

産業界の技術動向

ビッグデータ

NEC 中央研究所 支配人
高島 洋典

1. はじめに

大量の情報、すなわちビッグデータを利用することによって新たな価値を見出そうという動きが活発になっている。従来から、コンビニエンスストアの販売時点データ（POS データ）を分析することによって、売れ筋商品の見極めや陳列の配置を最適化するような試みが行われていた。最近になって、GPS を装備した携帯電話やカーナビの普及、また、IC カード乗車券や電子マネー、RFID など使われており、あらゆる人や物のセンシングが可能になってきた。また、クラウドコンピューティングの一般化や、ソーシャルネットワークサービスの急拡大により、これまでは分散していた情報が一元的な管理のもとに利用可能になってきた。

これまでは貯めるだけで精一杯であったが、大規模データ処理技術の発達により、これらのビッグデータを利用することによって、新しい価値を創造する試みが始まっている。本稿ではこのようなトレンドを紹介するとともに、今後の方向性や、対応する技術などについて述べる。

2. ビッグデータを取り巻く状況

マッキンゼーグローバルインスティテュートの報告によると、ビッグデータの創り出す価値は、

- ▶ 米国のヘルスケア産業に 3000 億ドルの価値を提供
(薬、患者、臨床などのデータ活用)
- ▶ ロケーションデータの活用により世界で 6000 億ドルの価値を消費者に提供
(行動データの利用)
- ▶ 米国の小売業の営業利益率を 60% 向上可能
(顧客、販売、運営データの利用)
- ▶ 欧州の政府機関に 2500 億ユーロの価値を提供
(公共データの公開による効率化、新サービスの提供)

などと言われている [1]。

また、表 1 のような領域においてもビッグデータの活用が検討されている。

表 1 ビッグデータの活用が検討されている領域の例

適用領域例	内容
パブリックセーフティ (ボードコントロール)	・数億人規模のリアルタイムな顔認証の実現により、待ち時間削減、処理件数増大
運輸	・配送車の位置、IC タグによる物品出入管理、交通情報等の分析処理による、予定時刻通りの物品配送、配送時間の短縮
交通	・プローブカーから得られる交通情報 / センサ情報の分析処理による、到着時刻予測、渋滞解消、安全確保、エネルギー効率最大化
金融	・銀行における入出金情報から口座不正利用を発見、犯罪利用を予防 ・世の中の様々な要因を考慮した金融商品のアルゴリズムトレード
通信	・携帯電話の通話・通信ログ解析による、通信ネットワークの混雑解消、輻輳回避
データセンタ	・サーバのセキュリティログ、アクセスログ、エラーログなどの収集・分析による、サーバ機器等の障害早期復旧、運用管理の省力化

このほかにも、例えば、温泉街における宿泊客の行動データを収集・分析することにより温泉街としての付加価値を高める試みや、天気予報、渋滞予測などすでに多くの情報活用が始まっている [2]。

画像認識や自然言語処理の領域では、従来は精緻なモデルとエレガントなアルゴリズムによって、例えば文字認識の精度を上げたり、機械翻訳の正確性を向上させてきた。しかし、最近では圧倒的な量の例文を学習することにより、人間がルールを書き下すことなくこれまでのものより正確な処理が可能になってきた。多少野蛮とも思えるやり方であるが、大量のデータが有する迫力を感じる。

3. サイバーフィジカルシステムとビッグデータ

物理システムとネットワーク化されたコンピューティングによって、新たなシステムが生まれつつある。これをサイバーフィジカルシステムと呼んでいる。サイバーフィジカルシステムでは、物理システムに埋め込まれた通信とコンピューティング機能により、物理システムが新たな能力を持つようになる。たとえば最近の自動車には多くのコンピュータが搭載され、各種の制御を行っている。従来の機械式の制御機構では得られないような高度な制御がソフトウェアによって可能になっている。自動車のほかにも、小さいものでは心臓のペースメーカーからロボット、大きなものでは化学プラントや電力ネットワークなどもサイバーであると同時にフィジカルなシステムである。これらは輸送や防衛、医療、農業、など様々な領域でますます重要になってくると思われる [3]。

特に、これからの社会基盤は高度にネットワーク化され、通信とコンピュータを有機的に結びつけ、安心・安全かつ効率的な生活を実現することが期待される。本来、サイバーフィジカルシステムとビッグデータは異なる概念であるが、両者が結びつくことによって、安全・安心な社会を実現したり、地球環境と人類の共存・共生を実現したり、あるいは産業競争力を実現するための社会のインフラとなることのできる (図 1)。



