

ビッグデータのビジネス活用を成功させる4つの必須ポイント

～NTTコムウェアならではの技術とトータルサポート～

ビッグデータを用いたビジネスの成功事例が毎日のようにニュースをにぎわしています。「自分たちのデータも金の卵なのでは?」との期待が高まると共に、それを実現する技術やパートナーへのニーズも高まってきています。ビッグデータビジネスの成功には、(1)データ、(2)分析手法、(3)分析プラットフォーム、および(4)データサイエンティストの要素が必要です。これら必須要素に対するNTTコムウェアの取り組みと、その強みを紹介します。

NTTコムウェアならではのトータルサポート

ビッグデータビジネス成功の必須要素とは、データ、分析手法、分析プラットフォーム、そして人(=データサイエンティスト)です。ビジネス自体の歴史が浅いこともあり、これらすべてを備えた企業は多くはありません。NTTコムウェアでは、これまで培った数千万通の請求書発行業務や勘定／基幹システムのオペレーションのノウハウを基に、データマイニング、テキストマイニングおよび機械学習などの分析モデルやアルゴリズム(データから情報を取り出すための手順)の研究だけでなく、その実装方式や必要となるツール、アプリケーション、プラットフォームの検証を各組織で協力して行っており、ビッグデータビジネスをトータルでサポートする研究開発

を積極的に行っています。

データの面で

ビッグデータとして注目されるデータには、ソーシャルメディアおよびTwitterなどマイクロブログ、およびクリック、検索や購入履歴などのログデータがあります。その理由は、(1)これらのデータを使うことで、従来のマーケティング調査に比べ、情報収集や効果測定などが安価・大規模かつ短時間で実行できること、(2)これらのデータの多くはクロールやAPI^{*1}などでWebから収集できるため、自前でデータを持たないユーザも容易に利用できること、が挙げられます。そのため、まずは、これらのデータを収集し蓄積することが必要です。データ収集の技術支援・代行から、収集データの保存先としてのクラウド環境提供まで、NTTコ

ムウェアでは技術と環境の両方でお客さまをサポートすることができます。

分析手法の面で

データを蓄積する目的は、それらを活用することです。ビッグデータにおけるデータは、既存のデータに比べてただ量が多いだけでなく、生成頻度が高く、データの種類・形式、そして構造もバラエティーに富んでいます。従って、従来のデータマイニングや機械学習がそのまま適用できない場合や、目的に応じて新しい分析手法が必要となる場合もあり得ます。NTTコムウェアでは、ビッグデータからの消費行動分析、マーケティングおよびCRMのアプローチとして、図1に示すように「AIDALo^{アイダロ}」というモデルとそれに対応した分析手法を提案しています。「AIDALo」は、ユーザ間

