

クラウドとビッグデータが作る新しい未来

畿央大学大学院教育学研究科 廣瀬 一海

1. はじめに

クラウドの普及は大量の計算リソースを手軽に、その場で確保できる事を実現しました。その中でも複数のコンピュータに処理を分割して、従来では不可能であった大きなデータ処理を行う、分散データ分析処理は、このクラウドの恩恵を最も享受できる分野の一つです。

従来であれば、このような処理はコンピュータと関連する機材が必要で、初期の設備投資や長い準備時間を割いて行うものでした。これがクラウドによって、必要な時に、その場で手軽に得られるようになりました。この結果、今までのビジネスのスピードを大きく変えようとしています。

そもそも、ビッグデータとは一体どういうデータなのでしょう？ そして、どのようにすれば、ビッグデータを活用できるのでしょうか？

2. ビッグデータとは

ビッグデータは、その名の通り「ビッグデータ(大きい情報)」というわけですが、どんなデータであっても、蓄積をしていけばいずれは大きくなるのは自明の理です。

各シンクタンクによるレポートで共通する一つの定義ですが、ビッグデータは従来の情報処理システムで蓄積・運用・分析を行った場合で「処理能力」を超えてしまい、「処理にとっても時間がかかるデータ」の事を指しています。これらのシンクタンクの将来予測では、私たちがインターネット上で蓄積され、取り扱うデータは全世界で40 ZB(ゼタバイト)となり、その内の90%のデータが特定のフォーマットや構造定義などによらない、WebサーバのログやSNSの投稿など、DBな

どの構造に格納されていない、非構造型のデータであると発表されています。

また、この5年の間で、「50%以上のデータ分析システムが機械・アプリケーション・人の行動によって生成されるイベント型のデータを対象とし、これらデータを活用した分析需要は3倍に増加する」との予測が発表されています。

これらの話を整理すると、以下のデータはビッグデータの性質を持っていると推察されます。

- ソーシャルメディアにおいて参加者が書き込むプロフィール・コメントなどのソーシャルメディアデータ
- ビデオ監視・動画・音声のデータと視聴記録などのマルチメディアデータ
- ECサイトにおける購入履歴、閲覧履歴やブログのエントリーなどのデータ
- 顧客データDM等の販促、店頭の購買志向情報、会員カードなどによるデータ
- GPS・NFC(ICの世界共通規格)・ICカード(Suicaなどの国内規格)の位置情報、乗車履歴・電気・水道・加速度から得られるセンサデータ
- メール・文書などのオフィスデータ
- Webサーバにおけるアクセスログなど、ゲームなどにおけるアプリケーションの行動データなどの履歴データ
- 販売管理等の業務システムにおいて生成されるPOSデータ、取引明細データなどのオペレーションデータなど

これらのデータのほとんどは、蓄積こそ何とかできていたかもしれませんが、積極的に分析するにはとても手間がかかっていたものばかりです。

JR東日本が関東圏域で展開している、駅のホームなどに設置されている「acure」という自動販売機があります。この自動販売機はカメラセンサを内蔵しており、季節、時間帯や温度によって液晶の中のコンテンツが入れ替わり、カメラを利用して年代や性別の判別を行い、その属性に合わせ、お勧めする商品が変わるというものです。

JR東日本ではこのPOSデータの解析により「小容量ペットボトル商品は午後女性・中高年の方に多く購入される」というデータが得られ、小容量ペットボトルの積極採用に至りました。



図1 自動販売機acure
(<http://www.acure-fun.net/acure/index.html>)

このように、ビッグデータの処理によって、私たちの生活を取り巻く分析を行う事で、今まで認識されていなかった課題を明らかにする事ができるのです。

3. ビッグデータが注目されている背景

ビッグデータが注目されるのはどうしてなのでしょう？ その背景には以下のような理由が考えられます。

1. データを取得可能なデバイスの普及
2. 人々が使うアプリケーションの普及
3. クラウドストレージの普及
4. データを解析する技術の普及

クラウドやインターネットの普及によって、世界中の人々がスマートフォンやタブレットに代表されるような、携帯デバイスを持ちまわるようになりました。無線や携帯電話網の整備もあり、これらのデバイスを使って自宅や職場などの場所を選ぶことなく、いつでも必要な時に作業をできる環境が整っています。カメラ、GPSや基地局、ジャイロによる位置情報、加速度センサによる走る、止まるなどの動作情報、テキストメッセージやインターネットのアクセスの閲覧履歴、ダウンロードしたアプリのデータなど大量のデータがユーザーによって生み出されています。

