

加速するスマートシティ化とビッグデータ活用

識者が解説！

ビッグデータの可能性

「情報を制する者は世界を制する」——この言葉がいよいよ現実味を帯びてきた。猛烈な勢いで増加する大量のデータを収集・分析して、実社会にどう適応させていくか、いわゆる「ビッグデータ」活用の新たな挑戦が始まっている。

近年急速に脚光を浴びているビッグデータ。その要因のひとつとなった象徴的な出来事が7年前に起きた。米国立標準技術研究所後援の機械翻訳コンテストが開催されたときのことである。難解かつ独特なアラビア語の大量の記事をいかに正確に英文記事へと翻訳できるか、その翻訳精度の高さを競う大会だが、この年、圧倒的勝利を収めたのは初参戦したグーグル。勝因は、従来のアプローチをひっくり返す驚くべき手法だった。

「彼らはアラビア語がわかるメンバーが1人もいなかったが、自然言語処理モデルやア

ルゴリズムを作ることもしなかった。何をしたかという点、単にグーグルのWeb検索で収集した1兆2000億にのぼる語句のデータに当てただけ。大量のデータを適用したことにより、精度の高い翻訳分析ができてしまった。大量データを使えば使うほど価値を生み出し、できなかつたことができるようになることが証明された結果、ビッグデータという概念が生まれ世界が注目するようになった」。

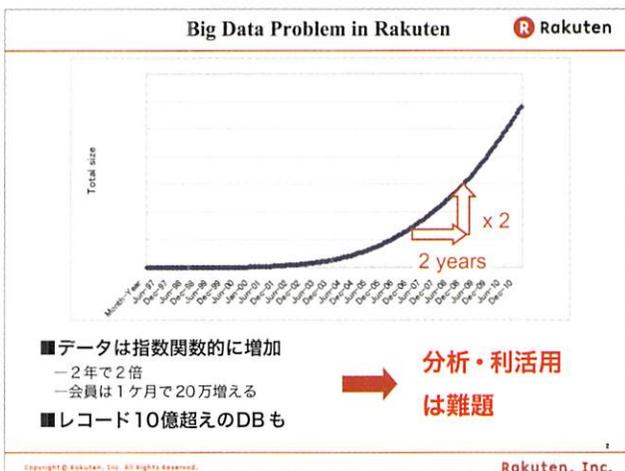
そう説明するのは、楽天技術研究所の森正弥所長だ。

eコマースの精度を高める楽天の取組み



楽天市場、楽天トラベルなど50以上の事業を展開し、2012年6月現在7800万人の会員数を抱える楽天は、購入履歴や宿泊情報など様々な大量のデータをデータベースに蓄積してきた。07年頃より、これらのデータを、ショッピングなどにおけるレコメンデーションやサーチといったサービスに活用している。

「大量データを使えば使うほど、組み合わせ合わせるほど、精度が向上することが



楽天が取り扱うビッグデータの量(資料提供:楽天)

わかり、特にeコマースにおいては売り上げに非常に大きなインパクトをもたらした」。

こうした大量のデータをビジネスに応用展開し、全体のクオリティを上げていくところが、楽天のビッグデータの定義だと森氏は強調する。事例を積み重ねる中で、ビッグデータは全社をあげて取り組むべき課題だという認識を新たに、今年2月「ビッグデータ部」を創設した。

従来の構造化データに加え、

楽天技術研究所
森 正弥 所長



多種多様な店舗と多種多様な人たちをつなぐのがネットの基本。そこにビッグデータを取り入れることで、さらにマッチングを最適化できる

スマートフォンなどの普及で利用範囲が広がるにつれ、非構造化データが爆発的に増えている。楽天は、今までデータベースに入れていなかったオンラインザクシヨシログ、検索ログ、アクセスログ、システムログを活用する際には、大規模データの分散処理を可能にするオープンソースのソフトウェアElasticsearchなどのツールを使ってアプローチしている。

さらに2月、「スマートデバイス部」も新設した。スマートデバイス向けのアプリケーションは数十種類あるが、現状では、そこで集積したデータは活用せず捨てていた。その有効利用対策として、ビッグデータ部のデータプラットフォームと連携させる基盤を年内に構築する予定だ。森氏は

