2022年1月10日印刷　2022年1月15日発行（毎月1回15日発行）
ISSN 0447-8053

IPSJ MAGAZINE

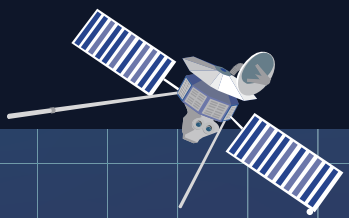
情報処理

2022
2
Vol.63 No.2
通巻 683 号

特集 オンライン スマートファクトリーは工場の何を変えるのか？

特別解説　国家公務員採用総合職試験における「デジタル区分」の新設について
―試験の概要と「デジタル区分」の試験問題例―

巻頭コラム
忘れやすい身体
菊川裕也



オンライン デジタルプラクティスコーナー：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～
教育コーナー：ペた語義
連　載：IT紀行／5分で分かる!? 有名論文ナナメ読み／ オンライン <Info-WorkPlace委員会企画>働き方を共有しよう!／
オンライン 教科「情報」の入試試験問題って?／情報の授業をしよう！／先生，質問です!／
ビブリオ・トーク
オンライン 会議レポート
委員会から

電子版もご覧ください

電子版を読む（会員無料）	iPhoneなどで読む（有料）	電子版を購入（有料）	Web公開（無料/有料）
情報学広場	Kindle	Fujisan	note

一般社団法人 情報処理学会
Information Processing Society of Japan

F8VPS FORUM8 VIRTUAL PLATFORM SYSTEM バーチャルプラットフォームシステム


Web会議機能


VRモード

Webプラットフォーム　3Dリアルタイム VR

F8VPSは、あらゆる空間のバーチャルシステムを構築！

基本ライセンス ¥550,000（税込）～

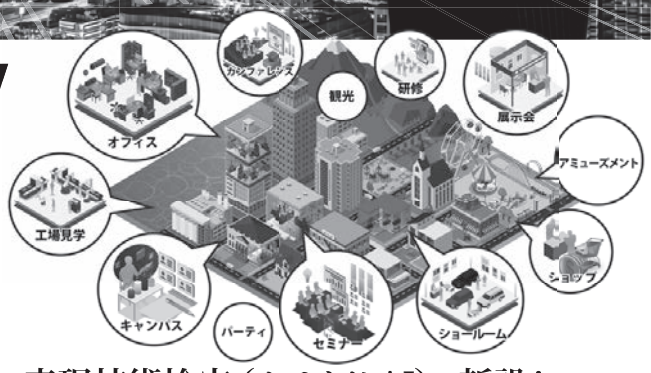
業界最先端の技術によって、御社のオープンプラットフォーム化を強力に推進。最小限のコストで、クラウド上での開発・展開から、テレワーク、商品PR・広報まで、DX時代に必須のバーチャルプラットフォームシステムを構築します。

F8VPS活用事例　スマートシティを実現するデジタルツイン

バーチャルツアーで敷地内を見学できるほか、実験施設の紹介ムービーの再生なども可能


フォーラムエイト東京本社ショールームを3DVRで再現


バーチャルキャンパス事例
（東京工業大学 Tokyo Tech ANNEX）



表現技術検定（クラウド-AI）　新設！

AI・クラウドの基本・活用事例を通し、実務に活かせる最新知識を習得

日時　2021年3月8日（火）9:30～16:30

場所　本会場：フォーラムエイト　東京本社　セミナールーム
大阪・名古屋・福岡・仙台・札幌・金沢・岩手・宮崎・沖縄＋オンライン

受講料　12,000円（検定証発行手数料込み、税込）

フォーラムエイトの出版書籍

書籍のお求めは

表現技術検定
公式ガイドブック
情報処理 / データベース

著者　石河 和喜
出版 フォーラムエイトパブリッシング
価格 3,080円（税込）

DX時代のビジネスの必須知識を基礎からわ学べる教習本シリーズ。「情報処理編」では確率・統計に加えてプレゼン表現やAI技術までを扱う。表技協「表現技術検定」受験者向け公式テキスト。

2022 FIA World Rally Championship Round 13
FORUM8 RALLY JAPAN 2022 AICHI/GIFU 11.10 THU - 13 SUN

フォーラムエイトは、FIA世界ラリー選手権 FORUM8 Rally Japan2022をタイトルパートナーとして応援しています

株式会社 フォーラムエイト 東京本社
東京都港区港南 2-15-1 品川インターシティ A 棟 21F
Tel （代表）03-6894-1888　（営業窓口）0120-1888-58
Fax 03-6894-3888 | E-mail f8tokyo@forum8.co.jp
◆ショールーム：東京・大阪・名古屋　◆セミナールーム：東京・大阪・名古屋・福岡・仙台・札幌・金沢・岩手・宮崎・沖縄／上海・青島・台北・ハノイ
FORUM8 フォーラムエイト®
www.forum8.co.jp

情報処理学会編集の教科書シリーズ！

IT Text（一般教育シリーズ）一般情報教育

情報処理学会一般情報教育委員会 編／稲垣知宏・上繁義史・北上 始・佐々木 整・髙橋尚子・中鉢直宏・徳野淳子・中西通雄・堀江郁美・水野一徳・山際 基・山下和之・湯瀬裕昭・和田 勉・渡邉真也 共著
A5判／266頁／定価2,420円（税込）ISBN 978-4-274-22595-6

情報処理学会一般情報教育委員会で編纂した、これからの一般情報教育に対応した標準テキストです。情報ネットワークや情報機器の基礎知識から、プログラミングの考え方、情報倫理、データサイエンス等、社会生活で不可欠な教養ともいえる知識を幅広く網羅します。

ソフトウェアエンジニアリング・スタンダードの第9版！

実践ソフトウェアエンジニアリング 第9版

Roger S. Pressman・Bruce R. Maxim 共著／SEPA翻訳プロジェクト 訳
B5判／548頁／定価8,800円（税込）ISBN 978-4-274-22794-3

本書は米国においての第1版の発行（1982年）以来、世界累積300万部を超えるベストセラーの最新刊である第9版の邦訳書です。ソフトウェア同様、改良が続けられているソフトウェアエンジニアリングの「最良の手法」を解説している書籍であり、現役のソフトウェアエンジニアならびに学生諸氏におすすめする1冊です。

機械学習のしくみをイラストや図解でやさしく学ぼう！

機械学習をめぐる冒険

小高知宏 著
四六判／184頁／定価2,420円（税込）ISBN 978-4-274-22761-5

機械学習に関するさまざまなトピックスを概説。人工知能における機械学習の位置づけを説明し、機械学習内の分野をマップ化し、マップ内の街（＝機械学習内の分野）を旅する形でやさしく解説していきます。「どんなしくみで、どこで使われていて、どう役に立つのか」という要点をわかりやすく示しています。

あの「天然知能」を情報科学として明快に解説！

セルオートマトンによる知能シミュレーション 天然知能を実装する

浦上大輔・郡司ペギオ幸夫 共著
A5判／256頁／定価3,740円（税込）ISBN 978-4-274-22782-0

オートマトンの基礎から解説を始め、セルオートマトンに見られる典型的な現象（相転移、カオスの縁）、セルオートマトンと人工知能との対応、非同期調整セルオートマトンと著者らの提唱する「天然知能」との対応、リザバーコンピューティングによる実装の手法までを、順を追って解説します。

データ活用社会に必須のデータ分析・解析の基本的な考え方と手法をわかりやすく解説！

データサイエンスの考え方 社会に役立つAI×データ活用のために

小澤誠一・齋藤政彦 共編
A5判／320頁／定価2,750円（税込）ISBN 978-4-274-22797-4

本書は、政府の「AI戦略2019」での議論を経て策定・公表された「数理・データサイエンス・AI（応用基礎レベル）モデルカリキュラム」に準拠した内容です。具体的な事例と分析手法を扱いながら、社会のさまざまな場面で必要とされるデータサイエンスの考え方を、関連する数学とともに丁寧に解説します。

オーム社 Ohmsha | 〒101-8460 東京都千代田区神田錦町3-1 www.ohmsha.co.jp
TEL 03(3233)0853 FAX 03(3233)3440 定価は変更になる場合があります。

【ご案内】会誌「情報処理」のオンライン記事について

会誌「情報処理」の特集記事は，これまで冊子，オンライン（電子図書館）の両方に掲載しておりましたが，次のとおり**オンラインのみへの掲載**に変わりました．また，オンライン限定記事の掲載も始まりました．

◆**開始月**：**2020 年 11 月号**（発行日：2020 年 10 月 15 日）
◆**閲覧方法**：会員区分によって異なりますので以下をご確認ください．

【個人会員の皆様】

電子図書館（情報学広場：https://ipsj.ixsq.nii.ac.jp/ej/）にログインし，該当記事の pdf をダウンロードしてください．
すでに電子図書館をご利用いただいている方は今までどおりです．
電子図書館を初めて利用される方は，会員としてのユーザ登録が必要になります．
未登録の方には毎月上旬に次の件名のメールを送信しておりますので，到着次第，登録してください．

- 件名：[情報学広場 : 情報処理学会電子図書館] ユーザー登録のご案内
- 差出：ipsj-ixsq@nii.ac.jp

【個人会員】
電子図書館
（情報学広場）

★詳細：電子図書館利用方法（個人用）－利用までの流れ（https://www.ipsj.or.jp/e-library/ixsq.html#anc2）
ご案内メールをお急ぎの方や閲覧方法が分からない方は，会員サービス部門（E-mail: mem@ipsj.or.jp）に会員番号を添えてご連絡ください．

【賛助会員各位・購読員の皆様】

賛助会員・購読員の企業・大学に所属されている方に「情報処理」（冊子）を貸し出しした場合，特集の閲覧方法について照会がございましたら，次の手順をお知らせください．

＜手順＞

(1)「情報処理」の特集ページ（扉または概要ページ）を開く．
(2) 閲覧申込の URL にアクセスする（または QR コードを読み取る）．
(3) 必須事項を入力し送信する．
(4) 次の件名（2 月号の場合）の受信メールに従って，電子図書館から特集の pdf をダウンロードする．

- 件名：情報処理 2022 年 2 月号（Vol.63, No.2）「チケットコード」とご利用方法のご連絡

★注意事項

- 法人アカウントではご利用いただけません．
- 閲覧される方が電子図書館のユーザ ID をお持ちでない場合は，ご自身でユーザ登録する必要があります．

本件に関する問合せ先：一般社団法人情報処理学会　会員サービス部門　E-mail: mem@ipsj.or.jp

〔前付最終〕

〔目次前〕

2

PREFACE

巻頭コラム

46 忘れやすい身体 菊川裕也

お知らせ

特集記事はオンラインのみの掲載となります（本誌には「編集にあたって」「概要」のみ掲載されます）．オンライン記事（電子図書館）の閲覧方法につきましては本誌前付最終に掲載しておりますのでご確認くださいますようお願いいたします．

SPECIAL ARTICLE

特別解説

48 ■国家公務員採用総合職試験における「デジタル区分」の新設について
—試験の概要と「デジタル区分」の試験問題例— 佐藤 壮

SPECIAL FEATURES

特 集

スマートファクトリーは工場の何を変えるのか？

54 編集にあたって 袖美樹子・田中功一
56 概要

DIGITAL PRACTICE

デジタルプラクティスコーナー

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～

58 編集にあたって 里 洋平・石井一夫
60 概要

連載：情報の授業をしよう！

64 ■高等学校（工業）でのスマートフォンを利用したデータ活用の授業 岸本有生

70 連載：Jr. 先生，質問です！

教育コーナー：ぺた語義

71 ■Jr. オンライン授業を快適に受講するには？ 越智 徹
72 ■シンポジウム「大学入学共通テスト『情報』が目指すもの」 稲葉利江子
77 ■大学入学共通テストにおける教科「情報」の導入を受けて 河原達也
79 ■国立大学入学者選抜制度への「情報」の追加について 中山泰一

連載：Jr. ビブリオ・トーク—書評—

84 データサイエンス入門 教養としてのデータサイエンス 石井一夫

連載：Jr. 5分で分かる!? 有名論文ナナメ読み

86 Lars Ole Andersen : Program Analysis and Specialization for the C Programming Language 内田公太

委員会から

88 今年度もやります!全国大会の“デリバリー” 坊農真弓

連載：IT紀行

90 VR作品の登竜門IVRCに行ってみた! 山本ゆうか

《記号の説明》

基 応 専 般 Jr. ■基礎 ■専門家向け ■応用 ■一般（非専門家）向け Jr. ジュニア会員向け

※各記事に指標がついていますのでご参考になさってください

IPSJ MAGAZINE
情報処理学会誌
Vol.63 No.2（2022 年 2 月号）
通巻 683 号

情報処理

常時更新中!

「情報処理」オンライン版　目次
https://www.ipsj.or.jp/magazine/contents_m_e.html
※オンラインでのみ掲載している記事の目次を掲載しております（目次から電子図書館の各記事へリンクしております）.

■ Vol.63 No.2
特集：スマートファクトリーは工場の何を変えるのか?
e1　1. 工場のスマート化を実現する最新の FA 技術と取り組み（楠　和浩）
e7　2. リアルタイム AI 技術の製造業への適用（櫻井保志）
e13　3. IoT プラットフォームの現状と未来―製造 DX の本質―（鈴木　聡）
e19　4. スマートファクトリーを支えるローカル 5G―導入に向けた制度や技術，留意点の理解―（柿元宏晃）
e26　5. 持続可能な社会に向けた今後の生産システムと産業基盤―「人」が主役となるものづくり革新推進（HCMI）コンソーシアム 2050 年に向けたロードマップから―（岩井匡代・谷川民生）

デジタルプラクティスコーナー：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～
1. ［招待論文］Apache Arrow による Ruby のデータ処理対応の可能性（村田賢太・須藤功平）
2. ［招待論文］大阪府の特定健康診査データの因果探索（大山飛鳥・古徳純一・土岐　博）
3. ［招待論文］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・北内　啓・高柳慎一・早川敦士・林　樹永・長田怜士）
4. ［招待論文］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・辻　智・田村光太郎）
5. ［招待論文］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）
6.「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～」座談会（進行役：里　洋平　インタビュイー：高柳慎一・安部晃生・飯尾　淳・牧山幸史　インタビュアー：石井一夫）
グロッサリ

連載：教科「情報」の入学試験問題って?
e33　「データの分析」分野の入試問題の分類と解法の一考察　入試センターのサンプル問題解説～第 3 問データの分析～（阿部百合）

連載：<Info-WorkPlace 委員会企画 > 働き方を共有しよう!
e50　CASE4：私のオフィスはオンラインに引っ越しました（伊東　香）

会議レポート
e57　FIT2021 イベント企画「ヒトゲノム・生体情報と情報処理の課題」会議報告（金子　格）

「情報処理」note
https://note.com/ipsj
※人気記事や最新記事のチラ見せ，無料で読める記事などさまざまなコンテンツを公開していきます.

目次前 【ご案内】会誌「情報処理」のオンライン記事について
63　2022 年度会誌「情報処理」モニタ募集のお知らせ
81　［重要］過去のプログラミング・シンポジウム報告集の利用許諾について
82　論文誌ジャーナル掲載論文リスト／論文誌トランザクション掲載論文リスト／デジタルプラクティス論文リスト／ IPSJ カレンダー
94　会員の広場
97　英文目次／アンケート
98　人材募集
99　会告
102　編集室／次号予定目次
103　掲載広告カタログ・資料請求用紙
104　賛助会員のご紹介

■情報処理学会事務局本部
〒 101-0062 東京都千代田区神田駿河台 1-5　化学会館 4F
Tel(03)3518-8374（代表）　Fax(03)3518-8375
E-mail: soumu@ipsj.or.jp　https://www.ipsj.or.jp/
郵便振替口座　00150-4-83484
銀行振込（いずれも普通預金口座）
みずほ銀行虎ノ門支店　1013945
三菱 UFJ 銀行本店　7636858
名義人：一般社団法人　情報処理学会
名義人カナ：シヤ）ジヨウホウシヨリガツカイ
■規格部　情報規格調査会
〒 105-0011 東京都港区芝公園 3-5-8　機械振興会館 308-3
Tel(03)3431-2808　Fax(03)3431-6493
E-mail: standards@itscj.ipsj.or.jp　https://www.itscj.ipsj.or.jp/
■支　部　北海道／東北／東海／北陸／関西／中国／四国／九州

電子版
-DIGITAL VER-
Kindle
Fujisan
情報学広場

忘れやすい身体

■ 菊川 裕也

最近物忘れが激しい．

何度も会ったことのある人でも名前が出てこなかったりする．久しぶりにイベントに出かけたら，菊川さん，と親しげに話しかけられた．何となく会ったことがあるような気もするけど，いつどこで会った誰なのかまったく思い出せない．非常に申し訳なく感じる．早くARグラスが発売されてその人の名前や履歴がオーバーレイで表示されるようになってほしい．

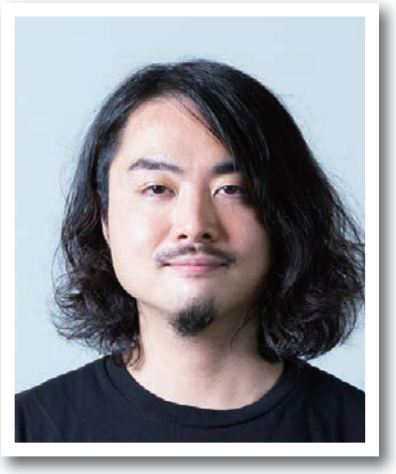
近頃はオンラインでしか会ったことのない相手と仕事をすることも多い．この間とあるプロジェクトの仲間と初めてオフラインで会って，画面上の顔と喋り方で勝手に小柄な人と思い込んでいたのだが，会ってみると思いのほか高身長で驚いた．

オンライン会議が一般化する中でバーチャル背景を使うことは当たり前になった．その一方で，アバターに変身したりフィルタで顔を変化させることに関しては現時点では日本の会議の場には適さないということになっている気がする．画像処理という観点ではほぼ同じ技術でも，人間社会に与える効果はまったく違うのが興味深い．

2021年11月の現在では猫も杓子もメタバースの話題でもちきりだ．今後はどうやらメタバース上で遊んだり仕事をすることになっていくようだ．メタバース上で仕事の会議をするとき，アバターは自分に似せていないといけないのだろうか？　動物とかの姿で出席すると失礼に当たるだろうか？

■ 菊川 裕也
（株）ORPHE 代表取締役

一橋大学商学部を卒業後，首都大学東京大学院にて芸術工学を専攻．2014 年よりスマートフットウェア「Orphe」の製品化をきっかけに（株）no new folk studio（現（株）ORPHE）を設立．



そもそもずっと同じ姿でいるのだろうか？　プラットフォームによって変える気もする．Aの仮想空間上では猫，Bの仮想空間上では寿司なクライアントとオフラインで会ったら意外と高身長だったとき，自分はちゃんと会釈できるだろうか？（現実で会った人すら覚えられないのに……）．

仕事上の認知を得るためにはできるだけ統一の姿の方がいいかもしれない．考えてみると人の顔はほとんどの場合唯一無二で，誰でも身分証明として使うことのできる便利な存在だ．仮想空間であれば自分の身体性から解き放たれて自由な姿になれるが，一方で肉体が自動的に行う唯一性の証明を失うことになる．遊びであればどんなアバターが相手でもいいが，たとえば不動産の購入のような信頼が重要な場面であれば，本当の顔を知らない相手から買うのは怖い気がする．そうなると結局メタバース上でも元の顔からは逃れられないのだろうか？　もしくはこの感覚が古いだけで，ほかの方法で信頼を獲得していくようになるのだろうか？

センサ内蔵の靴で人の歩行をセンシングすることを仕事にしている．この数年，スポーツ科学への応用を行ってきたが，今後は医療へも応用していくつもりだ．歩くことと健康は相関が強いし，遠隔医療の必要性も高まる昨今，やりがいは感じる．

メタバースが発展しても，人は歩き続けるだろう．歩くことは楽しいし，仮にアバターが自由気ままに取り替え可能だったとしても，自分の身体は1つで取り替えがきかないのだから（今のところ）．

一般

国家公務員採用総合職試験における「デジタル区分」の新設について
―試験の概要と「デジタル区分」の試験問題例―

佐藤　壮 | 人事院人材局企画課制度班

試験区分の新設・見直しの趣旨

2020 年 12 月，政府は今後のデジタル社会への対応やその司令塔となるデジタル庁の設置を示した「デジタル社会の実現に向けた改革の基本方針」☆1 を閣議決定した．その中で，デジタル庁を含む政府部門においてデジタル政策の中心となるような人材を確保する観点から，人事院に対して国家公務員採用総合職試験（以下「総合職試験」）に新たな区分（デジタル）を設けることが要請された．これを受けて，2021 年 4 月，情報系の専門的な素養を持つ有為の人材をこれまで以上に確保するため，2022 年度より，総合職試験に「デジタル区分」を新設するとともに，国家公務員採用一般職試験（以下「一般職試験」）の「電気・電子・情報区分」について，試験内容の見直しを行った上で，区分の名称を「デジタル・電気・電子区分」とすることを人事院は発表した．「デジタル区分」からの採用者には，情報系の知識を持って，各府省の政策の企画および立案または調査および研究に従事することが期待される．

今回は，この場を借りて，総合職試験「デジタル区分」と一般職試験「デジタル・電気・電子区分」の概要と，「デジタル区分」についてはその試験問題例を紹介する．

なお，国家公務員採用試験のより詳細な情報については，人事院の国家公務員試験採用情報 NAVI☆2 を参照されたい．

総合職試験「デジタル区分」の概要

総合職試験（院卒者試験・大卒程度試験）は，政策の企画および立案または調査および研究に関する事務をその職務とする係員を採用するための試験である．2022 年度の総合職試験では，その専門性に応じた表 -1 の区分の試験を実施する．

表 -1 におけるいずれの区分においても，第 1 次試験では，すべての受験者が共通して解答する基礎

☆1　デジタル社会の実現に向けた改革の基本方針，
https://www.kantei.go.jp/jp/singi/it2/dgov/dai10/gijisidai.html

☆2　国家公務員試験採用情報 NAVI，
http://www.jinji.go.jp/saiyo/syokai/digital_gaiyou.html

表 -1　総合職試験（春試験）の試験の区分

院卒	行政，人間科学，**デジタル**，工学，数理科学・物理・地球科学，化学・生物・薬学，農業科学・水産，農業農村工学，森林・自然環境
大卒程度	政治・国際，法律，経済，人間科学，**デジタル**，工学，数理科学・物理・地球科学，化学・生物・薬学，農業科学・水産，農業農村工学，森林・自然環境

＊このほかに，例年秋に実施する「法務区分」（院卒者試験），「教養区分」（大卒程度試験）がある．

能力試験（多肢選択式）と専攻分野に応じて必要な専門的知識が問われる専門試験（多肢選択式）が行われ，第 2 次試験では，専門試験（記述式）や人物試験等が行われる（表 -2）．

これらの試験に合格後，自身の志望する官庁を訪問し，各府省において採用されるかどうかが決定される．

2022 年度の総合職試験から，新たに設ける「デジタル区分」の専門試験（多肢選択式）および専門試験（記述式）では，情報系の試験種目の問題選択の柔軟性を高め，受験者の専門性に合わせて受験しやすくした．2021 年度まで，情報工学（ハードウェア）および情報工学（ソフトウェア）の問題は「工学区分」において選択問題として出題されていたが，「デジタル区分」の新設に伴い，「工学区分」では出題されないこととなる．なお，「デジタル区分」での出題と一部重複する分野のある数理科学系の問題は，「数学・物理・地球科学区分」の専門試験においても引き続き出題される．

2022 年度の総合職試験（春試験）のスケジュールについては図 -1 に示すとおりを予定している．

表 -2　総合職試験の試験種目

試験	試験種目	
	院卒者試験	大卒程度試験
第 1 次試験	基礎能力試験（多肢選択式） 専門試験（多肢選択式）	
第 2 次試験	専門試験（記述式） 人物試験	
	政策課題討議試験	政策論文試験
英語試験	外部英語試験（TOEFL (iBT)，TOEIC (L&R)，IELTS，英検）の結果を活用し，最終合格者決定の際に，成績に応じて加点	

総合職試験「デジタル区分」の試験問題の詳細

前述のとおり，総合職試験の専門試験には，多肢選択式と記述式があり，いずれも試験区分ごとに専門的知識，技術などを問うものである．このため，「デジタル区分」の新設にあたっては，情報系専攻の受験者の専門的知識，技術の学習達成度等を適切に測定できるように，大学のカリキュラム・シラバスの情報，有識者からの意見等を参考に出題科目の検討を行った．

情報分野は，技術革新の速い分野であり，大学および大学院においては，基礎から最先端まで幅広い分野の講義が行われている．このため，従来の「工学区分」や「数学・物理・地球科学区分」で出題している情報系の出題科目に限定するのではなく，出題内容や科目についても見直しを行った．

専門試験（多肢選択式）

多肢選択式試験は，幅広い専攻分野の受験者に対応し，また，デジタル技術の広い範囲から専門性

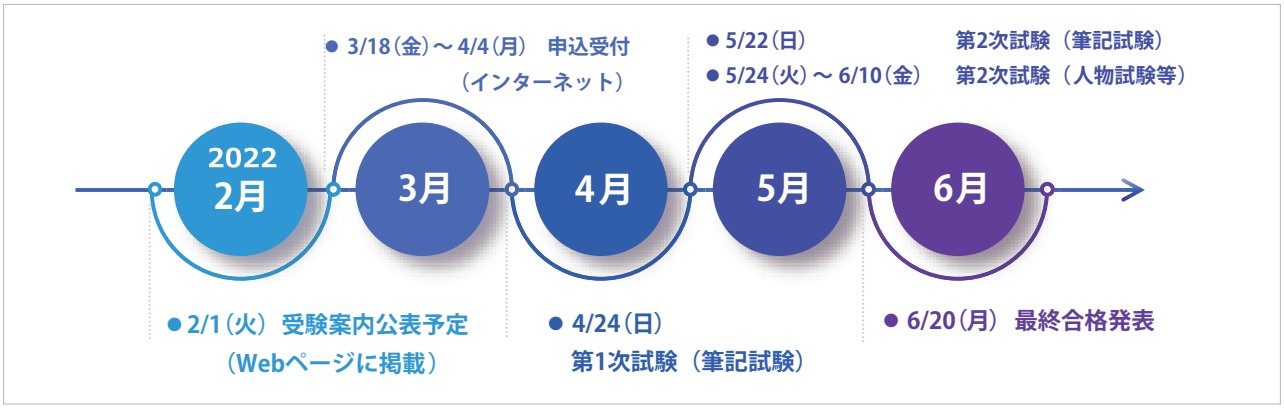


図 -1　総合職試験（春試験）のスケジュール（2022 年度）

に即した試験問題を選択できるよう，全問 63 題中から解答を要する問題を 40 題とし，このうち情報分野の基礎的な科目から受験者全員が解答を要する必須問題として 20 題を出題する．また，情報分野の中心的科目から選択必須問題として 17 題を出題し，この中から 10 題以上を選択して解答することとした．さらに，周辺領域の科目から選択問題を出題することとし，選択必須問題と選択問題を合わせて 20 題となるように解答する形式とした．表 -3 に出題科目と問題数を示す．

出題科目のうち，必須問題の「情報と社会」および選択必須問題の「情報技術」は，今回の区分新設により情報系の専門的な科目として新たに出題することとした．

「情報と社会」では，情報システムやデータ活用の進展と社会とのかかわりなどに関する基礎的な問題を出題する．また，「情報技術」では，計算機科学や情報工学の応用技術である情報セキュリティ，人工知能等に関する問題を出題する．

専門試験（記述式）

専門試験（記述式）は，計算機科学，情報工学（ハードウェア），情報工学（ソフトウェア），情報技術の 4 科目から 6 題出題し，2 題を選択して解答する（表 -4）．これらのうち，情報技術は，「デジタル区分」で新設する科目で，出題内容は多肢選択式試験と同様である．計算機科学は，従来から，「数学・物理・地球科学区分」において出題している情報科学を，より計算機に関連した内容にして出題することとした．情報工学（ハードウェア），情報工学（ソフトウェア）は，「工学区分」で出題していた科目を「デジタル区分」に移設する．「デジタル区分」では，従来の「工学区分」ではできなかった記述式問題の同じ科目から 2 題（情報工学（ハードウェア）から 2 題または情報工学（ソフトウェア）から 2 題）を選択することができることとし，これまでよりも受験者の専門分野から選択をしやすいようにした．

表 -3　専門試験（多肢選択式）の出題科目

専門試験（多肢選択式）【63 題中 40 題解答】	
必須問題【20 題解答】	基礎数学⑩，情報基礎⑦，情報と社会③
選択必須問題 【17 題中 10 題以上解答】	計算機科学③，情報工学（ハードウェア）⑤，情報工学（ソフトウェア）⑤，情報技術④
選択問題 【選択必須問題＋ 選択問題で 20 題解答】	線形代数・解析・確率・統計⑧，数学モデル・オペレーションズリサーチ・経営工学⑤，制御工学②，電磁気学②，電気工学③，電子工学③，通信工学③

○内の数字は出題する問題数

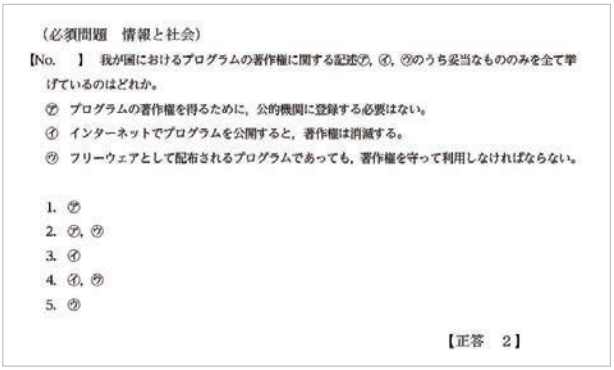
試験問題例

これまで説明した専門試験（多肢選択式）および専門試験（記述式）の試験問題例を紹介する．なお，本節で紹介する問題を含む「デジタル区分」の試験問題例は，人事院の国家公務員試験採用情報 NAVI にも掲載している．

■専門試験（多肢選択式）

今回の区分新設により新たに出題することとされた必須問題の「情報と社会」および選択必須問題の「情報技術」の試験問題例を紹介する．

○情報と社会



（必須問題　情報と社会）

【No.　】　我が国におけるプログラムの著作権に関する記述㋐，㋑，㋒のうち妥当なもののみを全て挙げているのはどれか。

㋐　プログラムの著作権を得るために，公的機関に登録する必要はない。

㋑　インターネットでプログラムを公開すると，著作権は消滅する。

㋒　フリーウェアとして配布されるプログラムであっても，著作権を守って利用しなければならない。

1. ㋐
2. ㋐，㋒
3. ㋑
4. ㋑，㋒
5. ㋒

【正答　2】

表 -4　専門試験（記述式）の出題科目

専門試験（記述式）【6 題中 2 題解答】	
計算機科学	①
情報工学（ハードウェア）	②
情報工学（ソフトウェア）	②
情報技術	①

○内の数字は出題する問題数

○情報技術

（選択必須問題　情報技術）

【No.　】 大規模なソフトウェアやコンピュータシステムの開発プロジェクトは管理が難しく，ややもすると納期が遅れたり，当初の予算を超過したり，予定していた機能を実現できないという事態が起こってきた。しばしば，発注者と設計開発者との間には，実現すべき機能の内容やその実装の難易度について認識にずれがあり，これを調整することに多大な時間と労力がかかって，開発は遅れがちになる。

その反省に立ち，近年では「アジャイルソフトウェア開発」と呼ばれる新しいプロジェクト管理が提唱されている。「アジャイル」とは，すばしこいや身軽なという意味である。

アジャイルソフトウェア開発で推奨されているプロジェクト管理の方針として最も妥当なのはどれか。

1. プロジェクトを，要件定義，概要設計，詳細設計，実装，テストの5工程に明確に分け，それぞれの工程での品質管理を徹底させることによって，工程の手戻りが起こらないようにし，プロジェクトの迅速化を図る。
2. 試作段階であっても一定程度は動作するソフトウェアを顧客に対して頻繁にリリースし，それを試した顧客からの意見を取り入れて開発計画を適宜修正することが望ましい。
3. 開発の各工程に要する「人月」(作業者数と作業月数の積)を，経験則から精度よく見積もり，人月に比例して各工程に開発資源を分配することによって，開発を迅速化し，バグの発生確率も最小化する。
4. アジャイルソフトウェア開発の最重要点は，PDCA (plan-do-check-action)サイクルを着実に実施することである。それゆえ，PDCAいずれの段階においても，次の段階に作業がスムーズに接続できるように配慮したドキュメンテーションを整備することが推奨される。
5. 機能ごとにソフトウェアのテストを繰り返すことが開発の基本方針であり，機能ごとに独立したテストが容易なプログラミング言語を用いることが望ましい。これに最も適するのは関数型言語であり，オブジェクト指向言語の使用は避けることが望ましい。

【正答　2】

■専門試験（記述式）

計算機科学，情報工学（ハードウェア），情報工学（ソフトウェア），情報技術の試験問題例を紹介する．なお，専門試験（記述式）の試験問題例は，国家公務員試験採用情報NAVIに掲載している各科目の試験問題例の一部を抜粋したものである．

○計算機科学

（計算機科学）

【No.　】 次は，C言語で書かれたクイックソートを行う関数である。a を配列，L，R を整数，a[L]，a[L + 1]，...，a[R]を配列要素とすると，関数呼出し quicksort(a，L，R)は a[L]，a[L + 1]，...，a[R]をソートする。

```
quicksort(int a[], int left, int right) {
  int p, i, pivot, temp;
  /*  (A)  */
  if (left < right) {
    pivot = a[left];
    p = left;
    for (i = left + 1; i <= right; ++i) {
      if (a[i] < pivot) {  /*  (B)  */
        ++p;
        temp = a[p]; a[p] = a[i]; a[i] = temp;
      }
    }
    a[left] = a[p];
    a[p] = pivot;
    quicksort(a, left, p - 1);
    quicksort(a, p + 1, right);
  }
}
```

(1) a[0] = 2，a[1] = 4，a[2] = 5，a[3] = 1，a[4] = 3 であるとき，関数 quicksort(a，0，4) を呼び出したとする。この計算が終了するまでに，quicksort が再帰的に呼び出される。各呼出しについて，呼び出したときの引数の変数 left と right の値はいくらか。整数値で答えよ。

(2) quicksort(a，L，R)は a[L]，a[L + 1]，...，a[R]をソートすることを説明せよ。

(3) quicksort(a，L，R)の計算は必ず停止することを示せ。

(4) quicksort(a，0，n − 1)を呼び出してからこの計算が終了するまでのコメント(B)の置かれた行の a[i] < pivot の不等号の比較回数を，再帰呼出しの実行の分まで合わせて，数えることを考える。例えば，a[0] = 0，a[1] = 1，a[2] = 2 のとき，quicksort(a，0，2)を呼び出したときの比較回数は 3 である。

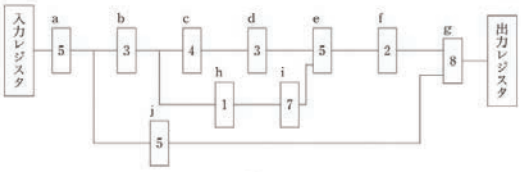
○情報工学（ハードウェア）

（情報工学（ハードウェア））

【No.　】 以下の設問に答えよ。

(1) 図Ⅰの回路は，a~jのモジュール，入力レジスタ，出力レジスタで構成されている。各モジュールは，左端を入力，右端を出力とする組合せ回路であり，示された数値はそのモジュールの入力から出力までのレイテンシ[ns]を表す。

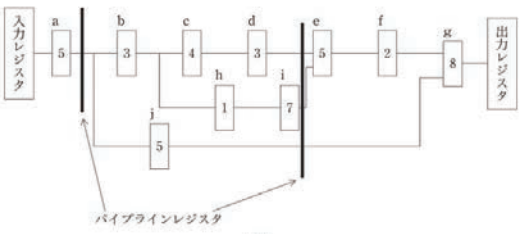
ただし，入力レジスタ，出力レジスタには同一のクロックが接続されているものとする。



図Ⅰ

(a) 図Ⅰの回路が正しく動作する最高クロック周波数は何 MHz か，整数値で示せ。ただし，出力レジスタのセットアップ時間は考慮しないものとする。

(b) 図Ⅱのようにパイプラインレジスタを挿入したとき，正しく動作する最高クロック周波数は何 MHz か，整数値で示せ。ただし，パイプラインレジスタには，入力レジスタ，出力レジスタと同一のクロックが接続されており，パイプラインレジスタ及び出力レジスタのセットアップ時間は考慮しないものとする。



図Ⅱ

○情報工学（ソフトウェア）

（情報工学（ソフトウェア））

【No.　】以下の設問に答えよ。

二つの系列がどのくらい似ているかを計測する方法として，最長共通部分列問題（longest common subsequence problem）を考える。まず，以下のように概念・記法を定義する。

【部分列】系列 $X=\langle x_1, x_2, \cdots, x_m\rangle$ に対して $Z=\langle z_1, z_2, \cdots, z_k\rangle$ が部分列であるとは，全ての $j=1, 2, \cdots, k$ に対して $x_{i_j}=z_j$ で，$i_1<i_2<\cdots<i_k$ となるようなインデックスの系列 $\langle i_1, i_2, \cdots, i_k\rangle$ が存在することである。例えば，系列 $X=\langle A, B, C, B, D, A, B\rangle$ に関して，インデックスの系列 $\langle 2, 3, 5, 7\rangle$ を考えると，$\langle B, C, D, B\rangle$ は X の部分列であることが分かる。

【最長共通部分列】二つの系列 X と Y に共通する部分列のうち，最も長いものを最長共通部分列と呼ぶ。同じ長さの最長共通部分列は複数存在し得る。

【接頭辞】系列 $X=\langle x_1, x_2, \cdots, x_m\rangle$ の長さ $p\,(p\in\{0, 1, \cdots, m\})$ の接頭辞とは $\langle x_1, x_2, \cdots, x_p\rangle$ のことであり，X_p と表す。例えば，系列 $X=\langle A, B, C, B, D, A, B\rangle$ に関して，長さ 4 の接頭辞は $X_4=\langle A, B, C, B\rangle$ である。ただし，長さ 0 の接頭辞は空系列と定義する。

二つの系列 $X=\langle x_1, x_2, \cdots, x_m\rangle$ と $Y=\langle y_1, y_2, \cdots, y_n\rangle$ の最長共通部分列の一つを $Z=\langle z_1, z_2, \cdots, z_k\rangle$ とすると，以下の定理が成り立つ。

1. もし $x_m=y_n$ ならば，$z_k=x_m=y_n$ であり，Z_{k-1} は系列 X_{m-1} と Y_{n-1} の最長共通部分列である。
2. もし $x_m\neq y_n$ かつ $z_k\neq x_m$ ならば，Z は系列 X_{m-1} と Y の最長共通部分列である。
3. もし $x_m\neq y_n$ かつ $z_k\neq y_n$ ならば，Z は系列 X と Y_{n-1} の最長共通部分列である。

この定理より，系列 X_i と Y_j の最長共通部分列の長さ $c_{i,j}$ に関する再帰的な式が得られる。

$$c_{i,j}=\begin{cases}0 & (i=0\text{ 又は }j=0\text{ のとき})\\ c_{i-1,j-1}+1 & (0<i\text{ かつ }0<j\text{ かつ }x_i=y_j\text{ のとき})\\ \max(c_{i,j-1}, c_{i-1,j}) & (0<i\text{ かつ }0<j\text{ かつ }x_i\neq y_j\text{ のとき})\end{cases}$$

C 言語を用い，最長共通部分列の長さを求める関数を，2 通りのアルゴリズム（lcs 関数と lcs2 関数）で実装した。ここで，系列 X は文字の配列 a，系列 Y は文字の配列 b で与えられ，それぞれの配列の要素数は m 及び n で与えられるとする。また，配列のインデックスは 0 から始まることに注意せよ。すなわち，系列 X の要素 x_i は a[i−1] で表される。

○情報技術

（情報技術）

【No.　】以下の設問に答えよ。

(1) 情報セキュリティに関する以下の問いに答えよ。

(a) Web サイトを設定するときには，http://で始まるよりも https://で始まるように設定されている方が一般に安全であるとされている。その理由をセキュリティに関する技術的な観点から 2 行程度で説明せよ。

(b) https://で始まる Web サイトにアクセスしたときの通信の手順について，セキュリティに関する箇所を次の語句を全て用いて 5 行程度で説明せよ。

ただし，用いた語句に下線を引くこと。

[語句：共通鍵，公開鍵，秘密鍵]

