

データサイエンティストとM2M ービジネス・技術・育成ー

尾崎 隆, 中村 伊知郎

データ分析がビッグデータの登場と共に、再び脚光を浴びているが、データ分析にこれから取り組もうとする企業にとって、データ分析の専門家であるデータサイエンティストを育成することは容易ではない。本稿ではデータ分析にこれから本格的に取り組もうとする企業が、データ分析に関して、どのような分野で活路を見だし、新たにデータ分析に参入するためにどのようなスキルを想定したか、また、データサイエンティストの育成のために、どのような取り組みをしようとしているかを紹介する。

キーワード：データサイエンティスト, M2M, 機械学習, ビッグデータ

1. はじめに

三菱電機インフォメーションシステムズ株式会社では、2013年7月にデータ分析に関する専門チームを設立し、データ分析関連事業の立ち上げを開始している。

データ分析専門チームの設立からはほぼ1年が経過したが、その間に、弊社がこの分野ですでに後発であることを踏まえ、ビッグデータ分析をはじめとする分析分野での現状と動向を分析し、弊社が目指すデータ分析分野での方向性を探ってきた。本稿では、弊社が認識するデータ分析分野の現状と動向を紹介し、それを踏まえて、弊社がどのような分野での分析事業の展開を考えているか、さらに、そのための技術要素、人材のスキル要素、それに基づく育成方法、組織のあり方等について、どのような考え方で取り組んでいるか等、取り組み方の概要を紹介する。

2. データ分析市場の現状と動向

弊社がデータ分析チームを発足させるに当たり、どの分野に可能性があるかを探るため、この市場での先進企業がどのように分析を実施しているかを考察し、以下のような分類を行った。

2.1 メガデータを扱う先進的なWEB系企業

(1) 技術

eコマースやWEB広告等のWEB系各社は単なる数理統計的処理に留まらず、行動科学等の最新科学に注目し、ビッグデータ技術を駆使した先進的な分析を

行い、WEB上のパーソナルマーケティング等に適用している。

(2) 人材確保

上記各社は人材確保のため、日米等の有名大学の博士号保有卒業生および博士課程在学からの採用を行う等、積極的な人材確保を行っている。

2.2 蓄積した分析技術を駆使するマーケティング系企業

(1) 技術の蓄積

大手広告代理店等では、マスマーケティングから、セグメント分析によるデジタルマーケティングに至る、長年の技術蓄積により、分析技術を確立している。

(2) 手法

これらも単なる数理統計ではなく、顧客プロファイリングやクラスタリング、機械学習、データマイニング等の分析技術要素の中から、最適な技術を抽出し、これらを組み合わせて、確実に結果を出せる手法を確立している。

(3) 人材確保

上記を遂行する分析技術者については、十数年以上前から確保・育成している。

2.3 リスク・コンプライアンスに関する分析を中心に据える金融系企業

(1) 技術

金融業界でのビッグデータ活用は、事業の総合的な観点から言えば、必ずしも先進的とは言えないが、与信や不正検知等、事業の根幹に関わるリスク・コンプライアンスを中心とした分野では、ほぼ確立された分析技術を保有しており、また、金融商品系分析についても、高度なドメイン知識と連携した分析技術を保有

