

# 地方自治体における データ・サイエンスの 導入について

三重大学教育学部教授  
萩原 克幸



発行所  
三重県地方自治研究センター  
三重県津市栄町2丁目361番地  
(一助)三重県地方自治労働文化センター内  
TEL059-227-3298  
FAX059-227-3116  
<http://www.mie-jichiken.jp/>  
info@mie-jichiken.jp

## ビッグ・データ・

## IoT・人工知能

近年の情報技術・計測技術の発展により、様々な分野で、ビッグ・データの活用が進んでいる。ビッグ・データは、それまでの「データ」と比較して、質的に多様かつ量的に大規模であり、高頻度で（リアルタイムに）アクセス可能なデータとして特徴づけられている。具体的には、医療分野での画像を含む診断情報や遺伝子情報、自然科学分野での位置情報や気象情報、経済分野での販売記録情報、電子商取引情報および物流情報などがある。また、インターネット上では、SNS上での画像・動画・音声・テキストなどのマルチメディア情報などの大規模かつ多様なデータが溢れている。この他、インターネットの閲覧履歴やヘルスケアアプリに代表される個人の情報や様々な場所に設置されたビデオ・カメラ映像もまたビッグ・データである。こうしたビッグ・データの活用方法として、最近、IoT（インターネット・オブ・シングス）という考え方が現れた。それは、計測機器・センサーなどにより収集されたビッグ・データをインターネットを経由で収集し、クラウド上で分

析し、インターネット経由でユーザ（あるいは機器）にフィードバックするという、ものがインターネットによりつながるといふ考え方である。こうしたビッグ・データの収集・処理・分析に至るすべてのプロセスは、コンピュータ・ネットワーク技術、ハードウェア技術（コンピュータ、計測機器など）、ソフトウェア技術（計算・データベースなど）などの基幹技術に支えられていることを忘れてはならない。一方で、ビッグ・データと関連して、「人工知能」という言葉が再び注目を集めている。人工知能とは、人間が行っている知的情報処理を実現するソフトウェアのことであるが、一般に、高度な認識・推論を行うソフトウェアのことを指すことも多い。現在、人工知能は、画像・音声に係わる分野において盛んに開発・応用がなされている。特に、人工知能ブームの火付け役であるディープ・ニューラルネットワークは人工知能をビッグ・データから自動的に構成できる技術であり、その応用が本格的に広がっている。

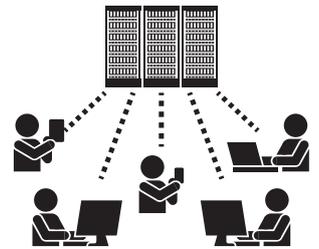
以下では、ビッグ・データ、IoT、人工知能の関係をもう少し具体的に説明するために、例として、気象予測システムを考えてみる。気象に関するデータとしては、気圧、気温、湿度、風速など観測地点の地図情報を含めた数値データだけでなく、衛星画像なども利用できるため、質的な多様性をもつ。また、莫大な数の観測地点から、時々刻々と変化するデータが計測されるため、規模の点においてもリアルタイム性においても典型的なビッグ・データである。一方で、それらのデータ

は、コンピュータ・ネットワークを経由して、データ・センターなどにストックされるとともに、高性能なコンピュータを用いて分析・予測に利用され、その結果が様々なユーザにフィードバックされる。すなわち、これは、大規模なIoTの具体例であると言える。農業・漁業をはじめとする様々な分野において気象情報は重要であることから、このIoT技術は我々の生活やビジネスに直結する。気象変動は物理法則に従うことから、現在の気象予測は、数値計算を用いた物理シミュレーションによる数値予測に基づいて行われている。そこに人工知能を導入することを考える。いま、ある時点において計測あるいは計算された気象状況（降水量・風向き・風速・雲の位置情報や衛星画像など）を入力として、その先に起きる気象状況（例えば、降水量など）を出力するシステム（プログラム）を構成することを考える。このシステムは、過去のデータとその後のデータとの間の関係性を表すため、これを使うと、現在の気象状況データをそのシステムに入力することで、未来の気象状況がそのシステムの出力として得られる。ここで、そうしたシステムをどのように構成するかが問題となる。人工知能技術では、すでに手元にある過去のデータから、ある時点とその後との時点でのデータの組をピックアップし、前者を入力として、後者を出力するように、システムを自動的に更新する。このプロセスは人間の活動に例えて「学習」と呼ばれている。すなわち、システムは、過去のデータを利用して、学習により、気象の時間変動（過去と未来の関係）を自