図1 NTTコムウェアが提唱するビッグデータ分析アプローチ

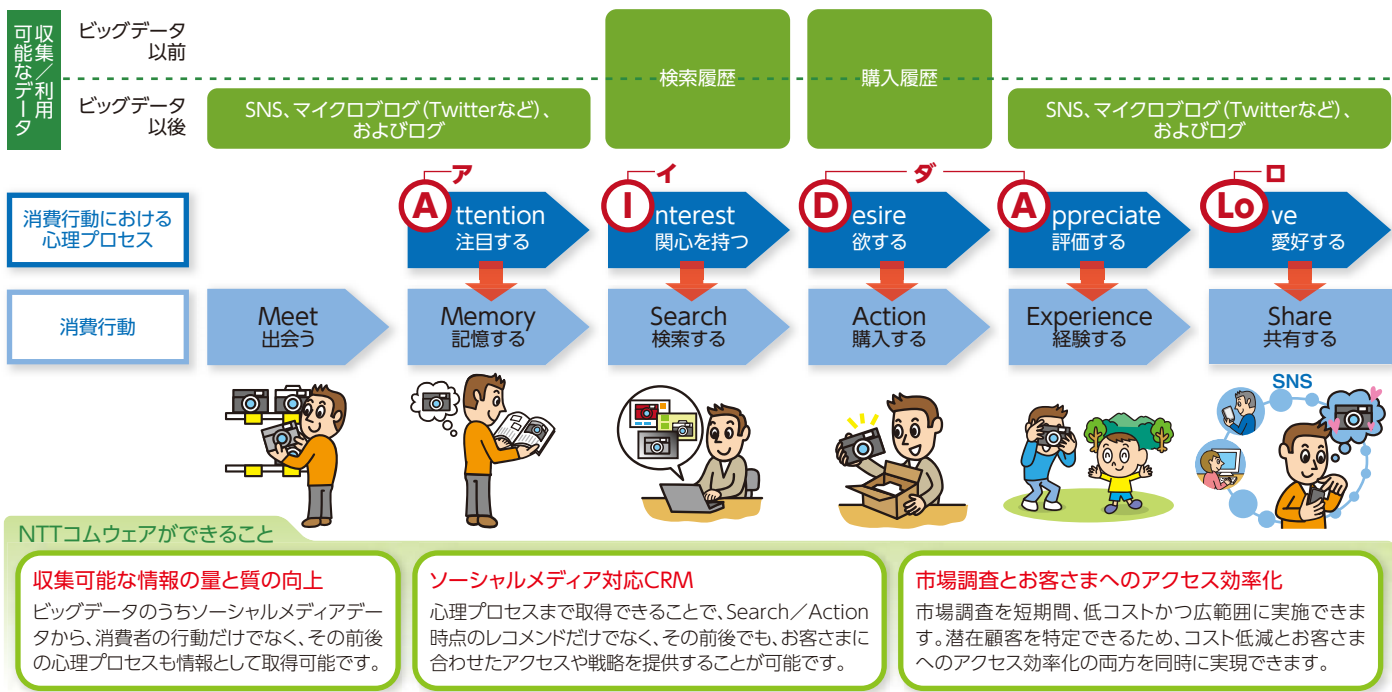
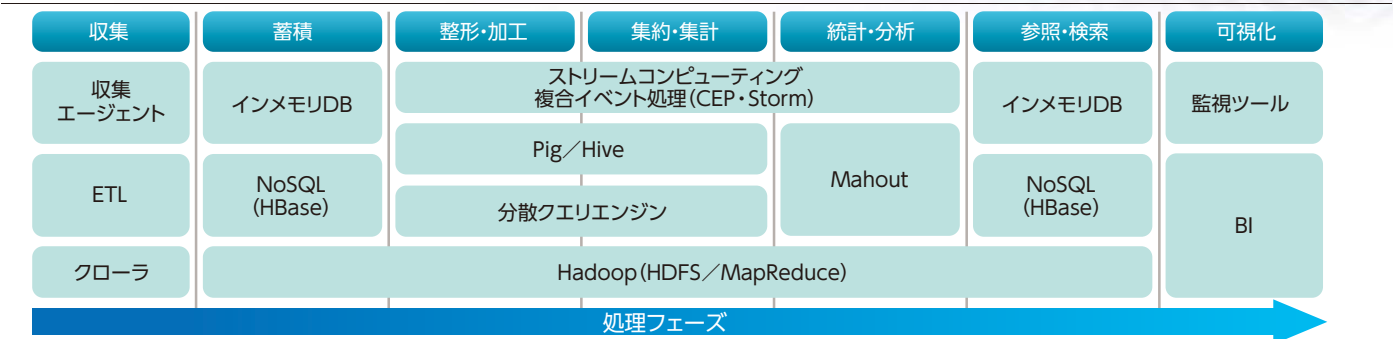


図2 ビッグデータにおける代表的な基盤技術



の情報共有と伝播過程などのネットワーク解析、およびSentiment analysis(感情分析)などのテキスト解析に着目／工夫をしており、ユーザの消費行動予測に資するモデルです。従来は検索クエリやクリック履歴、および購入履歴のデータが分析のメインで、ユーザが購入に至った背景や購入後の感想など一連の心理プロセスを取得することが困難でした。しかし、昨今、ソーシャルメディア上で発信される情報が増えてきたことにより、例えば、商品やサービスごとに、「どのような媒体を通じてどのようなユーザに認知され、どのような観点で評価を受け、誰に伝わるか」というユーザの消費行動における心理プロセスを明らかにするデータの取得が可能になり、企業のマーケティングや商品開発に生かすことが期待されています。「AIDALo」は、構造が違うこれらのデータを一括して分析し、個々の心理プロセスを明らかにすることができます。

