

ビッグデータの波と 求められる人材像

アイティメディア株式会社

@IT担当編集長

三木 泉

「どれくらいの量のデータをビッグ
データと呼ぶのでしょうか？」

—最近のある記者説明会での質問

クラウドの適切な定義とは？

「利用者が、利用したいものを、利用したいだけ、利用することに専念できるIT消費スタイル」

ではビッグデータの適切な定義とは？

「デジタルデータを駆使して、ビジネスやリサーチなどの価値を高める行為、またはその対象データ」

これではあいまいすぎる？ しかし
ビッグデータは比較級でしか語れない

Big Data Examples

米国歳入局や証券取引委員会における不正検知

日本でも例えば生活保護の不正受給検知に適用できるはず

さまざまなデータの突き合わせ

「例えば、人気商品のコロッケはどんな人が繰り返し買っているのか。意外な結果が出ました。最もリピーターになっていたのは実は、60歳以上。コンビニにとって課題のシニア層を呼び込める商品と分かったのです。そこでシニア層の客などを狙う店では、夕飯用にコロッケを大量に並べました。さらに、この店ではカードのデータから徒歩5分圏内の客が多いことが分かりました」

—NHK「社会を変えるビッグデータ革命」

ローソンポイントカードの例:「年齢、性別、住所の情報」を活用もっとやっているはずだが、「エビデンス」の提供が目標の1つだということを示している

顧客のマイクロセグメンテーションによる
新たなサービスの可能性

あなたがローソンにいたら何をしますか？

顧客の情報をどこまでとれるかが重要な
ポイントでもある

楽天は楽天カードの作成時に多様な情
報の入力を求めている

ヤフージャパンとTカードの提携が意味す
るものとは

ベネッセの情報優位性

WIDEはインターネットカー・プロジェクトで
天気や渋滞のリアルタイムデータを集め
ることをやっていた

リアルタイムでデータを提供することで
サービスが向上、あるいは新たなサービ
スが生まれる

Health Map

Google Newsで取得できるローカルニュース、WHOの最新情報、メーリングリストのディスカッション、Twitterなどを分析して数時間で世界各地の感染症アウトブレイクあるいはその可能性を検知

改めて、ビッグデータとは

データ

- ・これまでに入手できなかったデータ
- ・入手可能だったがデータ処理コスト、データ分析手法の未成熟などから入手をあきらめていたデータ
- ・これまで入手していたが(十分に)活用されていなかったデータ

データ分析

- ・大量なデータを分析
- ・新たなデータ分析手法を導入
- ・これまで関連付けて分析していなかった複数のデータを関連付けて分析
- ・データ分析をより短時間あるいはリアルタイムで行う

目的

- ・行為にエビデンスを与える
- ・行為の精度を高める
- ・将来を予測する
- ・製品やサービスの新たな販売機会を創出する
- ・それ自体が価値をもたらす新たなサービスを創出する

利用される技術

Map/Reduce、NoSQL/SQL DB, インメモリ DB、複合イベント処理(CEP)技術、R、BI ツール.....

非構造化データの扱いやテキスト分析が
広まる、ただし既存データおよびリレー
ショナルDBとの連携が必要とされる場面
も多い。統計解析手法は基本的には従
来のものを活用

広がるエコシステム

- ・データ分析関連ハードウェア
- ・パブリッククラウドサービス
- ・データベース、Map/Reduceメカニズムなどの進化
- ・BIツールの進化

求められる人々

- ・データの分析や活用のセンスがあるビジネス担当者（日本では特に不足）
- ・ビッグデータ分析基盤の構築・運用ができる人（社内、社外）
- ・データサイエンティスト（社内、社外）

「ビッグデータ分析基盤の構築・運用ができる人」にもいろいろな意味がある。クラウドでいえばIaaS、PaaS、SaaSのどのレベルで提供するか

データサイエンティストが解析プロセスのすべてを担うのか、ビジネス担当者にはどのレベルまでのデータ、分析機能を提供するか

データサイエンティストという新たな仕事

- ・「分析屋」、だがそれだけではない
- ・数値のみでなく定性的な情報をどう分析して意味のある結果を得られるか
- ・与えられた命題を解くだけでなく、命題自体を探していく役割も担う
- ・ビジネス自体への理解あるいはビジネス要求への理解をベースに、周りの人々とどう建設的な協力関係を築けるか。協調性と自律性、能動性が求められる

ビッグデータは、業界を変える可能性も
秘めている