

データサイエンス時代の横幹連合

横幹連合会長 北川 源四郎*



1. はじめに

P. F. ドラッカーによれば、西洋社会はこれまで数百年に一度大きな転換期を迎え、数十年をかけて世界観、価値観、社会構造、政治構造、技術や芸術も変えて、その後新しい世界が生まれてきた。そして彼は人類社会は今まさにそのような転換期の真っ只中にあり、2020年頃まで続くであろうと予言した [4]。それから25年が経過した現在、実際に人類社会は大きな変化に直面している。社会の隅々に配置されたセンサーからは日々大量のデータが生み出され、グローバルに連結したネット社会にはビッグデータが出現し、その結果、通信・情報発信の手段、情報検索の方法、販売・流通のシステム、さらには犯罪捜査の方法に至るまで社会のシステムが一昔前とは一変している。これらは単なる10年に一度程度の技術革新ではなく、蒸気機関が産業革命を引き起こし工業化社会を生み出したように、ビッグデータの出現とその利用技術の飛躍的発展がデータ駆動型社会への転換をもたらそうとしているように思える。

このようなデータ駆動型社会への転換にあたって、ビッグデータにもとづく知識獲得や価値創出が実現の鍵であり、データサイエンスやAIが脚光を浴びており激しい国際競争が始まっている。

データサイエンスのための科学技術も、横幹連合が目指す横断型の基幹科学技術と位置付けられるように、社会の発展に対応して横幹技術自体も急速に拡大している。横幹連合も今後の発展の方向や新たな可能性を改めて考えてみる必要があると考えられる。

横幹連合のこれまでの活動については「横幹連

合の歩み」[3]に詳しいが、横幹連合は20世紀の後半に科学技術が極限まで細分化され専門化された状況に対して、横断型基幹科学の重要性を訴え、推進することを目的に2003年に30の学会が結集して設立された。当初は横断型基幹科学技術や知の統合の定義づけやビジョンの構築が行われたが、それらの活動の成果は2005年の「ことづくり長野宣言」に結実した。そこでは、

- ・ 知の統合に向けた学問の深化とその推進
- ・ 横断型基幹科学技術を活用した社会問題解決
- ・ 知の統合を推進・定着させるための人材育成

が横幹連合の活動方針として宣言され、その後、横幹コンファレンスや横幹技術協議会と共催の横幹技術フォーラムの開催や、会誌「横幹」の発行が続けられてきた。2010年頃からは課題解決活動の実践のフェーズとなり、いくつかプロジェクトが推進されとともに、知の統合技術として知のプラットフォーム構築やシステム科学技術の新興に向けた活動が行われてきた。

ただし、設立以来15年が経過し、ビッグデータの登場により社会は大きく転換しようとする現在、横幹連合はより広範な横断型の科学技術を積極的に取り込み、次世代における更なる発展を目指す必要がある。本稿ではこのような検討の一助となることを期待して、現在のデータサイエンス特にデータサイエンティスト育成の動きを概観し、次に横幹連合の今後について私見を述べてみたい。

2. ビッグデータの登場とデータサイエンス

情報通信技術とくに高機能センサーの飛躍的発展によって、人間社会のあらゆる活動が網羅的かつ精細に記録できるようになり、ビッグデータの時代が出現している。その結果、グーグル、アマゾ

*東京大学 数理・情報教育研究センター 東京都文京区本郷7-3-1

ン、アップルに代表されるデータ駆動型の産業やネット関連企業が製造業、金融業、商社に代わって株価時価総額の上位を占めるようになった。これは、世界経済発展の原動力がものづくり革新からデータ活用に基づく価値創出に急速に転換していることを象徴している。

