

ビッグデータの利活用のための 専門人材育成について

平成 27 年 7 月 30 日

大学共同利用機関法人 情報・システム研究機構

ビッグデータの利活用に係る専門人材育成に向けた産学官懇談会

要 旨

1. 報告書の背景

- 情報通信技術の発展により、ビッグデータが登場し、社会の価値がビッグデータとそれを処理する情報技術へと大きくシフトしているにもかかわらず、我が国ではデータサイエンティストは大幅に不足しており、またデータサイエンティストを育てる体制もできていない。そのため、データサイエンティスト育成が喫緊の課題となっている。
- 当懇談会では、産業界やアカデミアからの要請を把握し、育成すべきデータサイエンティストが持つべきスキルや能力を特定し、その育成のための方策を検討した。

2. データサイエンティストのスキルレベルとその育成

- データサイエンティストに要求されるスキルとしては、
 - ・データサイエンス力（統計学、機械学習、最適化などを理解し、使える力）
 - ・データエンジニアリング（データサイエンスを実装する力）
 - ・ビジネス力（課題を理解し、問題設定し、解決する力）などがある。結果として、データサイエンティストは方法と領域の双方に通暁した T 型・II 型人間であり、さらに、コミュニケーション能力や研究倫理も持ち合わせた人となる。
- 本懇談会ではデータサイエンティスト育成を、すべての大学生を想定したリテラシーレベル（50 万人）から、見習いレベル（5 万人）、独り立ちレベル（5000 人）、棟梁レベル（500 人）、業界代表レベル（50 人）、世界トップレベル（5 人）までの、6 つのレベルで考えることにし、それぞれの育成方法を検討した。

3. 提言

- そのうち主要なものを提言として掲げると以下のとおりである。
 - ・我が国の問題の根源は、棟梁レベルの決定的不足にある。この解決のために国家レベルの拠点を設置して、年 500 名規模の「棟梁レベル」の人材育成をめざし、上層への成長や下層へのトリクルダウン効果も狙う。
 - ・リテラシーレベルや独り立ちレベルの大学教育を加速させるために、主要 10 大学程度で本報告書の提案に基づく人材育成をスタートすると共に、MOOC などのオンライン教材を整備し、全国への波及効果を狙う。
 - ・社会全体のリテラシーやアウェアネスを向上させるために、全学的教養教育の実施、国家レベルのフラッグシップ・プロジェクトの推進、コンテストの開催、映像素材の充実などの取組を行う。
- これらの方策を実現するにあたっては、データサイエンスを副専攻とするダブルディグリー制人材育成の推進やスキル認定制度も有効と考えられる。

目次

1. データサイエンティスト育成の必要性と我が国の課題.....	1
1.1. はじめに.....	1
1.2. 我が国の課題.....	1
2. 我が国におけるデータサイエンティストへの要請.....	2
2.1. 産業界からの要請.....	3
2.2. アカデミアからの要請.....	4
3. データサイエンス人材育成のあるべき姿と実現に向けた仮説.....	6
3.1. データリテラシー.....	7
3.2. 見習い(基礎能力)レベル.....	7
3.3. 独り立ちレベル.....	7
3.4. 棟梁レベル.....	8
3.5. 指導的データサイエンティスト.....	8
3.6. 「中抜き」仮説.....	8
4. 具体的施策.....	10
4.1. リテラシーの醸成(50万人規模).....	10
4.2. 見習いレベルの育成(5万人規模).....	10
4.3. 独り立ちレベルの育成(5000人規模).....	11
4.4. 棟梁レベルの育成(500人規模).....	12
4.5. 指導的データサイエンティストの育成(50人規模).....	13
4.6. トリクルダウン効果とスケールアウト.....	14
4.7. 出口戦略.....	14
5. まとめ.....	14

1. データサイエンティスト育成の必要性和我が国の課題

1.1. はじめに

ビッグデータと情報技術により、従来の質・量とは、全く異なる価値が提供されるようになった。ビッグデータの活用領域は異常検出・予測・自動化・最適化など多岐に広がっており、ネット広告におけるリアルタイムビidding、金融における高頻度証券取引、UberやAirbnbなどクラウドソーシングに基づく新しいビジネス、Google等による自動運転車など、データ・計算能力が爆発的に増加したことによって初めて可能になった新しい価値が社会の成長を牽引しつつある。これは、産業革命が肉体労働に対して起こしたのと同様あるいはもっと大規模な革命が、今回は知的労働の面で起こりつつあることを示している。また、2015年7月時点での世界の企業の時価総額ランキングのトップ3はアップル、マイクロソフト、Googleであり、これからも社会の価値がビッグデータとそれを処理する情報技術へと大きくシフトしていることが見て取れる。

この世界的な大きな動きを先導するには、伝統的な統計学の枠を越えて、機械学習、最適化、自然言語処理、因果推論など最新の技術を様々な分野に応用できる専門人材(T型・II型)が必要である。加えて、このような専門人材を実務や最先端の研究開発の場で活かせるマネジメントレベルのデータリテラシーも圧倒的に不足している。これらの課題を早急に解決しない限り、我が国は産業競争力・科学技術力において急速に力を失うだろう。本報告書では、ビッグデータの利活用に係る専門人材育成に関する課題とあるべき姿を整理し、そのための施策を提言する。

1.2. 我が国の課題

今後多くの価値がビッグデータの利活用から生まれることが実証されつつあるにも関わらず、我が国では、未だに主にKKD(勘と経験と度胸)によってビジネスが運営され、政策が作られている。その主要な要因の1つはデータの利活用に係る教育の不足にあると考える。ビッグデータの利活用に係る専門人材、すなわちデータサイエンティストの育成が急務であるが、我が国の現在の高等教育においては、本提案書で述べるようなスキルセットを持ったデータサイエンティストを育てる学部・専攻は存在しない。

次章で述べるデータサイエンティストに必要なスキルの中で、教えるべきカリキュラムが比較的明確であるのは、統計の分野である。この統計教育の観点から見ると、我が国では統計学は数学科のほかは、経済、医学、工学等の応用分野に埋め込まれた形で点在してい