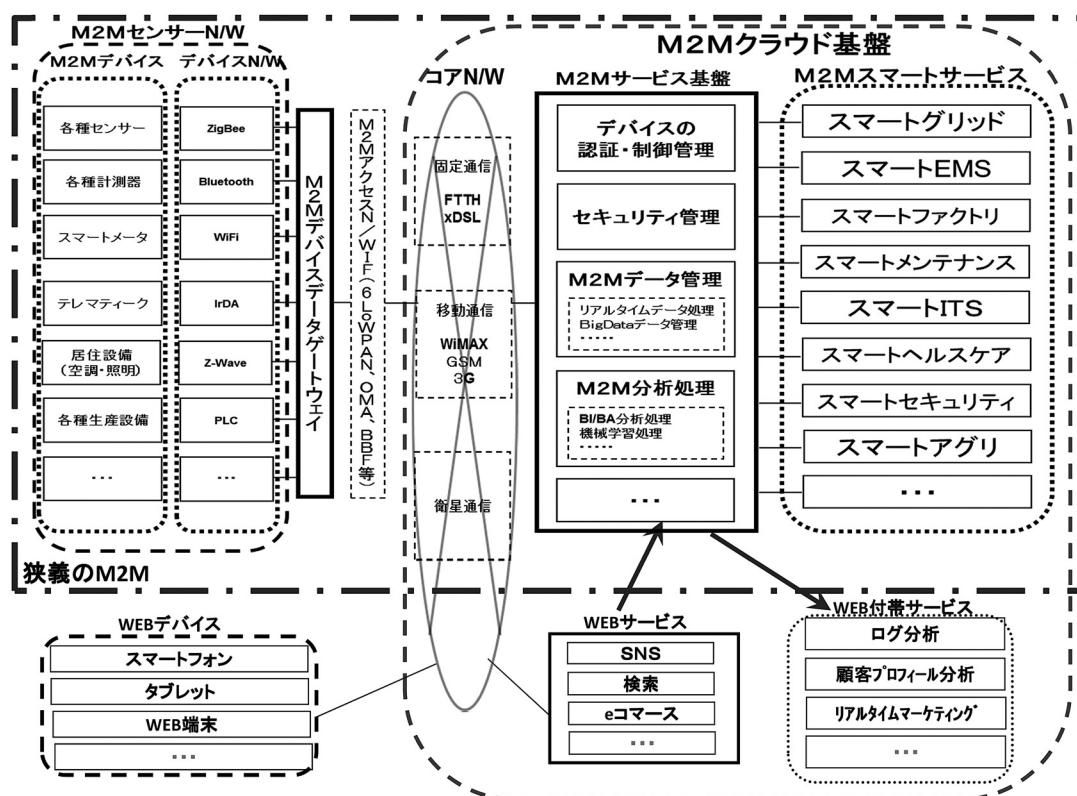


図1 IOT と M2M のアーキテクチャと適用可能性

している。

(2) 人材・基盤

金融系大手は基幹業務基盤と共に、分析基盤についても、オンプレミス上に巨大基盤を用意している。また、分析技術者についても、自前もしくは、ITおよびコンサル系子会社にて調達している傾向にある。

2.4 M2M に取り組み始めた製造系企業

(1) 製造業での動向

まず、製造業においては、従来からある、品質管理に関わる分析に加えて、GEのIndustrial Internetや独政府の製造業・ICT産業の国策であるIndustrie4.0のように、欧米ではすべての製品サイクルにおいて、IOT (Internet of things)・M2M (Machine to machine)の方向性を明確に示す大手企業が増加しており、M2Mに関するビッグデータ分析が今後、事業の中核に必須になってくる傾向が見て取れる。日本でも同様の傾向が見られるため、大手製造業各社はM2Mビッグデータ分析に関する取り組みを始めている。

(2) M2Mの多様性

ただし、M2Mの状況を見てみると、現状ではIOTという言葉が示すように、非常に多様な展開を見せて

おり、製造業に限らず、多くの産業での可能性がある。M2MとIOTは必ずしも同義ではないが、IOTを「もののインターネット」の中で、人に対して提供されるサービスという観点で広義のM2Mと捉えることができる。IOTとM2Mの適用可能性とアーキテクチャについては、図1 [6~10]に示す。従来のM2Mは狭義のM2Mの部分である。

(3) 製造業での適用可能性

製造業の観点で、M2Mの状況を俯瞰した場合、スマートファクトリやスマートメンテナンスを中心として、適用可能性が広がっている。製造業にデータ分析を適用する場合、M2Mに代表されるビッグデータの分析であっても、データにすべてを語らせ、データのみで仮説を確立する完全なデータ指向のアプローチが必ずしも適切であるとは言えない。開発・設計からアフターサービスまでの、製造業のバリューチェーンの中で、物理モデル等のモデリングとデータ分析の結果を融合した仮説に基づく数理モデルを作るほうが効率的で高精度である場合があることも否定できない。

