

ビジネス・アナリティクスの最新動向（4）

中川慶一郎（なかがわ けいいちろう）
 株式会社NTTデータ数理システム
 生田目 崇（なまため たかし）
 中央大学理工学部

1. はじめに

これまでの連載を通じて、ビジネス・アナリティクス（BA: Business Analytics）を分析シーンと分析目的という観点から整理したうえで、ビッグデータ時代における最新の動向を分析技術、IT基盤の両面から紹介した。

また、ビッグデータは以下の3つのパラダイム・シフトをもたらし、これを受けて、ビジネス・アナリティクス（BA: Business Analytics）も新たなチャレンジの時期を迎えていることを説明した。

- (1) 目をつぶってきた事実を目を向ける。
- (2) “どうにもならない時間間隔”を埋める。
- (3) 複雑な状況で合理的に判断する。

ここで、(1)、(2)については、実践的な取り組みがなされている。これに対して、(3)はGoogleの自動運転[4]のように、いまだ試験研究的な色彩が強く、分析技術から見ても十分な道具立てがそろっているとは言い難い。

BAの観点から見ると(3)のポイントは機械学習など人工知能の技術を有機的に結びつけることによって、意思決定品質を高める、あるいは自動化することであり、今後のBAの大きな潮流となることが想定される。

連載最終回である今回は、意思決定品質を高めるための枠組みを図1のように考え、これに従って今後の発展が期待される分析手法をいくつか紹介する。また、最後にBAの方向性について筆者らの考えをまとめる。

2. 進化する機械学習

現在でも多くの現場では、発見型、WHAT-IF型の分析シーンで、従来の多変量解析やデータ・マイ

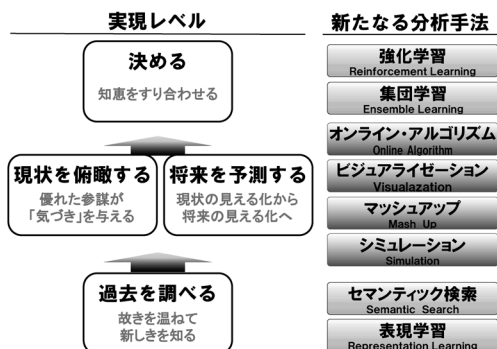


図1 意思決定品質を高める枠組み

ニングにもとづき、(1) データ間の相関関係や因果関係を事前に定義し、(2) 統計的な見地からその有意性について確かめる、といったことが広く行われている。

このとき「複雑な状況下で合理的に判断する」ためには、複雑に入り組んでいる因果関係を分析・評価することも必要となる。

また、プロアクティブ型の分析シーンでは、まだ起こっていない将来の事象に対して先回りして、気の利いたサービスを提供する。こうしたサービスを実現するには、市場や消費者の時々刻々の変化に対応した分析が必要となる。

図1はこういった状況での意思決定品質を高めるにあたり、実現レベルと対応する新たな分析手法を整理したものである。俯瞰すると第2回目まで整理した分析目的についてはこれまでと変わるところはないが、それに対応する分析手法については、近年さまざまなものが、人工知能研究の一分野にある機械学習を中心に登場しつつある。

これまで、増加し続けるデータに対する分析手法として、数々のデータ・マイニングの手法も提案されてきた。しかし、有効性に関する大きな枠組みと

してはこれらの手法も多変量解析の延長上にあるといえよう。

ここでこの分析手法に共通する思想としては、例えば回帰分析のように、

- (1) あらかじめ教師データとなる変数（目的変数）を定める。
- (2) その差を説明する候補となる変数（説明変数）を定める。
- (3) 特定の関数（回帰分析の場合は線形関数）について、統計的見地から影響のある変数を選択する。
- (4) その影響の度合いを定量的に評価する。

といったことである。

もちろん主成分分析や因子分析、クラスタ分析、自己組織化マップのように、教師データが存在しない分析手法もあるが、これらは変数間の関係の強さからサンプルの特徴を集約しているに過ぎない。

こういった問題点に対して、機械学習では、人間のもつ学習能力を計算機上で再現・代替する分析手法が研究されており、「複雑な状況下で合理的に判断する」ための道具立てとして発展が期待される。

計算機が人間の代わりに状況に応じて判断するというと、一昔前ならばSFのように聞こえるかもしれないが、計算機の性能向上に伴い、近年、機械学習分野においてさまざまな成果を見せ始めている。

2.1 強化学習

1 番目は、強化学習（Reinforcement Learning）である。強化学習は、逐次的な分析の中で、得られる利得を最大にするように試行錯誤しながらモデルを修正していく学習手法である。つまり、分析の置かれている環境に対して適合するようにモデルを変化させていく。

