

ビッグデータ利活用における現状と課題

横浜市立大学国際総合科学部准教授
国立国会図書館 非常勤調査員 永松 陽明

目 次

はじめに

I ビッグデータ利活用が活発となった背景

- 1 急増するビッグデータ
- 2 ビッグデータの種類と特徴
- 3 ビッグデータの動向

II 利活用事例を整理するためのフレームワークと基盤技術

- 1 ビッグデータ処理・利活用フロー
- 2 基盤技術

III ビッグデータの利活用事例

- 1 海外の先進事例
- 2 日本の先進事例

IV ビッグデータの利活用における課題

- 1 パーソナルデータの活用
- 2 オープンデータの活用
- 3 データ活用人材の育成
- 4 分析結果に基づく意思決定の困難性

おわりに

【要 旨】

ビッグデータ利活用の取組において、成果を挙げている事例が多く見受けられる。そこで本稿は、先進事例のポイントを整理するフレームワークの構築を行い、それを用いることで、ビッグデータ利活用に関する知見を得ることをねらいとする。まず、ビッグデータの利活用が活発となった背景として、SNS (Social Networking Service) などの普及やセンサの小型化、低価格化によるビッグデータの急増を説明した上で、その種類と特徴に触れ、ビッグデータの動向をまとめる。次に、先進事例を整理するフレームワークのビッグデータの処理・利活用フローを構築し、併せて基盤技術の概要を論じる。そして、フレームワークを用いて、国内外の先進利活用事例を整理し、知見を得る。最後に課題として、①パーソナルデータの活用、②オープンデータの活用、③データ活用人材の育成、④分析結果に基づく意思決定の困難性を説明する。

はじめに

今日、大量かつ多様なデータという意味を持つ「ビッグデータ」が新聞などのメディアに登場しない日は少ない。しかし、そのような状況になったのは平成24 (2012) 年以降であり、Buzzword⁽¹⁾として捉えられていることも多い。ただ、ビッグデータ利活用の代表的な事例は、平成24 (2012) 年以前から取り組まれたものが多く、一朝一夕でできたものは少ない。また、その取組に成功した企業は、高い競争力を手にしていることは周知の事実であり、ビッグデータという語彙が継続的に利用されていくかどうかは別にして、大量かつ多様なデータを解析し、ビジネスに利活用することは今後も続いていくと考えられる。

こうした中で、多くのビッグデータ利活用の取組がなされており、大きな成果を挙げている事例もある。その好例のポイントを整理できれば、参考になるところが大きいと考えられるが、まだ一般的な整理のフレームはない。そこで、本稿はビッグデータ利活用の事例についての新しい整理フレームワークを構築し、それを用いることでビッグデータ利活用の知見を得ることをねらいとし、論を進める。

ねらいを達成するために、まずビッグデータの急増要因や種類、特徴、ビッグデータをめぐる主な動きを整理する。次に、ビッグデータ整理のフレームワークとなるビッグデータ処理・利活用フローの内容を検討する。その中では、データを処理するための分析方法の説明やハードウェアやソフトウェアの機能をネットワーク上で提供するクラウドコンピューティングなどに代表される基盤技術の概要を論じる。そして、ビッグデータ処理・利活用フローを適用して、Amazon.comやコマツといった国内外の先進利活用事例やその他の企業の取組を整理し、知見を得る。最後に、利活用の課題となっている①パーソナルデータの活用、②オープンデータの活用、③データ活用人材の育成、④分析結果に基づく意思決定の困難性について説明する。

* 本稿におけるインターネット情報の最終アクセス日は、2015年2月1日である。

(1) Buzzwordとは、「流行語っぽい専門用語、時のことば」という意味を持つ語彙である。栗原潔「技術のライフサイクルを知り長期的な価値を見極める—第3部生き残る技術と消え去る技術の境界線—」『日経ITプロフェッショナル』32号, 2005.1, p. 34.

I ビッグデータ利活用が活発となった背景

1 急増するビッグデータ

SNSのFacebookやTwitter、動画投稿サイトであるYouTubeなどが、全世界的に利用されている。Facebookは、ユーザ数は13億人⁽²⁾を超えているだけではなく、アラブの春において重要なコミュニケーションツール⁽³⁾ともなった。YouTubeは2013年5月時点において、月60億時間の動画が視聴されていると主張している⁽⁴⁾。また、Amazon.comや楽天などが手掛けるオンラインショッピングも堅調な成長を続けている。

SNSの利用やオンラインショッピングなどのネットワークの活用だけではなく、スマートフォンを代表とする電子機器に搭載されている全地球測位システム（Global Positioning System: GPS）などの各種センサによる情報創出もビッグデータの急増に拍車をかけている。それらは、小型化、省電力化、低価格化によって、電子機器以外の自動車、建設機械、橋梁に代表されるインフラなどに積極的に組み込まれている⁽⁵⁾。加えて、中央演算処理装置（Central Processing Unit: CPU）やストレージ⁽⁶⁾などのハードウェア、ネットワークなどの性能も指数関数的に進化しており、それらの組合せであるクラウドコンピューティングの普及とも相まって、ビッグデータが増大している。

以上のような環境において、米国の調査会社であるIDC（International Data Corporation）は、2013年に生成・複製されたデジタルデータ⁽⁷⁾は4.4ゼタバイト⁽⁸⁾であったが、2020年になるとおよそ10倍の44ゼタバイトに到達すると予測している⁽⁹⁾。

2 ビッグデータの種類と特徴

ビッグデータの種類を図1に示す。

ビッグデータは、大きく「狭義のビッグデータ」と「広義のビッグデータ」に区分される。狭義のビッグデータはデータそのものを指し、広義のビッグデータは分析技術や人材育成など関連するものまでを含む。さらに狭義のビッグデータについては、定型的に整理されたデータである「構造化データ」と、音声、SNSなどの文字情報、動画などのデータサイズや更新される時間が一律ではなく整理しにくいデータである「非構造化データ」で構成される⁽¹⁰⁾。

(2) Facebookによれば、2014年12月31日時点での毎月1回以上利用するユーザ数は13億9000万人である。また、企業名についてはCorporationや株式会社などを省略して記載している。“Company Info.” Facebook Website <<http://newsroom.fb.com/company-info/>>

(3) 総務省編『平成24年度情報通信白書』2012, pp.143-145.

(4) “YouTube Official Blog,” 2013.5.1 YouTube Website <<http://youtube-global.blogspot.jp/2013/05/yt-brandcast-2013.html>>

(5) 機器に通信機能が組み込まれることを、Internet of Things (IoT) やMachine to Machine (M2M) と呼ぶ。

(6) ストレージとは外部記録装置のことであり、ハードディスクやフラッシュメモリなどで構成される。

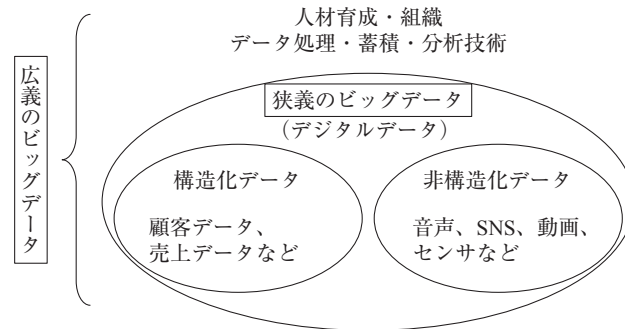
(7) デジタルデータとはコンピュータに蓄積される様々なデータのことであり、それらが大量に集まってビッグデータを形成する。

(8) ゼタバイトは、10の21乗を示す単位である。

(9) 「急増するセンサーにより、デジタルユニバースが拡大」（プレスリリース）2014.4.9. EMCウェブサイト <<http://japan.emc.com/about/news/press/japan/2014/20140410-1.htm>>

(10) 総務省情報通信国際戦略局情報通信経済室「情報流通・蓄積量の計測手法に係る調査研究報告書」2013.3, p.5. <http://www.soumu.go.jp/johotsusintokei/linkdata/h25_03_houkoku.pdf>

図1 ビッグデータの種類



(出典) 総務省情報通信国際戦略局情報通信経済室「情報流通・蓄積量の計測手法に係る調査研究報告書」2013.3, p.5. <http://www.soumu.go.jp/johotsusintokei/linkdata/h25_03_houkoku.pdf>を基に筆者作成。

また、図1に示した狭義のビッグデータは、「高解像度（データ粒度の詳細さ）」、「高頻度（データ収集をするタイミングが頻繁、もしくは随時）」、「多様性（データの種類の豊富さ）」といった特徴を持つ。以上の特徴から、ビッグデータを分析することによって、今まで見えなかったレベルで現実を可視化することができ、テーラーメードのようなアプローチ⁽¹¹⁾が可能になる。

3 ビッグデータの動向

(1) 関心が集まったきっかけ

近年ビッグデータに関心が集まるようになったきっかけは、2011年のMcKinsey Global Institute（以下、MGIと略す）⁽¹²⁾ のレポート⁽¹³⁾と、2012年のホワイトハウス声明⁽¹⁴⁾である。

