

# ビッグデータのデータサイエンス ～ニューノーマル時代のビッグデータ～

## 編集にあたって

里 洋平 | (株) Village AI / nat (株) / (株) Lupinus 石井一夫 | 公立諏訪東京理科大学

2020年春から、またたく間に世界中に広がった新型コロナウイルスの蔓延は、我々の生活を大きく変化させた。リモートワークや在宅勤務による新しい勤務形態も拡がり、DX推進が加速化している。本誌で前回ビッグデータ特集号（「ビッグデータ、IoT、AI：最新の事例と人材育成」）を企画したのは、2020年7月であり、ちょうどコロナ禍に突入し、今後の世界の変化と混乱を予感させるときであり、先行きが見えないときであった。以後、新型コロナウイルスの蔓延は、一進一退を繰り返し、依然先が見えない状況が続いている。同時に、世界中で熱波や豪雨など、気候変動による地球温暖化の顕在化が加速化してきており、その危機的状況はもはや見過ごすことのできないほどになっている。国内では、少子高齢化が急速に進み医療や職場を含む社会システムの崩壊が危惧されるようになってきている。

このような社会課題に対し、解決策を模索し提供する手段として、ビッグデータの重要性への認識が高まっている。いわゆるビジネス活動の推進や効率化のためのビッグデータ利活用という観点から、DXの推進を基盤にした社会課題解決の手段としてのビッグデータ利活用という形に、そのありようが少しずつ変化してきている。この中で、ニューノ-

マル時代のビッグデータを分析するデータサイエンスの在り方を探るという趣旨で、今回1年半ぶりに、「ビッグデータのデータサイエンス～ニューノーマル時代のビッグデータ～」と題して企画するに至った。

ビッグデータとは何かを、ここで議論するのは野暮であるが、最近再認識し、強調したいことは、ビッグデータ・アナリティクスの強み、あるいは特徴は「悉皆調査」を基本にしているということである。これにより、従来の推計統計学では難しかった社会課題の解決に向けた個別化対応とグローバルな未来予測が可能になってくる。新型コロナウイルスの蔓延の分析や、地球温暖化の影響調査などは、「悉皆調査」であるからこそ、その真髄に迫ることができる。本特集号で、社会課題の解決に向けたビッグデータのデータサイエンスについて考えつつ、我々がどこに向かおうとしているのか、何を成さなければならぬのか、ヒントとなるようなものを提供できれば幸甚である。

本特集号は、5つの招待論文と座談会で構成した。最初の招待論文は、「Apache ArrowによるRubyのデータ処理対応の可能性」と題し、プログラミング言語Rubyにおけるデータ処理環境の構築である。ご存知のとおり、データ分析で使用されるプログラ

## 【デジタルプラクティスコーナー】

各記事の概要のみ掲載しております。本文は電子版

<https://www.ipsj.or.jp/dp/contents/publication/49/S1301-index.html> を  
ご覧ください。



ミング言語は、フリーソフトウェアでは、Python や R が主流で、新しいものとしては Julia が注目されている。Ruby は、Python や R と同様に学びやすく、書きやすい言語で、Ruby on Rails の基盤があり Web アプリケーションとしては非常にポピュラーであるが、データサイエンスのためのプログラミング言語としては対応が遅れている。しかし、本来、Python や R と同じような使い方ができるので、環境さえ整えば非常に優れた言語になる可能性がある。本稿では、Ruby で本稿の Apache Arrow をはじめとする Ruby のデータサイエンス言語環境の整備に尽力している村田賢太（(株) Speee）らにその開発の現状を執筆いただいた。

2 番目の招待論文は、医療ビッグデータに関するもので、「大阪府の特定健康診査データの因果探索」というタイトルで大山飛鳥（大阪大学キャンパスライフ健康支援・相談センター）らによる、国保連合会が管理する国民健康保険のデータベース（KDB データベース）のデータ分析に関する報告である。レセプトデータ、健診データのデータベースは、巨大で専門性が強く、その処理には特別な配慮と技術が必要であるが、その処理に関するノウハウや知見が語られている。特に、線形回帰モデルの性質を利用した因果推論探索は、医療ビッグデータのデータ分析を行う上で参考になると思われる。

3 番目の招待論文は、マーケティングにおけるビッグデータ処理に関するもので、「Account-Based Marketing のためのターゲット企業推薦モデルの改善」というタイトルで新井和弥（(株) ユーザベース）らによるものである。本稿では、ターゲット企業推薦モデルにおいて、L2 正則化項付きのロジスティック回帰モデルを、ナイーブベイズ拡張モデルなどほかの方法と比較した上で、提案手法として選択した

プロセスを示している。機械学習全盛の時代に、あえて古典的手法であるロジスティック回帰モデルを選択したプロセスや考察は、ほかのビッグデータを用いた推薦モデルの検討にも役立てることができるであろう。