強化学習は、入力されるデータが逐次的に与えられたもとの、ある環境における最適応答について、試行錯誤を繰り返しながらその環境に適応させていく学習制御である。

回帰分析のような教師付きデータの分析とは異なり、入力に対する正しい出力が存在しない。その代わりに、意思決定のたびに受け取る報酬をもとに学習を行う。ただし、報酬にはノイズや遅れが含まれる。そのため、行動を実行した直後の報酬を見るだけでは、学習主体はその行動が正しかったかどうか

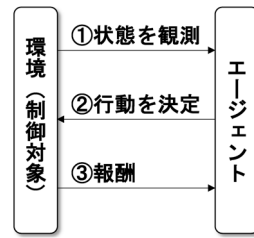


図2 強化学習

を判断できないという困難を伴う。例えばある市場で行動する主体が、環境の変化に応じてその都度最適に行動を決定していくような、最適な意思決定を行うための学習手法である。

強化学習の枠組は図2のようになり、ここで用いられる代表的手法としては、マルコフ決定過程やニューラル・ネットワークが挙げられる。

マルコフ決定過程は、時間ごとに状態の集合と状態間の推移確率が与えられた問題において、将来に得られる価値が最大となるような最適行動を逐次的に求める意思決定手法である。マルコフ決定過程は動的計画法の一種であるので、将来時点から現在にさかのぼって各時点各状態での期待価値を計算し、その期待価値が最大になるような方策を各状態について求める。

ニューラル・ネットワークは非線形の複雑な因果関係をデータに合わせて逐次的にパラメータを収束させていく手法である。したがって、計算過程での状況を加味しながら、各段階での決定を変化させていくことが可能となる。したがって、ニューラル・ネットワーク自身は教師付きデータの予測手法であり、強化学習そのものではない。しかし、強化学習では構造の複雑な問題を扱うことが多いため、その最適反応のルールも複雑となる。そのような問題における学習器としてニューラル・ネットワークは広く使われており、今後強化学習が注目されるに伴い、同時に注目される方法といってもよいであろう。複雑な関数関係がある、もしくはスパースなデータでも精度のよい評価ができるという点では強力な分析手法である。

2.2 オンライン・アルゴリズム

2 番目は、オンライン・アルゴリズム（Online Algorithm）である。オンライン・アルゴリズムは、

逐次的に発生するデータについて、分析を繰り返して行い、その都度パラメータを更新することで常に、データに対して適合するモデルを得る手法である [1].

第3回目で紹介した CEP (Complex Event Processing) も逐次発生するデータをリアルタイムに処理するという意味では思想は類似する。CEP では例えば「ある株式の価格が直近5分以内に3% 値上がりしたら保有する株式を売却する」といったように、あらかじめ設定された IF-THEN 型のルールに対して、発生したデータを判断し出力を決定する。したがって、CEP ではモデルそのものは事前に設定し、このモデルに沿った出力を得る。

これに対して、オンライン・アルゴリズムでは、発生するデータに対してモデルそのものを変化させていくことで、データによりフィットした出力を得ようとする手法である。従来の分析では、データを一度すべて DWH などに格納してからそのデータについてモデル分析を行っていたが、多頻度、大量データを分析するためには、こうしたデータ格納は現実的でない場合も多い。そこで、逐次的に学習を通じて最適なモデルに近づけようというようなアルゴリズムが注目されるようになった。

オンライン・アルゴリズムは 1980 年代に競合比解析を用いた評価によってスタートし、その後、情報工学分野で目覚ましい発達を遂げた。

近年、マーケティング・サイエンスの分野では多様な顧客の行動を説明するのに顧客ごとの個別のパラメータを推定するというモデルが多数提案されている。パラメータについてはベイズ推定を繰り返して行うことで推定値を収束させている。この手法では、顧客全体のパラメータに対して正規分布などのある分布を仮定し、各顧客に対するパラメータを初期値からベイズの法則を援用して更新する。このときにマルコフチェーンモンテカルロ (MCMC) が用いられるが、その収束計算は決して高速ではない。

これに対して近年では、粒子フィルタと呼ばれる手法も注目されつつある。粒子フィルタは分析の設計によっては MCMC よりも早く収束することも知られており、十分な量のデータが確保できるならば、高速に推定値が求められるのも一つの特徴である。

