

学会講演会

第36回 横浜経営学会講演会

ビッグデータとビジネス・アナリティクス

中 川 慶 一 郎

日時：2014年11月12日（水）14：40～16：10

場所：経営学部講義棟2号館（B棟）営-109教室

<司会> 今日は横浜経営学会講演会ということで、NTTデータの中川慶一郎さまにおいでいただきました。題目は「ビッグデータとビジネス・アナリティクス」ということで、マーケティングを非常に、根底から書き換えるような意味合いを持ってしまして、非常に貴重なお話を、我々は無料で聞けるということで、非常にラッキーな会になっております。

それでは、まず最初に、学部長の中村先生のほうから、ごあいさつをいただきたいと思います。よろしく願いいたします。

<中村学部長> 皆さん、こんにちは。経営学部長の中村です。私は経営学部長と、かつ経営学会の会長ということで、こちらでお話いたします。まず、今日は出席いただきましてありがとうございます。また、準備をいただいた各種連絡等を含めてお願いした先生方にも感謝いたします。ありがとうございました。

本日、タイトルのとおり「ビッグデータとビジネス・アナリティクス」ということで、先ほども中川さんとお話したんですけれども、出席の会員の皆さんは若いのですが、もはや好むと好まざるとにかかわらず、このビッグデータは毎日聞くことばです。皆さんは知らず知らずのうち、このビッグデータの枠の中で生きているわけです。

皆さんはこれから社会へ出て多分大成功をしてくれるはずですから、そういう社会で大成功するときに、ビッグデータとどういうふうにお付き合いするか、これは極めて重要だと思っています。

今日は色々な話、先ほど私も中川さんと楽しくお話させていただいたんですけれども、それを分かりやすくお話していただけたと思います。皆さんはこれから、じゃあビッグデータなるものというふうにお付き合いをするかということを考えながら聞いていただければ、なお一層楽しくなると思っております。今日、有意義な一日になると思いますので、皆さんも色々なことを考えてほしいと思います。

簡単ではございますが、これをもちまして私のごあいさつといたします。今日はわざわざ出席いただきありがとうございます。

<司会> それでは中川さま、ご準備ができましたら、よろしくお願ひします。それでは、拍手でもって中川さまをお迎え下さい。

<講演>

皆さん、初めまして。NTTデータ数理システムから参りました、中川と申します。本日は講演の機会をいただき、誠にありがとうございます。今日は「ビッグデータとビジネス・アナリティクス」というタイトルで、1時間20分ぐらいの内容を用意しています。通常、講演というと、大体40分から45分ぐらいなので、1時間20分というのは結構長いです。時間調整が難しいので、まずは一通りお話をして、後半、時間が余るようだったら、より具体的な事例をお話したいと思います。

まず初めに自己紹介ですが、私はNTTデータ数理システムという会社に勤務しています。皆さんもデータマイニングや機械学習という言葉聞いたことがあると思うのですが、そういった技術を使ったデータ分析を専門としている会社です。

私自身は、NTTデータという会社に22年前に就職しまして、ずっとマーケティングを中心にデータ分析の仕事をしてきました。というのは、NTTデータはシステム開発の会社なので、様々なお客様の情報システムを構築します。すると、データベースにデータが蓄積されてきますので、それをどう活用するか、お客様に提案する仕事をずっとやってきました。2年前にNTTデータ数理システムという会社がグループ入りし、現在はそちらで仕事をしています。

今日はビッグデータというテーマで話をしますが、格調の高い話は一切しませんので、皆さんもそのつもりで聞いて下さい。では最初に、事例を持ってまいりましたので、皆さんにお見せしたいと思います。これは少し前になりますけれども参議院選の事例です。前回の参議院選挙はネット選挙と言われました。ネット選挙はどうだったかという、実はあまり盛り上がりなかったというのが総括でした。ただ、個別に見て行くと、実は面白い現象がいくつかあったので、紹介します。

こちらの動画をご覧ください。IY候補のツイートした後、HA氏がツイートした様子を可視化したものです。最初にIY候補がツイートしたときは、リツイートが少ししかありませんでしたが、その8時間後にHA氏がリツイートした瞬間、画面でご覧いただくように急激にツイートが拡散していきました。こうした現象はバーストと言われたり、その中心となっている人はインフルエンサー (influencer) と言われたりしますが、そういった現象がつぶさに見てとれます。

次の事例は、NTTデータが提供している「なずきのおと」というサービスからです。このサービスでは、ツイッター、ブログ、ニュース・フィードなどの情報を集めて、高度な言語処理を行い、それを企業のマーケティング担当者やIR (Investor Relations) 担当者に配信します。

このサービスのポイントは、ツイッターのようにリアルタイムなメディアからインシデントを検知して、すぐ知らせることです。また、それを活用して、顧客満足やプロモーションの効果測定、場合によっては炎上検知へのアクションなどに結びつけていただくことを狙っています。

こちらをご覧ください。震災のときにあるチェーン店の店長さんが「営業再開しました。テレビは地震ばかりでつまらない。そんなあなたの来店をお待ちしています。」とつぶやきをしました。あまり適切ではないですね。こちらのグラフは20時くらいツイートしてから3~4時間後くらいの様子です。上のグラフはさきほどのツイートを受けたリツイートの数です。下のグラフはポジティブ、ネガティブに分けたときのリツイート数の推移です。これを見るとネガティブ・

ツイートが急激に増えている、いわゆる「炎上」という状態になっていることがよく分かります。「なずきのおと」は、こういったことを検知して、アクションに結び付けていただくサービスですが、こういった形で、ビッグデータを少し分析するだけでも、それなりに色々と見えてくるものがあります。

NTTデータ・グループでは、他の会社と同じように、ビッグデータの取り組みを行っています。ここでは、ソーシャル、エンタープライズ、後ほど詳細にお話するM2Mの3つの領域で、Go-to-Market戦略を展開しています。

このGo-to-Market戦略を支える強みとして、データ分析方法論、ビジネス・アナリティクス基盤、ビッグデータ基盤の3つがあり、また、ケーパビリティとしては、豊富な実践経験と専門家集団があり、これらを組合せてお客様のニーズに対応しています。

3つの強みを具体的に説明すると、ビッグデータ基盤とは、例えば大規模データ処理の基盤であるHadoopであったり、ストリーム・データを処理する基盤であるCEP（complex event processing: 複合イベント処理）であったり、ビッグデータを処理するためのIT基盤の技術です。NTTデータでは、こういったIT基盤の技術については、ベンダ製品にとらわれることなく、幅広くノウハウを蓄積しています。

データ分析の метод論については、私自身20年来やってきた分析コンサルティングの経験を9つの分析シナリオという形で体系化しています。そして、今回NTTデータ数理システムがグループ入りし、データマイニングや数理計画、シミュレーションといったビジネス・アナリティクス基盤を構成するパッケージ群が揃ったことで、三位一体でソーシャル、エンタープライズ、M2Mという3つのGo-to-Market戦略を推進することができるようになりました。NTTデータの説明はこのぐらいにして、本題に進みたいと思います。

今日は3つのお話を用意しています。皆さんもすでにビッグデータの話は聞き飽きたかも知れませんが、まず初めに「ビッグデータとは」ということについておさらいします。次に「ビッグデータ活用の発展」ということを少し考えてみたいと思います。

最後に、ビジネス・アナリティクス（BA: Business Analytics）が今、新たな挑戦のフェーズを迎えているということを皆さんにお話したいと思います。ビジネス・アナリティクスというと難しく聞こえるかもしれませんが、要はデータ分析だと思って下さい。

まず初めに「ビッグデータとは」ということについて説明します。皆さん、ビッグデータという言葉メディアで目にしない日はないですね。IBM、アクセンチュア、日立、富士通、NEC、みんなビッグデータ、ビッグデータと騒いでいますね。

実際我々の社会は、ビッグデータの時代、情報爆発の時代を迎えています。例えば、ツイッターでは一日に4億件のツイートがあるそうです。フェイスブックの場合だと、「いいね」の件数が一日27億件、写真のアップロード数が3億件だそうです。Amazonでは、Cyber Mondayという日本でいうところの年末商戦みたいな時期に、一日当たりの注文が3680万件にも及ぶそうです。LINEでは一日に70億件のショートメッセージが飛び交うといった具合に、まさにビッグデータの時代を迎えています。

調査会社のIDCジャパンの推計によると、全世界にある電子化された情報は2012年では2.8 zettabyte（ゼタバイト）で、それが指数的に伸びて、2020年には40ゼタバイトになるとのことです。また、その大半は、音声、動画、テキストといった、今までの表形式のデータ（構造化デー

タ)と違う、いわゆる非構造化データが占めるそうです。皆さんはゼタバイトがどのくらいの単位か分かりますか。皆さんが馴染みのあるはキロバイト、メガバイト、ギガバイトくらいだと思います。その上がテラバイト、さらにペタバイト、エクサバイトとなって、その上にゼタバイトが来ます。ブルーレイが大体25ギガバイトくらいなので、1ゼタというとは40億枚ぐらいです。

同じくIDCジャパンの最新の情報によると、ビッグデータ関連の国内市場は現在6千億円で、2017年には1兆2千億円ぐらいになるそうです。実際我々の回りでもビッグデータやビジネス・アナリティクスの案件というのは、確かに増えています。