る。この我が国特有の教育体制は統計をそれぞれの分野で生かしていくという面で、今まではメリットもあったが、機械学習、自然言語処理や大規模データ処理など最新の技術スキルを集中的に身につけるといっても、また急速に広がる新しい応用領域への対応においても、限界が見えてきている。また、現状の分野点在の仕組みでは、たとえカリキュラムを整備したとしても大量のデータサイエンティストを育てることは難しい。一方、統計教育が発達している欧米では、専門の統計学科の他に、生物統計学科があるなど、分野ごとの専門家も育成できるマスの育成体制になっている¹。

一方、スキルセットの観点からは、特に産業界は、統計の専門家というよりは問題解決能力を持つデータサイエンティストを欲しているが、現状はデータ解析経験の乏しい、数学を専門とする教員が統計教育を行っているなど、統計教育が数理統計に偏って形骸化しており、実務における問題解決能力と必ずしも結びついていない。

そもそも、我が国の場合は、人口当たりの理工系の学生数やアナリティカルスキル保有者数が極めて少ないことに代表されるように、理系の素養を持つ人が少ないことに根本的な問題があると考えられる²。我が国が学術やビジネスの世界で国際競争に勝つためには、統計的素養に加えて情報科学の知恵を持つ人、リアルな課題解決に熱中する人、課題を俯瞰し柔軟にビッグデータを処理できる人が必要である。また、高等教育だけでなく、国民全体に行き渡るリテラシーの涵養も重要である。分析的でデータドリブンな思考力が、従来の三種の神器(よく言われるのが、母国語、世界語、IT 技術の 3 つのセット。他のセットもありうる)に加えて必要という認識を、国全体で共有する必要がある。京都大学の共通教育改革の一環として行われた「統計入門」の取組はその一歩だが、長期的にみれば 10~15 年後の底上げには高校教育からの積み上げが必要で、そのためには、教える人を育てる仕組が必要である。

2. 我が国におけるデータサイエンティストへの要請

それでは、どのようなスキルセットを持った人材を育成すればよいのだろうか。この章では、データサイエンティストに要請されるスキルを産業界、アカデミアおよび地方自治体

¹我が国でも臨床試験に生物統計の専門家が必須とされ、製薬企業で爆発的に生物統計の専門家を雇うことになったが、その育成は生物統計講座において行われているので、多人数の育成ができる体制とはいえない。

² OECD の各国比較では、我が国の中学生の理科嫌いが長年指摘されているが、あまり教育政策等に反映されていない。情報系の学生数についても、我が国だけ情報系の区分に関する統計量が把握できていないという基本的な問題が根本にある。

の観点から整理する。

2.1. 産業界からの要請

2014年に発足した、主に産業界のメンバーからなるデータサイエンティスト協会は、データサイエンティストに要請されるスキルセットとして、ビジネス力、データサイエンス力、データエンジニアリング力の3つを定義した(図1)。

データサイエンティストに求められるスキルセット

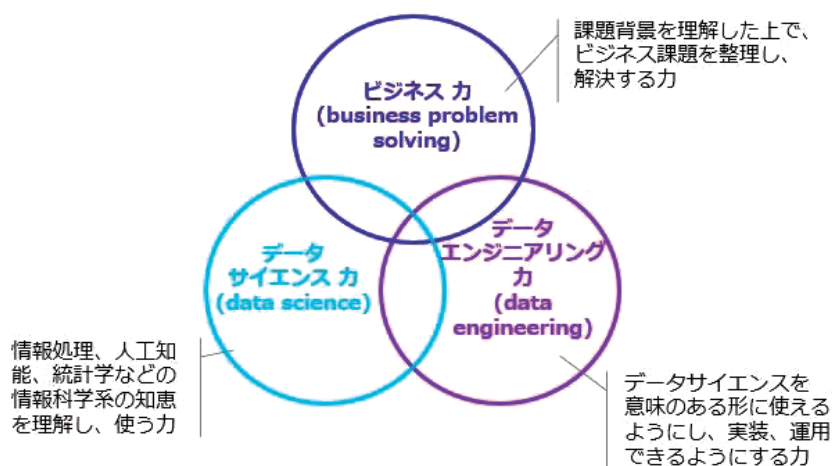


図1. データサイエンティストに求められるスキルセット
(データサイエンティスト協会スキル定義委員会の報告書より転載)

企業がデータサイエンティストに対して求めているスキルは、単にデータを収集してその統計分析をするだけでなく、むしろ「データに基づき予測し、その結果をビジネスに活用する」ことにある。統計学に加えて、機械学習や離散的最適化など最新の手法に精通すると共に、それらを用いたビジネス的な課題解決能力、すなわちデータと現実のビジネスをつなぐことのできる人材を求めている。このため、学部新卒者には基本的な問題解決能力、現実社会の課題との接近、実社会でのインパクトの認識、ITシステムへの実装能力などが求められる。さらに、アカデミアに比べて、ビジネスではスピード感がより重要であり、不完全な結果から素早く意思決定を行う、言わば「ザックリ感」も必要である。

一方、データサイエンティストは単なるITエンジニアでもない。ITエンジニアの主な仕事は、要件定義に対して、それを満たすシステムを納期・コスト・品質を担保しつつ作りこむことにあるが、データサイエンティストはむしろ、試行錯誤を通して新たな要件、新

たな価値を創造することが求められている。このため、データサイエンティストには、①顧客と建設的な会話をする能力、あるいは顧客の課題をデータ分析や情報技術に落としこむ翻訳力、②課題領域を見通して本質的な問題（イシュー）を見抜く能力、③課題解決のため各分野のエキスパートを動員できる能力の3つが必要である。

ビッグデータをビジネス価値につなげるには、専門人材であるデータサイエンティストの側だけでなく、国民一般の側でも、ビッグデータ分析がもたらす価値やデータサイエンティストの活動とその意味を理解するリテラシーが必要である。特に、データサイエンティストの成果をビジネスの中で活かしていくためには、その受け手となるミドルマネジメント層のデータリテラシーが不可欠であり、学校教育の成果を待つのではなく、ただちに再教育する場が必要である。

