

# ディスカバリーサイエンス再考

Discovery Science Revisited



富士通研究所フェロー  
有川ディスカバリーサイエンスセンター長

有川 節夫 Setsuo Arikawa

## はじめに

ディスカバリーサイエンス（発見科学, Discovery Science, 以下DSと略す）は、筆者が代表を務めた1998年から2001年にかけての文部省科学研究費補助金特定領域研究によるプロジェクト「巨大学術社会情報からの知識発見に関する基礎研究」の略称名として使ったのが始まりである。このプロジェクト（DSプロジェクト）には国内の第一線の研究者約60名が参加して、発見科学という新しい学問領域に関して様々な研究活動を行った。その中に、プロジェクト開始時から毎年開催してきた国際会議International Conference on Discovery Scienceがある。この会議は、プロジェクト終了後も継続して世界各地で開催され、今年20回目を迎える。

本稿では、DSの提唱から20周年の節目の年に当たり、構想時点における議論を振り返るとともに、機械学習や人工知能（AI）が大きく発展し、広く社会に定着しつつある現在において取り組むべき課題と方向性を示し、富士通研究所における有川ディスカバリーサイエンスセンターの活動方針について簡単に紹介する。

## 構想時におけるDSの基本的な研究項目

DSプロジェクトは、機械学習の基礎理論の展開、類推などの各種推論の研究、理論展開のための統一的な枠組みの構築・提案、具体的な問題解決のために展開した基礎理論とそれに基づく元の問題の解決、機械発見の基本的な理論展開など、組織だった研究成果<sup>(1)-(9)</sup>に基づき、世の中に科学技術者や巨大なデータが存在する限り必要とされ、発展し続ける新しい科学であるという考えのもとで、

想起されたものである。

当時のDSの研究は、(1) 科学的発見の新しい手法を与え、(2) 知識発見のための計算機ネットワーク環境を提供し、(3) 情報科学の新しい基礎分野を確立する、ことを目指したものであった。以下のような考えのもとで研究項目を設定して、各研究項目にそれぞれに一つの研究班を置き、研究活動を展開した。

### (1) 知識発見の論理

計算機による知識の発見、すなわち機械発見においては、個々の仮説や理論はもとより仮説空間自体の論駁可能性が問題になる。また、発見に関する研究は、事例に学ぶ実証的な側面を持つ一方で、厳密な表現形式と演繹体系を用いた理論的基礎付けを必要とする。

### (2) 推論による知識発見

(科学的)知識の発見には、上記のような演繹のほかに、帰納やアブダクションによる推論が必要である。アブダクションは、既知の知識では説明不可能な事象について仮説を立てて説明しようとする際に用いるもので、(科学的)発見の契機を与えるものである。

### (3) 計算学習理論に基づく知識発見

実験・観測データからの知識発見には、計算機にデータから学習させる機械学習が有効である。その基礎理論である計算学習理論に基づいた知識発見の方式について、特に学習に要する計算量に配慮して研究する必要がある。

### (4) 巨大データベースからの知識発見

こうした知識発見に関する研究は、最終的には統合してシステム化し、それを科学やビジネスにおける巨大なデータに対して実際に適用し評価し

て、次の研究に役立てるべきである。また、データマイニングの具体的な方法論と手法を確立する必要がある。

### (5) ネットワーク環境における知識発見

知識発見においては、人間と機械、人間と人間、機械と機械とのインタラクティブな共同作業が重視される。したがって、協調、競合、妥協、調停などの種々の様相を含むこうした共同作業を効率よく遂行・支援するための、ネットワークエージェントを含む計算機ネットワーク環境を研究開発する必要がある。

(1)の班には哲学者や数理論理学者も擁し、各項目はそれぞれの担当者を中心にした国際的にも注目された多くの独創的な研究成果に裏打ちされたものであり、新しい学問領域としての体系だった構成になっている。

この後、2001年度から4年間にわたる元田浩教授(大阪大学)による特定領域研究「情報洪水時代におけるアクティブマイニングの実現(略称:アクティブマイニング)」を経て、データの量が指数関数的に立ち上がるいわゆる情報爆発の時代を迎え、2005年から5年間にわたる喜連川優教授(東京大学)による特定領域研究「情報爆発時代に向けた新しいIT基盤技術の研究(略称:情報爆発)」が続くことになる。そして、2012年の米国オバマ前大統領による「ビッグデータ・イニシアティブ」の発表以来、一般社会も巻き込んで世界的な広がりを見せているビッグデータの時代が到来するのである。