ビッグデータとは何かというと、実は明確な定義はありません。総論として言えることは2つの系譜があるということです。一つは、例えばツイッター、ブログといったソーシャル・ネットワークワーキング・サービス (SNS) やインターネット通販など、いわゆる“生活回り”の履歴である「ライフログ」と言われている系譜です。

もう一つは、RFID (非接触型ICカードのように、電波を使って物品や人物を識別する技術) やリモート・センシングといったセンシング・データの系譜です。また、ここで注意して欲しいのは、これら2つの系譜の重なる領域があることです。例えばHEMS (Home Energy Management System) のように家中の電力をモニタリングしたデータは、センシング・データでもあり、ライフログ・データでもあります。皆さんの持っているSuicaもセンシング・データであり、移動の履歴が分かるという意味ではライフログ・データです。しかし、実際にはこのような系譜を区別することなく、大雑把にビッグデータと言っているのが現状です。

こういったデータを蓄積し、分析・活用することで、社会なり、企業なり、個人なりに何らかの価値のある情報を提供していこうというのが、今のビッグデータに向けられている最大公約数的な期待なのだと思います。

また、ネットでビッグデータという言葉調べると、Volume, Variety, Velocityの「3つのV」という言葉が出てきます。これは、以前のデータと比較してビッグデータが持つ大きな特徴です。Volumeというのは当たり前ですね。Varietyというのは、データの多様性です。Velocityというのは速度と直訳されるのですが、絶え間なく発生するデータの稠密性と思って下さい。具体的には次のスライドをご覧ください。

「多様なデータが集まる」とは、数値データ以外にも、例えばテキスト、それも皆さんがツイッターなどで書く平文のテキストや音声、画像といった多様な種類のデータが集まるということで、いわゆるVarietyのことを指します。これもVarietyの話になりますが、インターネット上のあちこちで情報が発信されており、色々なところからデータが集まります。また、「絶え間なくデータが集まる」とは、センシング・データのように連続的発生するデータが集まるということで、Velocityのことを指します。

こういった状況になるとどういった事態が起こるかということ、従来のデータベースではもはや溜めきれない、あるいは投資に耐えられないということになってしまいます。また、データの発生に処理が追いつかないということにもなってしまいます。それに伴って、従来型の情報分析・活用技術では、もう太刀打ちができないということになり、現在ビッグデータ時代の新たな技術が求められています。

一方、ビッグデータを支える技術は、IT基盤の技術と、分析関連の技術に大きく分けられます。IT基盤では、インメモリ、Hadoop、NoSQL、複合イベント処理といった単語が出てきます。

こういったIT基盤は、大規模データを高速に蓄積し、検索、処理することを目指しており、ITベンダは色々な方式を競っています。

例えば、大規模データの処理を1台のマシンで行うは無理なので、Hadoopという技術を使って並列・分散処理を行います。Hadoopでは、パソコンみたいな普通のマシンを、何百台、何千台と並べてデータを分散させ、一斉に「よーい、どん!」と検索やデータ処理をします。そうすることで速く返って来るといいう仕掛けです。

また、ストリーム・データといって次から次へと流れてくるデータに対しては、溜めてからまとめて処理するのではなく、来た球を打ち返すように処理するCEPという技術が使われます。それから、すべてのデータをメモリ上に持てば処理が早くなるので、インメモリという技術も使われます。

分析技術の方も、データマイニング、機械学習、自然言語処理、セマンティック検索といった色々な単語が出てきますが、やりたいことは何かというと、とにかく量を質に変えたいんです。ただ、IT基盤にせよ、分析技術にせよ、技術は生々流転なので、皆さんはあまり言葉に惑わされず、こんなキーワードがあるのだな、という程度に覚えていただければ結構です。

次に、ビッグデータはどこを目指しているのかということについて、私の個人的な考えをまとめたものを説明します。

社会、企業、個人はそれぞれ問題や課題ニーズを抱えています。社会でいくと、低成長、少子高齢化、省エネルギー社会、原発といった問題があります。そういった中で、社会としては限られたリソースをいかに上手く活用していくかということが課題になります。ビッグデータの活用は、それをどのように解決にしていこうかという方向に向かっていくことは間違いのないと思います。企業の方は、グローバル化し、かつ予測困難なマーケットの中で、少しでも変化の予兆を察知して機先を制するために、ビッグデータをどう使っていくかという方向に向かっていくと思います。個人については、ソーシャル・メディアやスマート端末が浸透していく中で、より気が利いたとか、より知的なとか、最近でいくと、経験価値というものをどう実現するかというところに、ビッグデータの活用というのがシフトしていくと思います。今後はますますこういった課題やニーズに応えるビジネス・アナリティクスというものが求められると思います。

続いて「ビッグデータ活用の発展」というお話をしたいと思います。まず初めに、こちらのスライドをご覧ください。昔アルビン・トフラーという人がいて、『第三の波』という本が話題になりました。それを真似して4つの波と表現しています。

なぜ、この絵をつくったかということ、4、5年ぐらい前から、メディアの人やユーザ企業様と、ビッグデータについて色々話す機会がありました。そうすると、どうも話が噛み合わない、ということがありました。なぜだろうと思ったのですが、よくよく考えてみると、イメージしている時間軸が人によって違うことに気が付きました。どこにビジネスの刈り取りの焦点を当てているかという時間軸が違うんです。それを4つに分類したものが、「ビッグデータ活用の発展モデル」です。

第1の波は、すでにビジネスとして成立しているものです。次に第2の波はというと、我々NTTデータとしては、センシング・データを活用するビジネスというのが来るだろうと考えています。第3の波は、本来業務やサービスで蓄積されたデータを何か別のサービス・業務に活用するビジネス。最後に、一見価値のないデータを組み合わせると、新たな価値を生み出そうと

するビジネスを第4の波としています。

ここで、4つの波について、少し細かくお話ししたいと思います。第1の波は、すでにビジネスとして成立しているものです。グーグルのような検索は、ビッグデータを活用した、すでに出来上がったビジネスです。我々NTTデータでも、例えば決済システムやNTTドコモのiモード・ゲートウェイのシステムなどを構築するお手伝いをしています。これらのシステムで発生するデータは巨大なデータで、お客様側でも現状の集計するような分析はすでに行われています。こういったシステムを開発しているメンバと話していると、「俺たちビッグデータの分析なんて、もうやっているよ。何言ってるの!?!」、みたいな感じですが、でも、これをビッグデータ分析という、ちょっと響きが違いますよね。これは単なるビッグデータ処理ですね。こういったビッグデータ活用はすでに行われており、システム開発としても収益は上がっています。とはいえ、まだプレ・ビッグデータというのが第1の波だだと思います。

こちらはテレコムや放送事業者向けのCRM分析で、第1の波で実際によく行われる分析の例です。サービスの利用度合いを横軸に取り、1人当たりの月額収入を縦軸に取って、お客様をプロットします。このプロットを縦・横2つに分割し、4つの象限に区切ると、当然右上が優良顧客になるというように「見える化」できます。次にその時間的な推移を追っていくと、どのくらいのお客様が上に上がって、横にシフトするのか、あるいは横にシフトしてから上に上がるのか、はたまた突然離反してしまうのか、というようにお客様の成長パス毎に動きが把握できます。さらにデータマイニングの手法を使うと、年齢、性別、利用の仕方といった顧客プロフィールから成長するお客様とそうでないお客様の違いを分ける要因が導き出されます。

実をいうと、この分析は10年前、15年前のCRMのパラダイムと何ら変わっていないんです。ビッグデータといっても、ただ単にデータが大きくなっただけで、考え方が何も変わってない昔ながらのCRMのアプローチです。こういう分析はすでにビジネスの場で使われています。

続いて第2の波はセンシング・データを活用するビジネスです。このビジネスを考える上で、M2M (Machine to Machine: 機器間通信) ネットワーク、最近ではIoT (Internet of Things: もののインターネット) といった方がいいのかも知れませんが、これを抜きして語ることはできません。M2Mとは、スマートフォンやパソコンだけでなく、車から、工場から、店舗から、橋から、あらゆるものをネットワークで繋いでしまおうというコンセプトで、スマート・コミュニティーやスマートシティ、いわゆるスマート××と名の付くようなものを、実現するインフラになります。このように色々なものを繋ぐことで、人、モノ、金の詳細な情報が時系列的に集まっていくという状況が、今まさに実現されようとしています。

M2M, IoTというのは、年率20%の成長ビジネスと言われています。例えば、General ElectricはIoTをビジネスの柱にしていこうと言っています。一方、M2Mのインパクトを考える上で、一つ間違えなく言えることは、センサーが身の回りを覆い尽くす社会が来る、いやもう来ているということです。皆さん持っているスマートフォンもセンサーの塊です。GPSの位置情報からジャイロ・センサの振動まで、色々なものが計測可能です。車もセンサーの塊です。そのうちヘルスケア・デバイスも普及してくるでしょう。しかし、繋がりますと、情報が集まりますというだけでは、M2Mの本当の価値を見誤ってしまいます。

