータサイエンティスト集団を抱えている企業は、既に、ビッグデ

ータの活用が相当進んでいる企業である。こうした企業は社内にも十分なノウハウが蓄積されており、ビッグデータを日常的に活用した経営改善に取り組んでいると考えられる。問題は社内に独自の大規模データを保有しているにも関わらず、専門の部隊を有しない企業である。データサイエンスに関する技術やノウハウは一朝一夕では蓄積することができないため、外部のコンサルティング会社等に分析業務を依頼しなければならないことが多い。

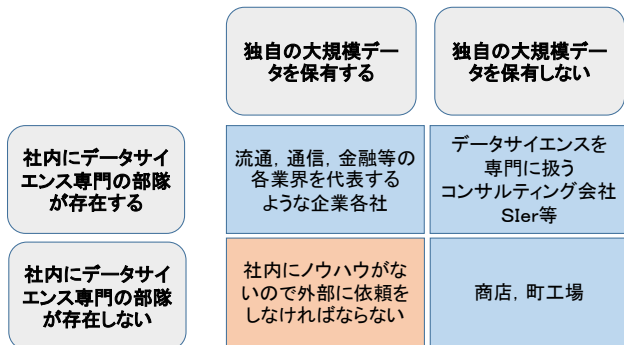


図 1 保有データの規模ごとに見たデータサイエンスへの取り組み

ところが、ここで是非次のことを記憶に留めておいて欲しい。データサイエンスに関する技術やノウハウは技術者やアナリストの中に蓄積されるものだ。そのため、業務を外部に丸投げするようなプロジェクトの進め方をしたのでは、時間が経っても社内には全くノウハウが蓄積されない、という事態が発生しかねない。こうした方法でプロジェクトを進めていくと、気付いた時には社内のシステムがブラックボックス化し社内だけではデータサイエンスに関するプロジェクトを回せなくなっていることもあり得るのだ。やむを得ずシステム開発や分析業務を外部に委託する場合には、何を最終ゴールとするのか、プロジェクトの目的を経営陣が明確にしておく必要があると言える。

ビッグデータの活用を社内で成功させるためには、3つのSが重要な役割を果たすことに気付かされる。それは、System(Database, Hadoop 等), Statistics(統計解析)そしてStrategy(経営戦略)である。これら3つの要素を全て兼ね備えた人材は地球上探し回っても数えるほどしかない。つまり、これらの要素のうちのいずれかに秀でた人材を集め、データサイエンスチームを社内に構築した上でプロジェクトを推進してゆくことが求められるのである。

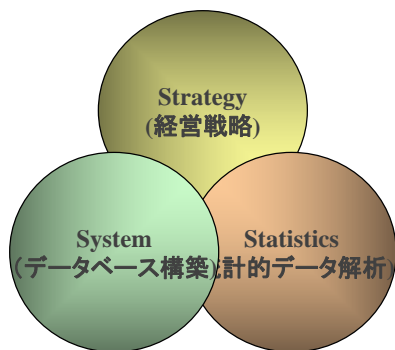


図 2 データサイエンスに必要な要素

どんなに高価な情報システムや統計ソフトを導入しても、戦略が揺らいでいてはよい成果は得られない。また、どんなに良い戦略を構築しても、数値的な裏付けが十分でなければよい成果は得られない。これら3つのSが融合した時に初めて、ビッグデータの活用は成功するのである。これを聞いて、初期の頃のデータマイニングを思い出す人がいるのではなかろうか。データマイニングがはやり始めた90年代の後半、データマイニングプロジェクトを成功に導くために必要な要素として、全く同じことが言われていたことに気付かされよう。

### 5. パズワードはいずれ消滅する

企業に蓄積された大規模データを分析し、企業の経営改善に活用する取り組みが現在様々な企業に於いて行われている。一方で、第3節で既述した通り労働市場に於いてデータサイエンティストと呼ばれる人材はそれほど多くないのが現状であり、今後、人材が不足することが予想される。一方で、起業に蓄積された大規模データを分析しデータの背後に潜む規則性や法則性を見つけ出す技術はかつて「データマイニング」と呼ばれており、20年以上前から存在するのである。「データサイエンティスト」や「ビッグデータ」という言葉は大手のメディアが騒ぎ立てたがゆえに、現在、パズワードになっているのではなかろうか。

これらの言葉の定義が未だに明確ではないことや、パズワードは必ず消滅する運命をたどってきた歴史を考えると、近い将来、「データサイエンティスト」や「ビッグデータ」という言葉は消滅するのではないかと考えられる。ところが、大規模データのビジネスでの活用は20年以上も前から世の中に存在していたのである。「データマイニング」から「ビジネスインテリジェンス」そして「ビッグデータ」「データサイエンス」、時間の流れと共に呼び方が変わっただけで、やっていることは20年以上前から何ら変わらないのだ。

### 6. まとめ

ビッグデータという言葉を目にするようになってかなりの時間が経ったような気がする。

分析ツールは複雑な計算をアイコン操作や短いプログラムで実行することを可能にする。つまり、与えられた問題の答えを見つけ出すのは非常に得意である。ところが、分析ツールは何が問題なのか、までは考えてくれないのである。それを考えるのは社内の人間に他ならない。分析ツールは複雑な計算を行うための道具でしかなく、決して分析ツールを導入しただけでは、結果を出すことはできないのだ。SASシステムを始めとした分析ツールに加え、それをどのように活用したらよいか、を考える立場の人材がビッグデータを活用する上で必要なのだということを肝に銘じておく必要があると言える。

以上