動的に獲得するのである。こうした気象予測システムは、人が作りこんできたシステムと異なり、過去の経験から気象状況の時間変化を自動的に学習し、未来についての推論を行うという意味で人工知能であると言える。ここで、強調したいことは、未来の気象状況を予測するためには、過去のデータからその関係性をシステムに獲得しておく必要があるという点である(すなわち、人工知能を過去のデータから構成しておく必要がある)。このため、新しくデータを観測すれば、その情報を追加して学習することで、システム(人工知能)をアップデートできるのである。

### データ・サイエンスとは

ビッグ・データの利活用の本質的な部分は統計学である。統計学の歴史は古いが、医学・工学的应用などで扱うデータの多様化に伴い複雑な手法を必要とするようになり、二〇〇〇年頃に、その理論・開発および応用を主とする「機械学習」という分野が確立された。最近のデータ・ラーニングを含むビッグ・データを扱う手法はこの分野でも研究されている。さらに、最近では、ビッグ・データの分析方法の理論と応用を扱う分野を指して「データ・サイエンス」といい、それを担う人材を「データ・サイエンティスト」というようになった。データ・サイエンティストとは、社会的背景を考えれば、統計学・機械学習の方法を的確に利用し、あるいは、新しく開発して、ビッグ・データの利



活用を實踐できる人材である。当然、その利活用の中には、IoT、さらには、人工知能の構成も含まれている。ただし、もっと一般的には、データ分析・処理およびそれに基づく予測の理論と応用を担う人材であると言える。データ・サイエンスにはコンピュータが必要不可欠であるため、データ・サイエンティストは情報技術にも長けた人材である必要がある。

大局的には、データ・サイエンスの目的は問題解決にある。すなわち、データを分析して知見を見出すだけでなく、問題解決のためにその知見を活かすことまでが重要なのである。このためには、一つのアプローチがある。一つは、トップダウン的(計画的)アプローチである。このアプローチでは、まず、問題点を明らかにし、それを解決するために必要なデータを吟味し、収集し、分析し、問題解決・改善につなげるというプロセスにより目的を達成する。ただし、このサイクルの中で、問題解決・改善の段階で十分な成果が伴わなければ、再度、それを踏まえて、データの吟味・収集から同じプロセスを繰り返すことになる。もう一つのアプローチは、ボトムアップ的(発見的)アプローチである。これは、すでに得られているデータから、統計学・機械学習の様々な方法を適用することで、データに潜む

規則性や特徴的な性質を見つけるアプローチである(これは、大量のデータに埋もれた規則性の発見を鉱物の採掘に例えた言葉である「データ・マイニング」に対応する)。そうした分析から、問題点を見つけ、解決・改善の方策を立てていくことが考えられる。このアプローチでも、分析が進むことで、新たなデータを収集・追加して、分析を深める必要があるなど、トップダウン的アプローチに移行する形をとることも考えられる。このことを逆の形で述べれば、トップダウン的アプローチでは問題ありきで考えているが、問題の発見も重要であり、そこにボトムアップ的アプローチを導入することになる。ただし、ボトムアップ的アプローチにおいて、データを与えられただけでは、有用な知見を見出すのは困難な場合が多い。いざそのアプローチにおいても、中心的な役割を果たすのはデータである。しかしながら、そのデータの背景を熟知している人間とデータの分析方法を熟知している人間は一般には異なり、データ・サイエンティストは後者にあたる。このことは、データ・サイエンスにより問題の解決を図る場合、問題をもつ分野の人間とデータ・サイエンティストの間のコミュニケーションが必要不可欠となることを意味している。例えば、遺伝子情報と病気の関連を調べる際に、データ・サイエンティストはある遺伝子と病気との間に特徴的な関係があることをデータからつきとめても、その定性的な妥当性の判断には、遺伝子や病気に詳しい専門家の知見が必要となる。あるいは、遺伝子情報は莫大であるから、それにつ