自動販売機を例に考えてみましょう。知っている人は知っていると思いますが、実は自動販売機はかなりの割合でネットワークが繋がっています。各自販機から送られる情報がセンターに集まり、そのデータが配送担当者に送られます。このようなビッグデータを活用することで、

欠品ロスの削減や配送の効率化ができるというのが、自販機オンラインのメリットです。

ここまででは、誰が考えたって分かるビッグデータ活用のメリットですね。しかし、自販機におけるM2Mの本質な価値は違うところにあります。というのは、自販機というのはコンビニの店頭と違って、つい最近まで在庫切れになっているかどうか、蓋を開けてみないと分かりませんでした。その中で、配送担当者は何をするかというと、例えば水三分の一、お茶三分の一、清涼飲料を三分の一といったバランスで荷物を積んで順番に回って行きます。そうすると途中で、「水がなくなっちゃったとか。」という、「じゃあ、しょうがないから帰るか。」となるわけです。

実際はここまで何も考えてないわけではなくて、配送担当者というのは、この季節は、あるいはこの地域では、大体こういうものが売れるというのが頭の中に入っていて、積み荷のバランスを変えていきます。しかし、よくよく考えてみると、これは人の頭の中に宿る知識でしょう。残念ながら、人の頭の中に宿る知識というものはその人がやめてしまうと遺失してしまいます。これが大きい経営課題だったのです。

それをM2Mでつまびらかにしました。どういうふうには考えて、どういうふうには積み荷のバランスをとるか。いわゆる暗黙知を形式知化することができたというのが、ここでのM2Mの大きい価値です。

続いてリモート・メンテナンスの例です。現在、複合機をはじめ、色々なものがネットワークに繋がっています。一番有名なのはコマツ様のKOMTRAXです。KOMTRAXでは建機やブルドーザなど、海外に輸出したもまで、全部ネットワークに繋がっており、それをリモートでメンテナンスしています。

このように、機器から集まるビッグデータを活用することで、故障の予兆を発見したり、故障の箇所を特定したり、さらに最寄りの代理店が迅速に駆けつけたり、といったメリットがあります。このメリットも誰が考えても分かりやすいのですが、ここでのM2Mの価値は何かというと、アフター・サービスでの顧客経験価値を高めるというものです。

少し詳しくお話すると、重要なことは、消費者は二度評価するということです。マーケティングをやる人にはとっては常識なのですが、一番目の評価は買うときです。二番目の評価が下されるのは買ってからです。

企業は、売るときは、「買って下さい、買って下さい」ともの凄いエネルギーを割きますが、いざ買ってもらうと、どうしてもエネルギーが細ってしまいます。

一方、消費者はというと、買うときは買うときで、もちろん心をときめかせます。その後、最初にメーカーの言葉で商品の利便性を理解します。要は説明書を読んで、メーカーの言われたとおり使います。ところが、消費者は商品を使っていく中でだんだん生活化していきます。そういった経験の中で価値を見出し、ロイヤルティが確立されていきます。これが二番目の評価です。この顧客経験価値が次の購買に大きな影響を与えるにも関わらず、実はメーカーは、一度店頭にものを置いてしまったら、もはやそこから先は、顧客接点は失ってしまいます。

簡単な例でいくと、お醤油のメーカーは売れたという情報は分かって、どのくらいの期間で使い終るかは分からないです。私もかつて食品調査をやっていたのですが、レトルト食品はある程度の期間保存する食品だと思っていたら、結構な世帯が2、3日で食べてしまいます。そういったことは通常、メーカーには分からないんです。

ここでのM2Mの価値は、二番目の評価が下される場面での顧客接点になるということで、そ

れによってリピートに向けた先手のアクションにつながるということです。この辺までくると、実は第2の波と第3の波のちょうど境目の話になってきます。

続いて、今後ビッグデータ活用の主戦場となるのが、第3の波、本来業務やサービスで蓄積されたデータを何か別のサービス・業務に活用するビジネスです。こちらの表に第3の波の事例をまとめてみたので、ご覧下さい。

一番目は先ほど挙げたりモート・メンテナンスです。大学でも同じだと思いますが、例えば複合機をリプレイスするとき、相見積もりを取って、一番安いものを選ぶと思います。メーカー側から見ると相見積もりになってからでは、価格競争になってしまうので、もう遅いんですよね。その前に指名買いしてもらえるように、なんとかして「顧客の囲い込み」をしたいんです。そのために複合機の使い方を提案していくのですが、そういったときに、M2Mから得られる情報を使ってユーザの日頃の使い方を分析していくことになります。

二番目にシステム・ログの分析とありますが、実はシステム・ログは隠れたビッグデータであるにも関わらず、これまであまり注目されることもなく、運用監視以外にほとんど使われてきませんでした。これを上手に活用しようというのが、生命保険会社様向けに提供している分析サービスの事例です。生命保険会社には保険加入の審査業務があり、この分析ではその監査ログを使って審査担当者の業務効率を分析します。

審査業務プロセスは手順化されており、明確です。また、金融庁から指導されていることもあり、詳細な監査ログを取っています。ここでは、「いつ・誰が・何を」というログが、溜まっているので、それを紐付けることでビッグデータ基盤を作り、例えば時間帯別や日別の作業量を見える化することで「ムリ・ムラ・ムダ」をあぶり出します。さらに、人別に処理速度がどのくらい違うのかとか、処理が早い熟練者のやり方は何が違うのかとか、ということ調べて、横展開していくこともできます。

こういったシステムは本来、事務処理の精度を向上するためあるわけで、システム・ログの一次活用もシステムの運用監視です。これを使ってどんな二次活用をしたいかという「ホワイトカラーの生産性向上」です。

三番目にCPFR (collaborative planning, forecasting and replenishment) とありますが、これはメーカーと小売が需要予測を共有して、協調して商品の品揃えや自動受発注をやっていく取り組みです。ここでの一次活用は、在庫の削減と欠品防止です。在庫を厚く持てば欠品はなくなります。が、一歩間違えると在庫過剰になってしまうので、この2つは二律背反なんです。これをなんとかバランスさせることを考えます。

その先にある二次活用は「商談改革」です。小売りのバイヤとメーカーの担当者は人気商品の供給やリベートなど、商談の駆け引きにもの凄いエネルギーを割いています。需要の見通しを共有することで、お互いに握りやすくするというのが商談改革のやりたいことです。

四番目は、医療レセプト分析です。レセプトも本来の一次活用は医療費の適正化なのですが、これを「予防医療」に二次活用していこうということです。

これらの事例を縦に見て下さい。一次活用というのは、みな業務目線の言葉です。業務におけるビッグデータ活用は採算次第です。例えば、自販機ネットワークは、通信コストが下がったので普及しました。一次活用は採算が合えば、どんどん普及していきます。

では二次活用はというと、顧客の囲い込みとか、ホワイトカラーの生産性向上とか、商談改革とか、予防医療とか。これはみな経営目線の言葉です。古くて新しいというか、企業にとっ

ては永遠の課題です。このように業務目線でなくて、経営目線の課題にきちんとビッグデータを使っていこうというのが、第3の波の本質です。

世の中にそういうビッグデータ活用の事例がいくつもあるかということ、実はまだ業務目線の課題しか扱っていないというのが現状です。コマツ様のKOMTRAXがなぜ凄いかということ、リモート・メンテナンスをやるだけでなく、あの情報を元に生産計画を立てるといって、経営目線の課題にまでビッグデータの活用を昇華させているからです。あの情報に部品メーカーがみなぶら下がるといって戦略的な情報システムになっています。

最後は第4の波で、一見価値のない無関係のデータを組み合わせると、新たな価値を創造するビジネスです。おそらく世の経営者の多くは、ビッグデータというとこれを期待していると思います。「その辺にいいデータがたくさんあるのだから、それを組み合わせればなんか付加価値が生まれるんじゃないの?」、みたいな話です。

第3の波までとは大きく違いますよね。なぜかということ、目的がないんです。第3の波までは何をしなければいけないという明確な目的があります。ところが、第4になると、そこにデータがあるからという理由だけで、目的がないんです。したがって、本格的になるのは少し時間が必要だと思っています。というのは、目的のないデータ分析は必ず迷走するからです。20年来ずっとデータ分析コンサルティングをやってきた経験から、これは間違いありません。データがあるから何とかしてくれない?と言われるのが一番困ってしまいます。

とは言いつつも、第4の波の萌芽ともいえるべきものが色々なところで現れています。一つはオープン・データの取り組みです。政府や自治体が公的な立場から集めているデータを民間に開放することで、産業創出に役立っていきこうというような動きがあります。欧米各国ではすでに始まっていて、日本でも総務省が国勢調査とか家計調査とか、色々な調査を開放しています。地図の検索だったり、自転車のシェアのサービスだったりとか、ルート検索だったり、とまだまだプリミティブなレベルですが、英国やフランス、ベルギーではオープン・データを使ったサービスが色々起こってきています。

もう一つの大きな萌芽としては、スマートフォンやタブレットといったスマート端末を使った情報提供サービスがあります。

一番分かりやすいのは、グーグルの「クライシス・レスポンス」です。このサービスでは、グーグル・マップ上に各自動車メーカーから提供された自動車のプローブ・データ（自動車のGPSから得られる移動軌跡情報）を重ねて見せています。震災のときに、この道は通れるとか、この道は通れないということが分かり、大変高い評価を受けました。総務省や経産省でオープン・データというものの機運が高まったのは、このインパクトが大きかったと思います。