MGIのレポートは、ビッグデータ利活用の経済的なインパクト、分析手法、分析に必要な人材の不足などを論じたレポートである。具体的に、経済的なインパクトについては、米国のヘルスケア産業と小売業、欧州連合（European Union: EU）の公共事業、世界の位置情報データ活用分野、世界の製造業分野の5つの部門におけるビッグデータ活用による経済効果や便益、コスト削減効果などの推計を行っている。分析手法については、回帰分析など20以上の手法を挙げている。人材不足については、2018年において米国におけるビッグデータ分野の必要な人材は44～49万人なのに対して、その不足は、14～19万人に達すると論じている。

ホワイトハウスの声明は、「ビッグデータ研究開発イニシアティブ（Big Data Research and Development Initiative）」と呼ばれ、国立科学財団（National Science Foundation: NSF）や国立衛生研究所（National Institutes of Health: NIH）、国防総省（Department of Defence: DoD）などの6つの政府機関に対して、ビッグデータの利活用に関する研究開発のため、5年にわたり2億ドルを超える予算

(11) テーラーメード (tailor made) とは、目的や注文に合わせて作る、もしくは調整するという意味を持ち、ビッグデータの分析により、一人一人の嗜好などに合わせたアプローチを採用することができる。

(12) MGIは、米国大手コンサルティング会社McKinsey & Company の調査研究機関である。

(13) J. Manyika, *et al.*, “Big Data: The next frontier for innovation, competition, and productivity,” McKinsey Global Institute, 2011.6. <http://www.mckinsey.com/~media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx>

(14) Office of Science and Technology Policy Executive Office of the President “Obama Administration Unveil “Big Data” Initiative: Announces \$200 Million in New R&D Investments,” 2012.3.29. <http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf>

を割り当てるものである。国立科学財団は人材育成のための大学でのプログラムなどを実施し、国立衛生研究所では1,000のゲノムプロジェクトのデータをクラウド上で入手できる取組を行っている。

以上のレポートと声明はビッグデータに関心を集めたが、関心を集めるに至った根底には、次の3点の指摘がある⁽¹⁵⁾。

- ・ センサの高性能化と低価格化
- ・ クラウドコンピューティング技術の進展とビジネス化
- ・ データ中心科学と言うべき科学研究の第4のパラダイムの登場⁽¹⁶⁾

第4のパラダイムとは、自然科学を中心とした研究方法の変遷をまとめるに当たって、4番目に現れた現在主流になる科学研究のパラダイム⁽¹⁷⁾であり、ビッグデータの利活用はこのパラダイムに沿ったものだと言える

(2) 各国の取組

EUでは、第7次研究枠組計画（The 7th Framework Programme funded European Research and Technological Development from 2007 until 2013: FP7）⁽¹⁸⁾においてBig Data Public Private Forum（BIG）の取組が実施された。本取組では、ビッグデータの獲得から利活用までの具体化や技術トレンドのロードマップ作成などが実施され、次期研究枠組であるHorizon2020策定に活かされた。⁽¹⁹⁾

中国では、2012年秋に開催された中国共産党第18回大会において「工業化・情報化・都市化・農業の近代化」の推進という大きな方針が示された中で、ビッグデータ利活用の実現への期待が述べられているため、中央政府をはじめ、地方政府も積極的に関連企業の誘致などを推進している。2012年12月に国家発展・改革委員会は、「2012年におけるハイテク・サービス業の研究開発と産業化に関する通知」を発表し、ビッグデータの分析ソフトウェア開発とビッグデータを活用したサービス創出を重点的支援の対象として指定した。2013年9月には工業・情報化部（工業和信息化部）は「情報化及び工業化の深度融合プロジェクト・アクションプラン（2013年～2018年）」を発表し、2018年度までの大手企業と中小企業におけるビッグデータ利活用計画の詳細を提示している。⁽²⁰⁾

日本では、平成25（2013）年5月に安倍晋三首相が成長戦略第2弾スピーチの中でビッグデータ活用に関する規制・制度改革を示した⁽²¹⁾。また、同年6月にIT総合戦略本部が「世界最先端IT国家創造宣言」⁽²²⁾の中で、個人情報を含むパーソナルデータの利活用と保護を両立する環境整備の方針を示し、同年9月に検討の場として「パーソナルデータに関する検討会」が設置さ

(15) 樋口知之「ビッグデータと個人化技術」『統計』63(9), 2012.9 pp.2-3.

(16) Toney Hey et al., eds., "The Forth Paradigm: Data-Insentive Scientific Discovery," Microsoft Research, 2009.10. <http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_complete_lr.pdf>

(17) 科学研究のパラダイムの変遷は、太古の時代において自然現象を記述する「経験」（第1のパラダイム）が生まれ、次に自然現象をモデル化し、汎用化する「理論研究」（第2のパラダイム）が登場し、計算機の出現によって複雑な現象をシミュレーションする「計算的方法」（第3のパラダイム）が台頭し、そして、第4のパラダイムが現れてきたとされる。

(18) "The 7th Framework Programme Funded European Research and Technological Development from 2007 until 2013," 2014.12.19. EU Website <http://cordis.europa.eu/fp7/home_en.html>

(19) BIGの詳しい取組は次を参照。Big Data Public Private Forum, "Welcome to BIG: Big Data Public Private Forum," <<http://www.big-project.eu/>>

(20) 総務省編『平成25年度版情報通信白書』2013, p.151; 総務省編『平成26年度版情報通信白書』2014, p.106.

(21) 「安倍総理「成長戦略第2弾スピーチ」（日本アカデミア）」2013.5.17. 首相官邸ウェブサイト <http://www.kantei.go.jp/jp/96_abe/statement/2013/0517speech.html>

れた⁽²³⁾。さらに平成26(2014)年6月には「パーソナルデータの利活用に関する制度改正大綱」⁽²⁴⁾が決定され、個人情報保護に関する法律の改正などの議論がなされている。平成26(2014)年に閣議決定された「産業競争力の強化に関する実行計画」⁽²⁵⁾においても、「世界最高水準のオープンデータやビッグデータ利活用の推進」が施策項目として取り上げられている。

(3) 企業の取組

企業の対応としては、ビッグデータ関連ビジネスの強化や活発な買収が展開されている。ビッグデータのITサービスにおいてトップ企業と言われているInternational Business Machines(以下「IBM」という。)では、これまで人間が担っていた知識の導出を、200社を超えるパートナー企業とタイアップして「Watson Developers Cloud」と呼ばれるサービスを提供することによって代替を狙っている。また、大手のITサービス企業は、ビッグデータの分析を得意とする企業買収を積極的に行っている。その主な買収は、MicrosoftによるFast Research & Transfer⁽²⁶⁾の買収、Hewlett-Packard(以下、HPという。)によるAutonomy⁽²⁷⁾の買収、OracleによるEndeca⁽²⁸⁾とBlueKai⁽²⁹⁾の買収、IBMによるSPSS⁽³⁰⁾とVivisimo⁽³¹⁾の買収、NTTデータによる数理システム⁽³²⁾の買収など枚挙にいとまがない。

(4) 技術の動き

技術の動きとしては、まず2003年と2004年に発表された2つのGoogleの論文⁽³³⁾がきっかけとなっている。これらの論文は、ビッグデータを高速に処理するための基本技術を記載したものであり、これらの技術のアイデアを踏まえて、2007年にApache Software Foundation⁽³⁴⁾によって「Apache Hadoop」が開発、発表された。「Apache Hadoop」がオープンソース⁽³⁵⁾で提供されたことにより、それ以降、ビッグデータに関連する技術開発が積極的に行われている。

以上、ビッグデータに関心が集まったきっかけ、各国の取組、企業の取組、技術の動きにつ

22 「世界最先端IT国家創造宣言について」(平成25年6月14日閣議決定)首相官邸ウェブサイト <<http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20130614/siryou1.pdf>>

23 「パーソナルデータに関する検討会の開催について」(平成26年8月8日高度情報通信ネットワーク社会推進戦略本部長決定)首相官邸ウェブサイト <<http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20140808/siryou2.pdf>>

24 高度情報通信ネットワーク社会推進戦略本部「パーソナルデータの利活用に関する制度改正大綱」2014.6.24.首相官邸ウェブサイト <http://www.kantei.go.jp/jp/singi/it2/info/h260625_siryou2.pdf>

25 「産業競争力の強化に関する実行計画」(平成26年1月24日閣議決定)首相官邸ウェブサイト <http://www.kantei.go.jp/jp/singi/keizaisaisei/pdf/housin_honbun_140124.pdf>

26 保存場所やデータ形式にとらわれずに企業の内部情報を検索するソフトウェアの提供を行うノルウェーの企業。

27 企業内部の情報を検索するソフトウェアやその表示、分析するソフトウェアを提供する企業。英国ケンブリッジで創業したのがルーツ。

28 企業内部の情報を検索するソフトウェアやその表示、分析するソフトウェアを提供する米国企業。

29 ビッグデータを活用しマーケティングに結びつけるソフトウェアの提供を行う米国企業。

30 高度な統計解析ソフトウェアを提供する企業。

31 保存場所やデータ形式にとらわれずに企業の内部情報を検索するソフトウェアの提供を行う企業。

32 様々なシミュレーションやデータ分析ソフトウェアを提供する日本の独立系ソフトハウス。

33 Sanjay Ghemawat et al., "The Google File System," 2003.10. Google Website <<http://static.googleusercontent.com/media/research.google.com/ja/archive/gfs-sosp2003.pdf>>; Jeffrey Dean and Sanjay Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," 2004.12. Google Website <<http://static.googleusercontent.com/media/research.google.com/ja/archive/mapreduce-osdi04.pdf>>

