

「ビッグデータに備える」特集号について

喜連川 優 (東京大学) 森 正弥 (楽天) 武田 浩一 (日本 IBM 東京基礎研究所)

1. はじめに

インターネットの発展、およびクラウド環境の普及に伴い、世界規模でデータの蓄積とその活用が進行しています。インターネットサービスの利用が一般に広がることで、データベースに格納されていた構造化データが増大するだけでなく、モバイル端末の普及やセンサ技術の発達も合わせ、様々なログやテキスト、画像や音声、動画等の非構造化データの収集も容易になり、その種類も増えつつあります。この変化の傾向は加速度的に強まり、データはますます大量かつ多様になり、いわゆる「ビッグデータ」の時代が到来すると言われています。

米国は2012年3月、ビッグデータに関して2億ドル規模の予算を投じると発表しました¹⁾。この発表の興味深い点は、ビッグデータをスーパーコンピュータやインターネットと同様の基幹技術になると位置づけていることです。今までもデータを分析することで有意な情報を抽出し、ビジネスや社会に活かす試みは様々行われてきました。例えば、小売・流通分野では商品の在庫や販売情報を分析し、金融・保険分野ではリスクを分析し、通信・放送分野でネットワークや視聴率を分析し、それぞれサービスの利便性向上や事業の改善に活用してきました。ですが、ビッグデータの問題は、その用語の広がりと共に、とにかく大量かつ多様なデータを収集し解析すれば、従来にない革新的な知見が得られるだろうとの期待が膨らみ、本来の利用目的が見えづらくなっているようにも思われます。

ビッグデータの捉え方はまだ混沌としています。そのため、技術基盤が整い、効果的な分析方法が認知され、ビジネスや社会に役立つ応用事例が蓄積されるまでにはまだ時間を要するのではないかと思います。ますます増大するビッグデータに備えて、長い期間に渡ってこつこつと研究開発を進めていかななくてはならないと言えます。

本特集には、大量なデータを収集し解析するための基盤構築に関する論文が寄せられました。また加えて、実際のデータ分析を行った事例に関する論文も寄せられました。これらの論文には、実践からしか得られない貴重なプラクティスが多く議論されています。

多様なデータが獲得できる現代においては増え続けるデータに対処できる技術を構築することは大事です。その技術基盤は様々な構築方法がありえます。そして何よ

り、実際に分析し、ビジネスや社会の改善に役立ていかなければ、いくらデータを保有していても宝の持ち腐れでしかありません。これらを解析基盤の構築や、データの分析だけに留まらず、ビッグデータ時代に備えて、広く価値を創り出す試みにいかに高めていくか。そのための知見を明らかにしていくのが、本特集の目的です。

2. 本特集の論文について

本特集は、招待論文5件と投稿論文2件、インタビュー1件、グロスリから構成されています。論文の内容は、ビッグデータを収集または解析する基盤の構築に関するものが3件、基盤システムの性能評価方法が2件、実際にデータを解析した分析事例が2件となっています。

まず、平手氏らの「楽天におけるビッグデータとその収集・解析基盤の構築」では、大規模ユーザ属性DBプラットフォームとグローバルイベント解析プラットフォームの構築事例が紹介されています。日々増え続ける多種多様なデータ群を統一的に収集・解析し、いかにビジネス側にフィードバックを行うのか。またそのための基盤はどうあるべきか。ビッグデータにおいて必須になってくるこの課題について、実践や運用における難しさは何かを浮き彫りにしつつ、それを解決するための技術的な工夫点について明らかにしています。

岡野原氏らの「大規模リアルタイム解析エンジン Jubatusの創り方」では、大規模かつリアルタイムにデータを解析するための基盤ソフトウェアについて紹介されています。本論文では、基盤ソフトウェアの構築に加え、それを複数の企業で開発したことについての背景やメリットについても述べられています。ビッグデータの処理基盤は、分散並列処理による高速化やリアルタイムに近づいた解析、機械学習の活用等、必要となる技術が幅広く、解決すべき課題も困難になることが多くあります。この論文のように、複数の組織がそれぞれの強みをあわせて必要となる基盤を構築するという手法は、今後、ますます取り入れられてくることでしょう。

桑田氏らの「ビッグデータ時代のビジネス・インテリジェンス～次世代ビジネス・インテリジェンス～」では、今後役立つデータ分析基盤について紹介しています。従来においても、データウェアハウスと共にデータを分析するビジネス・インテリジェンスといわれるデータ分析