(2) ソフトウェアの品質管理に関する以下の問いに答えよ。

ソフトウェアは，テストによってその品質を確保することができる。テストには，ソフトウェアを実行する動的テストとそうでない静的テストがある。動的テストには，ソフトウェアの構造（実装）に基づくホワイトボックステスト，ソフトウェアの仕様に基づくブラックボックステスト，結果に応じて次に実施するテストを決めていく探索的テストがある。

動的テストでは，ソフトウェアの状態と与えた入力に応じて，期待した出力が得られるかどうかを調べる。その際に使う状態と入出力の組合せをテストケースと呼ぶ。一連のテストケースがソフトウェアの動作のどの程度を網羅しているかを表したものをカバレッジと呼ぶ。

(a) ソフトウェアの内部構造が分かっている際には，命令文や条件式等のコードが一連のテストケースによってどれほど実行されたかをカバレッジとすることができる。ここでは，

1. ステートメントカバレッジ 100 % とは，一連のテストケースによって，全ての命令文が少なくとも一回は実行されること
2. デシジョンカバレッジ 100 % とは，一連のテストケースによって，全ての条件文の条件式が真となる場合と偽となる場合がそれぞれ少なくとも一回は実行されること
3. 条件カバレッジ 100 % とは，一連のテストケースによって，全ての条件文の個々の条件が真となる場合と偽となる場合がそれぞれ少なくとも一回は実行されること

とする。

このとき，次のプログラムについて以下の問いに答えよ。

ただし，x，y は整数型の変数とし，関数 printf は C 言語の関数 printf と同様にフォーマット文字列に従って引数を印字する。&& 演算子は第 1 項が真のときのみ第 2 項の評価を行うものとする。

```
if (x > 2) y = 1;
if (x < 4) y = 2;
if ((x > 1) && (x < 5)) y = 3;
printf ("x=%d, y=%d", x, y);
```

一般職試験「デジタル・電気・電子区分」の概要

一般職試験（大卒程度試験）は，定型的な事務をその職務とする係員を採用するための試験である．

2022 年度の一般職試験（大卒程度試験）は，その専門性に応じた表 -5 の区分の試験を実施する．

試験種目は表 -6 に示すとおりであり，一般職試験（大卒程度試験）についても，志望する官庁を訪問し，各府省において採用されるかどうかが決定される．

一般職試験「デジタル・電気・電子区分」は，2021 年度までの「電気・電子・情報区分」から名称が変更され，これまで必須問題であった一部の問題が選択問題となることで，情報系の受験生および電気・電子系の受験生がより受験しやすい区分となった．

表 -5　一般職試験（大卒程度試験）の受験区分

行政，**デジタル・電気・電子**，機械，土木，建築，物理，化学，農学，農業農村工学，林学

＊行政区分については全国 9 つの地域ごとに区分が分けられている

表 -6　一般職試験（大卒程度試験）の試験種目

試験	試験種目
第 1 次試験	基礎能力試験（多肢選択式） 専門試験（多肢選択式） 専門試験（記述式）
第 2 次試験	人物試験

2022 年度の一般職試験（大卒程度試験）のスケジュールについては，図 -2 のとおりを予定している．

2022 年度の国家公務員採用試験について

総合職試験「デジタル区分」および一般職試験「デジタル・電気・電子区分」の科目の検討や 2022 年度試験問題の作成において，多くの専門家の方々にご尽力をいただいたことに感謝申し上げる．

繰り返しとなるが，国家公務員採用試験の受験申込期間等の情報は，例年 2 月頃の発表となることから，2022 年度試験の情報については，人事院 Web ページで随時確認されたい．

(2021 年 11 月 25 日受付)
(2021 年 12 月 15 日 note 公開)

佐藤　壮　kohchan@jinji.go.jp
2006 年人事院に採用，財務省や内閣官房への出向，人事院事務総局総務課長補佐等を経て 2020 年より現職（人事院人材局企画課長補佐（制度班））．

図 -2　一般職試験（大卒程度試験）のスケジュール（2022 年度）

特集

スマートファクトリーは工場の何を変えるのか？

編集にあたって

袖美樹子 | 国際高等専門学校　田中功一 | 三菱電機（株）

数十年前豊富な労働力を求め工場が海外に移転し，国内での空洞化が起こった．しかし近年日本を代表する各メーカが工場の国内回帰の動きを見せている．工場の何が変わったのか？

AI や IoT などのデジタル技術を活用した，生産性を高めた効率的な工場，スマートファクトリーへの移行が急速に進んでいる．人手不足の解消，生産性の向上，品質の安定化，コスト削減，製品化・量産化の期間短縮，ノウハウの蓄積・共有，新たな付加価値の提供・提供価値の向上など多くの効果をもたらすスマートファクトリー．IoT センサなどでリアルタイムに取得したデータをクラウドまたはエッジサーバに送り，そのデータを AI が分析し工場現場に反映することで，製造ラインを効率化する仕組みだ．日本のスマートファクトリーの強みはどこなのか？　第一線で取り組んでおられる企業や大学の，研究者，専門家から解説をいただく特集を企画した．

「1. 工場のスマート化を実現する最新の FA 技術と取り組み(楠和浩)」では SMKL を紹介する．日本では，スマートファクトリーの現状や将来進むべき方向性を経営層と工場現場とで共有することが大切とされている．その考えから SMKL（Smart Manufacturing Kaizen Level）を国際規格として提案，スマートファクトリー化に活用している．SMKL を用いると現状レベルの把握，次に進むべき方向の決定を経営層と工場現場間で認識を共有し継続的に改善を行うことが可能となる．SMKL を用いたスマートファクトリー化の考え方，実例を紹介いただく．

「2. リアルタイム AI 技術の製造業への適用（櫻井保志）」ではデータ活用について解説する．工場では IoT センサからリアルタイムに膨大なデータを収集できる．この膨大な時系列データを活用することにより効率を高めることが重要である．スマートファクトリーの根幹である工場内で収集された時系列ビックデータを学習し，予測，要因分析，トラブル予知，行動最適化のための情報提供をリアルタイムに行うリアルタイム AI 技術を解説いただく．また，実例の紹介もいただく．

「3. IoT プラットフォームの現状と未来―製造 DX の本質―（鈴木聡）」では IoT プラットフォームについて解説する．日本の工場現場はこれまでも徹底した効率

化，納期短縮の取り組みを行ってきた．また日々の業務に追われる工場現場にとって新たな技術導入は容易ではない．そのような状況下においても経営層と工場現場が認識を同じにし，スマートファクトリー化を進めることを可能とするIoTプラットフォームについて解説いただく．

「4. スマートファクトリーを支えるローカル5G―導入に向けた制度や技術，留意点の理解―（柿元宏晃）」では5G対応の現状を解説いただく．スマートファクトリーでは他の生産拠点とのリアルタイムな情報連携やビッグデータ解析などが不可欠である．5Gを活用すれば遅延のない他拠点との接続，大容量データの解析が可能になる．また，多接続という特徴を活かしてAGV（無人搬送ロボット）の導入が可能となり，工場の無人化を進めることが可能となる．スマートファクトリーを支えるローカル5Gに関して解説いただく．

「5. 持続可能な社会に向けた今後の生産システムと産業基盤―「人」が主役となるものづくり革新推進（HDMI）コンソーシアム　2050年に向けたロードマップから―（岩井匡代，谷川民生）」では持続可能な社会に向けた今後の生産システムと産業基盤について解説いただく．日本は世界に先立ち高齢化社会を迎え，生産年齢人口が半減すると予測されている．そのような状況においても継続して産業を発展させ成長していくことが重要である．そのためにはAI，ロボットを活用し，人はロボットと協働し生産を行うことが大切で，労働の質を高め新しい働き方を育んでいくことが重要である．2050年に向けた人が主役となるものづくりの考え方を紹介いただく．

日本経済の持続的発展の原動力となる製造業が，高い付加価値を創出し，国際競争力の維持・向上を実現することは非常に重要である．国土が狭く，資源に乏しい日本は原料を輸入，加工し，海外に販売を行うことにより成長してきた．ものづくりは我が国の国際競争力の源泉であり，工場は正にその本丸である．スマートファクトリーの考え方は改善を繰り返し発展してきた日本に非常にマッチしているように思える．

なお，本会と協力関係にある人工知能学会では「特集スマートファクトリーとAI」（人工知能学会誌2022年5月号掲載）を予定している．本特集とも関連性の高い，AI技術の詳細が解説されているので，ぜひ併読をお勧めする．

日本人は勤勉で継続力がある．新型コロナの感染者数が減少しても手を緩めることなくマスクをし，体温測定を行い，手の消毒を怠らない．これらの振舞いは工場においても重要な特性であり日本のものづくりの強さだと思う．スマートファクトリー化においても日本らしさを強みとし推進されていることが本特集から理解していただけるのではないかと思う．日本のものづくりの強さを感じていただける特集になれば幸いである．

（2021年12月8日）

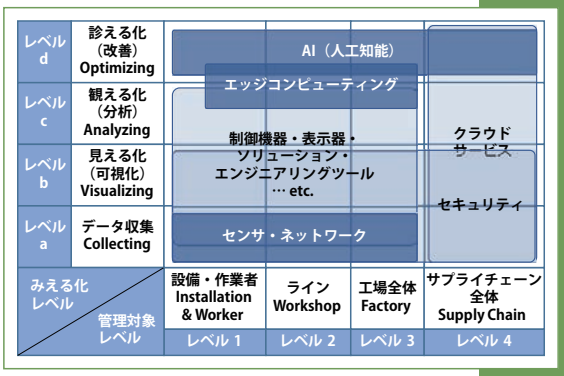
概要

① 工場のスマート化を実現する最新の FA 技術と取り組み

応用 一般

楠　和浩 | 三菱電機（株）情報技術総合研究所

スマートファクトリーの実現のためには，数多くの技術が必要である．一方で，スマートファクトリーは概念が幅広いため，技術が適用されるシステム上の階層や役割などの共通認識を合わせる必要がある．そこで，まず，参照システムアーキテクチャについて述べる．次に，代表的な技術として「ネットワーク」「エッジコンピューティング」「AI」を取り上げ，それぞれの技術動向について例を交えて解説する．

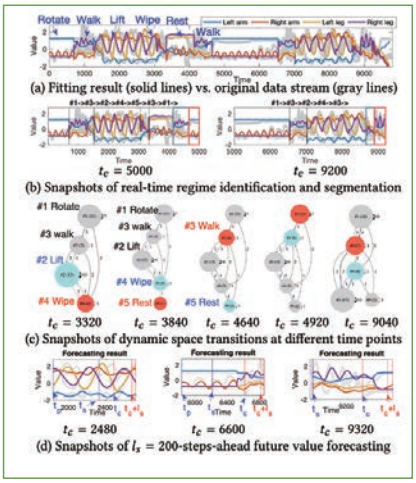


② リアルタイム AI 技術の製造業への適用

応用 一般

櫻井保志 | 大阪大学 産業科学研究所

本稿では，IoT ビッグデータを高速に学習し，予測，要因分析，トラブル予知，行動最適化のための情報提供をリアルタイムに行う AI 技術基盤について述べる．その中でも特に 3 種類の要素技術，(1) リアルタイム予測と動的要因分析，(2) リアルタイム時系列テンソル解析，(3) リアルタイム特徴自動抽出，について概説する．また，実際の工場の生産ライン支援や自動車運行管理など，製造業 DX に関連する取り組みを実例とともに紹介する．

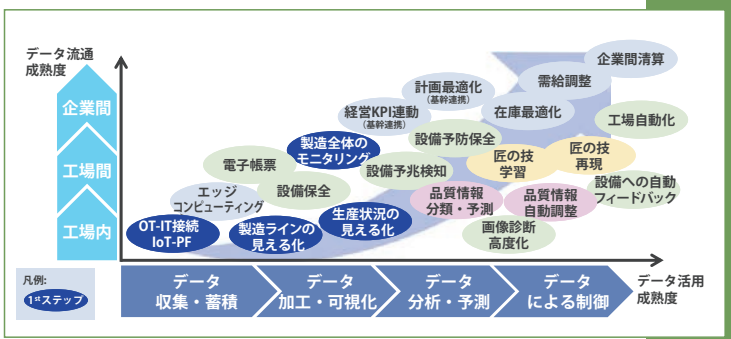


③ IoT プラットフォームの現状と未来
—製造 DX の本質—

応用 一般

鈴木　聡 | （株）NTT データ

大きく変化する市場の中で，製造業は従来の強みを認識した上で，変革を進めていくことが求められている．社会的変化からモノづくりを超えさまざまなステークホルダの期待が加わる．デジタル技術を活用した製造 DX を実現する手段である IoT プラットフォームは企業内・企業間のあらゆるデータをつなげ，新しい価値を生み出す可能性を秘めている．IoT プラットフォームの未来を考察し，現状について技術的アプローチを含めて概説する．



④ スマートファクトリーを支えるローカル 5G
―導入に向けた制度や技術，留意点の理解―

応用

柿元宏晃　NTT コミュニケーションズ（株）

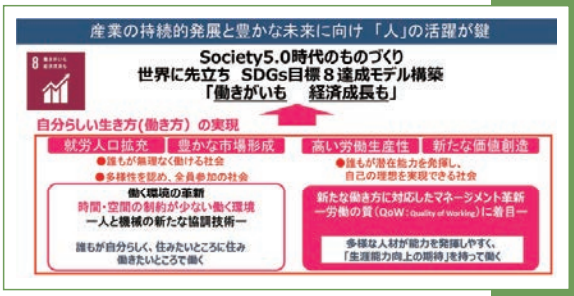
昨今，ローカル 5G はスマートファクトリーを実現するためのテクノロジーの 1 つとして注目され，2019 年の制度施行以降，実証実験や企業での導入等の事例が着々と増えている．一方，導入に際してはいくつかの障壁がある．技術面・コスト面はもとより，免許制度面も併せて理解しておく必要がある．本稿ではローカル 5G の免許制度や関連する主要技術，現場に導入する際の進め方やコストの考え方等について紹介する．

⑤ 持続可能な社会に向けた今後の生産システムと産業基盤
―「人」が主役となるものづくり革新推進（HCMI）コンソーシアム 2050 年に向けたロードマップから―

応用

岩井匡代　谷川民生　国立研究開発法人　産業技術総合研究所
HCMI コンソーシアム

将来産業の環境は，人口動向などの社会課題や自然環境が産業基盤に及ぼす影響が無視できない状況にある．持続可能な社会への対策として消費社会を形成する就労人口の拡充と活躍が不可欠であり，さらに災害時も生活や事業の継続が重要課題になる．その課題解決のためには，生産システムを人モデルなどを導入した Human-in-the-Loop に進化させ，人と機械が時空を超えて協調する新たな（遠隔）協調型協働システムと労働の質（QoW）に着目したマネジメントシステムの実現が求められる．これを踏まえ，産総研 「人」が主役となるものづくり革新推進（HCMI）コンソーシアムのロードマップを策定した．



特集
Special Feature

［スマートファクトリーは工場の何を変えるのか?］

① 工場のスマート化を実現する最新の FA 技術と取り組み

応
般

楠　和浩　三菱電機（株）情報技術総合研究所

スマートファクトリーとは何か

スマートかどうかは別にして，ファクトリー（工場）と聞いて，皆さんは何を想像するだろうか？

ある人は，たくさんの作業者が，流れてくる自動車にドアや内装を装着しているような自動車工場（最終組み立て工程）を思い浮かべるかもしれない．あるいは，同じ自動車でも，塗装ロボットが自動車のボディを囲んで塗料を吹きかけているような場面を想像するかもしれない．また，人によっては，真っ赤な鉄が流れていく工場（鉄鋼の圧延ライン）を思い浮かべるかもしれないし，いわゆる家族経営で営まれているような金属加工工場を想像する人もいるかもしれない．

このように，製造する製品によって，製造設備も違えば製造の仕方も異なる．また，最終製品は同じでも，その途中工程で使われる機器や装置などがまったく違うのは当然である．

2011 年にドイツで提唱されたインダストリー 4.0 を契機として，また，最近では製造業 DX と絡めた話題として，「スマートファクトリー」という言葉は，さまざまな場面で取り上げられている．しかしながら，先に示した通り，ファクトリー（工場）はさまざまであり，結果，スマートファクトリーの定義は難しく，実は公式な定義は存在していない．

さまざまな論文や記事で議論がなされているが，ここでは，デロイトトーマツ合同会社の Deloite University Press「The smart factory」において定義された下記を「スマートファクトリー」の定義とする．

「スマートファクトリーとは，広範なネットワークで自らパフォーマンスを最適化し，リアルタイムまたは，ほぼリアルタイムで新たな状況に自ら適応して学習をし，自律的に生産プロセス全体を動かすことができる柔軟なシステムである」

しかしながら，この定義は一般的な概念としてスマートファクトリーを定義しており，したがって，この定義を基に関係する技術，あるいは標準化などに関する議論をしようとすると混乱が生じる可能性がある．つまり，全体概念としてはスマートファクトリーに関係していても，実は，話し手によって想定しているシステムや適用対象が異なっている状況が発生し，その結果，議論そのものがかみ合わない場合が存在する．

そこで，まず技術的な「アーキテクチャ定義」が必要になる．

たとえば，インダストリー 4.0 においては，2015 年に発行された「インダストリー 4.0 実現戦略」において「インダストリー 4.0 リファレンスアーキテクチャモデル（RAMI4.0）」（図 -1）が発表されている．まず，この RAMI4.0 について，簡単に解説する．

見た目の通り，RAMI4.0 は，3 次元の構造になっている．それぞれの軸の定義と概要は以下の通りである．

(1) 垂直方向ー Layers

生産システムの構成要素を情報通信工学的視点から分類したものである．具体的には，事業層（Business），機能層（Functional），情報層（Information），通信層（Communication），統合層（Integration），物体層（Asset）の 6 層に分かれている．

(2) 左横軸方向ー Life Cycle & Value Stream

製品のライフサイクル管理を示す軸であり，国際標準 IEC 62890 に基づいている．

この軸では，「タイプ」と「インスタンス」という概念を導入している．「タイプ」とは，製品になる前の段階，つまり設計図・試作品であり，「インスタンス」が製品と考えてよい．

(3) 右横軸方向ー Hierarchy Levels

ネットワークでつながった生産システムの階層レベルを示している．ビジネス・製造システム統合の国際標準 IEC 62264 で規定されている階層レベルを拡張した定義となっている．

具体的には，つながる世界（Connected World），企業（Enterprise），ワークセンタ（Work Centers），作業ステーション（Station），制御装置（Control Device），フィールド機器（Field Device），製品（Product）という 7 つの階層が定義されている．

この参照アーキテクチャを使うことで，研究開発ターゲットのインダストリー 4.0 における位置づけを明確にすることができる．

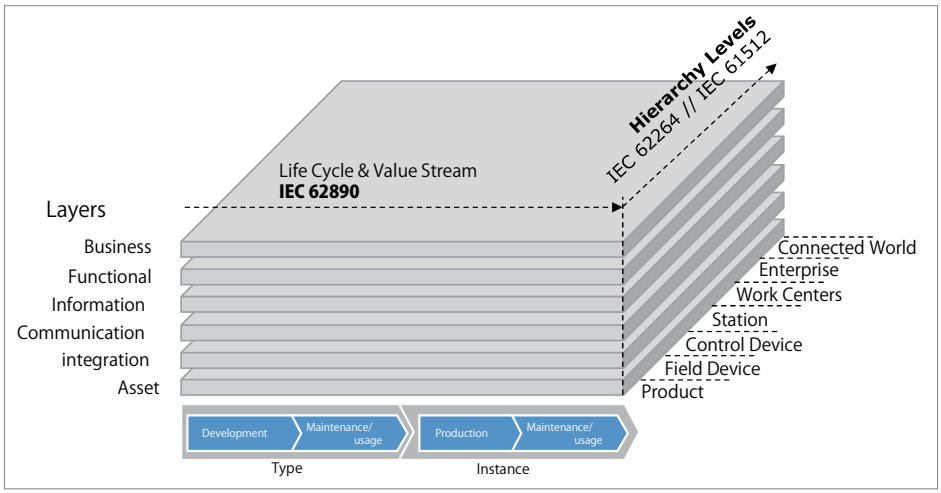
一方，すでにある製造現場をスマート化，もう少し正確に言うと IoT 化するために，経営層と現場管理者との間で投資効果判断をするための指標を作る動きもある．

スマート工場の実現には，スマート工場化の現状や，将来進むべき方向性を，経営層と現場とで共有することが重要である．共有することで，スマート工場実現のための適切な投資計画を継続的に立案・実行することが可能となるからである．

共有のための手段として，日本から提案し，現在 ISO/TC 184 で国際規格として審議中の SMKL（Smart Manufacturing Kaizen Level）（図 -2）がある．

SMKL は，「工場をどう IoT 化していけばよいか分からない」という悩みを持つ製造現場の IoT 化推進を支援する目的のために作られたものであり，図 -2 のように IoT 導入レベルを，4 つの「みえる化」段階と，4 つの管理対象レベルで区分けされた 16 マスで表すことで，対象とする製造現場がどの段階まで IoT 化が進んでいるかを判断できるようにするものである．

縦軸のみえる化レベルは，レベル a「データ収集」，レベル b「見える化（可視化）」，レベル c「観える化（分析）」，レベル d「診える化（改善）」とし，いずれも電子化かつ自動化されていることを到達条件としている．



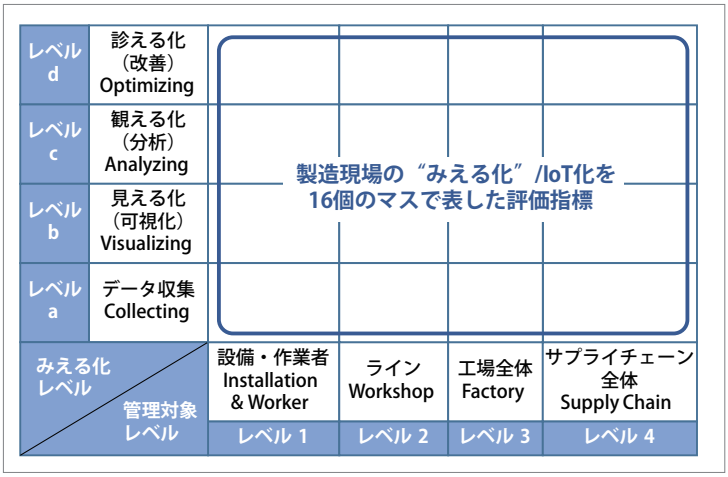
■図 -1　インダストリー 4.0 リファレンスアーキテクチャモデル（RAMI4.0）

また，横軸の管理対象レベルは，いわゆるディスクリート系の工場を対象とした場合，レベル 1「設備・作業者」，レベル 2「ライン全体」，レベル 3「工場全体」，レベル 4「サプライチェーン全体」とするが，工場の種類に応じた定義を行う必要がある．

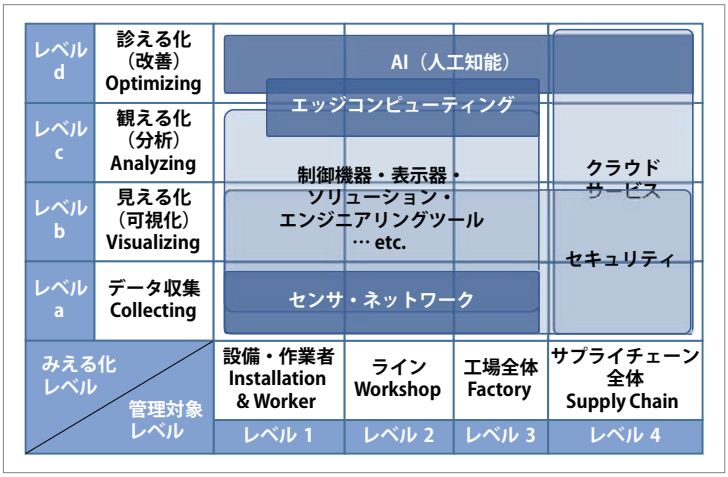
SMKL に基づいて，現状のレベル把握と次に進むべきレベルの決定を行い，そのための投資計画を立案することで，経営層と現場との間で理解を共有しながら継続的にスマート工場化を前進させることができる．

工場スマート化のキー技術

では，次に，工場のスマート化におけるキー技術について考えてみたい．図 -3 は，先に述べた SMKL のそれぞれのステージにおいて，重要と考えられる技術名を示したものである．以降，この中で，ネットワーク，エッジコンピューティング，AI について簡単に解説する．



■図 -2　Smart Manufacturing Kaizen Level（SMKL）



■図 -3　工場スマート化のキー技術

ネットワーク

工場の製造現場で利用されるネットワークは，製品製造にかかわる機械や装置あるいは設備の制御を行うために利用される．具体的には，複数の機械や装置の間の作業を連携させたり，同期をとるなどの目的で利用される．

製造現場における制御は，周期的かつ高速に実施されるため，たとえば制御用コントローラ（製造現場内のセンサやアクチュエータを制御する PLC（Programmable Logic Controller）や，工作機械の制御を司る NC（Numeric Control）等がある）間を接続するネットワークでは，数ミリ秒から数十ミリ秒での応答性能が必要になるとともに，応答性能の揺れ（ジッタ）も数マイクロ秒単位で要求される場合がある．

また，最近では，同一ネットワーク上に複数の制御周期を持つ制御用コントローラを接続できるようにしたい，という要求もでてきている．

一方，先ほど述べた SMKL の「みえる化レベル」の「レベル a（データ収集）」では，製造現場あるいは工場全体の稼働効率を高めるために，製造現場内ネットワークを利用して，機械，装置，設備などから製造状態に関するデータを収集する場合がある．この場合には，先に述べた制御の実行とは異なり，応答性能は比較的遅い代わりに転送できるデータ量が多いネットワークが必要である．

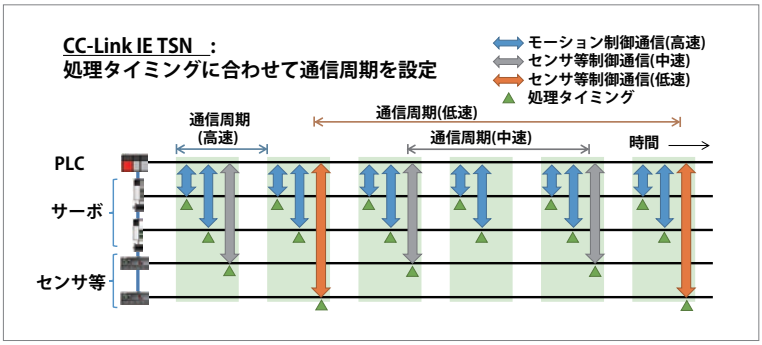
工場の製造現場内のネットワークは，この両方の要求を満足するような仕様になってい

るものが多い．

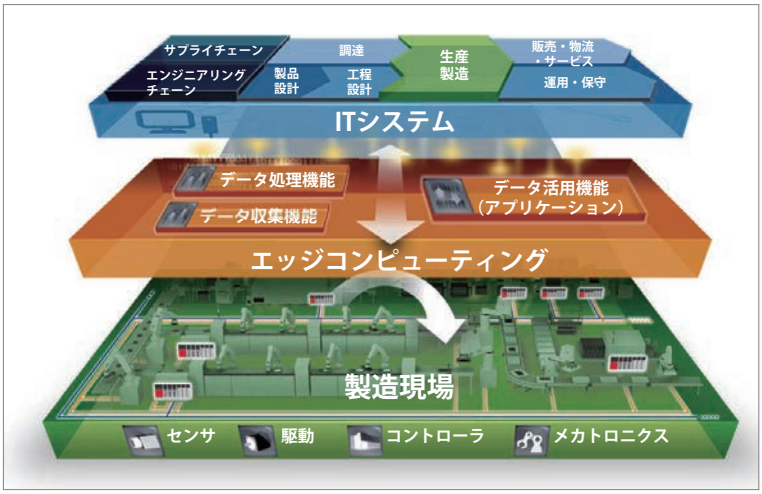
代表的な工場内ネットワークとしては，CC-Link IE TSN（CC-Link 協会），PROFInet（PROFIBUS & PROFINET International），EtherNet/IP（Open DeviceNet Vendor Association），EtherCAT（EtherCAT Technology Group）などがある．

代表的なネットワークである CC-Link IE TSN におけるデータ転送制御方法は以下の通りである．

まず，応答時間の違いや応答性能の揺れ（ジッタ）の違いを同一のネットワーク上で実現するための基本方式としては，IEEE（米国電気電子学会）のイーサネットデータリンク層である TSN（Time Sensitive Network）を活用する．具体的には，IEEE802.1AS（時刻同期）および IEEE802.1Qbv（スケジュールされたトラフィックの拡張）技術を利用し，図 -4 にあるように，制御コントローラ（図では PLC）から，処理タイミングの異なるサーボやセンサなどに対するデータ転送を実施する．



■図 -4　工場の製造現場で利用されるネットワークにおける転送制御（ex. CC-Link IE TSN）



■図 -5　3 階層システムアーキテクチャ（ex. 三菱電機 e-F@ctory）

一方，最近では，製造現場における無線通信の活用も活発に議論されている．変種変量生産に対応するために，製造現場も製造する製品に対応する必要が生じており，装置の変更，あるいは生産ラインの組み換えや製造中の製品の流れを柔軟に変更するなどの必要性が生じている．その結果，これまで，ほぼ固定であった装置や機械などもレイアウトフリーで稼働できるようにする必要があり，あるいは無人搬送車（AGV）を含む可動ロボットの活用なども考える必要がでてきている．その結果，まず，通信を行う対象物の位置が変化するために無線通信を利用する必要が生じてきている．

また，スマートファクトリー実現のためには，装置・機械の間，装置・機械と人，装置・機械と製造物（ワーク）などインタラクションする関係が増えてくることも無線化が必要な一因と考えられる．

特に，最近では，ローカル 5G の活用が盛んに議論され始めている．5G 仕様の高度化・環境構築の容易化（低価格化）のトレンドに合わせて，作業効率化→協調作業→自律制御への適用範囲が広がると期待されている．

エッジコンピューティング

スマートファクトリーを実現するためのシステムアーキテクチャは，当初は，クラウド＋製造現場の 2 階層システムアーキテクチャである場合が多かったが，最近では，これにエッジコンピューティングを加えた 3 階層システムアーキテクチャ（図 -5）を前提とした議論が多くなってきている．

これには，次のような理由がある．

まず，製造現場内には，機械，装置が

数多くあり，また，さまざまな状態を検知するセンサ類も数多く存在する．制御周期が短くなる（きめ細かい制御を実施する）ことに比例して，収集すべきデータ量も膨大になる傾向がある．この大量のデータをクラウドに直接あげて処理するのは，ネットワーク負荷の観点から得策ではない．また，特に，分析結果を制御そのものに反映させようとすれば，データ解析を含めた全体性能にリアルタイム性が求められ，結果，クラウドに接続するためのネットワーク遅延が問題になる．

したがって，データの発生源および活用の場に近いところで情報処理を行う「エッジコンピューティング」が重要になってきている．

エッジコンピューティング層は，以下に示す機能から構成される．

(1) データ収集機能

製造現場には，製造メーカや接続プロトコルが異なる数多くのセンサ・アクチュエータ・装置・設備が存在する．製造現場からのデータ収集においては，通信プロトコルの差異を吸収し機器メーカやネットワークを問わずにデータを収集できる機能が必要である．

(2) データ処理機能

製造現場から集めたデータの処理(データの収集・加工・診断）を行うための機能群．特に製造現場の機械・装置・ラインを抽象化し，かつ，それらの役割を階層的に管理できる仕組みが必要である．

(3) データ活用機能（アプリケーション）

データ処理機能を使って処理された製造現場データを基に，現場の稼働監視や予防保全を行う，あるいは AI を活用した最適制御を行う等のアプリケーションが配置される．

なお，このエッジコンピュータは，サイバーフィジカルシステム（CPS）と関連付けて言及されることが多い．

つまり，現実空間（ここでは製造現場）からリアルタイムに得られる情報（データ）を，サイバー空間（ここではエッジコンピューティング層）で処理し，現実空間での生産システムの最適化を実現することがサイバーファクトリーである，と考えられていることによる．

AI

工場の製造現場に対する AI の適用には，2 つのパターンがある．

1 つは，加工精度の向上，予防保全による稼働率の向上，あるいは，最近では匠の技を自動化して継承するなどの目的のために，製造する機械や装置の制御に AI を適用する場合である．

また，もう 1 つのパターンは，工場全体の歩留まり向上や，生産性向上を目的とし，人を含めた製造工程に適用する場合である．

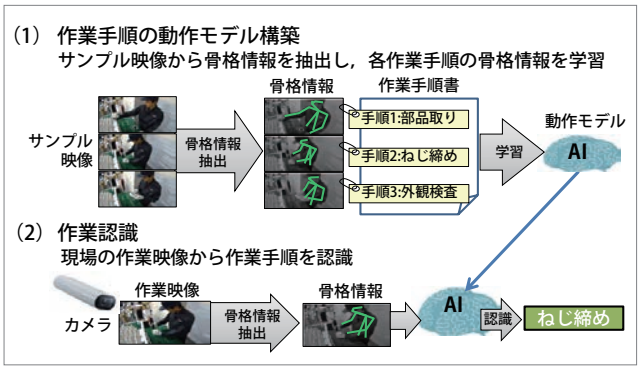
ここでは，後者の場合について解説する．

すでに多くの工場で，設備や機械のデータを活用したデータ分析の取り組みは多く行われている．しかしながら，さらに生産性を向上するためには「人」や「作られるモノ」の動きを解析できるようになる必要がある．

ここでは，カメラ画像を基に作業分析を実現する技術を例について述べる．

この技術は，AI を用いて，カメラ映像から人の骨格情報を抽出・分析し，特定の動作を自動検出するものである．

図 -6 に示すように，この技術は，サンプル映像



■図 -6 カメラ映像を用いた作業認識の仕組み

から抽出した作業者の骨格情報を，部品取り，ねじ締め，外観検査といった作業手順単位で AI に学習させる．その後，現場の作業映像から同様に骨格情報を抽出し，作業者がどの作業手順を実施中であるかを AI が自動的に認識する．

実際に工場で実施した検証によると，作業認識率は，目視と同程度の 90% であり，人間による目視作業の代替が可能なレベルにある．認識結果は，作業バランスの確認や作業手順誤りの自動検出など，さまざまな用途で活用できる．

また，動作研究の先駆者である Frank Gilbreth 氏が提唱した「動作経済の原則」（疲労を最も少なくして有効な仕事量を増やす，人間のエネルギーを効率的に活用するための約 30 項目からなる経験的な法則）に基づき，骨格の動きを分析することで，作業者の無理・無駄な体の動きの課題を見える化（可視化）することもできる．

これにより，異なる監督者であっても同じ課題を見つけることができるため，属人性を排除した標準的な作業改善が可能となる．

このような技術を利用することにより，生産現場の作業者の動きをカメラで撮影するだけで作業内容を認識・特定し，作業時間や作業ミス，無駄を自動検出することで作業分析を効率化でき，生産現場の生産性向上が期待できる．

今後の展望

本稿では，スマートファクトリーを実現するためのキー技術について述べた．

最初に述べたように，スマート工場の厳密な定義はない．したがって，当然，何をやればスマート工場になるかはさまざまである．また，同時に「スマート」は，実現するための技術の事を指すわけではなく，各人が目指すファクトリー，さらには製造業の姿そのものを表す言葉である．

変種変量生産への対応，つながる工場などスマート工場を表す言葉は数多くあり，それらを実現するためには，必ずしも最新の技術を活用する必要がないことも事実である．

ただし，製造現場で発生するさまざまな事象をデータとして取得し，それを分析・解析し，製造現場だけでなく，サプライチェーンも含めた製造業全体システムをある目的に向かって改革していくために，これまで以上に，「情報処理技術」が重要になってきていることは確かである．

（2021 年 10 月 27 日受付）

■楠　和浩（正会員）
Kusunoki.Kazuhiro@ea.MitsubishiElectric.co.jp
三菱電機（株）情報技術総合研究所長．静岡大学情報学部客員教授．博士（工学）．

特集
Special Feature

［スマートファクトリーは工場の何を変えるのか?］

② リアルタイム AI 技術の製造業への適用

応用 一般

櫻井保志　大阪大学 産業科学研究所

製造業 DX のためのデータストリーム解析

製造業におけるデジタルツインやデジタルトランスフォーメーション，自動車分野のコネクティッドカー・サービス，デバイスや材料開発におけるマテリアルズインフォマティクスなど，産業や社会は大きく変化し，このような状況において AI やビッグデータ解析は第 4 次産業革命を支える技術として期待されている．その中でも AI 関連の技術については深層学習を用いたソフトウェアの開発が活発になっており，製造業，自動車，金融，医療・ヘルスケア，小売業など広範囲に渡る事業に活用されている．深層学習による AI 実用に関しては基本的なアプローチとして，自然言語処理による文書解析や対話サービス，画像処理に基づく判別処理などの課題に対して，大量のデータから 1 つの予測モデルを学習し，その固定的なモデルから解析処理を行うような取り組みが主流である．

一方，これから本格的に到来するデータ駆動型社会においては IT, IoT などによりさまざまな端末やセンサからデータを収集することになるが, Beyond 5G 時代においては通信環境がきわめて充実し，大容量・高速・低遅延・多点同時接続の通信が実現するとともに，次世代 IoT 技術の進化に伴いデータ量は爆発的に増大する．ネットワークから大量に流れてくるビッグデータ，すなわちデータストリームには，刻々と変化する環境や突発的な外的要因の影響などにより，時間とともに変わり，時には急激に変化する時系列情報が含まれている[☆1]．実社会において多種多様な設備やデバイスから発生する IoT データストリームには，設備やデバイスに共通する特徴を含む場合もあれば，個体差を示すこともあり，さらにさまざまな状況において発生する事象間の関係性や因果関係なども含んでいる．そのため，データストリーム解析においてはモデルを固定することなく，時系列パターンの急激な変動や設備個別の特徴や傾向の変化を高速に検出し，モデル学習やモデル更新をリアルタイムに処理することが重要であり，このような新しい課題に対応することが，製造業をはじめとする事業の成否を左右する大きな鍵となる．

製造業と IoT を巡る環境が急速に変化し，生産性，品質向上，省エネルギー化のためのスマート工場に関する技術開発は必要不可欠になっている．スマート工場の実現は，設備から稼働状態などの情報をリアルタイムに収集し，人の指示を介さず自律的に判断し，工程を効率化／最適化して，より優れた製品を創出する．特に近年，サイバー空間上に実際の製品や製造工程を再現してリアルタイムに現実世界と連動させようとするデジタルツインに関する取り組みが注目されている．ただ現状，シミュレーションを通して 3 次元データで製品の特性をテストするなど，生産のリードタイムの短縮にとどまっている．物理法則に沿った数式として表現できるタスクではシミュレーションツール

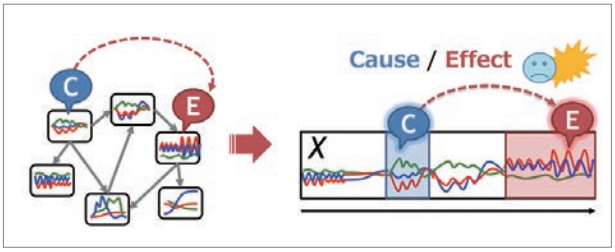
☆1 データストリームの基礎技術については，有村博紀，喜田拓也：データストリームのためのマイニング技術，情報処理，Vol.46, No.1, pp.4-11 (Jan. 2005)，櫻井保志：時系列データのためのストリームマイニング技術，情報処理，Vol.47, No.7, pp.755-761 (July 2006) を参照．

が使えるものの，大規模な製造システムの稼働状況を解析するようなタスクを想定した場合，多数の設備の動作状況や周辺環境が複雑に影響し合い，それら影響の度合いや因果関係が自明ではないため，シミュレーションツールによって課題を解決することは難しい．そこで，生産工程における設備稼働データを統合的に解析し，製造工程を効率化し，リアルタイムに最適化するような高度な AI ソフトウェアが求められている．

本稿では，筆者らがリアルタイム AI 技術と呼んでいる,IoT ビッグデータを高速に学習し，予測，要因分析，トラブル予知，行動最適化のための情報提供をリアルタイムに行う AI 技術基盤について述べる．その中でも特に 3 種類の要素技術，(1) リアルタイム予測と動的要因分析，(2) リアルタイム時系列テンソル解析，(3) リアルタイム特徴自動抽出，について概説する．また，実際の工場の生産ライン支援や自動車運行管理など，製造業 DX に関連する取り組みを実例とともに紹介する．

リアルタイム予測と動的要因分析

図 -1 と図 -2 はリアルタイム将来予測・動的要因分析技術[1), 2)]の概要を示している．IoT/ センサデータストリームをはじめとする大規模な時系列データから，リアルタイムに特徴や潜在的なトレンド（レジーム）を検出し，各レジーム間の動的な関係性を見つけることにより，長期的かつ継続的に時系列イベントストリーム内の重要な動的要因を監視し，将来のイベント予測を行う．



