

ビジネス・アナリティクスの最新動向 (2)

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

前回は、ビジネス・アナリティクス (BA: Business Analytics) を巡る近年の話題について説明した。第2回目の今回は分析手法の側面から見た近年の潮流について解説する。

はじめに、実務においてデータ分析を実施する「分析シーン」と「分析目的」という2つの側面からビジネス・アナリティクスを類型化する。次にこれらの分析の中で使われる手法について整理したうえで、今後の方向性についてまとめる。

2. 分析シーンから見た類型

ビジネスの場で実際にデータ分析を行うシーンから BA を図1に示すように4つのパターンに類型化する。以下では各分析シーンとそこでの BA について概説する。

2.1 集計分析型

1番目は、蓄積されたデータをさまざまな角度から集計・分析するシーンであり、ここで行われる BA を集計分析型と呼ぶこととする。集計分析型

BA の類型	分析シーン
集計分析型	蓄積されたデータを様々な角度から集計・分析、見える化 例: 計画業務, 経営管理, 需給調整
発見型	蓄積された膨大なデータから、隠れた関係性や規則性を発見 例: 顧客のスコアリング, リスクの計量化
WHAT-IF型	新しい業務方式をデザインし、業務方式の変更に伴う効果を事前に試算 例: SCM最適化, BPOに伴うリソース最適化
プロアクティブ型	ユーザー行動を理解し、一歩先回りして、気の利いたサービスや機能を提供 例: レコメンド, 不正検出

図1 分析シーンから見た4つの類型

BA では、個々人が表計算ソフトなどを用いて簡単な分析を行うレベルから、組織的に「見える化」を行うレベルまで幅広く対象範囲となる。

特に「見える化」は従来型のビジネス・インテリジェンス (BI: Business Intelligence) で行われてきた取り組みであり、すでに多くの企業で導入されている。BIシステムでは、データ・ウェアハウス、データマートの整備から始まり、集計・分析にあたっては、多次元データ分析ツールである OLAP (Online Analytical Processing) が利用される。

集計分析型 BA は人手による要因分析を基本としており、情報分析活用の第一歩といえる。

2.2 発見型

2番目は、蓄積された膨大なデータの中から隠れた関係性や規則性を見つけ出そうとするシーンである。例えば、生産データから歩留りに影響を与える要因を特定したり、購買データから顧客の特徴を抽出したりするとき、人手を前提とする集計分析では対応が困難な状況がこれに該当する。

このようなタイプの BA はデータをかき回して埋もれた知識を見つけ出すという意味で、発見型と呼ぶことができる。一般に発見型 BA では、多変量解析やデータ・マイニングといったより高度な統計技術の活用が有効である。

発見型 BA には、顧客のスコアリングや信用リスクの計量化といった、従来「勘・経験・度胸」になりがちであった業務を、客観的な「根拠に基づく意思決定」に変える役割があり、金融、通信、流通における CRM (Customer Relationship Management) など、特定業務ではすでに定着している。

2.3 WHAT-IF 型

3番目は、新しい業務方式や業務プロセスをデザ

インして、その中にデータ分析を組み込み、さらにこれに伴う効果を試算するシーンである。こういった状況は業務改革を実現しようとする際にしばしば見受けられる。

例えば、米国流通最大手のウォルマートは、メーカーと協調して商品計画、需要予測、在庫補充を行い、欠品防止と在庫削減を同時に実現する CPFR (Collaborative Planning Forecasting and Replenishment) という取り組みを行った。

このような業務改革を進める場合、① CPFR の例ではどのような情報を共有して需要を予測し、どのような発注方式にするのか、②いくつか考えられる業務方式の中でどの方式が効果的か、といったことが問題となる。

上記のうち、①を実現するための個別要素技術(需要予測や在庫・発注計算)として BA が有効であることは、異論の余地はない。これに対して、②において従来とは異なる業務方式を導入したときの効果を確認するためには、現場での実証実験が有効な手段となる。しかし、現場を巻き込んだ大規模な実験となると失敗が許されない場合も少なくない。そのようなとき、業務上の要件、制約をリアルに描写するシミュレーション技術が大きな力を発揮する。

一般に、シミュレーションや感度分析によって前提条件の差異が結果に与える影響を比較する分析は「WHAT-IF 分析」と呼ばれる。新しい業務方式や業務プロセスを試行錯誤的に検討し、その導入がもたらす結果を比較することで最適な業務を考える分析を、ここでは WHAT-IF 型と呼ぶことにする。

