NTT DATA Technology Foresight 2013

戦略的なデータ収集と分析

インターネット通販を始め多くのサービスで顧客の購買履歴やサービス利用履歴を活用したレコメンドサービスが提供されている。また、最近では選挙戦略の立案やダウ平均株価の予測など、これまでなかった分野でデータを活用して高い成果を得る事例が出てきている。今後は、ログデータや環境データ、ノウハウ情報などを戦略的に収集し分析することが競争力の源泉となるだろう。本稿では、技術トレンドのひとつである“戦略的なデータ収集と分析”と、それに対する当社のアプローチを紹介する。

競争力の変遷

産業革命以来、工業化、商業化、サービス化、情報化の進展により、社会は大きく発展してきた。製品が市場の中心だった時代には、大量生産によりいいモノをより安く市場に提供する労働集約型のビジネスが全盛だったが、その後、製品産業からサービス産業へと時代は進み、IT 化による自動化やサービスの高度化が競争力の中核をなす時代となった。

現在では、既に多くの企業が IT 化による自動化・高度化を実現しており、新たな競争力の源泉として、知識やデータを戦略的に活用してふかん的な意思決定を迅速に行う知識集約型のビジネスが求められる時代となっている。

図 1：競争力の変遷

進化の方向性

知識集約の時代においては、様々な現象や環境、行動や知見等がデータとして収集・蓄積され、これらのデータが膨大なコンピュータパワー／リソースを使って解析されることによって、新たな知識や知見が発見、獲得できるようになる。これにより、従来は限定された情報と人的経験に基づき行われていた意思決定を、客観的なデータに基づいてふかん的、且つ、高精度に行うことができるようになる。こうした状況を支えるのは、“(A) 分析サイクルの短縮”と、“(B) 多種多様なデータを活用”する技術である。「戦略的なデータ収集と分析」は、大きくこの二つの軸で進化すると考えられる。

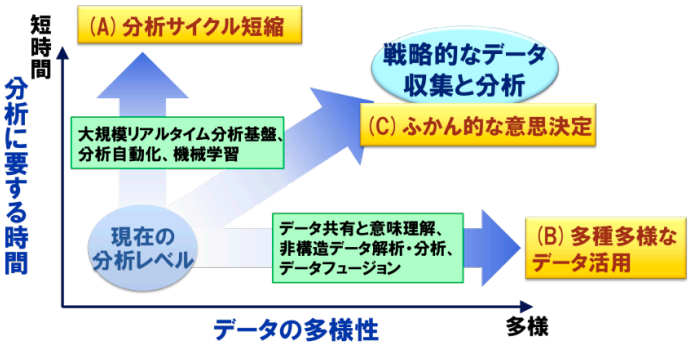

図 2：進化の方向性

(A) 分析サイクルの短縮

ひとつは、データ分析に要する時間を短縮化する軸である。増え続ける多量なデータを短時間に分析する技術の必要性は高い。この軸に該当する技術としては、大規模リアルタイム分析基盤、分析自動化、機械学習などが挙げられる。

(B) 多種多様なデータ活用

もうひとつは、多種多様なデータを活用する軸である。種類やフォーマットが異なるデータも対

象となってくる。この軸に該当する技術としては、データ共有と意味理解、非構造データ解析・分析、データフュージョンなどが挙げられる。

NTT データのアプローチ

NTT データでは、「戦略的なデータ収集と分析」の進化の方向性である 2 軸に該当する技術について研究開発を進めると共に、様々なサービスに対して実際にビジネス活用を始めている。特にビッグデータ分析に基づく業務改善や、新たな価値を創出する取り組みは既に多数の実績がある。具体的にどのような事例があるのか紹介する。

1. エネルギーマネジメント

「(A) 分析サイクルの短縮」に該当する事例。近年、日本では電力自由化に関する議論が進んでおり、エネルギーマネジメントの一つとして、電気などのエネルギー使用状況を把握し、電力需給量を予測・制御することで電力需給バランスの最適化を行うことが求められている。

NTTデータでは、アイルランドをモデルケースとして、電力市場に関わるビッグデータを分析し、エネルギーマネジメントに活用する研究[1]をアイルランドの研究機関と共同で実施している。

電力自由化が進んでいるアイルランドでは、電力卸取引市場にて国内の総発電量のほとんどが取引されており、電力供給業者が卸取引市場に売電し、電力小売事業者が卸取引市場から電力を調達して電力消費家に販売するビジネス構造が既に構築されている。

このビジネス構造の中で、本事例では電力小売事業者に着目した。一般家庭や企業などの電力消費家と直接つながりを持つ電力小売事業者は、電力需給バランスを調整する上で重要な役割を果たす立場として期待されている。国策として風力発電の導入を推進しているアイルランドでは、風力発電量の増加に伴い気候等の外部要因によって供給量が大きく変動するため需給バランスの調整の必要性が高い。電力小売事業者が、電力需要の変化を高精度に予測し、例えば、電力需要が高い時間帯に電力価格を高く設定する、あるいは電力需要が低い時間帯に価格を安くするといった制御を行うことができれば、電力消費変動は抑えられ電力需給バランスが最適化されると考えられる。

