

5 データの活用

データ活用における課題と対策

「Things Cloud®」を使うと、多様なデバイス／センサーからの情報取得・可視化をスムーズに実現できる。情報収集・可視化でビジネス効果が見込めるケースも多いが、取得データのストリーム処理・分析・機械学習による予測・制御が、さらなるIoT活用の鍵となる。この章では、そうしたデータ活用を行う際に抑えておくべき課題と対策についてご紹介する。

「Garbage in, garbage out」

「AI/機械学習は、とにかく集めたデータを渡せば、人間が気づかないような素晴らしい予測をしてくれる」と考えがちだ。しかし、残念ながら今の機械学習にはそこまでの能力や精度はない。人が計画を立て、適切なデータを収集し、適切なアルゴリズムを適用する必要がある。

やみくもに集めたデータをそのまま機械学習に渡しただけでは、精度の低い、あるいは不適切な予測モデルしかできない。「Garbage in, garbage out」という言葉は、ゴミデータを渡すと、ゴミのような結果しか返らないことを表している。

データ分析において、機械学習にかける工程は全体の僅かな部分だ。図1は「Hidden Technical Debt in Machine Learning Systems (機械学習システムの隠れた技術負債)」(D.Sculleyら NIPS 2015)の引用である。図中央の黒い四角「ML

Code」が、いわゆる機械学習にデータを処理させる部分である。しかし、ほかにさまざまな考慮が必要だ。

中でも、収集したデータを機械学習に理解できる形に整形・抽出する「前処理工程」と、前処・アルゴリズム適用・結果活用までを行う、「データ利活用のためのインフラ整備」が重要である。

前処理工程では膨大なデータを探索的に分析し、機械学習に有効と思われるデータを抽出、整形する。ここで統計学、機械学習に加え、データベースなどのエンジニアリングスキルを要する。

データを利活用するには、インフラとして、こういったデータベース・ストレージ・BIツール・機械学習サービスを採用し、運用するかの知見を要する。

データ利活用のための「Things Cloud®」

「Things Cloud®」は、そうした



NTTコミュニケーションズ株式会社
経営企画部 IoT推進室

【左から】担当課長 増田 知彰氏
担当課長 竹中 康博氏

インフラ整備の中で、データ収集・蓄積を担う。プリセットのデバイスの管理体系、データ構造を使い、迷わずPoCに踏み出せる。閾値超過検知など一般的なストリーム処理も、標準で備わっている。さらに「Things Cloud®」標準のAPIで、BIツール・機械学習サービスなどの外部サービスと容易に連携できる。いわばデータ活用のハブ機能を果たせるのだ。こうして、多様なサービス／ツールと組み合わせることにより、お客さまのデータの利活用を推進する。

本格利用でも多くの実績がある。例えばNTTデータの「iQuattro®」は、データ収集基盤に「Things Cloud®」を採用している。「Things Cloud®」には、Microsoft Azureと連携し、Power BIでのインタラクティブ分析、PaaSを活用した業務アプリケーションとの連携ソリューション実績もある。その他の外部クラウドとも柔軟に連携ができる。このように「Things Cloud®」は、顧客ビジネスのデータ活用を後押しできる。

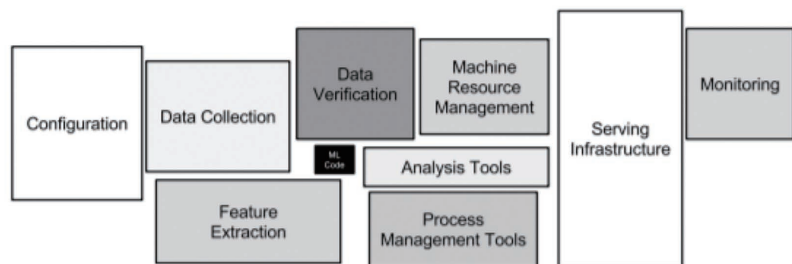


図1 「Hidden Technical Debt in Machine Learning Systems (機械学習システムの隠れた技術負債)」(D.Sculleyら NIPS 2015)の引用

※出典：<https://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems>