

データ分析読本

(株) NTT データ数理システム
データマイニング部

目次

1. はじめに.....	1
2. データ分析全般について	2
2.1. データ分析がもてはやされる理由	2
2.2. どうしてデータから知見が出せるのか.....	3
2.3. データ分析手法の発展	4
2.4. データ分析で可能なこと.....	8
2.5. データ分析は「動機」まで決めてくれない.....	9
2.6. データ分析に必要となるもの.....	10
2.7. データ分析が適用可能な範囲（事例）	13
3. 具体的な分析の流れ（小売業のケース）	14
3.1. 集計で現状を把握する ～ 最初のステップ.....	16
3.2. 集計レベルの分析を超えて ～ 様々な手法	18
3.3. さらに施策に打って出る	21
3.4. 業務の中でのデータ分析.....	23
4. NTTデータ数理システムの役割.....	24

1. はじめに

この読本では、データをお持ちの方がどのようにデータ分析をしたらよいかについて簡単にご説明をします。その前に、データ分析を取り巻く環境について概観します。

近年、ビジネスの様々な現場でデータが蓄積され、その勢いは日々ますます加速しています。また、蓄積されるデータの量もさることながら、データの種類も多種多様となっています。このような状況を反映した**ビッグデータ**という言葉は広く浸透しています。またデータを分析するスキルを持った**データサイエンティスト**に対するニーズが高まっています。

日々蓄積されるデータには、ビジネスの改善のヒントになる知見が大量に埋もれています。小売店の販売データであれば、思わぬ商品が売れ行きを伸ばしていることがわかり、売上の向上につなげることができるでしょうし、営業マンのデータであれば、営業成績の低い営業マンと高い営業マンの動き方を比較して、どういう要素が成績を高くしているかを見出すことができるかもしれません。

とはいえ、これらのデータの溜まるスピードは加速するばかりで、このデータを業務に生かすための知見に変える時間的な余裕がないという現実もあります。**ビッグデータ**という用語は、データから様々な知見を活かして、業務プロセスを改善することのできるという期待が込められていますが、なかなかそこまで到達できないまま、データをため続けるケースが多いのが現実です。

この読本では、データに基づいた分析とはどのようなものかをご説明し、具体例として実店舗を持つ小売店のケースを考え、小売店で蓄積されるデータに対して分析をおこなう際の分析の流れを基に、データ分析が業務の中でどのように利用されるものかをご説明します。

2. データ分析全般について

2.1. データ分析がもてはやされる理由

データ分析がもてはやされる理由としては、ハードウェアの劇的な進歩により、様々な状況においてデータがどんどん蓄積されているという背景があります。これにより、意図したかどうかにかかわらず、様々な業種の現場の担当者が、結果として大量のデータを手にすることとなりました。

その実情を反映して「たまったデータを活用すれば、データから何か結論が出てくるかもしれない！」という期待感もどんどん増してきています。

データ分析が近年重要視される理由は、期待感だけではありません。データから何かしらの知見を導き出すには、それなりのデータ量が必要とされます。従来はデータが少ないために、十分な裏付けが取れないことからデータ分析から意味のある結果を導き出せないこともありました。データ量が十分にそろふことで、以前は手にすることができなかった結論を導き出すことが可能になってきています。

但し、データ量が増えると、今度は分析に必要となる計算量が膨大となり、分析は徐々に困難になってきます。近年ではデータ量が増大するのと同じ勢いで、データに対する処理技術が進歩してきています。技術の進歩には次の三つの側面があります。

① ハードウェアの進歩

まず、大量のデータを蓄積するために必要となるハードディスク、また分析の際に必要な CPU・メモリが比較的安価に購入できるようになってきたという点が挙げられます。これにより、意味のある結論を出すためにかかる時間が短くて済むようになってきました。

② 分析手法を支える理論の進歩

次に、分析のための理論の進歩が目覚ましい点が挙げられます。分析を実施するためには、分析手法が動作するハードウェアのほかに、分析プログラムを構成するための各種の理論が必要です。使われる理論としては、統計的技術、データマイニング、機械学習などがあります（詳細は6ページ「データ分析手法の発展」に記載しました）これらの理論に関する研究が、近年爆発的な勢いで発展しています。

③ 便利なソフトウェアの登場

分析手法の理論的バックグラウンドもさることながら、それらの技術を組み合わせて手軽にデータ分析を実施するためのソフトウェアも徐々に充実してきていることが挙げられます。数理システムの **Visual Mining Studio** を中心としたソフトウェアによって、理論に基づいた分析をすることが可能です。

これらの「データ量の増大」という環境要因に対して、「データ分析に必要なとされる各種技術の進歩」をベースとしてデータ分析ができる技術的素地が固まってきた、というのが **ビッグデータ** を中心としたデータ分析の現在の立ち位置と言えるでしょう。

2.2. どうしてデータから知見が出せるのか

では、そもそもどうしてデータ分析から業務改善につながる知見が得られるのでしょうか？ 過去の事実だけの積み重ねに見えるデータから何かしらの知見が得られるというのは、よくよく考えると非常に不思議な気がします。

これを考えるヒントとして、日常業務におけるよくあるシチュエーションを考えてみましょう。職場によっては長年の経験をもったベテラン職員が存在します。ベテラン職員は、多くの経験を積み重ねてきたことにより、次のようなスキルを持っていることが多く、現場で大変重宝されます。

- 何かが起きた場合に、その事象について原因を察知する能力
- この先どうなりそうかという予測

さて、ベテランはどうしてそのようなことができるのでしょうか。これは難しい問題ですが、ベテランが「過去の経験に基づいて」判断を下しているというのが一つの大きなポイントです。経験あるベテランは、過去の経験に基づいて以下のようなことが実施できるスキルを持っていると考えられます：

- 大量のデータの中から今考えている事象について関係のある部分についてのみのパターンを見出す能力がある。
- 過去のデータにおける事象間の関連性パターンを導き出して将来に適用する能力がある。

データ分析の話に戻しましょう。データ分析手法のよりどころとなる理論は、過去に起きた事象が蓄積された「データ」から、ベテラン職員が経験や頭で考えて行っている（と考えられる）**パターン抽出**を、人の頭でなくコンピュータ上（正確にはコンピュータ上で動作するソフトウェア上）で実施するという考え方に基づいています。

データ分析は、人が経験に基づいて知見を見出すのと同じように、データに基づいて知見を見出す機構を利用するものです。過去のデータに対して何らかの仮定を置き、その仮定をあてはめて将来に適用することで、過去の状況の把握のみならず、将来の適用までできる、というのがデータ分析の基礎となる考え方なのです。

2.3. データ分析手法の発展

データ分析が一躍脚光を浴びている理由として、データ分析技術が近年著しく発展していることが挙げられます。ここでそれらの技術を概観してみましよう。＜続きは小冊子をお読みください＞

小冊子ご希望の方は、vmstudio-info@msi.co.jp まで下記を明記の上、ご連絡ください。

- ・ 件名「データ分析読本発送希望」
- ・ ご所属先名（企業名・大学名）
- ・ 部署名
- ・ お名前
- ・ 郵送先ご住所
- ・ お電話番号
- ・ 質問/ご相談事項（ございましたらお気軽にお書きください）

NTT DATA

株式会社NTTデータ数理システム

Visual Mining Studio 担当

〒160-0016 東京都新宿区信濃町 35 番地 信濃町煉瓦館 1 階

TEL : 03-3358-1701

FAX : 03-3358-1809

本書の一部あるいは全部において(株)NTT データ数理システム

からの文書による許諾を得ずに、いかなる方法においても無断で

複写・複製することを禁止します。