その他に、ウェザーニューズ様の「ゲリラ豪雨防衛隊」のように予測困難なゲリラ豪雨の情報をスマホでかき集めて配信したり、クックパット様のように、小売りのID付きPOSデータとレシピを連動して、購入した商品に合ったレシピをレコメンドしたり、あるいは三井住友海上様の「スマ保『運転力』診断」といって、スマートフォンのジャイロ・センサを利用して、ブレーキやアクセルを踏むタイミングから運転力を診断したり、といったサービスが登場しています。

こういったサービスは、みな情報提供型のサービスで、下手をすればただなんですよ。例えお金をもらえたとしても、1アクセス何銭とかいう世界かも知れません。けれども、とにかく何百万人、何千万人という生活者の幅広い利用がビジネスを牽引していくところに大きなポイントがあります。

もう一つのポイントは、技術的には情報を単純に重ね合わせるだけで、必ずしも高度な分析というのは必要ないかもしれないという点です。よくマッシュアップと言うのですが、先ほどのグーグル・クライシス・レスポンスは、地図の上に車の位置情報を重ね合わせているだけです。このように第4の波は技術より知恵比べと捉えていただければいいかなと思います。

皆さん、少し疲れた感じがあるので、ショート・ブレイクを入れたと思います。あるテレビを見ていたら、人工知能で有名な東大の松尾豊先生が、「ビッグデータとはカンブリア爆発だ」と言っていました。皆さんカンブリア紀って知っていますか。私もネット調べた知識ですが、約5億数千年前のカンブリア紀に、生命の種のようなもの（動物門）がそれまでと比較して爆発的に増え、これをカンブリア爆発というそうです。

ちょうど三葉虫が出てきたのもこの時期で、三葉虫は初めて目を持った動物なんだそうです。それまで触覚だけで食べ物を探していた生物が、初めて目を持つことで何が食べ物か分かるようになり、周りの動物を一気に食べるようになった。逆にそれに対抗するために殻を持った生物が増え、結果的に化石が残りやすい種が増えたというのがオチで、それ以前は種が少なかったかという点も必ずしもそうではないようですが…。それはさておき、松尾先生は、「ビッグデータによって企業がそれまでにない目を持った」という意味で、カンブリア爆発だと言ったのだと思います。

私は少し違う見方をしています。こちらのスライドは、ホモ・サピエンスの軌跡を示しています。ホモ・サピエンスは今から10万年前にアフリカ大陸で生まれました。それからユーラシア大陸に渡り、このときはまだネアンデルタール人がヨーロッパで生きていて、戦ったりしているんですね。その後ユーラシア大陸から北米に来て、1万2千年ぐらい前に南米に行きついたそうです。

DNA考古学という分野が確立されて、このダイナミックな動きが初めて分かったのです。DNA考古学では、あちこちに点在するホモ・サピエンスの化石からDNAの差分を比較し、似ている・似ていないという関係から、どこをどう流れていったのかということ調べるそうです。本当に欧米流ですよ。力技でしらみつぶしに調べてしまう。

私はこれから何を言いたかったかという点、実はDNA解析が考古学という業界を変えてしまったのだということです。それまでは多分、古文書を読んだりとか、発掘したりとか、放射性同位元素を調べたりとかとしていたと思うのですが、そんなことをやってもこういうダイナミックな動きは分からないと思います。DNAはビッグデータの権化のようなもので、ビッグデータ解析のスタートはDNA解析です。DNAという異分野の技術を引き下げた人達が参入してきて考古学のやり方を変えてしまったんですよ。

学生の方は実感が湧かないかも知れませんが、「考古学」というものを自分のいる業界だと思って下さい。経営者の立場になってみると、自分たちが、ここで言うDNA解析のような技術を持てば、競争を一網打尽にできるぞと思う反面、異分野の人達が全く新しい技術を引き下げて来たら、完膚なきまでにやられてしまうと思うわけです。グーグルなんかそういうことをやるじゃないですか。これまでとは違うアプローチで自動車の自動運転なんかに参入してくるわけですね。このように新しい技術で業界を席卷できるのではないかという期待、逆にぱっと来られてやられちゃうんじゃないかという恐怖、この期待と恐怖が一体となった焦りが、今の経営者をビッグデータに駆り立てているのかなというのが、私の考えです。皆さんはどう思いますか。

ではここから「ビジネス・アナリティクスの新たな挑戦」という話を進めていきたいと思えます。ビッグデータといっても色々な側面があります。例えば、サービスにどう活用するのか、業務の中でどう効率的に処理するのか、分析をどうするのか、といった具合です。そこで、ここからは情報分析・活用という側面からビッグデータを考えていきます。

ビッグデータとかビジネス・アナリティクス（以下、BAと略す）という言葉が出るずっと前から、似たような概念としてビジネスインテリジェンス（BI: Business Intelligence）という言葉がありました。これは、1990年代にガートナーという調査会社が言い始めてから、IT業界が強力でプロモーションしたこともあり、普及して行きました。

BIとは何かというと、企業の内外に散在するデータを分析して、経営の意思決定に活用する取り組みであったり、ITであったり、方法論であったり、といったものを総称するコンセプトです。

BIのポイントとしては、まずは分析専用のデータベースである、データ・ウェアハウスというものを構築します。これも少し余談になるのですが、企業には様々な業務系の基幹システムがあり、そこにはデータベースがあるのですが、通常は分析には使いません。なぜかというと、そんなものを誰かに触らせて、壊されてたりでもしたら業務が止まってしまうので、特定の人以上には触らせません。そこからデータを抜き取って分析用に別のデータベースを作りましょうというのがデータ・ウェアハウスです。次に、検索とか分析とかレポートなどの機能を持つ、OLAP（online analytical process）ツールと言われるものを活用していきます。

このBIは、2000年代に入ってから急速に浸透しました。なぜかというと、多くの企業に「見える化」という考え方が普及したからです。実際、NTTデータでもBIは大きなビジネスになっています。

ここで、皆さんにYoutubeの動画をお見せします。これは、Youtubeにアップされているtableauという、最近人気のOLAPツールのデモ動画です。OLAPツールとは、グラフや表を簡単に作成したり、見せたりすることができるツールです。ユーザが縦軸、横軸を指定すると、それに合わせてグラフを書いてくれます。最近だと、地図連動の機能があります。例えば日本地図上で都道府県の売上が色分けされていて、東京をクリックすると、23区の売上が色で分かり、さらに練馬区をクリックすると、詳細に地区の売上が分かるといった具合です。これがタブレットにも対応していて、社外からも見ることができます。皆さん、マイクロソフト・エクセルのピボット・テーブルは知っていますか。ピボット・テーブルの機能が拡充した専用製品だと思ってくれればいいかと思います。

最近では、このBIに代わってBAという言葉がよく使われます。北米でガートナーのアナリストから聞いた説明によると、BIに加えて、データマイニングや機械学習のような高度な分析、英語で言うとDeep Analyticsと言うのですけれども、それにパフォーマンス・マネジメント、すなわち企業の業績管理、この3つを合わせたものを、BAと呼んでいるようです。

ビッグデータは、ややもすれば技術の話だというふうに扱われがちです。単なるはやり言葉だとおっしゃる方もたくさんいらっしゃるでしょう。確かにそういう面があることは間違いありません。しかし、ここではビッグデータは、ものの見方、捉え方、考え方というものを変えるパラダイム・シフトだという論陣を張って、皆さんのご批判を受けたいと思います。また、BAもデータ・サイエンティストなる人たちや、オープン・ソースの統計言語であるR言語などがブームになった瞬間、単なるハウツーの論議になってしまうきらいがあります。そこで、BA

もこのパラダイム・シフトに合わせて、新たなチャレンジの時を迎えているというお話をしたいと思います。

まず一番目は何かということ、「今まで目をつぶっていた事実を目を向けていこう」とするパラダイム・シフトです。今ビジネスの世界では、これまで到底無理だとして、目を背けてきた、あるいは見ることを諦めていた事実を積極的に見ていこうとするパラダイム・シフトが起っています。

具体的に見ていきましょう。1990年代から2000年にかけて我々がやってきた従来型のBIとビッグデータ時代のBAを比較して何が違うか。結論から言うと、従来型BIというのは、結局食べやすいデータしか食べてこなかったんです。食べやすいデータとは、要は扱いやすいデータということなのですが、我々BIをやってきた人間からすると、その反省が大いにあります。

分かりやすい例がPOSデータです。POSデータは、典型的な構造化データで、扱いやすいデータではあるのですが、例えば水が売れば、水が売れたという記録しか残りません。でも「本当はコーラが飲みたかったのに、売ってないからしょうがない、水を買おう。」といっても、記録に残るのは水なんですね。

我々マーケティングやデータ分析をやっている人間は、みなそんなことは百も承知なんです。POSデータはそういうものだということを、けれども「それを言っちゃあ、おしまいよ」みたいなところがあり、あえて目をつぶっているんですね。ところが、「いや、それじゃ駄目なんじゃないの、見えるんじゃないの」という大きなうねりと、それに対処する新しいBAの動きが起っています。

