

AI・OT・セキュリティ分野で活用が進む 大規模データ利活用の研究 —Data is Money—

アブストラクト

1. 研究背景

近年、AI、OT(Operational Technology)、セキュリティなど、幅広い分野でデータの重要性が認識され、その利活用が望まれている。また、IT技術の進展により、大規模データ利活用のための環境も整いつつある。例えば、データ蓄積コストが大幅に低減し、Rなどのデータ可視化/分析ツールも充実し、Hadoopなどの並列分散処理やクラウドサービスも使える状況にある。一方でデータ利活用がうまく進んでいない状況がある。分科会参加メンバー企業12社で調査したところ、全ての企業でデータ利活用に着手できていない、または有効に利活用できていないという現実が明らかとなった。

我々は、自らの経験から従来のデータ利活用における課題を洗い出し、課題を解決する新しい手法を提案し、身近なテーマで実証することを目指して分科会活動を実施した。

2. 研究の進め方

第1ステップは、文献調査と併せて、当分科会メンバー企業に於けるデータ利活用の状況を分析し、現状整理と、現実としてデータ利活用を阻害する課題の洗い出しを実施した。見えてきた従来のデータ利活用は、目的達成に向けて仮説を設定し、その仮説をデータ分析により確認する手法であり、着手のハードルが高い、必要なデータが揃わない、失敗時の手戻りが多い、など、多くの課題が見つかった。

第2ステップは、抽出した課題を解決するための新たなデータ利活用手法の検討である。そこで、データ利活用の原点に戻って発想を転換し、「データを中心に捉え、使えるデータの価値を最大限に抽出して利活用する」ことを考え、最近言われ始めている「データ駆動経営」からヒントを得た「データ駆動によるデータ利活用」という手法を考案した。対して従来手法を「目的駆動によるデータ利活用」として再定義を行い、両者に共通のプロセスとして、(1)データ収集、(2)データ蓄積、(3)データ抽出/加工、(4)データ可視化/分析の4フェーズに分けて、メリット、デメリットを整理した。

第3ステップは、考案した手法の実証である。自分たちの身近なテーマについて、「データ駆動」および「目的駆動」の手法を用いて実際にデータ利活用を実践し、整理した両手法の特徴を確認しつつ、新たな気づきを得て肉付けを行い、「データ駆動」の使い方の整理および、有効性の検証を実施した。

3. 研究成果

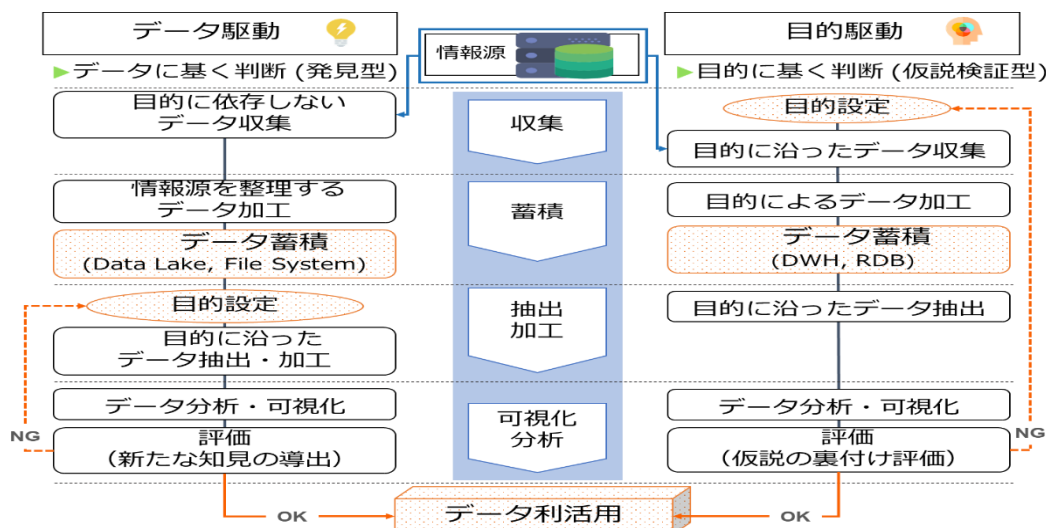


図1. 「データ駆動型」と「目的駆動型」の比較

我々の研究成果は、新たなデータ利活用手法として「データ駆動」を提案し、従来手法である「目的駆動」との対比を通じて、「データ駆動」の特徴および使い方を明確にし、更に実証を通じて有効性を確認した上で「データ駆動」によるデータ利活用を提言するものである。図1に基づき、我々が整理した両方式を(1)(2)に示し、実証結果を(3)に示す