いての知識が乏しいデータ・サイエンティストが闇雲に分析を試みても専門家にとって意味のある結果が得られない可能性が高く、予め専門家の知見により絞り込まれた遺伝子情報を利用することで効果的な分析ができる場合も少なくないだろう。こうした連携においてよく見られるのは、データの提供者がデータ・サイエンティストにデータを丸投げするケースである。これは、ボトムアップ的アプローチにより何かが見つかるといふ期待によるものかもしれないが、それは一般には難しい。したがって、データの利活用においては、問題解決を図りたい分野の専門家もデータの吟味・分析への理解が必要となる。一方で、データ・サイエンティストについても、情報技術分野におけるセールス・エンジニアと同様に、問題のデータ化をサポートするためのコミュニケーション能力が必要となる。すなわち、データ・サイエンスによる問題解決においては、両者の歩み寄りが必要不可欠であると考えられる。

ビッグ・データの利活用の必要性の高まりに反して、データ・サイエンティストの数は足りておらず、その確保が国家的な戦略の上でも重要な課題となっている。内閣府による第五期科学技術基本計画では、「サイバー空間とフィジカル(現実)空間を高度に融合させたシステムにより、経済発展と社会的課題の解決を両立する、人間中心の社会」として定義される「ソサイエティ5.0(超スマート社会)」の実現が掲げられている。これは、「IoTが普及し、そこで得られるビッグ・データを、人工知能の実現を含めて、データ・

サイエンスにより活用する社会であると言える。こうした流れは世界的なものであり、国際競争を考えると、これらの新技術の進歩は日本の未来にとって重要な役割を果たすであろう。こうした国策を支えるために、文部科学省でも、小中高におけるプログラミング教育の推進、高等学校における統計教育の充実、大学におけるデータ・サイエンス・ト育成事業など、情報教育や統計教育を推進している。情報教育は現状および将来を見据えた教育として重要だけでなく、データ・サイエンスには情報技術が必要であることから、データ・サイエンスの育成には、統計教育と情報教育が必要となる。実は、最近まで、統計学を主体とする学部は日本の大学には存在しなかったが、近年、いくつかの大学でデータ・サイエンスや人工知能に関わる学部が創設され、多くの大学でそれに関わるセンター組織が立ち上げられている。一方で、統計学会においても、近年、統計学についての知識・理解および実践力の標準化と人材育成のために、統計検定が始まっている。

### 地方自治体におけるデータ・サイエンスの導入

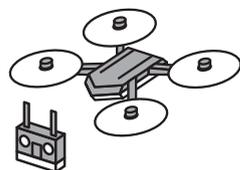
こうした中、自治体という単位でのデータ・サイエンスの導入も検討され始めている。例えば、組織における業務の効率化を図るためにデータ・サイエンスを利用することが考えられる。これは、組織のコンサルティングをデータに基づき行うということであり、地方自治体に限らず、一般企業にも当てはまる。特

に、役所関係の業務の効率化には、定型作業を人工知能やルール・ベースのソフトウェアで行うロボティクス・プロセス・オートメーションの導入が効果的であると考えられており、これからの試みは具体化していくであろう。ただし、自治体の業務には住民への様々な公的サービスがあり、こうした業務は定型作業とならない特殊なものも多い。こうした業務へのデータ・サイエンスの利用については、個別の検討が必要である。その中の一つとして、自治体が担うべき地域振興には、データ・サイエンスを活かすチャンスが多くあり、すでに、それに関係するビジネスも始まっている。以下では、まず、この方向で、三重県に適したデータ・サイエンス普及の可能性を探るとともに、それを実現していくための行政の役割を考える。まず、最初に述べておくが、データ・サイエンスを導入して、人工知能を組み込んだIoTを構築することは、様々な場面での効率化・自動化を実現できることから、特に、三重県のような人口減少の問題が深刻な地方自治体においては有益であると考えられる。三重県はよく知られているように、南北において産業の質が異なっている。北部では、電子部品・デバイス・電子回路製造業を筆頭として製造業が集積しており、三重県の製造品出荷額は国内有数である。一方で、南部では、農業・漁業・林業および観光を中心とした産業形態となっている。製造業はIoTと相性が良いため、そこに人工知能を導入して生産性を向上させる試みは容易なように思われる。しかしながら、それは、工場に導入された口