また、ビッグデータ分析の専門人材を育成すると共に、その価値を国民に広く知らしめるには、国家主導のフラッグシップ・プロジェクトの推進とその成果の発信も効果があるだろう。

以上をまとめると、産業界からはデータサイエンティストに関連して、具体的に次の3点による抜本的底上げと桁違いの人数の育成を期待されている。

- ① 大学でのデータサイエンティスト育成のためのプロフェッショナル教育
- ② 中等・高等教育を含む理系素養・データリテラシーのテコ入れ
- ③ 国家レベルのビッグデータ活用フラッグシップ・プロジェクト

2.2. アカデミアからの要請

まず、「あらゆる分野の研究者は、同時にデータサイエンティストでなければならない」という認識を持つ必要がある。大量のデータを産出する生命科学や天文学を始め、すべての科学の分野において、ビッグデータとその分析が科学の主要な手段のひとつとなりつつある。特に、宇宙・地球物理、高エネルギー物理学、物質・材料科学、生命科学、医療など最先端のサイエンスでは、ビッグデータの分析が研究の要となってきた。いわゆる「第4の科学」と呼ばれるこの方法論を実践するこれからの研究者は、すべからくビッグデータ利活用の専門家、すなわちデータサイエンティストでもある必要がある。そうでなければ我が国の科学技術の面での競争力を維持できない。このため、それぞれの分野の専門課程を学ぶ学生が、必ず並行してデータサイエンティストとしてのスキルも学べる仕組みをつくる必要がある。

機械学習や並列計算などビッグデータ分析の手法は 21 世紀に入ってから長足の進歩を遂げている。このため人材育成と並行して、我が国のアカデミアにおいても、常に最新の手法・応用の研究開発を怠ってはならない。すなわち、最先端のビッグデータ分析手法を研究し、第 4 の科学を牽引できるトップレベルのデータサイエンス研究者が必要であり、そのための育成拠点を設け、ビッグデータ分析手法・方法論の分野でも世界をリードしていく必要がある。現状わが国においては、個別の研究者は、世界に伍して戦えるが、組織的に大量に育てるという意味では全く不十分である。

一方、リテラシーのレベルでは、ピラミッドの底辺を広げる必要があり、リベラルアーツ教育の一環としてデータに基づく思考ができるようにすることが極めて重要であるが、残念ながら現状の統計教育では、学生に継続して興味を持たせることが難しい。ビッグデータ分析が社会にどのようにインパクトを与え社会を変えつつあるか、またデータサイエンティストがどのように活躍できるのかを実例を通して学べる機会を与えるべきである。特に、データ分析の実習においては、今までの統計教育にありがちな小規模の人工（サンプル的な）データではなく、リアル感に満ちたビジネスで得られた実データを産業界等から提供してもらえれば、迫力ある実習ができ、学生の興味を掻き立てることができるのではないと思われる。なお、学生に興味を持たせる入門の講義としては、京都大学の「統計入門」の取組が一つの参考になる。

産業界におけるイノベーションでもそうだが、科学技術の世界でも複数の分野にまたがって議論するところに大きな発見や発明がありうる。データサイエンスは複数の分野の研究者が視野を共有するための共通言語になりうるし、データサイエンティストは分野を結ぶ本質的な媒介者になりうる。このため、分野と手法の様々な組み合わせを議論できるマルチディシプリンの研究機関もイノベーションには欠かせない。人材育成の面では、ポストドクの T 型・II 型化再教育を目指している統計数理研究所の統計思考院の取組が良い先例になろう。

2.3 地方自治体からの要請

我が国に約 1700 ある市町村の多くでは今後急速な人口減が予想され、その中で地方創成をどう図っていくかが大問題となっており、地方自治体においてもデータサイエンティストを求める機運が急速に高まっている。今後、オープンデータの動きはますます加速すると予想され、それらも含めてビッグデータを活用した、エビデンスに基づく効果的施策の立案や、施策実施後の評価・再検討を担う人材の確保は、地方自治体にとって急務である。地方自治体にも産業界同様、データを戦略的に活用する部署を設置し、社会や組織を

横断する中核部隊として位置づけることが求められる。その部署を束ねる人材は企業における CAO³のような棟梁レベル（後述、3.4 節参照）が望ましいが、町村レベルだと、独り立ちレベル（3.3 節参照）のデータサイエンティストが外的機関（組織）に指導を仰ぎながら実務を担当することも有効であろう。一方で、ビッグデータビジネスや、データサイエンティスト人材育成にかかわるイベントや機会が東京に集中しており、地方自治体は自ら人材育成することが大変困難となっている。従って、全国レベルの底上げの必要性は明らかであり、例えば、全国的に一定数の拠点を設置する取組を進めることも考えられる。

3. データサイエンス人材育成のあるべき姿と実現に向けた仮説

産業界とアカデミアからの要請をまとめると、我が国においてデータサイエンティスト育成のあるべき姿は、図2のようになると考えられる。スキルレベルとしては、対数スケールで6段階としたが、それと前出のデータサイエンティスト協会スキル定義委員会が発表した4つのスキルレベルとの対応を図に示した。これらの人材が、大学や企業において継