北米のシンクタンクであるForester Researchは、2012年のレポートで「17か国のインフォメーションワーカーの内52%が3つ以上のデバイスを使って作業」を行うだろうと報告しています。

また、デバイスはタブレットやスマートフォンだけに限りません。この数年の話題である、Internet of Things (モノのインターネット) では、日々昇降するエレベータやエスカレータ、トンネルを走る地下鉄、チケットの券売機、入場記録、地下鉄のホームを照らす照明器具、車に搭載されたGPSやセンサなど、人々の生活によって今では無くてはならないものからもデータを取得できるようになっています。

デバイスが普及すれば、当然ながらそのデバイス対応のアプリケーションが普及する事となります。FacebookやTwitterなどのアプリを使って、多くの方は日々の生活を記録したり、写真をアップロードしたりしている事でしょう。この際の行動時刻やGPS情報、投稿されたテキストなど、利用者は多くのデータを生み出していきます。

近年発表されているApple WatchやFitbitに代表される腕時計型デバイスは、活動量計センサとして稼働しており、歩いた歩数、心拍数、睡眠時間、食事による取得カロリーなどのデータも生み出しています。アプリケーションを起動した時間すら、生み出すデータの一つとなります。

様々なデバイスとアプリケーションから、能動的なデータを生み出すと問題になるのはデータの保存場所です。当然ながら、これらのデバイスの記録領域には限りがあります。そこで、クラウドのストレージにデータを保存する事で、デバイスの記録領域の制約を受けずに、多くのデータを扱えるようにするのです。クラウドによって提供されるストレージは、ほぼ無尽蔵と思われるほどに記録可能なデータ最大容量を拡大しています。

マイクロソフトが提供しているクラウドプラットフォーム「Microsoft Azure」では、2015年4月に50兆のオブジェクトデータが格納され、昨年と比較して3倍のデータアクセスが行われたと発表しています。

また、HDDやSSDなどのストレージ部品の大量購入の結果、提供価格も下がり、年々容量に対しての価格は下がり、容量は大規模化を続けていきます。

データが格納されれば、それらのデータを使って統計や解析を行い、新しい知見や動向を得ようと試みるのは自然な事です。しかし、これらのデータはあまりにサイズが大きく、手軽に解析するにも膨大な時間がかかっていましたし、従来型のデータベースに仮に投入して格納できたとしても、解析には時間がかかり現実的ではありませんでした。

2004年にGoogleが大量のコンピュータを並べ、データを各コンピュータに振り分けて効率的に処理を行う「Map and Reduce」を発表しました。2010年には大量のコンピュータが分散で協調動作するデータベースシステムの「Dremel」について論文を発表し、その2年後の2012年に数千台のコンピュータで高速に結果を得られる「Google Big Query」として提供を開始しています。

大手インターネット通販サイトであるAmazon.comが運営するAmazon Web Servicesも、2012年に統計解析に適した列志向データベースシステムである「Amazon Redshift」を発表しています。

また、これらのデータをコンピュータ自身が学習し、プログラミングを行うことなく、自律的に機能する「機械学習」と呼ばれる分野があります。最近のコンパクトカメラには、顔を認識する機能が付いているものがありますが、これも「顔を認識する」機械学習の事例の一つです。この機械学習には、大量のデータを学習させる必要がありますが、ビッグデータは正にうってつけの素材データなのです。これらのデータの学習をさせることによって、認識、分類、予測などを行う事ができるようになる技術も普及しました。

2014年にマイクロソフトは「Azure Machine Learning」を発表しています。従来の機械学習は、精度を向上させる為にも膨大なデータと計算量が必要とされ、十分な機能を持たせるまでにかかりの時間を必要としてきました。これがビッグ