「多種多様な店舗と多種多様な人たちをつなぐのがネットの基本。そこにビッグデータを取り入れることでさらにマッチングを最適化できる」とポテンシャルを強調する。

今後のカギを握る 個人情報保護法

ビッグデータの秘める可能性は研究者も期待を寄せる。東京大学大学院情報理工学系研究科の江崎浩教授は、

「00年頃から研究開発してきたIPv6は、今というビッグデータそのものである」と

と解説する。例えば当時、名古屋で行った実証実験では、約1500台のタクシーにセンサーを取り付け、ネットにつないで情報データを収集・分析した。その結果、道路の混雑情報は、国より精細でリアルタイムなデータが得られた。同時に取得した雨量センサーのデータが思わぬところで役立つこともわかった。

「クルマのセンサーは、雨量計が検出しないような雨でも感知したので、天気を取取りできるデータも得られた。これはコンビニ等のサービス業にとっては非常に欲しい情報。実は、これこそがビッグデータの定義だ。少しの情報や精度が不足している情報で

は意味がないが、大量の情報が集まってはじめて意味を持つ、そしてある目的で取得した情報と別の目的で取得した情報がどこかで統合化・解析されて別の意味を持つ、この2つが大きなポイントになる」と江崎氏は分析する。

このようにフィールドで起こっていることを実測定し、真実をその中から抽出していくことを江崎氏は「エビデンス・ベースト・リサーチ」と呼び、ビッグデータの利活用で重要視している。また、実フィールドでの経験がなければデータも活かされないのが、「エクスペリエンス・ベースト・リサーチ」も必要になると解く。

「膨大なデータの中にある隠れた本質的なものを発見していくことが求められる。その集めたデータをどう組み合わせるかの目的に使用できるかが勝負になるだろう」。

従って、今後はどれだけデータを収集できるかが大事なポイントになる。だが、その前に大きな障壁が立ちほだかつていると江崎氏は危機感を募ら

せる。個人情報保護法だ。

「ある目的以外にデータを使用することが不可能と解釈している日本の企業は、過剰なほど真面目に遵守している。だが本来、保護法は、貴重な情報はイノベーションの基になるので、プライバシーを守る機能処理を施した上で、安全にみんなが使用できる環境を整えることが基本のはずだ」。

ビッグデータに関していえば、アプローチを間違えたとイノベーターな発見につながらないと指摘。実は、これこそが日本が今抱える問題である。他国にも日本と同じ個人情報保護法があるが、その壁を乗り越えているんな発見を生んでいる。たとえばアメリカのストリートビューが最たる例で、

「要は実行上の問題。経営者がきちんとリスクを取れるかどうかにもかかってくる」。

楽天では、どんな対策をとっているのだろうか。森氏は、「そもそも個人情報はビッグデータと呼んでおらず、完全に分けている。新設した部でも個人情報には触れることができ

早稲田大学
理工学術院
山名早人教授



ビッグデータ活用のポイント

- 個別の情報を統合するために、データをオープン化
- 膨大なデータを収集し、少なくとも数年分を蓄積
- オープンソースのソフトウェア “Hadoop”等を使ってデータを解析
- データの中に隠された有益な情報を抽出

期待される成果

- eコマースの売り上げ増
- リアルタイムで精細な道路や天気の情報を獲得
- ゲノムデータの分析で生活習慣を含めた病気の予防 …etc

現状の課題

- 個人情報保護法の捉え方
- 人材不足

プライバシー保護の観点からいっても、今後、法の整備が必要になる

「ない設計を構築済みだ」と説明する。

一方、データマイニングなどを研究している早稲田大学理工学術院の山名早人教授は、「今の個人情報保護法は、名前や電話番号がリンクしなければOKという定義だが、現在の技術をもってすれば複数のデータから個人を特定することは不可能ではない。プライバシー保護の観点からいっても、今後、法の整備が必要になるだろう」と課題を示す。

そうしたセンシティブな問題をほらみつも、海外に目を向ければ、ビッグデータの利活用に向けて国家を挙げたプロジェクトが進行中だ。アメリカでは3月、ビッグデータプロジェクトに2億ドルを投じることを発表した。その中でも、1000人のゲノムデータをアマゾンのクラウドに載せ無料での利用を可能とし、研究活性化を図る「NIHプロジェクト」に山名氏は注目している。膨大な量が集まればそこからいろんなパターンを見出せ、生活習慣を含めた病気の予防に活かせるからだ。