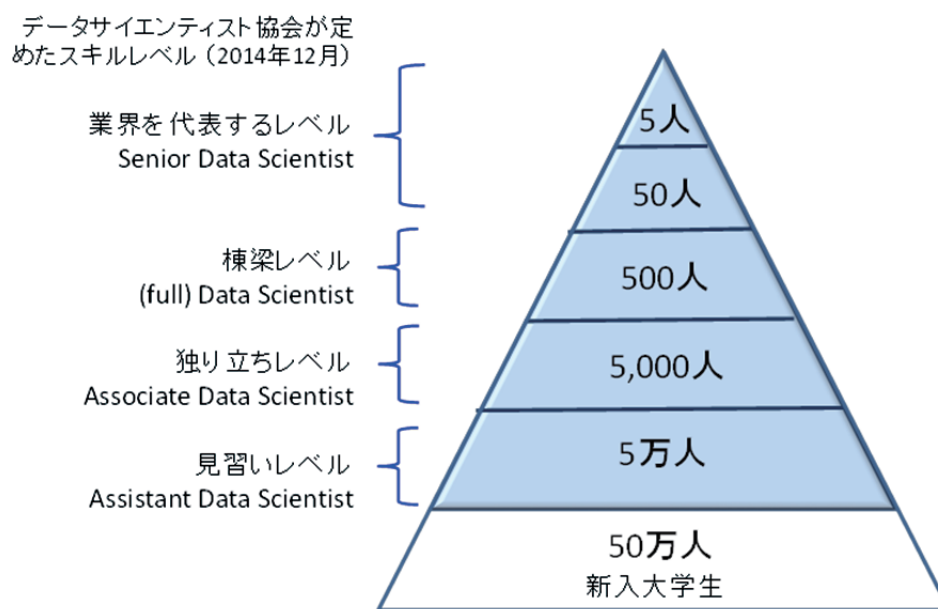


図2. 育成レベルと、毎年の育成目標人数

³ 企業において複数の応用分野を俯瞰的にマネージすることができ、データサイエンスの観点から全体最適な戦略を策定し実行できるようなリーダーシップを持った人材を CAO (Chief Analytics Officer) と呼ぶこともある。

続的に育成され、かつ効果的に利活用されている姿を、あるべき姿と考える。なお、毎年の育成目標数は概数であり、桁としてこの程度必要であると解釈されたい⁴。

3.1. データリテラシー

高校から大学学部レベルの素養（リテラシー）で、文系・理系を問わず全ての学生が持つべきであり、ベースになる統計的概念、データに基づく思考や問題解決の基礎概念、ITリテラシー、データに関する研究倫理などの内容を含む。我が国の大学の入学者数はおおよそ60万人であり、目標育成数はほぼその全数と考える。

3.2. 見習い(基礎能力)レベル

全てのデータサイエンティスト（実務家・研究者を問わず）が持つべき共通能力。この能力は同時に、多くのミドルクラスマネジメントにも必須である。理系の修士は全て、また文系でも経済・経営などの社会科学系や言語学や心理学等の専攻ではすべからず身につけるべき能力であり、この能力を持つ者は、適切な指導の下でビッグデータ利活用プロジェクトの一部分を担当できる。データサイエンティスト協会のスキル定義では、見習いレベルに相当する。我が国の理系の修士入学者数は、年間およそ5万人であり、これが目標育成数の1つの目安となる。

我が国には、データサイエンスに関する学部・学科は存在しないが、そのような学部・学科が今後設置されたならば、学部生が卒業までの3年・4年次に習得するレベルがこの見習いレベルになる。

3.3. 独り立ちレベル

データサイエンティストとしての専門能力、すなわちビジネス、データサイエンス、データエンジニアリングのいずれかの分野で専門的な能力を持ち、自らのイニシアチブで高度なデータ分析・問題解決能力を発揮する。実務経験が必須であり、大学においては修士ま

⁴参考としてデータサイエンティストの必要数に関する推測(公表されたもの)を下記に記す。

- ・ McKinsey Global Institute Report on Bigdata (2011): "deep analytical talent"が2018年までにUSで14万人から19万人不足。また、"data-savvy manager and analyst"については150万人不足。
- ・ ガートナー(2011):「日本ではビッグデータ関連雇用が36万5000人分増える見込みだが、実際に雇用条件を満たせる人材は11万人程度しかいない」
- ・ 日経(2013.7.7):「国内ではデータサイエンティストが約25万人不足する」

たは博士課程での具体的な PBL(Project-Based Learning)の経験が求められる。データサイエンティスト協会のスキル定義における独り立ちレベルに相当する。データサイエンスを主とする大学院の専攻においては、修士課程修了時までにはできれば習得しておいて欲しいレベルが独り立ちレベルになる。我が国における資本金 10 億円以上の会社はおよそ 6,000 社であり、その企業の中で毎年 1 名程度ずつ採用を検討するとすれば、産業界向けだけでも 5,000 名程度が目標育成数の目安となる。

3.4. 棟梁レベル

データサイエンティストのチームを率いて、組織におけるビッグデータ利活用を先導できる能力をもった人。複数の応用分野を俯瞰的にマネージすることができ、データサイエンスの観点から全体最適の戦略を策定し実行するリーダーシップが求められる³。主に実務を通して育成される能力であるが、専門知識については大学院における社会人コースを用いて集中的に育成(例:情報学研究所における高度 IT 人材育成「トップエスイー」プログラム)することも考えられる。毎年 5,000 名の「独り立ちレベル」を指導統括するこのようなリーダーは、組織論の観点から見れば、「独り立ちレベル」6~15 人につき 1 人程度必要と考えることができる⁵。したがって、目標育成数は、「独り立ちレベル」より 1 桁少ない年間 500 名程度であろう。

3.5. 指導的データサイエンティスト

学術においてはデータサイエンスの最先端を切り開くワールドクラスの研究者・開発者として指導的な能力を発揮する者、また産業界においては、業界におけるビッグデータ・データサイエンスに基づくイノベーションを牽引できる者。年間数名から数十名が現実的な目標育成数と考えられる。なお図での世界的トップタレントとは、グーグル CEO のラリー・ページ級の人物を指す。このようなトップタレントを系統的に育成することは困難であるが、棟梁レベルを対象に、そのための素地をつくる取組が必要である。

3.6. 「中抜き」仮説

図 2 において、懇談会が最も深刻な問題と考えるのが、「棟梁レベル」の人材が育っていないことである。

⁵ Span of Control (1 人の管理職がマネージする直接の部下の数) は業種や組織によって異なるが、平均 6-7 人と考えられている。GE 社のガイドラインは、10-15 人とされている。