業務改革には常に大きなリスクが伴う。したがって、そのインパクトを事前に推計して分析することで、経営的なリスクをできるだけ軽減する WHAT-IF 型 BA は、業務改革を進めるうえで必要不可欠な分析である。

2.4 プロアクティブ型

近年、ユーザの行動や文脈を理解し、一步先回りする知的なサービス・機能が現れている。

例えば、EC (Electronic Commerce) サイトへ行くと、次から次へとお勧め商品が表示される。顧客が商品を検索したときに、その人が興味を持ちそうな商品も併せて表示されることで、今まで知らな

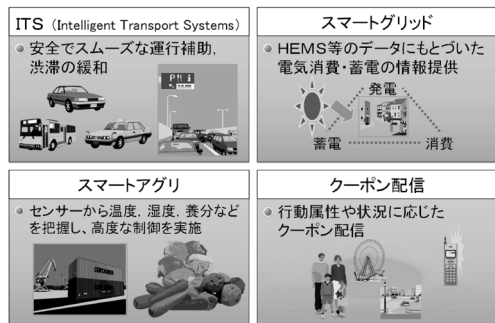


図2 プロアクティブ型 BI

かった商品にも出会うことになる。このようなレコメンド・サービスも一種の知的サービスと言うことができる。

また、高度道路交通システム (ITS: Intelligent Transport Systems) における安全でスムーズな運転の補助など、スマート・ビジネスと呼ばれるものには、多かれ少なかれこのような機能が組み込まれており、今後さらに広がっていくことが想定される。

4 番目は、このようなサービス・機能を提供するために、情報分析を活用して、裏で動くロジックを構築するシーンである (図2)。

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. 分析目的から見た類型

分析に関するニーズは現在多様な業界で高まっており、これまでビジネスデータの活用が未開拓であった分野にも拡大しつつある。その一方で、分析という観点から見ると、業界が違っても根本的な課題意識、データ構造、使われる分析手法、導かれる分析結果の形式が類似したケースが多々あることも事実である。その点を踏まえて、ここでは、どのような分析をするのかという分析目的から見た類型を考えてみる。

例えば、通信会社は顧客の離反を防ぐためにさま

表1 分析目的から見た9つの類型

BAの類型		分析の目的
予兆発見型		行動変化や状態変化の監視による予兆の発見
異常検出型	不正検出型	不正・異常の定義と合致／類似する行動・状態の検出
	外れ値検出型	標準的な行動・状態の定義と逸脱の検出
予測・制御型	収益シミュレーション型	予測モデルの構築と意思決定変数の変化による増収効果の試算
	リスク・シミュレーション型	業務のモデル化と不確実要素によるリスクの試算
	リスク・ヘッジ型	業務のモデル化とリスク分散手法を用いたリスク低減策の提示
	最適化型	業務のモデル化と最適化手法を用いた意思決定策の提示
ターゲティング型		見込み顧客など重点アプローチすべきターゲットの抽出
リスク・スコアリング型		顧客の解約や企業の倒産などのリスクの試算
評価・要因分析型		さまざまな対象の比較評価と改善要因の特定
マーチャンダイジング型		さまざまな視点での売れ筋ランクの作成と品揃えの決定
コンテキスト・アウェアネス型		行動履歴・嗜好の分析から一歩先回りしたサービスの提示
プロセス・トレース型		成長・発展プロセスの抽出と促進・阻害の特定

さまざまな解約防止策を実施している。そこでは、顧客の利用履歴やプロフィールなどのデータを分析して解約しそうな度合いを数値化し、その値が高い顧客から順にアプローチをすることになる。

一方、保健医療の分野では健康保険組合が医療費の支出増を抑制するために、生活習慣病の発症リスクの高い保険者に対してカウンセリングや禁煙指導・食事指導などを実施している。ここでも、保険者ごとに発症リスクを数値化して、その値が高い保険者から順にアプローチすることになる。

これらの事例が目指すところは、個々の対象のリスクを数値化することによって、損失を防止することであり、ここでは「リスク・スコアリング型」としてまとめることができる。

こうした事例を積み上げて、整理したものが表1である。これらの類型の詳細については、例えば[3]を参照いただきたいが、いくつかの手法については簡単に紹介したい。

ターゲティング型は、ダイレクトメールにヒットしやすそうな対象を絞り込むなど、ターゲットを絞り込むための分析であり、発見型BAの中心的な分析手法といえる。異常検出型は、不正パターン、あるいは外れ値による検知を通じては異常に対するアラートを通知する分析であり、プロアクティブ型BAの範疇に属する。