分析プラットフォームの面で

プラットフォームは、機器のアラートのリアルタイム検知といったNTTコムウェアのノウハウが最も生かされる分野です。図2にあるように、ビッグデータに対する代表的な分散処理ソフトウェアであるHadoopは、安価なマシンの上で、従来よりも容易に分散処理を実現するプラットフォームであり、データフロー言語であるPig、分散データウェアハウスであるHive、機械学習のライブラリセットであるMahoutなどの関連製品によるエコシステムが充実している点で優れているため、広く利用されています。しかし、一括処理(バッチ処理)に特化しているために、リアルタイム処理が求められる場合には、それらを補完するソフトウェアが必要です。

そこでNTTコムウェアは、リアルタイム処理をHadoopで実現する製品の検証を、利用シーンと組み合わせて行ってきました。例えば災害検知といった利用シーンでは、入力時に機器センサーからのストリームデータ(無限に到来する時刻順のデータ)処理が求められるため、ストリーミング処理にはStorm^{*2}、出力したデータの蓄積と参照・検索にはHBase^{*3}をはじめとするNoSQL^{*4}製品などを検証し、既存環境の改善提案、および導入・オペレーション支援を行っています。加えて、ビッグデータに対応可能な検索システムの検証を実施中であり、今後はセキュリティやマルチテナンシー^{*5}などクラウド環境上での運用に必要な課題を解決すべく研究開発を進めています。また、最近では、トレンドの抽出に機械学習の利用が着目されています。抽出するトレンドには、最近人気の商品だけでなく、現在、あるイベントがどの場所で、あるいはどの番組が盛り上がっているかなど、リアルタイム性が強く求められています。このニーズに対し、私たちはHadoopと分散型メモリキャッシュシステム^{*6}を併用することで主要な機械学習のアルゴリズムの高速化を実現できるのではないかと考え、お客さまとの共同検証でその効果を確認しました。

データサイエンティストの面で

NTTコムウェアの現場主義と全員営業はビッグデータのビジネスについても息づいています。実際に、システム構築だけでなく、その前段階である企画や、後工程である運用中のサービス改善の局面でお声掛けいただく機会もあります。NTTコムウェアは、既存の機械学習に精通するだけでなく、お客さまとゴールを共有し、コンサルティングや分析手法の

提案、および分析結果のプレゼンテーションやビジネス適用まで支援可能なデータサイエンティストを育成しています。

- *1 Application Programming Interface。OSやアプリケーションが持つ機能の一部を外部から簡単に利用できるようにするインタフェース。
- *2 リアルタイム分散処理のフレームワーク。Twitter社がオープンソースとして公開。
- *3 Hadoop上に構築したカラム型データストア。
- *4 リレーショナルデータベース管理システム(RDBMS)以外のデータベース管理システム。
- *5 ソフトウェア、データベースなどを複数の顧客企業で共有する事業モデル。
- *6 複数のサーバーに搭載される物理メモリーを連携し、巨大な論理メモリーとして利用するシステム。

VISION

さまざまなビジネスニーズに総合力でお応えします。

ビッグデータというと、Hadoopを中心としたソリューションの話題が多いですが、それに加えて、データ、サービスの目的に応じたHadoop上での分析手法の選択、設計や開発、そして分析結果の解釈までの連携が必要です。場合によってはHadoop以外のアプリケーションの選択肢も必要です。NTTコムウェアではさまざまなビジネスニーズに応えられるように、システム構築だけでなく、企画、コンサルティング、オペレーションおよびデータサイエンティストの面でご協力できると考えています。

品質生産性技術本部 研究開発部
川前 徳章(かわまえのりあき)
工学博士、研究開発部 勤務。専門は情報検索、統計的機械学習、マーケティングサイエンス。現在は感性検索とコンテンツジェネレータの研究と開発に従事。東京電機大学 安田研究室 協力研究員。

ビッグデータのビジネスはNTTコムウェアにとってもビジネスチャンスの到来だと考えています。国内発の成功事例を、海外発のものに数でも質でも匹敵できるよう、私たちがお手伝いいたします。