翻って日本では、データを収集・分析し、実社会へ適応すべく企業がしのぎを削るが、「今のビジネスで扱っているデータは多くてテラバイト。ビッグデータでは、その3けた上のペタ、さらに3けた上のエクサになった時に何ができるかが重要」と山名氏は現状を語る。では、企業は今から何を始めた方がいいのか。「まずはデータ蓄積すること。あまり古いデータは参考にならないが少なくとも数年間ストックするべきだ。さらに目的に合わせてlog等のものでオープンソースを用いて解析すること」(山名氏)

その際には人材を含めた投資が必要不可欠となる。データの扱いに関しては、江崎氏は、いかに効率的にアウトソースできるかが1つの勝負となるとみている。

スマートシティ化に向けて進む研究開発

こうしたビッグデータを巡る動きは、とりわけ「スマートシティ」の利活用で注目が集まる。ただ、スマートグリッドがそうであるように、「各フィールドで集まってきた情報をどう活かすがまだ見えない」と江崎氏は指摘する。とはいっても、今から企業がとるべき対策はある。「スマートシティは言うなれば情報の海だ。そこに手をつかめる状態にしておくことが大事」。

この情報の海には重要なポイントが2つ潜んでいる。1つは情報を囲いこまない点。「国境の海を作ってしまったら、入ってくる情報も限られたものになってしまう」

もう1つは、データのオープン化だ。セバレート・フラグメントしている情報が統合できるよう標準化されているかどうか勝負を決める。「データ統合はコストが小さいほど成功する確率が高くなる。従って初めからインターフェイスで使えるようにしておくことが重要」。

そうした観点で東大が開発したオープン通信規格「IEEE 8088」は、あらゆる産業セグメントにまたがって使用できるように設計されている。まさに全産業がステークホルダーになると予想されるスマートシティを見据えた研究開発が進んでいるのである。

オープン通信規格の開発を巡っては、会津若松市が「スマートシティ推進協議会」を立ち上げ、積極的にスマートグリッド化を推進。スマートメータ

ーの標準化などを目指す。

このように、企業だけではなく、社会インフラとしても大きな効果が期待できるビッグデータ。しかし、新たな課題も浮上している。喫緊の問題が人材不足だ。

楽天の森氏は、

「求められるスキルレベルは非常に高い。統計知識、機械学習、データ分析の専門知識に加え、ビジネスセンスも必要になる。そういう人材確保が難しい」

と厳しい現状を説明。既に世界レベルで人材の奪い合いが起きている状況だ。その対策の1つとして、同社はビッグデータ部を海外拠点にも設置することを模索している。

東大の江崎氏は、「アカデミックな研究と違い、業務に引張られる企業では自由なセンスがなかなか見せられないのは事実。ただ、ビッグデータは基本的に、過去インターネットでやってきたことをもう一回実践する形。ネットがデータに変わるだけの話」と持論を展開する。さらに、ロングテール理

論の80対20の法則を挙げ、

「ビッグプレーヤーが市場全体の2割の情報を集めているとすれば、残りの8割は市場に眠っている。そのすべてを集めればフェイスブックもグーグルもひっくり返せてしまっ」

と説明、ビジネスチャンスは平等に広がっていると無限の可能性に触れた。

膨大な情報をどう制し、正しい分析のもとに実社会の利活用につなげていくか、ビッグデータを巡る戦いの火ぶたは切られている。

東京大学大学院
情報理工学系研究科
江崎浩教授



膨大なデータの中にある隠れた本質的なものを発見していくことが求められる。
その集めたデータをどう組み合わせるかが勝負になるだろう

111th
Anniversary

MEDIA of 感動伝達

Made by **Mitsumura**

光村印刷が提供するの、製版・印刷技術をもとにしたさまざまな製品。新聞、ビジネスフォーム、カタログ、美術書などの紙製品はもちろん、IT化を支えるタッチパネルやエレクトロニクス製品にいたるまで、社会に感動をひろげる多様なメディアを、心を込めてお届けします。

 光村印刷株式会社 〒151-8567 東京都品川区大崎1丁目15番9号 <http://www.mitsumura.co.jp>