■図 -1　時系列データストリームにおける動的関係（要因―結果）の解析の様子

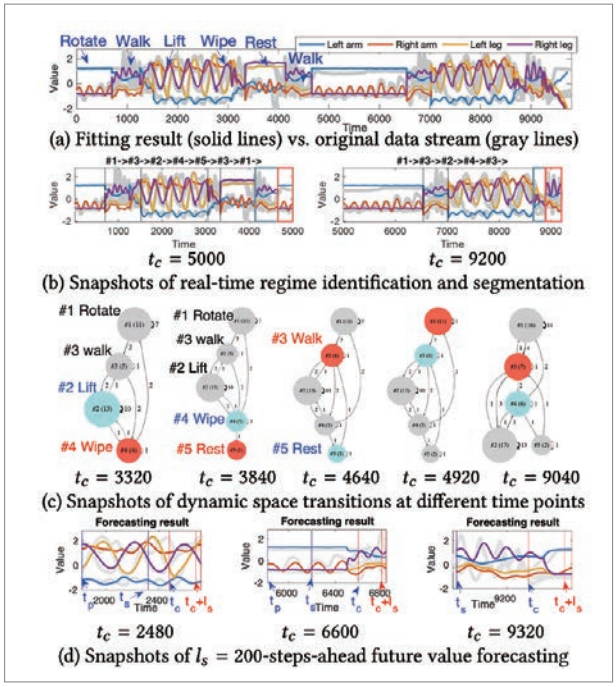
データストリームのためのリアルタイム予測と動的要因分析

開発技術は次の 3 つの特長を有する．

非線形動的システムに基づくリアルタイム学習：実世界における時系列データストリームは，さまざまな時系列パターンから構成され，外的要因などによって突発的に変化していく．本技術はデータストリームの最新時刻の潜在的トレンドや時系列パターンを動的に把握，モデル化し，将来値を継続的に予測し続ける．特に，さまざまな時系列ダイナミクスを非線形動的システムによって表現するとともに，時系列トレンドの急激な変化を変化点としてリアルタイムに検出し，モデルパラメータを瞬時に切り替え，柔軟に予測を行う．

アルゴリズムとしては，まずデータストリームから学習したさまざまな非線形方程式のモデルをデータベース（モデル DB）に格納する．予測処理においては，現在の時系列パターンに合うモデルを DB から探索し，探索したモデルを用いて予測値を推定する．

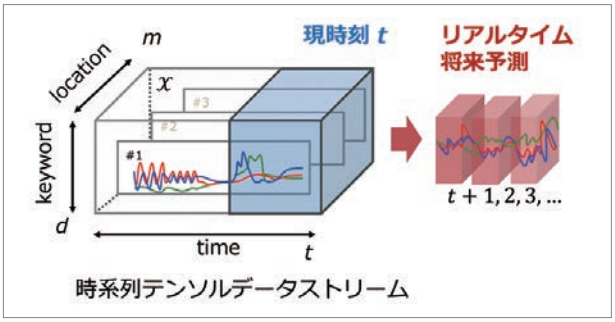
動的要因分析:一般に，大規模な時系列データストリームは，自然現象や人々の社会活動，さらにはさまざま



■図 -2　リアルタイム要因分析・将来予測技術

な設備の動作状況など，さまざまな事象を表現している．本技術は，時系列ビッグデータから時系列モデル間の前後関係（要因－結果関係）を捉え，それらの事象の連鎖を動的空間遷移ネットワークとしてモデル化する．さらに，要因分析と動的空間遷移ネットワークを用いることにより，予測精度を向上させている．図 -2 は本技術を用いたリアルタイム要因分析の出力例である．ここでは，加速度センサデータを用いて解析している．より具体的には，工場における作業者の両手足 4 カ所に加速度センサを設置し，100Hz（ヘルツ）で加速度を計測している．図 -2 (a) は，オリジナルのデータストリームの学習結果を示し，図 -2 (b) は，各時刻におけるリアルタイムパターン検出とモデル生成，図 -2 (c) は各時刻におけるネットワークの成長の様子を示す．より具体的には，作業者の行動の間のつながり（回転する→歩く→持ち上げるなど）をネットワークとして示している．図 -2 (d) は，学習した動的モデルとネットワークを用いたリアルタイムの様子を示している．ここでは，200 単位時刻先（つまり，2 秒先）の行動を予測している．現時刻 t_c において，時刻 t_c+l_s を予測している．ここで，l_s は予測する長さを示す．本技術によってリアルタイム要因分析を実現することができ，たとえば，スマート工場における装置故障，自動車走行における急ブレーキや急なハンドル操作など，さまざまな事故やトラブルの兆候（サイン）をビッグデータから高速かつ自動的に抽出することが可能となる．

リアルタイム行動支援：本技術は要因分析と予測に基づいてさまざまな状況を引き起こす要因を検出することで，リアルタイムに最適な行動を選択，推薦情報として提示する．技術の応用先としては，スマート工場，すなわち製造業の設計製造設備の高度化，さらには建設現場や車両運転における事故の防止など多岐にわたり，生体データや視線データを用いた介護やヘルスケアのサポートにも有効な技術である．



■図 -3　リアルタイム時系列テンソル解析

リアルタイム時系列テンソル解析

リアルタイム時系列テンソル解析技術 [3] は，リアルタイム予測技術 [1, 2] をベースに，テンソル解析技術として発展させることによって開発された基礎技術であり，図 -3 は，複合データストリームの時系列テンソル解析，特にリアルタイム将来予測の様子を示している．本技術は，現時刻 t における動的パターン（図 -3 青色個所）を解析することにより，将来発生するイベント（図 -3 赤色個所）をリアルタイムに予測し続ける．ここでの例は，オンライン活動データを用いているが，本技術は，オンライン活動データにとどまらず，製造業における生産工程の設備データなど，さまざまな複合時系列ビッグデータストリームのリアルタイム解析・予測に適応することが可能である．

図 -4 は，Web 上におけるキーワードアクセス件数推移データに対し本技術を適応した例を示している．時間，地域，キーワードのように複数の属性を持つテンソルデータストリームが与えられたとき，最新の観測データ（図 -4 (a) 青）を監視しながら潜在的なトレンドを発見し，適応的にモデルを変化させながら長期先のデータ（図 -4 (a) 赤）を予測し続ける．このとき，図 -4(b) のように各地域で共通する周期（季節）パターンを抽出し，それらに基づき，図 -4 (c) に示すように類似パターンを有する地域のグループ化を行う．このように本技術は，非線形微分方程式に基づき，非線形性を有する時系列ダイナミクス，長期トレンド，周期性を同一のモデル空間で表現し，そして時系列テンソルの内部において類似した潜在トレンドを持つ属性データのグループ化を自動的かつ効率的に行う．

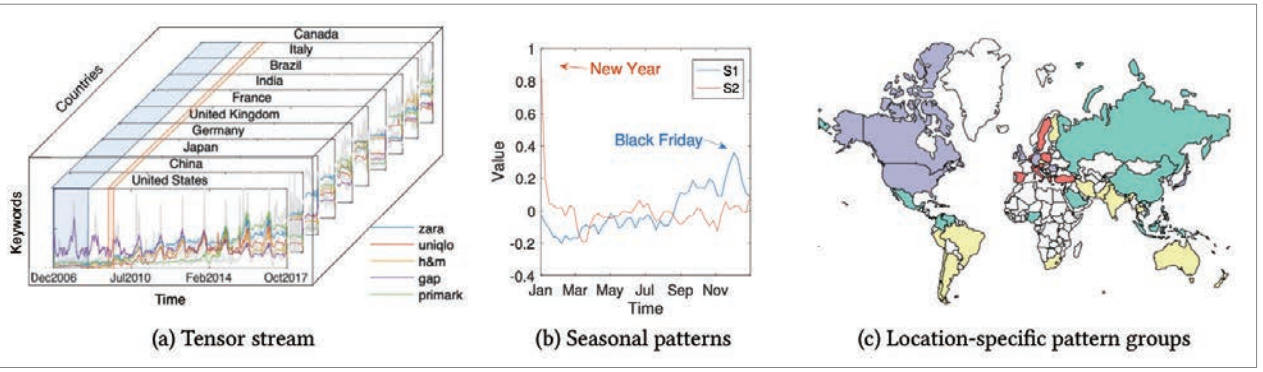
リアルタイム特徴自動抽出

リアルタイム特徴自動抽出技術[4), 5)]は，大規模時系列データストリームに含まれる典型的なパターンを発見するための技術であり，図 -5 は技術の概要を示している．与えられたデータストリームをリアルタイムに解析し，時系列パターンの種類と変化点を発見し，それらの特徴をモデルパラメータとして表現する．

特徴自動抽出：大規模データストリームの解析には高度なマイニングアルゴリズムが求められるが，複雑な演算による計算コストに加えて高度なパラメータチューニングなどによる時間的・人的コストが高くなり，実用化の際にはそれらが大きなボトルネックとなる．本技術では符号化理論に基づくモデル評価基準を応用し，解析データに関する事前知識を必要とせず，データの要約情報（時系列パターンの種類・変化点）を自動的に取得する．より具体的には，時系列シーケンスの符号化コストとコスト関数を定義し，コスト関数を最小化するようなモデル数，セグメント分割や割り当て，モデルパラメータを推定することにより，時系列データの特徴を自動抽出する．

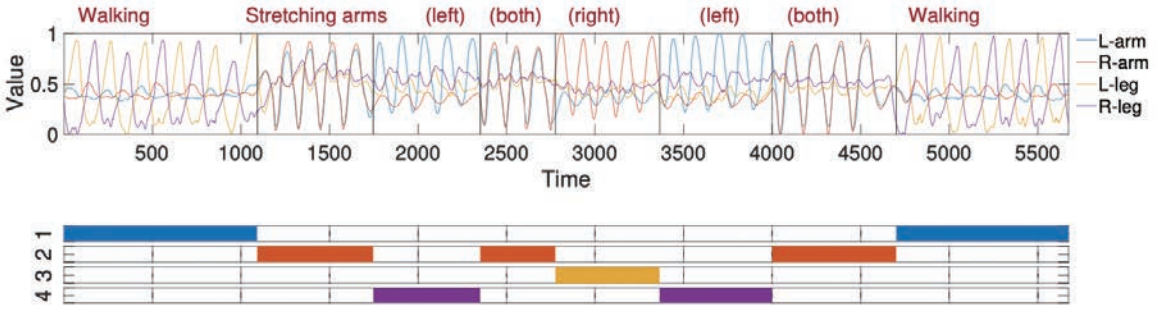
図 -5 は本技術を用いたリアルタイム解析の出力例である．図 -5 上段はヒトの両手足 4 カ所に加速度センサを設置し，歩行と 3 種類（両腕，右腕，左腕）のストレッチで構成される合計 4 種類の動作を捉えたデータストリームを示す．図 -5 下段の番号は，本技術が検出したパターンの ID を示し，長方形の両端は各パターンの開始点・終了点を示す．本技術はセンサデータをリアルタイムに監視し，歩行，両腕のストレッチ，と次々にパターンの変化点を検出する．このとき，モデル評価基準を用いて新たなモデル生成の必要性を自律的に判断することにより，パターンの種類（モデルの数）を自動的に決定する．図 -5 に示すように，本技術は 8 つのパターン変化点を検出し，それらを 4 つの動作へと分類することに成功している．

高速パターン検出：本技術は階層構造を有する確率モデルを用いて過去に検出したパターンの特徴を表現し，新たなデータが観測されるたびに類似パターンの



(a) Tensor stream　(b) Seasonal patterns　(c) Location-specific pattern groups

■図 -4　複合ビッグデータストリームのためのリアルタイム解析技術



■図 -5　モーションキャプチャデータストリームに対するリアルタイム特徴自動抽出技術の出力結果

検索とモデルパラメータの推定を逐次的に行う．このとき，最後に観測したパターンのデータと過去に推定したモデルパラメータのみを保持するため，一度に大量のデータを処理する必要がなく効率的にデータストリームを処理することができる．

データストリーム解析技術の産業への展開

多角的テンソル特徴抽出技術の車両走行データ解析への応用

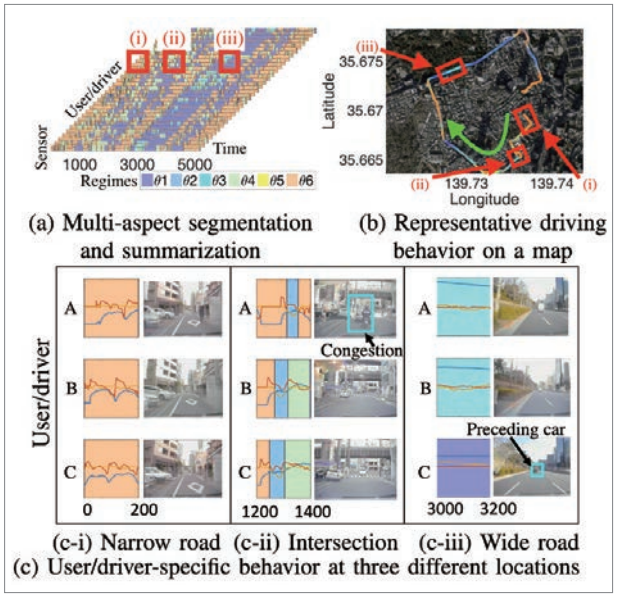
筆者らの研究グループではトヨタ自動車と連携し，多角的テンソル特徴抽出技術[6]を共同開発した．運転行動の予測に基づく走行支援サービスの実現に向けて，本プロジェクトでは各ロケーションにおけるドライバ共通の運転パターンやドライバの個人差を把握し，モデルとして学習／表現するための要素技術を開発することを目的としている．開発技術はデータに含まれる重要なパターンの時間遷移やユーザ特性を自動的に抽出する．センサ，ユーザ，時間など複数ドメインで構成される時系列テンソルデータから各ドメインにまたがる複雑な特徴を時系列パターンとして検出し，各パターンをモデルパラメータとして要約する．

多角的特徴抽出：実社会において収集されるビッグデータは，センサ，デバイス，ユーザ，時間など複数のドメインを持つ時系列テンソルデータとして表現でき，本技術は，時系列テンソルに含まれる複雑な特徴を多角的に解析し，パターンの時間遷移とドメイン間の個体差を同時に抽出する．図 -6 は車両走行データの解析結果である．図 -6 (a) は（センサ，ドライバー，時間）で構成される時系列テンソルからのパターン抽出結果であり，同色のセグメントが類似パターンのグループを表している．図 -6 (b) は，出力結果を実際に走ったコース上にプロットしたものであり，図 -6 (c) では詳細な出力結果を示している．本技術は，直進，右左折，徐行など車両走行における運転行動の時間遷移のみならず，ドライバごとの特性も同時に解析し，車両走行のさまざまな共通パターンを抽出すると同時に，車両走行のグループ化と，モデル化を完全自動で行う．

社会実装のための事業化に向けての取り組み

筆者らの研究グループでは，さまざまな企業と連携し，社会実装に向けて製造業 DX 技術の実用化に取り組んでいる．その中でも本稿ではソニーセミコンダクタマニュファクチャリングとの取り組みを紹介する．

半導体製造工程における設備故障予測：上述した要素技術を統合，発展させ，複合時系列データからイベント予測を行うための AI ソフトウェアを開発した．センサデータの潜在的な動的パターンを時系列モデルとして要約し，特徴量として抽出することにより，イベントの要因分析を行いながら長期先のイベントの発生確率を予測することを可能とする．開発したソフトウェアを活用して，CMOS イメージセンサの半導体製造工程における DRY 装置のターボ分子ポンプの故障予測に関する評価実験を行っている．DRY 装置のターボ分子ポンプの突発故障は製造ラインへの影響が大きく，また高額パーツのライフ適正化を進めることも可能となるため，故障予測のニーズはきわめて高い．



■図 -6　車両走行データに対する多角的テンソル特徴抽出技術の出力結果

開発ソフトウェアによって，DRY 装置のターボ分子ポンプ故障を事前に予測し，計画保全を実現する．

図 -7 は，コントロールユニットにおける管理データ（軸の正常位置からのブレ，電流値，回転数など）およびターボ分子ポンプの振動計測データを解析し，装置状態を推定した結果である．図 -7 の上段はオリジナルの入力情報となる 22 次元の設備管理および稼働データであり，最右端の時系列データの切れ目において故障が発生している．開発ソフトウェアは，センサデータに潜在する特徴的パターンやその変化点を時系列として捉え，装置状態を推定しており，図 -7 下段はその解析結果である．装置状態を正常（青），注意（オレンジ），水色（故障直前）に分類しており，水色のセグメント幅は約 13 日である．すなわち，開発したソフトウェアによって，約 13 日前に故障の兆候を検出している．評価実験では，故障データは 5 件しかないものの，5 件すべての故障の兆候を完璧に捉え，最短でも 8 日前に，最長で 15 日前に設備故障の予測に成功している．

革新的な AI 生産プラットフォームの開発に向けて

本稿では，最近のデータストリーム解析技術と製造業 DX への応用について述べた．日本において製造業は GDP の 22% を占め，今も日本経済を支える重要な産業である．そして，日本が組込・制御機器分野で強みがある点に鑑みれば，その技術や製品の価値を高めるために IoT ビッグデータをリアルタイムに解析する技術はきわめて重要となる．

今後の製造業においては，Beyond 5G/6G の技術を活用して設備を連携させ，消費者／利用者個々のニーズにフィットさせる多品種少量生産の取り組みも必要となるため，性能，機能，計算速度を劇的に向上させた AI ソフトウェアは，設計から製造，保守まで，製造システム全体を変革するキーテクノロジーになる可能性を秘めている．

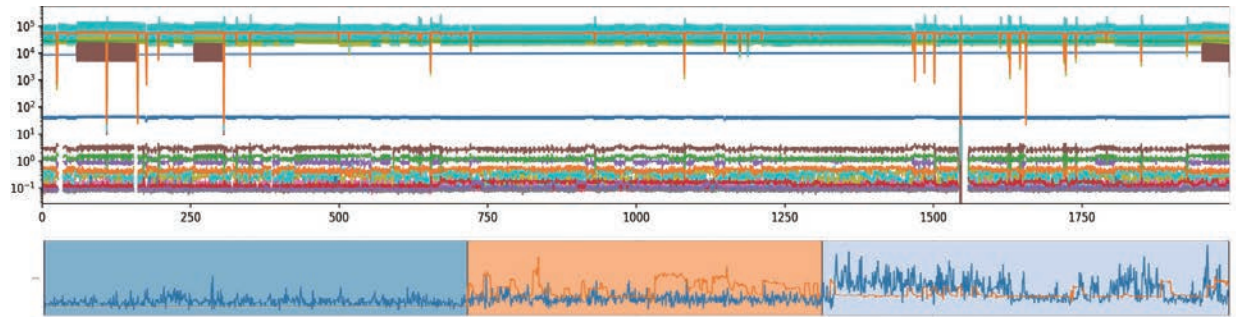
参考文献

1) Matsubara, Y. and Sakurai, Y. : Regime Shifts in Streams: Real-time Forecasting of Co-evolving Time Sequences, ACM SIGKDD Conference (KDD) , pp.1045-1054 (Aug. 2016).
2) Matsubara, Y., and Sakurai, Y. : Dynamic Modeling and Forecasting of Time-evolving Data Streams, ACM SIGKDD Conference (KDD) , pp.458-468 (Aug. 2019).
3) Kawabata, K., Matsubara, Y., Honda, T. and Sakurai, Y. : Non-Linear Mining of Social Activities in Tensor Streams, ACM SIGKDD Conference (KDD) , pp.2093-2102 (Aug. 2020).
4) Matsubara, Y., Sakurai, Y. and Faloutsos, C. :AutoPlait: Automatic Mining of Co-evolving Time Sequences, ACM SIGMOD Conference, pp.193-204 (June 2014).
5) Kawabata, K., Matsubara, Y. and Sakurai, Y. : Automatic Sequential Pattern Mining in Data Streams, ACM Int. Conf. Information and Knowledge Management (CIKM) , pp.1733-1742 (Nov. 2019).
6) Honda, T., Matsubara, Y., Neyama, R., Abe, M. and Sakurai, Y. : Multi-Aspect Mining of Complex Sensor Sequences, IEEE Int. Conf. on Data Mining (ICDM) , pp.299-308 (Nov. 2019).

（2021 年 12 月 5 日受付）

■櫻井保志（正会員） yasushi@sanken.osaka-u.ac.jp

1991 年同志社大学工学部卒業，同年 NTT 入社．1999 年奈良先端科学技術大学院大学博士課程修了．工学博士．NTT 研究所，熊本大学を経て，2019 年より現職．ACM KDD best paper awards（2008 年，2010 年）など受賞．AI・IoT データストリーム処理，Web や医療情報解析の研究に従事．



■図 -7 設備稼働データの解析と故障予測

特集
Special Feature

［スマートファクトリーは工場の何を変えるのか?］

③ IoT プラットフォームの現状と未来
―製造 DX の本質―

応用 一般

鈴木　聡　（株）NTT データ

製造業の課題と目指す方向性

現在の製造業にはモノの製造高度化だけではなく，人間の価値観・倫理観を持ち込んださまざまなステークホルダの期待に応えることが求められている．製造業を取り巻く課題は以下のようにさまざまである．

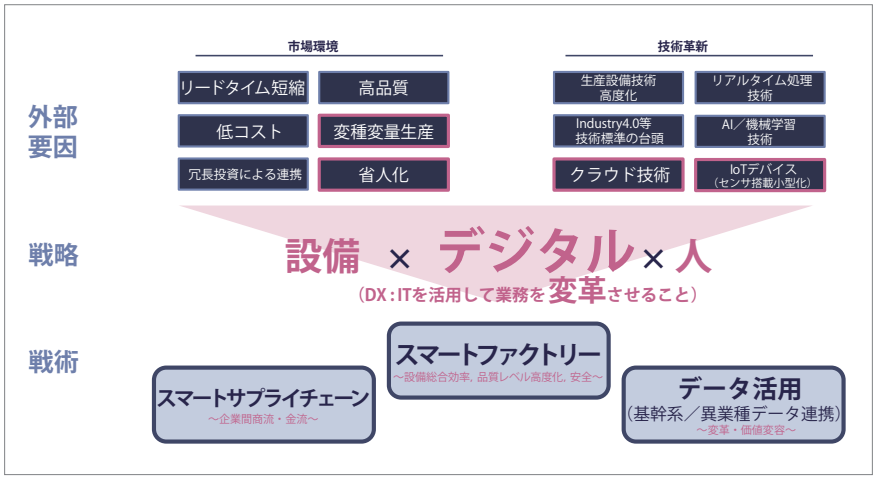
- 生産性向上，稼働率向上，製造自動化
- リードタイム短縮
- 人財不足，省人化，技能伝承，ノウハウ形式化
- モノづくり変革（マスカスタマイズ☆1，在庫レス）
- 災害リスク（COVID-19 等，供給網遮断のための在庫確保）
- 紙帳票の廃止，データ化
- 安全対策

さらに，SDGs（Sustainable Development Goals）に代表される社会貢献目標はこれまでの大量生産，売るために作る，という製造の在り方を見つめなおし，社会に対して何が還元できるか，地球環境に対して循環型で継続性のある事業ができているか，社員の自己実現を通じて活力から前向きな改善活動が行えているかが問われている．

企画，マーケティング領域に比べて製造領域はコストの対象となり，現状から削ぎ落していく，洗練していく変化を継続的に実施する傾向にある．その中で，品質が高いものを提供して当たり前とされ，日々品質管理レベルの向上を求められている．一方で，人財不足，担い手不足，技能伝承にかけられる時間は限られるといった制約の中で，徹底した省力化が求められている．一見矛盾が発生しているともとれる．

ここで期待されるのが，製造 DX（Digital transformation）である．図 -1 に示すようにデジタル技術を活用し，業務自体を変革し時代の変化に対応していく方向性（戦略）が求められる．デジタル技術は手段とし，業務自体を変革していくことが重要

☆1 大量生産に近い生産性を保ちつつ，顧客ごとのニーズに合う商品やサービスを提供すること



■図 -1　製造業を取り巻く状況

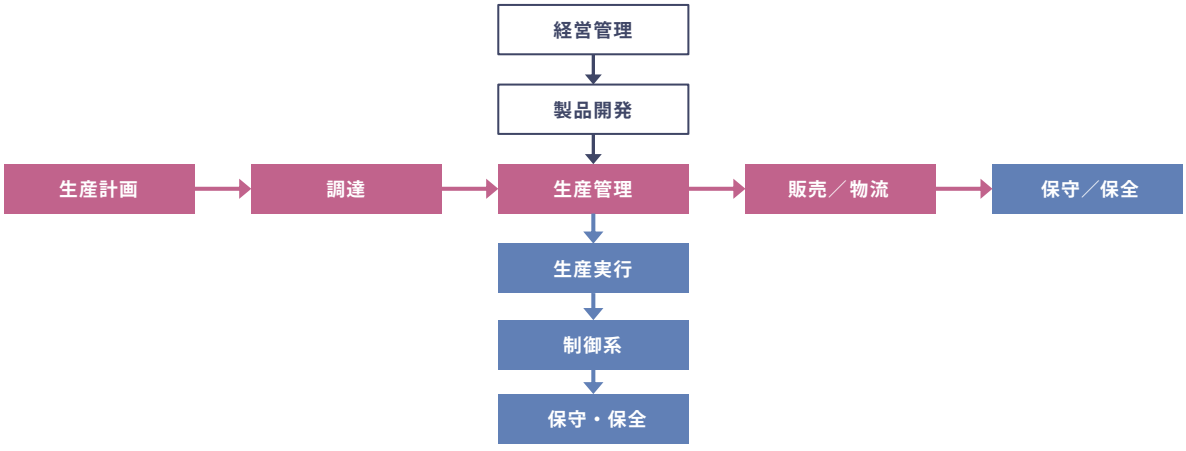
である．製造現場に従事する社員からさまざまな発想が生まれ，作るものは同じでも，作り手が状況に応じて変化を加え，品質は一定化させる．単なるコスト要求に応える部門ではなく，現場からさまざまな意見やアクションが生まれる風土を養うのである．そのためのビジョンを明確に発信することが肝要である．この手段（戦術）として，スマートファクトリーが注目されている．

IoT プラットフォームの位置づけ

生産の効率化だけに目を向けてしまうと計画が進まない可能性がある．なぜなら，日本の生産現場はこれまでも徹底した効率化，納期短縮の活動をしてきているためである．OT（Operational Technology, 制御機器を制御し運用するシステムや技術）は専門性が高く，時に企業ごと，ラインごとに個別に作られる．生産現場の可視化だけであれば SCADA（Supervisory Control And Data Acquisition, 産業制御をコンピュータによって統合的に監視，プロセス制御を行うシステム）に代表されるような生産設備と一体的なシステムによってほぼリアルタイム（マイクロ秒～ミリ秒）に状態を可視化できる．

近年導入が進んでいる IoT プラットフォームは OT の世界を置き換えるものではない．リアルタイム（ミリ秒以下）にデータを収集しエッジに設置されたゲートウェイを介して高速にデータをアップロードする．ゲートウェイではクラウドに上げるデータを選別してフィルタリングすることで通信量を低減しつつ，工場ノウハウを用いて意味解釈を加える．ただし，この場合，全体のコストは運用体制面を含めて上がることに留意が必要である．

工場設備は耐用年数が長く，事業の隆盛に応じて投資される．そのため，設備やライン単体で状態を見ることはできても，生産全体，工場全体，企業全体での状況のベンチマークは難しい．図 -2 にあるようにプラットフォームの適用範囲は企業の生産活動全般に目を向けるべきである．ビジョンは大きく持ち，IoT が中心となる生産実行領域だけでなく，SCM（Supply Chain Management）全体をプラットフォーム上で運営することを構想することが肝要である．これはデータの流通性としてのビジョンを持つという意味である．トレーサビリティにも寄与する．どこでどのようにして作られたものかが追えれば，顧客からの問合せに即座に応え，顧客満足度を維持するだけでなく，品質管理レベルを上げること



■図 -2　製造業におけるプラットフォーム

にもつながる．ビジョンは生産領域だけにとどめずに，設計へのフィードバック（部品改善等），物流との連携による在庫最適化，委託企業との部品・予備品コントロール，果ては金流ともつなげ設備投資に対する予算管理，検収といったステークホルダとの連携を可能にする．つまり，IoT プラットフォームとは，「データ連携性」である．

設備やライン単独で監視をして効率化や予防保全を行うものではなく，企業のサプライチェーン全体でデータ連携を実現し，データを流通させ，企業活動のさまざまなプロセスの中でデータを掛け合わせて新たな変化を発想していくことこそが製造 DX の本質である．

IT と OT の融合

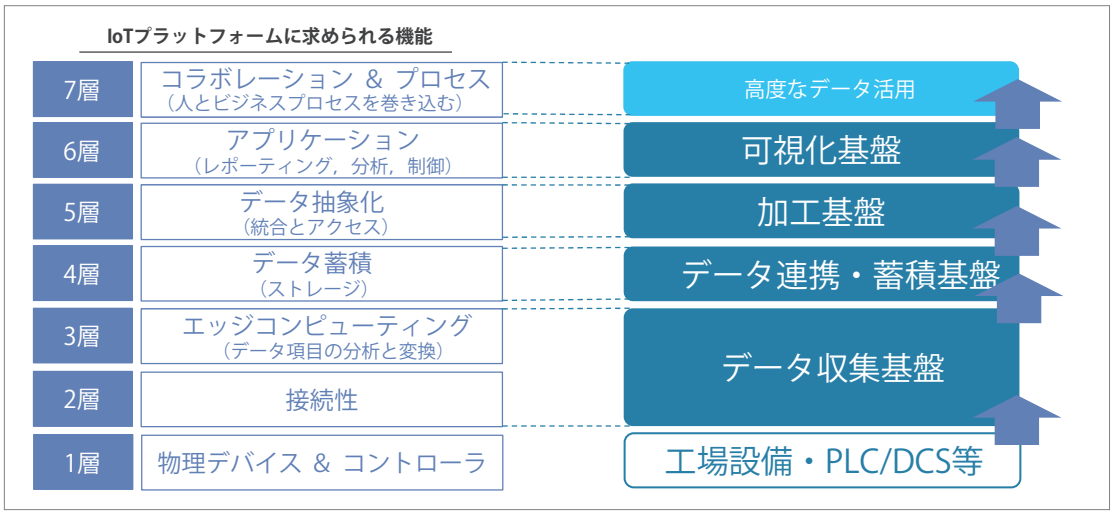
IoT プラットフォームの導入は工場現場と経営部門の価値観を合わせながら進めていく非常に難易度の高いプロジェクトになる傾向にある．

経営部門がトップダウンの指示を出しても，設計部門・研究開発部門から定められた品質基準を落とすわけにはいかない現場には響かない．現場も製造業務で繁忙，先を見るような余裕はなく，経営部門からの需給調整要請に日々追われ，中長期的な企業の課題に触れる機会が少ない．経営部門が導入を進め管理する IT と工場現場が運用する OT の融合を果たすものが IoT プラットフォームである．導入を通じた双方の理解，企業内でのデータリテラシーの均一化を目指すものである．

レイヤ化アーキテクチャ

デジタル技術は常に進化する．役割を明確にした階層構造を採用し，必要に応じて部分的に技術を置き換え可能な変化に強い構造にする必要がある．このためにマイクロサービス（アプリケーションをビジネスの機能に応じて複数の小さい疎結合なサービスの集合として構成するソフトウェア開発の手法），API（Application Programming Interface）の導入が有効である．また，将来 IOWN（Innovative Optical and Wireless Network, NTT が 2030 年の実現を目指す高速大容量通信，計算リソースの提供を行うネットワーク・情報処理基盤構想）のようなネットワーク技術の飛躍的な進歩が進めば，制御全体がクラウド側に集中するかもしれない．クラウドとエッジのトレンドは行き来する．この場合でも層の役割分担ができていれば対応が可能である．

図 -3 にクラウドをベースとする IoT プラットフォームのレイヤ化アーキテクチャを示す．図の



■図 -3　レイヤ化アーキテクチャ

左は IoT World Forum が共同で作成した IoT プラットフォームの構成要素を示した IoT Reference Model[☆2] である．図の右は筆者がそれと工場基盤の対応付けを行ったものである．

データ収集基盤，データ連携・蓄積基盤は設備のネットワークへの接続，紙帳票の電子化をはじめとした物理世界をデータ化するレイヤである．設備のネットワークへの接続は一般的に Ethernet を介して行うことになるため，ポートを保有しない旧機種の設備は PLC（Programmable Logic Controller, 工場等の自動機械の制御に使用されるプログラミング可能な制御装置）等の設備制御装置を介して接続することになる．

加工基盤はデータを蓄積するだけでなく人が使える形に加工するためのレイヤである．データ分析の基本は「比較（前日, 前週, 前月, 前年等）」,「変化（時系列）」,「分類構成」である．事業として重視する事柄に対して課題になっている個所を洗い出し，可視化の方向性を導き出す．

可視化基盤は可視化によって業務プロセスに対し気づきを経て次のアクションを起こすためのレイヤである．基本的にはデータから予測を立てていくことになる．画像や動画の活用は，一般的に機械学習による物体認識で行われる．専用設備（外観検査機等）の導入に際しては，品質管理レベルを落とさずに置き換え可能でコストメリットがあるかといった比較が行われる．

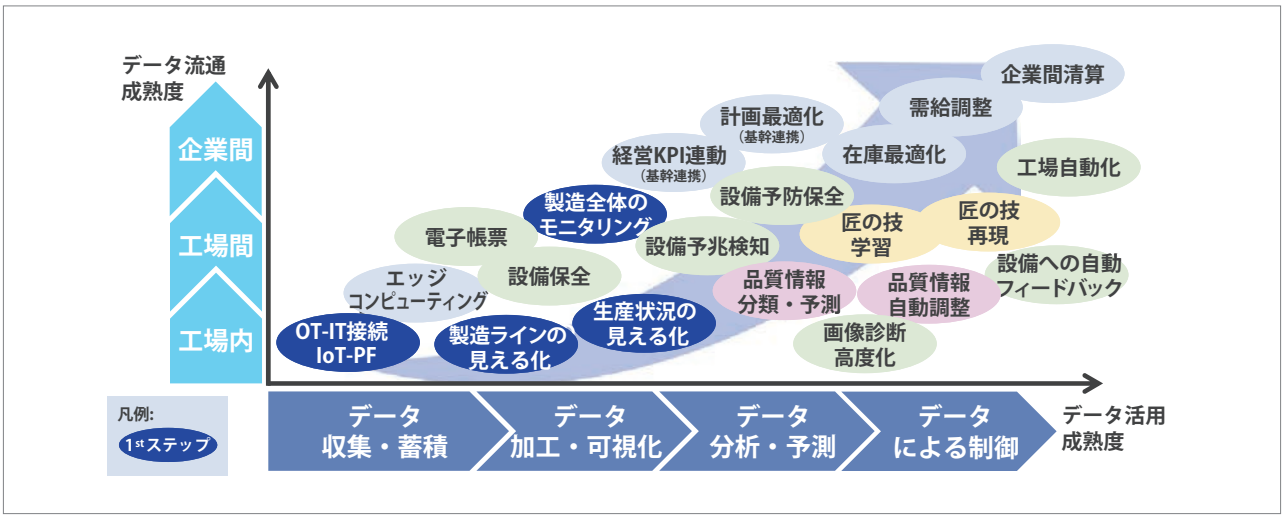
高度なデータ活用は，設備の自動制御のレイヤである．導入には運用による実証が必要となるケースが多い．最終的な人の判断を伴ってオペレーションされなければ大規模な機会損失につながる可能性があるためである．

図 -4 にデータ活用のロードマップを示す．これまでに説明したレイヤをデータ活用の成熟度として横軸に，工場内，工場間，企業間といったデータ流通の成熟度を縦軸に取った．データ活用の高度化は工場内の工程間にとどまらず，工場間での情報連携（品質管理，保全技術継承，在庫最適化等），果ては企業間での情報連携（企業間精算等）に向かっていくことが望まれる．

技術的アプローチ

すべての工場のすべてのデータを集約して統合的に管理したいという要望が強い場合は，データを高効率にクラウド側に送信することが求められ

☆2 出典 IoT World Forum Reference
http://cdn.iotwf.com/resources/72/IoT_Reference_Model_04_June_2014.pdf



■図 -4　データ活用ロードマップ

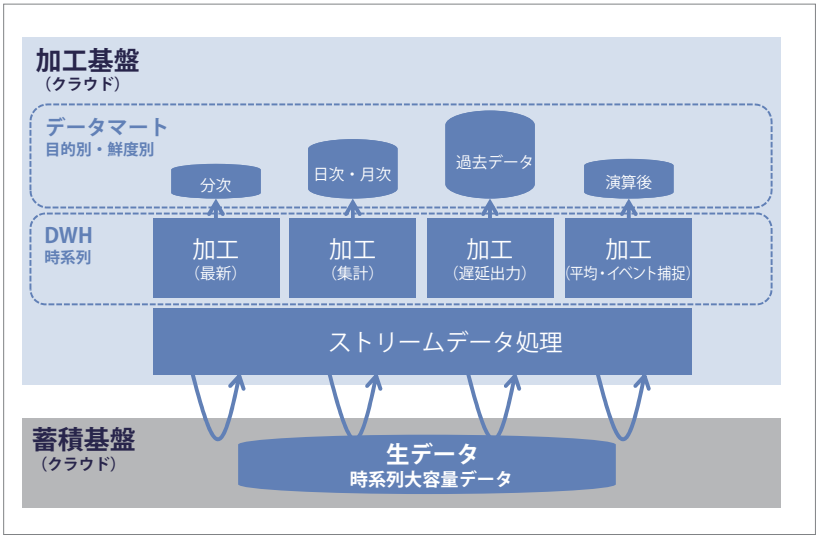
る．日々大量に発生するデータを常に保管し，常時5年～10年のスパンで経年変化や比較を行いたいとなるとクラウドを選択肢にせざるを得ない．工場内のデータは秘匿性が高く，工場外にセキュリティ観点で出せないという課題に直面するケースもある．工場からクラウドへの接続はインターネット経由ではなく，専用線を通じて企業内の集約データセンタに収容し，そこからパブリッククラウド事業者への接続等が選択肢になる．

組み立て加工業で，加工精度を上げるためにミリ秒のアナログ変化を捉えたい，1秒未満でアラーム発生と同時に通知を出したいといった場合，エッジでの処理が選択肢になる．エッジではストリームデータ処理を行い，すべてのデータを残すのではなく，変化点（イベント）のみを切り取る方法となる．ストリームデータ処理は大量にセンサから発生するデータを人間が認識できる，もしくは関心を抱くイベントとしてデータ化する処理である．特徴づけの方法にはいくつかあるが，絶対値をある断面で最新値として切り取る，単位時間当たりの平均値，積算値，ビット信号に応じてイベントを補足して値を導出するといった方法が考えられる．階層化されたデータの考え方は人間が限られた時間で洞察を得るためには必要なアーキテクチャである．図-5にそのシステム構成例を示す．

エッジにミドルウェア等を導入するケースが多く工場設備の癖や現場プロセスの特徴に合わせて収集するイベントを選別できるというメリットがある．クラウド側で加工を行う場合，ストリーム処理やETL/ELT処理 （Extract（抽出）/Transform（変換）/Load（書き出し），Extract（抽出）/Load（書き出し）/Transform（変換），データ加工処理手法）のロジックを組み込む必要がある．

エッジ側機能の特徴としては，機械学習による推論結果を設備にフィードバックすることが考えられる．学習用の大量のデータはクラウド側に送信し，モデルを作成，その作成されたモデルをエッジにデプロイ（配備）し，推論（予測）をエッジで行うことでリアルタイムな制御が可能となる．

さらに，生データは永続保管され，上記の分析の結果，より詳細なデータ（秒間データ等）を解析したい，機械学習のための学習データを得たいとなったときにいつでも取り出せる必要がある．ポイントとなってくるのはいかに取り出しやすいかという点である．



■図-5　ストリームデータ処理

可視化パフォーマンス

KGI（Key Goal Indicator，重要目標達成指標）に対して，KPI（Key Performance Indicator，重要業績評価指標）を設定し，それを可視化することが求められる．KPIを見ただけで，なぜKGIが達成できないのか，その間にはCSF（Critical Success Factor，重要成功要因）を手掛かりにし仮説に従ってデータを解析し，アクションを取っていくことになる．この仮説は基本的に業務上の経験や過去の蓄積されたノウハウが手掛かりになる．すべてのデータを可視化すれば相関が見えてくるといった手法はシステム処理能