34 Apache Software Foundationとは、最先端技術のソフトウェアをオープンソースで提供する非営利組織である。

35 オープンソースとは、開発組織がソースコード(プログラム)を公開し、その組織がライセンスを与えた第三者に対しソフトウェアの使用、改良、再配布を認めるものである。

いて主な動きを表1に整理する。

表1 ビッグデータをめぐる主な動き

西暦	動 向
2000年	・ビッグデータの原点となった論文が発表される（8月）。 F. X. Diebold, "Big Data Dynamic Factor Models for Macroeconomic Measurement and Forecasting," Discussion Read to the Eighth World Congress of the Econometric Society, Seattle, 2000. 8.
2002年	・埼玉県の交通政策に活用された本田技研工業のinternavi(双方向のカーナビゲーションシステム) サービス開始（10月）。
2003年	・GoogleによるGoogle File System論文が発表される（10月） S. Ghemawat et al., "The Google File System," 2003.10. Google Website < http://static.googleusercontent.com/media/research.google.com/ja/archive/gfs-sosp2003.pdf >
2004年	・Facebook創業（2月）。 ・GoogleによるMapReduce論文が発表される（12月）。 J. Dean and S. Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," 2004.12. Google Website < http://static.googleusercontent.com/media/research.google.com/ja/archive/mapreduce-osdi04.pdf >
2005年	・T. O'Reilly氏によるWeb 2.0のコンセプトが発表される（9月）。 T. O'Reilly, "What Is Web 2.0, Design Patterns and Business Models for the Next Generation of Software," 2005.9.30. O'Reilly Media Website < http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html >
2006年	・GoogleのCEO、E. E. Schmidt氏によってクラウドコンピューティングが提唱される（8月）。 "Search Engine Strategies Conference" 2006.8.9. Google Website < http://www.google.com/press/podium/ses2006.html >
2007年	・Apple、iPhoneを初市場投入（6月）。 ・Apache Software FoundationによるApache Hadoopが発表される（9月）。
2008年	・米国大統領選挙においてオバマ陣営がSNSを活用する。 ・MicrosoftによるFast Search & Transfer買収（4月）。
2009年	・IBMによるSPSS買収（7月）。 ・MicrosoftによるForth Paradigmレポートが発表される（10月）。 T. Hey et al. eds., "The Forth Paradigm: Data-Intensive Scientific Discovery," Microsoft Research, 2009.10. < http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_completehr.pdf >
2010年	・SNSやソーシャルゲームが急拡大。
2011年	・MGIによるBigDataレポートが発表される（6月）。 J. Manyika, et al., "Big Data: The next frontier for innovation, competition, and productivity," Mckinsey Global Institute, 2011.6. < http://www.mckinsey.com/~media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.aspx > ・HPによるAutonomy買収（8月）。 ・OracleによるEndeca買収（11月）。
2012年	・NTTデータによる数理システム買収（2月）。 ・ホワイトハウスによる「ビッグデータ研究開発イニシアティブ」声明（3月）。 ・IBMによるVivisimo買収（4月）。 ・平成24年度情報通信白書に「ビッグデータ」の項目が記載される（7月）。
2013年	・安倍首相による成長戦略第2弾スピーチの中でビッグデータ活用に関する規制・制度改革が示される（1月）。 ・IBMによる「Watson Developers Cloud」サービスが発表される（12月）。
2014年	・「産業競争力の強化に関する実行計画」閣議決定（1月）。 ・OracleによるBlueKai買収（2月）。 ・IT総合戦略本部「パーソナルデータの利活用に関する制度改正大綱」を決定（6月）。

(出典) 小林孝嗣『ビッグデータ入門』インプレスジャパン, 2014, pp.8-9; 梶谷健一「海外におけるビッグデータのビジネス活用」『CAIJ Journal』54(3), 2014.3, pp.8-15; 総務省編『平成25年度版情報通信白書』2013, p.150; 総務省編『平成26年度版情報通信白書』2014, p.105; "Obama Administration Unveil "Big Data" Initiative: Announces \$200 Million in New R&D Investments," 2012.3.29. Office of Science and Technology Policy Executive Office of the President <http://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf>; 「安倍総理「成長戦略第2弾スピーチ」(日本アカデミア)」2013.5.17. 首相官邸ウェブサイト <http://www.kantei.go.jp/jp/96_abe/statement/2013/0517speech.html>; 「産業競争力の強化に関する実行計画」(平成26年1月24日閣議決定); 首相官邸ウェブサイト <http://www.kantei.go.jp/jp/singi/keizaisaisei/pdf/housin_honbun_140124.pdf>; 「パーソナルデータに関する検討会の開催について」2013.8.8. 首相官邸ウェブサイト <<http://www.kantei.go.jp/jp/singi/it2/kettei/pdf/20140808/siryoku2.pdf>>を基に筆者作成。

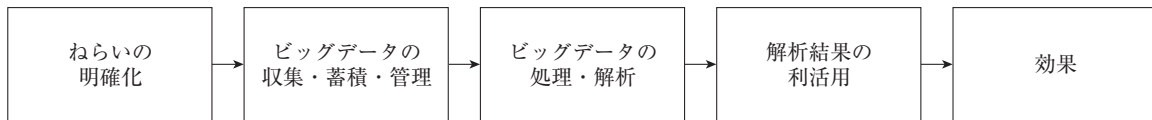
II 利活用事例を整理するためのフレームワークと基盤技術

1 ビッグデータ処理・利活用フロー

ビッグデータの利活用事例を整理するための一般的なフレームワークはないため、まず事例整理のフレームワークの構築を行う。それをを用いることによって、ビッグデータ利活用の知見を得ることができる。

フレームワークを考えていく上で、データを処理し、利活用するフローが重要であると考えられる。そこで、ビッグデータを利活用していくためのフローを図2に整理する。そのフローは5つのステップで構成され、以下でそれぞれの解説を行う。

図2 ビッグデータ処理・利活用フロー



(出典) 筆者作成。

(1) ねらいの明確化

先進企業・組織は、図2のフローをタイムリーに実施している。これによって競争力を獲得している。そのフローにおいて、まずビッグデータの処理・解析結果の利活用についてのねらいを明確化する必要がある。なぜならば、必要なデータ量、解析の精度を始めに決めておかないと、投資金額が設定できないためである。

(2) ビッグデータの収集・蓄積・管理

Iで述べたように各種センサやハードウェア、ネットワークの性能向上やそれらの低価格化によって、これまで収集できなかった様々なデータが収集、蓄積できるようになっている。これにより、ねらいに沿ったビッグデータの処理・利活用が可能となった。

図1 (I-2) において、ビッグデータは定型的に整理されているもの（構造化データ）とそうでないもの（非構造化データ）に区分することができることを述べた。データが構造化されている場合は、リレーショナルデータベース（Relational Database: RDB、以下RDBと略す）⁽³⁶⁾にデータを格納、管理することができる。

一方、データが音声や動画などの非構造化データである場合は、RDBが利用しにくい。なぜならば、データサイズや更新される時間が一律ではないためであり、データを表形式で持つことが難しいからである。そこで、非構造化データを処理できるデータベースを利用する必要があるため、NoSQL（No Structured Query Language、「ノーエスキューエル」と読む。）データベースを利用することになる。NoSQLデータベースについては、次節で概説する。

また、ビッグデータを高速に管理するには、読み書きに時間がかかるハードディスクではなく、半導体メモリにデータを持っておくインメモリデータベースも必要となってくる。インメ

⁽³⁶⁾ RDBは、データを表形式で持ち、その表それぞれに関連性を定義付けるため、データの加工が容易に行え、データを活用しやすいデータベースである。

モリデータベースについても次節で概説する。

(3) ビッグデータの処理・解析

ビッグデータを収集・蓄積・管理できるようになった一方で、それらのデータには処理・解析に使えないデータ、いわゆるノイズも多く含まれる。データの処理・解析を行うためには、不必要なデータやノイズを除去することが必須であり、ビッグデータの処理・解析においてはこの段階が非常に重要となる。

また、データの処理・解析については、様々な手法がある。表2に代表例を列挙する。ただし、表2に挙げる手法は同じ粒度にそろえた体系化ができていないことに留意されたい。例えば、データマイニング（Data Mining）には、クラスタリングを含む場合もある。

表2に示した処理・解析手法はビッグデータ特有の方法ではなく、従来から利用されているものである。しかし、「高解像度」、「高頻度」、「多様性」というビッグデータの特徴に合わせ用いるところにビッグデータ処理・解析の難しさがある。

処理・解析手法の使い方は、3パターンある⁽³⁷⁾。1番目のパターンは、表2に示した手法のうち一手法を用いて結果を得る方法である。2番目のパターンは、アンサンブル分析と呼ばれ、一つの手法で結果を得るのではなく、複数の手法を用いて、それらの結果を勘案して最終的な