### 新しいDSの展開

創設期のDSでは、巨大とはいえ、それはせいぜい数テラから数十テラバイトを指していた。文字どおり隔世の感がある。しかし日本においては、上述した一連の特定領域研究を中心にして、ビッグデータ時代が到来するまで、「ディスカバリーサイエンス」をはじめ「情報洪水」や「アクティブマイニング」、「情報爆発」などの新しいコンセプトを提示・提唱しながら世界の研究を先導してきた。その後のDSの研究は、DS関係者の地道な努力により国際会議が現在まで継続して開催され、固有の展開を続けているが、これからのDSには、以下のような新たな視点に基づく研究が必要である。

### ● 人間が理解できる機械学習

現在を第3次のAIブームというならば、DSプロジェクトは第2次AIブームのときに蓄積された技術や知見に基づいて着想された新しい学問分野としてスタートしたものである。当時のAIは、医療診断システムMYCINに代表される知識獲得を一つのゴールにしたエキスパートシステムの構築にその主眼があった。したがって、知識の表現や知識の獲得が中心的な話題であった。知識の獲得は、人間のエキスパートからのインタビューなどによる機械可読な形での獲得と、巨大なデータベースからの知識の自動獲得に大別されていた。

知識の表現としては、ルール形式の知識や確定節などの論理式の集合としての知識、決定木などの木構造のものなどがあり、いずれも機械だけでなく人間にも理解できるものであった。こうした巨大なデータを説明するための知識の獲得や機械学習の方法について、様々な研究がなされていた。

そういう状況にあったことから、機械学習と知識発見は近接関係にあり、何が学習されたかについては、複雑さや単純さの問題は別にして、その結果は人間にも理解できるものであった。したがって、学習で得られた知識を人間が理解できるように表現するという事は、問題にならなかった。しかし、近年のディープラーニングによる機械学習や認識などでは、学習の結果から、何が学習され、どのように認識されたのかを、人間に理解できる言葉で表現することは難しいという大きな問題がある。

これからの機械学習に基礎を置くDSの研究では、こうした学習や認識の結果を人間にも可読で理解できる表現に変換することが重要なテーマとなる。また、AIの第2次ブームの時代のような明確な知識の表現を獲得するために、超大規模な実データをふんだんに駆使できる最近のビッグデータ時代に相応しい、新しい機械学習の手法の研究開発も重要な課題である。そうしたことが可能になれば、機械学習の結果は、すなわちデータからのディスカバリーということになる。

### ● ビッグデータ時代における各種推論の再考

DSの分野を提唱する際に、帰納推論や類推をはじめとする各種推論に関して、定式化や議論のための統一的な枠組みの提唱・構築<sup>(2)</sup> 推論の可能性

などに関して理論的な研究を展開してきた。例えば、九州工業大学の篠原武教授は、正データからの帰納推論の能力がそれまで認識されていたよりはるかに強力であることを、その統一的枠組みを用いて証明した。<sup>(3)</sup> これは、帰納推論における記念碑的な貢献である。

類推に関しては、北海道大学の原口誠教授が、通常の演繹推論の拡張の形でその定式化に成功し、基礎理論を展開した。<sup>(4)</sup> 類推は、問題解決や課題発見において日常的に使っている推論である。与えられたいくつかの対象間に類似性を発見し、その類似性に基づいて一方の対象で成立する事実や知識をもう一方の対象に変換することにより、問題解決の手がかりを得たり、未知の事実や知識などを予測推論したりする推論の方式である。こうした類推において最も重要な役割を果たす「類似した事実や知識」に関して、この理論展開を行った時代と違って、今日では数多くの候補がネット上に存在している。したがって、それらを有効活用することによって、類推の効果的かつ実用的な利用が可能になると考えられる。

同様のことはほかでも考えられる。例えば、帰納的論理プログラムなどの第2次AIブームの頃の事実や知識を意識した学習や推論を、現在の計算機やネットワークの能力、ビッグデータの時代感覚で捉え直すことによって、DSとしての新たな展開が期待できる。

### ● データサイエンスとオープンサイエンス

DSに関連の深い領域として、データサイエンスが注目されている。データサイエンスは、数理統計学やデータ解析などの伝統的な領域の延長上にデータ工学や機械学習などの人工知能分野も巻き込んで、ビッグデータから新しい価値を発見・創造するものである。その略号が同じというだけでなく、我々のDSと近接関係にある領域である。特に、それを駆使して課題解決や課題発見を行うデータサイエンティストに社会的需要と関心が高まり、近年その育成に総務省や文部科学省なども注力している。<sup>(10)</sup>

また、国や自治体などから出される公的データだけでなく、公共交通機関などの産業界から提供される各種のデータ、大型の実験・観測・計測装置から提供されるデータ、更には公的研究資金を