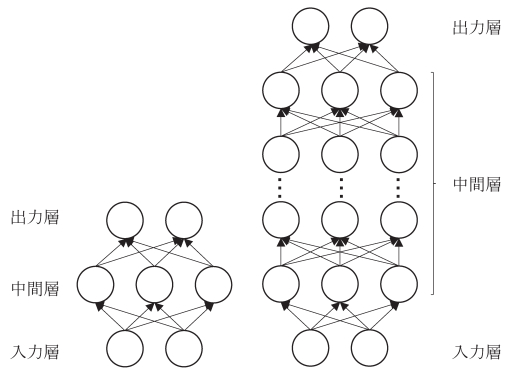


図3 浅いNNと深いNN

2.3 ディープ・ラーニング

3 番目は、ディープ・ラーニング (Deep Learning: 深化学習) である [2]. ディープ・ラーニングは形式的にはニューラル・ネットワークを多層化してより複雑な因果関係を分析しようという手法である。ニューラル・ネットワークは 1990 年代に非線形の因果関係を分析する手法として盛んに研究されたが、その後いったん下火の時代が続いた。しかし、人工知能分野において、ディープ・ラーニングとして再度脚光を浴びることになった。

ディープ・ラーニングが注目を浴びることになったのはコンピュータが自律的に判断することに成功したという報告である。特に Google が猫の画像を認識させることができたという記事は驚きをもって迎えられた [3].

通常のニューラル・ネットワークと大きく異なる点は、通常のニューラル・ネットワークが中間層を挟むものの入力層と出力層を合わせた 3 層構造を基本としたものなのに対して、ディープ・ラーニングは接点数が場合によっては 10 億以上持つという深いニューラル・ネットワークである点が挙げられる (図3)。また、識別のための教師データを事前に与えずに学習をさせた点も特徴として挙げられよう。

特に後者については、猫とラベリングした画像や、猫がどのようなものかを事前に情報を与えることなく、数多くの画像を解析しながらシステム自らが猫を猫以外のものと識別した点にある。こうした学習は自己教示学習 (Self-taught Learning) と呼ばれる。現状では最先端の人工知能というべきディープ・ラーニングは、自律的にルールを生成しながら

判断する特長がある。今後生成されるデータが増加する中で、教師データを与えるような、これまでの分析を大きく超える新たな方向性として期待できる。

近年では企業が主催して、顧客行動履歴をもとにその後の行動（例えば購買や閲覧）を予測するといったデータ分析のコンペティションも開催されるようになった [5]。ここでは、判別ルールとして従来使われていたロジスティック回帰分析や決定木などに加えて、ディープ・ラーニングを使ったという報告も出現している。

現状の研究では、画像や文字の認識が主流であるが、上記のようにBAの分野でもその本格的な登場が期待される。また、以外にも、音声認識や文章のセンチメント識別能力向上などにも応用されよう。

実際 Python などの科学技術言語上でディープ・ラーニングを実装するためのライブラリなども公開されており、大規模なモデルでなければ、PCを利用して試行することも可能である。

このとき行われるBAはユーザ行動を先読みしてサービス提供をすることを志向しており、プロアクティブ型と呼ぶことができる。プロアクティブ型BAは、ビッグデータへの関心が広がるとともに急速に発展と遂げており、今後も業務やサービスを革新的に変えていく根幹技術となっていくことが予想される。

3. 今後のBA

これまで、さまざまな角度からビッグデータ時代のBAについて論じてきた。ビッグデータやBAがITベンダによる流行言葉だという指摘は、ある側面では正しいかもしれない。しかし、なぜこれほどまでに経営者をひきつけるのか、いささか突飛かもしれないが、ホモサピエンスの軌跡を例に考えてみよう。

ホモサピエンスはおよそ10万年前にアフリカで出現して、その後地上を移動しながら現在のように地球全体に広がってきた(図4)。その変遷を知るためにこれまで何をしてきたかという、実際にその場にいたことを示す遺跡の発掘、その地層がいつのものかを放射性同位体によって特定するといったいわゆる伝統的な考古学の方法で行われてきた。