表2 処理・解析手法の代表例

手法名	概要
機械学習	人間が行っている学習能力をコンピュータに実装する手法のことで、データから規則性を見つけ、予測まで行うことを実現。音声認識や画像認識、機器の故障予測など広範囲に適用可能。
データマイニング	Mining（採掘）の持つ意味の通り大量に蓄積されたデータを深く掘り下げて、相互関係や因果関係を探る手法。具体的には、「クラスタリング」など下記の手法群を用いて、機械的に関係性やパターンなどを発見。
クラスタリング	データの中で似ているもの同士をまとめて、いくつかのグループに分類する手法。
ニューラルネットワーク	脳の神経回路（ニューロン）の計算方式を模倣した手法であり、機械学習を実現する計算方法の一種。
回帰分析	ある変数（従属変数）がいくつかの変数（目的変数）によって関係付けられると想定される場合、その関係を定量的に求めることができる手法。
決定木 （ディジジョンツリー）	予測や分類のために、意思や行動を決定するまでの条件をツリー状にした手法。1つの条件に対して、YESの場合とNOの場合それぞれの分岐（処理）を記述。
アソシエーション分析 （相関分析）	マーケットバスケット分析とも呼ばれ、2つ以上の変数の相関関係を明らかにすることができる手法。ただし、因果関係のない場合にも、相関関係が高い結果になる場合もあるため、区別して用いる必要がある。
自然言語処理	言語をコンピュータで解析する手法。具体的には、文を意味が分かる最小意味単位に分解する「形態素解析」や各単語の出現頻度をカウントする「頻度分析」など。
セマンティック分析	検索文章内の品詞間の関連性などから言葉の意味を解析して精度を向上させる検索手法。
リンクマイニング	SNSやWebページのリンク構造、論文の引用関係など様々なネットワークのつながりを分析する手法。ネットワーク分析あるいは、SNSを分析する場合はソーシャルグラフなどとも呼称。
A/Bテスト	ウェブサイト最適化のため、複数のバージョンを同時に提供して、どちらかが好評かテストする手法。

（出典）城田真琴『ビッグデータの衝撃』東洋経済新報社, 2012, pp.76-80を基に筆者作成。

⁽³⁷⁾ ビル・フランクス（長尾高弘訳）『最強のビッグデータ戦略』日経BP社, 2012, pp.262-272.（原書名: Bill Franks, *Taming The Big Data Tidal Wave: Finding Opportunities in Huge Data Streams with Advanced Analytics*, 2012.）

結果を得る方法である。3番目のパターンは、コモディティモデル (Commodity Model) と呼ばれ、分析を自動化して結果を得る方法である。1番目及び2番目のパターンは、統計学を踏まえて、有意性が高い結果のもののみを利用するが、コモディティモデルでは処理・解析に採用を決めた手法から得た計算結果をそのまま信頼し利活用する。コモディティモデルは、膨大なビッグデータを効率的かつ低コストに処理するために採用される。

また、ビッグデータを分析する実務者や研究者の分析に対する考え方は2種類ある。1つ目の考え方は、指標間の因果関係を踏まえた統計学的な考え方であり、因果推論をベースとするものである。この考え方では、データの持つ時刻情報を取り込むことで、結果から因果関係を推論することができる⁽³⁸⁾。2つ目の考え方は、複雑な計算をせず迅速に結果を求め、その結果について利用者が妥当性や因果関係を想定し、その活用を考える情報論的なものである。この考え方では、大規模なデータ処理・解析において、迅速に計算を行えるため、探索的なアプローチ⁽³⁹⁾が可能となり、膨大なデータに埋もれた傾向を抽出できると主張している⁽⁴⁰⁾。

(4) 結果の利活用

このステップでは、各種の分析手法によって導出された結果をビジネスなどにおいて活用するものである。

(5) 効果

結果の利活用のステップが適切に実施された場合、最初のステップで明確化したねらいがこのステップにおいて実現されている。そのため、この段階ではこれまでのステップでかかった費用が回収できている状態になっている。

以上の5つのステップによって構成されるビッグデータ処理・利活用フローを用いて、Ⅲにおいて先進事例を整理する。

2 基盤技術

図2で示したフローは、多くの基盤技術の上で成り立っている。そこで、そのフローにおいて欠けてはならない重要な基盤技術に的を絞って概説する。

(1) クラウドコンピューティング

クラウドコンピューティングとは、インターネットにつながるPersonal Computer (以下「PC」という。)やタブレット端末から、ウェブブラウザなどを通じて、IT (Information Technology) リソースを利用できるサービスの総称であり、表3に示す種類がある。

⁽³⁸⁾ 統計数理研究所などはこのような考え方を主張している。丸山宏「データに基づく意思決定」『ESTRELLA』No.231, 2013.6, pp.3-4.

⁽³⁹⁾ 探索的なアプローチとは、良好な結果が出ると想定される計算式を用いるのではなく、多くの計算を行うことによって良好な結果を導き出すアプローチである。

⁽⁴⁰⁾ 主にデータマイニングを行う研究者はこのような考え方を主張している。芝村良「数理統計学とデータマイニング」『九州経済学会年報』No.43, 2005.12, pp.112-113.

表3 クラウドコンピューティングの種類

種類名	概要
IaaS (Infrastructure as a Service)	サーバなどのコンピュータシステムのインフラを利用できるサービス。 例 Amazon EC2
PaaS (Platform as a Service)	アプリケーションソフトウェアを稼働させることができる環境（インフラも含む）を利用できるサービス。 例 Google App Engine、Amazon Web Service
SaaS (Software as a Service)	アプリケーションソフトウェアの機能を利用できるサービス。 例 Salesforce.comによるSales Cloud

(出典) 筆者作成。

クラウドコンピューティングは、利用者にとってハードウェアやネットワークを購入する必要がなく、それらの機能をネットワーク経由でサービスとして利用するものであり、安価で迅速なサービスを利用できるため、急速に利用が増加している。

以上のクラウドコンピューティングによって、ユーザにとって低コストと導入スピードを兼ね備えたビッグデータを処理・解析するための環境が整備されてきている。

(2) Apache Hadoop

GoogleはWeb検索を高速に行うために、Google File SystemとMapReduceと呼ばれる分散処理技術を開発し、それらの概要を2003年と2004年に発表した。そして、Apache Software Foundationは、Googleの分散処理技術のコンセプトを踏まえて、並列分散処理基盤のApache Hadoopをオープンソースとして開発してきている。Google、Apache Software Foundationの技術とも多くのサーバを並列に動かすことによって高速処理を実現するものである。Apache Hadoopは、HDFS (Hadoop Distributed File System) とHadoop MapReduceから構成される。HDFSは多数のサーバにデータを分散し持たせることで、処理の集中によるスピードの低下が抑制できると共に、データ拡張性の高さも実現する。また、Hadoop MapReduceは多数のサーバが同時にデータを処理できる仕組みであり、これらによって並列分散処理基盤が作られている。

以上の技術は、ユーザにとってビッグデータを高速に処理できるメリットをもたらす。

(3) NoSQL データベース

NoSQLデータベースとは、非構造化データを様々な形で持つデータベースの総称であり、RDBと比べて高度なデータ加工はできないが、一つの情報に対して一つの値のみを持つなどのシンプルな構造で、データを大量に持つことができるデータベースである⁽⁴¹⁾。ビッグデータの利活用において、NoSQLデータベースが利用される理由は、RDBのようにデータを表形式で持たず、表それぞれに関係性を定義することなく利用できる上に、シンプルな構造である分、データ拡張を容易に実現できるためである。

(4) インメモリデータベース

インメモリデータベースは、ストレージなどにデータを格納するデータベースとは違い、データ処理を行うコンピュータの内部のメインメモリ上にデータを格納し処理を行うものである。

(41) 本橋信也ほか『NoSQLの基礎知識』リックテレコム, 2012, pp.50-55.

ビッグデータの利活用において、インメモリデータベースが利用される理由は、高速にデータを処理するためである。半導体であるメインメモリは、機械で構成されているハードディスクよりも入出力時のデータ転送速度スピードが速いため、高速化が実現できる。また、容量が限られるメインメモリ上で全データを扱うため、データの構造が単純化されていることも検索の高速化に貢献している。

III ビッグデータの利活用事例

本章では、ビッグデータ処理・利活用フローを用いて事例を整理し、知見を得る。

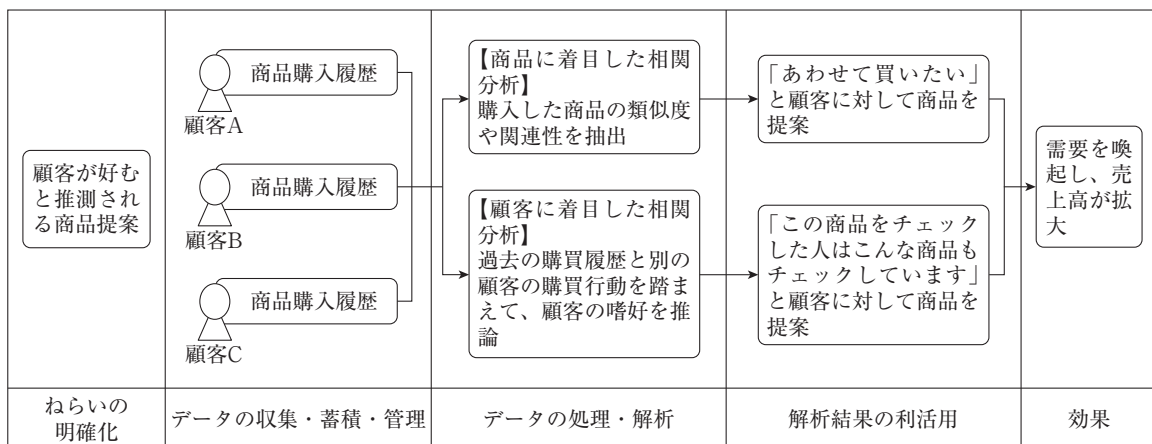
1 海外の先進事例

(1) Amazon.com の取組

ビッグデータ活用でよく知られているのは、Amazon.comである。これまでAmazon.comは様々なビッグデータを分析し、「ユーザ購買履歴に基づくおすすめ商品決定とその情報掲示」、「商品の自動価格設定」などの革新的な情報システムを多数開発している。