力的にも，人間が理解し得るといった情報としても無理がある．

基本的に工程の流れによって製品が製造されていくため，時系列でのデータを確認することになる．データ分析はトライアンドエラーを行うことになる．

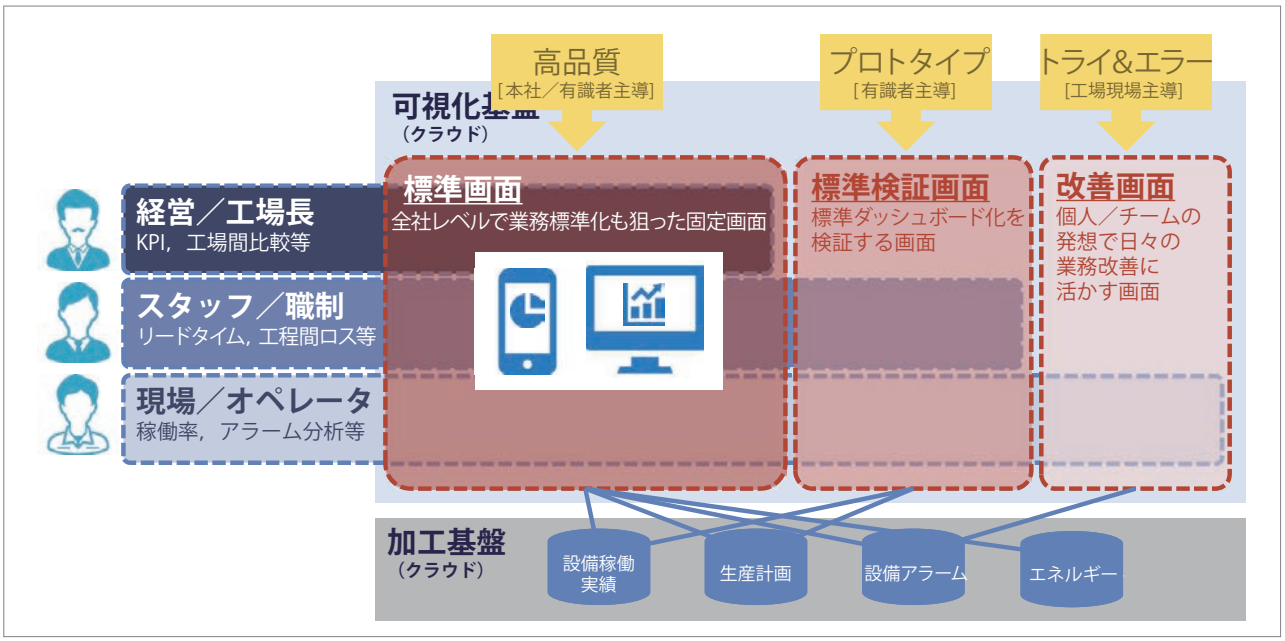
図 -6 は「誰が見るか」を縦軸に，「どのような柔軟性で見るか」を横軸にステークホルダが見るダッシュボード画面の分類を示したものである．「誰が」「なぜ」そのデータを見る必要があり，どのような頻度，抽象度で見たいデータかによって公開範囲も変わってくることに留意する必要がある．

製造 DX の先に

IoT プラットフォームは製造業の業務に変革をもたらすために，データを活用して新しい価値を見出すきっかけの手段にすぎない．今まで見えていなかったデータを見えるようにしてアクションにつなげる，データとデータを掛け合わせて新しい業務プロセスを生み出す，データがつながることで顧客に新しい価値を提供する．経営高度化に資する目標，工場間連携による最適化目標，製造現場の品質／生産性／リードタイムに資する目標，あらゆる場面で活用が可能である．必ず目標，仮説を立てることが必要である．必ずしも正解が用意されているわけではない．デジタル技術を活用して，「変革」をしていくことこそが製造 DX の本質であると考える．日本の製造業の強みをデジタルで補完し，新たな強みを生み出す一助になることを願う．

（2021 年 10 月 29 日受付）

■鈴木　聡　Satoshi.Suzuki@nttdata.com
シニア IT アーキテクト．（株）NTT データの提供する DX プラットフォーム iQuattro® を活用し製造業を中心に顧客の DX を進める責任者を務める．



■図 -6　可視化パフォーマンス

特集
Special Feature

［スマートファクトリーは工場の何を変えるのか?］

④ スマートファクトリーを支えるローカル 5G
―導入に向けた制度や技術，留意点の理解―

柿元宏晃　NTT コミュニケーションズ株式会社

ローカル 5G を理解する

昨今，高信頼なプライベートワイヤレスネットワークの 1 つであるローカル 5G は，製造業の生産現場における多量なデータの収集や，自動搬送車（AGV : Automatic Guided Vehicle）をはじめとしたロボットの群制御等，スマートファクトリーを実現するためのテクノロジーとして，注目を集めている．2019 年度の制度施行から現在にかけて，ローカル 5G を用いた実証実験の事例が多数公表され，さまざまな企業や公共団体において実フィールドでの取り組みが盛んになってきた（図 -1）．

一方，ローカル 5G の導入を検討していく際にはいくつかの障壁がある．技術面やコスト面はもとより，ローカル 5G はライセンスバンド（免許取得が必要な無線）であることから，制度面も複合的に理解しておく必要がある．そういった現状において，少しでも多くの方にローカル 5G の技術的特徴や，その可能性について興味・関心・理解を深めていただくため，本稿では NTT コミュニケーションズ株式会社がこれまで検証や導入の提案等，実プロジェクトを通じて得た，ローカル 5G の免許制度やかかわる主要技術，ローカル 5G を現場に導入するにあたっての進め方，コストの考え方等，留意するべき事項を紹介したい．

■図 -1　ローカル 5G を用いた検証の様子

ローカル 5G は，日本における呼称であり，5G 技術をプライベートネットワークとして使うための技術・制度のことを指す．海外では Private 5G と呼ばれることが多い．国によっては同様の制度がなく携帯キャリアの 5G のみが提供される場合や，使用できる周波数や免許が与えられる対象が日本と異なる場合がある．本稿で取り扱う内容は日本における制度にフォーカスする．

制度の理解

免許制度

日本では，2019 年 12 月と 2020 年 12 月にローカル 5G にかかわる制度が段階的に施行された．ローカル 5G を使用するためには無線局免許が必要となり，無線局免許は，特定の場所と特定の用途において特定の機器から無線を出力する許可を得るためのものとなる．使用場所，目的，無線機の数量・

電波の出力等を明確にし，設計を行った上で初めて免許の申請を行うことができる．無線局免許は人や企業に対して包括的に与えられるものではないため，「今後どこかで使えるようにとりあえず免許を取っておく」ということはできない．

弊社が取得した無線局免許の 1 つを例に挙げる（図 -2）．無線局（電波を出すもの）の単位で免許が与えられる．免許を取得すると，許可された場所（設置場所・移動範囲）と用途（目的），機器（識別番号／出力等）が総務省によって公表される．情報は一部マスクされた状態で公開される．図 -2 の例では設置場所が「東京都港区」となっているが，免許の申請書には地図上で詳細に位置や範囲を示し，電波伝搬のシミュレーションをした上で，提出する必要がある．

免許人の氏名又は名称	エヌ・ティ・ティ・コミュニケーションズ株式会社		
免許人の住所	＊＊＊＊＊		
無線局の種別	基地局	免許の番号	＊＊＊＊＊
免許の年月日	令3.4.1	免許の有効期間	令7.5.31まで
無線局の目的	一般業務用		運用許容時間 常　時
通信事項	一般業務用通信に関する事項		
通信の相手方	免許人所属の陸上移動局		
識別信号	＊＊＊＊＊		
無線設備の設置場所又は移動範囲			
東京都港区			
電波の型式、周波数及び空中線電力			
99M9X7W　4849.98 MHz　1.2 W			

■図 -2　総務省による無線局免許の公表例

ミリ波とサブ 6

ローカル 5G には，28GHz（ギガヘルツ）帯から 28.2 ～ 29.1GHz（ミリ波と呼ばれる），4.7GHz 帯から 4.6 ～ 4.9GHz（サブ 6 と呼ばれる）が割り当てられた（図 -3）．帯域幅で比べると 28GHz 帯は 900MHz（メガヘルツ），4.7GHz 帯は 300MHz と差がある．無線通信では使用できる帯域幅と通信スピードは比例し，周波数が高いほど帯域幅を確保しやすい．同じローカル 5G でも 28GHz 帯は帯域幅を大きく確保できているため，4.7GHz 帯に比べ通信速度を重視するような用途でメリットを出しやすい．2021 年 11 月現在では，28GHz 帯の 900MHz 幅のうち 400MHz，4.7GHz 帯の 300MHz 幅のうち 100MHz に対応したシステムを手に入れることができる．

一方，周波数が高いことで，遮蔽物に対する透過や回析は期待できず，到達距離も短いといった特性があるため，基地局のアンテナと受信端末の間に遮るものがない理想的な無線環境に近い見通し内

■図 -3　ローカル 5G の周波数帯（サブ 6 とミリ波）

（LOS : Line Of Sight）でのスポット的な使用が主な用途と見込まれる．4.7GHz 帯は，28GHz 帯よりも遠くまで電波を飛ばしやすく，遮蔽物に対してもある程度の透過や回析をするため，28GHz 帯での置局設計よりは工場の敷地内など一定のエリアをできるだけ少ない基地局でカバーすることができる．

通信スピードや伝搬以外に，2 つの周波数帯には，免許制度上の制約の違いがある．28GHz 帯は衛星通信と，4.7GHz 帯は公共無線と隣接する周波数が含まれている．28GHz 帯のうち一部の周波数では衛星通信と干渉する可能性があり，4.7GHz 帯の一部周波数では，所要電力によって一部エリアでの設置が許可されないケースがある．前述の電波の特性に加えて，制度上の制約も考慮して使用する周波数を選択していく必要がある．詳細は総務省が公表する「ローカル 5G 導入に関するガイドライン」を参照されたい．

技術的特徴の理解

高信頼の無線通信技術

ローカル 5G の技術的特徴として，高速大容量（eMBB : enhanced Mobile Broadband），高信頼低遅延（URLLC : Ultra-Reliable and Low Latency Communications），多数同時接続（mMTC : massive Machine Type Communication）が一般的に挙げられる．これら 3 点の特徴はシステムへの実装において発展途上にあり，実環境では他の通信規格にパフォーマンスが劣るケースも実際にある．ただ，そういった状況の中でも，ローカル 5G が必要となるケースがある．現時点で実導入を検討する際に特に重視すべきポイントとして，安定性・信頼性にかかわる特徴を追加で 3 点紹介したい．1 点目は免許制による安定性である．「制度の理解」で述べたように，免許なくしてローカル 5G の周波数帯を使用することはできないため，免許不要の Wi-Fi で起こるようなアクセスポイント乱立による干渉が発生しづらい．2 点目は移動する物体への通信の安定性である．5G はモバイル機器向けに開発された通信規格であるため，通信対象がアンテナ間を移りゆくようなケースにおいても途切れず，通信を継続できる特徴を持つ．3 点目は強力な認証の機構である．ローカル 5G においては，公衆無線網と同様に SIM（Subscriber Identity Module：利用者を特定する情報が格納されたモジュール）を用いて通信の認証を行うため，なりすましや不正アクセスに強い．これら，ローカル 5G の安定性・信頼性を支える 3 点の特徴が，「現在のローカル 5G」を選択する上での重要なポイントとなる．

高信頼かつ低遅延で通信ができることは，工場などを持つ企業にとって魅力的である．工場の生産ラインのネットワークは，遅延に対して非常にシビアで，高い信頼性が求められる．パケットの到達順まで指定されるケースもある．こうした条件を Wi-Fi で満たすことは難しいため，有線 LAN が用いられているが，有線の場合は生産ラインを柔軟にレイアウト変更できないため，無線化したいという声は多い．信頼性の高い無線通信が可能なローカル 5G は，特定の領域で Wi-Fi や有線 LAN の代替として期待されている．

また，ローカル 5G のメリットは，利用者の用途に合わせて柔軟にプライベートネットワークをカスタマイズできる点もある．たとえば，利用者が低遅延の無線通信を求めている場合，要件に合わせてローカル 5G 環境をチューニングすることができる．4K 映像や LiDAR（Light Detection and Ranging：光を用いたセンシング技術）データ等，大容量データの収集を行うために利用するのであれば，アップリンクを優先して通信の帯域を制御する．時間帯やサービスごとにデータの行先をコントロールするといった仕組みも実現可能．プライベートなネットワーク環境であれば，その時々の最新機能を取り入れることができる．

利用者の中で通信が完結できる点もローカル 5G の大きなメリットの 1 つである．各種機器のコン

トロールを行う場面において，通信速度・遅延時間の改善やセキュリティ対策として，外部のネットワークを経由させずに，ローカルネットワーク内に閉じて通信を行いたいという声が，工場を持つ企業から多くあがる．

MEC とスライシング

企業ネットワーク全体から見れば，ローカル 5G での通信は一区間にすぎない．ローカル 5G の特性を活かすためにはエンド・ツー・エンドでネットワークの構成・品質を考える必要がある．ここでは，MEC（Multi-Access Edge Computing）とネットワークスライシングという 2 つのアプローチを紹介する．MEC はクラウドよりも手前にあるネットワーク上でデータの処理を行うエッジコンピューティング技術のことである．ローカル 5G を用いて収集したデータの処理を行う際，データ処理にかかる時間やセキュリティに関してシビアな性能が要求されるケースがある．データを処理する MEC を工場内等，オンプレミス（施設の構内）に配置して，データの収集から処理までのネットワーク上の距離を近くすることで，通信にかかる遅延時間や，リスクにさらされる区間を短くできる．さらに，高度な構成として，MEC をオンプレミスだけでなく，ネットワークエッジにも配置して，クラウドと合わせて，多段にコンピューティングリソースを配置して使い分ける方法もある．MEC を上手く活用すれば，低遅延を実現しながらクラウドへ送信するデータ量を削減できるほか，災害などによってクラウドやインターネットにトラブルが発生しても処理を継続し可用性を高めることができる．

ネットワークスライシングは，仮想的にネットワークを分割する技術のことである．物理的には 1 つしかないネットワークシステムを分割し，複数の仮想的な領域として，多様なユースケースに応じた通信を実現させることができる．たとえば，高速大容量を優先するネットワークと，超低遅延を優先するネットワークを使い分けるといったことが可能になる．製造現場では製造にかかわる制御信号や，従業員のコミュニケーション，監視カメラの映像等，ネットワークに求められる要件はさまざまな通信が混在する．これらを適切に選り分けることで，限られた無線リソースを効率的に使用することができる．

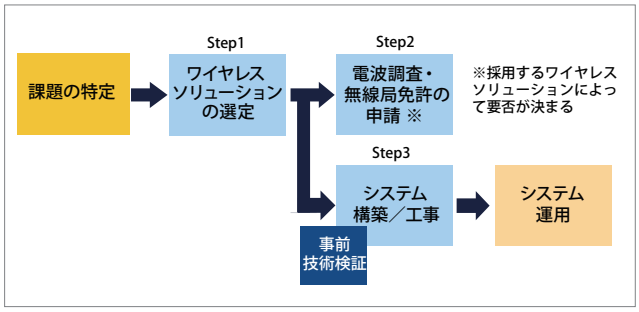
ただし，ここで一点留意が必要である．上記の現時点のローカル 5G にかかる製品では上記の特徴すべてが実装されているわけではない．5G 技術の最新技術の標準仕様については，3GPP（移動通信システムの仕様標準化プロジェクト）にて策定されているが，製品への反映と改善は，日進月歩で進行中であり，5G の技術標準動向と，メーカの製品対応状況については留意する必要がある．5G 電波を出力する基地局システムと，ユースケースへの適用に直接的に影響する各種デバイス（ローカル 5G 受信端末等）はメーカが異なる場合もあるため，両面での確認が必要となる．

導入の進め方

導入プロセス

本章では導入にあたってのプロセスについて紹介する．ワイヤレスネットワークには，ローカル 5G のほかにも Wi-Fi やプライベート LTE 等，多くの選択肢がある．それぞれ特徴が異なるため，ユースケースに合わせて，適切なワイヤレスネットワークシステムの選定が必要となる（図 -4）．

Step1 では，ローカル 5G や Wi-Fi 等，ワイヤ



■図 -4　ワイヤレスネットワークシステムの導入プロセス

レスソリューションの中で，ユースケースを実現する上で，最適なテクノロジーを選択する．ローカル5G においては前述の技術面・制度面を理解した上で，ユースケースの実現において真にローカル 5G が必要となるか見極める．Step2 では，無線局免許の取得に向けて，事前に現場周辺の電波状況の確認や，必要書類などを準備し，申請手続きを行うこととなる．最後に Step3 で，現場へのローカル 5G システムの構築・工事を行い，無線システムとデバイスやアプリケーションを含めたシステム接続試験を行うことで，ユースケースを実現し，システム運用を開始する．

なお，現在のローカル 5G の導入事例においては，本格導入の前に事前実証を実施するケースが多く，まずは小さく早く実証を始めることで，現場のユースケースへの事前技術検証や効果測定を行うのが一般的である．

以降，順に各 Step について解説する．

ワイヤレスソリューションの選定

まず，現場導入を行うにあたっては，解決したい課題やユースケースに合った技術の選定が重要である．5G のような最新技術も万能ではなく，ユースケースに適合していなければ現場で採用されることはない．表 -1 に，製造現場で活用されるワイヤレスネットワークの例とローカル 5G の比較について紹介する．通信速度に加えて，信頼性や，通信先（移動体かどうか）が選択のポイントとなる．

ローカル 5G の特徴である高信頼性を活かしたユースケースとして，製造現場における生産ラインの制御機器，生産を支えるロボット（自動搬送車等）の無線制御が挙げられる．また，高速大容量の特徴を生かした映像伝送もローカル 5G のユースケースとして期待が高い領域の 1 つである．実際に，これまで行われてきたローカル 5G の実証事例や導入事例として，ロボット等移動体から映像伝送等の大容量データの送信を行うユースケースが多い．

ワイヤレスソリューションを選定するにあたっては，最終的にユースケース全体を通して，ワイヤレスネットワークで収集した生産現場データを，活用用途やセキュリティの観点から，生産現場やクラウド環境等のどこに蓄積・配置するのかを並行して全体設計することも肝要である．

電波調査・無線局免許の申請

ローカル 5G の導入を決定した後は，無線局免許取得の手続きを始める（図 -5）．ローカル 5G システムの機器選定から免許の申請や構築までの期間の中で，大きなウェイトを占めているのは免許にかか

■表 -1　製造現場で活用されるワイヤレスネットワークとローカル 5G の比較

	Wireless HART	ISA100.11a	Wi-Fi	ローカル5G
標準化規格	IEEE802.15.4	IEEE802.15.4	IEEE802.11	3GPP
使用周波数	2.4GHz(ISM)帯	2.4GHz(ISM)帯	2.4(ISM)/5GHz帯	4.7/28GHz帯
免許	免許不要（共用）	免許不要（共用）	免許不要（共用）	免許要（占有）
トポロジ	メッシュ	スター／メッシュ／ハイブリッド	スター	スター
通信速度	250Kbps	250Kbps	～9.6Gbps ※Wi-Fi6	～10Gbps*1
概要	HART*2を無線化した工業用無線通信規格．膨大な対応デバイスが特徴であるが，TCP/IPとの互換性がない．	工業用無線通信規格．国際計測制御学会(ISA)が主導，規格化．ISA100 WCIが相互接続性を確保．	無線LANのこと，一般的な無線通信規格として普及．Wi-Fi認証により相互接続性を確保．	第5世代移動通信システム「5G」を自営無線として構築・運用．移動通信を実現し，周波数を占有して運用可能．

*1：5G通信規格としての通信速度であり，現行ローカル5Gの通信速度はDL:数百Mbps/UL:数十Mbps程度．
*2：Highway Addressable Remote Transducerの略．プラント制御大手のmerson Electric(米)が開発．

わる工程であり，期間短縮に向けては，必要な手続きをあらかじめ理解しておくことが重要である．

免許申請に向けてまずは，電波利用目的を固め，目的に合わせた免許種別（実験試験局免許・実用局免許）の選択，具体的なシステム運用場所・機器構成の決定，周辺事業者との干渉調整を行っていく．検証等の一時的な利用となる場合，免許種別としては実験試験局免許での対応となる．商用利用の場合は実用局免許を選択する．導入場所や機器構成等が決まったタイミングで，総務省の地方支分部局である各地方の総合通信局に事前の相談を行う．相談の際に，導入予定場所において，同一周波数や，隣接する周波数帯において無線干渉が起こる可能性のある事業者を通知され，免許申請前に，各事業者と，予定している構成・出力で，干渉等の影響がないか，調整を行う．電波干渉を与えるような場合には干渉調整等を設計において考慮しなければならない．

免許の申請手続きや，事前に行う事業者との調整に必要なシステム構成の検討や電波設計においては，特別な技能を必要とすることから，システムインテグレータ等が提供する支援サービスの利用等を検討するのもよい．

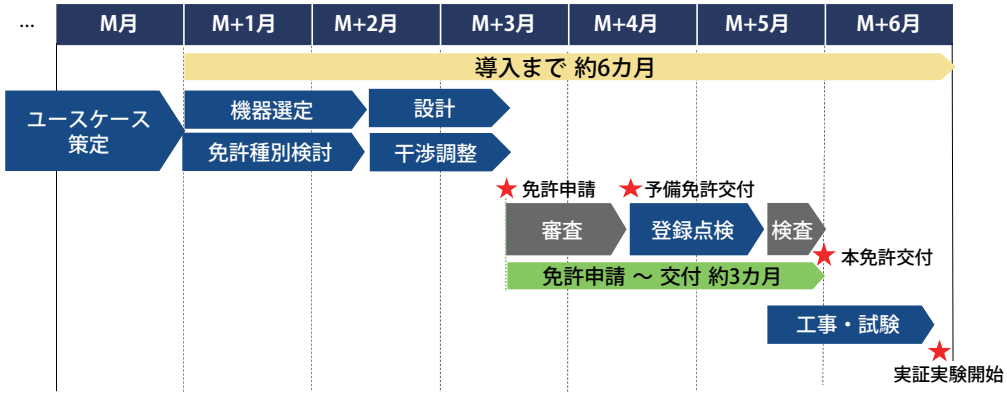
システム構築・工事の実施

電波調査・無線局免許の申請のパートでも触れたが，システム構成や電波設計の検討については免許申請の段階までに実施しておく．免許取得のタイミングを考慮し，実際の無線通信テスト等を行う必要はあるが，システム構築については，通常のネットワーク構築工事と基本的には同様である．ただし，機器メーカをはじめ，ローカル 5G の各種機器の取り扱いについては，現時点では導入実績が少ない状況でもあるので，サポート等の体制については，事前に十分に確認をしておいた方がよい．

コスト面の課題

ここまで，ローカル 5G の導入における技術的特徴や免許制度の観点で話を進めてきたが，現時点で導入にあたっての最大の課題となっているのが導入コストである．ローカル 5G の導入に必要となるコストとは，主に，初期費としてローカル 5G の機器（基地局装置，受信デバイス）にかかる費用，免許申請にかかる費用（代行委託含む），構築・試験にかかる費用があり，システム稼働後のランニング費として，システム保守・運用，電波利用料にかかる費用が挙げられる．また，ユースケースによって，エッジコンピュータやクラウド等のシステム基盤費用や，クラウドサービス利用にかかる費用も考慮する必要がある．

もちろん，今後数年の技術普及により，機器，役



■図 -5　ローカル 5G 現場導入のスケジュール例

務を含めたコスト低廉化には期待するが，従来の Wi-Fi と比較すると，現時点ではまだコストとして高額であり，単純に Wi-Fi 代替のネットワークとして考えると費用対効果の面でも導入には高い障壁となっているのが実情である．ただ，ローカル 5G の技術的特性はスマートファクトリーを実現する上で，非常な強力な武器となる．将来性に着目し，その特性を使いきるような形で，現場で複数のユースケースを適用できるような使い方を検討し，取り組み実績を積み重ねていくことが肝要だと考える．また，プライベートネットワークを構築した方がよいのか，公衆網（いわゆるパブリック 5G）で対応するのかといった点も留意が必要であり，携帯電話事業者の 5G 設備の展開状況や，サービス内容も注視した上での判断が必要となる．

導入に向けた取り組み

現在，ローカル 5G については初期費の低廉化を目的に，一部，基地局システムを共同利用するような形態でのローカル 5G サービスも提供が進んでいる．現場導入にあたっては，こういったサービス活用の検討も一考の価値があると思われる．また事前検証等においては，実現場での実証をシミュレーションできるように，機器メーカやシステムインテグレータによっては検証環境も準備されている（図 -6）．そういった環境を上手く活用し，ローカル 5G の電波特性や，デバイス機器類に触れてみるとよい．

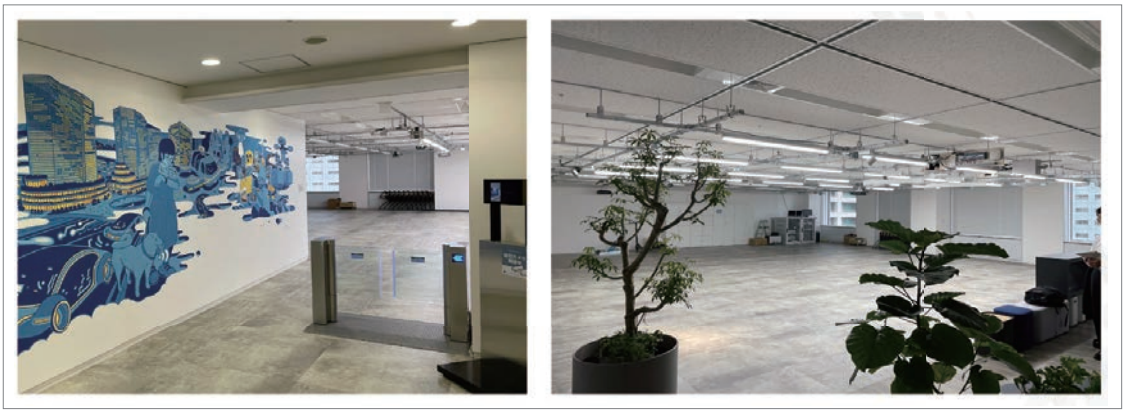
まずは，機器メーカや，ソリューション提供会社などのパートナ会社と共創を行うような協力関係を構築し，技術やコストの見極めを行い，ローカル 5G のみならず，他の無線ソリューションの技術動向や，製品・サービス動向に着目し，課題の解決に向けた仮説を立て，実際の技術に触れる事前検証等を通じ，現場の未来に繋がるユースケース作りのチャレンジを一緒に進めていくことが重要である．

本稿では，ローカル 5G の導入に向けて必要な知識として，技術・制度・導入プロセスや留意事項について解説してきた．現在，ローカル 5G を活用した新しい業務形態の実現に取り組んでいる企業が多く，特に製造業は，先進テクノロジーの活用を先導する業界として，他の業界からも注目を集めている．ローカル 5G という新たな技術が，生産現場の変革の一助となることに大きく期待したい．

（2021 年 11 月 22 日受付）

■柿元宏晃

2010 年 NTT コミュニケーションズ（株）に入社．2019 年の制度化前からローカル 5G のソリューション企画や実証実験に従事．現在は，企業へのローカル 5G の導入支援や，ローカル 5G の特徴を活かす組合せソリューションの開発を進めている．



■図 -6　ローカル 5G が利用できるラボ環境の例

特集
Special Feature

［スマートファクトリーは工場の何を変えるのか?］

⑤ 持続可能な社会に向けた今後の生産システムと産業基盤

―「人」が主役となるものづくり革新推進（HCMI）コンソーシアム 2050 年に向けたロードマップから―

応般

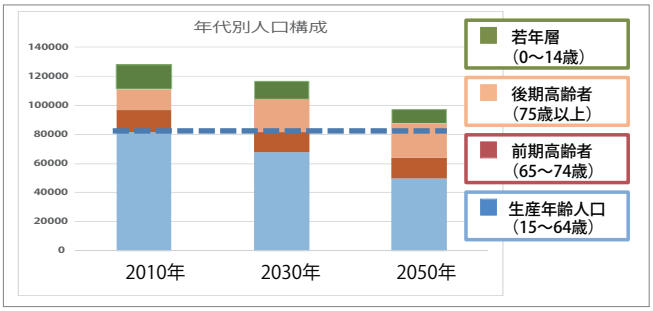
岩井匡代　谷川民生

国立研究開発法人　産業技術総合研究所　HCMI コンソーシアム

「人」の活躍が持続可能な社会の鍵

「人」が主役となるものづくり革新推進（HCMI）コンソーシアムは，2019 年 4 月 10 日に産業技術総合研究所 臨海副都心センター内に事務局を設置して発足した．産業競争力懇談会（COCN）で産学官が集結し，次世代のものづくりと豊かな未来に向けた産業基盤の課題について協議した結果，「人」は労働力だけでなく，消費社会を形成する重要な財産であり，「人」の活躍が今後の産業基盤の進化に不可欠であるとし，「人が主役となる新たなものづくりの確立」を提言し，その実現に向けた産学官協働プラットフォームとして発足した．

当コンソーシアムでは，2020 年度，2050 年の社会・産業の環境を分析した結果を踏まえて，2030 年までに目指すべき姿の確立に向けたロードマップを策定した．本稿ではこの活動を通じて検討した持続可能な社会に向けた今後の生産システムと産業基盤について紹介する．



■図 -1　日本の人口と年齢構成推移 [1]

未来の産業基盤の在り方考察

産業基盤を揺るがす社会課題

昨今，消費者のニーズが「モノ」から「コト」に移行し，この傾向は国内だけでなくインバウンド市場にも強まっている．産業動向だけでなく，COVID-19 により，在宅勤務などを余儀なくされるなど，社会課題が産業基盤に大きく影響している．

生産年齢人口減が産業・経済の減退を招く

我が国は世界に先立ち，超高齢化社会を迎え，生産年齢人口は半減する（図 -1 参照）．2030 年にはこの傾向が世界中に拡大する．すでに深刻な人手不足が，「人手不足倒産」（後継者不足，従業員退職，求人難，人件費高騰が原因の倒産）が増加傾向となり，事業継続問題になっている．

同時に，消費層を担う重要な就労世代が縮小することにより，国内市場の消費経済の大幅な縮退を招く．さらに，市場があるところでものを作る地産地消の傾向は，消費経済の縮退に連動してものづくり産業需要も縮退する．このままでは国内産業の危機につながる．

人口偏在による市場の不確定さの加速

従来，消費動向は，「消費の重心」（人口×経済力が大きい国，地域）の需要動向で予測してきた．しかし，将来の人口動向分析では，現在経済力がある欧米，日本，中国などの国は大幅に生産年齢人口が減少し，人口増の８割はアフリカなど消費社会が未成熟な国である（図 -2 参照）．そのため「消費の重心」が不安定になり，消費動向の予測は困難になり不確定さが加速すると予想される．

もちろん消費社会が未成熟な国においても，生産年齢人口増による，経済成長期待が高まっているが，地球環境の容量を超えた資源の消費が見込まれ，社会の持続可能性にも大きな課題がある．

すなわち，消費動向はもはや予想するのではなく，想定して，環境を整え創造する時代になると言える．

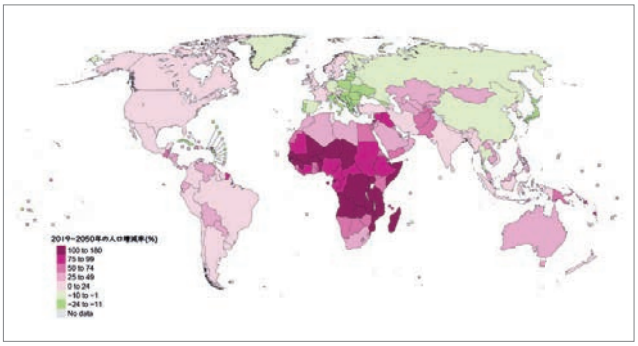
自然環境の変化が及ぼす社会持続性課題

地球温暖化の影響により，気温の上昇が５～６℃まで達すると異常気象の頻度や強度が高まり，海面上昇による居住地域の制限が出てくるなど，自然災害リスクが高まる．感染症の媒介動物の生態地域にも影響し，感染症リスクも高まることが確実である．

もはや，緊急事態対策では対処できないレベルになり，生活や事業を継続できる対処が重要課題となる（表 -1 参照）．

また，枯渇性資源の逼迫問題は，単に材料調達が困難であるというレベルではなく，もはや消費主導経済（大量消費，大量破棄）の限界を示している．この対策は企業の枠を超え，重要な産業課題である．

一方，生活様式や経済の大きな変化を加速する側面もある．COVID-19 では緊急事態対応である在宅勤務が通常時にも広がり，国内外のデジタル社会の急激な進化を促した．また，資源枯渇についてはサーキュラーエコノミー（循環経済）という新たな経済潮流を生み，欧州を中心に国策として進められている取り組みの中には，単にリスクに対する防御的施策だけでなく，新たな価値創造や，新たな雇用創造につながるとして，積極的に国策として環境と経済成長の両立に向けた政策が展開されている．国内でも，経済産業省が循環経済 2020 として政策展開を強化している．



■図 -2　世界人口変動マップ [2]

産業の持続的発展に向けた課題

当コンソーシアムでは，上述の社会課題が及ぼす産業基盤への影響を踏まえ，「人」の活躍を経済成長に効率的に転換すること，自然災害・感染症リスク発生時に生活や事業を継続しやすいこと，さらに質的な豊かさを提供できることを目標として設定し，産業の持続的発展に向けた課題を分析した．

生産人口増と労働生産性向上

「モノ」から「コト」へ移行し，市場ニーズの多様化

■表 -1　温暖化による環境変化と健康影響 [3]

温暖化の健康影響		
	温暖化による環境変化	人の健康への影響
直接影響	暑熱、熱波の増加	熱中症、死亡率の変化（循環器系、呼吸器系疾患）
	異常気象の頻度、強度の変化	障害、死亡の増加
間接影響	媒介動物等の生息域、活動の拡大	動物媒介性感染症（マラリア、デング熱など）の増加
	水、食物を介する伝染性媒体の拡大	下痢や他の感染症の増加
	海面上昇による人口移動や社会インフラ被害	障害や各種感染症リスクの増大
	大気汚染との複合影響	喘息、アレルギー疾患の増加

に対応するため，生産方式は多品種少量生産からさらに変化に追随する変種変量生産への転換が急がれる．

これまで生産性を牽引してきた機械中心の自動化は，開発期間が追いつかず，投資効果を得にくくなり，人件費支出の方が効率的という結果に陥りやすい状態になる．一方，人依存が大きくなると人手不足の影響をダイレクトに受けるため，求人対策や労働生産性の維持向上対策が重要課題になる．

人口偏在の緩和と生活・事業継続

自然災害や感染症対策として，人口偏在の緩和や，少ない移動で日常生活が維持できる自立した地域環境づくりと，自立した地域同士の連携が有効であり，これはサスティナブル社会の実現にも貢献する．

また，豊かな未来に向け，人々の生活も地域に密着しながらグローバルなつながりが得られるなど開かれていることがリスクの緩和適応と産業基盤の進化を促すと考える．そのためには時間・空間の制約が少ない自分らしい生き方（働き方）を実現する働く環境革新が求められる．

生涯能力向上の期待を持つ豊かな働き方

就労人口の拡大を豊かな消費社会形成に発展させるためには，単に労働の場の提供にとどまらず，生涯能力向上の期待を持った豊かな働き方の実現が不可欠である．また在宅勤務などの定常化から，現在の労働集約型を前提として充実してきた労働安全の仕組みを進化させ，個々人の状況に応じて能力を発揮しやすい状態を確保する新たなマネジメントシステムが必要であると考える．

Society5.0 時代のものづくり —SDGs 目標 8「働きがいも経済成長も」

上記の考察を踏まえ，当コンソーシアムでは図 -3 に示す，活動の狙いを定めた．

ものづくりの進化と働く環境革新

誰もが無理なく働くことができ，自分らしく住みたいところに住み，働きたいところで働ける，時間・空間の制約が少ない働く環境の整備に向け，これまでの機械中心で，機械化が及ばないところを人がサポートするというのではなく，人が主体性を持って働けることに集中して技術習得や成長への希望が持てるように，人と機械が柔軟に役割を調整しながら働く新たな人と機械の協調型協働システム基盤の構築に取り組む．さらに時間・空間を超えてつながる遠隔協調型協働へ進化させ，働く場所にかかわらず協調できるものづくり IoA（Internet of Ability）の構築が必要であると考える．その前提として生産システムを Human-in-the-Loop に進化させる必要がある．

労働の質に着目した新しい働き方

産業の持続的発展と豊かな未来に向けた「人」の活躍が，日本における Society 5.0 時代のものづくりの鍵になるという考えのもと，高い労働生産性や新たな価値創造には「働きがい」が非常に重要であると認識している．そこで，労働の質を Quality of Work ではなく，労働者の就労生活の質や能力の向上と組織価値（労働生産性）の双方に着目した，Quality of Working（以降の QoW は，

■図 -3　2030 年に向けた HCMI コンソーシアムの活動の狙い[4]

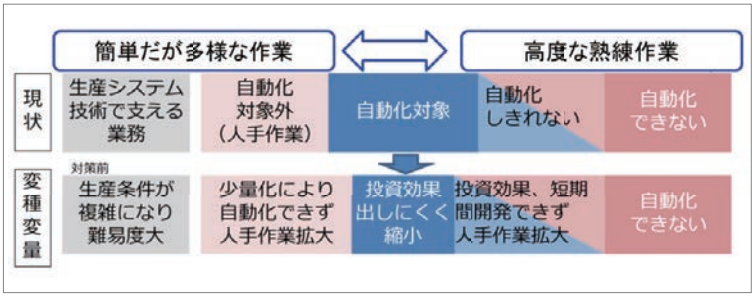
Quality of Working の略称）の確保・向上を実現するための新たな働き方に対応したマネジメントの革新が不可欠と考えている．就労人口拡大に向け，多様な人材が働きやすく，就労者の労働寿命を延伸し，生涯能力向上の期待を持って働けるように各人の状態に応じたきめ細やかなセルフマネジメントと組織マネジメントを実現するためのマルチタレントマネジメントの手法確立とオンジョブでシステムインするための技術確立に取り組む必要があると考えている．

Human-in-the-Loop のつながるものづくり

人と機械の協調システム基盤

生産分野において，人手でなければならない作業として大きく 2 種類考えられる．

図 -4 における左側が，簡単ではあるが，多様な動作や手順を必要とする作業，右側が，作業自体が高度であり，主に高度な熟練技術を要する作業である．従来の大量生産方式であれば，設備投資をしても十分投資回収が可能であったが，作業工程が頻繁に変わる変種変量生産においては，ロボットによるモーションの再設定や生産ラインの改変等，自動化におけるコストが増大し，費用対効果について人件費のほうが良い結果となっている．しかし，労働生産性や人手不足の影響を考慮すると，AI 等を活用し，自動化技術をより高度化することが必要であると考えられる．特に，簡単だが多様な作業においては，今までのようにロボット技術と AI 技術を活用し，ロボットの知能化を向上し，多様な作業に適用できるようにすることは重要ではあるが，すべての作業において，完全に自動化することは困難であると考えられ，多様な作業に対応できるという人の柔軟性に依存する部分は残らざるを得ない．すなわち，1 人の作業者をロボットで置き換えるという考えではなく，2 人の作業者のうち，1 人をロボットに置き換え，人とロボットが協調作業することで，1 人あたりの生産性を向上するという考え方で進めることが現実的と考えられる．現在でも，ロボットメーカ各社から協働ロボットというコンセプトの下，人と同じ空間で作業できる安全性の高いロボットが販売されている．しかし，基本的には，安全柵等の物理的な作業空間を分離することなく安全センサを備えたロボットであるだけで，労働安全衛生法上は，ロボットが作業者にツールを手渡しするようなインタラクションの強い協調作業をすることは認められていない．そこで重要な技術は，人の作業動作の認識技術の高度化である．協働ロボット等の安全を担保する機能安全の考え方においては，人間のモデルは不確定なものとしてリスクアセスメントされていた．よりインタラクションの強い協調作業をさせるには，作業者個々のパーソナライズされたモデル化およびその作業者の行動推定を行い，タイミング等を予測する技術が必要となる．パーソナライズされた作業者の予測動作通りに作業者が動いている間は，安全率よりも効率性を上げ，モデルと食い違う動作になれば，安全率を上げていくといった，人の動作予測制御と組み合わせて，安全性と効率性を両立した人・機械協調制御技術を開発していくことが求められる．



■図 -4　変種変量生産におけるものづくりの変化

さらに，昨今の COVID-19 の緊急対策として，在宅勤務という業務が浸透してきている．会議や PC を活用したオフィスワークについては，作業場所によらない業務形態が取り入れられてきているが，工場や物流等のモノを扱う物理作業においては，現場での作業の必要性は変わらない．しかし，ロボットの遠隔操作技術を活用することで，物理

作業においても，在宅勤務のように，現場に依存しない就労が可能であることが期待できる．これは，人と機械の協調という「1 人あたりの生産性向上」と場所に依存しない就労の実現による「働きやすさ向上による潜在的労働者の活用」の効果が期待でき，今後の就労人口の低下の大きな解決策として期待できる．