(1) 「目的駆動型」データ利活用

総務省の指針[1]で示されているように、従来は「まず目的を決める」ことから始まり、何事も「目的に即して判断する」手法が用いられる。具体的には、最初に目的を決め、目的達成のための仮説を設定し、仮説立証に必要なデータの収集を行う。収集したデータは分析目的に適した形に加工して蓄積する。データの蓄積には、一般的にRDBMSによるデータウェアハウスを利用する。蓄積したデータを分析することで最初に設定した仮説を立証する。仮説が立証できなかった場合、必要により目的や仮説を変更して、開始点に戻って再度データ収集からやり直す必要がある。この手法の主な課題は以下のとおりである。

- ×費用対効果を明確にしないと着手しにくい（分析しないと効果が分からない現実とのギャップ）
- ×分析目的や仮説が変わった場合、データ収集からやり直す必要がある。
- ×分析目的に適合するデータが収集できないと先に進めず、非構造データの扱いも難しい。
- ×データ蓄積時点で振り落とされているデータ項目もあり、当初想定した目的以外では使いにくい
- ×想定以上の知見（新たな気づき）が得られる可能性は低い。（仮説が実証されるか否か）

(2) 「データ駆動型」データ利活用

データ中心に考え、何事も「データに基づいて判断する」手法である。具体的には、目的は後で決める。データは多目的に使うことを前提に収集し、情報源の性質を損なわない形で整理して蓄積する。データの蓄積には様々なデータ形式に対応するファイルシステムによるデータレイクを使用する。蓄積したデータの利活用時に、初めて目的を決める。ここでは仮説の設定は必須ではなく、目的はデータ分析結果により柔軟に変更できる。次に、目的に必要なデータをデータレイクから抽出し、分析に必要な形式に加工して可視化/分析を行う。評価では、分析結果を検討して、データ利活用（施策への採用）、目的の変更、データの再選定、分析手法の変更などを再帰的に実施して、有効な知見を探る。

(3) 実証

身近なテーマについて、使えるデータの範囲で実際に分析を行った結果、分科会メンバー12名全員が「データ駆動」で分析に着手することができ、うち8名については分析を通じて新たな知見を得ることができた。この中には、「目的駆動」で失敗した事例を「データ駆動」で成功できたケースも含まれる。このことから、我々が提唱する「データ駆動型」データ利活用の有効性が確認できた。具体的な実証内容については本論文および事例集に記載している。実証した「データ駆動」の主な利点を以下に示す。

- ◎データ収集/蓄積が目的に依存しないので、貴重なデータを多目的に共有しながら活用できる。
- ◎小さな目的、少ないデータから着手し（スモールスタート）、分析しながら段階的に充実させる（スケールアウト）ことができ、「とにかくやってみよう」が容易になり、そのリスクも最小化できる。
- ◎蓄積データの範囲で、その価値を最大限に引出して利活用でき、データ不足による破綻が生じにくい。
- ◎分析結果が悪い場合でも、蓄積データの抽出（選定）から再試行できるので、手戻りが少ない。
- ◎データ収集/蓄積を現場のSEが担当できるので、データサイエンティストの負荷が軽減される。

4. 総括

我々の提言は、「データ駆動型」のデータ利活用を新たな手法として採用することである。「失われたデータは戻らない」「データの蓄積には膨大な時間とコストを要する」を教訓として、「Data is Money」という考え方で、価値のあるデータを多目的に使えるように蓄積することから始めることが重要である。留意点として、データ収集/蓄積は具体的な利活用目的が後で決まるため、全社方針としてデータ利活用の推進を掲げて（トップダウンで）データ収集/蓄積に着手することを推奨する。その上で、蓄積されたデータを組み合わせ、AIなど、今後生まれる新たな分析手段も使いながらデータサイエンティストが試行錯誤する中で、様々な目的で有効にデータを利活用して頂きたい。

本研究では、企業内データの蓄積と利活用を中心に議論したが、利活用範囲を1企業に閉じる必要はなく、将来は企業間でデータを流通させることで、より有効にデータを利活用できる可能性がある。ここでは、企業内に蓄積したデータの価値が更に高まることが期待できる。