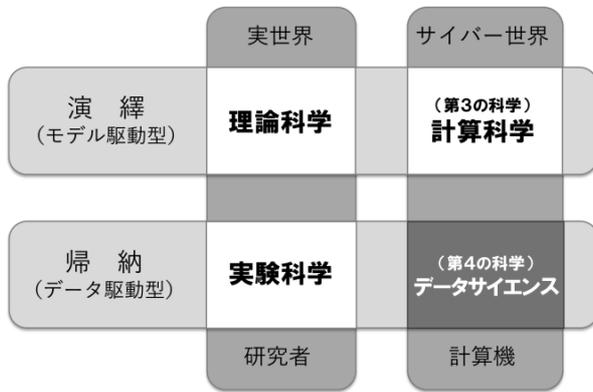


Fig. 1: 4つの科学的方法論

このようなデータが遍在する社会においては、個々のデータの価値は崩壊し、ビッグデータの中から如何に意味を見出すか、如何に価値を創出できるかに勝負がかかっている。そのための大規模データにもとづく価値創造の方法論として、データサイエンスが理論科学、実験科学、計算科学に続く第4の科学的方法論として出現している。20世紀までの科学技術は実験科学と理論科学の二つの科学的方法論によって駆動されてきた。言うまでもなく、実験科学は帰納的・データ駆動型の方法論、理論科学は演繹的・モデル駆動型の方法論であるが、いずれも実世界における研究者個人の経験や才能に依拠した方法であった。しかし、20世紀の終盤になると、複雑な現実に対応する非線形系や複雑システムの理解や予測において、理論的・解析的方法では対応できなくなり、飛躍的發展を遂げた計算機を活用して、シミュレーションを中心とする計算科学が確立し、流体解析や創薬などで大きな成果を挙げるようになった。

そして、21世紀の現在、ビッグデータの出現によって、ICTをフルに活用した新しい帰納的方法であるデータサイエンスが、データ駆動型の社会の

確立に不可欠な第4の科学的方法論として登場してきた。計算科学とデータサイエンスはサイバー世界における科学的方法論である、実空間とサイバー空間における演繹と帰納の4つの方法を完備することによってこれからの知識創造社会は推進されることになる。

今後のデータサイエンスの推進において、最も大きな課題はそれらの一連のプロセスを担うデータサイエンティストの不足である。このような新たな社会の到来を見通して、欧米では既に2005年ごろから統計教育の強化が図られ、アメリカの修士学位授与者数は21世紀初頭の年間1000名から、現在では4000名を超えるまでに増加している。また、統計学修士の学位授与大学数も現在では200を超えている[5]。統計教育の強化は欧州および近隣アジア諸国でも同様であり、統計学科を設置した大学数は中国で200以上、韓国でも50以上になっている[1]。

さらに、2012年のオバマ・ビッグデータイニシアティブ[6]のインパクトを受けてその翌年から欧米諸国ではデータサイエンスの専門教育プログラムが急速に開始され、現在ではその数は580を超えて、ほとんどの主要大学ではデータサイエンス教育体制の立ち上げが完了したものと考えられる[7]。

一方、日本においては、これまで大学に独立した統計学科や統計学部を置かず、経済学部、工学部、数学科、医学部などに小規模の講座を置いて、それぞれの領域の課題に即した統計教育を行ってきた。この日本の統計教育体制は、実は領域研究分野と方法論研究分野をクロスさせ現実世界の課題発見・課題解決ができる人材育成を目指すデータサイエンティスト育成の方法を先取りしてきたとも言える。ただし、この特定分野に特化した教育には2つの大きな欠点があった。まず第1に、この方式では分野に適合した人材育成ができる反面、過剰適合の結果、横展開を積極的に担う人材が育ちにくいという問題があった。更に深刻な第2の問題は、従来の徒弟制度的な丁寧な教育ではスケール化が困難で現在のデータサイエンティストの大量需要に全く追いつかないことである。