また、コンテキスト・アウェアネス型は行動履歴などから購買行動の文脈を理解したうえで、レコメ

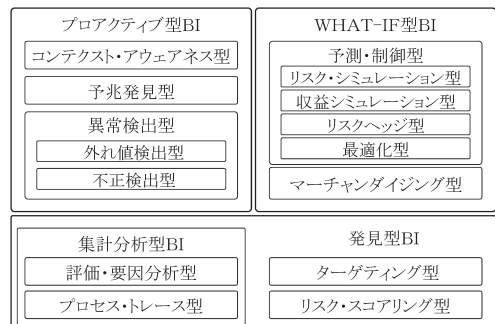


図3 類型間関係

ンドを実施する分析であり、同様にプロアクティブ型の分析といえる。

ここで、前述した分析シーンから見た4つの類型と分析目的から見た9つの類型の関係を整理すると図3のようになる。

4. 近年の分析トレンド

ここまでは「分析シーン」と「分析目的」という2つの側面からBAの類型化した。実際にBAを行う際は、例えばリスク・スコアリング型と一言で言っても場合によってはロジスティック回帰分析、別の場合ではCox比例ハザード・モデルを用いるというように、分析目的や状況に合わせて、適切な分析手法を選択し、適用していくことが必要とされる。

そこで、ここからは前述の類型化を踏まえうえで、既存の分析手法の特徴を整理し、今後の方向性を考える。

4.1 従来からの分析手法

データ分析といえば従来から統計分析、多変量解析の各手法が広く用いられてきた。表2にその代表的な分析手法をまとめる。共通する特徴としては、変数間の共分散もしくは相関係数をもとに、因果関係や相関関係を分析しようというものである。したがって、全体のデータを統計量という形に要約している。

多変量解析の各手法は、統計学を背景にしており、モデル構造の理論的背景が明確であることや分析結果の評価もしやすいことから発見型BA、WHAT-IF型BAにおいて広く使われてきた。その反面、集計された統計量をもとにしていることや線形モデルを中心としていることから、プロアクティブ型BAのように、

- さまざまな要因が複雑に絡み合い、その構造を理解するよりも、結果としてのアラートやシグナルが重視される（予兆発見型、異常検出型）。
- ワン・トゥ・ワン・マーケティングのようにサンプル間の異質性を考慮しなければならない（コンテキスト・アウェアネス型）。

といった状況では、これらの手法では不十分な場合もある。こうした問題に対して、機械学習やパター

ン解析などの手法が脚光を浴びるようになった。

データ・マイニングはこうした分析手法の総称ともいえるが、従来の多変量解析と比較したときの大きな特徴は、複雑な関係のモデル化や、先にデータを要約せずにそのまま評価するといった、データの複雑化、大規模化に対応している点である。

また、もう1つの特徴は、従来の分析は仮説検証に主眼が置かれてきたのに対して、仮説発見型プロセスを考慮されている点である。これは、分析者の経験などからは導き出せないような思いもよらない関係を、分析を通して「発見」という試行錯誤のプロセスが重要視されてきたためであろう。

表3に主なデータ・マイニング手法についてまとめる。このうち、決定木分析はもっとも広く利用されているデータ・マイニング手法といわれる [5]。その理由としては、結果がIf-then形式のルールベースで得られることで理解がしやすいことと、のちに紹介するランダムフォレストのような精度向上の拡張が目されたことも一因であろう。

ニューラルネットワークは複雑な非線形の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り、いずれも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ような手法である。また、現在注目が集まりつつある強化学習の学習器としてもニューラルネットワークは再度注目されている。

集団学習は単一の分析手法では全体の精度が望めない場合に、データを分割もしくはリサンプリングし、分析手法を組み合わせることで精度向上を図ろうという手法である。決定木分析を拡張したブースティング、バギング、ランダムフォレストが代表的な手法として挙げられる。

例えばランダムフォレスト [1] は、データセッ

表2 代表的な多変量解析手法

分析の方向性	主な分析手法
因果関係	回帰分析
	一般化線形モデル
	非線形回帰分析
	数量化理論I類
	Cox 比例ハザード・モデル
クラス判別	判別分析
	ロジスティック回帰分析
	数量化理論II類
構造解析	主成分分析
	因子分析
	コレスポンデンス分析
	多次元尺度構成法
	数量化理論III類

表3 代表的なデータ・マイニング手法

分析の方向性	主な分析手法
因果関係	ニューラルネットワーク
クラス判別	決定木分析
	サポートベクターマシン
	ナイーブベイズ
構造解析	非階層型クラスタ分析
	自己組織化マップ
パターン解析	相関ルール