例えば、Amazonは、ユーザが商品を探索する履歴を使って、レコメンドに反映させています。POSとは違い、探索履歴を見れば、どういうプロセスで商品を探し、目的の商品にたどり着いたかという文脈が分かります。他にも、JR東日本ウォータービジネス様は自販機に顔画像認識機能を付けました。例えば、私みたいなおじさんが、実は夕方にオレンジ・ジュースを買っていたという事実を発見したということがニュースになっていましたけれども、そういったことは顔画像認識で初めて見えてきたことです。先ほども言ったとおり、これまで自販機というのは、誰が何を買っているのか本当に何も分かりませんでした。これも、うまくやれば実は分かるというよい例です。

最近では、売り場の動線分析も盛んです。技術的にはまだまだ問題があるのですが、顧客がどの棚の前に立って、どんな商品を手取るかということまで認識しようとしています。例えば、アパレル業界を考えてみて下さい。皆さん、服などは興味を持って手にとって、また置いちゃうでしょう。業界関係者は、よく手に取られた商品は何か、あるいは全然見向きもされない商品は何か、ということに非常に興味があります。

こういったことをやろうとすると、画像とか音声、テキストのような非構造化データを扱うことになります。非構造化データは目や耳で判断することは簡単でも、コンピュータ上で分析しようとするときに実に扱いづらいデータです。テキストの平文もそうですが、プログラムで解釈するのは結構面倒です。オーソドックスなテキスト分析ではそのための辞書を作成します。皆さんがツイッターで書く文章も、表記ゆれや同義語がたくさんあります。例えばデーターとデータとか、マネージメントとマネジメントといった表記ゆれや、パソコンとPCのような同義語など、こういったものはみな辞書を作って、直します。画像、音声、テキストの分析というのは、技術の面でいえば情熱の世界なんです。とにかくどれだけ情熱を傾けたかによります。分析を

やる人間からするとできれば避けて通りたいのですが、得てしてそういった情報の中に、価値ある情報が隠れており、それを浮き彫りにしていくことが今のBAに求められています。

こちらの事例は、我々が釜石市でお手伝いをしている漁港のスマート・コミュニティ実験です。ここでは、魚市場や製氷工場、冷凍倉庫、水産加工場などの施設を集めて、個々の施設の活動量や電力消費を見える化したり、全体としての電力消費のピークをシフトさせたりします。ここでも、色々なセンサーのデータを分析して、いままで無理と思われていたことを浮き彫りにするチャレンジしています。

二番目は「どうしてもない時間間隔を埋める」というパラダイム・シフトです。ビジネスの世界に限らず、世の中にはどうしてもない時間の壁というものがあります。私は今朝8時半から打ち合わせがあったのですが、南北線が止まってしまいました。昔だと電車が止まると、状況がはっきりするまでは、迂回すべきか、待つべきか判断に迷ってしまったと思います。こういうのもどうしてもない時間間隔の一つです。今は皆さん何をするかというと、ツイッターで状況を確認するんですね。ツイッターを見ると、もう動き始めたとか、まだまだ止まっていそうだとか、下手するとアナウンスよりも早く分かります。

分かりやすいのは、タミフルですね。インフルエンザが流行ったと分かった後でタミフルを配ってももう遅いです。その他、設備投資もそうですが、例えば日銀短観や消費者物価指数は最短でも一ヶ月の時間が必要で、景気動向を見てから設備投資を判断しようとしても、調査結果がでるまでは待っていなければしょうがありません。

このように、状況が分かるまで時間がかかり、その間手が打てないという時間の壁は結構あります。これに対して、ビッグデータを上手く使えば、この時間間隔を埋められるのではないかというパラダイム・シフトが起こっており、それを克服するBAの挑戦が始まっています。

例えば、グーグルはインフルトレンドという指標を作っており、特定キーワードの検索状況から、地域別にインフルエンザ流行の兆しをつかむことで、どうしてもない時間間隔を見事に埋めています。

実は我々も似たようなことに挑戦しています。NTTデータはツイッター社と日本語ツイッターの再販契約をしており、そこで収集している3年分の全ツイートを分析して、ツイッター・センチメント指標というものを作りました。具体的には、株式に関連するツイートを抽出して、ツイート毎にポジティブ/ネガティブを評価し、センチメント指標を構築しました。それから、この指標と株式指標と相関を色々見ていきました。結論から言うと、日経平均との相関は全くありませんが、日経平均ボラティリティー・インデックス(日経VI)とはそこそこ相関があるということが分かりました。

これを新聞社の人に話したら、「NTTデータ、株価予測」という記事を書かれて、みんなから袋だたきに遭いました。はっきり言いますが、株価予測はできません。もし、これができるのであれば、とっくにこの会社を辞めて独立しています。

ボラティリティーとは、簡単にいうと株価などの分散、振れの大きさ(に関する市場参加者の見通し)です。ボラティリティー・インデックスは別名「恐怖指数」とも言われ、振れの大きさが市場の動揺を表しているとされています。なぜかという、株価は上がる時はじわじわ上がるのですが、落ちるときは一気に落ちてから激しく振れます。したがって、下がる局面の方が上がる局面に比べて分散が大きくなります。

我々が作ったツイッター・センチメント指標と日経VIは、相関が見られ、翌週のボラティリ

ティーともそれなりの相関が見られました。でも、実は予測が当たっているわけではありません。なぜかという、ボラティリティー自体が市場のセンチメント、心理を表す定量的な指標です。一方、ツイッター・センチメント指標はその定性的な指標です。そうすると、震災や世界同時株安のようなインシデントが起こったときに、それがどのくらいのインパクトがあるのかということが、ツイッターでなんとなく分かるんですね。あくまでどのくらい尾を引くかを見ているだけで、決して予測しているわけではありません。

こちらの事例は、ツイッターと購買の関係を分析したものです。ここでは、時事に関心のある人や芸能好きな人など、ツイッター・ユーザをいくつかのグループに分け、新商品に関するコメントを分析しました。また、一部モニタとして登録されている方の購買をみると、積極的にツイートするの方が、あるタイプの商品ではトライアル率が高いというのが分かってきました。

新商品は下手をすると市場に出て1週間で死筋と判断されてしまいます。発売後一定時間経ってから調査したのでは、遅すぎます。我々もこういった分析をすることでどうしようもない時間間隔を埋めようとしています。実際にコンビニでプライベート・ブランドを担当している方とお話をしていると、みな新商品を市場に出す前に、ツイッターでの盛り上がりを見て、売れ行きを予想しているそうです。

このように時間の壁を超えるというのは、今BAに突きつけられている大きな課題ですが、そのときのポイントは、多少の精度というのは犠牲にしても、状況を把握して、今そこにある現実をあぶり出すということです。

三番目に、「複雑な状況でもビッグデータを使ってなんとか合理的な判断をしたい」ということが、今新しいパラダイムとして起きつつあります。IBMは最近ワトソンというシステムを作りましたが、おそらく目指しているのはこういったことだと思います。

複雑な状況での合理的な判断、これを分かりやすい例でいうと、“消費者は昔に比べて格段に進化している”ということです。私がまだ皆さんと同じくらいの年のころは、インターネットなんかありませんでした。電気製品、例えばパーソナル無線を買いたいと思うと、まず一生懸命に秋葉原を回るんですね。この店と決めると、一生懸命、店員さんに値切り交渉をするんです。100円、200円かけてもらうのに、2時間も3時間もかけて粘るんです。今からいうとばかな話ですよ。その時間でバイトした方がよっぽど経済的だと思います。

今は皆さん、絶対にそんなことはやらないですよ。なぜかという、まず値段、価格.COMを調べれば一発ですよ。一番安いのはどこだかすぐ分かる。商品の仕様、メーカーのサイトに行けば分かります。それから、買った人の感想や評判、その人は他に何を買っているのかということは、楽天、Amazonを見れば一発で分かります。こういうのを見て判断すれば、ある程度のことは分かってしまう。さらに最近はそれでも買わないんですね。ショールーミングとって、さらに店頭まで行っちゃうんですね。そこでやっと買うものを決めてから、またネットで買うというプロセスを辿ります。

何を言いたいかというと、思考の順序、プロセスというのがすっかり変わってしまったということです。30年前の私は、限られた情報の中で判断し、その場で価格を聞いてから買っていました。今は最初にネットで広く調べてから実物を見て、またネットで確認してから買う。思考のプロセスが格段に進化していて、結果的に、判断が圧倒的に合理的になっています。30年前の私よりも今の皆さんの方が圧倒的に合理的です。消費者はAmazonや価格.COMにあるビッグデータというものを、知らず知らずのうちにうまく使いこなしています。私はデフレの

原因の一つはこれだと思っているのですが….