2.5 弊社の目指す方向

上記のようなデータ分析の市場を考慮して、弊社とし

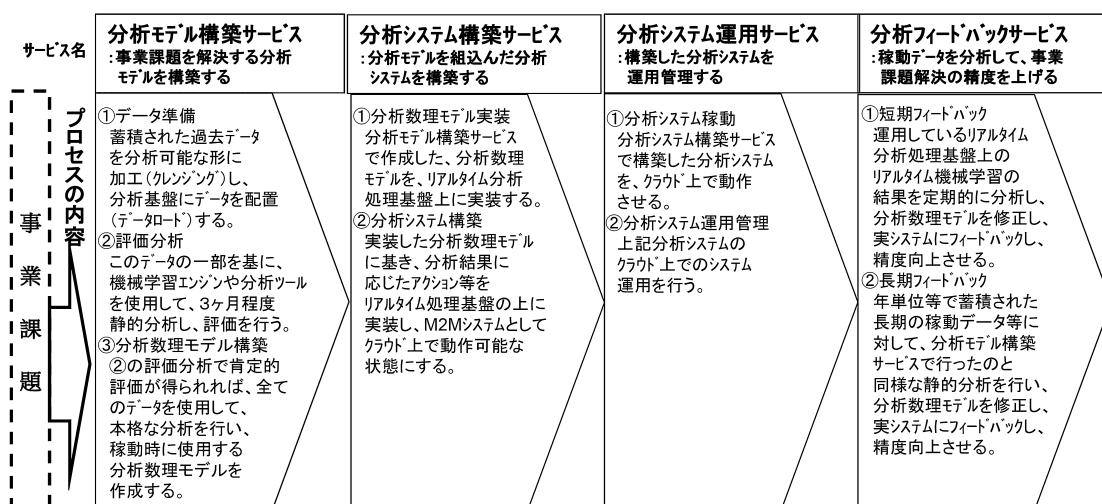


図2 M2M 分析サービスメニュー

ては、製造業のM2Mの分野での参入を検討している。その理由は以下の3点である。第1点は上記の2.1～2.3節のような分野に進出するにはかなりの期間がかかること、第2点は製造業のM2Mは今後、成長が期待できる分野であり、市場拡大が見込めること、第3点は弊社が製造業のグループ企業であるため、ある程度短期間で市場参入の可能性があること、である。

この分野でのデータ分析は、2.4節で先述したように、対象領域(ドメイン)に密着した物理モデルに精通した人材との協力によるデータ分析が必要になる。そのため、弊社では、グループ企業の研究部門と協力して、図2に示すようなデータ分析メニューを検討している。このメニューは、データ分析によって、M2Mに関する各事業ドメインでの事業課題を解決するためのプロセスに対応したサービス提供を目的にしている。本来であれば、事業ドメインでの事業課題を発掘することが、最初のプロセスとなるが、事業課題発掘に関しては、データ分析とは別の、ビジネスアナリシスのアプローチが必要となり、具体的な案件個別の事情に大きく左右されるため、本サービスのスコープからは、外している。

3. データ分析の為の人材像

データ分析全般にわたる人材像として、昨今、データサイエンティストが取り沙汰されている。データサイエンティストについては現時点では、厳密な定義は確立されていないのだが、データサイエンティストが職業として取り上げられるような現状では、前提や範囲を明確にしたうえで、何らかの定義をしない限り、人