データとクラウドによって簡単に実現可能となったのです。

2015年にマイクロソフトは「Project Oxford」を公開しました。機械学習済みのモデル（学習結果）を手軽に使えるようにするプロジェクトです。例えば、この機能を使ったウェブサイト「How-Old.net」は、ユーザがアップロードした写真から「年齢を判定」するウェブサイトです。

また、最近では写真の分析から、楽しいのか悲しいのかななどの「感情認識」が可能になりました。



図2 写真から顔と感情を認識する機械学習モデル (<https://www.projectoxford.ai/demo/Emotion#detection>)

画像から得られる学習結果だけでなく、機械学習は「言葉の認識」や「音声の認識」も可能となってきています。つまり、ユーザが「何を話しているか」を認識するようになりました。最近の利用ケースでは、日本マイクロソフト株式会社がLINEで提供を開始した女子高生AI「りんな」などにも応用されています。

最新のOSであるWindows 10で提供を開始したマイクロソフトの音声対応パーソナルアシスタント「Cortana」(コルタナ)は、会話やメールのデータ、予定のデータなどを解析、認識する事で「そろそろ打ち合わせですよ?」と近い未来の提案までしてくれます。

この近い未来の予測は、エレベータやエスカレータ、駅の券売機にも応用されています。今までの機器の電圧や稼働時間のデータを学習させる事で「故障する時期」を教えてくれるようになりました。これによって機器が故障する前に部品を交

換する事が可能になったのです。

150年の歴史を誇るロンドン地下鉄（UNDERGROUND）は、Microsoft Azure Intelligent Systems Serviceを用いて、行先案内板や運行状況の表示パネルなど、運行業務に必要となる機器の故障の予測を可能にしており、年間10億人の利用者を支えています。



図3 UNDERGROUNDの故障検知マップ
(Microsoft CityNext https://www.youtube.com/watch?v=f1_t1tsgvFIより)

すでに私たちの周りには、センサが搭載された機器が多数存在しています。トヨタ自動車が開発を進めているT-Connectナビでは、400万台を超える車両にこのナビゲーションシステムとセンサが搭載されています。自動車に搭載されたセンサは、渋滞箇所の分析や将来の渋滞箇所と時間の予測、新しい自動車の利用シーンの開拓などを可能にしました。例えば、集計されたデータに応じて、事故発生箇所の危険度分析を行えば、自動車の運転中に予測と警告などが可能になります。

4. クラウドとビッグデータが作る新しい未来

まるで夢物語であり、SFの世界のような話ですが、これまでの話のいずれもが既に使われている技術です。近年になって、クラウド・デバイス・ビッグデータという、3つの材料が出そろった事によって実現できた事がわかります。これまでは活用しないままに埋もれていたデータが解析の材料になる可能性は至る所にあります。

今までは、キーボードやタッチパネルで操作をしてきたコンピュータでしたが、近い将来は「話しかける」だけで、家の中の照明機器を操作したり、冷蔵庫の中の材料から作れる料理のレシピを提案してくれたり、そんな生活が待っていると思

うととても楽しみです。

引用・参考文献, Webサイト

- Forrester Research, Inc.: “BT Futures Report: Info workers will erase boundary between enterprise and consumer technologies,” Aug. 30, 2012
<https://www.forrester.com/Info+Workers+Will+Erase+The+Boundary+Between+Enterprise+And+Consumer+Technologies/fulltext/-/E-res77881>
- Google BigQuery
<https://cloud.google.com/bigquery/?hl=ja>
- Google Dremel
<http://research.google.com/pubs/pub36632.html>
- Map and Reduce
<http://research.google.com/archive/mapreduce.html>
- Microsoft finally talks about how many customers are REALLY using its Amazon-competitor cloud
<http://www.businessinsider.com/microsoft-azure-usage-doubled-2015-4>
- THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadow s, and Biggest Grow the in the Far East
<https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>
- 女子高生AI「りんな」
<http://rinna.jp/rinna/>
- トヨタのCIOが語るConnected CarとAzure, これからのIT屋
<http://ascii.jp/elem/000/001/012/1012038/>
- ロンドン地下鉄を丸ごと可視化するマイクロソフト
<http://www.microsoft.com/enterprise/ja-jp/the-lounge/cases/201408-lodon.aspx#fbid=fz1I8oyJUXE>