以上のように，「AI によるロボットの知能化」から，作業者の動作予測技術を活用した「人と機械（ロボット）の協調技術」，さらに遠隔操作技術を活用した「時空間を超えた人とロボットの協調技術」へと展開していくことが必要であると考えて，当コンソーシアムでは図 -5 のロードマップを策定した．

ものづくり IoA システム基盤

IoA（Internet of Abilities）とは，「能力のネットワーク」を意味する概念であり，人がネットワークを介して他の人間や機械と知覚および動作を共有することで，意識や能力を拡張するものとされる．この IoA の概念を技能継承に適用し，熟練技術者の広範な技能や知識をネットワークを介して集約し，集合知化して展開することによって人材育成やロボット活用を推進するものづくり IoA システム基盤を，我々は提案している．

ものづくり IoA システム基盤実現のための技術開発は，以下の段階を経て実施することを当コンソーシアムのロードマップを策定した（図 -6 参照）．

(1) AI を活用した技能転写技術

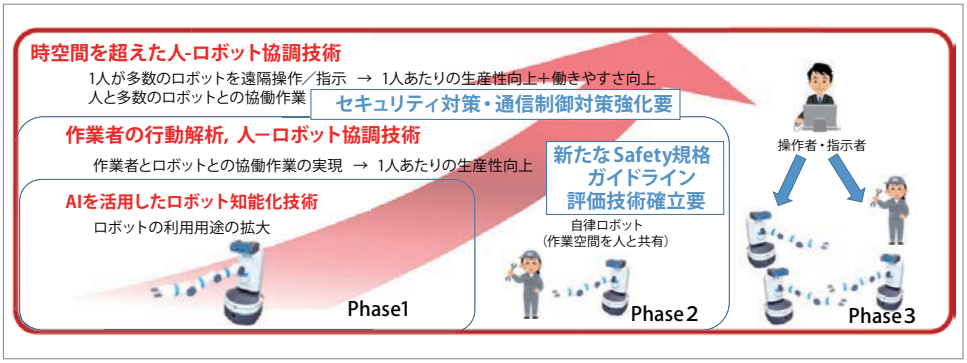
熟練作業は，定まった条件の下で行われる定型的熟練作業と，変化する状況に応じた柔軟な判断が求められる適応的熟練作業に分類される．ここでは，定型的熟練作業にロボットを適用するための技能転写技術を開発する．

(2) 熟練知識の構造化と高度化プロセスの定式化

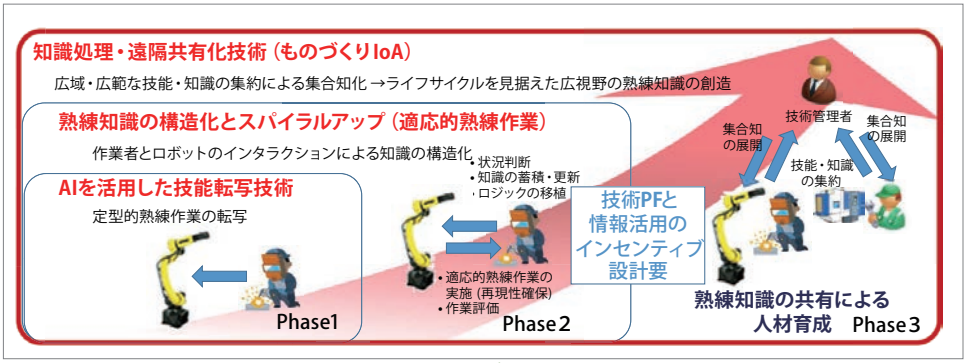
ここでは，適応的熟練作業を対象として，作業者とロボットのインタラクションによる知識の構造化を進める．適応的熟練作業における技能継承の難しさの 1 つに，作業の再現性が挙げられる．ロボットに熟練者の知識や判断ロジックを転写し，再現性の確保された動作を実施・評価することで，知識やロジックの更新を行い，知識の構造化を進める．そして，この過程を技能の高度化プロセスとして定式化する．

(3) 知識処理および遠隔共有化技術

(1) と(2) の成果をベースとして，ネットワークを介した広域・広範な技能・知識の集約による集合知化を行う．そして，それまで散在していた知識を結び付けることによって，ライフサイクルを見据えた広視野の熟練知識の創造を図る．



■図 -5　協調型協働から遠隔協調型協働へのロードマップ[5]



■図 -6　ものづくり IoA 基盤構築に向けたロードマップ[5]

特集
Special Feature

QoW マルチタレントマネジメントシステム

当コンソーシアムでは，労働の質とは，健康（労働寿命の延伸），働きやすさ（労働環境），働きがい（成長意欲）を指標として構成されていると定義し，労働の質向上のためマルチタレントな多様な人材の「労働生活の Well-being」「多様な働き方でも主体的に労働参加」「能力を発揮しやすい状態の維持向上」を各人の状態に応じて支援することが必要であると考えている．このため，チャレンジしたい労働に必要な，労働生活の能力の見える化に向けた「労働生活能力分類の策定」の検討，寝込むほどの健康悪化の前の中間的な虚弱状態（フレイル）を予知し，早期に適切な改善を行い健康を維持するためのフレイル予知予防，つらさの度合い測定，組織との相互信頼に着目した QoW 指標の検討やセンシング方法，改善フィードバック方法の開発，物理的な身体モデルを簡易的な計測で評価できる人モデル基盤の開発，生産システムと人モデルを融合するコンセプトの立案に取り組んでいる．パーソナルマネジメント手法の確立と並行して，ものづくり IoA 型マルチタレントマネジメントシステム手法の確立を進めることにより，NEXT 健康経営として QoW 経営の社会実装を目指している．QoW を定量的かつ客観的に見るための，3K（きつい，汚い，危険）に代表される働きやすさを阻害する作業のつらさの状態を示す指標の検討やフレイルの予知予防の視点で見る健康指標の検討，またこれらの計測方法の検討は，これらの活動における基盤的な取り組みであり，まさに今取り組むべき課題であると認識している（図 -7 参照）．

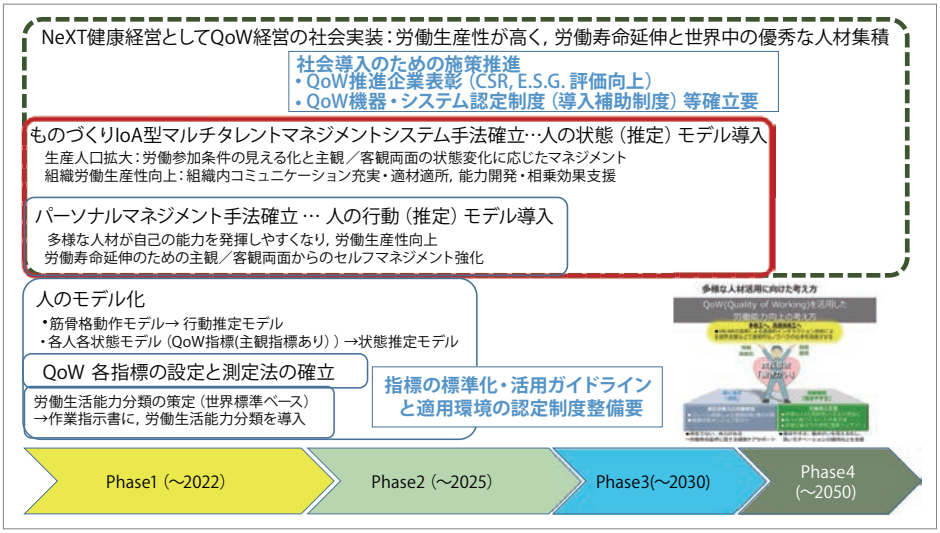
2050 年に向けた人が主役となる循環経済のものづくり

欧州の CE（サーキュラーエコノミー）政策に代表されるように，現在の使用済み製品から材料を再生する資源循環から，2050 年にはシェアリング経済，製品のサービス化が進展するとともに，より高付加価値な資源循環メカニズムが構築されると予想される．

本コンソーシアムでは，産業横断的に資源循環を情報管理する 21 世紀政策研究所が発行した「欧州 CE が目指すもの」に定義されている「循環プロバイダー[6]」の登場により起こり得る社会変化として，以下の 3 つのシナリオを想定した．

- 消費者が主体となるネットワーク型の循環経済へ移行する．地域市場ニーズに柔軟に対応した地域内資源循環の最大化（資源循環のローカライズ）．
- 物流合理化による余剰部材削減，資源投入量削減．
- 製品ライフサイクルを通じたデータの利活用（標準化／流通）により，資源循環の効率化を進め，産業の育成，発展に寄与．

以上のシナリオの実現に必要な技術や制度をバックキャストし，人と資源循環のかかわりについて具体的な移行シナリオの検討を進めるとともに，上述の働く環境革新やマネジメント革新を製品の再利用を担う産業へも適用することで，製品開発産業と再利用を担う産業を融合し，産業横通しの循環メカニズムを構築する



■図 -7　QoW 経営に向けたロードマップ[5]

準備を早々に始める必要があると考えている．

モノ・コトづくりの両輪で活躍する産業基盤へ

社会課題や自然環境が及ぼす産業基盤への影響が強くなる中，これからの産業競争力は単に技術力の強化や事業力の強化だけでなく，同時に産業基盤の根幹である消費社会を形成する就労人口の拡充と活躍を促す対策を行うことが後世に豊かな未来を残しているために必要不可欠である．

そのため，生産システムに人モデルなどを導入し，Human-in-the-Loop に進化させ，協調型協働基盤，遠隔協調型協働基盤，ものづくり IoA 基盤の形成と労働の質（QoW）に着目したマネジメントシステムの社会実装が必要であると考えている．

産業基盤としても，コトづくりの進展に向けより消費地ベース地域密着型と，基盤製品・部品については，極力量を集めて生産効率を上げ，市場価格を安定させるべく，部材生産密着型の産業クラスタ構築もやはり必要になり二極化が進むと考える．この間をつなぐ広域物流や資源循環などに新たな産業や基盤づくりの可能性があると考えるが，上記の基盤がこの可能性を広げることに貢献できると考えており，2050 年に向け，日本はコトづくり・モノづくりの両輪で世界のニーズに応える新たな産業構造と DX 革新の実現により，災害時の生活・事業継続にも貢献する，日本型の循環経済の新たな産業モデル構築に発展させたいと考えている（図 -8 参照）．

■図 -8　HCMI コンソーシアム活動の狙い（2050 年）[5]

なお，本 HCMI コンソーシアムのロードマップは，運営委員会（産業技術総合研究所　加納誠介氏，増井慶次郎氏，澤田浩之氏他）を中心に会員と協議して作成し，各界を代表するアドバイザー[5]にご指導をいただき作成した．

参考文献

1）https://www.stat.go.jp/data/jinsui/2.html#series，総務省統計局データから編集．
2）世界人口統計 2019 年版データブックレット．
3）環境省地球温暖化に関わる影響に関する懇談，https://www.env.go.jp/earth/ondanka/pamph_infection/full.pdf
4）NEDO モノづくり日本会議「TSC Foresight」オンラインセミナー「産業の持続的発展と豊かな未来にむけて」講演資料，https://www.nedo.go.jp/content/100937372.pdf
5）「人」が主役となるものづくり革新推進コンソーシアム活動概要　2021 年 6 月 10 日版，https://www.hcmi.cons.aist.go.jp/document/introduction_pdf.pdf
6）欧州 CE 政策が目指すもの～ Circular Economy がビジネスを変える～，21 世紀政策研究所 報告書，2019 年 3 月，http://www.21ppi.org/pdf/thesis/190405.pdf

（2021 年 11 月 25 日受付）

■岩井匡代（正会員）　iwai-msy.hcmi@aist.go.jp

国立研究開発法人　産業技術総合研究所　HCMI コンソーシアム事務局長 兼 三菱電機先端技術総合研究所開発戦略部連携推進 G 担当部長．2016 ～ 2017 年産業競争力懇談会「人」が主役となる新たなものづくりプロジェクトの事務局代表として提言書策定取り組み，その提言を実現すべく，2019 年より事務局長に就任し現在に至る．三菱電機では産学官連携推進の担当部長として組織的連携の推進に取り組んでいる．

■谷川民生　tamio.tanikawa@aist.go.jp

国立研究開発法人　産業技術総合研究所　HCMI コンソーシアム運営委員長 兼 産業技術総合研究所情報・人間工学領域インダストリアル CPS 研究センター研究センター長．マニピュレーション技術，空間型ロボット，Society 5.0 等社会デザインの研究を進めてきた．現在，労働生産人口低下における社会課題の解決として CPS を基盤とした人機械協調技術の研究を進めている．

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

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，ニューノー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ないのか，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て書きやすい言語で，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 創立者．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）**kazuoishii2014@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

[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

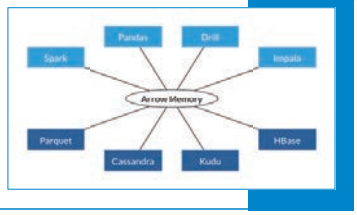
1 Apache Arrow による Ruby のデータ処理対応の可能性

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．

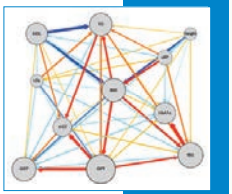
本稿では，Ruby の Apache Arrow への対応がこの弱点の解消に有効であることを示す．



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科）・**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，実体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．



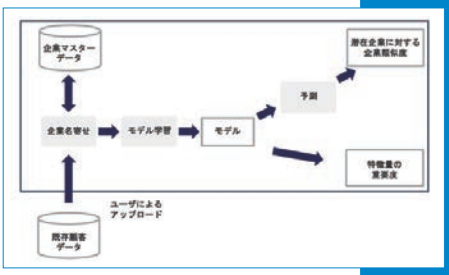
本編はこちら ▸▸

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) ニューズピックス）・
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 Busuness）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．

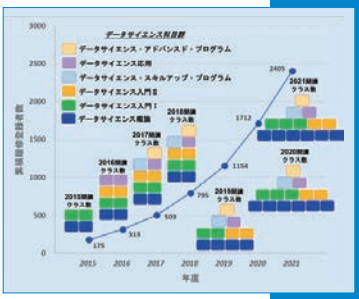
本稿においては，ABM をソフトウェアによって実践する 1 つの方法，および筆者らによる実装と現状の課題について紹介し，その課題を解決するための研究結果を報告する．



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） https://www.ipsj.or.jp/dp/
	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） https://www.ipsj.or.jp/dp/DPreport/index.html

2022 年度会誌「情報処理」モニタ募集のお知らせ

会誌編集委員会

会誌「情報処理」をより良くするために編集委員一同努力を続けておりますが，会員の方々の評価や希望をうかがい，今後の改善に役立てるために，モニタ制度を設けております．関心のある方はぜひふるってご応募ください．

応募の資格　本会会員で，モニタの役割を積極的に果たしていただける方．

モニタの役割　学会 Web ページ（https://www.ipsj.or.jp/magazine/enquete.html）から，毎月アンケートに回答する．
◇記事に対する評価　◇記事に対する感想　◇意見　◇記事テーマの提案
◇そのほか全般的な意見・提案など
注）記事をすべて読むといったことは必ずしも必要ではありません．自分の立場や問題意識，得意とする分野などを基準とした「独断と偏見による」自由な意見を期待します．

期　　間　原則として 1 年間（2022 年 4 月～ 2023 年 3 月）．**＊最長 3 年までとします．**

対 象 号　会誌「情報処理」63 巻 5 号～ 64 巻 4 号

謝　　礼　貴重なご意見をいただいた方には，モニタ任期終了後薄謝または記念品を贈呈します．

募集人員　特に定めませんが，応募者数によっては当委員会で調整させていただくことがあります．

応募締切　**2022 年 2 月 25 日（金）必着**

そ の 他　ジュニア会員で，会誌（冊子体）の送付を希望される方には，モニタ期間中会誌を送付いたします．
（**先着 50 名**，アンケート（12 回）に必ず回答いただくことを条件とします）
希望する場合は，申込書の要望欄に＜会誌送付希望＞とお書きください．

申　　込　以下 Web ページ内＜ 2022 年度 会誌「情報処理」モニタ申込フォーム＞よりお申し込みください．
https://www.ipsj.or.jp/magazine/topics/2022monitor.html

照 会 先　情報処理学会　会誌編集部門（モニタ係）　E-mail : editj@ipsj.or.jp

連載

情報の授業をしよう！

本コーナー「情報の授業をしよう！」は，小学校や中学校で情報活用能力を育む内容を授業で教えている先生，高校で情報科を教えている先生や，大学初年次で情報科目を教えている先生が，「自分はこの内容はこういう風に教えている」というノウハウを紹介するものです．情報のさまざまな内容について，他人にどうやって分かってもらうか，という工夫やアイディアは，読者の皆様にもきっと役立つことと思います．そして「自分も教え方の工夫を紹介したい」と思われた場合は，こちらにご連絡ください．
（E-mail：editj@ipsj.or.jp）

高等学校（工業）でのスマートフォンを利用したデータ活用の授業

基般

岸本有生　大阪電気通信大学高等学校

工業科でのデータサイエンス教育とプログラミング教育の必要性

2022 年度の新学習指導要領「情報 I」では，問題解決を学習目標として，基礎的なプログラミングや，データの特徴を分析するデータサイエンスが注目されている．筆者は高等学校の工業科の中で，ゲームプログラミングのコースを担当している．普通科では「情報 I」でプログラミングが必履修になったが，工業科では主に C 言語を利用した制御プログラミングを学習する．その中で，データサイエンスを取り入れつつセンサ計測と結びつける授業を考え，2021 年 9 月に 6 コマの授業を実施した．

実施した授業の目的は，「統計の必要性を学ぶ」「結果から意味を考察する」「公開されているオープンデータの利用方法を学ぶ」「センサの計測を利用する」である．特に統計的な計算だけで終わらずに，説得力のある答えを導き出せる考察力を身につけてもらうことに注視した．本稿では，実践した授業について紹介する．

■表 -1　授業内容

時限	内容	学習環境
1，2	データ分析と統計の基礎	Connect DB
3	オープンデータの利用	
4，5	スマートフォンのセンサを使用した行動の分析	
6	スマートフォンのセンサを使用したゲーム制作	Bit Arrow

授業全体の構成

本授業は，工業科 3 年生のゲームプログラミング基礎の授業に対して行った．生徒人数は 38 名である．内容は，表 -1 の通りである．1 コマ 50 分で，図 -1 のような本校のコンピュータルームで行った．学習

■図 -1　授業の様子

環境は，データ分析学習環境 Connect DB[1] とプログラミング学習を支援する実行環境 Bit Arrow[2] を使用した．Connect DB は，手計算では不可能なビッグデータの統計処理が行え，操作が簡素であることが特徴である．図 -2 には，それぞれの学習環境の Web サイトを表示している．まず，1，2 限目に統計の必要性を学ぶことにした．ここでは，Connect DB にあらかじめ用意されているサンプルデータを利用する実習形式とした．次に，3 限目にオープンデータの利用方法を学習した．ここでは，寝屋川市の公衆トイレの位置を表示させた．4，5 限目は，スマートフォンに内蔵されているセンサを計測した．ここでは，加速度センサを使用して，人の動作を分析した．最後に Bit Arrow を利用してスマートフォンのセンサを使用したゲームを制作した．

データ分析と統計の基礎

度数分布を用いた定性データの分析

1 限目の授業では定性データの分析を扱った．サンプルデータから「購入履歴」を選択すると，架空のスーパーマーケットの購入履歴が表示される．購入履歴には，図 -3 のように購入者の年齢（「10 代」「20 代」「30 代」「40 代」）と，購入した商品として「肉」「魚」「お菓子」のデータが含まれている．

生徒たちは Connect DB を使い，「購入したもの」を度数分布として集計した．結果として，図 -4 のような購入個数が表示された．続いて，集計した結果を図 -5 や図 -6 のようにグラフで視覚化した．棒グラフからは「肉」が「お菓子」よりも 75 個多く購入されているのが分かり，円グラフからは「肉」が全体の 44% を占めていることが分かることを確認した．

授業では続いて，「購入したもの」と「年齢」のクロス集計を扱った．生徒は図 -7 のようなクロス集計を行い，結果

Connect DB

Bit Arrow

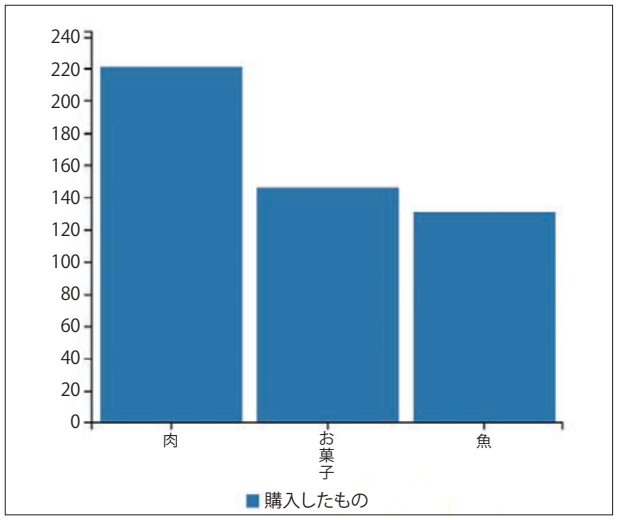
■図 -2　学習環境

■年齢	■購入したもの
40代	肉
40代	肉
30代	肉
30代	肉
40代	魚
30代	お菓子

■図 -3　購入履歴

■階級値	■購入したもの
肉	221
お菓子	146
魚	131

■図 -4　度数分布による定性データの集計分析

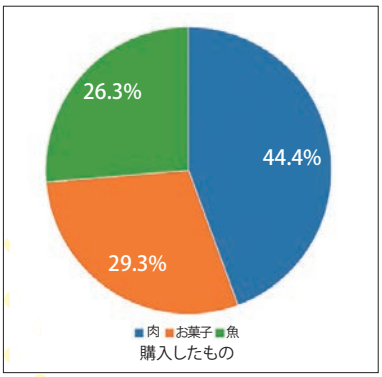


■図 -5　棒グラフによる度数分布の表示

を視覚化して比較検討するために，図 -8 や図 -9 のような，円グラフや帯グラフを表示した．そして，円グラフでは年代別などの内訳を読み取ることができ，帯グラフでは全体の比率を一覧して比較できることを学習した．

散布図と相関係数

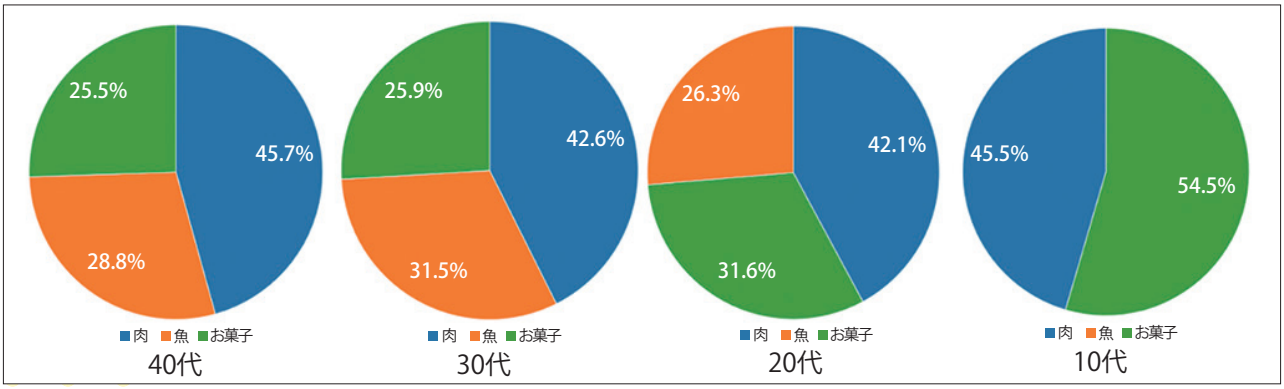
2 限目の授業では定量データの分析を扱い，気温とアイス・お茶の売り上げ数を使用して相関係数を求めた．「気温」と「アイス売り上げ数」や，「気温」と「お茶の売上数」の散布図を調べると相関係数が計算される．図 -10 のように，気温に対してアイスの売上数の相関係数は 0.8622 と正の相関を持っており，気温が高くなるとアイスの売上数が上昇する傾



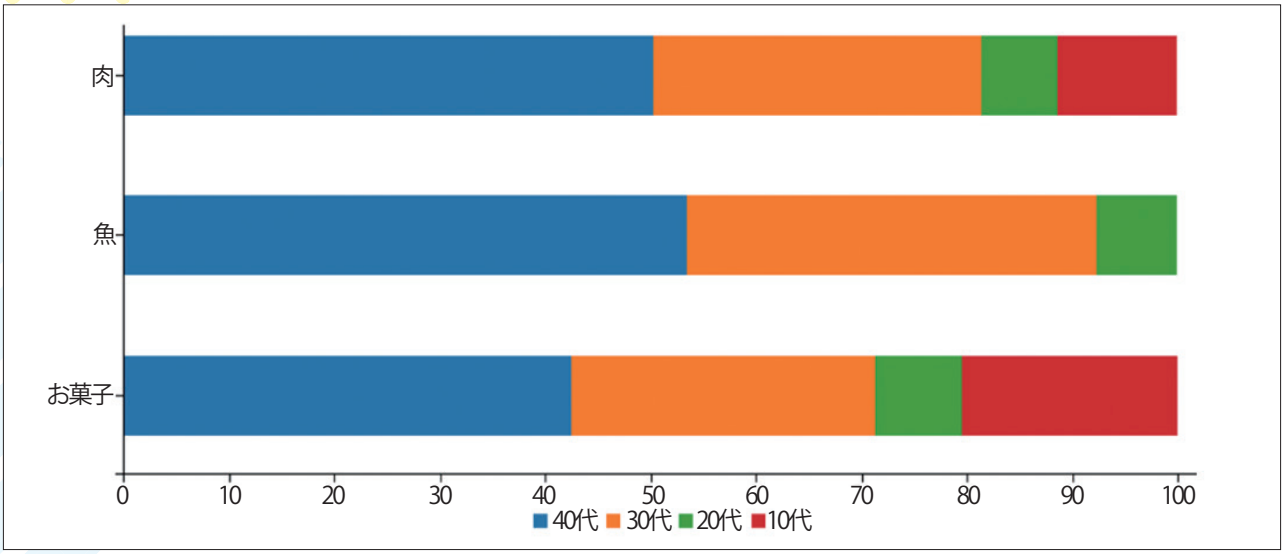
■図 -6　円グラフによる比率の表示

階級値	40代	30代	20代	10代
肉	111	69	16	25
魚	70	51	10	0
お菓子	62	42	12	30

■図 -7　クロス集計による年齢層ごとの集計



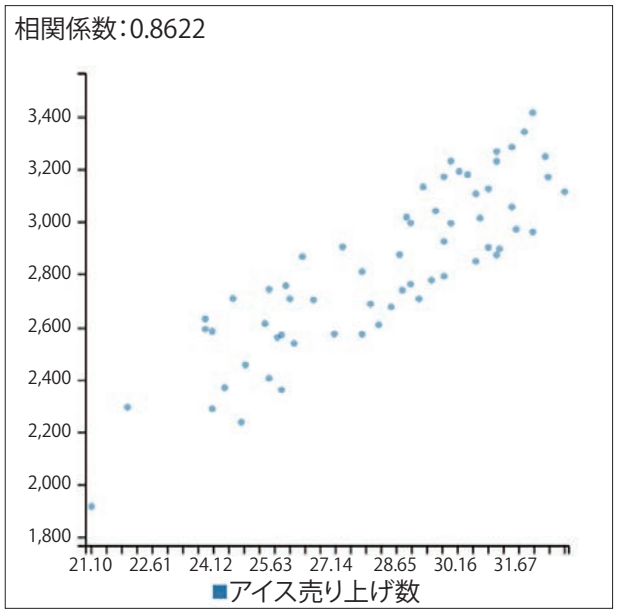
■図 -8　複数の円グラフによる年齢層ごとの比率表示



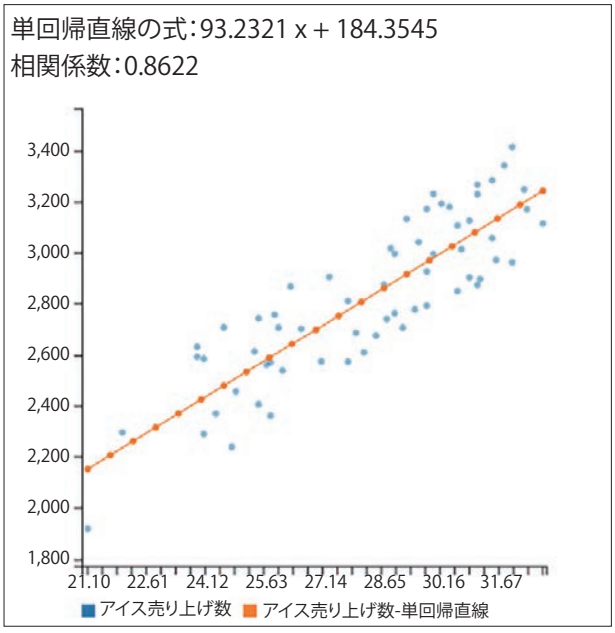
■図 -9　帯グラフによる比率の表示

向が見られる．同様に，気温とお茶の売上数も相関があった．そこで，アイスの売上数とお茶の売上数の相関係数を計算した．相関係数が 0.8256 であったため，生徒たちは「アイスが売れるとお茶も売れる」と考えていたが，生徒同士で議論することで，疑似相関であることを発見して理解することができた．

続いて，散布図と相関係数を調べた後に，図 -11 のような，各点の距離が最短になる「単回帰直線」を表示させた．「単回帰直線の式を使用すると，調べていない点の値を予測できる」ことを説明し，予測について議論した．生徒は「この直線式からは，0℃の時は約 180 本売れると予測できるが，本当に真冬に 180 本になるのかは予想が難しい」などの声があり，範囲外の数値を予測するのは単回帰直線でも簡単でないことを理解することができた．



■図 -10　散布図と相関係数による気温と売上の分析



■図 -11　単回帰直線による気温と売上の分析

オープンデータを使用したデータ分析

Connect DB はオープンデータを読み込んで使うことができる．3 限目の授業では，近隣の大阪府寝屋川市の公衆トイレのデータをダウンロードして使用した[3]．ダウンロードしたオープンデータを Connect DB に登録してマップを表示すると，図 -12 のように寝屋川市の公衆トイレの位置がマップとして表示される．この課題は生徒全員が問題なく作業を行うことができた．

授業でデータ分析を体験した生徒の感想を図 -13

■図 -12　寝屋川市の公衆トイレマップ

- グラフの組合せで見やすくなって面白かった．
- POS データは買った人の好みが分かるので，それに合わせて商売していると考えると面白かった．
- オープンデータの存在は前から知っていたが使ったことがなかったのでよかった．
- 普段の日常で必ず使うものが地図で表示できるのは便利だと思った．もっとデータを知りたい．
- 地図にコンビニの位置や POS データを保存すれば，簡単に行動が分析できそうで面白い．

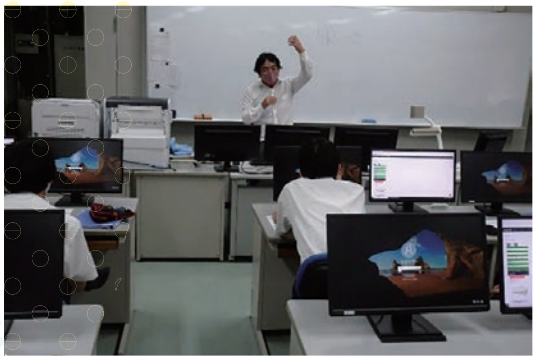
■図 -13　生徒の感想

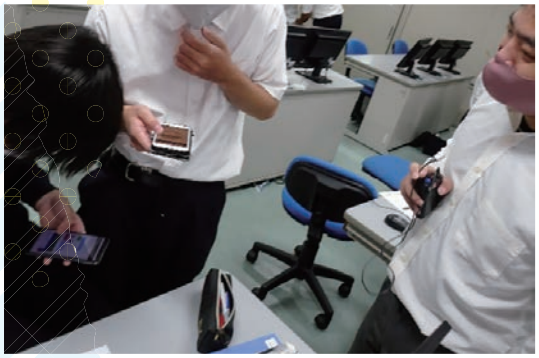
に載せる．全体的な感想は，「面白かった」といった肯定的な意見が多く，生徒が興味を持ってくれたことが分かった．オープンデータの利用も，「名前は知っていたが利用方法が知れてよかった」といった声があった．ほかにも，「地図データとPOSデータから生活習慣を分析したい」と考える生徒もいた．

スマートフォンの加速度センサを使用した動作分析

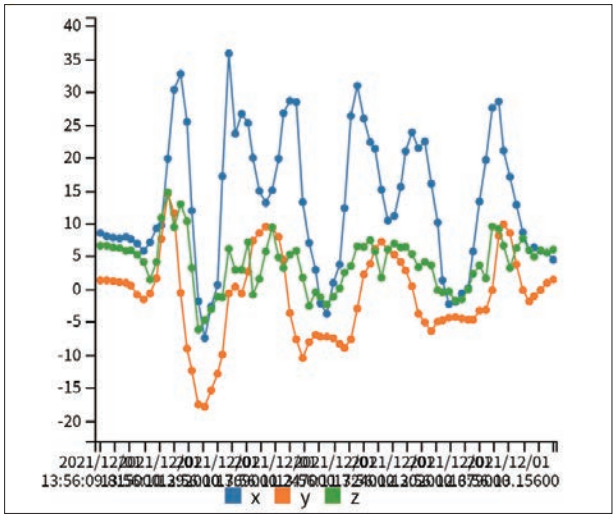
4限目と5限目の授業では，スマートフォンでConnect DBのWebサイトにアクセスした．生徒は普段と違い，スマートフォンを使う授業に新鮮な興味を示していた．授業では最初に加速度の意味を説明し，加速度センサの計測を行うと，持っているだけでも地球の重力に引かれ重力加速度として数値が表示されることを伝えた．そして，スマートフォンを持ちながら「歩く動作」「走る動作」「オリジナルの動作」について試しながら，それらの特徴を考察した（図-14）．

「歩く動作」は，図-15（a）のように加速度センサが振動している様子が分かる．「走る動作」は，図-15（b）

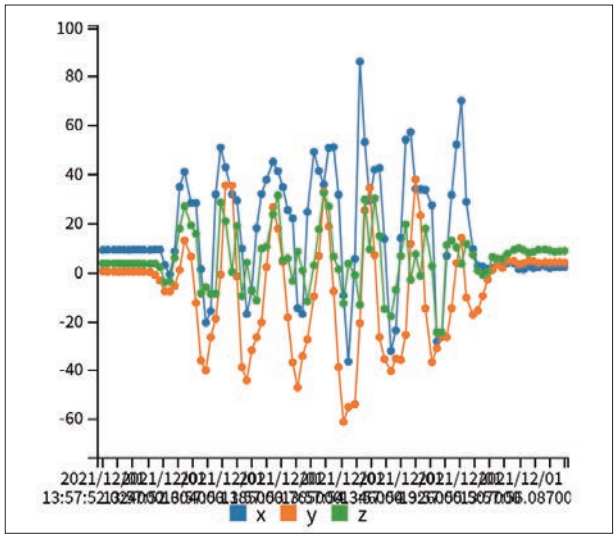




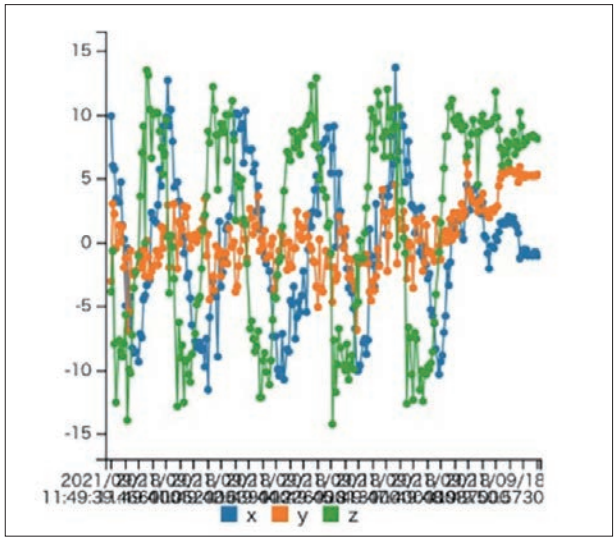
■図-14　スマートフォンの加速度センサを用いた動作の計測の様子



（a）歩く動作



（b）走る動作



（c）オリジナルの動作（くるくる回す）

■図-15　加速度センサのグラフ

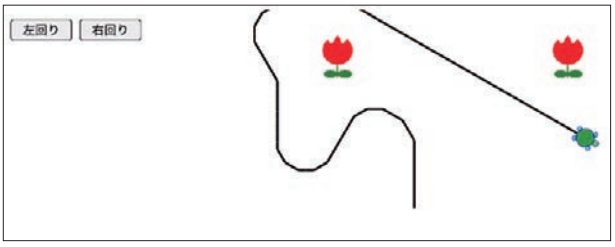
のように細かい振動であり振れも大きいことが読み取れる．「オリジナルの動作」は，図 -15（c）のように「スマホをくるくる回した」生徒がいた．この場合，グラフからは歩いているような波形が読み取れる．生徒たちは「加速度センサだけでは，歩行と回転の動作の特徴を見つけるのが難しい」ことに気づいたため，授業ではジャイロセンサの必要性についても説明した．

スマートフォンのセンサを使用したゲーム制作

4，5 限目のセンサの計測実習から，センサの計測データも数値であることを生徒達は学んでいる．6 限目の授業では，センサから得られた数値データを制御プログラムに活かせるようにゲームを制作した．制作環境は，Bit Arrow（ドリトル言語）[2] である．完成したプログラムを図 -16 に示す．このプログラムでは，スマートフォンの傾きをジャイロセンサで検出することで画面上のキャラクタを操作することができる．PC の画面で Bit Arrow からプログラムを入力し，画面に表示した QR コードを読み込むと，入力したゲームがスマートフォンで起動する（図 -17）．ゲームを完成させた生徒からは，自分たちの入力したプログラムがスマートフォンで動作する学習を体験することで，喜びの声が上がっていた．

```
かめた＝タートル！作る ９０ 左回り.
ボタン１＝ボタン！ "左回り" 作る.
ボタン１：動作＝「かめた！ 30 左回り.」.
ボタン２＝ボタン！ "右回り" 作る.
ボタン２：動作＝「かめた！ 30 右回り.」.
時計＝タイマー！作る １００ 時間.
時計！「
    かめた！ 10 歩く.
    x= ジャイロセンサ！ヨー？.
    かめた！（x/3）右回り.
」実行.
タートル！作る ペンなし "tulip.png" 変身する -100 200 位置.
タートル！作る ペンなし "tulip.png" 変身する 200 200 位置.
かめた：衝突＝「｜相手｜ 相手！消える.」.
```

■図 -16　ジャイロセンサのゲームプログラム



■図 -17　ドリトルでゲームの実行画面

今後の展開

生徒たちは，本授業を通して統計的な考え方を理解できた．さらに，スマートフォンから加速度センサを計測することで，動作の特徴を考察することができた．今後，より内容を広げていくには，スマートフォンの内蔵センサの利用だけでは終わらずに，さまざまなセンサを利用してデータを集めて分析する IoT を利用した実習をすることが必要であると考えられる．

参考文献

1）Connect DB, https://cdb.eplang.jp（参照 2021-10-24）.
2）Bit Arrow, https://bitarrow.eplang.jp（参照 2021-10-24）.
3）寝屋川市オープンデータ，https://www.city.neyagawa.osaka.jp/organization_list/keieikikaku/johosuisinka/open_data/opendata/index.html（参照 2021-10-24）.