その大きな原因は、棟梁レベルの人材が活躍できる場がこれまで我が国になかったことである。アカデミアにおいては最先端科学技術に関わる研究プロジェクト推進におけるデータサイエンスの重要性は、かねてからある程度認識されていたと思われるが、これまでの研究プロジェクトでは、データサイエンティストは分野を専門とするプロジェクトリーダーの指示に従いつつ研究開発を行うことが通常であり、棟梁レベルのデータサイエンティストがリーダーシップを発揮して、複数の分野の専門家とともに革新的な研究開発をすすめる機会が十分に存在しなかった。同様の問題は、部門縦割り主義の弊害が叫ばれている産業界においても多く見られ、データサイエンスを横串としたイノベーションを妨げる壁となっている。この状況は、企業におけるビッグデータ利活用に責任を持つ CAO (Chief Analytics Officer) が急速に普及しつつある米国⁶とは大きく異なる。

我が国におけるこの「中抜き」の状況をすみやかに解消しなければならない。棟梁レベルの人材が年 500 人規模で育てば、その中から世界的なトップタレントが現れてこの分野全体を引っ張っていくことが期待でき、また同時にこの棟梁レベル人材が、実務の傍ら随時大学や企業で人材育成を支援することで、独り立ちレベル以下の人材育成が促進され、スケールアウトが進むことを期待することができる(図 3)。

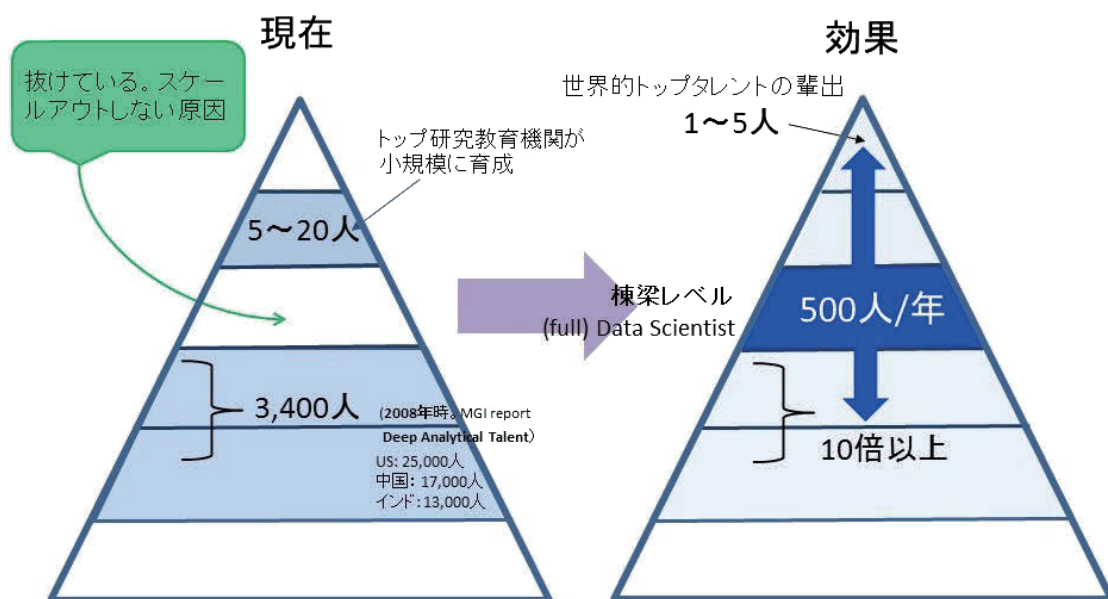


図 3. 育成が遅れ、最も欠けている棟梁レベル(左)と育成実現時の波及効果(右)

⁶ 米国における CAO の現状については、例えば *Computerworld* の以下の記事を参照。
<http://www.computerworld.com/article/2688352/chief-analytics-officer-the-ultimate-big-data-job.html>

4. 具体的施策

前章で述べた「あるべき姿」に向けて、懇談会が考える具体的な施策の例を示す。これらの施策は個別に実施すべき性質のものではなく、適切に組み合わせることで量と質の双方を担保した適切なポートフォリオを考えるべきである。

4.1. リテラシーの醸成（50万人規模）

- ・ 高校教育・大学教養での講義において、世界がデータとその利活用で大きく変わっている重大性を教える。たとえば、
 - 高校生・大学生がワクワクするような啓発書・教科書をつくる。
 - 社会でどう使われ役立っているかを示す事例のビデオ素材をつくる。
 - 活躍中のデータサイエンティストをプールし、大学や高校にこれらのデータサイエンティストを適宜派遣し講義を担当してもらう。
- ・ 大学基礎教育にデータサイエンスを取り入れる。
 - 大学 124 単位の内、共通教育で例えば 4 単位、専門教育では専門に応じて例えば 2 から 6 単位をデータサイエンス（統計）に割り当てる、と定める。この際、核となる週 1 時間の講義にコンピュータ実習や問題を解く演習もセットにする。
 - これに合わせて、基礎統計教育も見直す。
- ・ 社会一般の興味を惹くための施策を実施する。
 - 全国的なデータサイエンスに関するコンテストを実施する。スーパーグローバルハイスクール指定校等の先進的な取り組みを行っている高校にも積極的参加を呼びかける。
 - gacco における「社会人のためのデータサイエンス入門」のように、MOOC (Massive Open Online Course) を利用した一般向け教材を充実させる。

4.2. 見習いレベルの育成（5万人規模）

- ・ データサイエンス力の基本的スキルとしての、統計学、機械学習、最適化およびビッグデータ解析に必要なプログラミングやデータ可視化の基礎を学習させる。学部・大学院におけるデータサイエンスの教育カリキュラムに関しては、統計教育参照基準・情報教育参照基準に相当するものを早急に作成する⁷。