基に展開された研究に活用されたデータと研究成果としてのデータなど、そのオープン化が推奨され、進行しつつある。<sup>(11)</sup> こうしたデータ収集に直接参画せずに、オープンになったデータを利活用する研究やビジネスも活発になってきた。学術的にはサイエンスのオープン化であるが、これには古典籍などのデジタル化・オープン化によってもたらされる人文系分野におけるオープンサイエンス、あるいはシティズンサイエンスも含まれる。

### ● 人間とインタラクティブに行うディスカバリー

ネットワークが浸透し高速化した現在、多くのAI応用においてもそうであるが、DSに関してその全てを最終的な形で機械が自動的に行う必要はない。既に定着している日本語ワープロにおける仮名漢字変換や、機械学習を使った最近の高度な医療診断で行われているように、可能性のある候補をユーザーや医師に提示して、その中から選択し、あるいはそれを参考にして判断し、決定してもらう、という機械の使い方が健全であり、多くの可能性を含んでいると考えられる。またDSに関しては、こうした機械とのやり取りは1回である必要はなく、相互に影響を受け合う人間と機械、あるいは人間を含む情報環境や社会環境との連携やインタラク션을繰り返して行ってもよい。更に、ネットワークの先にあるものが機械学習などのAIである必要もなく、ネットワークに人間がつながっていて、人間が判断して反応・応答してもよい。そのようなところから始めて、システムをコストや経験、蓄積されたデータなどを基にして、徐々に機械化していくことも現実的である。

### ● ディスカバリーインフォマティクス

DSに関する国際的な活動としては、冒頭で述べたDS国際会議が唯一のものという状況が久しく続いてきたが、2012年から数年にわたってアメリカ国立科学財団(NSF)がディスカバリーインフォマティクス(Discovery Informatics, DI)というワークショップを開催している。情報科学技術を駆使した科学的発見に強い関心を示し、組織的な議論を行って報告書を公開している。<sup>(12)</sup>

そこでは、知識の同化(assimilation)と推論を必要とする科学的発見のプロセスを解明し、こうしたプロセスの様々な側面の理解、自動化、改良、イノベーションのために知的な計算や情報システ

ムを適用することを目指している。重要課題として、情報の抽出とテキスト理解、モデル形成、様々な形の科学的知識を使った推論、実験の計画・実行・推進、モデルに基づく (model-guided) データ収集など、10個の課題を挙げている。また、基礎研究の領域として、情報と知識、インタラクション、オートノミーを挙げ、DIの主なテーマとして、発見の過程を計算面から支援すること、データとモデル、発見のためのソーシャルコンピューティングを掲げている。

**富士通研究所における活動**

2015年12月、筆者の富士通研究所フェロー就任に際して、有川ディスカバリーサイエンスセンター (以下、DSセンター) が開設された。このDSセンターは、上記のような経緯で開始された日本発世界初のDSという野心的な研究領域を、企業の研究所で新たな戦略的な観点から発展させることを目指してスタートしたものである。

DSセンターには、北海道大学、東北大学、東京大学、京都大学、九州大学、九州工業大学などの、世界の第一線で活躍中の著名な教授陣の参加を得ている。DSに関して各教授の専門に関係の深い研究課題を委託し、富士通研究所では主に人工知能研究所の研究員が連携対応して研究活動を行っている。研究所における連携セミナーに加えて、DSセンター長と研究所側の担当者が大学の研究室を訪問し (サイトビジット)、教授およびその研究室

メンバーとのセミナーを開催するなど緊密に連携して、常にゴールを意識した新しい形の産学連携活動を行っている。そのほかに、富士通研究所における通常の業務以外の社会的な課題解決を目指したユニークな活動に関しても、DSの観点から捉えた方向性を検討するなどの活動も行っている。

2017年度は、与えられた課題を解決する従来型のAIを、解くべき課題自体を発見するAIへ発展させるために、課題発見の汎用基盤技術を開発し、発見した課題を社会の制度設計やオープンサイエンスなどに活用することを目標としている。具体的には、以下の計画に基づいて研究活動を行うことにしている (図-1)。

- (1) マルチエージェント：ゲーム理論やマッチング理論などのマルチエージェント基礎理論を社会制度の課題発見と設計に応用する。
- (2) 学習と発見：説明可能な学習により学習結果を検証し、人や機械による課題発見を補助する技術を開発する。
- (3) データアーカイバ：IoT (Internet of Things) で発生したビッグデータから特徴・構造を抽出し、圧縮して格納する課題発見基盤を開発する。