（2021 年 10 月 27 日受付）

岸本有生（正会員）
t-kishimoto@dentsu.ed.jp

大阪電気通信大学高等学校教員．2006 年から工業科を担当している．

連載

先生，質問です！

今回は倉橋先生からゲーム理論について網羅的にご回答いただきました。

「ゲーム理論」て、何ですか？（スマホゲームではないだろうことは分かりますが，ネーミングが面白いので聞いてみました）

匿名希望
【正会員】

Q

倉橋節也
【正会員】
筑波大学

ゲーム理論の「ゲーム」とは，PlayStation や Nintendo Switch などと同じ意味のゲームで，生物や人間が複数いる中で行われるさまざまな意思決定を，数学を使って表現したものです．この理論を使うと，競争や対立，協力，交渉などが，どのようにして行われているのかを説明することができます．たとえば，囚人のジレンマというゲームでは，一緒に罪を犯した 2 人の囚人が，それぞれの取り調べの中で，黙秘をするか自白するかを，相棒がどうするかが分からない状態で決める場合に，どちらが有利かを判断したり 2 人の判断がどうなるのかを推定したりすることができます．経済学での応用も有名で，多くのノーベル経済学賞の受賞者が，このゲーム理論を使って研究をしています．たとえば，共有地の悲劇と言われる，多数の人が共有する資源の乱獲をどうしたら防げるのかといった研究で，オストロム（Elinor Ostrom）が 2009 年に受賞しています．これは，地球温暖化の原因となっている温暖化ガスの排出問題につながる研究でもあり，私たちの社会と直結した科学理論として，ゲーム理論は注目を集めています．

A

「先生，質問です！」・「先生が質問です !!」への質問・回答募集

▸ **Web から質問する：** 下記の Web ページ内の投稿フォームから質問をご記入ください．
「先生，質問です！」 https://www.ipsj.or.jp/magazine/sensei-q.html
「先生が質問です !!」 https://www.ipsj.or.jp/magazine/senseiga-q.html

先生，質問です！　先生が質問です !!

▸ **回答募集：** 情報処理学会 Facebook ページ（@IPSJ.official） Twitter アカウント（@ipsj_shinsedai）

ぺた語義

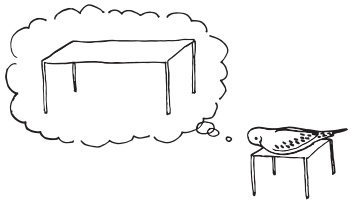
Vol.125

CONTENTS

【コラム】オンライン授業を快適に受講するには?…越智 徹
【解説】シンポジウム「大学入学共通テスト『情報』が目指すもの」…稲葉 利江子
【解説】大学入学共通テストにおける教科「情報」の導入を受けて…河原 達也
【解説】国立大学入学者選抜制度への「情報」の追加について…中山 泰一

COLUMN

オンライン授業を快適に受講するには？



2020 年度の前期授業はコロナ禍の中，オンライン授業で幕を開けた．年が明けて 2021 年度の大阪工業大学の前期授業は，当初は対面形式で始まったものの，第 5 波による 3 度目の緊急事態宣言発出によって，再びオンライン授業に移行した．2020 年度はほぼ誰もが経験したことのない中でのオンライン授業だったが，2021 年度は，1 年生は「初めて」でも教員側は「再び」のオンライン授業になる．なお本学では，BYOD（Bring Your Own Device）によって，全員ノート PC を所持していることが前提となっていたため，PC 環境についてはあまり心配する必要がなかったのは幸いだった（もちろんある程度のサポートは必要だったが）．

筆者は，前期授業期間中では 1 年生の情報リテラシー系授業を担当しているが，2020 年度の経験を経て，学生には受講環境に関するさまざまなアドバイスを与え，オンライン授業での受講環境や疲労について調査した．詳しくは，情報教育シンポジウム 2021 での発表を参照いただきたいが [1]，特にオンライン配信画面を見つつ，PC 作業を伴うような演習授業の場合は，ノート PC の画面だけでは実質的に半分の画面領域しか使用できない．オンライン授業の快適さを向上させるにはこの点の解決が必要と考え，学生には次のように連絡した．

1）外付けモニタは安価なものだと 1 万円前後で購入できる．
2）通常の液晶テレビでも HDMI 端子で接続すれば PC の外付けモニタとして使用できる．

さて，学生は外部モニタを導入しただろうか．アンケート調査を実施したところ，回答者 161 人中，外付けモニタもしくはテレビを使用したと回答したのは 25 人と 15% 程度だった．また，「設置場所や購入費用をまったく考慮しないと仮定して欲しいものは」と質問すると，外付けモニタが 23 人，もっと広い机という回答が 37 人という結果となった．新しく外付けモニタを購入するとしても，ノート PC や教科書類，さらにモニタのスペースが必要になってしまう．そのため，モニタよりもまず広い机が欲しい，という回答になったのではないか．

現在，後期授業中も筆者の一部担当科目はオンライン授業を継続している．前期に引き続き，外付けモニタの導入を推奨しているが，何人かはその後購入し「買ってよかった．とても快適です」と感想を送ってきた学生もいた．オンライン授業をどのように快適に受講してもらうか，まだまだ暗中模索，道半ばである．

参考文献

1）越智　徹，館野浩司：初年度情報リテラシー教育のオンライン授業における受講環境と疲労の調査，情報教育シンポジウム論文集（SSS2021），Vol.2021, pp.61-68.

越智　徹（大阪工業大学）（正会員） toru.ochi@oit.ac.jp

大阪工業大学情報センター講師．情報工学，情報教育が専門．情報センター教員として，2018 年度より導入した BYOD 運用の学生向けマニュアル作成や初年次情報リテラシー教育などを担当している．また，企業と合同で AI や IoT の教材開発や講座の実施も手がけている．

LOGOTYPE DESIGN...Megumi Nakata, ILLUSTRATION&PAGE LAYOUT DESIGN...Miyu Kuno

ARTICLE

応
般

シンポジウム「大学入学共通テスト『情報』が目指すもの」

稲葉利江子
津田塾大学

大学入学共通テスト「情報」

大学入試センターは，2021 年 3 月 24 日に公表した 2025 年に実施する大学入学共通テストの教科・科目の再編案において，「情報」を新たに導入し，国語や数学などと並ぶ基礎教科とする方針を示した．

これを受け，FIT2021（第 20 回情報科学技術フォーラム）において，日本学術会議情報学委員会情報学教育分科会，情報処理学会，電子情報通信学会が主催して，公開シンポジウム「大学入学共通テスト『情報』が目指すもの」が，2021 年 8 月 26 日にオンライン開催された．当日は 300 名を超える多くの参加者があった．

文部科学省が 7 月 30 日に「令和 7 年度大学入学者選抜に係る大学入学共通テスト実施大綱の予告」で，「情報 I」が 2025 年の大学入学共通テストから独立した科目として実施されることを公表したことが影響したと思われる．

本稿では，公開シンポジウムの内容について報告するとともに，大学入学共通テスト「情報」の動向について述べる．

シンポジウムの概要

❑ 開会挨拶

徳山　豪氏（日本学術会議情報学委員会情報学教育分科会委員長，関西学院大学）

開会挨拶として，シンポジウムのテーマの趣旨と背景について説明がなされた．

日本学術会議では，これまで，情報学の位置づけと情報教育の設計の在り方について検討し公表してきた．中でも，2016 年に公表された大学教育における「情報学分野の参照基準[☆1]」では，「情報学とは何か」という理想の形を示し，小学校から大学の共通教育，専門基礎教育までの各教育段階において，「情報学の何を学ぶことが必要なのか」ということを示した．さらに，2020 年に公表した「情報教育課程の設計指針―初等教育から高等教育まで[☆2]」では，メタサイエンスとしての情報学の位置づけや，情報社会において市民の一人ひとりが情報技術に関する知識を有することが求められることを示した．つまり，情報学は，IT の科学技術の専門家だけの学問領域だけではなく，すべての市民に必要な教育であり，教育現場の裁量で実現していかなければならない．

徳山氏はさらに，「入試と教育は『ニワトリと卵』の関係にある」と述べられた．入試により教育目標を明確化し，教育の充実を良いサイクルで回すことで，情報教育の人材育成や教育環境の整備を喚起し，情報先進国を支える人材の育成につながるのである．

❑ 講演 1「新学習指導要領に対応した令和 7 年度大学入学共通テストの出題教科・科目について」

前田幸宣氏（文部科学省高等教育局大学振興課大学入試室長）

政府における「情報 I」の出題に関する閣議決定等の変遷が示された． 2021 年 7 月に「大学入試のあり方に関する検討会議」から「新たに必履修科目となる『情報 I』を出題すべき」と提言がなされた．さらに，「『情報』については，問題の発見・解決に向けて情報

☆1 https://www.scj.go.jp/ja/info/kohyo/pdf/kohyo-23-h160323-2.pdf
☆2 https://www.scj.go.jp/ja/info/kohyo/kohyo-24-h200925-abstract.html

技術を活用する力を見る出題を工夫することが期待される」ことなども示された．

大学入試においては，個別学力検査および大学入学共通テストにおいて課す教科・科目を変更する場合には，2年程度前に予告する必要があるという，いわゆる「2年前予告ルール」がある．そのため，2021年7月30日に「『令和7年度大学入学者選抜実施に係る大学入学共通テスト実施大綱の予告☆3』及び『令和7年度大学入学者選抜実施要項の見直しに係る予告』」が通知された．内容としては，6教科30科目から7教科21科目となり，「情報Ⅰ」を新たに加え，試験形態は引き続き，紙ベースで試験を行うことが公表された．あわせて，試験時間と現行の教育課程を履修した入学志願者(浪人生)への対応については，大学入学者選抜協議会で議論し公表されることが示されている．

【注】2021年9月29日に，「令和7年度大学入学者選抜に係る大学入学共通テスト実施大綱の予告（補遺）☆4」が通知され，「情報Ⅰ」については，試験時間が60分，現行の教育課程履修者に対応した経過措置が実施されることが発表された．

❑ 講演2「大学入学共通テスト『情報』サンプル問題について」

水野修治氏(大学入試センター試験問題調査官)

お伝え(お願い) したいこととして，図-1に示されている4点の内容の説明がなされた．特に，2点目については，「情報Ⅰ」の試験がPBT (Paper Based Testing) で実施されることになったことで「暗記科目になるのではないか」という懸念に対する伝えたいことである．そもそも大学入学共通テストの問題作成方針では，以下2点が示されており，これをもとに作問がなされる．

- 知識の理解の質を問う問題や，思考力，判断力，表現力等を発揮して解くことが求められる問題を重視
- 授業において生徒が学習する場面や，社会生活や日常生活の中から課題を発見し解決方法を構想する場面，資料やデータ等を基に考察する場面など，学習の過程を意識した問題の場面設定を重視

したがって，PBTであっても情報に関する深い知識や，思考力，判断力，表現力等を測ることができる試験を目指しており，懸念されることはないことが伝えられた．

次に，2021年3月に大学入試センターが公開したサンプル問題☆5について解説がなされた．なお，サンプル問題は，「情報Ⅰ」の問題のイメージを共有するため，有識者に短期間で作成いただいたものであり，セットとして作成したものではなく，教科書は検定中であるため照合したものではないという点に注意が必要である．

また，入試におけるプログラミング言語について気になる方も多いと思われる．授業で多様なプログラミング言語が利用される可能性があること，共通テストとして実用性よりも教育的で，公正・公平なプログラミング言語が求められることから，大学入試センター独自の日本語表記の疑似言語DNCLで出題すること，教科書などで利用されているプログラミング言語をしっかりと学習すれば，DNCLの仕様を知らなくても無理なく理解できるようにすることを検討しているとのことである．

最後に，大学では，文系・理系問わず，数理・データサイエンス教育強化が現在，進められているが，「情報Ⅰ」は大学におけるデータサイエンス・AI教育をさらに充実させるための基礎となり得ることを話された．

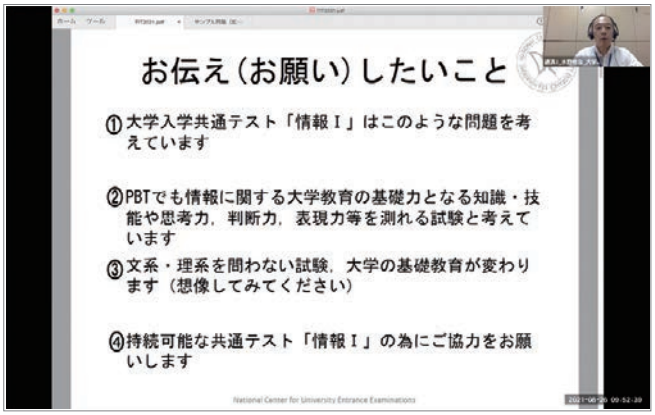


図-1　水野氏の講演の様子

❑ 講演3「高等学校情報科と高大接続，教員養成について」　鹿野利春氏(京都精華大学)

文部科学省で情報科の教科調査官を務められ，新学習指導要領をまとめられた鹿野氏より，まずは新

☆3 https://www.mext.go.jp/content/20210729-mxt_daigakuc02-000005144_1.pdf
☆4 https://www.mext.go.jp/content/20210929-mxt_daigakuc02-000005144_1.pdf
☆5 https://www.dnc.ac.jp/kyotsu/shiken_jouhou/r7ikou.html

学習指導要領により，何がどう変わるのかについて，説明がなされた．

2003年に教科「情報」が設置されてからの変遷が説明され，2022年度からは，「情報Ⅰ」を共通必履修科目として日本の高校生全員が学ぶことが示された．図-2中のスライドの橙色の個所はプログラミングを学ぶ個所になっているが，2022年度からは全員が学ぶことになり，さらに発展的な科目の「情報Ⅱ」も設置される．

「情報Ⅰ」については，文系・理系にかかわらず，国民的素養として皆が身につけていかなければならない内容を厳選し，まとめられている．内容としては，「問題の発見解決」を目指して，「コミュニケーション」「コンピュータネットワーク」「情報モラル」といった知識なども大切にしながら，「情報デザイン」「プログラミング」「データの活用」といったツールをしっかり使いこなしていく形となっている．入試等では，そういった知識・技能に関したものとともに，思考・判断・表現に関したものが問われるのではないかと考えられ，これらがバランス良く出題されることを希望されていた．

また，新しい学習指導要領の「情報デザイン」「プログラミング」「統計に関連した学び」については，小学校，中学校，高校と積み上げていく形になっており，高校で急に高度になったのではなく，小学校からの積み上げで設計されていることが説明された．また，高大接続については，高校に「情報Ⅰ」ができ，それを大学でどのように活用するのか，大学入試も挟み一体的に改革するとなれば，大学入試の意義は大きいと述べられた．

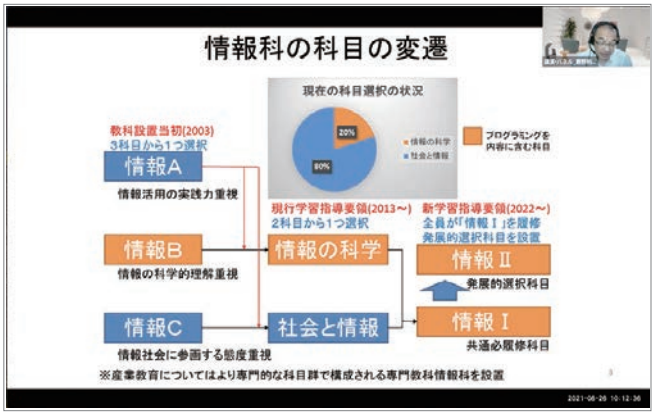


図-2　鹿野氏の講演の様子

そして，高等学校の教員養成については，新学習指導要領での新しい内容についても，現在，各教育委員会や情報科の研究会で活発に研修が行われており，情報科の教員は2022年4月からの授業を十分にやっていけるのではないかと，おっしゃっていた．

❑講演４「大学入学共通テスト新科目案『情報』への期待」　河原達也氏（京都大学）

本会には，情報処理教育委員会という組織があり，高等教育だけでなく初等中等教育においても展開すべく，さまざまな活動を行っている．たとえば，高校教科『情報』シンポジウム（ジョーシン）の開催，情報入試の模擬試験の試行，情報科の教員免許更新講習などの活動である．そして，2020年からは，「情報Ⅰ」のオンライン教材「IPSJ MOOC[☆6]」の制作・公開も行っている．河原氏は，2014年から2016年に本会の教育担当理事として，これらの活動のサポートを行ってきたとともに，2020年12月には8大学情報系研究科長会議から，「情報Ⅰ」の大学入試での取り扱いについて賛同する声明[☆7]を出されている．

「情報」を入試に導入することによる大学にとってのメリットは，現在どの大学でも採用している初年次の基礎情報処理教育のかなりの部分を，「情報Ⅰ」がカバーしているということであると述べられた．2022年度から「情報Ⅰ」が高校生全員に必修化され，さらに入試科目に課されることになれば，これまで大学で教えてきた情報リテラシーは大丈夫だろうという前提をおくことができる．その結果，初年次教育の内容が，より高度なデータ科学やAIの教育にシフトでき，充実させることができるのではないかという期待である．

さまざまなメディアでも高校の「情報Ⅰ」の必修の意義は評価されている一方で，課題として教育体制が挙げられている．特に，高校では専任で教えられる教員の不足や地域格差の問題もある．教員不足の根源的問題は，これまで教科「情報」の位置づけがそれほど重要視されていなかったため，教員採用が抑えられていた

☆6 https://sites.google.com/a.ipsj.or.jp/mooc/
☆7 https://www.i.u-tokyo.ac.jp/proposal/information_8universities.shtml

こともあるが，一方で，情報系を専攻とする学生に教員になる選択肢，インセンティブがあまりないというのも事実であろう．こういった問題に対しては，外部人材やオンライン教材の活用などにより問題が軽減されることを期待したいと述べられた．

情報学，情報技術の持続的な発展においては，情報教育の裾野の広がりが重要であり，今回，大学入試に情報が入ることによってもう一段，押し上げられることを期待したい，と述べられた．

❑総合討論

司会：筧　捷彦氏（東京通信大学）

パネル討論では，本会情報入試委員会委員長の筧氏の司会のもと，2025年に実施する大学入学共通テストに，「情報」を教科として，単一の時間枠で採用されたことを踏まえ，大学入学共通テストにおける情報科の果たすべき役割，そして，大学は今後どのように活用していくかについて議論された．まず，パネリストからご自身の立場と情報入試の位置づけについて話題提供いただいた．本稿では，各パネリストのご意見と興味深い討論について記載する．なお，講演者の発言は，「所属大学の意見を代表するものではない」ことに注意いただきたい．

■徳山　豪氏（関西学院大学）

情報教育をきちんとやっていかないと，近い将来，「日本は情報の後進国になってしまう」ことを一番懸念している．情報教育をしっかり行うためにも，情報入試は必要である．

「情報」とはどういうものなのか．正しく扱うとはどういうことか．どのように大切なものなのかということを教えることが情報教育の基盤だと思う．情報入試に関しても，そういったことを意識していくことが必要である．

■須田礼仁氏（東京大学大学院）

入試というのは，大学教育をどのように進めるかということに直結する．情報入試をパスした学生が入学してくることによって，初年次教育で，いわば「大学らしい」情報教育をスタートできるようになることが期待される．そのためには，大学のカリキュラムの再構築や組織化に，それなりの労力や時間を掛けることが必要になるが，それが進められる大きなきっかけになると思う．

個人的には，情報入試は文系・理系問わず，ぜひ受けてほしいと思う．文系の学生であっても，「情報」をきちんと学ぶ必要がある．

情報技術はきわめて重要で，情報のハード・ソフトがなくては社会が回らないという状況にある今，もし国同士の対立や大規模な事故で使えなくなったらどうなるか，想像するのは難しいことではない．欧州では，以前から「情報技術の自給率」を踏まえて政策を立てている．

新学習指導要領の高校の「情報Ⅰ」，大学入学共通テストへの「情報」の導入が1つのきっかけになることを期待している．

■高岡詠子氏（上智大学）

大学入試が変わる効果として，高校のカリキュラムが変わると，大学のカリキュラムが変わる．初年次教育での情報リテラシーの内容は撤廃され，データサイエンスや人工知能について，より深く学ぶことができるようになるだろう．そして，一番重要と思うのは，情報の素養を身につけているということを社会が評価することである．

「情報Ⅰ」をしっかり勉強して，普段から情報活用能力を身につけるようにすれば，試験対策のためだけの暗記など必要なくなる．そもそも「情報が暗記ものである」という間違った認識をしないでいただきたいと思う．

■中野由章氏（工学院大学附属中学校・高等学校）

情報入試はもう実施が決まったので，いまさらネガティブなことを言っても仕方がない．すでに，どのように上手くやっていくのかというフェーズに変わった．そして，大学入学共通テストが，高校の情報科の授業内容の1つの基準になるだろう．つまり，「うちの生徒たちには，どんな授業をしていかなければならないか」の基準が情報入試である．

日本学術会議が示した「情報教育の設計指針」が目

指したものが実現し，その内容が改訂・充実していくことに期待している．

■鹿野利春氏（京都精華大学）

「情報Ⅰ」は入試のために作ったものではない．これは，国民的素養ということで，初等中等教育で必要な情報活用能力の総仕上げとして，「こういうものが必要である」ということを形にしたものである．しかし，大学でも当然必要なものであり，今後大学の教育を大きく変えていくことになるだろう．

■情報科教員問題について

筧：高校によっては情報科を担当する先生の数が足りていないとか，専任の先生がいないといった問題があります．これはなぜなのか？

中野：批判を覚悟で言えば，教科「情報」が必履修になったのは 2003 年で，すでに 20 年近く経っているにもかかわらず，まだ情報科の先生が足りていないというのは，もはや教育行政の不作為と言われても仕方ないと思う．一方で，きちんと計画的に教員採用を行い，育成してきた自治体や学校もある．ただ，不利益を被るのは子供たちなので，きちんと対策をしてあげなければいけない．

■大学から見た情報入試について

筧：「情報」を大学入試に採用する大学はたくさん出るのでしょうか．

徳山：多少様子見もあるかもしれませんが，情報入試を取り入れる大学はかなり多くなるのではないかと思っている．学術会議も，そういったことを目指して動いていくということになっている．

須田：2021 年 12 月に，8 大学情報系研究科長会議として，情報入試をサポートしていきたいという声明を発表した．これは，大学入学共通テストに「情報」が入っただけでは意味がなく，それを大学が活用して，大学の教育が改善され，日本の社会が変わっていくことが大前提となる．こういったメリットを，いろいろな場や機会で説明していきたいと思っている．

高岡：私自身，学会の情報入試の活動に携わってきているので，多くの大学が入試に取り入れてくださるのが願いであり，期待している．これからもこういった活動を地道にやっていくのがよいと思っている．

大学入学共通テスト「情報」への期待

今回のシンポジウムは，大学入学共通テスト「情報」を取り巻くさまざまな立場の関係者が一堂に会し，それぞれの視点での考えを共有いただけた貴重な機会であった．シンポジウムを通して，登壇者全員が述べられたこととして，以下の 2 点が挙げられる．

- 情報は，情報社会に生きる市民が共通して身につけておくべき素養であり，基礎的な教科である
- 「情報」を入試に導入することにより，大学における全学的な数理・データサイエンスを含む情報教育の改革につながる

これらはまさに，情報の活用・提供が巨大な価値を生む 21 世紀を生きる私たちにとってとても重要なことである．徳山氏が述べられたように，「情報後進国になってはいけない！」のである．

2021 年 10 月 1 日には，大学入試センターから，新学習指導要領を踏まえた問題作成の方向性について 2022 年度中に公表し，出題方法および問題作成方針について 2023 年 6 月までに公表することが発表された．「情報」については，新課程，旧課程の受験生を対象とした出題科目の全体構成が分かる配点付きの試作問題も作成され公表されるとのことである．

今後の日本の将来を見据え，多くの大学で入試に「情報」が採用されることを期待したい．そして，日本の将来が，「情報先進国」であってほしいと願うばかりである．

参考：情報入試に関する本会会誌関連記事

1) 筧　捷彦，中山泰一：情報入試のすゝめ，情報処理，Vol.59, No.7, pp.632-635 (2018).
2) 萩谷昌己：未来投資会議における大学入学共通テストに情報の試験を入れる方針に賛同する提言について―大学情報教育体系化の必要性―，情報処理，Vol.59, No.9, pp.778-781 (2018).
3) 高岡詠子：100 回の重さ，情報処理，Vol.61, No.1, pp.80-84 (2020).
4) 髙田真弥：大学入学共通テスト「情報」サンプル問題を題材とした研究協議～令和 3 年度愛知県高等学校情報教育研究会研究協議を通して～，情報処理，Vol.62, No.11, pp.610-613 (2021).

(2021 年 10 月 31 日受付)

稲葉利江子（正会員） inaba@tsuda.ac.jp

津田塾大学学芸学部情報科学科准教授．メディア情報学，教育工学に関する研究に従事．現在，本会情報処理教育委員会，情報入試委員会，セミナー推進委員会などの委員として活動．

ARTICLE

応般

大学入学共通テストにおける教科「情報」の導入を受けて

河原達也
京都大学

ここまでの経緯

社会の高度情報化，いわゆるデジタル化が進展する中，我が国の「AI 戦略 2019」においても，「すべての高等学校卒業生（約百万人／年）がデータサイエンス・AI の基礎となる理数素養や基本的情報知識を習得する」という目標が掲げられている．これに対応して，高等学校で「情報 I」が 2022 年度から必修化されることとなり，さらにその 3 年後から大学入試共通テストに導入されることが決定された．その経緯については，中山の記事[1)]を参照されたい．

筆者は 2014 ～ 2016 年にかけて，本会の教育担当理事として，情報教育の裾野を広げるためのさまざまな活動にかかわった．その中には，高校教科「情報」シンポジウムの開催，教員免許更新教習や情報入試模擬試験などが挙げられる．これらの活動が結実していったのは感慨深いとともに，関係者の多大な努力に敬意を表したい．

今回のような大学入試における新教科の導入は，大学共通第 1 次学力試験以来，前例がないものである．そのため，各大学・学部においてこの取り扱いについて新たに検討を行っていると思われる．本稿では，主に大学教員の立場から，大学入学共通テストにおける教科「情報」の導入の意義と期待についていくつかの観点から述べる．

入学者選抜の観点

大学入学共通テストにおいて「情報」が追加され，「6 教科 8 科目」のパッケージとして扱われるのであれば，そのまま受け入れるのが自然と考える向きもあるが，既存の科目についても配点を傾斜している場合は配点の検討も必要となる．その場合は，入学者選抜において「情報 I」で何を見るのか考慮することになる．

「情報 I」は以下の 4 つの内容から構成される．

(1) 情報社会の問題解決
(2) コミュニケーションと情報デザイン
(3) コンピュータとプログラミング
(4) 情報通信ネットワークとデータの活用

特筆すべきは，プログラミングを含む科学的理解やデータ科学の基本的考え方が必修になったことである．これらから，論理的思考力やデータ解析力，そして情報リテラシーなどを見られると考えられる．論理的思考力などは数学で十分と考える向きもあるが，実世界に則した問題で見られるのは情報の強みであろう．しかしながら，高校で「情報 I」が 2 単位しかないことを考慮すると，他の科目と比べて，配点が小さくなってもやむを得ないだろう．

大学の情報教育の観点

大学にとって，情報科を入試に課すことの明白なメリットは，学生が一定の情報リテラシーを有する

ことが担保できるので，入学後の基礎的な情報処理教育（の一部）が不要になることであろう．筆者の大学でもほぼすべての学部で事実上必修の扱いで，数単位を割いているが，コンピュータやネットワークの基礎やワープロや表計算ソフトの使い方などを教えている．これらは，高校の情報科で十分カバーされるものである．したがって，より高度なデータ科学やAIの教育から始めることができると期待できる．

実際に，情報系以外の理工系・医薬・農学や人文社会系でも，データ科学の知識だけでなく，ICTやAI関連のプログラミングができるとよいと考える向きが増えており，これを機に大学における情報教育のカリキュラムを見直す必要がある．

初等中等情報教育充実の観点

その反面，大学入試センターで作成された試作問題[2]やサンプル問題[3]に対する反応・評価を見ると，特にプログラミング関連の問題が難しく，高校の教員の間では不安を感じているようである．

これまでも．情報科は単位数が少ないことから専任教員の採用が少なく，他の教科の担当者が掛け持ちしていたり，臨時免許などで対応していることが多かった[4]．令和2年度時点の調査[5]においても，全国の情報科担当教員約5.000名のうち24%が免許外教科担任か臨時免許状での担当となっている．また，その半数を8県で占めており，地域間格差も大きい．これが大学入試において情報科を実際に採用する際の障壁と指摘される．

しかし，大学入試に採用することで，高校における情報科の教育体制の充実を促すことも期待される．実際に，これまで専任教員の採用のなかった県でも今年度初めて採用されたとのことである．

入試科目に採用されるということは，高校だけでなく，中学における情報教育の充実にも影響を与えることが期待される．このように，情報教育の推進力を高めることが，情報科を入試に採用することの大きな意義と考えられる．情報学が，数学や物理学と同様に「学問」と認知されることにもつながる．情報の専門家集団である本会には，さらなる関与・貢献が期待される．その一例として，「情報I」に対応したオンライン教材MOOC[6]の構築が挙げられる．

参考文献

1) 中山泰一：大学入学共通テストへの「情報」の出題について．ニューサポート高校「情報」，Vol.18, pp.6-7 (2021), http://id.nii.ac.jp/1438/00009894/
2) 井手広康：大学入学共通テスト「情報」試作問題に対する教育現場の想い，情報処理 Vol.62 No.5, pp.254-257 (2021), https://ipsj.ixsq.nii.ac.jp/ej/?action=repository_uri&item_id=210701
3) 水野修治：大学入学共通テスト「情報」のサンプル問題について，情報科学技術フォーラム（FIT）(2021), https://www.ipsj.or.jp/event/fit/fit2021/FIT2021_program/data/html/event/pdf/eventB2_347.pdf
4) 中山泰一：高等学校情報科の教員採用と免許外教科担任の現状，情報教育資料，Vol.50, pp.14-16 (2020), http://id.nii.ac.jp/1438/00009464/
5) 文部科学省：高等学校情報科担当教員の専門性向上および採用・配置の促進について（通知）（令和3年3月），https://www.mext.go.jp/content/000102780.pdf
6) IPSJ MOOC 情報処理学会 公開教材，https://sites.google.com/view/ipsjmooc/

（2021年10月28日受付）

河原達也（正会員） kawahara@i.kyoto-u.ac.jp

京都大学情報学研究科教授・研究科長．2014年～2016年本会理事．2017年から日本学術会議連携会員．

ARTICLE

基般

国立大学入学者選抜制度への「情報」の追加について

中山泰一
電気通信大学

高等学校情報科と情報入試のながれ

2018年3月30日，2022年度から高等学校で実施される新学習指導要領が告示された．情報科は，情報の科学的な理解に重点を置き，「情報Ⅰ」を必履修科目とした上で，その発展的内容を扱う「情報Ⅱ」を選択科目として設置することになった．内容は，次のとおりである．

●情報Ⅰ（必履修科目，2単位）
(1) 情報社会の問題解決
(2) コミュニケーションと情報デザイン
(3) コンピュータとプログラミング
(4) 情報通信ネットワークとデータの活用

●情報Ⅱ（選択科目，2単位）
(1) 情報社会の進展と情報技術
(2) コミュニケーションとコンテンツ
(3) 情報とデータサイエンス
(4) 情報システムとプログラミング
(5) 情報と情報技術を活用した問題発見・解決の探究

そして，2021年7月30日に文部科学省は，2025年の大学入学共通テストから「情報」を出題教科として，「情報Ⅰ」をその科目とすることを決定した（表-1）．それまでの経緯は文献1），2），3）を，また，大学入試センターが同年3月24日に公表した「情報」のサンプル問題は文献4）を参照されたい．

さらに，文部科学省は同年9月29日に，「情報Ⅰ」を独立した時間帯に60分で行うことと，2025年の大学入学共通テストでは既卒者のために旧学習指導要領（2009年3月告示，情報科は「情報の科学」と「社会と情報」の選択必履修）に対応した経過措置問題を出題することを決定している[4]．

国大協による「6教科8科目」の原則の検討

現在，国立大学協会（国大協）で，2025年に実施する入学者選抜制度が議論されている．これまで，国立大学は一般選抜においては，第一次試験として大学入学共通テスト（原則5教科7科目）を課して

表-1　令和7年度大学入学者選抜に係る大学入学共通テスト実施大綱において定める出題教科・科目

教科	グループ	出題科目
国語		『国語』
地理歴史		『地理総合，地理探究』，『歴史総合，日本史探究』，『歴史総合，世界史探究』，『地理総合，歴史総合，公共』
公民		『公共，倫理』，『公共，政治・経済』，『地理総合，歴史総合，公共』（再掲）
数学	①	『数学Ⅰ，数学A』，『数学Ⅰ』
	②	『数学Ⅱ，数学B，数学C』
理科		『物理基礎，化学基礎，生物基礎，地学基礎』 『物理』，『化学』，『生物』，『地学』
外国語		『英語』，『ドイツ語』，『フランス語』，『中国語』，『韓国語』
情報		『情報Ⅰ』

きた．これに「情報」を加えた「6 教科 8 科目」を原則とすることが検討されている．高等学校新学習指導要領で 2022 年度から「情報Ⅰ」が必履修科目となること，2018 年 5 月 17 日に開催された第 16 回未来投資会議で「大学入試においても，国語，数学，英語のような基礎的な科目として，情報科目を追加，文系，理系を問わず理数の学習を促していく」とされたことが背景にあると考えられる．

「情報」を加えた「6 教科 8 科目」の原則が検討されていることは，2021 年 11 月 12 日開催の国大協総会後の記者会見で示されており，2022 年 1 月 28 日開催予定の国大協総会で審議される予定とのことである．本稿の掲載は，その審議の前であり，予断を許さない状況ではあるが，筆者は「情報」を加えた「6 教科 8 科目」の原則が決定されることを強く願うとともに，その願いが叶うと信じている．

国大協入試委員会が大学入試センターに宛てた経過措置についての意見では，(1) 経過措置問題の出題，(2) 得点調整の実施，(3) 現在の高校生への周知を求めている（図 -1）．これに対し，大学入試センターは 2021 年 12 月 17 日に「情報」の出題方法の詳細（図 -2）を公表するとともに，「情報Ⅰ」と「旧情報（仮）」の間で得点調整を行うと公表している [4]．

デジタル社会を生きる生徒には，文系，理系を問わず，大学入学時点で情報活用能力を身に付けていることが求められる．国立大学の入試科目に「情報」が加わることの意義は大きいと筆者は考えている．

以下の条件を満たした上で実施される場合には，「適当である」と考えられる．

1. 旧教育課程「情報」に対応した経過措置問題（以下，『旧情報』）については，旧教育課程における教科「情報」の選択科目である，「社会と情報」「情報の科学」いずれの履修者も回答できるような問題内容，あるいは選択問題の設定が行われること．
2. 『旧情報』と『情報Ⅰ』との間で難易度に差が出ないような作問がなされること．なお，『旧情報』と『情報Ⅰ』で一定の平均点差が生じた場合には，得点調整が実施されること．
3. 現在旧教育課程を履修している高校生に対して，令和 7 年度入試においては，現在大学共通テストで出題されていない『旧情報』および『情報Ⅰ』が出題されることについて，十分な説明がなされること．

図 -1　国大協入試委員会が大学入試センターに宛てた経過措置についての意見（第 7 回大学入学者選抜協議会（2021 年 9 月 13 日開催）の配布資料（参考資料 4「『情報Ⅰ』の経過措置についての関係団体からの意見」）からの抜粋）[5]

新教育課程（平成 30 年 3 月告示の高等学校学習指導要領に基づく教育課程）に対応した『情報Ⅰ』とは別に，現行の教育課程（平成 21 年 3 月告示の高等学校学習指導要領に基づく教育課程）の「社会と情報」及び「情報の科学」の内容を出題範囲とする経過措置科目『旧情報（仮）』を出題する．

なお，『旧情報（仮）』では，高等学校等において「社会と情報」，「情報の科学」のいずれの科目を履修していても不利益が生じないよう，両科目の共通部分に対応した必答問題に加え，「社会と情報」に対応した問題及び「情報の科学」に対応した問題を出題し，選択解答させる．

図 -2　令和 7 年度大学入学者選抜に係る大学入学共通テスト「情報」の出題方法について [4]

参考文献

1) 中山泰一：大学入学共通テストへの「情報」の出題について，ニューサポート高校「情報」，Vol.18, pp.6-7 (2021).
2) 萩谷昌己：大学入学共通テスト実施大綱の予告に関する本会の意見について，情報処理，Vol.62, No.11, pp.e62-e66 (2021).
3) 河原達也：大学入学共通テストにおける教科「情報」の導入を受けて，情報処理，Vol.63, No.2, pp.77-78 (Feb. 2022).
4) 大学入試センター：令和 7 年度以降の試験に向けた検討について，https://www.dnc.ac.jp/kyotsu/shiken_jouhou/r7ikou.html
5) 大学入学者選抜協議会：議事録・配付資料，https://www.mext.go.jp/b_menu/shingi/chousa/koutou/112/giji_list/

（2021 年 12 月 1 日受付）

中山泰一（正会員） nakayama@uec.ac.jp

1993 年東京大学大学院工学系研究科情報工学専攻博士課程修了．同年より電気通信大学において，計算機システム，並列分散処理，情報教育の研究に従事．現在，同大学院情報理工学研究科教授．2017 年度科学技術分野の文部科学大臣表彰科学技術賞受賞．2020 年より本会教育担当理事．日本学術会議特任連携会員．

［重要］過去のプログラミング・シンポジウム報告集の利用許諾について

2020 年 12 月 18 日
プログラミング・シンポジウム委員会

情報処理学会発行の出版物著作権は平成 12 年から情報処理学会著作権規程に従い，学会に帰属することになっています．

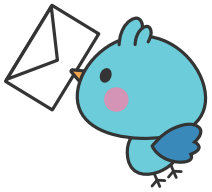
プログラミング・シンポジウムの報告集は，情報処理学会と設立の事情が異なるため，この改訂がシンポジウム内部で徹底しておらず，情報処理学会の他の出版物が情報学広場 (= 情報処理学会電子図書館) で公開されているにもかかわらず，古い報告集には公開されていないものが少からずありました．

プログラミング・シンポジウムは昭和 59 年に情報処理学会の一部門になりましたが，それ以前の報告集も含め，このたび学会の他の出版物と同様の扱いにしたいと考えます．過去のすべての報告集の論文について，著作権者（論文を執筆された故人の相続人）を探し出して利用許諾に関する同意をいただくことは困難ですので，一定期間の権利者捜索の努力をしたうえで，著作権者が見つからない場合も論文を情報学広場に掲載させていただきたいと思います．その後，著作権者が発見され，情報学広場への掲載の継続に同意が得られなかった場合には，当該論文については，掲載を停止いたします．

この措置にご意見のある方は，プログラミング・シンポジウムの辻尚史運営委員長 (tsuji@math.s.chiba-u.ac.jp) までお申し出ください．

加えて，著作権者について情報をお持ちの方は事務局（jigyo@ipsj.or.jp）まで情報をお寄せくださいますようお願い申し上げます．

情報処理学会著作権規程
https://www.ipsj.or.jp/copyright/ronbun/copyright.html



● 論文誌ジャーナル掲載論文リスト

Vol.63 No.1（Jan. 2022）

【特集：社会課題を解決するコラボレーション技術とネットワークサービス】

■ 特集「社会課題を解決するコラボレーション技術とネットワークサービス」の編集にあたって　市川裕介
■ オンラインの学会発表におけるプレゼンテーションスタイルの印象評価　越後宏紀 他
■ 会話の流れの可視化によるビデオ会議への効果　今井　廉 他
■ 実空間の点群情報を用いた空間接続表現の提案　本信敏学 他
■ 新型コロナウイルス感染症流行時における Twitter 上の流言訂正情報に関する分析　平林（宮部）真衣 他
■ Baseless-Rumor Alert Bot to Promote Reliability of Information　Ryota Nishimura 他
■ University-level Mathematics Pre-enrollment Education Combining Individual and Group Works in a Perfectly Distributed Asynchronous Environment †　Yuki Hirai 他
■ タッチ操作ログに基づいた操作形態推定手法　平部裕子 他
■ 客動線分析のための ID-POS データを用いたエージェントシミュレーションシステムの提案　中村綾乃 他
■ 施設管理支援に向けた常時型人流予測　角田啓介 他
■ カラー管理された電力需給における高効率な電力利用を可能とする充放電および送電管理方式　鈴木敏明 他
■ 単語分散表現による類義語統一と単語 N-gram によるフレーズ抽出に基づくセキュリティ要件分類手法　宮崎智己 他
■ 作業手順内の行為の目的を表出し構造化する方法の提案　－介護現場での目的指向知識構造化－　伊集院幸輝 他
■ 高齢者の自立支援介護における遠隔技術を用いた知識・データ融合の実践と分析　吉田康行 他

【特集：ニューノーマル時代の高度交通システムとパーベイシブシステム】

■ 特集「ニューノーマル時代の高度交通システムとパーベイシブシステム」の編集にあたって　清原良三
■ Congestion Control Algorithms for Collective Perception in Vehicular Networks　Susumu Ishihara 他
■ 拡張 NTMobile を用いたアプリケーションレベルで実現するシームレス IP Flow Mobility　松岡　穂 他
■ 人口統計データを用いた高需要時の飲食店需要予測　篠田謙司 他
■ Accurate and Efficient Driving Intention Inference Based on Traffic Environment Information and FES-XGB Framework　Shuo Wang 他
■ Traffic Prediction During Large-Scale Events Based on Pattern-Aware Regression　Takafumi Okukubo 他
■ 自動走行車両の進行方向提示と搭乗者の安心感の関係性調査　坂村祐希 他
■ レーン別渋滞検知技術の提案とフィールド実験への適用評価　森　皓平 他
■ Carrying-Mode Free Indoor Positioning Using Smartphone and Smartwatch and Its Evaluations　Tomoya Wakaizumi 他