⁷ HP の一覧で確認できるだけでも現在 279 のデータサイエンス教育プログラムが存在する（巻末の表 1）。そのうち、修士レベルが 214。

- 大学院に、ダブルディグリー・ジョイントディグリー⁸の考え方を導入し、専門科目と共にデータサイエンスを副専攻などの形で学べるようにする。逆に、データサイエンスを専攻とする学生が、データサイエンスの適用分野を副専攻として学べる機会も提供する。データサイエンスのコースを持たない大学では、他大学の大学院とのダブルディグリーも視野に入れる。
- スケールアウトの方策として、MOOC(Massive Open Online Course)を積極的に利用すると共に、コンテンツの充実を図る。
 - gacco における「統計学 I, II, III」を参考に、データサイエンス向けのコンテンツの充実を図る。
 - また、Coursera における Data Science の証明書の発行などを参考に、MOOC 修了者に対して修了書を発行する、あるいは、大学における単位認定を行うなど、MOOC の位置付けを社会的に認知させる。
- 社会人に対しては、特に機械学習など新技術に関して、ミドルマネジメント層を含む広い対象に、再教育のプログラムを提供する。統計数理研究所の公開講座(主に初級)の取組を参考に、地方を含む各大学に展開する。この際、必要に応じて産業界からの寄付を募ることも検討する。社会人の学び直しを推進するため、文科省において検討が進められている「職業実践力育成プログラム」認定制度(大学等における社会人や企業等のニーズに応じた実践的・専門的なプログラムを国が認定する制度)の活用(見習いレベルの育成に限ったものではなく、各レベルでも活用しうる)も効果的と考えられる。

4.3. 独り立ちレベルの育成 (5000 人規模)

- 問題設定能力、問題解決のための戦略立案能力、データの収集・キュレーション能力、データ分析結果の業務や事業への実装能力、異分野研究者や事業者との連携・コミュニケーション能力、研究倫理、情報セキュリティの能力を備え、独立してデータサイエンスを推進できるレベルを目指す⁹。
- データ解析スキルとしては、MCMC (マルコフ連鎖モンテカルロ) 法、データ同化、インプューテーション技術、高次元空間の構造探索とモデル化、異種情報統合による個人化技術、隠れた関係の検出、特異性の発見、因果推論など、またデータ可視化技術

⁸ Double Degree は 2 つの大学が 1 つの学位論文に対して異なる学位記を発行すること、Joint Degree は、1 つの学位記に対して 2 つの専門分野を認定すること、である。これらに限らず、学位記とは別に、修了書を発行する形でもよい。

⁹ 日本学術会議 提言「ビッグデータ時代に対応する人材の育成」。米国の修士プログラムのカリキュラム例を巻末の表 2・4 に掲載。

としては、次元圧縮、特徴抽出、パターン認識などが重要とされているが⁹、データサイエンスを巡る状況は急速に変化しているので、不断の検討が必要である。

- 大学院において PBL に基づく専門育成プログラムを推進する。
 - 「分野・地域を越えた実践的情報教育協働ネットワーク(enPiT)¹⁰」、「先導的 IT スペシャリスト育成推進プログラム¹¹」などにおける PBL の取組が参考になる。
- 企業との連携を通して、企業の事例を PBL 化したり、インターンシップ・プログラムを通して実務を学べる機会を与える。
 - 産学連携では経団連主体の CeFIL（高度情報通信人材育成支援センター）¹²が参考になる。

4.4. 棟梁レベルの育成（500 人規模）

- 分野を俯瞰し戦略の立案・実行ができる棟梁レベルの人材は、常にデータサイエンスの最新の手法群(現時点でいえば例えばノンパラメトリックベイズ、スパースモデリング、カーネル法、グラフィカルモデル、深層学習、分散リアルタイム計算フレームワーク Storm¹³など)に加えて、新しい応用分野（例えば広告におけるリアルタイムビディング、マーケティングにおける協調フィルタリング、ロボティクスにおける画像・映像認識、質問応答システムの医療応用など）に精通している必要がある。このような「ホットな」情報は、実際にビッグデータを扱い、データサイエンスを研究開発している専門家が集中しているところに集まってくる。このため、ビッグデータ・データサイエンスの最先端の手法・応用の研究・開発及びそれらに精通した人材育成を行う国家的な拠点を設置する必要がある。
- データサイエンティストとしての実務経験のある社会人を、棟梁レベルに育成する集中的プログラムを上記拠点で開講する。ここでは、最先端の手法を PBL を通して実地で経験し、各応用領域での最新の成果をケーススタディとして学ぶと共に、拠点の特質を活かして棟梁レベルデータサイエンティスト間の人的ネットワークも形成する。
 - 社会人に 1 年間の集中的な教育をほどこす情報学研究所「トップエスイープログラム¹⁴」を参考にし、社会人を受け入れてデータサイエンスにおける経験と人脈を形成させる統計数理研究所「データサイエンス・リサーチプラザ¹⁵」の取組を量的に拡充する。

¹⁰ <http://www.enpit.jp/>

¹¹ http://www.mext.go.jp/a_menu/koutou/it/

¹² <http://www.cefil.jp/>

¹³ <https://storm.apache.org/>

¹⁴ <http://www.topse.jp/>

¹⁵ http://www.ism.ac.jp/shikoin/overview/pdf/DSRP_lo_0323.pdf

- ・ 科学の諸分野において、既にビッグデータを活用した研究方法論をある程度身につけた人材（博士号取得者等）に対して、産業界やアカデミアにおいて活躍できるキャリアパスを見据えて再教育する。
 - 米国の Insight Data Science Fellow Program¹⁶では、物理や天文学など諸科学における博士号を持つ科学者をデータサイエンティストとして訓練して産業界に送り込んでいる。この取組を参考にし、情報・システム研究機構「データ中心科学リサーチコモンズ¹⁷」、統計数理研究所「統計思考院¹⁸」で行われている、データサイエンス研究者の育成プログラムを量的に拡充し、アカデミア及び産業界で活躍できる棟梁レベルのデータサイエンティストを育成する¹⁹。