図 1. 社会インフラとしてのサイバーフィジカルシステム

多種多様な情報の取得・収集を経て、それらの蓄積・管理を行い、さらに分析を加えることにより、実世界に対してより効果的な働きかけが可能になる。このそれぞれの段階において、ビッグデータとしての大規模性、リアルタイム性、データの多様性・複雑性への対処が必要になる。

以下では、一例として NEC の中央研究所で取り組んでいる、ビッグデータに対応するための技術開発について述べる。

4. 大規模データ処理技術の概要

ビッグデータの分析には、2つのタイプがある。一つは一定期間蓄積した大量データを一括処理して知見や法則性を見出す蓄積データ処理である。一方、データを蓄積せずに通信の途中で処理し、データを収集し終わった段階で処理が完了するような方式もある。これをデータストリーム処理と呼ぶ(図2)。これらは単独で行われることもあるが、図2の右側に示したように、蓄積データ処理で得られた結果をルールという形でストリーム処理に利用するということもある。以下、それぞれの技術について紹介する [4]。

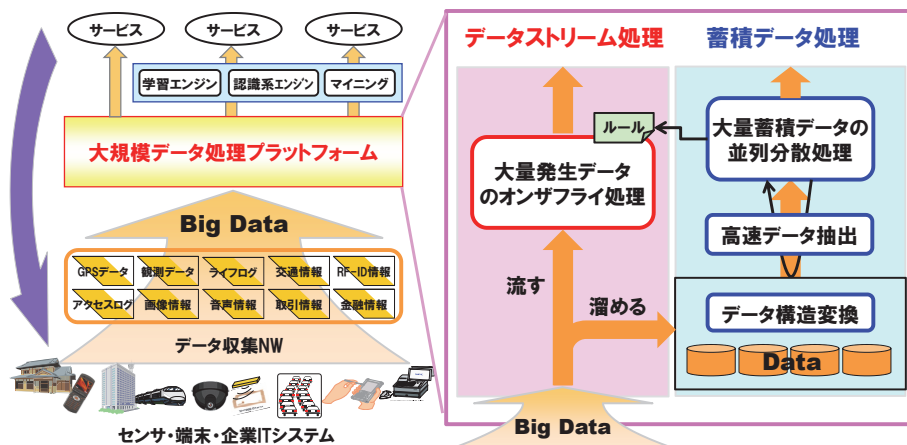


図2 大規模データ処理技術の構成

4.1 蓄積データ処理技術

蓄積データ処理には3つの要素がある。

(1) データ構造変換

データ分析アプリケーションには、それぞれに適したデータ構造があるため、必要に応じて分析前にデータ構造を変換しなければならない。たとえば、時間順に並んでいる利用者ごとの位置データを地区別に並べ替えたりすることがある。このデータ構造変換をデータの蓄積と並行して実行できるようにし、従来は前日の蓄積データを分析していたのに対して、約5分前の直前のデータの分析を可能とした(図3)。こうすることによって、きめ細かな対応が行えるようになる。

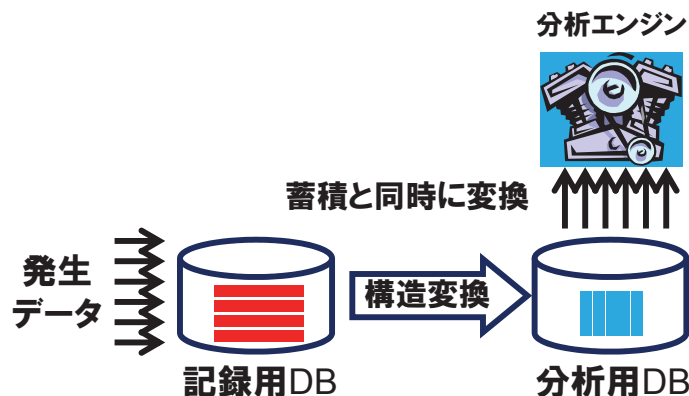


図3 データ構造変換

(2) 高速データ抽出

データが複数ノードに分散格納されている場合、位置情報等の多次元で表現されるデータの検索を高速化するためには、検索空間を絞り込み、アクセスするノード数を最小化することが鍵となる。多次元データの分散配置の最適化およびインデックスの最適化により、従来技術と比較して10～100倍の高速化を実現した。

(3) 大規模データの並列分散処理

大規模データを高速に処理するためには、処理を分割して並列に実行する並列分散処理が有効であるが、並列化の効果を最大限に引き出すためには、通信オーバーヘッドを削減する通信最適化が重要になる。特に、大規模分散システムにおいては、ネットワークスイッチがボトルネックになることが問題となっている。この問題を解決するため、ネットワークスイッチを考慮してデータとサーバ間の関係をモデル化し、このモデルから通信コストが最小となるデータ配置の最適解を求める手法を開発した。これにより、データとサーバ間の通信コストを最小化する処理配置を求めることが可能になった（図4）。