【一般論文】

■ CBR-ACE : Counting Human Exercise Using Wi-Fi Beamforming Reports　Sorachi Kato 他
■ 分散協調型の電波電力伝送における位相最適化アルゴリズム　林健太朗 他
■ iBeacon を用いた位置推定における人体の影響による誤差の軽減 *　宮崎喬行 他
■ A Proof of Work Based on Preimage Problem of Variants of SHA-3 with ASIC Resistance*　Takaki Asanuma 他
■ マネージドセキュリティサービスのための受動的なログを用いたネットワーク構成情報検証方法 *　上川先之 他
■ Knowledge Graph Attention Network に基づく購買行動分析モデルに関する一考察　伊藤史世 他
■ 機械学習アプローチに基づく中古ファッションアイテムの価格保持期間適正化モデルの提案と実証的効果検証　桑田　和 他
■ 対話型顧客アクターによるマルチモーダル接客訓練 VR システム　古野友也 他
■ 仮面劇のためのプロトタイピングが容易な動的外見拡張手法 †　増井元康 他
■ なげれる君：ボウリング初心者のための投球フォームリフレクション支援アプリケーションの設計と実装 †　浦谷成敏 他

*：推薦論文　Recommended Paper
†：テクニカルノート　Technical Note

● 論文誌トランザクション掲載論文リスト

（Jan. 2022）

【論文誌 プログラミング Vol.15 No.1】

■ Automatic Optimize-time Validation for Binary Optimizers　Motohiro Kawahito 他
■ RL78 マイコン向けCコンパイラにおける procedural abstraction の実装　千葉雄司 他
■ Scalar Replacement Considering Branch Divergence　Junji Fukuhara 他

【論文誌 データベース Vol.15 No.1】

■ 実空間のユーザ行動分析に基づく潜在的興味分析方式　大村貴信 他

【Transactions on Bioinformatics Vol.15】

■ A Novel Metagenomic Binning Framework Using NLP Techniques in Feature Extraction　Viet Toan Tran 他

【論文誌 デジタルプラクティス Vol.3 No.1】

■ レガシーシステム移行時の性能劣化を改善するリファクタリング支援手法の提案　岡田譲二 他
■ Performance Evaluation of High Availability Database Systems Using Low-latency I/O Device　Shinji Fujiwara 他

■ 空調機ソフトウェアを対象としたSPLE開発におけるブランチ・マージプロセスの改善と考察　長峯　基 他

■ SFC GO：学生同士の繋がりを支援するオンライン体育授業サポートシステム　佐々木航 他

【論文誌 数理モデル化と応用 Vol.15 No.1】

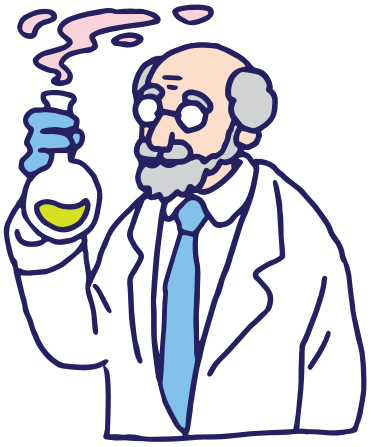
■ ナッジ理論を用いたコミュニティ活動活性化モデル　木村隆大 他

■ Modeling Imperfect Information TANHINMIN with Structural Oracle　Hironori Kiya 他

■ Accelerating the Numerical Computation of Positive Roots of Polynomials Using Suitable Combination of Lower Bounds　Masami Takata 他

【論文誌 コンシューマ・デバイス＆システム Vol.12 No.1】

■ 現場作業でのドキュメント閲覧に適したスマートグラス用ジェスチャ操作UI　尾崎友哉 他



◎ IPSJ カレンダー◎

学会イベントの最新情報を下記URLでご案内しています. 新型コロナウイルス感染症拡大を受け, 開催方法の変更, 開催中止などの可能性がありますので, 最新情報をご確認いただきますようお願いいたします.

https://www.ipsj.or.jp/calendar.html

…▸ 石井一夫（公立諏訪東京理科大学）

データサイエンス入門 教養としてのデータサイエンス

北川源四郎，竹村彰通 編
内田誠一，川崎能典，孝忠大輔，佐久間淳，椎名洋，中川裕志，樋口知之，丸山　宏 著
講談社サイエンティフィク（2021），1,980 円（税込），240p., ISBN：978-4-06-523809-7

モデルカリキュラムリテラシーレベルのテキストとして

本書籍は，「数理・データサイエンス・AI（リテラシーレベル）モデルカリキュラム」（以下，モデルカリキュラム）に準拠し，認定が進められている「数理・データサイエンス・AI 教育プログラム認定制度（リテラシーレベル）」のテキストとして使用することを目的としたものであると考えられる．本書籍の構成は，モデルカリキュラムと完全に一致しており，第 1 章［導入］社会におけるデータ・AI 利活用，第 2 章［基礎］データリテラシー，第 3 章［心得］データ・AI 利活用における留意事項の 3 章構成となっている．

モデルカリキュラムにおいてリテラシーレベルは，文系理系共通の導入科目の位置づけであると思われる．本書籍においても，統計学を学ぶにしても，プログラミングを学ぶにしても，何かを学ぶというよりは，ガイダンス的な内容で広く浅く説明している．したがって，本書籍を読めばデータサイエンスとは何か，どういう内容か，それで何ができるのかをさわりだけ触れることはできるが，データ分析ができるようになることは難しい．一方で，書籍そのものは易しいので，さらっと読んで概要をつかみ，興味が出てきた項目についてはほかの資料を読み進む本と考えるとよい．

社会におけるデータ・AI 利活用

「1.1 社会で起きている変化」，「1.2 社会で活用されているデータ」，「1.3 データと AI の活用領域」，「1.4 データ・AI 利活用のための技術」，「1.5 データ・AI 活用の現場」，「1.6 データ・AI 利活用の最新動向」と，モデルカリキュラムの項目に沿って，キーワードの解説がされている．ビッグデータや AI の利活用促進に基づく第 4 次産業革命や Society 5.0 による社会の変化が紹介されており，そこで利用されているデータや応用分野，応用技術についての最新動向の，用語も含めた解説が試みられている．モデルカリキュラムがまとめられたのがコロナ禍の前のことであり，最近ますます問題が大きくなっている気候変動・地球温暖化をはじめとする SDGs（Sustainable Development Goals）については，本書籍では触れられておらず，テレワークを中心としたニューノーマルな生活様式や，脱炭素社会に向けたデータサイエンスの在り方についても触れられていない．しかし，これらは，従来のものづくりを中心とした産業や生計の在り方や，シンギュラリティなど AI と人間との関係にも再考を強いるものであり，今後考慮する必要があると思われる．

データリテラシー

第 2 章は，「2.1 データを読む」，「2.2 データを説明する」，「2.3 データを扱う」という，実際にデータと接する局面での対応について，データとのかかわりを基に書きつづられている．まず，「2.1 データを読む」では，データの種類，データの分布と代表値，ばらつきと誤差の扱い，相関と因果関係，母集団と標本抽出，データセットの扱いや統計の考え方など統計の基礎に関する項目について説明がなされている．もっとも，数式的な扱いなどは触れられておらず，雲をつ

かむような感じがするのは否めない．統計学についてよく理解したい方は，ここをベースにより専門の書籍を手にすることをお勧めする．「2.2 データを説明する」では，グラフ（ヒストグラム，棒グラフ，円グラフ，散布図，ヒートマップなど）を用いた表現とその解釈，取り扱いについて述べられている．「2.3 データを扱う」では，表形式のデータ，すなわちスプレッドシートのデータの説明とその扱い，たとえば，データの集計や，並べ替え，ランキング，散布図の描き方などの説明である．いわゆるアプリケーションソフトを使ってデータを眺めたり，スプレッドシートを使った簡単な処理について，要点を説明するにとどめている．分析者の使う統計ツールには R や Python があるが，本章では R が要点だけ紹介されている．本章だけでは，R や Python を使いこなすというレベルには到底到達し得ないので，これらのツールを使いこなしたい方は，より専門的な書籍に当たる必要がある．

データ・AI 利活用における留意事項

第 3 章は，「3.1 データ・AI を扱う上での留意事項」，「3.2 データを守る上での留意事項」からなる．「3.1 データ・AI を扱う上での留意事項」は，まず，ELSI（Ethical, Legal and Social Issues），すなわち，倫理，法律，社会的側面から，データを扱う上での留意事項について述べている．次に，データの品質保証の問題を取り上げている．具体的には，AI のブラックボックス化，説明可能性，アカウンタビリティ，透明性，公平性などの問題である．データの品質保証の問題については，データの正確性を担保する上でも重要な事項であり，注意を払う必要がある．これらは，最近の AI の運用上でも話題になっている事項である．「3.2 データを守る上での留意事項」は，データサイエンス上の情報セキュリティとプライバシーの問題を取り上げている．情報セキュリティについては，機密性，完全性，可用性の 3 点から述べられており，プライバシーについては，プライバシーの定義と匿名化，匿名加工の問題について述べられている．特に，ビッグデータを扱う場合には，患者データや顧客データなど個人情報に配慮しなければならない場面は増えており，今後ますます重要性が増してくると思われる．

どういう人に薦めるか

AI，ビッグデータ，データサイエンスで何が話題になっているかを知って，これから何を学ぶか決めたい人や，データサイエンスの全容をざっくりとつかみたいという人，特に高校生や，大学生の初学年の人には向いている．また，「数理・データサイエンス・AI（リテラシーレベル）モデルカリキュラム」や，「数理・データサイエンス・AI 教育プログラム認定制度（リテラシーレベル）」の活用や運営にあたって，どのようなことが教えられているかを知りたい大学教員や，企業の人事担当者，データ分析部門の担当者などにお薦めする．しかし，これだけで，データサイエンスを学んだというには，あまりに薄い内容であるので，より高度な知識を身に付けたい学生には，ここをベースにほかの書籍を当たって知識を深めてほしい．

追記）本記事の執筆後，データサイエンスに大きなトレンドが見られるので，強調しておきたい．本文中でも少し触れているが，SDGs で表されるような社会的重要課題の解決への手段と指針としてのデータサイエンスの重要性が高まっていることである．具体的には，コロナ禍や，気候変動，少子高齢化などの地球規模や国家規模の課題へどう対峙していくかという側面である．これは，人類の叡智と行動力が試される厳しい選択を迫られるものになりそうである．

（2021 年 8 月 9 日受付）

石井一夫（正会員）
kishii@rs.sus.ac.jp
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・血管内科部門客員准教授．専門分野：ビッグデータ分析，計算機統計学，データマイニング，数理モデリング，機械学習，人工知能．医療ビッグデータ，気象ビッグデータ研究に従事．2015 年度本会優秀教育賞受賞．日本技術士会フェロー，APEC エンジニア，IPEA 国際エンジニア．

連載

5分で分かる!? 有名論文ナナメ読み

内田公太（サイボウズ・ラボ）

Lars Ole Andersen : Program Analysis and Specialization for the C Programming Language

PhD thesis, DIKU, University of Copenhagen (1994)

プログラム解析と最適化

プログラムの振る舞いを調べることは，バグ発見や最適化にとって重要な前処理である．配列の添え字として使われる変数に配列の長さ以上の数値が代入される可能性を発見できれば，バッファオーバーフローの防止に役立つ．特定のプログラム行において，ある変数の値が定数だということが分かれば事前に式を計算できる．本稿で紹介するポインタ解析は，たとえばリスト1でポインタ変数p，q，fpが指すオブジェクトをそれぞれ特定する手法である．

リスト1：解析対象のコード

```
int main(void) {
  int x, y, *p, **q, (*fp)(char *, char *);
  p = &x;
  q = &p;
  *q = &y;
  fp = &strcmp;
}
```

本論文の主テーマはC言語で書かれたプログラムを最適化することである．背景には製品試作から納品までの乖離がある．顧客要求を満たす試作品を作るにはLispのようなリッチな言語（論文執筆当時）を用いる．しかしLispは動作が非効率なため，納品用にはC言語で書き直したいが，試作品に盛り込んださまざまな要件を漏らさずに手作業で移植するのは大変な作業である．そこでLispをC言語に自動変換したいが，当時の技術で自動変換したC言語プログラムは，職人が手書きしたものよりずっと非効率だった．

そこで，C言語プログラムを自動的に最適化しよう，というのが本論文の主旨である．Generating extension generatorなど，最適化のための面白いアイディアがたくさん詰まっている．だが，引用されるのはもっぱら「ポインタ解析のアルゴリズム」である．ポインタ解析の話題は論文の1つの章を構成するに過ぎないのだが，それだけ色あせない画期的なアルゴリズムだったということだろう．

本論文はポインタ解析手法の基礎を築き，今でもプログラムの静的解析手法の文脈で引用される．ポインタ解析の結果を利用し，さらに高度なプログラム解析を行う，というような使われ方をする．著者名をとって「Andersen's analysis」や「APA（Andersen's Pointer Analysis）」などと呼ばれることが多いようだ．

Andersenのポインタ解析

プログラム解析の中でも，ポインタ解析はそのほかの解析の前提となる基礎的な情報を与える重要な役割を持つ．ポインタ解析はポインタ変数が指し得るオブジェクトを求める．「ポインタ」という名前であるが，C言語のポインタだけではなく，Javaの参照など，メモリ上のオブジェクトを参照するような機能を持つプログラミング言語に適用できる．

ポインタ変数pが変数xを指し得るということを，APAでは[p↦{x}]と表記する．これをpoints-toマップと呼ぶ．各ポインタ変数についてこのマップを求めることがAPAの目的である．リスト1における解析結

果は [p ↦{x,y}, q ↦{p}, fp ↦{strcmp}] である．

プログラムの各点においてポインタ変数が指すオブジェクトが変化する可能性がある．リスト1において，*q = &y を実行する前では p は x を指すが，代入後では p は y を指す．このように，プログラムの各点（あるいは関心のある点）に固有の情報を計算する手法はフローセンシティブである．一方，APA はフローインセンシティブな手法であり，p に対し唯一の points-to マップ [p ↦{x,y}] を出力する．C 言語のプログラムは小さな関数からなるケースが多く，変数に関して解析結果をまとめてしまってもそれほど問題にならず，それよりもフローセンシティブな解析に必要となる記憶域のコストの方が問題である，と本論文は主張する．

リスト2：ポインタを受け取る関数

```
int *inc_ptr(int *p) {
  return p + 1;
}
```

APA は関数の呼び出し文脈を区別する．たとえばリスト2のように定義した関数 inc_ptr があったとき，inc_ptr(a) と inc_ptr(b) を考える．a，b はポインタである．関数の呼び出し文脈を区別しない解析では2つの呼び出しはマージされ，inc_ptr は「a または b を指すポインタ」を返すと判定されてしまう．APA ではそれぞれの文脈を区別し，inc_ptr(a) は a だけを指し得ると判定できる．このおかげで，関数の呼び出し元で誤検出（本来指すはずがないオブジェクトを指す可能性を検出する）を抑制でき，ポインタ解析の結果を利用するほかのプログラム解析の精度が向上する．

ポインタ解析の苦労

C 言語には整数とポインタを変換する機能がある．たとえば (int*)0xabcd という変換が可能だ．ただ，この結果は処理系定義であるので，アーキテクチャ非依存の解析を目指す APA では結果を Unknown，すなわち「どのオブジェクトも指し得るポインタ」として，安全側に倒して扱う．

また，C 言語は分割コンパイルが可能である．関数はソースコード外から呼ばれ得るし，グローバル変数はソースコード外で書き換わり得る．すなわち，コンパイラが通常行うようにソースファイル1つを見ただけでは引数の値を特定できない．そこで，APA では関連するすべてのソースファイルをまとめて解析し，高精度な結果を得る．

本論文は現実の C 言語に真っ向から向き合うため，各文法に対応した泥臭い定式化を行うなど，大変な労力が注ぎ込まれている．また，本論文の主テーマだがあまり引用されない話題であるプログラムの自動変換も面白い．書式文字列をソースコードだと思えば，printf はインタプリタと見なせる，というのは目からうろこであった．本論文の公開から 28 年ほどの間にプログラム解析は大きく進展している．ただ，本論文を引用する新しい解析手法の多くは APA が出力する結果を基礎情報として用い，さらに高度な解析を行うものが多い．言語処理系基盤である LLVM に CFLAndersAliasAnalysis.h というファイルがあるなど，APA は時代を超えて利用されている．プログラム解析や自動変換などに興味があれば読んでみてはいかがだろうか．

(2021 年 10 月 22 日受付)

内田公太
kota-uchida@labs.cybozu.co.jp
2014 年に東京工業大学大学院の計算工学専攻修士課程を卒業し，サイボウズへ入社．SRE として 6 年間，インフラ基盤のソフトウェア開発に従事に従事し，2020 年にサイボウズ・ラボへ転籍．高校時代から趣味で OS を作り続け，2021 年 3 月に『ゼロからの OS 自作入門』を出版．

Info-WorkPlace 委員会企画「お届け *Info*」

今年度もやります！全国大会の“デリバリー”

坊農真弓 | Info-WorkPlace 委員会委員長／国立情報学研究所

お届け***Info*** とは，学会で発表されたできたてほやほやの情報や知識を，有志の取材をしてくださる会員（以下，デリバリー会員）がその場の臨場感とともに，学会会員の皆さま（以下，カスタマー会員）のお手元にお届けするサービスです．家族のイベントで参加できない，家族の病気やさまざまな事情があって参加できない……．このようなカスタマー会員のアクセシビリティを確保します．デリバリータイプは 3 種類を予定しています．

●**ビデオデリバリー**：発表録画を配信します（3 日以内に届きます．50 件募集）

●**速報メールデリバリー**：メールで発表のサマリーを送ります（3 日以内に届きます．5 件募集）

● **note 記事デリバリー**：ブログ形式で発表概要を掲載します（3 カ月以内に掲載します．5 件募集）

デザイン：木塚あゆみ

※全国大会参加者のオーダーは無料で承ります．
※デリバリーは抽選とさせていただきます．抽選結果はデリバリーの発送をもってかえさせていただきます．

申込締切 2 月 22 日

詳細・お申込みはこちら

https://forms.gle/nYFqE8fnN5hLx6Tm8

お届け*Info* 使ってみた！

出産・育児で大忙し編

漫画：ヤシ

デリバリー会員さんの知識を拝借！編

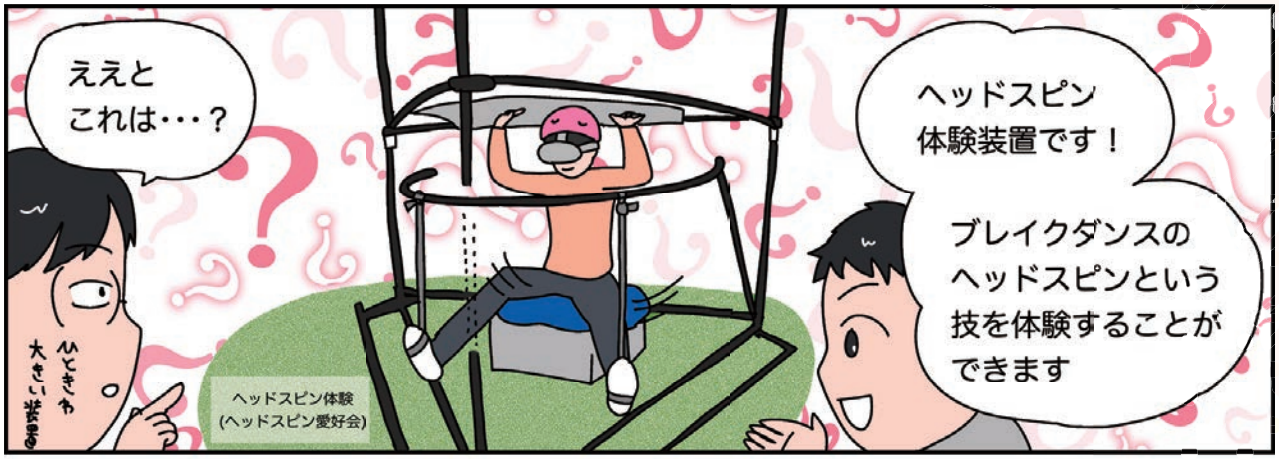
録画ビデオだけ送ってもらお！編

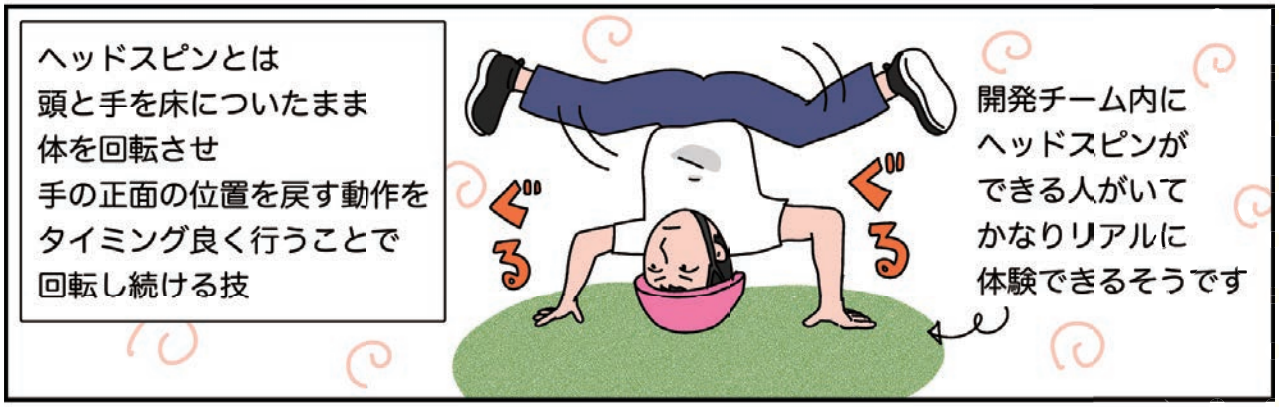
漫画：nam

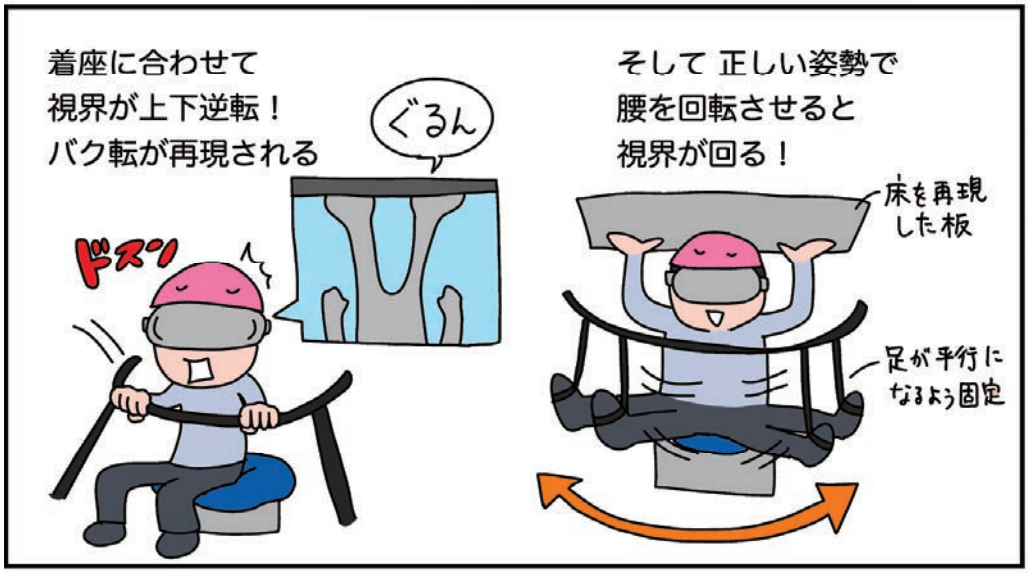
その16 VR作品の登竜門IVRCに行ってみた！

漫画：山本ゆうか（Twitter @ymmox）

※1 詳しい募集要項はこちらをご覧ください https://ivrc.net/2021/callforchallenges/


ええと
これは・・・？
ひときわ大きい装置
ヘッドスピン体験
(ヘッドスピン愛好会)
ヘッドスピン
体験装置です！
ブレイクダンスの
ヘッドスピンという
技を体験することが
できます


ヘッドスピンとは
頭と手を床についたまま
体を回転させ
手の正面の位置を戻す動作を
タイミング良く行うことで
回転し続ける技
ぐる
ぐる
開発チーム内に
ヘッドスピンが
できる人がいて
かなりリアルに
体験できるそうです


着座に合わせて
視界が上下逆転！
バク転が再現される
ぐるん
ドズン
そして 正しい姿勢で
腰を回転させると
視界が回る！
床を再現した板
足が平行になるよう固定


手を離す
タイミングが
難しい！


稲見編集長
私がヘッドスピンを
する世界線が
あるなんて!!
ぐる
ぐる
めちゃめちゃ
撮影されて
いました


ゲスト審査員の
わくわくさんもかなり
楽しまれたようです
わくわくさんが
やってるとこ
見たかった!!
アハハハ
ヒィ〜!


「転生したらコウモリだった件」
(バットマン)
あーーー！
いくつかのチームで見かけた扇風機
声を出して
壁を避けていくゲーム
顎だけで操作するのが
印象的でした！
ゲームオーバーで
顎が落ちる仕掛けに
驚かされた
コウモリの落下をイメージしたそうです
ガクン
肩の動きだけで
レースゲームが
できるんだ！
操作性を良くするために
チュートリアルと同時に
キャリブレーションを
行っているそうです
肩り手
(肩操作)
肩を回して
シールドを出す
操作
げっ コントローラーに
ウイルスが！
ウイルス可視化したよ！
フランスからの招待作品でした
Rebirth-20
(Arts et Métiers)
ゴリゴリゴリ
締切を
ぶっ壊す！
締め切り
言葉を壊す
手ごたえが
気持ちイイ!!
ストレスをぶっ壊す！
(ストレスから国民を守る党)
わんこ～
なでなで～♡
犬が来ている
Meta Dog
(TKKK)
大阪会場の
体験審査会にも
参加したかったな～
人類山化計画
(人類山化委員会)で
山になりたかった
炒飯道
(NEXT ゆめくじら)
はデバイスがすごく
凝っていた！
いるの？いないの？
ゆら・・・
未然たるあなた
(チーム有明)は
どんな気分になるのか
予想がつかない
楽しんで
ますか？
大阪芸術大学
安藤 英由樹先生
はい！
個性的な
作品ばかり
ですね！

取材してほしい人・イベントなどを募集中！ Twitter ハッシュタグ：#IPSJ_IT 紀行／メールアドレス：editj@ipsj.or.jp

「データの分析」分野の入試問題の分類と解法の一考察　入試センターのサンプル問題解説～第3問データの分析～

♡ 7

情報処理学会・学会誌「情報処理」
2021年10月20日 10:21

阿部百合（早稲田大学高等学院）

連載「教科「情報」の入学試験問題って？」第4弾です．前回までの入試問題解説に関してはこちら（https://note.com/ipsj/n/n81737ef872ec）をご覧ください．

大学入試センターによる共通テスト『情報』のサンプル問題[1] の大問3を見てみましょう．

（以下，サンプル問題中の記号や文字を**太字**で表現しています）

目次

問題設定
問1
問2
問3
問4
まとめ

問題設定

実際の状況や，生徒に身近なデータが使われています．今回の問題では実際のサッカーワールドカップのデータ [2]を使って決勝進出チームと予選敗退チームの違いを分析します．問に入る前に2つの分析資料（**表1**，**図1**）が提示されています．データの分析問題で重要，かつ見落としがちな点に生データの数値変換や補完があります．図表にまず目がいきがちですが，資料の前後にある説明文をしっかりと読

み「図表中の数値（図示されている場合は点やグラフ）の意味を理解する」ことが肝要です．数値の意味について書かれている部分を赤線で示しました．グラフの場合は軸やラベルに注目し，単位も必ず見ます．入試問題演習では，数値の意味の部分を自身でチェックするとよいでしょう．

決勝進出チームと予選敗退チームの違いを調べるために，決勝進出の有無は，決勝進出であれば 1，予選敗退であれば 0 とした。また，チームごとに試合数が異なるので，各項目を 1 試合当たりの数値に変換した。

表 1　ある年のサッカーのワールドカップのデータの一部（データシート）

	A	B	C	D	E	F	G	H	I	J	K
1	チーム ID	試合数	総得点	ショートパス本数	ロングパス本数	反則回数	決勝進出の有無	試合当たりの得点	1 試合当たりのショートパス本数	1 試合当たりのロングパス本数	1 試合当たりの反則回数
2	T01	3	1	834	328	5	0	0.33	278.00	109.33	1.67
3	T02	5	11	1923	510	12	1	2.20	384.60	102.00	2.40
4	T03	3	1	650	269	11	0	0.33	216.67	89.67	3.67
5	T04	7	12	2257	711	11	1	1.71	322.43	101.57	1.57
6	T05	3	2	741	234	8	0	0.67	247.00	78.00	2.67
7	T06	5	5	1600	555	9	1	1.00	320.00	111.00	1.80

▲問の前の資料**表1**（赤線は筆者追加）

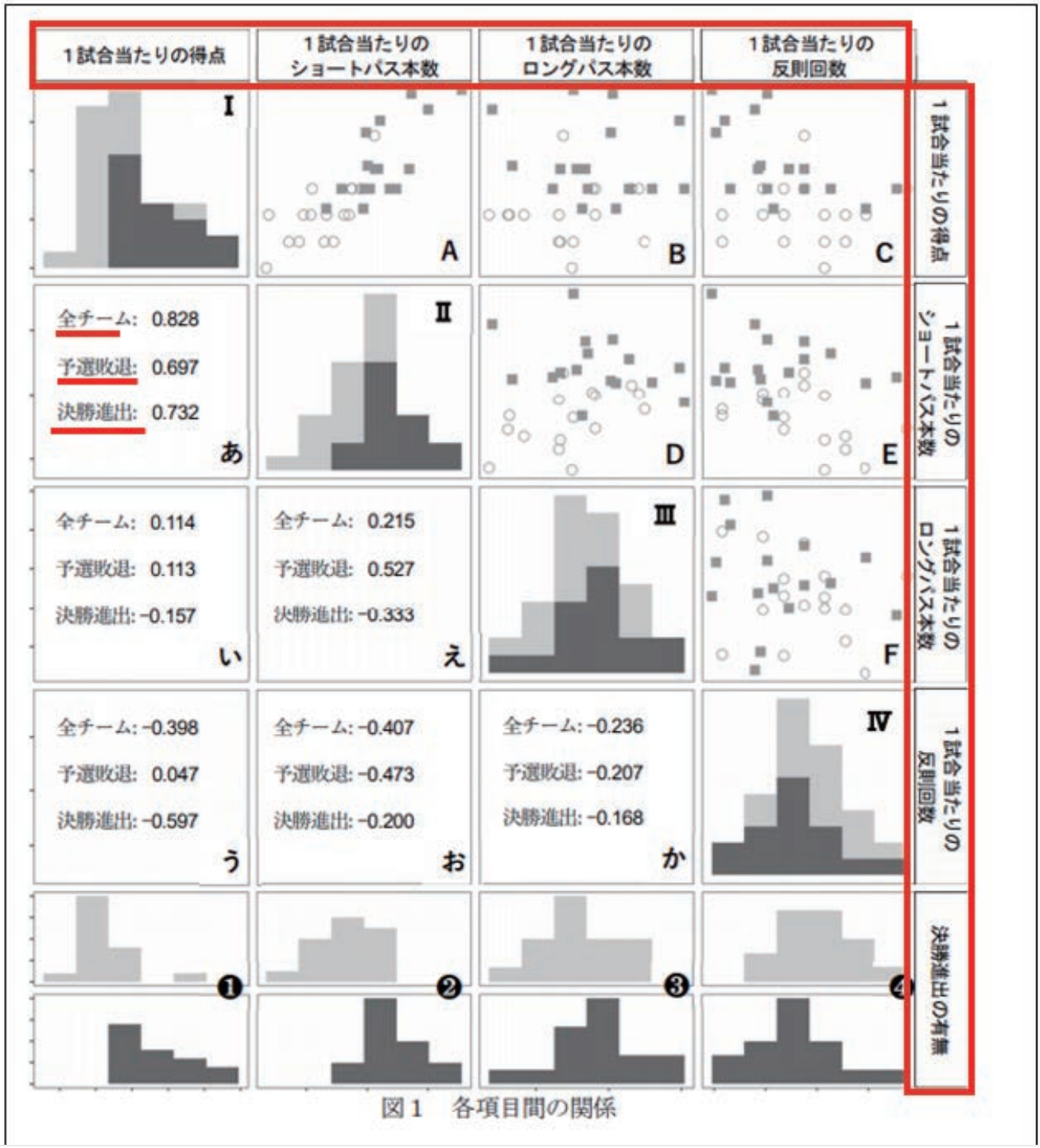


図1　各項目間の関係

▲問の前の資料**図1**（赤線は筆者追加）

問1

資料**図1**を読めるか問う問題です．ここでは散布図行列がプログラムで作れるかは問題ではありません．重要なのは「図を読めること」です．2次元行列の形に面食らった方もいるでしょう．図の下に以下の説明がありますが，かえって読みづら

いかもしれません．むしろ一度でも**図1**のような行列（総当たり戦の対戦表など）や散布図行列を扱った経験があれば，**図1**は読めます．この機会に演習しておくと安心です．

図1のⅠ～Ⅳは，それぞれの項目の全参加チームのヒストグラムを決勝進出チームと予選敗退チームとで色分けしたものであり，❶～❹は決勝進出チームと予選敗退チームに分けて作成したヒストグラムである。**あ**～**か**は，それぞれの二つの項目の全参加チームと決勝進出チーム，予選敗退チームのそれぞれに限定した相関係数である。また A～F は，それぞれの二つの項目の散布図を決勝進出チームと予選敗退チームをマークで区別して描いている。例えば，図1の A は縦軸を「1試合当たりの得点」，横軸を「1試合当たりのショートパス本数」とした散布図であり，それに対応した相関係数は**あ**で表されている。

▲問の前の資料**図1**の下にある説明文

a問題

a問題では文章中に問いが隠れています．赤い線の部分が問いです．

図1を見ると，予選敗退チームにおいてはほとんど相関がないが，決勝進出チームについて負の相関がある項目の組合せは，1試合当たりの［ア］と［イ］である。また，決勝進出チームと予選敗退チームとで，相関係数の符号が逆符号であり，その差が最も大きくなっている関係を表している散布図は［ウ］である。したがって，散布図の二つの記号のどちらが決勝進出チームを表しているかが分かった。

［ア］・［イ］の解答群

⓪ 得点　① ショートパス本数　② ロングパス本数　③ 反則回数

［ウ］の解答群

⓪ A　① B　② C　③ D　④ E　⑤ F

アと**イ**の組合せは，**図1**の**あ**～**か**の予選敗退と決勝進出の数値を比較すれば**⓪得点**と**③反則回数**であるとすぐに解けます．「相関」の資料が**I（ヒストグラム）・A（散布図）・あ（相関係数）**のどの形式か分かっているかを，問われています．（図-1.1）

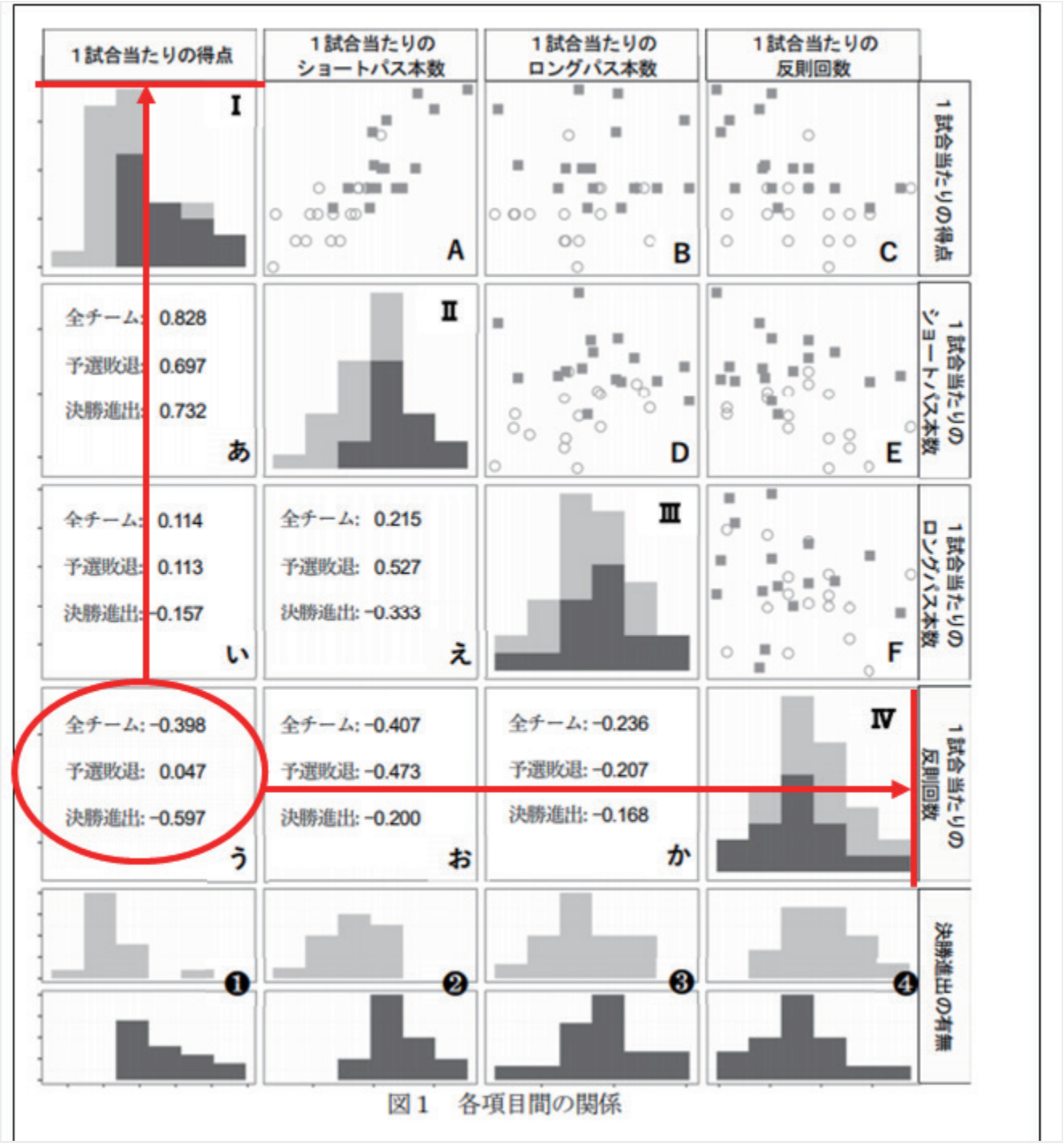


図１　各項目間の関係

▲図-1.1　問の前の資料**図1**からの読み取り1（赤線部分）

アと**イ**の答え(順序問わず)

⓪得点

③反則回数

ウは，問題文に釣られて散布図を見ても解けません．まずは**あ**～**か**の相関係数を見て，対応する散布図が答えとなります．総当たり戦の対戦表同様，左上から右下への対角線が対称軸となって，データの組合せが同じです．（図-1.2）