材確保や育成も難しくなる。

一般社団法人「コンピュータソフトウェア協会(CSAJ)」では、データの収集・加工・分析に注目し、データに関わる、従来から米国等を中心に定義されてきた、データアーキテクト、データスチュワード、データアナリストを包含する形で、データサイエンティストを定義している[5]。これも1つの定義の方向性である。また、筆者が参加している、一般社団法人「データサイエンティスト協会」ではビジネスとの関わりを重視し、ビジネスアナリシスをデータドリブンに行える人材という観点で、スキル定義委員会を設置し、データサイエンティストに求められるスキルを定義しようとしている。弊社でも、データ分析を事業の中で取り上げるに際し、今後のデータ分析技術者に求められるスキルについて、ある想定をした。これはかなり一般的な見方と思われるのだが、実際の分析業務を考えた場合に、ビジネスに関するスキル、分析に関するスキル、ITに関するスキル、の3スキル要素に関して、基本的なスキルの保有を仮定せざるをえない。そのため、弊社ではそれらの3スキル要素に関して、以下のような考えに基づき、人材確保や育成を実施している。

3.1 ビジネスに関するスキル

データサイエンティストを含め、データとビジネスに関わる職業的な地位と役割が確立している米国等に比べ、我が国ではデータサイエンティストに対して、ビジネスへ関与することが、かなりの期待を持って語られている。その意味では、ビジネスアナリスト的な要素がかなり求められるのだが、ビジネス・分析・ITの3要素すべてに関し、エキスパートであるようなスー