「ユーザ購買履歴に基づくおすすめ商品決定とその情報掲示」は、ユーザの過去の購買履歴を相関分析することで、顧客が好むであろう商品を提案するシステムである。具体的な内容としては、購入又は購入しようとしている商品に類似したものや関連性の高いもの（プリンターとインクカートリッジなどの検討している商品と同じタイミングで購入される商品）を購入検討画面に掲示、提案する機能と、購入又は購入しようとしている商品を購入した別の消費者が購入したものを購入検討画面に掲示、提案する機能の2つで構成される。それらの機能を図3に示す。それらによってもたらされる売上高は、2011年の全売上高において約35%を占めている⁽⁴²⁾。以上の機能は一般に「リコメンデーション機能」と呼ばれるが、Amazon.comだけではなく、楽天のウェブサイトやリクルートキャリアが展開するリクナビなどでも実装されている。

図3 Amazon.comによるリコメンデーション機能の取組

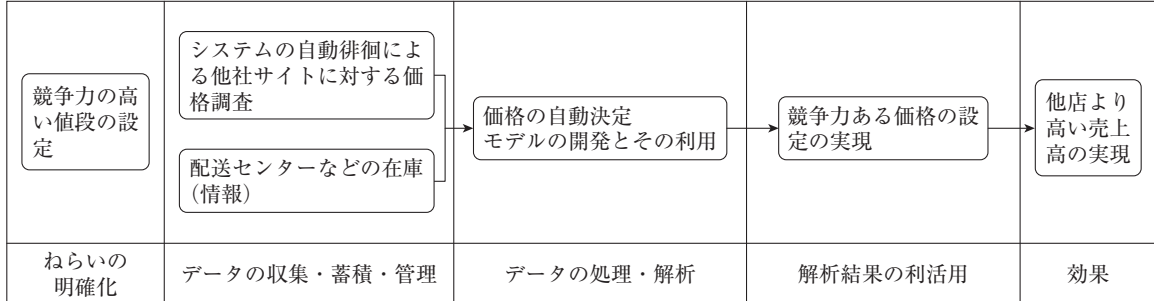


(出典) 春木良且『ソーシャルグラフの基礎知識』新曜社, 2012, pp.106-116を基に筆者作成。

(42) ポール・マクナーニ, ジョシュア・ゴフ「ビッグデータが日本企業に迫るもの」『ダイヤモンド・ハーバード・ビジネス・レビュー』38(2), 2013.2, pp.72-83.

「商品の自動価格設定」については、そのシステムがインターネットを自動徘徊し、他社価格を収集し、併せて自社の在庫情報を踏まえ価格を決定するものであり、他社に対して高い競争優位を保っている。具体的な内容を図4に示す。

図4 Amazon.comによる商品の自動価格設定の取組

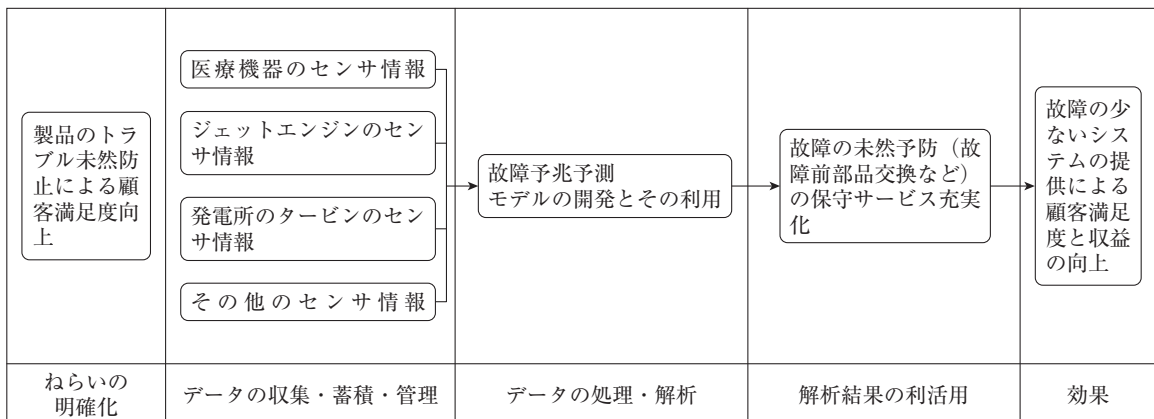


(出典) ポール・マクナーニ, ジョシユア・ゴフ「ビッグデータが日本企業に迫るもの」『ダイヤモンド・ハーバード・ビジネス・レビュー』38(2), 2013.2, pp.72-83を基に筆者作成。

(2) General Electric の取組

General Electric (GE) では、医療機器やジェットエンジン、発電所のタービンなどにセンサを取り付け、振動や温度などを細かく監視し、異常の兆候があれば、部品交換などの措置をとる保守サービスの提供を開始している⁽⁴³⁾。その一連の保守サービスを「インダストリアル・インターネット」と名付け、GEが展開する様々な機器、システムにそのコンセプトを適用している(図5参照)。これにより、顧客はトラブルを未然に防ぐことが可能となるため、GEは顧客満足度を向上させることができる。また、故障の未然防止などによって、全世界でGEが提供する機器の効率を1%改善すると、300億ドルのコストを削減できるとの試算も行っている。

図5 GEによるインダストリアル・インターネットの取組



(出典) 「データを一元管理・処理「レイクシステム」 GE、産業機械向け提供 解析高速化で効率運用」『日経産業新聞』2014.10.27, p.7を基に筆者作成。

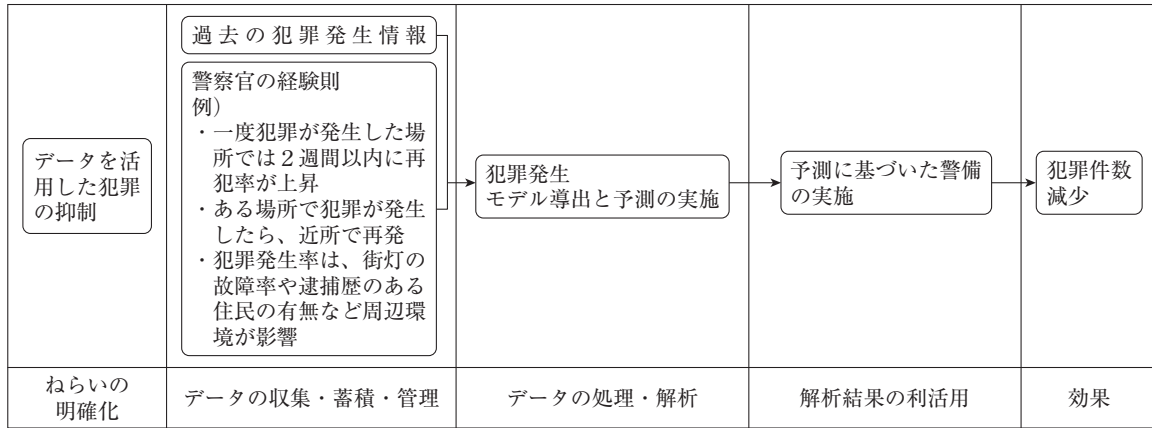
(3) サンタクルーズ市警の取組

米国カリフォルニア州のサンタクルーズ市の犯罪件数が、2011年と2013年を比較して13%減

(43) 「データを一元管理・処理「レイクシステム」 GE、産業機械向け提供 解析高速化で効率運用」『日経産業新聞』2014.10.27, p.7.