4 番目の招待論文は、文系大学におけるデータサイエンスの数理リテラシー教育の現状を紹介するので、「人文・社会科学系大学におけるデータサイエンス教育」というタイトルで増川純一（成城大学）らに執筆いただいた。特に、数理系科目に苦手意識を持つ文系学生に対するビッグデータを意識したデータサイエンス教育の整備状況について紹介している。初学者に向けた数理リテラシー教育、統計学を中心としたデータ分析に関する科目、自然言語処理や画像認識を中心とした機械学習と、その範囲を広げて教育環境の整備が進んでいる様子が伺える。

5 番目の招待論文は、農業におけるリモートセンシングによる画像解析に関するもので、「ドローンによる作物の表現型計測と機械学習による作物バイオマス・収量の予測」というタイトルで辰己賢一（東京農工大学）氏に執筆いただいた。いわゆる、農作物の背丈などの直接の表現型計測データに加え、ドローンによる画像データから得られる表現型計測データを元に農作物の重量（バイオマスともいわれる）や収穫量を予測する機械学習モデルを作成しようとするものである。今後のスマート農業など、環境計測によるスマート農業への応用推進が期待される。

最後に、「ビッグデータのデータサイエンス～ニューノーマル時代のビッグデータ～」と題し、本会ビッグデータ解析のビジネス実務利活用（PBD）研究グループ（略称：ビッグデータ研究グループ）の運営委員メンバによる座談会を企画した。本座談会では、ビッグデータに関するデータサイエ

## [特集:ビッグデータのデータサイエンス～ニューノーマル時代のビッグデータ～] 編集にあたって

ンスについて最近のトピックを語っていただいた。トピックとして、データと法律、データの流通、データの質、人材育成に関する議論がなされた。

今回の「ビッグデータのデータサイエンス」特集では、データ分析そのものを中心に意識しており、その実務面での活用促進の一面を垣間見れる。本特集が、今後の読者諸氏のニューノーマル時代のビッグデータ利活用推進のためのヒントになれば幸甚である。

(2021年11月1日)

### ■里 洋平 (正会員) y.sato@villageai.jp

R言語の東京コミュニティTokyo.R創業者。ヤフー(株)で、推薦ロジックや株価の予測モデル構築など分析業務を経て、(株)ディー・エヌ・エーで大規模データマイニングやマーケティング分析業務に従事。その後、(株)ドリコムにて、データ分析環境の構築やソーシャルゲーム、メディア、広告のデータ分析業務を経て、DATUMSTUDIO(株)を設立。2021年7月に退任し現在は、(株)Village AI代表取締役、nat(株)取締役、(株)Lupinus社外取締役。本会ビッグデータ解析のビジネス実務利活用研究グループ幹事を兼任。

### ■石井一夫 (正会員) kazuoshii2014@gmail.com

公立諏訪東京理科大学工学部情報応用工学科教授、久留米大学医学部内科学講座心臓・血管内科講座客員准教授。少子高齢化および地球温暖化問題の克服に向けた医療ビッグデータ、環境・農業ビッグデータの教育研究に従事。本会ビッグデータ解析のビジネス実務利活用研究グループ主査。

論文誌 デジタルプラクティス「特集:ビッグデータのデータサイエンス～ニューノーマル時代のビッグデータ～」はこちらでご覧いただけます(電子図書館)  
[https://ipsj.ixsq.nii.ac.jp/ej/?action=repository\\_opensearch&index\\_id=10669](https://ipsj.ixsq.nii.ac.jp/ej/?action=repository_opensearch&index_id=10669)



## [特集:ビッグデータのデータサイエンス～ニューノーマル時代のビッグデータ～] 概要

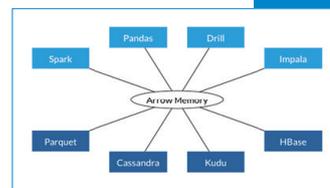
### 1 Apache ArrowによるRubyのデータ処理対応の可能性

村田賢太 ((株) Speee / Red Data Tools) ・ 須藤功平 ((株) クリアコード / Red Data Tools)

RubyはWeb業界で広く浸透しているが、データ処理分野ではほとんど利用されていない。Rubyの分析的データ処理への対応が弱いことが原因である。

この弱点は、Rubyで書かれたシステムを分析的データ処理に対応させる際の足枷となるだけでなく、昨今のDX対応推進の流れに乗り遅れる要因にもなり得るだろう。

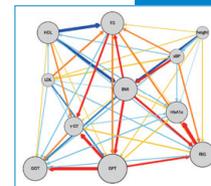
本稿では、RubyのApache Arrowへの対応がこの弱点の解消に有効であることを示す。



### 2 大阪府の特定健康診査データの因果探索

大山飛鳥 (大阪大学キャンパスライフ健康支援・相談センター) ・ 古徳純一 (大阪大学キャンパスライフ健康支援・相談センター / 帝京大学大学院医療技術学研究所) ・ 土岐 博 (大阪大学キャンパスライフ健康支援・相談センター)

我々は、大阪府民60万人の大規模特定健診データの提供を受け、健診データに対して因果探索の数理モデルを用いた因果ダイアグラムの構築を行い、主に理論的な側面について論文にまとめた。本稿では、そこに含めることができなかった健診データ解析の実際について、実験を交えながら赤裸々に語る。プログラムの実装方法や、実際に大規模健診データを取り扱う際の注意点や対処法についても紹介する。