企業でも、こういう問題はいっぱいあります。例えば、企業と契約をしようとする、相手の信用調査をします。昔だったら、担保があるとか、資産があるとか、財務状態を見さえすればよかったのですが、今はさすがにそれだけでは世の中は許してくれません。例えば、コンプライアンス面で問題ないかとか、相手企業の出資元や取引先に問題はないかといったことまで考えなくてはなりません。

ある銀行は、関係した相手企業が反社会的勢力と関わりがあっただけで、メディアから批判を受けました。そんなこと言われたって分からないものは分からないですよ。我々 NTT データも、取引相手が問題のある企業とお付き合いがあるだけでも怒られてしまう、そういう世の中になっています。それから、“ブラック企業”のようなネットでの評判も調べないといけないですね。こういう総合的な判断、合理的な判断を短時間に求められるというのが、今の企業の信用調査の現状だと思います。

世の中が複雑になって分からないことだらけなのですが、そういった中でも少しでも合理的に判断したいという状況になっています。これをビッグデータで何とかできるのではないかな、というのが三番目のパラダイム・シフトで、その根幹にBAがあります。

また皆さん少し飽きてきたころでしょうから、ここでショート・ブレイクにしたいと思います。皆さん、最強のチェスプレーヤーは何だかご存知ですか。この中で“ディープ・ブルー”という名前を聞いたことがある人は手を挙げて下さい。誰もいないんですね。先生方はご存じかと思いますが、ディープ・ブルーはIBMが昔、チェス専用にしたスーパー・コンピュータです。今から15～20年ぐらい前ですかね。ディープ・ブルーが、チェスの世界チャンピオンだったガルリ・カスパロフに勝ったんです。これは当時、一大ニュースになり、世界を駆け巡りました。

実は今、最強のチェスプレーヤーは人間かマシンかという論議は終わっています。なぜかという、先ほどのカスパロフさんが、その後何でもありのチェスゲームの団体を作ったのだそうです。そこで2005年にチャンピオンになったのは、3台のPCを駆使した二人のアマチュアでした。

つい最近も将棋ソフトがプロ棋士を破ったなんていう話がありましたが、探索プロセスという意味でいくと単純で、将棋ソフトを並べて多数決で決めるそうです。人工知能 (AI) でいうところの集団学習とも言えいいのか、それも一つの探索プロセスです。

いずれにしても重要なことは、(平凡な人間+平凡なマシン) × よい探索プロセス、というのが、今チェスの世界では最強で、これをAIの業界では知能増幅 (Intelligence Amplification) というそうです。将棋も囲碁も、そのうちコンピュータにやられるでしょうね。複雑な状況での合理的な意思決定という面で、これは面白い事例ですね。こういった技術が、これから先はどうなっていくのか皆さんも注目して下さい。

さて、話を戻して、我々 NTT データも、三番目のパラダイム・シフトに答えるBAとして、人間の知能を増幅して、合理的な意思決定を少しでも支援する、そういったコンセプトを色々な形で表現しようとしています。その一つがトラフィック・コントロールです。ここでは、車載のカーナビから出るプローブ・データを集めて、現状の見える化をします。ここまでは中国もシンガポールもみんなどこでもやっています。我々はさらにマルチ・エージェント・シミュレー

ション (MAS) という技術を使って、渋滞を予測し、信号や車線の制御を考える技術の開発に取り組んでいます。それでは、こちらのデモ動画を見て下さい。

ここでは、100万台規模のプロープ・データを使ったMASを行っています。このシミュレーションは184コア（1つのプロセッサの内部に組み込んだ複数プロセッサの機構をコアという、マルチスレッド処理を実現する）の環境で実行しました。大雑把に184個のマシンが並んでいると思って下さい。これはハイパフォーマンス・コンピューティング (HCP) と言われている技術分野で、HCP環境上でMASという、一個一個の車=エージェントが、意志=ロジックを持って動くシミュレーションをえています。そうすると、100万台なので、山手線の内側を走る車の台数に相当する規模のシミュレーションになり、その中で各車がルールに従って動き回ります。

まず初めに、現実の世界と数値の世界が、できるだけ同じ挙動をするように、一個一個の車の動きをロジックとして定義します。ここで、様々な箇所の渋滞などを見て、全体としての様相が同じようになれば、定義したロジックが妥当だということになります。

次にある場所で事故が発生したとして、その影響を見ています。これを見ると周辺の交差点の方が、渋滞が激しいことが分かります。このように周辺の影響も、全部シミュレートできるというような仕組みです。

では、実際に渋滞を解消するためにはどうしたらよいかという、信号制御の仕方を検討してみましよう。現行の信号は赤、青、黄色があらかじめ定義された秒数で切り替わる時差式という方式で動いています。これを到着する車の台数に合わせてダイナミックに切り替える新しいロジックにした場合、渋滞はどう変わるかと見ています。ある交差点では渋滞が著しく解消している様子が見て取れます。

このデモのポイントは、タイミングや粒度や形式の違うデータを統合した上で、大規模なシミュレーションをやることです。そうすることで、信号制御の仕方、“やり方の良し悪し”の判断をサポートすることで、知能増幅を実現しようとしています。

次のデモ動画は、海運ロジスティクスの事例です。ここでは、インシデントが起こったときにこのままいくとどうなるか、どうすればよいかということについて総合的な判断をサポートします。

船舶にも車と同じように船舶自動識別装置 (AIS: Automatic Identification System) というGPSのようなものがあり、プロープ・データが取得できます。デモでは、これを地図上でモニタリングしながら、各世界のニュース・フィードが確認できます。何か重大なニュースがあった場合には、そのニュースと関連するエリアが強調されます。ここでは日本が強調されているので、地図で日本をクリックすると、東京湾の地図と日本語のニュース・フィードに絞られたフィードが表示されます。脇には気象の情報も表示されます。

今回は、木更津でインシデントが起こったという設定で、これが重要インシデントだとしたときに、東京湾の状況がどうなるかシミュレートします。他の船舶は周辺の港に退避していきます。次に対策を考えていきます。そこで、周辺の港の予約状況と寄港コスト、配送のリードタイムを見ていきます。

上のグラフは周辺の港の予約状況です。実際に他の船舶が退避すると赤いところまで発着枠が埋まってしまっていますが、まだ余裕があることが分かります。左のグラフは周辺の港の寄港コストを示しており、これを見ると清水港も鹿島港もコスト面では遜色ないことが分かります。右のグラフで荷主毎のリードタイムを確認しても、一社以外はなんとか間に合うので、鹿島港

に寄港すればよいとなります。

このデモの一番目のポイントは、世の中の情報には堅い情報と柔らかい情報があり、その中からインシデントの予兆を発見することです。堅い情報とはいうと、例えば船舶のプロブ・データは、センシング・データなので堅い情報です。天候データも堅い情報です。それに対して柔らかい情報とはいうと、事故レポートとか、新聞です。例えば、北朝鮮は日本と国交がなく、なかなか情報が入ってこないで、イスラム圏の新聞を見た方が早いと言われていました。しかし、新聞の言っていることって曖昧ですよ。こういった堅い情報や柔らかい情報を総合的に見て、テキスト処理技術を使ってインシデントの予兆を発見します。

次に現状の見える化と将来の見える化を行い、枠の確保を意思決定していきます。例えばイランが政情不安でシーレーンが止まったときに、海運会社は発着枠を確保しなければなりません。そういったときにどうしたらいいのかということを見ていきます。

このデモの二番目のポイントは、とにかく様々な情報を重ね合わせて大局的な意思決定を支援することです。震災のときもそうでしたが、重要インシデントが起こると対策本部ができます。そういったリスク管理のときは、みな意思決定をするんですね。人間は分からないなら分からないなりに意思決定をします。ITを使って、少しでも合理的な意思決定ができないのかというのが、このデモの考えている知能増幅のコンセプトです。

少し時間が余ったので、「データ活用のポイント」ということについて、事例を交えてお話したいと思います。この事例は医療レセプトの分析です。レセプトを見ると、どんな病気にかかったかということが、直前の病気とあわせて分かります。そこで、データマイニングの技術を使って、何の病気の次に何の病気にかかりやすいかという情報をつなぎ合わせて、ネットワーク図を作りました。我々はこれを重症化モデルと呼んでいます。生活習慣病は併発しやすく、併発すると医療費もうなぎ登りに上がっていきます。こういった関係を明確にしてから、重症化予備軍を見つけ出し、予防医療を推進していきます。

実はこれはすごくもめる資料なんです。なぜかという、重症化モデルは疫学上、何の根拠もありません。けれども、これはある健保組合の全レセプトを使った事実データです。重症化モデルに従うと、生活習慣病はお金がこれだけかかるというのは事実です。やはりこれはこれで迫力があるんですよ。

次はデータテック様、NTT・サービスイノベーション総合研究所と一緒にいった事例を紹介します。データテック様は営業車などに設置する業務用ドライブ・レコーダから「安全の達人」という運転状況のレポートを配信するサービスを提供しています。これは、ユーザ企業様向けに行った、運転者の運転日報（ヒヤリ・ハット・シート）とドライブ・レコーダの情報を重ね合わせた分析です。分析の結果、いくつか面白いことが分かったので、皆さんにもお話ししたいと思います。

「安全の達人」では、注意挙動や危険挙動というものが定義されています。そこで、最初にこれらのデータ間の相関関係を分析したところ、速度と注意挙動には負の相関がある、つまり速度を出せば出すほど注意挙動が減りますという結論が出てしまいました。

さすがにそれはないということで、少し細かく見たのがこちらのグラフです。これは注意挙動と普通のときの速度分布を比較したもので、赤が注意挙動で、青は普通のときです。これを見ていくと、60~70キロあたりで青が赤を上回っています。なぜかという、60~70キロは高