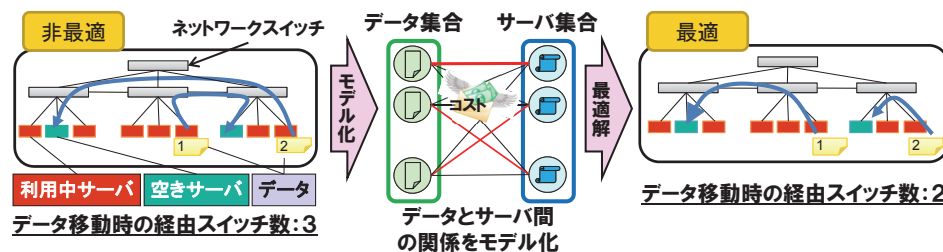


図4 大規模データ並列分散処理における通信経路の最適化

4.2 データストリーム処理技術

クラウドサービスに集められる情報を活用して、ユーザの位置ベースのレコメンデーションや、道路の渋滞情報、電力の使用状況の提供など、「いま」の状況に即したサービスの要求が高まっている。

(1) リアルタイムなマッチング処理

センサ等のデータソースから収集されたデータ（イベント）と、アプリケーションが予め設定したデータの配信条件（ルール）とを高速にマッチングし、データを蓄積せずにアプリケーションにとって必要な情報のみを提供する技術を開発している。これによりユーザは、刻一刻と変化する情報をリアルタイムに把握できるようになる。

(2) スケーラブルなシステムアーキテクチャ

スケールアウト型データストリーム処理アーキテクチャ（図5）により、データのマッチング処理は、サーバの追加に比例して性能を高めることができる。

このアーキテクチャは、大量に発生するデータを高速に振り分けるディスパッチャと、ディスパッチャで振り分けられたデータを複合イベント処理（CEP：Complex Event Processing）するイベントプロセッサの2段階で構成されている。ディスパッチャ間、イベントプロセッサ間で相互に情報を参照せずに、独立に処理を実行できるようにルールを配置することで、スケールアウト可能なアーキテクチャを実現した。従来は各サーバが独立に処理できず、スケールアウトできなかったという課題を、ディスパッチャレイヤでの構造化P2P技術を応用したルール分散配置方式と、システム全体でのルール配置の最適化アルゴリズムを開発することにより解決している。

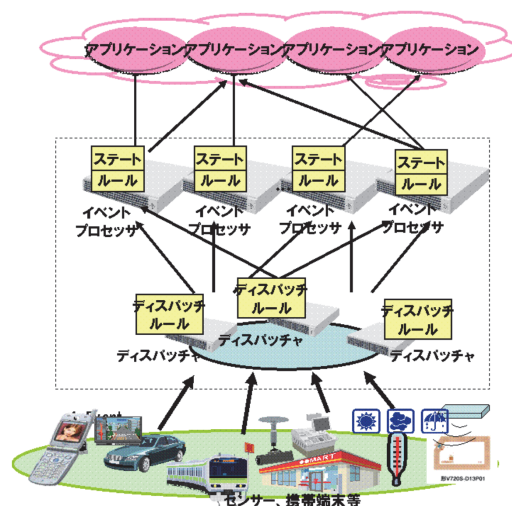


図5 スケールアウト型データストリーム処理

5. おわりに

ビッグデータを取り巻く状況と、それを支える技術について述べた。ここで重要なことは、技術が新たな価値を生み出すのではなく、ビッグデータをいかに活用して新たなサービス、価値を創り出すかということである。情報通信技術は交通や物流、医療や農業、あらゆるサービス産業の新しい価値を創造することができる。さらにこれまで個別に存在していたサービスを統合し、まったく新しいサービスを提供できる可能性も持っている。個別の産業の高度化とともに、融合による新たな産業創出の可能性を提供しているのである。

これらの動きは当然ながらグローバルなレベルで進展しており、わが国の競争力の維持、向上のためにも新産業の創出とそれを支える技術開発に取り組んでいかなければならない。今後とも、産業界としては大学とも連携し、研究開発に取り組む所存である。

参考文献

- [1] http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation
- [2] 「ビッグデータ大作戦」 日経コンピュータ 2012. 2. 2
- [3] “Cyber Physical Systems”, <http://varma.ece.cmu.edu/summit/CPS-Executive-Summary.pdf>
- [4] 「ダントツで骨太な成果創出に取り組む NEC 中央研究所」 ビジネスコミュニケーション 2012. 1