実際、日本学術会議は2014年の提言「ビッグデー

「データサイエンス時代の対応する人材の育成」において、日本の統計教育の特徴であった分野点在型の教育では、急速な分野拡大に対応できる汎用性を持った人材を大量育成できないことを指摘し、データサイエンスの専門育成機関の設置を提案している [1]。さらにこの提言では、データサイエンスの担い手は、基本的スキルとして、データ解析能力、データ可視化能力、ビッグデータ処理能力が必要だが、それに止まらずに、領域科学の知識と理解、課題設定・企画立案能力、異分野交流のためのコミュニケーション能力、さらには研究倫理や個人情報保護に通暁していることが必要とした。結果として、データサイエンティストはT型・II型人材となり、そのような専門人材の育成のためにはデータサイエンスを主専攻、領域科学を副専攻とする教育体制が効果的であると指摘している [1]。ただし同時に、これを一から始めるのでは現在の産業界の要請や国際的潮流には対応できないことから、速成の方法として、領域科学の博士号取得者に対して、統計・数理・情報の再教育を行うことを提案している。

この日本学術会議の提言を受けて、情報・システム研究機構に設置された産学官懇談会の報告「ビッグデータの利活用のための専門人材育成について」[2]では、データ活用人材に対する産業界、アカデミア、地方自治体からの要請を整理したうえで、データサイエンティストのレベルを定め、それぞれのレベルの人材の育成方法について検討して、最後に3つの提言を行っている。想定したレベルに関しては、一般社団法人データサイエンティスト協会が定めた4つのレベル（見習いレベル、独り立ちレベル、棟梁レベル、業界を代表するレベル）[8]を全面的に採用しているが、懇談会では今後はすべての国民がデータサイエンスを理解し、活用する能力を持つべきという観点から、底辺にリテラシーレベルを追加している。

この懇談会の提言の特徴は、それぞれのレベルにおいて毎年育成すべき人数を明記したことで、現在の社会の要請にこたえ、今後のデータ活用社会の発展を推進するためには、すべての大学生を対象とする年間50万人規模のリテラシーレベルの教育、5万人規模の見習いレベル、5000人規模の独り立ちレベル、500名規模の棟梁レベルの4段階の

データサイエンティスト育成が早急に必要であることを指摘している。さらに、データサイエンティストのもたらす成果や新しい価値を社会が受け入れ活用できるためには、マネジメント層を始め、すべての国民のデータリテラシーが不可欠であることも指摘している。

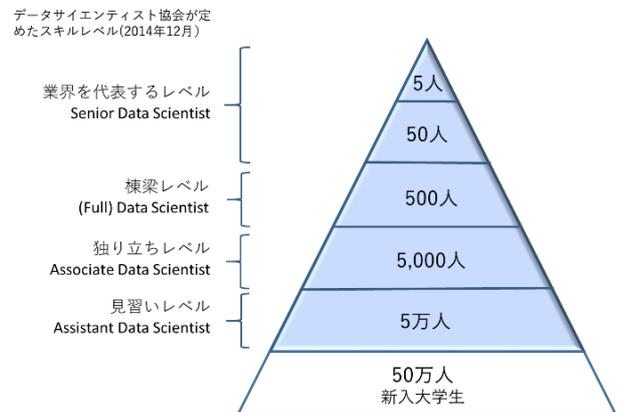


Fig. 2: 育成レベルと、毎年の育成目標人数：(情報・システム研究機構 ビッグデータの利活用に係る専門人材育成に向けた産学官懇談会、「ビッグデータの利活用のための専門人材育成について」の図2を転載)

この報告書では、まとめとして特に重要な、①年間500人規模の棟梁レベル人材の育成、②リテラシーレベルおよび独り立ちレベルの大学教育の加速のために、主要10大学程度で提案に基づく人材育成の開始、③全学的教養教育の実施と国家レベルフラグシッププロジェクトの推進の3点を提言している。このような提言を受けてようやく我が国でも、データサイエンス学部の設置、公募による全国6大学における数理・データサイエンスの全学教育のための教育研究センターの設置と標準的カリキュラムの構築とそれらの成果の全大学への波及を目指すコンソーシアムの設立など、一貫したデータサイエンティスト育成が開始されている。