少した⁽⁴⁴⁾。その理由は、過去の犯罪発生情報や警察官の経験則を基に、カリフォルニア大学ロサンゼルス校とサンタクララ大学の研究者が「犯罪発生モデル」を導出し、犯罪が発生しそうな場所を予測し、プレディクティブポリシング (Predictive Policing) と呼ばれる予測に基づいた警備を行ったためである (図6参照)。

図6 サンタクルーズ市警による犯罪抑制の取組



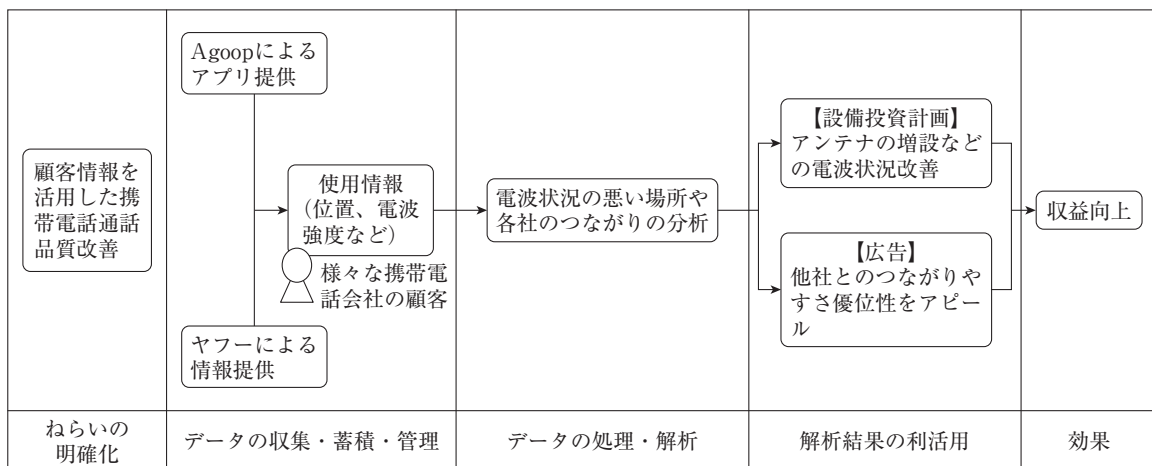
(出典)「特集格差広げるビッグデータ100」『日経コンピュータ』865号, 2014.7.24, p.35を基に筆者作成。

2 日本の先進事例

(1) ソフトバンクモバイルの取組

ソフトバンクモバイルは、顧客の携帯電話使用情報を収集し、電波状況などを把握した上で、通話品質改善を行っている。顧客の携帯電話使用情報は、グループ企業のAgoopが展開するアプリ「ラーメンチェッカー」などの各種アプリケーションやヤフーの防災速報から得られるユーザ使用情報を分析して把握している⁽⁴⁵⁾(図7参照)。

図7 ソフトバンクモバイルによる通話品質改善の取組



(出典) ソフトバンクモバイル「電波品質改善の取り組み：ビッグデータの活用」2014.4.9 <<http://www.softbank.jp/mobile/info/personal/news/service/20140409a/>>を基に筆者作成。

(44) 「特集格差広げるビッグデータ100」『日経コンピュータ』865号, 2014.7.24, p.35.

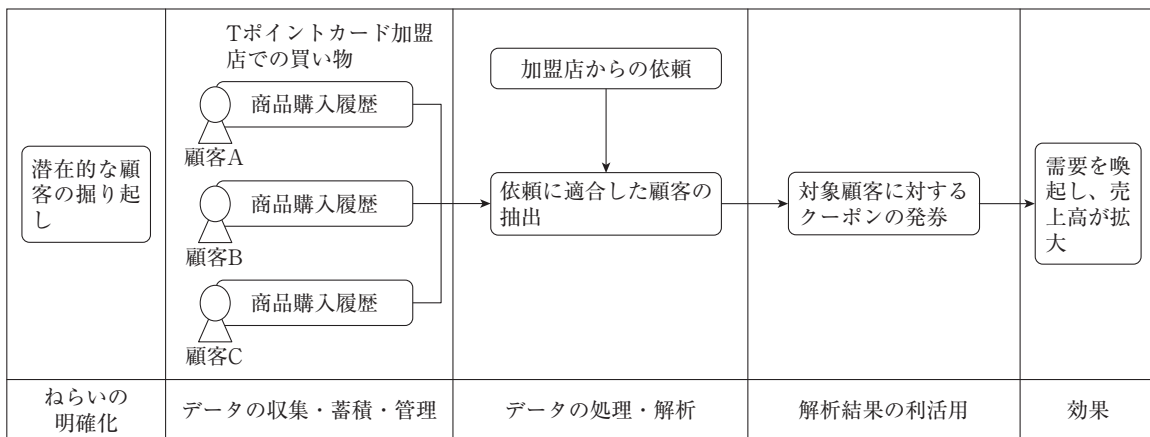
(45) ソフトバンクモバイル「電波品質改善の取り組み：ビッグデータの活用」2014.4.9. <<http://www.softbank.jp/mobile/info/personal/news/service/20140409a/>>

「ラーメンチェッカー」は、ユーザの現在位置近くのラーメン店を検索できるアプリであり、契約会社を問わず、無料で携帯端末に配信されている。その他にも病院や温泉などの探索アプリも提供している。それらのアプリとヤフーが提供する防災速報に関する使用情報を、ユーザの同意を得た上で収集し、位置情報、電波強度、データ通信の遅延速度などを分析し、結果を把握している。その結果を勘案し、設備投資計画などに反映し、携帯電話自身のつながりやすさに結び付けている。加えて、把握した他社の電波状況から、自社のつながりやすさが比較でき、そしてそのつながりやすさの優位性を、広告に活用している。

(2) Tポイント・ジャパンの取組

Tポイントカードを運営するTポイント・ジャパンでは、親会社であるカルチュア・コンビニエンス・クラブ（CCC）が展開するTSUTAYAに加えファミリーマート、マルエツ、カメラのキタムラなどの加盟店から得る顧客の商品購入履歴を分析し、潜在的な顧客の掘起しを行っている⁽⁴⁶⁾。加盟店が販売促進を行うために、Tポイントに対して潜在的な顧客へのクーポン発券の依頼があった場合、Tポイントは顧客のこれまでの購買履歴から対象顧客を抽出し、TSUTAYAなどで発券を行う。それによって、需要が喚起され売上高の拡大を実現している（図8参照）。ファミリーマートでは、対象を絞らないばらまき型のクーポン発券の利用率は2～3%であるが、分析を行って対象を絞ったクーポンの利用率は10%程度になっている。

図8 Tポイント・ジャパンによる潜在的な顧客の掘り起しの取組



(出典)「ファミマ、POSクーポン活用 多様なチェーンから誘客 売る技術光る技術」『日経MJ（流通新聞）』2012.8.22, p.5を基に筆者作成。

(3) コマツの取組

コマツでは、油圧ショベルなどの建設機械に温度などの各種センサや通信機器を組み込み、稼働情報を収集・解析するシステムを構築している⁽⁴⁷⁾。2013年3月末において、世界70か国で搭載車両が30万台を超えている。

このシステムは、①部品状況が把握可能になるため、故障・消耗する前に交換が可能になり、鉱山などでの24時間稼働の実現に寄与、②地域での稼働状況が把握できるため、需要予測がで

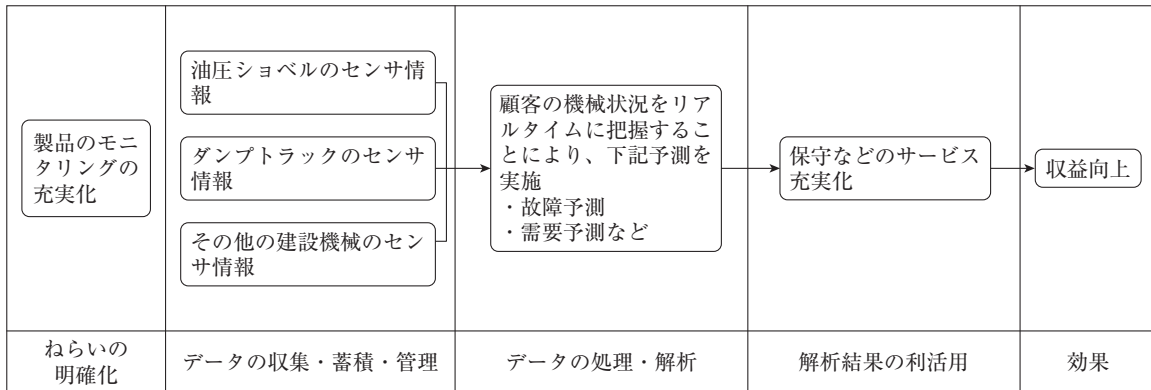
(46) 「ファミマ、POSクーポン活用 多様なチェーンから誘客 売る技術光る技術」『日経MJ（流通新聞）』2012.8.22, p.5.

(47) 山根宏輔「コマツの経営とIT戦略」『日本情報経営学会第66回全国大会予稿集（春号）』2013.5, pp.1-8.