ボット・センサー・コンピュータからのデータの収集が容易であるということから生じる期待であって、実際は、どのようなデータを収集し、どのように分析して、生産性の向上につながるかは、簡単な問題ではない。このためには、上述したように、製造現場を熟知する人材とデータ・サイエンスの連携が必要となる。すなわち、人材の問題が生じる。三重県北部の規模の大きい企業については、多くの人材を抱えており、また、データ・サイエンスを雇用できる環境にあると考えられ、プロジェクト・チームを構成することで問題の発見からその解決に至るまでのプロセスが実現できる。一方で、製造業においては、中小企業の方が業務の改善・効率化を必要としているが、これに反して、データ・サイエンス導入のハードルは高いように思われる。これは、そもそもデータ・サイエンスを活用するという考えに至らないことや、その効果に懐疑的なことが考えられる。さらに、興味はあっても、どうしたらよいか分からない、人材・経費の面において具体的に話が進まないなどが考えられる。一方、南部の中心産業である農業・漁業・林業においては、三重県に限らず、高齢化と人口減少が深刻な問題であり、業務の効率化は非常に重要な課題である。こうした産業では、現状、従事者の経験により培われた知識に基づく業務が多いことから、それを自動化するために情報技術とデータ・サイエンスを導入することは当然考えられており、ビジネス化も進んでいる。例えば、農業においては、気象状況などを考慮した栽培管理やドロー

ンによる農作物の監視など、漁業においては、養殖場の管理や漁の効率化など、林業においてもドローンによる情報収集・分析などである。また、南部を中心とした観光業においては、自治体によるグローバルな支援として、観光案内の自動化への人工知能の利用などが考えられる一方で、各企業においては、観光客の動向調査（観光場所・時期やお土産の購入状況など）や意識調査などの分析に基づいて、より細やかな的確なサービスを行うなど、集客戦略にデータ・サイエンスが利用できると考えられる。しかしながら、やはり、小さい経営規模では、情報技術とデータ・サイエンスを導入するには、意識・人材・経費の面でハードルが高い。

以上のことから、データ・サイエンスを活用した三重県の行政による地域活性化においては、産業を担う企業・団体の規模に応じて異なる方策を考える必要があると思われる。製造業を担う程度規模の大きな企業では、データ・サイエンスによる問題解決を導入することへの優遇措置をとることが考えられる。この事業により成果ができれば、その経験がモデル・ケースとしてフィードバックされ、中規模の企業にも、データ・サイエンスの活用を波及させることができると考えられる。一方で、規模の小さい企業については、データ・サイエンス活用ハードルを低くすることが必要となる。このためには、まず、モデル・



ケースを多く用意し、企業へのセミナー・研修を充実させることが必要となる。これにより、企業側で、データ・サイエンス導入の効果に対する期待が生じるとともに、具体的なイメージも作られると考えられる。これと同時に、具体的なデータ・サイエンス活用については、自治体の単位で、データ・サイエンスを配置した窓口を設置して、技術相談という形で対応する方策が考えられる。ここでの重要な問題は、モデル・ケースの確立と行政組織におけるデータ・サイエンスの配置である。現時点では、データ・サイエンスは足りていないため、データ・サイエンスをサポートする企業を充てることも考えられるが、コストの問題・継続的な活性化の問題は残る。ここでは、国策にも沿う形で、その育成の仕組みを考える。その方策として、インタースhip・プロジェクトとコンペティション・プロジェクトを考えてみる。まず、育成の対象として大学生や行政職員を考え、企業の協力を得て、そうした人材に対する長期インターンシップにおいてデータ・サイエンスによる問題解決を実践する仕組みを構築することが考えられる。これは、単なる人材育成にとどまらず、受け入れ企業の意識を高めるとともに、モデル・ケースの確立にもつながる。実際、三重県南部では、「フィールドスタディ型政策協働プログラム」が実施されたこともあるが、三重県を担う人材の登用とデータを中心としたプログラムを考えるとよいだろう。ただし、実践に参加する育成対象へのデータ・サイエンス教育や指導が必要となる。これに