基盤は有用でした。ビッグデータにおいては、データの種類がライフログ・データやセンシング・データへと広がり、またその量や生成速度においても変化があります。このような中、次世代のビジネス・インテリジェンスにどのような機能が必要かについて、4つのタイプ分けを行い、そのうちの一つプロアクティブ型が核となると、その技術モデルおよび事例について解説しています。

このような基盤システムの構築に対して、廣川氏らの「商用サービス適用のための大規模分散処理システムの性能評価」では、その評価手法を述べています。ビッグデータの解析を担うシステムを、商用サービスに適用する際に性能評価はどう行うかという観点です。商用サービスへシステムを導入するには、サービスに求められる要件を満たす必要があります。本論文は、構築したシステムに対する性能評価結果を元に、一般的な指標に加え、安定運用性や故障に対する耐性等の観点にも従った性能評価が有用であると明らかにしています。また運用や故障に伴うスループットの低下等に対する対策にも踏み込み、実用的な論文となっています。

この論文と同じシステムを取り扱いつつ、坂井氏の「大規模分散処理システムのソフトウェア試験とその実践」では、全般的な品質検証について考察しています。実運用する環境やユースケース、またシステムの特性によって品質検証の観点は異なります。この論文では、「資源効率性」「障害許容性」「回復性」の3つが重要と指摘し、実際の検証で確認しています。システム構築時は検証方法についても事前に確認し、決めておくことが大事です。大規模分散処理システムにおいては検討項目も増えるため、示された試験方法は価値ある内容になっています。

一方、実際にデータを分析した古関氏らの「金融分野におけるビッグデータ分析」は、分析のための主要な技術手法を概観しつつ、金融分野での事例を紹介しています。ビッグデータにおいては、テキストを中心とした非構造データの蓄積があり、解析技術には、テキスト解析、データマイニング技術、分散処理技術等、いくつかの種類がありえます。これらの技術は単独でも重要ですが、ビッグデータの解析においては、その組合せがより重要になります。この論文では金融分野において技術を組み合わせた分析例が複数紹介されています。各例にてデータから価値を生む上でのプラクティスが考察されており、様々な観点から応用を考えることができます。

森永氏らの「ビッグデータ価値化への挑戦 薬剤副作用分析と航空機着陸システムの安全性設計から」では、薬剤副作用の分析と航空機着陸システムにおける異常検出というタスクを題材に、データから有意な情報を抽出

する上での課題について紹介されています。データの入手やデータの実態、その分析に関して直面した様々な課題に対し、領域の専門家と分析の専門家が協力してデータ解析に取り組んでいくことの必要性が指摘されています。

このように、本特集に寄せられた論文は、いずれも単なるデータの収集・解析を行う基盤の構築に関する手法だけではなく、基盤を評価する手法や、またデータを実際に分析し、いかにビジネスや社会へと活かしていくかについての方法論が述べられています。これらのように、大量にあるデータから価値を生むための知見を活かし、実際に効果を出していくことが、ビッグデータに対する実践的挑戦という観点から重要と言えるのではないのでしょうか。

3. おわりに

本特集に寄せられた論文のうち、基盤システム構築に関するものからは、幅広い技術を組み合わせ、実際の課題をいかに解決していくか、またどう安定運用させていくかの困難さとその重要性がよく分かります。ビッグデータは大量かつ多様であり、その生成速度は極めて速くなります。それらデータをニーズにあった形とスピードで解析していくことは、ビッグデータの解析を担う基盤システムが目指すべき姿の一つであります。

また、データの実際の分析・活用に関する2件からも、実際には複数の技術の組合せが必要であること、またそれだけでなく、ドメイン知識と分析技術等、複数領域の専門家の連携も大切であることが読み取れます。いずれの研究も、従来の手法にこだわらず、果敢に技術や領域を組み合わせるといふ、積極的なアプローチがあります。受動的な研究開発ではなく、目標と成果を見据え、事例を積み重ねておくことが肝要といえるでしょう。

本特集ではビッグデータという今後本格化する動向に対しての現時点での事例を扱ってきましたが、このような方法論は、どれも実際の成果をふまえた洞察に富み、ビジネスや社会に着実に応用が効くものと予想されます。最初に述べているように、本特集に寄せられた、現場を経験したものでしかわからない知見が、ビジネスに限らず応用され、広く社会における価値創出へとつながっていくことを编者として願います。

参考文献

- 1) 科学技術政策研究所:「財政緊縮下における米国の科学技術政策:2012年 AAAS 科学技術政策年次フォーラム報告」科学技術動向 2012年7・8月号(2012)。