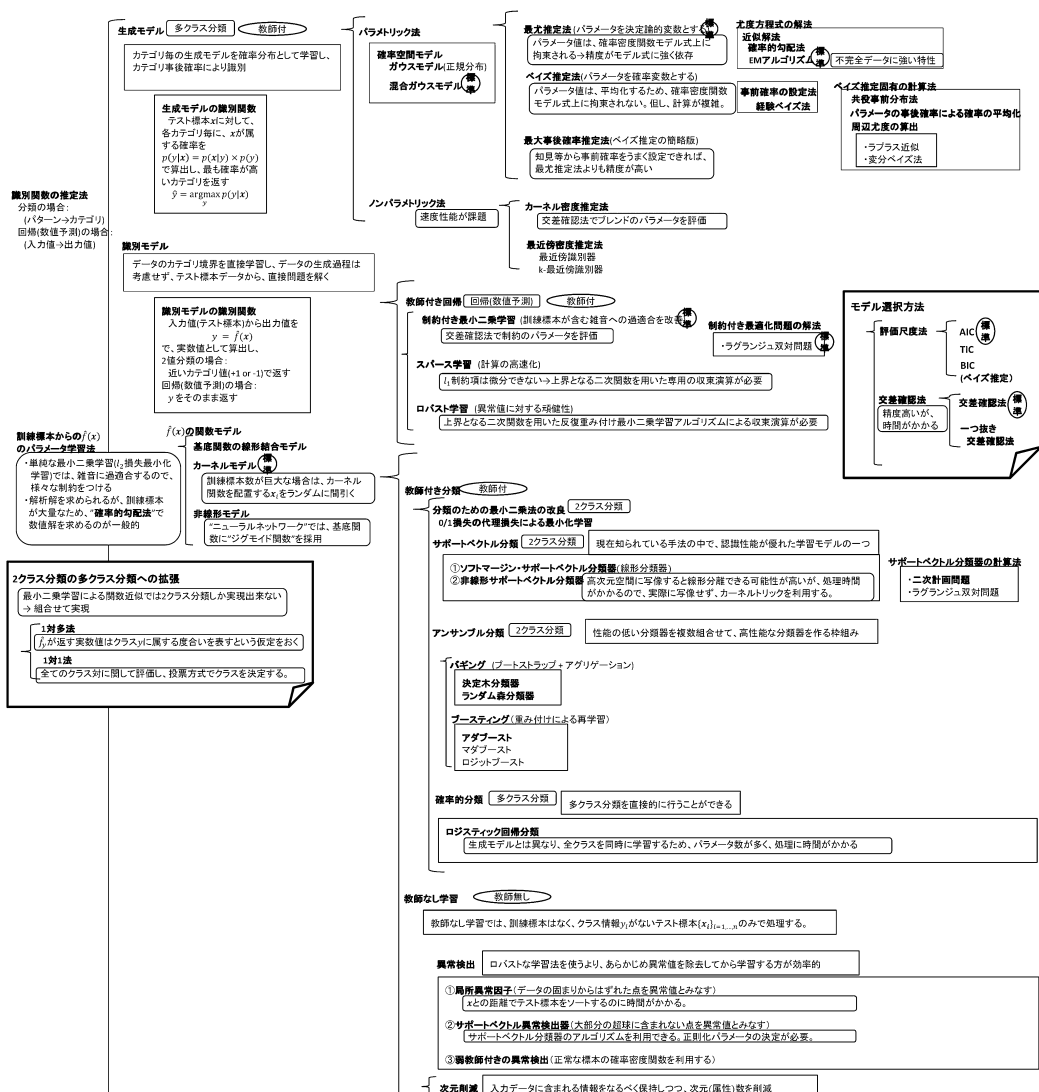


図 3 機械学習手法マップ (サマリー) の一部

パー人材の確保については、甚だ困難であると言わざるをえないので、まずは、以下の最低限のスキルを想定した。それは、ビジネスにおけるコミュニケーション・スキルであり、ビジネスの各対象領域のドメイン・エキスパートとコミュニケーションがとれ、ドメインでのビジネス課題を理解し、それをデータ分析に結びつけることができることが、最低限必要だと考えている。

3.2 分析に関するスキル

分析に関しては、最低限のスキルが必須であり、少なくとも、数理統計全般に関する基礎的な知識は必須である。M2M の分野を想定した場合 (M2M に限らず、ビッグデータを扱う場合)、基礎的な数理統計のみでは不十分である場合も多いので、機械学習やデータマイニングの手法が必要になるが、機械学習に関して

は、対象の理論や手法が多岐にわたり、短期間で、それらの手法の内容を詳細にわたって、論理的に理解することは相当困難であると言わざるをえない。機械学習を自由に使いこなす技術については、弊社内ですべてを調達することは当面困難であるため、グループ企業の研究部門等からの協力を仰ぐことを想定している。しかし、分析のプロセスの中で、分析の流れを正確に理解するためには、機械学習が使用された場合を想定して、俯瞰的に機械学習の体系を理解しておくことが必要になる。そのため、要員育成等において、機械学習の手法体系全般を理解することを助けるために、機械学習のさまざまな手法を、全体的に俯瞰できる「機械学習手法マップ」[11~13]を作成した。図3に、そのサマリー版の一部を掲載する。