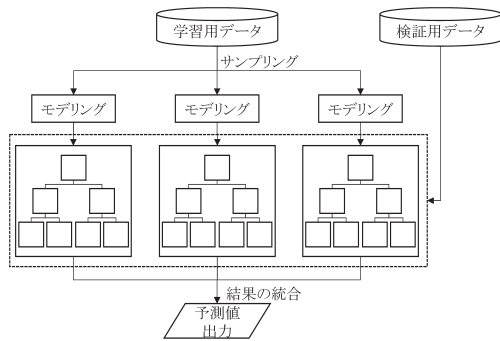


図4 ランダムフォレスト

トを分割してそれぞれについて決定木分析を行い、その結果を統合して、出力を決定するというものである（図4）。単一の決定木によるルールではデータのノイズや偶然により、結果に偏りが生じる場合も想定されるが、こうした集団学習では多数決原理が働くため、通常の決定木よりも安定した解が得られるという実験結果が報告されている。

これらいずれの方法も、変数の間に複雑な関係が想定される場合や、範囲によってルールが異なるような場合に対して、より高い精度で解を導きたいという状況において注目されている手法である。したがって、現在の消費者の購買行動のように、多様性や時間推移が想定されるような場合に効果を発揮することが期待される。

これらの分析手法はそれぞれ提唱されてから10年程度が経過しているが、注目されてきたのは最近のことである。これは、ビッグデータの時代といわれるようになってから、分析対象となるデータが取得されるようになったことと、そうしたデータの活用が社会全体で問われるようになったことと無関係ではないであろう。

4.2 非構造化データを利用した分析

もう1つ、近年の分析で注目されるようになってきたのは、文章などの非定型データを活用しようとする動きである。これは、SNSやミニブログのような生活情報や口コミ情報がまとめてデータとして取得できるようになった、またほかのサービスのIDとの連携ができるようになったといった情報環境の進化と密接に関係している。

文章解析の技術は、従来からこうした情報はテキ

スト・マイニングと総称され研究されてきたが、近年ではビジネス領域で本格活用しようという動きも加速している。そこでの分析手法は従来の標準的なテキスト・マイニングの範囲を超えていないとはいえ、前回述べたビッグデータによるパラダイム・シフトが大きく影響しているという過言ではない。

例えば、今日ではインターネットが商品販売サイトやプロモーション・メディアなどのマーケティング活動のためのプラットフォームとして広く使われているのはご承知のとおりである、近年では、SNSにおける口コミやつぶやきのような非定型データをベースに、消費者の購買意向や商品に対する評価を知ろうとする試みもなされている [6]。

一般に、新製品は発売前の製品発表会など、事前に情報が周知される。ただしその後、実際の流通までにはある程度の期間があり、その間に新商品の情報はニュースや広告などを通じて市場に広がる。消費者はそういった情報から、製品への期待や思いをSNSやミニブログで発信する。

企業は新商品の販売に際して、発売前につぶやきのデータを収集し、消費者の関心領域やつぶやきの頻度を分析することで、どのようなタイプの消費者がどのような発言をしているか把握できる。

さらに「ある商品が特定の顧客セグメントの関心領域とうまくマッチすれば購買が期待できる」というように、製品に対する事前の評価と顧客像との関係を結びつけることも可能である。

このように発売前に消費者の評価を知ることが、前回述べた「“どうにもならない時間間隔”を埋める」ことになる。また、こうしたデータの分析を通して、インターネット上でのプロモーション活動の最適化を図ったり、発売後の販売戦略に活かしたりすることも可能であろう。

こうした分析をするためには、大規模なつぶやきデータから、リアルタイムに関連するデータを抽出するIT基盤と、個々のつぶやきにセンチメントを付与する分析技術が必要となる。特にIT基盤の進展については近年目覚ましいものがあり、BAのブレークスルーとなった原動力である。BAにおけるIT基盤については次回に詳説する。

参考文献

- [1] L. Breiman, "Random Forests," *Machine Learning*, Vol. 45, pp. 5–32 (2001)
- [2] 金森敬文, 竹之内高志, 村田 昇『パターン認識』共立出版 (2009)
- [3] (株) NTT データ技術開発本部ビジネスインテリジェンス推進センタ『BI 革命』NTT 出版 (2009)
- [4] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, MIT Press (1998)
- [5] X. Wu and V. Kumar (eds.) *The Top Ten Algorithms in Data Mining*, Chapman & Hall (2009)
- [6] Twitter データを用いた金融マーケット向け「Twitter センチメント指標」を開発, NTT データ・プレスリリース (2014年3月17日)
http://www.nttdata.com/jp/ja/news/services_info/

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.