

# 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<sup>1</sup>。

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<sup>2</sup>。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

---

<sup>1</sup>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。

<sup>2</sup> 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

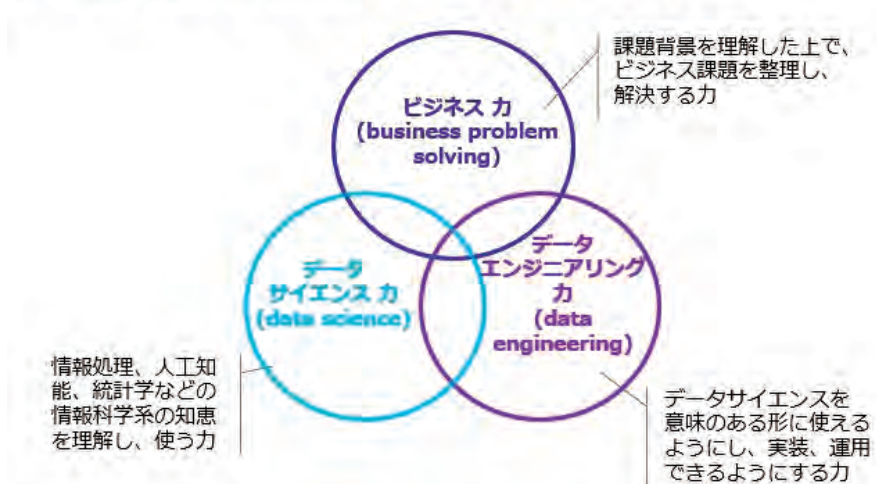


図1. データサイエンティストに求められるスキルセット  
(データサイエンティスト協会スキル定義委員会の報告書より転載)

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