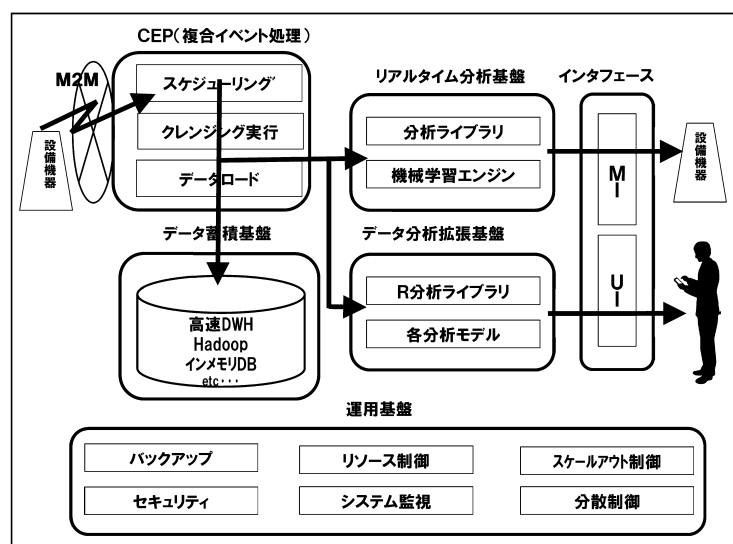


図 4 M2M 分析リアルタイム処理基盤

3.3 IT に関するスキル

M2M の分析を行い、構築された分析モデルをメンテナンスシステム等に組み込むためには、分析関連の IT システムやクラウド技術に関する IT スキルが必要になる。そのためのスキルを、ビジネスおよび分析のスキルと共に習得するには、やはり、かなりの時間がかかる。そのため、分析ツールや分析モデルをシステムへ組み込むことがある程度容易にでき、システム展開が属人的にならない、基盤の構築が必要になる。それらは、基本的にシステム運用が容易なクラウド基盤のうえに、収集したデータを分析ソフトに取り込めるようにするデータ連携基盤や、機械学習を始めとするデータ分析ソフトウェア等から構成される。この基盤は、過去のデータを分析して、分析モデルを作り上げる静的分析時と、作成済みの分析モデルを実際のシステムに適用して、リアルタイムに実行させ、実行時のデータからさらに分析精度を高めるリアルタイム分析時の、双方において必要になる。図 4 にリアルタイム分析基盤の例を示す。弊社では、このような考え方で、基盤を検討中である。

4. データ分析に関する組織

日経 BP ビッグデータ総覧 2013 によれば、データ分析を中心とした、データ活用組織は、おおむね、以下の 6 つに分類できるという [14]。

(1) 専門組織型

人材を専門組織に集約し、各事業部門を支援する

(2) 事業部門型

人材を各事業部門で育成する

(3) 専門組織 + 事業部門型

専門組織から各事業部門に人材派遣し、連携する

(4) システム部門型

人材をシステム部門に集約し、育成する

(5) 全社参加型

全社従業員に対し、教育を実施し、現場での分析を推進する

(6) 協業型

複数企業でコンソーシアムを形成し、各企業に分析結果を提供する

弊社の場合、(1) 専門組織型と (2) 事業部門型の中間的な組織を設立した。ただし、(3) 専門組織 + 事業部門型の組織にはなっていない。組織の内容を一言で表すと、(1) 専門組織を特定の事業部門に所属させるというものである。この組織では、データ分析に関する若手人材は全社的な観点から選抜し、育成する。また、データ分析に関わる案件についても、全社的な観点で取り扱い、所属する事業部門のみの案件のみを取り扱うわけではない。では、どのような理由で上記のような組織にしたのか？ その理由は以下の 3 点である。1 点目は弊社にとって、データ分析の取り組み自体が新たな試みであり、(3) のような折中案では専門組織結成後、直ちに事業部門へ派遣できる人材を用意できないこと。また、2 点目は (2) のように事業部門で人材育成ができるほど、投資効果の面で、各事業部門に余裕がないこと。さらに、3 点目は独立した専門組織にすると事業との結びつきが薄くなる危険性があり、

それを回避するためには、現時点で最もデータ分析案件の発掘の可能性のある部門の中におき、データ分析案件の発掘を、人材育成と同時に行う必要があること。このような組織で発足し、ほぼ1年が経過したなかで、現在、データヘルス系の分析案件と通信系WEB設備系の分析案件について、育成した若手分析技術者も投入して、データ分析を始めている。今後は、目指す所の、遠隔監視データの分析等M2M関連の分析案件の発掘を行っていく予定である。

3節で述べた、ビジネス、データ分析、ITの3スキル要素と組織の関係については、データ分析の組織がハブとなって、業務SEや客先、基盤SE、グループ企業の研究所等を結びつけ、組織として、データ分析に関するプロジェクトを遂行することを想定している。本来の意味でのデータサイエンティストなるスーパーマンの役割を、組織としてカバーしていくことが当面の目標である。