なお、図において「地域コミュニティ」や「専門士コミュニティ」などが関係する社会制度設計に関しては、富士通研究所の社員による社外活動を位置付けたものである。

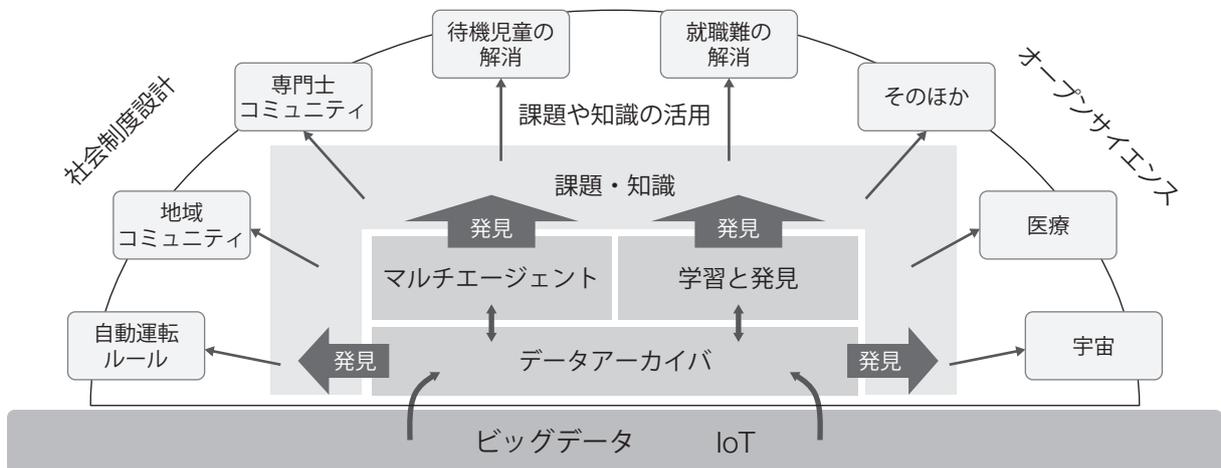


図-1 DSセンターにおける研究の狙いと計画の方向性

## おわりに

以上、DSに関して、その生い立ちから現在の課題、米国における動向、富士通研究所における研究活動などについて簡単に説明した。これからのDSの目標は、当DSセンターが掲げているように、課題の解決のための発見だけでなく課題の発見を明確に意識して展開すべきであると考え。そのためには、理論的な研究によって方式の限界を明らかにすることも重要である。その方式を棄却して新たな方式の探索に取り掛かるための重要なステップとなるからである。また、こうしたDSの研究は、現在行政レベルで注目されている「政策のための科学」としても機能し得るものと期待される。

## 参考文献

- (1) 原口 誠, 有川節夫: 類推の定式化とその実現. 人工知能学誌, 第1巻, 第1号, 132-139, 1986.
- (2) S. Arikawa et al.: Learning Elementary Formal Systems. Theoretical Computer Science, Vol.95, No.1, p.97-113, 1992.
- (3) T. Shinohara: Rich Classes Inferable from Positive Data: Length-Bounded Elementary Formal Systems. Information and Computation, Vol.108, Issue 2, p.175-186, 1994.
- (4) S. Shimozono et al.: Knowledge Acquisition from Amino Acid Sequences by Machine Learning System BONSAI. Trans. Inf. Proc. Soc. Japan, Vol.35, No.10, p.2009-2018, 1994.
- (5) Y. Mukouchi et al.: Towards a Mathematical Theory of Machine Discovery from Facts. Theoretical Computer Science, Vol.137, No.1, p.53-84, 1995.
- (6) 有川節夫: 巨大学術社会情報からの知識発見に関する基礎研究(平成10年度発足科学研究費補助金「重点領域研究」申請書). 1997.
- (7) 有川節夫ほか: 発見科学の構想と展開. 人工知能学会誌, Vol.15, No.4, p.595-607, 2000.
- (8) 森下真一, 宮野 悟(編): 発見科学とデータマイニング. 共立出版, 2000.
- (9) S. Arikawa et al.: Progress in Discovery Science. Lecture Notes in Artificial Intelligence, LNAI 2281, Springer-Verlag, 2002.
- (10) 文部科学省: 大学の数理・データサイエンス教育強化方策について. 数理及びデータサイエンス教育の強化に関する懇談会報告書, 2016.
- (11) 内閣府: 我が国におけるオープンサイエンスの推進の在り方について. 国際的動向を踏まえたオープンサイエンスに関する検討会報告書, 2015.
- (12) Y. Gil et al.: Final Report on the 2012 NSF Workshop on Discovery Informatics. National Science Foundation Workshop Report, 2012.  
<http://www.discoveryinformaticsinitiative.org/diw2012>