4.5. 指導的データサイエンティストの育成（50人規模）

- ・ 3.5節で述べたように、指導的データサイエンティストは、ワールドクラスの人材であり、系統的に育成するのは難しい。むしろ施策としては、上記ビッグデータ・データサイエンス拠点において、世界最先端の手法・応用の研究・開発を推進することによって、このような指導的データサイエンティストが生まれてくる土壌を醸成することが大切である。
 - このため、国レベルで、ビッグデータのフラッグシップ・プロジェクトを実施し、その中で指導的データサイエンティストが活躍できる場を提供する。
- ・ 同時に、このような才能のポテンシャルを持つ者を若いうちに、発掘し、十分な機会を与える必要がある。このための施策として、以下の2点を提案する。
 - 定期的なデータサイエンス・ハッカソンを実施する。ITセキュリティの分野では、同様の試みがIPA「セキュリティ・キャンプ²⁰」として行われている。これを参考に、データサイエンスにおけるプログラムを企画する。
 - 産業界で活躍する指導的データサイエンティストになれる人材を育成するために、才能のある若い者にメンターをつけ、資金等の援助を与えると共に人脈を形成する機会を提供する。例えばIPA「未踏IT人材発掘・育成事業²¹」やJST「さきがけ」を参考に、指導的データサイエンティスト候補者向けのプログラムをつくる。

¹⁶ <http://insightdatascience.com/>

¹⁷ <http://rc.rois.ac.jp/>

¹⁸ <http://www.ism.ac.jp/shikoin/>

¹⁹ 米国では Insight Data Science プログラムを含め 5 週 - 13 週程度の bootcamp が 17 プログラム実施されている。<http://datascience.community/bootcamps>

²⁰ <https://www.ipa.go.jp/jinzai/camp/index.html>

²¹ <https://www.ipa.go.jp/jinzai/mitou/>

4.6. トリクルダウン効果とスケールアウト

上記の具体的な施策の中で、最も重要なのが、現在わが国には決定的に不足し、ボトルネックとなっている棟梁レベルの育成であり、このレベルの人材が育ってくれば、独り立ちレベルのための PBL の指導者、見習いレベルのための専門科目の講義、リテラシーのための講師派遣などを通して、本報告書で述べたデータサイエンティスト育成の全体像がスケールしていくものとする。このためには産業界とアカデミアの密接な連携が欠かせない。

なお、これらの棟梁レベルデータサイエンティストについては、それらの間で情報交換、相互研鑽のためのコミュニティを作り、このコミュニティが、次世代のデータサイエンティストを育成する指導的役割（高校・大学等での講義も含む）を与える。

十分な数の「棟梁レベル」が存在しない現時点では、まずは国家的拠点の設立を急ぎ、あわせて全国主要 10 大学程度で他のレベルの人材育成をスタートすると共に、MOOC などのオンライン教材の利用を積極的に勧め、地方を含めた全国的な波及効果を狙うべきである。

4.7. 出口戦略

育成された人材は、民間・アカデミアを問わず広く活用されなければならない。このためには、ビッグデータ利活用とデータサイエンティストの重要性を広く社会に発信していくことが重要である。また、データサイエンティスト教育を受けた者、あるいはある一定のスキル基準を満たす者に対してスキル認定を行い、雇用する側とデータサイエンティストのスキルのミスマッチが起こらないようにすることも重要と考えられる。

5. まとめ

我が国のビッグデータ利活用専門人材育成の課題の解決に向けた取組みの考え方を下記の 3 点にまとめる。

- 我が国の問題の根源は、棟梁レベルの決定的不足にある。この解決のために国家レベルの拠点を設置して、年 500 名規模の「棟梁レベル」の人材育成をめざし、上層への成長や下層へのトリクルダウン効果も狙う。
- リテラシーレベルや独り立ちレベルの大学教育を加速させるために、主要 10 大学

程度で本報告書の提案に基づく人材育成をスタートすると共に、MOOC などのオンライン教材を整備し、全国への波及効果を狙う。

- 社会全体のリテラシーやアウェアネスを向上させるために、全学的教養教育の実施、国家レベルのフラッグシップ・プロジェクトの推進、コンテストの開催、映像素材の充実などの取組を行う。

これらの方策を実現するにあたっては、データサイエンスを副専攻とするダブルディグリー制人材育成の推進やスキル認定制度も有効と考えられる。

参考資料

公開情報

- 日本学術会議 情報学委員会 E-サイエンス・データ中心科学分科会提言「ビッグデータ時代に対応する人材の育成」平成 26 年 9 月
- 日本学術会議 数理科学委員会 数理統計学提言「ビッグデータ時代における統計科学教育・研究の推進について」平成 26 年 8 月
- 統計検定出題範囲表 (4 級 - 1 級)
- J. Manyika, M. Chui, et al., Big Data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute, (2011).

各委員からの提出資料 (内部資料)

- 安宅委員提出資料 「データ社会に求められる新しい才能とスキル」
「データ時代に向けたビジネス課題とアカデミアに向けた期待」
- 岡本委員提出資料 「データキュレーター育成と大学教育への期待」
- 北川委員提出資料 「日本学術会議提言について」
「ビッグデータ解析技術」
「海外大学におけるデータサイエンスプログラムについて」
- 佐藤委員提出資料 「統計教育に関する取組」
「統計家の行動基準」
- 樋口委員提出資料 「分野の壁を突き崩すビッグデータ」
「データに関連した数理分野の俯瞰図」
- 丸山 (宏) 委員提出資料 「データサイエンティスト - 現状と課題」
- 渡辺委員提出資料 「統計科学・データサイエンス教育の体系化」
「ヘルスデータサイエンス」

表1 海外におけるデータサイエンス関連教育プログラムの数

	US	GB	IE	FR	NL	ES	Others	Total
Bachelors	15	2	0	1	0	0	2	20
Masters	126	37	7	5	6	5	28	214
Doctorate	11	1	0	0	0	0	2	14
Certificate	29	0	1	0	0	0	1	31
Total	181	40	8	6	6	5	33	279

Colleges with Data Science Degrees (<http://datascience.community/colleges>) より編集

表2 修士課程のカリキュラム例：

Master of Science in Analytics, North Carolina State University

Summer		Fall			Spring		
0	1	2	3	4	5	6	7
Statistics Primer	Analytics Tools	Practicum: Data Security & Privacy, Data Cleaning, Geospatial Analysis, Project Management, Leadership & Teamwork, Professional Writing, Legal Issue, Data Visualization, Presentation Skill, Conflict Resolution and Negotiation					
	Analytics Foundations	Logistic Regression	Optimization	Simulation & Finance	Risk Analytics	Financial Analytics	Advanced Data Mining
		Linear Algebra	Text Mining	Customer Analytics	Design of Experiments	Web Analytics	Special Topics
		Data Mining		Machine Learning	Survival Analysis	Advanced Exploratory & Outliers	
		Advanced Programming					
		Time Series & Forecasting		Big Data			