速道路だからです。みな緊張して運転するので、注意挙動が少なくなります。これが悪さをして、スピードを出そうみたいな話になったわけです。むしろ20キロ付近を見て下さい。ここでは赤が青を大きく上回ります。出会い頭の事故が起こる20キロあたりが、実は本当に危ないスピードです。この分析から言えることは、速度を出すと注意挙動が減るのでも、速度を落とせば注意挙動が増えるのでもありません。高速での事故は起これば大事故になりますが、一般道に比べると相対的に事故の起こる率は低く、高速=危ないという単純な思い込みは危ないということです。

また、ヒヤリ・ハット・シートを分析していて、分かったことがあります。IT業界にいと、ヒヤリ・ハットはとにかくなくさない、なんです。情報セキュリティではインシデントの裏側に何百ものヒヤリ・ハットがあると言われます。とにかくヒヤリ・ハットをゼロにしろということなんですが、こと車に関しては、ちょっと事情が違うようです。

こちらのグラフを見て下さい。ヒヤリ・ハットが増えると注意挙動が減っているんですね。どういうことかという、ヒヤリ・ハット・シートのような取り組みをやっていくと、運転者の感度がどんどん上がっていきます。そうすると、今までヒヤリともしてなかったことに、ヒヤリとするようになるんです。実はヒヤリ・ハットが増えるというのは、悪いことじゃないみたいですね。減らすべきは注意挙動であって、ヒヤリ・ハットは別に減らさなくてもいいのではないかという結論が出てきます。これは結構面白い話です。

ビッグデータの活用のポイントというのは何か。一番目は「事実データで思い込みを排除する」ということです。先ほどの重症化モデルは疫学的には根拠がなくても、やはり事実データとしての迫力があるんですね。みなそれは否定できません。また、単純に高速=危ない、ヒヤリ・ハットは駄目と考えるのは思い込みに過ぎません。事実データを丹念に見ていくことでこういった思い込みを排除することができます。

最後は昔NTTドコモで久保田さんという方がなさった分析です。NTTドコモでは契約者の年齢分布を見ると、20代と40代で2つのピークがあるそうです。なぜだか分かりますか。単純に10代の子どもは親が契約するからです。久保田さんたちは悩んだ末、利用者のマクロな傾向を分析して、40~50代でも若者のような特徴をもつユーザーを“みなし若者”としました。ある若者向けサイトでも40代のアクセスが多くなっていたそうですが、みなし若者を考慮して見ると、実につじつまが合ったそうです。

みなし若者がなぜ重要かという、家族契約の解約はみなし若者、すなわち子どもが発端になるからです。周囲の友だちがみな他社の携帯を使っていると、そこから解約が起こるので、みなし若者の動きを特に注視していたそうです。

ここで言いたいことは「事実は必ずしも真実ではない」ということです。先ほどのサイトで40代のアクセスが多いのは事実です。そうすると、これからは私みたいな40代のおじさん世代を攻めるべきとなってしまいます。しかし、これは経営上の大きな判断ミスですよ。真実は神のみぞ知る世界かも知れませんが、仮説検証を繰り返して事実の向こう側にある真実に迫ることがビッグデータ活用の二番目のポイントとなるということ、締め言葉にさせていただきます。ちょうど時間が来ましたので、この辺で終わりにしたいと思います。

(講演終了)

<司会> 質問の時間を取っておりますので、質問があれば受けたいと思います。それではどなたか、ここ、分かりにくかったのもう一回お願いしますでも結構です。あるいは、もっと何かさらに発展した質問でも結構です。いつも言っているとおり、質問する練習という側面があります。

どうですか。先ほど、ディーブ・ブルーを知っている人はたくさんいたと思うんだけど、急に手を挙げてと言われた感じで手を挙げないと同じで、何か質問しにくいかな。あまり質問しないとこっちから行くよ。

じゃあ、手始めに、どうぞ。

<質疑応答>

<質問者1> 今日は貴重なお話をありがとうございました。一つ質問ということですが、ビッグデータとか情報がすごく多くなってくると思うんですけど、多少の精度は犠牲にしても状況を知ることが、パラダイム・シフトのところであったと思います。それってなんか速度も重視される中で、そういった情報の多さを企業の中で処理していかなくちゃいけなくなると思うんですけど、それが企業にとって、速度を重視すべきであったりとか、確実性を重視すべきだったりとか、難しいところだと思うんです。そういった点で、ビッグデータは速度についてはちょっとマイナスの面があると思うんです。情報が多すぎるのが、そういった点はどういう議論がなされているのかなと思ったんですけど。

<中川> そうですね。私は、経営にとってスピードは本当に命だと思います。今まで情報分析・活用という分野に関しては、特にBIと言われている時代はバッチ処理が中心で、正直言うとあまりスピードということを考えていなかったと思います。やはりそこに目が本当にいくようになってきていると思います。

これは自分が企業の中において思うのですが、自分たちも気を付けてないと、スピードはどんどん遅くなります。トップの人ほどせっかちなので、スピードというものを求めるんですね。ビッグデータと速度、スピードの経営を結びつけている語人は、世の中にあまりいません。そういう意味では、すごくいい観点をしているなと思います。スピードというと、すぐリアルタイムのことは完全に頭が行ってしまいますが、リアルタイムでなくても他社よりも少しでも早ければ価値なんです。そういうスピードという感覚でもう一度捉えなおす時期が、今まさに来ていると思います。ストレートな答えになってないかも知れませんが、よろしいですか。

<司会> 聞いた以上のことを言っていたら、他にもどなたかいかがですか。せっかく前に座っているあなたはどうですか。

<質問者2> 技術的な特異点を越えるのに、ビッグデータが鍵になるとは思いますか。人工知能が人間より頭がよくなるときに、そのときに、ビッグデータが鍵になるとは思いますか。

<中川> 要素の一つであることは間違いありません。しかし、私はビッグデータといっても所詮データのことを指しているのだから、手段でしかないと思っています。その前提にたつて、鍵ですかと聞かれると、要素の一つではあるけれども、鍵はやっぱりもっと別にあるような気がしますね。

<質問者2> 僕も同じ意見です。

<中川> ビッグデータやBAもやはりそうだと思うんですけど、技術の面から捉えて、まだ制御しきれている技術という気はしていません。セキュリティや個人情報とかも含めて、もう手

なずけた技術だとは思いません。原子力なんかも同じだと思うんですけど、気を付けないと暴走してしまう技術だと思います。

<質問者3> スマートフォンとかパソコンとか、全然使わないような人からデータって取れるんですか。

<中川> ヘルスケア・デバイスなんかは、中年の人は、そのうちに否でも応でも付けられてしまうかも知れません。

<質問者3> マーケティングとかでデータを使うとなったときに、年配の方向けに要件をつくるときに、ビッグデータは多分少ないと思うんです。そういうときって、もう諦めて、感覚でつくるのかとか。

<中川> 少なくともある程度年配層、特に後期高齢と言われる人がみなパソコンを使いこなす姿はなかなか想像しづらいですね。一方で、ある年齢まではそれなりにパソコンも使うし、データも取れると思います。これはむしろマーケティング自体の問題だと思っています。マーケティング・リサーチをやっていると、65歳以上を老人と括ってしまいます。でも、今はもうそんな時代じゃないですよ。65歳より上の年齢層は分割すべきだと思います。下手すれば75歳ぐらいでパソコン使う人はいますからね。それなのにマーケティングの人たちは、今まではそれでよいとして、65歳以上を老人にしちゃっていたんです。今はむしろ、それを見なおす時代に来ているなと思います。

<質問者3> ありがとうございます。

<司会> カメラのデータなんかもそのために使えないんですか。

<中川> 画像認識の素材としては使えると思います。顔画像認識は、だいぶ身近な技術になりつつありますね。3.11で米軍がかなりの予算をつぎ込んだので、顔画像認識がかなりポピュラーになったそうです。

<司会> うちの学生は、なんか恥ずかしがり屋なんですけど、こうやって聞くと、どんどん出てくるんですけれども、そろそろお時間もいっぱいになりましたので、ここで終えたいと思います。今日はどうも本当にありがとうございました。

[なかがわ けいいちろう 株式会社NTTデータ数理システム取締役]

ビッグデータとビジネス・アナリティクス

2014年11月12日
株式会社NTTデータ数理システム
中川 慶一郎

NTT DATA

アジェンダ NTT DATA

- ビッグデータとは
- ビッグデータ活用の発展
- ビジネス・アナリティクスの新たな挑戦

本資料に掲載の会社名、製品名またはサービス名は、それぞれ各社の商標または登録商標です。

ビッグデータとは

ビッグデータの時代 NTT DATA

- Facebook**: 27億件/日 (Facebookサイト内外で毎日クリックされるLike件数)
- Twitter**: 4億件/日 (1日の投稿数)
- Google**: 10億件/日 (1日にウェブサイトで見られる検索回数)
- Amazon**: 3680万件/日 (ピークの注文件数)
- LINE**: 70億件/日 (1日に送られるメッセージ数)
- eBay**: 50テラバイト/日 (1日にウェブサイト上で発生し、DWHに蓄積されるデータ量)
- 楽天**: 62万件/日 (1日の注文数)