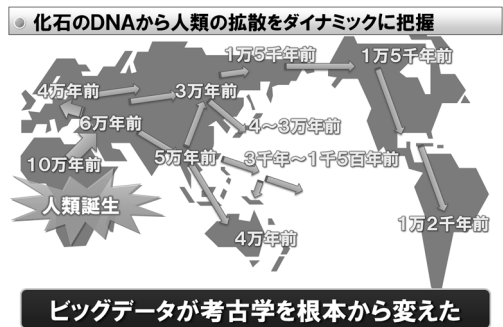


図4 DNA考古学

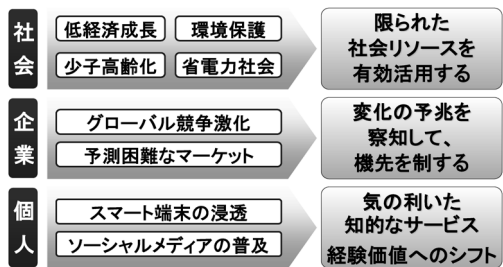


図5 BAの目指す方向性

近年これらのブレイクスルーとなったのがDNA考古学である。DNA考古学は、発掘された骨のかけらなどそれ単体ではどのような生物かが特定できないような微小な発掘物から、そこに含まれるDNAを解析することで、どの時代のどのような生物がいたのかについて特定しようという新たな考古学である。

DNAも身体の進化とともに徐々に変化すると考えられることから、人類に関するさまざまな試料のDNAの差分を分析することで、どの試料とどの試料が近い関係にあるかといったことを分析することができるようになった。これらを追うことで、人類がどの時代にどのようにそれぞれの地域に進出していったかといった、それまでの考古学では知りえなかった事実が明らかにされた。

生物学の世界の技術であり、ビッグデータの権化ともいふべきDNA解析を考古学の世界に適用することで、これまでの考古学の“やり方”を大きく変えてしまったことは注目に値する。

例えばGoogleが自動運転自動車の試作をするというように、一見すると全く関係がない異業種の人たちがその分野に入り込んでそれまででは全く考え

られなかったような新たなことをやり始めるという
ようなことが多くの分野でこれからも起こりえよ
う。

「考古学」というものを自らの業界に置き換えた
場合、DNA解析のような技術を持てば、競合を一
網打尽にできると思う反面、異分野の人たちが全く
新しい技術を引き下げて来たら、完膚なきまでにや
られてしまう。そういった期待と恐怖が一体となっ
た焦りが、今の経営者をBAに駆り立てているもの
と思われる。

最後に、BAが目指す方向性について見てみよう
(図5)。社会においては、少子高齢化の流れはこれ
からも進むとともに、経済成長を偏重しない社会づ
くりが必要となる。これまでのトレンドとは異なる
新たな社会づくりのためには、限られた社会的なリ
ソースをいかに有効活用するかの視点が必要となる。

企業においては上述のようにこれまでは考えつか
なかったような新たな市場参入や、予測不能なマー
ケットの動きに対する迅速な対応が必要になる。さ
らにミクロに見ると個人においてもスマートフォン
などに代表されるように粒度の細かい多様なデータ
が取得できるようになり、そうした行動やデータに
対する気の利いた知的なサービスを提供していくこ
とが望まれる。

このように、さまざまな対象において、これまで
とは全く異なるサービスや分析が求められるように
なる。データ・サイエンティストの分野において
も、従来の統計やORの専門家にとどまらない多彩
な人材が入り込んできている。こうした人材による
新たなアプローチが今後さまざまな分析のブレーク

スルーになる日が来るかもしれない。

参考文献

- [1] 徳山 豪 オンラインアルゴリズムとストリーム
アルゴリズム, 共立出版 (2014)
- [2] 八木康史, 斎藤英雄 (編)『コンピュータビジョン
最先端ガイド6』アドコム・メディア (2014)
- [3] Using large-scale brain simulations for machine
learning and A.I., Google official blog (2012 年 6
月 12 日) [http://googleblog.blogspot.jp/2012/06/
using-large-scale-brain-simulations-for.html](http://googleblog.blogspot.jp/2012/06/using-large-scale-brain-simulations-for.html)
- [4] A. Fisher, "Inside Google's Quest To Popularize
Self-Driving Cars," Popular Science, Retrieved 18
September (2013) [http://www.popsci.com/cars/article/
2013-09/google-self-driving-car](http://www.popsci.com/cars/article/2013-09/google-self-driving-car)
- [5] (株) オプト「ネット広告のクリック予測コンペ
ティション」<https://datasciencelab.jp/compel/6>

略歴

中川 慶一郎 (なかがわ けいいちろう)

2000年早稲田大学大学院理工研究科博士課程満期退
学。博士(工学)。2012年株式会社NTT データ数理シ
ステム取締役。マーケティング・エンジニアリングお
よびビジネス・アナリティクスに関する研究開発、コ
ンサルティングに従事。

生田目 崇 (なまため たかし)

1999年東京理科大学大学院工学研究科博士課程修了。
博士(工学)。2013年中央大学理工学部経営システム工
学科教授。マーケティング・サイエンスおよび経営科
学に関する研究に従事。