き、生産計画にフィードバック可能、③顧客の燃費動向及び稼働情報を把握できるため、効率的な運転をコンサルティング、④位置情報が把握できるため、盗難防止が可能、⑤債権回収ができていない顧客の機械を停止させ、その回収に活用など、様々に利用されている。

以上のサービスをKOMTRAX (Komatsu Machine Tracking System) と名付け、全世界に展開している (図9参照)。

図9 コマツによるKOMTRAXの取組

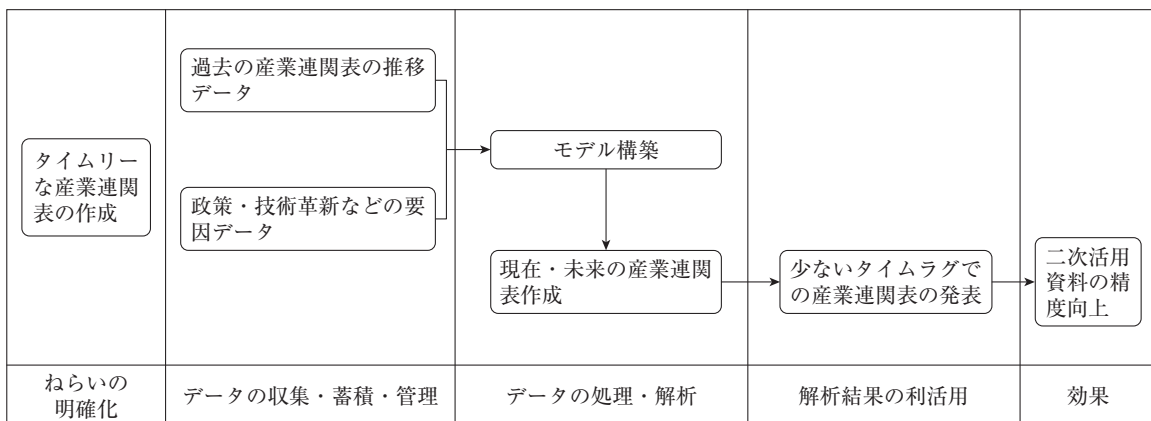


(出典) 山根宏輔「コマツの経営とIT戦略」『日本情報経営学会第66回全国大会予稿集 (春号)』2013.5, pp.1-8を基に筆者作成。

(4) 鳥取県の取組

鳥取県は、東京大学及び産業技術総合研究所と共同で産業連関表の精度向上に取り組んでいる⁽⁴⁸⁾。従来、産業連関表の作成には多くの時間がかかり、情報収集から発表までのタイムラグが5年程度あった。この改善のために、産業連関表の推移データや影響を与えるデータを考慮したモデルを構築し、タイムラグの少ない産業連関表を作成し、発表する計画である (図10参照)。

図10 鳥取県による産業連関表精度向上の取組



(出典) 森本浩之「鳥取県/産業連関表のエビデンスを基にした県政の展開—ビッグデータ活用による産業連関表精度向上の取組み—」『月刊LASDEC 地方自治情報誌』vol.42, no.12, 2012.12, pp.15-19を基に筆者作成。

(48) 森本浩之「鳥取県/産業連関表のエビデンスを基にした県政の展開—ビッグデータ活用による産業連関表精度向上の取組み—」『月刊LASDEC 地方自治情報誌』vol.42 no.12, 2012.12, pp.15-19.

(5) その他の取組

その他の活用例を表4にまとめる。

表4 その他の取組

組織名	業種（国内外）	事業などの概要及び取組内容
グリーン&ライフ イノベーション	農林水産 （国内）	北海道大学のベンチャー企業。大学の研究成果を応用し、人工衛星情報から魚の生息海域を予測する漁場予測情報サービスを提供。
本川牧場	農林水産 （国内）	大分県日田市にある乳牛、肉牛の生産牧場。平成20年からSalesforce.com社のクラウドサービスを用いて、牛に取り付けたセンサなどから収集したデータを管理。生育状況や健康に問題がある牛の検出、牛の状態の予測などを実施。
トヨタ自動車	製造（国内）	ナビサービスを利用している車から情報を収集。急ブレーキ多発情報などを愛知県警に提供。同様に他の自治体や企業にデータを提供。
マツダ	製造（国内）	新型エンジンに使用する鋳造部品の精度を向上させるため、製造過程における加工条件や品質情報などを大量に収集、分析。それにより精度の高い製造が可能になったと同時に精度の高い設計も実現。
トライアル カンパニー	流通（国内）	福岡に本社を置きディスカウントストアなどを展開する流通企業。店舗内にセンサとカメラを設置し、天井部より顧客の購買行動をモニタリング。
あきんど スシロー	サービス （国内）	回転すしチェーン運営企業。寿司皿にセンサを取り付け、個々の動態を把握。来店客や需要の予測などに反映。
アンデルセン	サービス （国内）	広島に本社を置くパン製造、販売企業。POS（Point Of Sales: 販売時点管理）情報に基づき顧客の販売履歴情報から来店客数を予測し、販売計画、製造計画を作成。
イーグルバス	サービス （国内）	埼玉県川越市に本社を置くバス運営企業。車載機器情報や乗客アンケートやコスト情報などから運行ダイヤの最適化を実施。
NTTデータ	サービス （国内）	橋梁などにセンサを設置し、変位やひずみなどをモニタリング。異常値が検知された場合、管理者へ通報するシステムを構築。東京ゲートブリッジやベトナム・カントー橋などで展開。
関西電力	サービス （国内）	電気メータをスマートメータに転換し、時間ごとの電力の計量を実施可能化。電気使用量の見える化サービスなどを実施。消費量に見合った設備投資が可能になる見込み。
マイクロアド	サービス （国内）	インターネット広告関連サービス企業。閲覧者の情報を基にWebサイトの広告枠のオークション価格を設定。
石川県 羽咋（はくい）市	行政（国内）	人工衛星などで撮影した刈取り前の圃場画像を分析し、タンパク質含有量を測定。それにより、品質の高い米の選別が可能となり、ブランド米として販売。
埼玉県	行政（国内）	本田技研工業の「internavi」情報を活用して、急ブレーキや事故が多発する地域に対策を実施。急ブレーキは対策後に約7割減少し、人身事故は対策後約2割減少。
Vestas Wind Systems	製造（海外）	デンマークに本社を置く世界的な風力発電機メーカー。気象データ、衛星写真などのデータを蓄積し、風力発電機の最適設置場所の解析を実施。
Progressive	金融（海外）	米国の自動車保険会社。契約者の自動車に運転状況を常時監視できる機器を搭載させ、その代わりに保険料が割引かれるサービスを実施。半年間、契約者の運転状況と事故リスクを分析し、特性に合った保険を提供する仕組み。
Climate Corporation	金融（海外）	米国の農家・農作物専門保険会社。米国国立気象局がリアルタイムに提供する気象データや米国農務省が提供する土壌情報を活用して、地域や作物ごとの収穫被害発生確率を予測し、保険料の算定を実施。
シカゴ市警	行政（海外）	犯罪発生を未然に防ぐため、SNSのつぶやき情報を分析。2012年に行われた国際会議の警備に活用。

（出典）野村総合研究所「ICT分野の革新が我が国社会経済システムに及ぼすインパクトに係る調査研究」2013.3, pp.132-142。総務省ウェブサイト<http://www.soumu.go.jp/johotsusintokei/linkdata/h25_04_houkoku.pdf>; 野村総合研究所「データの高度な利活用による業務・サービス革新が我が国経済および社会に与える波及効果に係る調査研究」2014.3, pp.7-16。総務省ウェブサイト<http://www.soumu.go.jp/johotsusintokei/linkdata/h26_06_houkoku.pdf>; 「特集 ビッグデータと電子自治体」『月刊LASDEC 地方自治情報誌』vol.42 no.12, 2012.12, pp.4-30; 「特集 日米最新事例 ビッグデータと電子自治体」『日経ビジネス』1709号, 2013.9.30, pp.28-50を基に筆者作成。

IV ビッグデータの利活用における課題

日本経済新聞と日経BP社によるビッグデータ関連調査⁽⁴⁹⁾によると、日本の経営者が国に期待する取組として、「データ活用とプライバシー保護を両立するルールづくり」、「国や地方公共団体が持つデータの公開」、「データ人材の育成」を挙げている。期待が高いことは、裏がえせばその点において課題があるとも考えられる。また、経営者がビッグデータを活用しきれていないとの指摘⁽⁵⁰⁾もある。そこで、本章では①パーソナルデータの活用、②オープンデータの活用、③データ活用人材の育成、④分析結果に基づく意思決定の困難性を課題として論じる。

1 パーソナルデータの活用

IDC (I-1参照) の調査によると、プライバシーなど保護されるべきデータの比率は40%以上になると予測されている。加えて、そうした情報のうち、現状は半分しか保護されていないとの指摘も行っている⁽⁵¹⁾。

そのような状況下において、平成24(2012)年7月にCCCのTポイントカード⁽⁵²⁾は、ドラッグストアのウェルシアなどから得た顧客データを用いて販売促進を行った件が新聞で大きく報道された。薬の情報は非常に高いプライバシー保持が必要であるとともに、顧客の承諾を得ずに活用すると刑法などにも抵触するおそれ⁽⁵³⁾があり、反響は大きかった⁽⁵⁴⁾。また、平成25(2013)年6月には、東日本旅客鉄道(JR東日本)が展開するICカード乗車券の「Suica」の乗車や購買履歴情報を日立製作所に提供し、同社がその分析結果を小売業や広告企業に販売しようとしたことも問題となった。JR東日本が日立に提供したデータは、一定の匿名化処理がされていたが、抗議が殺到した⁽⁵⁵⁾。Suica履歴データの販売は事前に会員に対する説明がされていなかったことや、オプトアウト(データベースから自身の情報を除外すること)が可能であることも告知されていなかったことなどが大きな要因になったと考えられる。以上のように、パーソナルデータの活用には、顧客からの大きな抵抗が存在する。一方、診療報酬明細書(レセプト)のデータを解析することで、医療費を増加させている要因を見つけ、医療費適正化を図ることもできる⁽⁵⁶⁾など、パーソナルデータの活用による効果も大きいと考えられている。そのため、パーソナルデータの活用とプライバシー保護が両立するルールづくりが求められている。

(49) 「ビッグデータ、悩みは人材 活用企業53% 本社・日経BP調査 9割で専門家不足」『日本経済新聞』2014.12.11, p.3.

(50) 丸山 前掲注(38), p.4.