North Carolina State University の HP より作成

表3 修士課程のカリキュラム例：

Master of Science in Analytics, Northwestern University

FALL	WINTER	SPRING	SUMMER	FALL
Professional Practicum			Internship (minimum 10 weeks)	Capstone Design Project
Analytics for Competitive Advantage	Analytical Consulting Project Leadership	Data Visualization		Leadership or Analytical Organizations and Functions
Introduction to Databases & Information Retrieval	Data Mining	Analytics for Big Data		Elective 2: Choose from Healthcare Analytics, Predictive Models for Credit Risk Management, or Text Analytics
Statistical Method for Data Mining	Predictive Analytics	Introduction to Data Warehousing and Workflow Management		
Optimization and Heuristics	Introduction to Java Programming	Elective 1: Choose from Analytical Topics with Python, Social Network Analysis, or Marketing Models		

Northwestern University の HP より作成

表4 修士課程のカリキュラム例：

Master of Science in Statistics, Data Science, Stanford University

(1) 標準的な5クォーターのスケジュール

Autumn	Winter	Spring	Autumn	Winter
Linear Algebra with Applications to Engineering Computations	Advanced Programming for Scientists and Engineers	Modern Applied Statistics: Data Mining	Numerical Linear Algebra	Discrete Mathematics and Algorithms
Introduction to Programming for Scientists and Engineers	Convex Optimizations I	Stochastics Models in Engineering	Introduction to parallel Computing using MPI, openMP, and CUDA	Modern Applied Statistics: Elements of Statistical Learning
Introduction to Statistical Inference	Introduction to Regression Models and Analysis of Variance			

Stanford University の HP より作成

(2) 5つのコアエリアに関する単位要件

Requirements 1	Requirements 2	Requirements 3	Requirements 4	Requirements 5
Mathematical Core (12 units)	Advanced Scientific Programming and High Performance Computing Core (6 units)	Statistics Core (12 units)	Domain Specializations or Preparatory Courses (9 units)	Practical Components(6 units)
<ul style="list-style-type: none"> Linear Algebra Numerical Optimization Discrete Mathematics 	<ul style="list-style-type: none"> Advanced Programming for Scientists and Engineering Software Design in Modern Fortran for Scientists and Engineering Computer Organization and Systems Large Scale Software Development Introduction to Parallel Computing using MPI, openMP and CUDA Parallel Methods in Numerical Analysis Parallel Computing Parallel Computer Architecture and Programming Parallel Computing Research Project 	<ul style="list-style-type: none"> Introduction to Statistical Inference Regression Models/Statistical Modeling Learning Data Mining 	<ul style="list-style-type: none"> Parallel and Distributed Data Management Topics in Computer Graphics Social and Information Network Analysis Modern Statistics for Modern Biology Human NeuroImaging Methods Human and Machine Learning, Analytics from Big Data, Data Driven Medicine, Geostatistics, Representations and Algorithms for Computational Molecular Biology 	<ul style="list-style-type: none"> A capstone project (computational) The new Data Science Clinic Statistical Consulting

Stanford University の HP より作成

ビッグデータの利活用に係る専門人材育成に向けた産学官懇談会の審議経過

第1回

日時: 2015年4月18日(土) 13:00~15:00

場所: 情報・システム研究機構 URAステーション(城山トラストタワー33F)

第2回

日時: 2015年5月8日(金) 10:00~12:00

場所: 情報・システム研究機構 機構本部(ヒューリック神谷町ビル 2F)

第3回

日時: 2015年5月20日(水) 10:00~12:00

場所: 情報・システム研究機構 機構本部(ヒューリック神谷町ビル 2F)

第4回

日時: 2015年6月10日(水) 10:00~12:00

場所: 情報・システム研究機構 機構本部(ヒューリック神谷町ビル 2F)

第5回

日時: 2015年6月24日(水) 10:00~12:00

場所: 情報・システム研究機構 URAステーション(城山トラストタワー33F)

第6回

日時: 2015年7月17日(金) 15:30~17:30

場所: 情報・システム研究機構 URAステーション(城山トラストタワー33F)

ビッグデータの利活用に係る専門人材育成に向けた産学官懇談会出席者

座長	北川 源四郎	情報・システム研究機構 機構長
委員	安宅 和人	ヤフーCSO、データサイエンティスト協会 理事
	榎本 剛	文部科学省 研究振興局 参事官(情報担当)
	岡本 青史	富士通研究所
	北山 浩士	文部科学省 高等教育局 専門教育課 課長
	佐藤 俊哉	京都大学医学研究科 教授
	長谷川 真理子	総合研究大学院大学 理事・副学長(教育担当)
	樋口 知之	統計数理研究所 所長
	丸山 宏	統計数理研究所 教授/データサイエンティスト育成ネットワーク事業 実施担当責任者
	丸山 文宏	富士通研究所
	渡辺 美智子	慶應義塾大学健康マネジメント研究科 教授/(独)統計センター理事
陪席	栗辻 康博	文部科学省 研究振興局 数学イノベーションユニット次長 基礎研究振興課 融合領域研究推進官
	金井 学	文部科学省 高等教育局 専門教育課 情報教育推進係長
	栗原 潔	文部科学省 研究振興局 参事官(情報担当)付専門官
	土生木 茂雄	文部科学省 高等教育局 専門教育課 視学官
	山路 尚武	文部科学省 高等教育局 専門教育課 課長補佐
事務局	北村 浩三	情報・システム研究機構 シニア URA
	岡本 基	情報・システム研究機構 URA
	竹ノ内 広和	情報・システム研究機構 URA ステーション
	由村 弘子	情報・システム研究機構 総務課総務係