ビッグデータ活用への期待 NTT DATA

● 社会・企業・個人に新たな価値を提供

社会 <ul style="list-style-type: none"> ● 電力・交通 ● 農業、資源エネルギー ● 環境・防災・保安 ● インフラクラシス 	企業 <ul style="list-style-type: none"> ● マーケティング ● ファイナンス ● 業務プロセス改善 ● コンプライアンス 	個人 <ul style="list-style-type: none"> ● 便利 ● 安心/安全 ● 省エネ ● ヘルスケア
---	---	---

ビッグデータの蓄積・分析/活用

センシング・データ ライフログ・データ

RFID (モノの運搬) HEMS (エネルギー消費履歴) 交通系ICカード (電車の移動履歴) EC (商品探検プロセス)

各種センサ (機器や構造物の状態) GPS (車や人の位置) インターネットバンキング (ATM 法注履歴) SNSやブログ (個人が発信する多様なデータ)

リモートセンシング (環境、災害) ヘルスケアデバイス (健康状態)

ビッグデータの3つの特徴 NTT DATA

Volume, Variety, Velocity

“ビッグデータ”とは、どういうことか? NTT DATA

● ビッグデータを捉える視点

多様なデータが集まる 色々な所からデータが集まる 絶え間なくデータが集まる

新しいタイプの情報分析技術が求められる

従来のDBでは貯めきれない データ発生に処理が追いつかない

“ビッグデータ”を支える技術 NTT DATA

IT基盤 <ul style="list-style-type: none"> インフラ インメモリ DB Hadoop NoSQL DWH アプライアンス 通信ネットワーク データ処理 <p>大規模かつ高速な蓄積・検索・処理</p>	分析技術 <ul style="list-style-type: none"> データマイニング 機械学習 自然言語処理 セマンティック検索 <p>量を質に転化させる</p>
--	--

例えば・・・自販機オンライン・サービス

NTT DATA

● 自販機から随時データを収集し、分析活用

ソリューション・イメージ

状況把握
 実売本数 滞留時間 高品別売上内訳
 在庫本数 故障情報

データセンタ

配車

高品A
高品B
高品C
高品D

波及効果

販売機会ロスの削減
在庫期間の短縮化／鮮度維持
 配送の効率化
ロケーション開拓

暗黙知的なノウハウの形式知化による継承

例えば・・・リモートメンテナンス(1)

NTT DATA

● アフター・サービスにおけるセンシング・データの活用

モニタリング
 故障の予兆発見
故障個所の把握

迅速な故障対応
 最寄りの代理店に通知し
修理備品を事前に準備

アフター・サービスで顧客経験価値を高める

例えば・・・リモートメンテナンス(2)

NTT DATA

● 消費者は2度評価する

企業	消費者	
商品選び	商品選び	一番目の評価
販売・契約	購買・契約	購買まで
アフターサービス	利便性の体験	二番目の評価
	日常への組み込み	購買されてから
	ロイヤルティ確立	

M2Mが顧客との接点となる

ビッグデータ活用によるビジネス変革

NTT DATA

事例	1次活用 (業務課題)	2次活用 (経営課題)
機器の稼働状況のモニタリング	リモートメンテナンス	先手の営業による顧客の囲い込み
システム・ログ分析による業務効率化	事務処理の精度向上	ホワイトカラーの生産性向上
POSデータ活用によるCPFR [®] と自動発発注	在庫削減 欠品防止	需要の見通し共有による商談改革
医療レセプトによる重症化予備群抽出	保険料支払の適正化	予防医療の推進による医療費削減

欧州各国のオープンデータ・ポータルサイト

NTT DATA

● 公共データを公開するポータルサイトの整備

英国 <http://data.gov.uk> フランス <http://data.gouv.fr> ハリ市 <http://opendata.paris.fr> ヘルリン <http://daten.berlin.de>

世界のオープンデータサイト開設国

- 世界30ヵ国が開発済み
- OECD、国連など国際機関もデータポータルを開設。EUも開設予定。

公共データを活用した民間サービスの例

NTT DATA

Multimap <ul style="list-style-type: none"> 郵便番号から地図を検索するサービス イギリス政府のデータ(有償)を活用 2人で検索→120人に拡大、マイクソフトへ売却 	ITO World社の交通案内サービス <ul style="list-style-type: none"> ロンドンのリアルタイム交通データを活用 地下鉄の運行情報から、電車が止まった際の代替ルートや所要時間を提供
Local Authority Dashboard <ul style="list-style-type: none"> コミュニティ・地方自治体が保有する自治体の情報を提供 様々な指標を箇の平列と比較して表示 Linked Data フォーマットでデータが提供されており、SPARQLエンドポイントも提供 	London Cycle Hire LIVE <ul style="list-style-type: none"> ロンドンではすでに自転車のシェアリングシステムがある(通称ポリス・バイク) このレンタル用自転車の置き換えをリアルタイムで検知できるサービス

第4の波の萌芽

NTT DATA

● スマホ等を使った情報提供サービス

	事例
1	「Google Crisis Response」 自動車・通行実績情報マップ
2	ウェザーニュース様 「ゲリラ雷雨防衛隊」
3	クックパッド様 「レシピ検索+ID付POS連動」
4	三井住友海上様 「スマ保・運転力診断」

生活者の幅広い利用がビジネスを牽引する

ビジネス・アナリティクスの新たな挑戦

【参考】ビジネスインテリジェンスとは NTT DATA

定義 Business Intelligence

- 企業内外に散在する膨大なデータを分析して、経営意思決定に活用する取組みを総称するコンセプト

分析用DBであるデータウェアハウス(DWH)を構築する

「データ検索」、「分析」、「レポート」などの機能を持つ「BIツール」を活用する

↓

「見える化」の普及により、広く浸透

BIからBAへ NTT DATA

ビジネス・アナリティクス

ビジネスインテリジェンス

高度な分析

パフォーマンス管理

ビッグデータを取り巻くキーワード NTT DATA

Hadoop NoSQL Volume Variety Velocity

インメモリDB 複合イベント処理 技術

データサイエンティスト ビジュアライゼーション How To

↓

もの見方、捉え方、考え方は変わったか？

ビッグデータがもたらすパラダイムシフト NTT DATA

目をつぶっていた事実を目を向ける

↓

どうしようもない時間間隔を埋める

↓

複雑な状況でも合理的に判断する

BI vs. BA...何が違うのか？ NTT DATA

● 従来型BIは、食べやすいデータしか食べてこなかった。

従来型のBI **ビジネス・アナリティクス**

【POSデータ分析の例】 **購買層の認識**

商品探索履歴 水 茶 茶 茶

出入り口 若い女性

店舗内導線の把握

↓

扱いづらいデータに価値のある情報が隠れている

どうしようもない時間の壁 NTT DATA

電車が事故! どのくらい足止めされるか見当がつかない

投資をするべきか? 景気判断材料である消費者物価指数や日銀短観は数か月かかる

タミフルをどう備蓄・配布? どの地域でどのくらいの人が感染しているか把握できない

新商品発売後のアクション! 発売後1週間で商品の成否を判断されてしまう

↓

多少の正確さは犠牲にしても、状況を把握したい

【参考】経済の体温を知るには!? NTT DATA

内閣府 経済産業省 消費者物価指数 (CPI)

内閣府 四半期GDP速報

内閣府 気象動向指致 気象庁気圧計

財務省 財務省 国際収支(確報) 国際収支(速報)

日本統計 日本統計 日銀統計 日銀短観

1年後 四半期後 2ヶ月後 1か月後 1週間後 1日後 取引時

リアルタイム性 (低) → リアルタイム性 (高)

↑ 向上か? ↓ 下向か?

↓ どうしようもない時間

【事例】Twitterセンチメント指標 NTT DATA

● 市場のセンチメント(市場心理)を定量化

アベノミクスで最近日経平均株価が急上昇。中国金融、円高に警戒感、後場売りで押し下げ反落。

Tweet データ抽出 → Twitter センチメント指標値

公開情報 → 株式指標

センチメント分析 → 関連性検証


イケイ(ポジティブ) 反落(ネガティブ)

↓

金融取引で有効な判断指標となる可能性

進化する消費者行動 NTT DATA

● 消費者の“思考”プロセスが格段に進化している



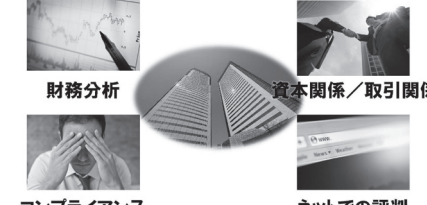
価格比較サイト
ECサイトの
カスタマ・レビュー

値段
製品の感想
製品の仕様
他の人は何を買っているか

消費者の判断は合理的になっている

取引企業を審査をする NTT DATA

● 財務以外にも多くのリスク要因が存在する



財務分析
資本関係／取引関係
コンプライアンス
ネットでの評判

さまざまな角度から網羅的に判断する

おわりに NTT DATA

- ビッグデータとは
社会・企業・個人に新たな価値を提供する
- ビッグデータ活用の発展
ビッグデータ活用で経営目線の課題を解く
- ビジネス・アナリティクスの新たな挑戦
ビッグデータはもの見方を変えるパラダイムシフト



NTT DATA
Global IT Innovator