(51) John Gantz and David Reinsel, “The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East,” 2012.12. p.11 <<http://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf>>

(52) Tポイント・ジャパンは平成24(2012)年10月に設立されているため、新聞報道時点(平成24年7月)では、CCCがTポイント事業を手掛けている。

(53) 刑法(明治40年法律第45号)第134条第1項の秘密漏示罪に抵触するおそれがある。

(54) 「医薬品購入データ取得 Tポイントで提携企業から CCCが販促利用」『朝日新聞』2012.7.17, p.38.

(55) 「「スイカ」データ外部販売 JR東、希望者は除外 利用者の不安に配慮」『日本経済新聞』2013.7.26, p.11.

(56) 安藤公一「データヘルスの推進について」『統計』65(10), 2014.10, pp.2-3.

2 オープンデータの活用

オープンデータとは、国や地方公共団体が持つ様々なデータを二次利用が可能な形で民間に開放することにより、民間主導で多様な公共サービスが創造されることを目指す動きを指す⁽⁵⁷⁾。前述した「世界最先端IT国家創造宣言」において、その推進が盛り込まれており、各省庁から二次利用可能なCSV（Comma Separated Values）やXML（Extensible Markup Language）などの形式で多くのデータの公開が始まっている。また、地方自治体においても神奈川県横浜市、福井県鯖江市、千葉県流山市などで先行的に取り組が行われている。例えば、鯖江市では、避難所などの位置や人口、気温、道路工事情報などをXML形式などで公開し、併せて第三者が開発したオープンデータを利用する携帯端末用アプリもホームページに掲示している⁽⁵⁸⁾。

以上のオープンデータ化を進めるに当たって、地方自治体では、「具体的な活用のイメージやニーズの明確化」、「提供側の効果・メリットの具体化」などの活用イメージが把握されていないことや、「個人情報等の機微情報の扱いに関する整理」、「提供情報の内容詳細・費用負担等の調整」に問題が存在するとの調査がある⁽⁵⁹⁾。

3 データ活用人材の育成

I-3で述べたMGIのレポート⁽⁶⁰⁾では、平成20（2008）年における各国のDeep Analytical Talent⁽⁶¹⁾と呼ぶ大学卒業生の数を挙げており、米国の24,730人に対して日本は3,400人で遥かに少ないことを指摘している。日本では、大学・大学院に統計学科や統計学専攻を設けずに、分野別の研究テーマに取り組むことにより、関連人材を育成してきた⁽⁶²⁾。この方法は、現場に精通した実践的な人材育成の観点からは大きな成果があったとされる。しかし、ビッグデータの利活用が活発化する中で、データ科学の知識のある人材の争奪戦が起こるなど、統計学の知見がある人材に対する需要は急増している。加えて、分野別の研究テーマに取り組むことから、応用が利きにくい人材が育成されがちであるとして、統計学の系統的な教育が必要とされている⁽⁶³⁾。

また、具体的に育成すべき人材として、3種類の人材が求められている⁽⁶⁴⁾。まず、様々な手法を駆使し解析する「データサイエンティスト」である。次に、多くのデータから役立つものを選び出し、それらに必要な加工を施し、処理・解析手法を決定する「ビッグデータアーキテクト（データキュレータ）」⁽⁶⁵⁾が必要とされている。その具体的な作業としては、欠損値や外れ

(57) 総務省編『平成25年度版情報通信白書』前掲注(20), pp.195-219.

(58) 「データシティ鯖江（XML,RDFによるオープンデータ化の推進）」2014.4.8. 鯖江市ウェブサイト <<http://www.city.sabae.fukui.jp/pageview.html?id=11552>>

(59) 野村総合研究所「地域におけるICT利活用の現状等に関する調査研究」2014.3, p.46. 総務省ウェブサイト <http://www.soumu.go.jp/johotsusintokei/linkdata/h26_07_houkoku.pdf>

(60) Manyika, *et al.*, *op.cit.*(13), p.105.

(61) 統計学や機械学習などの分析方法を習得したビッグデータを処理・解析できる人材を指す。 *ibid.*, p.85.

(62) 中西寛子「学習指導要領の改訂後の統計教育の動きと課題」『ESTRELLA』No.247, 2014.10, p.4. また、同論文では日本において専門的な統計学の専門教育を受けることのできる大学院は、総合研究大学院大学複合科学研究科統計科学専攻だけであると指摘している。

(63) 樋口 前掲注(15).

(64) マクナーニ・ゴフ 前掲注(42), pp.72-83; 丸山 前掲注(38), pp.4-6.

(65) ビッグデータ処理・利活用フローにおいて、根幹部分の詳細設計を行うため、建築家を指すアーキテクト（Architect）と名付けられている。また、本来は博物館などの学芸員を指すキュレータ（Curator）もデータの収集・蓄積・管理が解析結果に大きく影響を与えるため、その重要性を勘案して名付けられている。

表5 ビッグデータ利活用に必要な人材

専門人材名	役割	典型的なプロフィール
データサイエンティスト	・課題に最も適した技術やアプローチを用いた処理・解析を実施。	・統計学又は関連する学位保持者。 ・データ解析ソフト会社が認定するソフトウェアに関する有資格者。
ビッグデータアーキテクト (データキュレータ)	・新しいデータセットを構築、精緻化、管理して既存データを統合。 ・ビッグデータ戦略の実施に必要な処理・解析方法を定義。	・統計学又はコンピュータ工学の学位保持者。 ・10年以上のビッグデータに関連するマネジメント経験者。
ビッグデータストラテジスト	・会社として重要な意思決定の抽出とビッグデータ全体の活用戦略の策定。 ・組織要件の決定。	・当該業界における10年以上のマーケティングとITマネジメント経験者。

(出典) ポール・マクナーニ, ジョシュア・ゴフ「ビッグデータが日本企業に迫るもの」『ダイヤモンド・ハーバード・ビジネス・レビュー』38(2), 2013.2, pp.72-83; 丸山宏「データに基づく意思決定」『ESTRELLA』No.231, 2013.6, pp.4-6を基に筆者作成。

値などを処理し、異なるフォーマットや単位のデータを揃えるなどの分析の前処理である。この作業はデータのクレンジングと呼ばれ、ビッグデータの分析には非常に重要な部分となる。最後に、ビッグデータのねらいの明確化や結果を企業戦略に活用する「ビッグデータストラテジスト」も求められている。以上の人材の役割と典型的なプロフィール(経歴等)を表5にまとめる。

4 分析結果に基づく意思決定の困難性

ビッグデータによって意思決定の参考となる情報の収集及び可視化が容易になりつつある。それらの情報をモニタリングしながら意思決定できるシステム(Business Intelligence、経営コックピットなどと呼ばれる)が脚光を浴びている。前述したコマツもKOMTRAXで収集した情報を工場などでモニタリングできる仕組みを構築している。しかし、ビッグデータである稼働情報を活用できているコマツにおいても、インドネシアにおける需要予測が実績とかい離したため、平成26(2014)年3月期において大幅な業績下方修正を行った⁽⁶⁶⁾。同様にシャープも「エグゼクティブ・コックピット」と呼ばれる経営コックピットを保有しており、全世界の販売、在庫情報を把握しているが、液晶などの価格変動の大きい市場において成功しているとは言い難いとされている⁽⁶⁷⁾。

ビッグデータの分析、可視化は容易であっても、ビッグデータを利活用し、効果を上げ続けていくことは難しい。このような経験をモデル化していくことが、今後求められる。

おわりに

以上、本稿ではビッグデータ利活用の事例についての新しい整理のフレームワークを構築し、それを用いてビッグデータ利活用の知見を得ることをねらい、論を進めてきた。

(66) 「「お家芸」需要予測揺らぐ コマツ、東南ア、見誤る 為替・資源価格変動速く 利益率20%へ巻き返し ビジネスTODAY」『日本経済新聞』2014.4.26, p13.

(67) 丸山 前掲注(38), p.4.

具体的には、まずビッグデータの急増やビッグデータの種類、特徴、ビッグデータをめぐる主な動きを整理した。ビッグデータの急増についてはSNSやYouTubeの全世界的な利用とセンサの普及などが要因であることや、その種類に関しては「構造化データ」、「非構造化データ」があることを述べた。特徴は、「高解像度」「高頻度」「多様性」の3点があり、これらの特徴からビッグデータを処理・解析することにより、テーラーメイドのようなアプローチが可能になることを説明した。

次に、整理のフレームワークとなるビッグデータ処理・利活用フローの内容を検討した。そのフローは(1)ねらいの明確化、(2)ビッグデータの収集・蓄積・管理、(3)ビッグデータの処理・解析、(4)解析結果の利活用、(5)効果の5つのステップから構成される。そして、その(3)のステップにおけるデータ処理・解析の代表的な手法として、機械学習やデータマイニングなどがあることを説明した。また、ビッグデータの利活用を支える基盤技術であるクラウドコンピューティング、Apache Hadoop、NoSQLデータベース、インメモリデータベースについて概説を行った。

次に、ビッグデータ処理・活用フローを利用して、Amazon.comやコマツといった国内外の先進利活用事例を整理し、ビッグデータ利活用の知見を得ることができた。

そして、本稿の最後では、利活用の課題となっている①パーソナルデータの活用、②オープンデータの活用、③データ活用人材の育成、④分析結果活用時の意思決定の重要性について説明を行った。とりわけ、パーソナルデータの活用に関するルールの早急な確立が待たれる。

(ながまつ あきら)