は、企業や大学の講師による介入が考えられる。三重県でも「ビッグ・データ分析人材養成講座」が開催されたことがあるが、単発的ではなく、こうしたインターンシップ・プログラムに組み込むことで、実質的な人材育成が可能となると考えられる。一方、自治体が企業と連携してデータを準備し、それに対してデータ分析コンペティションを開催することも考えられる。このとき、中高生を対象としたものから、大学生以上を対象としたものなどを考えることで、様々な層で、その層に応じた実践的な人材育成が可能となる。また、この企画は、データ・サイエンスの普及およびその導入の促進につながることも、自治体全体としてデータに対する意識を高めることができると考えられる。さらに、こうしたコンペティションで創出されたデータ分析のアイデアの蓄積は、モデル・ケースと同様に、後の発展において重要な役割を果たす。ただし、ここでの育成を人材確保につながる道筋についても詰めて企画する必要がある。こうした試みにより、データ・サイエンスが育成され（あるいは、その意識が育まれれば）、その人材がまたプロジェクトに関わり、指導的立場になる可能性もある。また、先に述べたように、こうした試みで生じたモデル・ケースやデータ分析法の蓄積と公開は、さらに、企業の意識改革、人材育成とデータ・サイエンスの普及にフィールドバックされ、よい循環を生み出す。なお、こうした中で、自治体にデータ・サイエンスを配置できれば、観光産業のサポートなど自治体組織の様々な業務改善にも

大きな貢献が期待できる。ただし、ここで述べたことを実現するために、ケースやデータを扱うための（部分的に）オープンな環境構築への企業・自治体の理解が必要であることを忘れてはならない。

一方で、教育機関におけるデータ・サイエンス教育は重要であり、部分的には、行政が担う役割でもありと考えられる。まず、データ・サイエンス教育にはレベルがあると考えられる。教養的な教育、専門的な教育および高度に専門的な教育である。教養的な教育とは、データ・リテラシー教育である。すなわち、データを含む情報の活用能力を養う教育であり、これには情報教育も含まれる。前述したように、企業などにおける個々がこうした教育を受けていれば、課題のデータ化やデータから課題を見つめるなど、科学的観点からの効率化・改善が可能となる。さらには、本格的なデータ・サイエンス導入において、データ・サイエンスと連携の窓口になることが可能となり、先述した人的資源の問題を解決できる。専門的な教育とは、データ・サイエンスを実践できる人材、まさに、データ・サイエンスの育成である。こうした人材は、データ・サイエンスの数理を含めて理解している必要がある。情報技術にも長けている必要がある。さらに、高度に専門的な教育とは、データ・サイエンスの新しい方法を作り出し、この分野の将来をけん引する人材である。専門的な教育は大学・高専専攻科レベル、高度に専門的な教育は大学院以上のレベルとなる。これらは理系分野において担われるべき教育であると考えられる。一方



プロフィール

三重大学教育学部教授  
はぎわら かつゆき  
**萩原 克幸**

統計科学・機械学習・信号処理の研究に従事するとともに、小学校におけるプログラミング教育など、情報教育にも尽力している。

で、教養的な教育は、高校生・大学生が教養レベルで受けるべきものと考えられる。実際、そうした人材が、広く様々な分野に就職するため、就職先の組織において、データ・リテラシーを発揮できれば社会が大きく変わると考えられる。また、文部科学省の動向からも分かるように、小学校からの情報教育・統計教育は、現代社会における実質的な教育としても重要であるが、すべてのレベルの人材育成において裾野を広げる役割を担うという意味でも重要である。しかしながら、小中高の教育現場、特に、小中の現場では、情報教育・統計教育を実践できる人材は非常に少ない。こうした人材の採用・育成は行政の役割であると考えられる。また、上述したが、高等教育において、データ・サイエンスを主とした地方企業・組織でのインターンシップ・プロジェクトを実施すれば、それは、データ・サイエンスの育成になり、自治体での人材確保にもつながる。自治体は、このためのコディネーターを担うことを考えてもよいのではないだろうか。