共同研究では、電力需給バランス最適化の第一段階として、電力小売価格の設定に影響する要因である電力需要量、電力供給量、電力市場取引価格を予測するロジックを構築した。

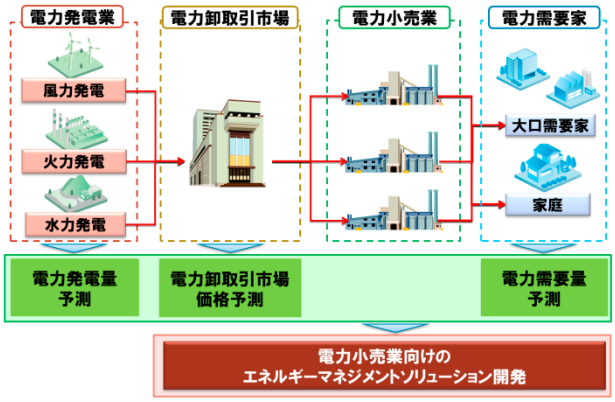

図 3：需給バランスの最適化の実現

本事例では、15 分単位で 24 時間先まで逐次予測し、実績データとの乖離が発生した場合に予測ロジックを軌道修正するフィードバックループロジックを適用させることで、予測精度の高い分析を実現した。

本事例の研究成果は、近年家庭に導入されつつあるスマートメーターなどから取得されるビッグデータ分析のための重要な技術となる。

今後は、共同研究を拡大し、電力ビジネスにおける課題の解決に向けたソリューションの開発を進める予定である。そして、ますます増大するビッグデータを分析・活用し、電力自由化された際に電力会社が導入できる需要側のエネルギーマネジメントシステム構築を目指す。

[1]共同研究開始時のプレスリリース（2012 年 11 月）
http://www.nttdata.com/jp/ja/news/release/2012/110101.html

2. 非定型データ活用基盤の研究開発

「(B) 多種多様なデータ活用」に該当する事例。企業の内外に存在する膨大な非定型データを分析し、ビジネスチャンスの発見やリスクの予兆検出に活かす「非定型データ活用基盤」の研究開発事例である。

企業の内外には様々な形態でデータが時々刻々と生成され増加している。また、現在に至るまでの企業活動で蓄積されてきたデータも膨大である。データ分析のための情報源として、このような膨大かつ多様な形態のデータを自由自在に扱えれば、様々な分析を進めることができ、より多くの発見や気づきを得ることが可能になると期待される。NTT データでは、フォーマットを定められていない多様なデータを活用するための「非定型データ活用基盤」ついて研究開発を進めている。この基盤を用いると、オフィスで作成されたデータ、業務システムに蓄積されているデータ、インターネットを介して取得したデータの 3 種類を組み合わせた分析など、従来は多大な労力を要していた分析についても、システム的にデータ分析・活用ができるようになる。

非定型データ活用基盤は、(1)取り込み、(2)エンリッチ、(3)分析、(4)可視化の 4 つからなる“データ活用プロセス”を実現するシステム基盤である。使われている主要な要素技術は、エンリッチ技術、XML データベース活用技術、可視化技術の 3 つがある。NTT データでは、これらの要素技術に関する研究開発を進めてきた。

エンリッチ技術は、データの意味解析と記録を行うための技術である。文章中から例えば「人名」や「会社名」などの単語、「故障発生時の状況」や「故障の内容」などを表現している句を判定し、意味情報をタグ付けした上で記録することが可能である。NTT データが長年に渡り蓄積してきた文書解析の技術と、NTT 研究所の研究成果を活用することで実現している。

XML データベース活用技術は、非定型データ

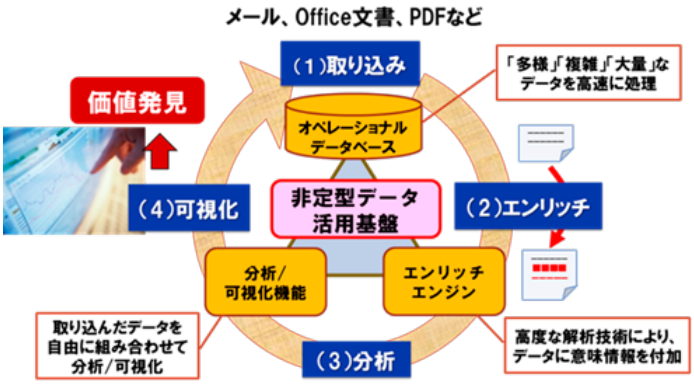

図 4：非定型データ活用基盤

活用基盤におけるデータベースを「XML データベース」に担わせるために必要な技術である。XML データベースを使う理由は、前述のエンリッチ技術を通じて記録したデータを“利活用しやすい形態”で蓄積するためである。「XML 形式によってタグ構造化」し、「XML データベース管理下に置く」ことが最適と考え、NTT データでは XML データベース専業ベンダである MarkLogic 社との協業・協力を進めている。