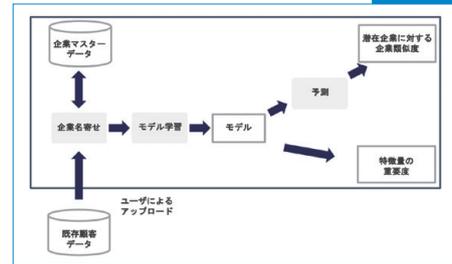
### 3 Account-Based Marketing のための ターゲット企業推薦モデルの改善



新井和弥 ((株) ユーザベース)・北内 啓 ((株) ニュースピックス)・  
高柳慎一・早川敦士・林 樹永・長田怜士 ((株) ユーザベース)

B2B (Business To Business) 領域における企業情報活用が著しい飛躍を遂げており、企業情報を用いた新たな B2B マーケティング手法として ABM (Account-Based Marketing) の活用が広がっている。

本稿においては、ABM をソフトウェアによって実践する 1 つの方法、および筆者らによる実装と現状の課題について紹介し、その課題を解決するための研究結果を報告する。



### 4 人文・社会科学系大学におけるデータサイエンス教育



増川純一 (成城大学)・辻 智 (成城大学データサイエンス教育研究センター)・  
田村光太郎 ((株) 野村総合研究所データサイエンスラボ)

人文・社会科学系の 4 学部からなる成城大学においても、数理科学のリテラシーを持ち、データ分析のできる人材の育成は教育目標の大きな柱の 1 つである。本学は 2015 年度に全学共通教育科目としてデータサイエンス科目群を設置しその教育課題に取り組んできた。本稿では 6 年間実施してきたデータサイエンス教育プログラムの成果と課題を整理したい。また、プログラムの次のステージに向けて今後の展開を述べる。



### 5 ドローンによる作物の表現型計測と機械学習による作物バイオマス・収量の予測



辰己賢一 (東京農工大学)

ドローンによる空撮画像から導出したトマト株の草高、植生指数から収量に影響の大きい説明変数を算出および選択し、複数の機械学習モデルを用いて収量予測精度の検証を実施した。その結果、単純統計量だけでなく GLCM によるテクスチャ情報を変数として考慮することで予測精度が大きく向上し、また収穫の約 1 カ月前のテクスチャ情報が果実重や果実数を予測する上で重要度の大きい説明変数であることが明らかとなった。



[特集：ビッグデータのデータサイエンス～ニューノーマル時代のビッグデータ～] 概要

● 座談会／ビッグデータのデータサイエンス  
～ニューノーマル時代のビッグデータ～



進行役：里 洋平 (株) Village AI / nat (株) / (株) Lupinus

インタビュイー：高柳慎一 ((株) ユーザベース)・安部晃生 ((株) コネクトデータ)・  
飯尾 淳 (中央大学)・牧山幸史 ((株) ヤフー)

インタビュアー：石井一夫 (公立諏訪東京理科大学)

本特集は、「ビッグデータのデータサイエンス」というタイトルで、ビッグデータを対象としたデータサイエンスについて、特に、コロナ禍や気候変動時代におけるビッグデータのデータサイエンスの在り方を意識しながら企画した。それを受けて、今回の座談会では、本会ビッグデータ解析のビジネス実務利活用 (PBD) 研究グループ (略称：ビッグデータ研究グループ) の運営委員メンバにより、「ニューノーマルにおけるデータサイエンス」と題して、最新の関連トピックについてお話しいただいた。本企画が、日々、目まぐるしく社会状況が変化していく中での、データサイエンスの今、これから、について、日々の業務のヒントになれば幸いである。

会誌「デジタルプラクティスコーナー」が始まりました

論文誌デジタルプラクティスは、2020年10月に生まれ変わりました。

この度、論文誌デジタルプラクティスを改め、新設する論文誌トランザクション デジタルプラクティスと、会誌デジタルプラクティスコーナー、既存のDPレポートを通じて、質の高い論文、速報性の高い論文をより分かりやすく皆様にお届けして参ります。

会誌「デジタルプラクティスコーナー」は概要を本誌に掲載し、論文本体は電子版として公開いたします。

会誌デジタルプラクティスコーナー (電子版) の購読は無料ですのでみなさまぜひ御覧ください。



	2020年7月刊行分まで	2020年10月刊行分以降
論文誌デジタルプラクティス	特集号投稿論文、一般投稿論文、推薦論文 [採録審査あり]	論文誌トランザクション デジタルプラクティス [採録審査あり] (電子版) <a href="https://www.ipsj.or.jp/dp/">https://www.ipsj.or.jp/dp/</a>
	特集号招待論文 (共同編集あり)	会誌デジタルプラクティスコーナー (共同編集なし)、 概要を会誌紙媒体に掲載し、論文本体は電子版として公開
	JISA 招待論文 その他招待論文	
DPレポート [採録審査なし]		DPレポート [採録審査なし] (電子版) <a href="https://www.ipsj.or.jp/dp/DPreport/index.html">https://www.ipsj.or.jp/dp/DPreport/index.html</a>