3. データサイエンス時代における横幹連合の役割

すでに冒頭で述べたように、横幹連合は2003年の設立後、横断型科学技術の確立を目指して、横断型科学技術の再定義、ことづくり長野宣言、知

の統合の推進, 課題解決活動の実践を行ってきた。これらは横幹連合の設立の意義や活動方針を明確化して, 横幹技術の可視化に重要な役割を果たしてきたといえる。しかし, 近年, 画像認識, 音声認識, 機械翻訳, ロボット, 自動運転, ゲームなどの分野における深層学習や人工知能の応用の成功は目覚ましく, 特にその進歩のスピードには圧倒されるばかりである。設立から15年が経過し, 社会が急速に転換しようとしている現在, 社会の動きを改めて捉えなおし, 関連学会やスコープの拡大や多様化を図るなど横幹連合の在り方を原点に立ち戻って考え直してみる必要があると考えられる。横断型の学会連合には何よりも時代の変化を取り込み, さらに先導できる先見性や多様性が不可欠と思われる。

現在, 2つの側面で研究方法の連結や統合が重要な課題となっているように思える。実世界とサイバー世界の連結と演繹と帰納の連結である。実世界とサイバー世界の連結はサイバー世界の方法論である計算科学とデータ科学に不可欠なもので, 現実世界の課題をモデル化によってサイバー世界に投影し, サイバー世界で知識発見, 価値創出, 最適化を行い, それを実世界に反映させることによって, 現実課題の解決を行おうとするものである。このようなサイクルは, 実世界→モデル空間→実世界のサイクルやスパイラルを想定してきたリスク科学やシステム科学技術とも類似したものである。

一方の, 演繹と帰納の統合も重要な課題である。認識の科学においては, 仮説の提示と仮説の検証のプロセスは峻別され, 演繹と帰納は意識的に分離されていたと思われる。しかしながら, 予測や知識発見などを含む設計の科学においてはこれらを分離する必然性はないと思われる。気象学や海洋学で始まり, 近年は工学や生物学でも使われるようになったデータ同化はシミュレーションの限界を

データの情報によって補完しようとしている。一方, 近年進展が著しいマテリアルインフォマティクスでは, 機械学習や深層学習などの帰納的方法の限界を物理モデルに基づくシミュレーションによって補完しようとしている。

このように, 演繹的方法と帰納的方法の距離はサイバー世界においては著しく狭まっており, 今後はむしろ不可分のものになると考えられる。あるいは, 4つの科学的方法論を繋いで知識発展のスパイラルを実現することが重要なかもしれない。これは知識創造社会の発展に不可欠な方法論的基盤と考えられ, 横幹連合にとっても積極的に取り組むべき重要な課題と考えられる。

横幹連合が社会の転換に伴って生じるこのような新しい課題に果敢に挑戦するとともに関連学会を積極的に取り込んで, 益々発展することを願っている。

参考文献

- [1] 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, 「ビッグデータ時代に対応する人材の育成」, 2014年9月.
- [2] 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, 「ビッグデータの利活用のための専門人材育成について」, 2015年7月.
- [3] 横幹連合10年史編纂委員会, 「横幹連合10年の歩み」 2013年10月.
- [4] P. F. ドラッカー, 「ポスト資本主義社会」, ダイヤモンド社, 1993年7月.
- [5] Statistics and Biostatistics Degree Data, <http://www.amstat.org/asa/education/Statistics-and-Biostatistics-Degree-Data>
- [6] Big Data is a Big Deal, the White House, <https://obamawhitehouse.archives.gov/blog/2012/03/29/big-data-big-deal>
- [7] College & University Data Science Degrees, <http://datascience.community/colleges>
- [8] プレスリリース, データサイエンティスト協会, <http://www.datascientist.or.jp/news/2014/pdf/1210.pdf>