可視化技術は、分析結果を視覚的に表現する技術である。例えば、蓄積された膨大な文書データの関連性や類似性を求め、その結果をアイコンと線によって表現し、関連性や類似性のある文書群を視覚的に示す、といった表現方法を開発している。エンリッチのプロセスを経て XML データベースに蓄積された多種多様な大量データを様々な組み合わせで可視化し、ビジネス上の仮説検証や新たなビジネスチャンスの探索の助けとなることを目指している。

NTT データでは、携帯電話に関するトラブル報告の実データに非定型データ活用基盤を適用し、「携帯電話の故障発生時の状況」や「故障の内容」を表している句を判定しタグ付けが行えることを実証した。非定型データ活用基盤を活用し、刻々と蓄積される多種多様なデータからチャンスやリスクを発見することで、適切なビジネス判断が可能となる。

3. EC サイトの相互横断利用の促進

「(B) 多種多様なデータ活用」に該当する事例。複数の EC (Electronic Commerce「電子商取引」) サイト上で蓄積された総計テラバイト級の購買履歴を合わせて分析することで、EC サイトをまたぐ「クロスレコメンド」を可能とし、各ユーザの複数サイト相互利用に対する需要を発掘した事例である。

クロスレコメンドが必要となる場面には、次のような例がある。例えば、事業部ごとに EC サイトを運営しているような場合、新規に立ち上げたばかりのサイトと古くからあるサイトでは、そのユーザ数に大きな開きが生じる。あるいは、年齢層ごとにターゲットを絞り、ターゲットごとにサイトを運営している場合では、サイトをまたぐ年齢で他の競合サイトに顧客が流れ出てしまうなど、特定の年齢層以上からサイト利用が極めて減少する現象が生じることがある。このような状況に対して、多くのユーザ数を持つサイトからユーザ数が少ないサイトへ“誘導する”ことで、サイト提供側にとっては顧客増加となり、ユーザにとってはより多くの価値が得られる場合がある。

具体的なアプローチの 1 つとして、別のサイトの商品をレコメンドすることで、サイトの移動を促すことが考えられている。例えば、映画の DVD を閲覧しているユーザに原作本をレコメンドすることで、CD/DVD の EC サイトから書籍の EC サイトへユーザを誘導することが期待できる。この例のように、複数のサイトをまたぐ特徴から、「クロスレコメンド」と呼ばれる。既存のレコメンドとクロスレコメンドとの最大の違いは、ターゲットとなるユーザにおいては、レコメンドしようとする商品が存在するサイトの購買履歴が存在しない点である。ターゲットとなるユーザの購買履歴が存在しない状況下で、レコメンド対象サイトの商品群に対する、当該利用者の嗜好を如何にして把握するかが重要となる。

サイトをまたぐ共通情報やユーザ情報、商品情報などを最大限に活用することで、購買履歴の無いサイトの商品についてもレコメンド可能な仕組みを開発した。クロスレコメンドの導入により、ユーザのページビュー率が約 1.5 倍に増加した商品が存在するなど、売上増加や顧客の囲い込みに繋がる成果が実際に確認されている。

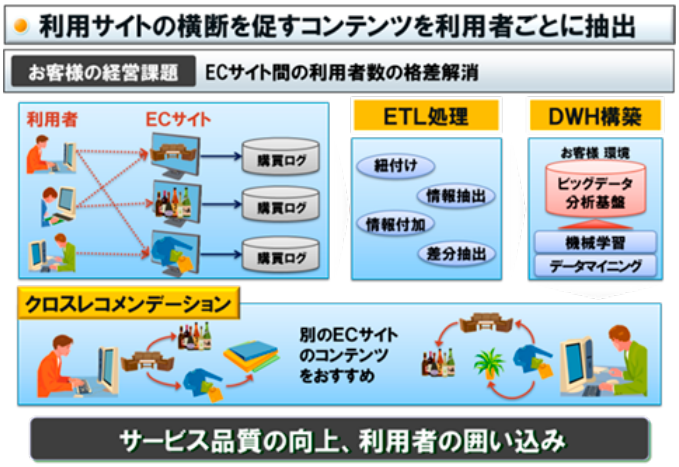

図 5：クロスレコメンデーション概略

今後のビジョン

IDC のデジタル・ユニバースに関する調査では、2020 年までに世界の情報量は 40 ゼタバイト（40 兆ギガバイト）に到達すると予測されており、マシンが生成するデータの量は現在の 15 倍に増加する。政府・自治体によるオープンデータ化が加速するとともに、データそのものが売買されるデータマーケットプレイスも大きく成長すると予測される。

これらにより、企業が活用できるデータは飛躍的に増大すると考えられるが、これを戦略的に活用し大きな成果を得るためには、優秀なデータアナリストの確保と、変化に即応しリアルタイムに高度な解析を実行し続ける分析基盤が必要となる。

NTT データでは、これらの両面で取り組みを進め、顧客企業に対し戦略的なデータ収集と分析サービスを提供していく。

2013 年 8 月版