5. データ分析に関する要員育成

前節までに、人材育成についてはかなり、触れているが、弊社の場合、データ分析については、新たに事業展開することになるので、自社内での教育だけでは、育成に限界があると感じていた。人材育成のためには、分析スキルに限ってみても、基礎的な数理統計から、多変量解析やデータマイニング、機械学習に至るまで、多岐の項目についての技術習得が必要である。分析スキル育成に関しては、社外講座を積極的に利用した。また、社内でも、若手とベテランが一緒になって、輪講と演習を実施し、基礎的な分析スキルに関しては、これらにより、かなりの成果を上げることができた。ビジネススキルに関する育成はかなり難しいし、講座等により、一朝一夕で身につくものでもないが、データサイエンティストとしての前提となる、事業課題をデータ分析によって解決するためのプロセスやデータ分析による事業課題の解決事例等、上位の概念知識等については、社外講座を利用した。ITスキルについては、選抜した若手がSE出身であることもあり、基礎的な知識はあるので、特別な教育は行わず、OJTによる実践で実施している。また、R等の分析ツールを使った分析の実践についても、一部、アッセンフェルターのワイン方程式について実データをRによって分析させ、分析の迫体験をさせる等、社内での実習を行った

りしたが、基本的には、OJTの中で実践中である。現在は選抜した若手に対し、OJTによる実践分析により、クレンジング等データ準備を含めた分析プロセスの実際、分析のためのIT関連スキル、分析がビジネスにもたらす効果、等を学んでいる。

また、前章で述べたデータ分析の組織が案件展開のハブとなる一步として、各事業部のSEに対し、初歩的なデータ分析の知識を身につけてもらうための社内教育も開始している。

データ分析チーム全体のスキルアップのためには、今後、チーム内の既存人材に対する、より高度な分析技術教育の実施、IT基盤開発の技術部門との連携、OJTによるビジネススキル体験の蓄積等を検討中である。また、チームに配属された新人等の新たな人材に関しては、昨年度の教育の成果を、新人材の教育に反映し、社内での教育比率を高めていく予定である。

参考文献

- [1] ネイト・シルバー（西内啓 解説、川添節子 訳）、『シグナル&ノイズ 天才データアナリストの「予測学」』、日経BP社、2013。
- [2] デイミトリ・マックス、ポール・ブラウン（馬淵邦美 監修、小林啓倫 訳）、『データ・サイエンティストに学ぶ「分析力」』、日経BP社、2013。
- [3] 河本薫、『会社を変える分析の力』、講談社、2013。
- [4] 統計数理研究所、『文部科学省委託事業「データサイエンティスト育成ネットワークの形成」平成25年度事業報告書』、2014。
- [5] 一般社団法人コンピュータソフトウェア協会(CSAJ)、『平成25年度データサイエンティスト育成研究会(報告書)』、2014。
- [6] 大澤達蔵、『現場のモノとクラウドをつなぐM2Mの実践』、FUJITSU、62(3)、282-287、2011。
- [7] 清尾克彦、『M2M(Machine to Machine) 技術の動向と応用事例』、サイバー大学紀要、5、1-22、2013。
- [8] 松田尚久、佐藤裕一、小関昌一、永井信行、『M2Mサービスプラットフォームの開発』、NEC技報、64(4)、18-21、2011。
- [9] 山田勝彦、塩尻浩久、『M2M標準化動向と遠隔管理技術の標準化活動』、NEC技報、64(4)、26-30、2011。
- [10] 稲田修一、M2Mシステムを活用した付加価値創造とイノベーションの本質、<http://it.impressbm.co.jp/articles/-/11095> (2014.4.21 閲覧)
- [11] 上田修功、『NII 軽井沢土曜講話会 2011 年 11 月 4 日講演資料「統計的機械学習入門」』、NTTコミュニケーション科学基礎研究所、2011。
- [12] 杉山将、『統計的機械学習—生成モデルに基づくパターン認識—』、オーム社、2009。
- [13] 杉山将、『イラストで学ぶ機械学習—最小二乗法による識別モデル学習を中心に—』、講談社、2013。
- [14] 日経BP ビッグデータ・プロジェクト、『ビッグデータ総覧 2013』、日経BP、2013。