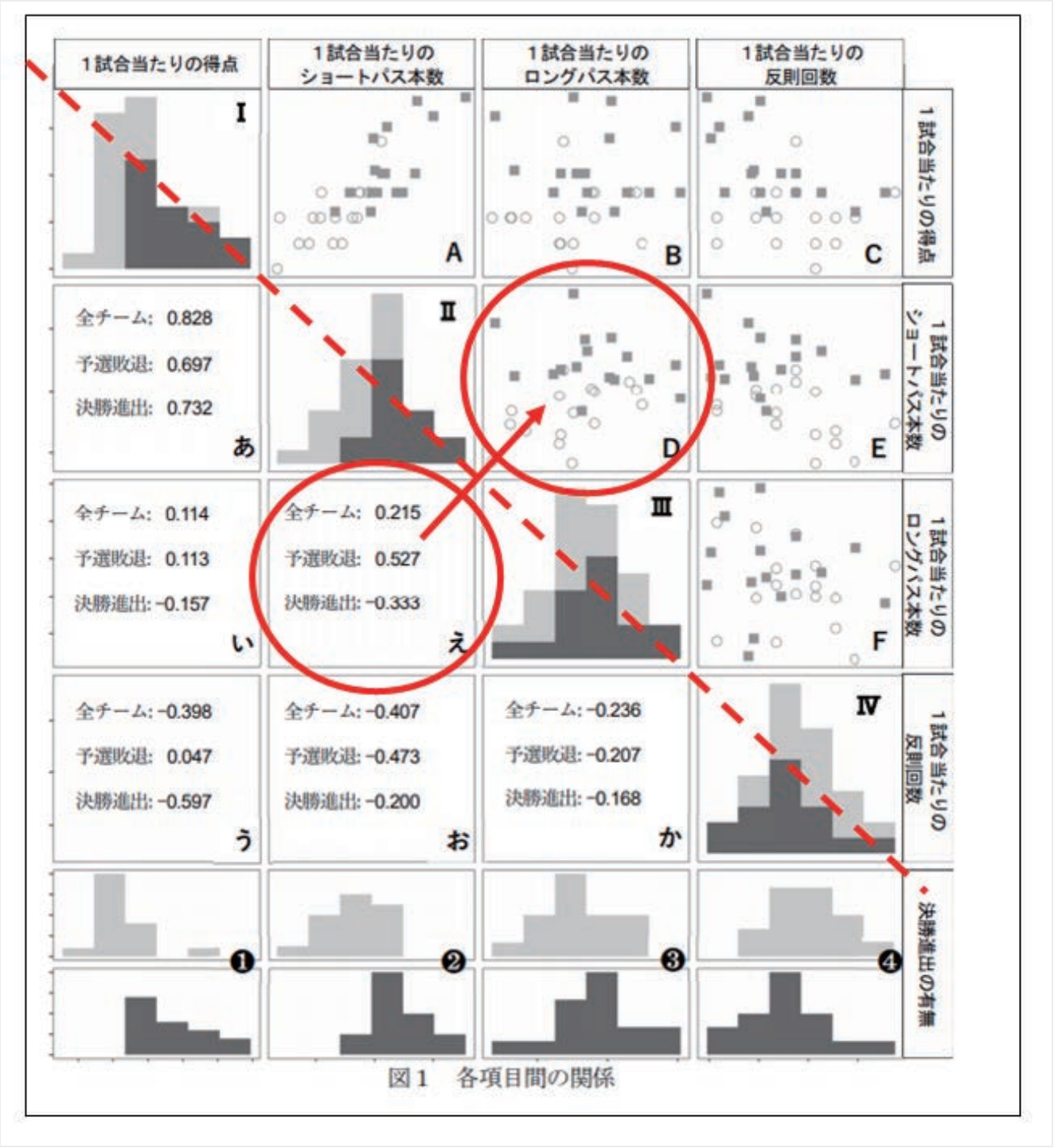


図1　各項目間の関係

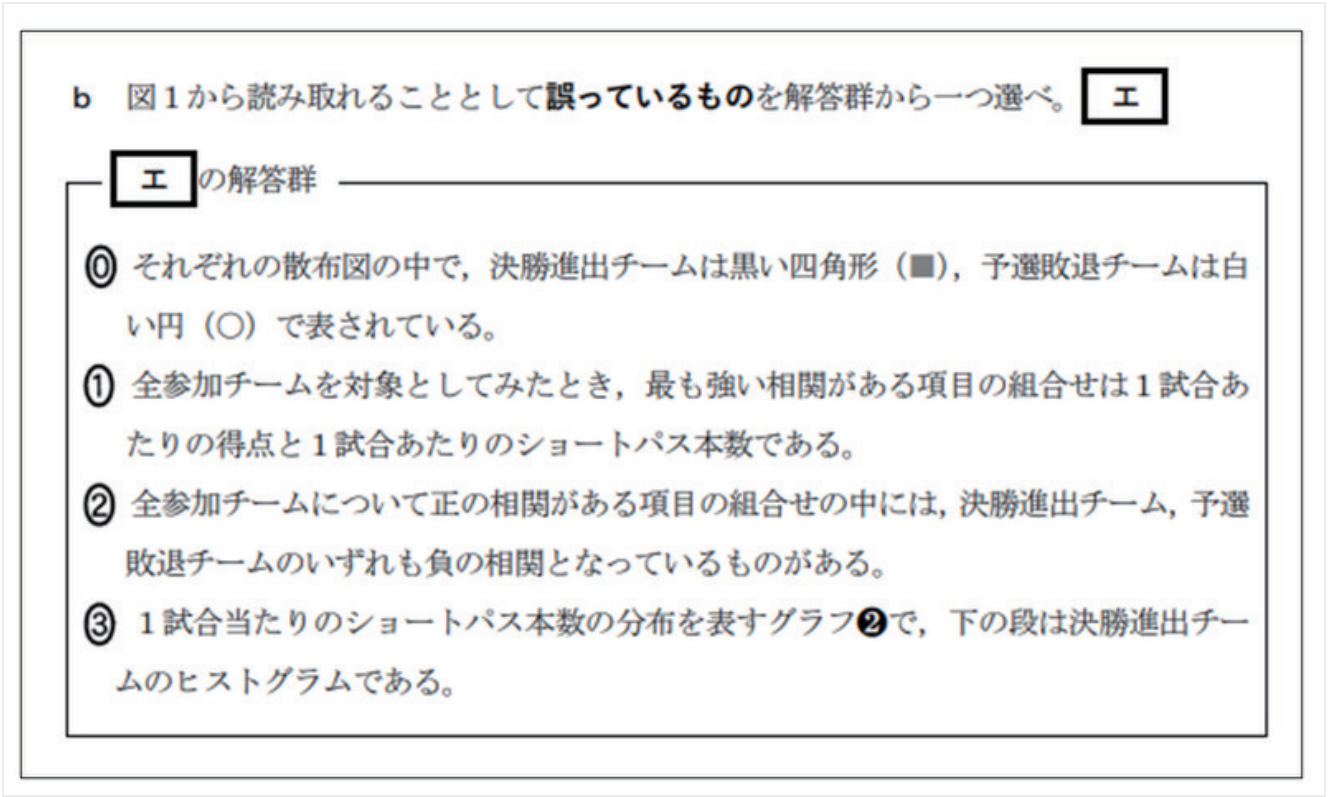
▲図-1.2　問の前の資料**図1**からの読み取り２（赤線部分）

ウの答え

③ D

b問題

b問題では，選択肢から誤っているものを選びます.


b　図１から読み取れることとして**誤っているもの**を解答群から一つ選べ。 エ

エ の解答群

⓪ それぞれの散布図の中で，決勝進出チームは黒い四角形（■），予選敗退チームは白い円（○）で表されている。

① 全参加チームを対象としてみたとき，最も強い相関がある項目の組合せは１試合あたりの得点と１試合あたりのショートパス本数である。

② 全参加チームについて正の相関がある項目の組合せの中には，決勝進出チーム，予選敗退チームのいずれも負の相関となっているものがある。

③ １試合当たりのショートパス本数の分布を表すグラフ❷で，下の段は決勝進出チームのヒストグラムである。

定石では資料**図1**をもとに各選択肢の正誤を判定しますが，選択肢すべてに先に目を通すと疑わしい選択肢②「全体の相関が正だが，個別の相関が２つとも負」があるため，まずは選択肢②を確認します．資料**図1**の**あ**～**か**の相関係数の符号だけ取り出してみます.

	あ	い	う	え	お	か	選択肢②
全チーム	+	+	−	+	−	−	+
予選敗退	+	+	+	+	−	−	−
決勝進出	+	−	−	−	−	−	−

全チームが正で予選敗退と決勝進出両方が負のものはなく，他の選択肢を確認せずとも答えは②です。

エの答え

②

問2

鈴木さんは，この結果からショートパス 100 本につき，1 試合当たりの得点増加数を決勝進出チームと予選敗退チームで比べた場合，0. **オカ** 点の差があり，ショートパスの数に対する得点の増加量は決勝進出チームの方が大きいと考えた。

また，1 試合当たりのショートパスが 320 本のとき，回帰直線から予測できる得点の差は，決勝進出チームと予選敗退チームで，小数第 3 位を四捨五入して計算すると，0.0 **キ** 点の差があることが分かった。鈴木さんは，グラフからは傾きに大きな差が見られないこの二つの回帰直線について，実際に計算してみると差を見つけられることが実感できた。

さらに，ある決勝進出チームは，1 試合当たりのショートパス本数が 384.2 本で，1 試合当たりの得点が 2.20 点であったが，実際の 1 試合当たりの得点と回帰直線による予測値との差は，小数第 3 位を四捨五入した値で 0. **クケ** 点であった。

問2では回帰直線の理解，計算力を確認しています．資料**図2**にある回帰直線の式に値を代入し計算します．このあたりは情報Iの授業で扱われるので点が取れる問題となるはずです．問1のa問題同様，問2も問題文の中に問いが含まれる形式です．

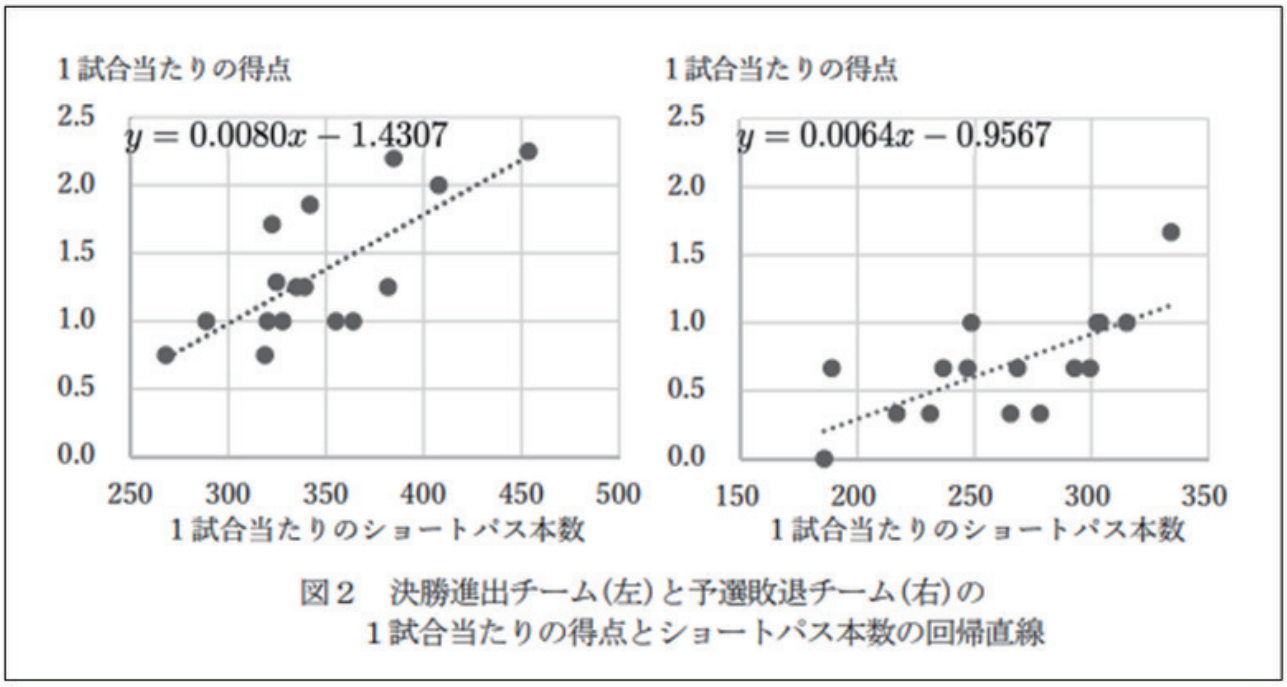


図2　決勝進出チーム(左)と予選敗退チーム(右)の
1試合当たりの得点とショートパス本数の回帰直線

▲サンプル問題の問2の資料**図2**

オカ：資料**図2**の決勝進出チーム（左）と予選敗退チーム（右）のショートパス100本につき，1試合あたりの得点増加数の差を聞いています.

単純にxに100を代入して計算しますが「増加数」を聞かれているので，直線の傾き（つまりxの前の数）を100倍して比較すれば十分です.

よって，0.8-0.64=0.16，答えは0.16です.

オカの答え

16

キ：問題文の誘導通り，1試合あたりのショートパスが320本のときを計算します.

x=320を代入して計算すると，決勝進出チーム1.1293，予選敗退チーム1.0913よって差は0.0380で小数第3位を四捨五入し答えは0.04です.

キの答え

4

クケ：決勝進出チームの実際の得点と回帰直線による予測得点の誤差の確認です．図中の式のｙが予測得点であることを理解していれば，計算は単純です．**図2**の左の式ｙ=0.0080ｘ−1.4307にｘ=384.2を代入した値と，実際の得点2.20の差を求め，小数第3位を四捨五入すると答えは0.56点です．

クケの答え

56

問3

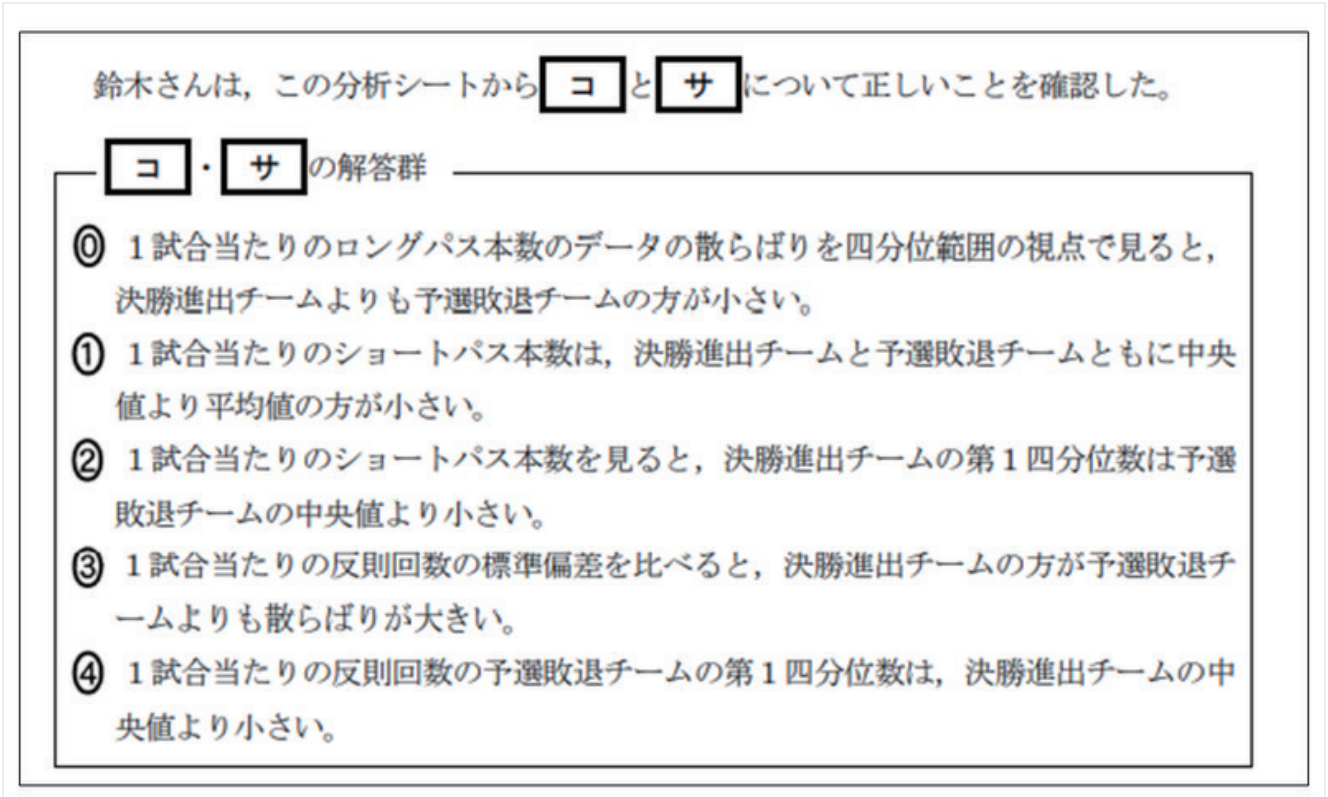


表2　1試合当たりのデータに関する基本的な統計量（分析シート）

	A	B	C	D	E	F	G	H	I
1		決勝進出チーム				予選敗退チーム			
2	統計量	1試合当たりの得点	1試合当たりのショートパス本数	1試合当たりのロングパス本数	1試合当たりの反則回数	1試合当たりの得点	1試合当たりのショートパス本数	1試合当たりのロングパス本数	1試合当たりの反則回数
3	合計	21.56	5532.21	1564.19	41.30	11.00	4213.33	1474.33	48.00
4	最小値	0.75	268.00	74.40	1.50	0.00	185.67	73.67	1.67
5	第1四分位数	1.00	321.82	92.25	2.10	0.33	235.25	87.67	2.58
6	第2四分位数	1.25	336.88	96.02	2.40	0.67	266.83	91.67	3.00
7	第3四分位数	1.75	368.33	103.50	3.00	1.00	300.08	98.00	3.42
8	最大値	2.25	453.50	118.40	4.50	1.67	334.00	109.33	4.67
9	分散	0.23	1926.74	137.79	0.67	0.15	1824.08	106.61	0.61
10	標準偏差	0.48	43.89	11.74	0.82	0.38	42.71	10.33	0.78
11	平均値	1.35	345.76	97.76	2.58	0.69	263.33	92.15	3.00

▲サンプル問題の問3の**表2**

表2の考察問題です．2つ選ぶ必要があるため定石通り選択肢を1つずつ吟味します。中高の数学で学んだ統計を理解していれば難しくないでしょう．

問3の答え

⓪と③

問4

複数の資料，情報から必要な情報を見つけ，答えを導く問題です．クロス集計表が読めることが大前提です．クロス集計表は数学でも条件付き確率で同様の表を使っており，見たことがあるでしょう．

シ：

鈴木さんは，作成した図１と表２の両方から，［シ］ことに気づき，決勝進出の有無と１試合当たりの反則回数の関係に着目した。そこで，全参加チームにおける１試合当たりの反則回数の第１四分位数（Q1）未満のもの，第３四分位数（Q3）を超えるもの，Q1 以上 Q3 以下の範囲のものの三つに分け，それと決勝進出の有無で，次の表３のクロス集計表に全参加チームを分類した。ただし，※の箇所は値を隠してある。

選択肢の吟味は先頭からする必要はありません．確認しやすい選択肢から見ます．選択肢③は，どの資料を見るか書いてあり，しかも読み取るものはヒストグラムと視覚的です．今回はちょうど見やすい選択肢が正解でした．

シの答え

③

ス：

表３　決勝進出の有無と１試合当たりの反則回数に基づくクロス集計表

	1試合当たりの反則回数			
	Q1 未満	Q1 以上 Q3 以下	Q3 を超える	計
決勝進出チーム	※	※	※	16
予選敗退チーム	2	※	［ス］	16
全参加チーム	8	※	7	32

▲サンプル問題の問4の**表3**

スの値は**表3**の中の数字から求められないので，続く問題文を見ます．

この表から，決勝進出チームと予選敗退チームの傾向が異なることに気づいた鈴木さんは，割合に着目してみようと考えた。決勝進出チームのうち1試合当たりの反則回数が全参加チームにおける第3四分位数を超えるチームの割合は約19%であった。また，1試合当たりの反則回数がその第1四分位数より小さいチームの中で決勝進出したチームの割合は**セソ**%であった。

「決勝進出チームのうち1試合当たりの反則回数が全参加チームにおける第３四分位数を超えるチームの割合は19%」とあります．決勝進出チームの内訳は表の1段目です．決勝進出チームの全数は1段目右端の16，「決勝進出チームの第3四分位数を超えるチーム数」は表の1段目右から2列目スの上の※です．つまり（**ス**の上の※の数字）÷16×100=19ということです．この方程式を解くと**ス**の上の※の数は3と分かります．よって表の**Q3**を超える列に注目して7-3=4です．

スの答え

4

セソ：「第1四分位数より小さいチームの中で」と問題文にあるので左端の「**Q1未満**」の列を見ます．左端の列は上から※，**2**，**8**です．よって※は8-2で6と分かるので，その割合は6÷8×100＝75，75%です．

セソの答え

75

まとめ

ここまで「設定に沿って解くタイプの問題」を見てきましたが，慶應義塾大学や明治大学では「設定や分析の不適切な点を考えさせる問題」「調査・分析の修正を

記述させる問題」も出されています[3].

「設定に沿って解くタイプの問題」は，分析データや問題設定にやや不自然さを感じることがあっても割り切って筋書きに乗って解きましょう．ただし，ただ誘導に乗っていればいいのではありません．情報分野を目指す場合や，より難易度の高い大学を目指す場合，「設定や分析の不適切な点を考えさせる問題」，「調査・分析の修正を記述させる問題」に対応できる必要があります．今回の問題でも，予選敗退チームは当然ながら決勝進出チームより試合数が少ない（つまりチームによって試合数が大きく異なる）可能性がある，データサイズが十分でない可能性がある，といったことがあります．常に問題の設定や，「扱われているデータに不適切な点はないか」「どのようにしたら適切か」などの視点から注意して眺める癖も付けたいものです．問題を見るポイントは，

①大問全体の流れ

②データ，図表の軸，単位，ラベル，前後の問題文といった細かいところ

を見ることです．

情報の入試問題演習では，

①長い問題文や複数の資料を読むことに慣れること

②試験時間に焦らずじっくりと問題と向き合うこと

が必要です．

今回扱った問題は試験時間を考慮していない[4]とのことですが，もともとセンター試験『数学I』『情報関係基礎』でのデータの活用分野は問題文や図表が多いです．文章中に問いを含む形式もあり，問題文を読んで聞かれていることを理解する速さも必要であり，ある程度解けるようになってきたら

③制限時間を決めて解く練習

をするとよいでしょう．

また，問題を見て知らない形式の図表で見方が分からない場合や分からない単語が出てきたときは，いったん問題から離れて分からない部分を調べ理解することが基本となります．

参考文献

1）大学入試センター，共通テスト『情報』サンプル問題，https://www.dnc.ac.jp/kyotsu/shiken_jouhou/r7ikou.html

2）科学の道具箱，FIFAワールドカップ2006データの集計，https://rika-net.com/contents/cp0530/contents/04-14-01.html

3）情報入試研究会　資料，2018年度明治大学情報コミュニケーション学部問IV他，http://jnsg.jp/?page_id=108

4）全国高等学校情報教育研究会第14回大会(大阪オンライン)基調講演「大学入学共通テスト　新科目「情報」～サンプル問題等とそのねらい～」（2021-8-11），https://www.zenkojoken.jp/14osaka/2021053913/，FIT2021公開シンポジウム大学共通テスト「情報」が目指すもの（2021-8-26），https://www.ipsj.or.jp/event/fit/fit2021/FIT2021_program/data/html/event/pdf/eventB2_347.pdf

（2021年10月11日受付）
（2021年10月20日note公開）

■阿部百合（正会員)

都内私立高校の情報の授業を担当している．本会CE研委員．

IPSJMooc（https://sites.google.com/view/ipsjmooc/）第4章作成に携わった．

情報処理学会ジュニア会員へのお誘い

小中高校生，高専生本科～専攻科1年，大学学部１～3年生の皆さんは，情報処理学

会に**無料で入会**できます．**会員になると有料記事の閲覧，情報処理を学べるさまざまなイベントにお得に参加できる等のメリットがあります**．ぜひ，入会をご検討ください．入会は**こちら**から！

CASE4：私のオフィスはオンラインに引っ越しました

♡ 5

情報処理学会・学会誌「情報処理」
2021年11月18日 20:48

伊東　香（ヤフー（株）サイエンス統括本部 産学連携推進室）

ヤフーでは，2020年2月より段階的にリモートワークでの勤務に移行し，10月には正式に制度化しました．セキュリティの関係や物理的な問題でオフィスでの業務は一部残っていますが，9割以上の社員がオフィスに出社せず仕事ができる環境が実現しています．

現在に至るまでにさまざまなオンラインツールを試したり，どこでもオフィス手当と通信費補助で最大月9,000円の補助が受けられたり，会社の組織制度も大きく変化しました．今回はコミュニケーションを活性化するためのちょっとした工夫などをご紹介いたします．

目次

リモートワークへの移行
オンラインへの引っ越し
コミュニケーションツールや制度
オンラインコミュニケーションの工夫
今後の働き方

リモートワークへの移行

ヤフーにはコロナによる緊急事態宣言の6年前である2014年からオフィス以外の好きな場所で働ける「どこでもオフィス」というリモートワークの制度がありました．月に5回までオフィス以外のどこからでも働いてよいという制度です．社員の創造性や働き方の自由度を高めることが目的で大規模アクセスに耐え得るVPN接続

の環境構築はこのタイミングで始まっています．

オンラインへの引っ越し

「ヤフー株式会社は働く共通の環境をオンラインに引っ越して，オンラインからサービスを提供していきます．」2020年7月にこのようなプレスリリースをいたしました．実際にリモートワークへの全面移行により我々の環境は大きく変化することとなります．

※プレスリリース　https://about.yahoo.co.jp/pr/release/2020/07/15a/

コミュニケーションツールや制度

社内のコミュニケーションツールはZoomやSlackが主ですが，チームやプロジェクトによりさまざまなツールを利用しています．バーチャル空間でアバターを動かすツールや，オンラインチャット上でアバター同士が吹き出しで会話するような

ものなどさまざまなツールが試されています．すべてのコミュニケーションをオンラインに移行することは未知のチャレンジであり，今なお常に新たな工夫や試行錯誤が続いています．一例をご紹介します．

・全社朝礼

月に一度開催される全社朝礼では自宅等のネットワーク品質がさまざまである状況を踏まえ，さまざまなツールで参加することができます．ストリーミング配信は（映像＋音声），（音声のみ）の2種，Slackによるテキスト＋資料の実況中継，時間が合わない人向けにアーカイブ動画，資料の共有等がされています．また質問ツールを使ったオンライン質疑応答に対する「いいね！」ボタン，「あんまり」ボタンを実装するなどリアルタイムでの参加感が得られるように工夫されています．

・組織内イベントや開発業務

先日，私の所属する部門では新入社員による発表会がありましたが，Zoom上でニコニコ動画風のテロップを流し応援して参加するなど双方向感が感じられる工夫がされていました．

コロナ前からエンジニアによるもくもく会等はZoom上でもSlackでもさまざまなツールで開催されていましたが，Gather.Town等の仮想空間上で集まるものが最近は人気のようです．複数のツールをかけ合わせるなど，さまざまな創意工夫がされています．開発業務の現場でもモブプログラミングを導入し実践したりしています．

※モブプログラミングの導入と実践

https://techblog.yahoo.co.jp/entry/2021083030177162/

・社内懇親会

社内コミュニケーション促進施策の1つとして社内懇親会が挙げられます．社内のカフェテリアからオンライン懇親会セットを各自宅あてにデリバリーしてもらうことも可能で，参加者が同じものを食べたり飲んだり，同じ時間を共有する仕組みがあります．私のチームではこのZoom懇親会の際は全員が居酒屋風のバーチャル背景でまるで向かい合っているように工夫しています．

オンラインコミュニケーションの工夫

オンライン上だとどうしても相手の目線がずれるなど反応が分かりにくく，よく知っている相手でも話が盛り上がらない場合がありますが，相手に目線を合わせて相槌を多く入れることで会話が活性化しより自然に話ができるようになりました．Zoomでオフィシャルな会議をする際には［セルフビューを非表示］にして，自分の画面を消して会議に集中するようにしています．セルフビューがあると知らない間に自分の顔を見てしまうので，実際に会っているように自分の目線をコントロールするための工夫です．ギャラリービューで複数の人たちと会話する際には，一番

よく発言している人をギャラリービューで上部に移動させカメラに近い位置に動かして目線を上にしたりしています．オンライン飲み会でもこの小技を使うだけで，コミュニケーションがとりやすくなりました．

今後の働き方

社内では新たなコミュニケーションツールを開発したり，元々フリーアドレスのオフィス環境もさらに情報交換や交流がしやすいように変更したりしていく計画が進んでいます．コミュニケーションの在り方やオフィスの在り方はこれからもどんどん変化していきます．リモートワークに限らずオフィスワークにおいてもさまざまなコミュニケーション手法や新しい働き方の模索が常に続いています．情報技術と創意工夫でより働きやすい環境を，新たな世界を構築していけると考えています．

※「新しい働き方」に対応できるオフィスとは？
https://about.yahoo.co.jp/info/blog/20210325/yjoffice.html

（2021年9月7日受付）
（2021年11月18日note掲載）

■伊東　香（正会員）

ヤフー（株）サイエンス統括本部産学連携推進室に所属．AI人材の育成と女性エンジニア活躍の場を創出することに奔走中リモートワークが快適すぎて仕事の傍ら野菜作りにはまっています．

★働き方について，もっと考えたい人はこちら→『Info-WorkPlace』note

FIT2021 イベント企画「ヒトゲノム・生体情報と情報処理の課題」会議報告

情報処理学会・学会誌「情報処理」
2021年11月15日 09:02

金子　格（東北大学）

目次

開催の目的

ゲノム情報処理と国際標準化の取り組み

ヒトゲノム研究者から見たデータ利用の実際

犯罪×ゲノム・生体情報

～データ保護に関する国際的な議論から～

パネルディスカッション

開催の目的

2021年8月27日のFIT2021でイベント企画「ヒトゲノム・生体情報と情報処理の課題」を開催した☆**1**．ゲノム情報やその生体情報の利用は急速な技術進歩に支えられ，今後急拡大が予想される．このような時期，各方面の専門家に，それぞれの専門にとどまらない学際的な議論をいただく場を設けることは，ヒトゲノム・生体情報に関する情報処理の未来を展望するために有用だろうと考えられる．そこで，本会電子化知的財産・社会基盤研究会の協力を得て，国際標準化，科学研究，犯罪予防，法制度という4つの分野の専門家にご参集いただき，多方面からの討論を行った．学際的に大変面白い，未来志向のディスカッションが行えたので，その概要を紹介する．

ゲノム情報処理と国際標準化の取り組み

まず東北大学 金子（本稿筆者）からゲノム情報処理と標準化の状況を次のように紹介した．

「人間のDNA全配列読み取り（全DNA配列の読み取り）コストは2000年に100億ドル規模であったが，ムーアの法則をはるかに超えるスピードで高速化，低コスト化が進み，現在では個人が数万円程度で行える（**図-1**）．今後爆発的に利用が広がり，デジカメで顔写真を撮るのと同じ手軽さで利用できる時代はすぐそこに来ている．そのような時代に備え，MPEG標準化グループの中でもゲノム情報符号化伝送の標準化の取り組みも行われている」

このあと，国際標準は作成に5年程度，利用は30年程度続くこともあり，数十年先のニーズに応え得る将来の社会制度を見据えた規格作成が必要になることや，各国のプライバシー法制度などの法制度自体も発展途上であるため，標準自体は，こうした制度を幅広く調査しつつ，各国の異なる法制度や今後の社会の変化に適応できるよう，柔軟な規格を目指して開発が進められていると紹介された.

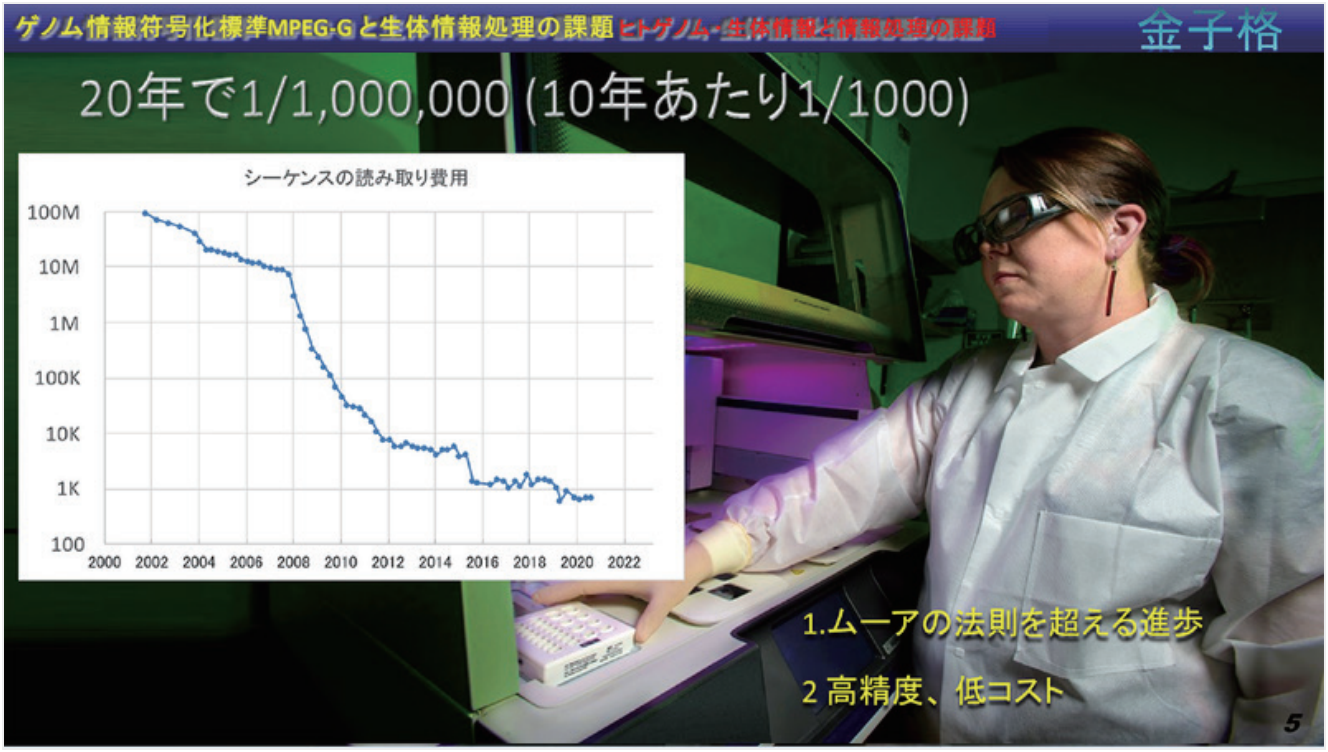


図-1

東北大学 金子「ゲノム情報処理と国際標準化の取り組み」のスライドから

20年で1/1,000,000（フルシーケンスの読み取り費用）

ヒトゲノム研究者から見たデータ利用の実際

東京大学 小金渕氏からは，ゲノム研究におけるゲノムデータ利用の概要が解説された.

次に，個人情報をどこまで守れるかについて，ゲノムからどういったことが分かるかという例を紹介いただいた.

「ゲノム情報を取得すれば大体どこの地域の出身かは分かる．多くの集団は特定の地域から大幅に移動することなく生活をしており，それらの集団間で婚姻がなされる．それを起因として生じた遺伝子流動の影響で遺伝子の特徴がグラデーションになってくるからだ．ゲノム情報を主成分分析で解析して，PC1（1番目のPrincipal Component）とPC2でプロットすると，遺伝的変異の分布と個人個人の地理的な分布がほぼ一致することが，2008年のNovembre等の論文で発表されている（**図-2**，右図）.

また，日本人にとってより身近な例として，下戸遺伝子である2型アルデヒド脱水素酵素（ALDH2（遺伝子はイタリック体で表記されるが、noteの仕様で通常書体としている．以下同様））を挙げる．たとえば，飲み会などですごい酒豪の人がいたら「もしかして九州出身？」などと推測するだろう．私がお酒に弱いタイプの遺伝子であるALDH2の日本列島内での頻度分布を調べた結果，実際のところ，九州や沖縄にはお酒に弱い（アルデヒド分解酵素が不活性化している）遺伝子型を持

っている人が，他の地域に比べて少なかった．これは，私たちの感覚と遺伝子の地理的分布が一致する例と言える（図-2，左図）．したがって，完全な匿名化は難しく，DNAからある程度の情報は得られてしまう」

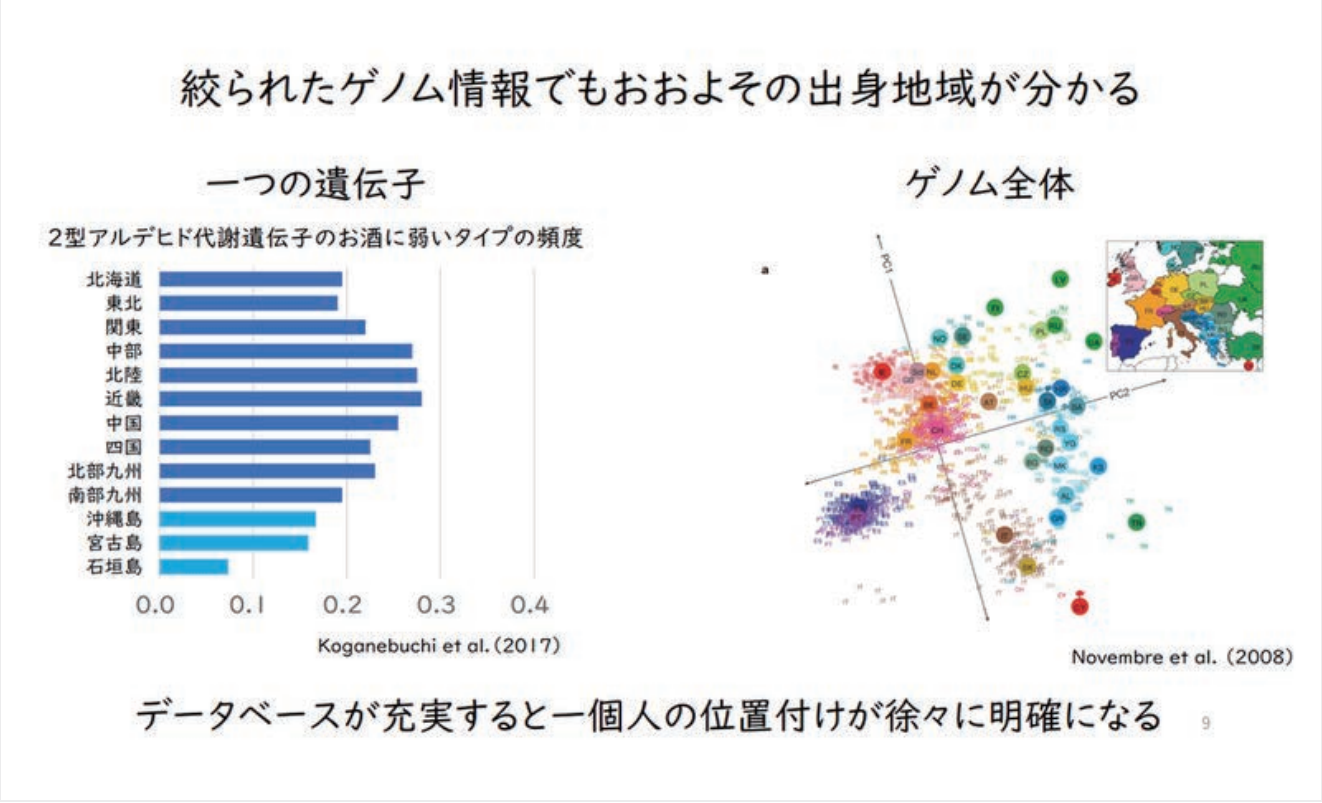


図-2

東京大学 小金渕「ヒトゲノム研究者から見たデータ利用の実際」のスライドから

遺伝子による個人の位置付け（左-小金渕等，右- Novembre等，による）

このあと，ヒトゲノム研究における検体収集はインフォームドコンセントと匿名

化に加え，データへのアクセス自体厳密な管理下で扱われていること，特定の人しか入れない閉ざされた場所やオフラインで分析している研究機関もあること，その一方で英国では登録によりオンラインで公開される例もあること，などが紹介された．

犯罪×ゲノム・生体情報

東北大学 荒井氏からは，犯罪とゲノム・生体情報という表題で，主に犯罪という文脈におけるゲノム利用についてご紹介いただいた．

「犯罪の一次予防（広報活動，環境設計により犯罪を起こさせないようにすること），二次予防（加害リスクを予見して早期介入すること），三次予防（問題の再発を防ぐこと）の考え方がある．ゲノム利用と関係が深いのは，特に二次予防である」

このあと，加害リスクを予見して早期に介入するためには，潜在的加害者を精度良く予測するための要因を明らかにする必要があり，近年では神経犯罪学（Neurocriminology）の台頭により，犯罪における遺伝的要因の重要性が再認識され始めていると言及された．具体的な例として，ドーパミン神経系遺伝子，酵素活性遺伝子，セロトニン遺伝子などが反社会的行動に関連し得ると報告されている（**図-3**）ことが挙げられた．ただし，こうした考え方の注意点として，遺伝的要因

が存在するだけで犯罪者になるのではなく，遺伝的要因と環境的要因が相互作用することで初めて犯罪という複雑な反社会的行動が発現する，という視点を忘れてはいけないことが指摘された．それに加えて，遺伝的要因が予測するのは衝動性や攻撃性であり，犯罪そのものの予測ではない点に留意が必要であることが示された．

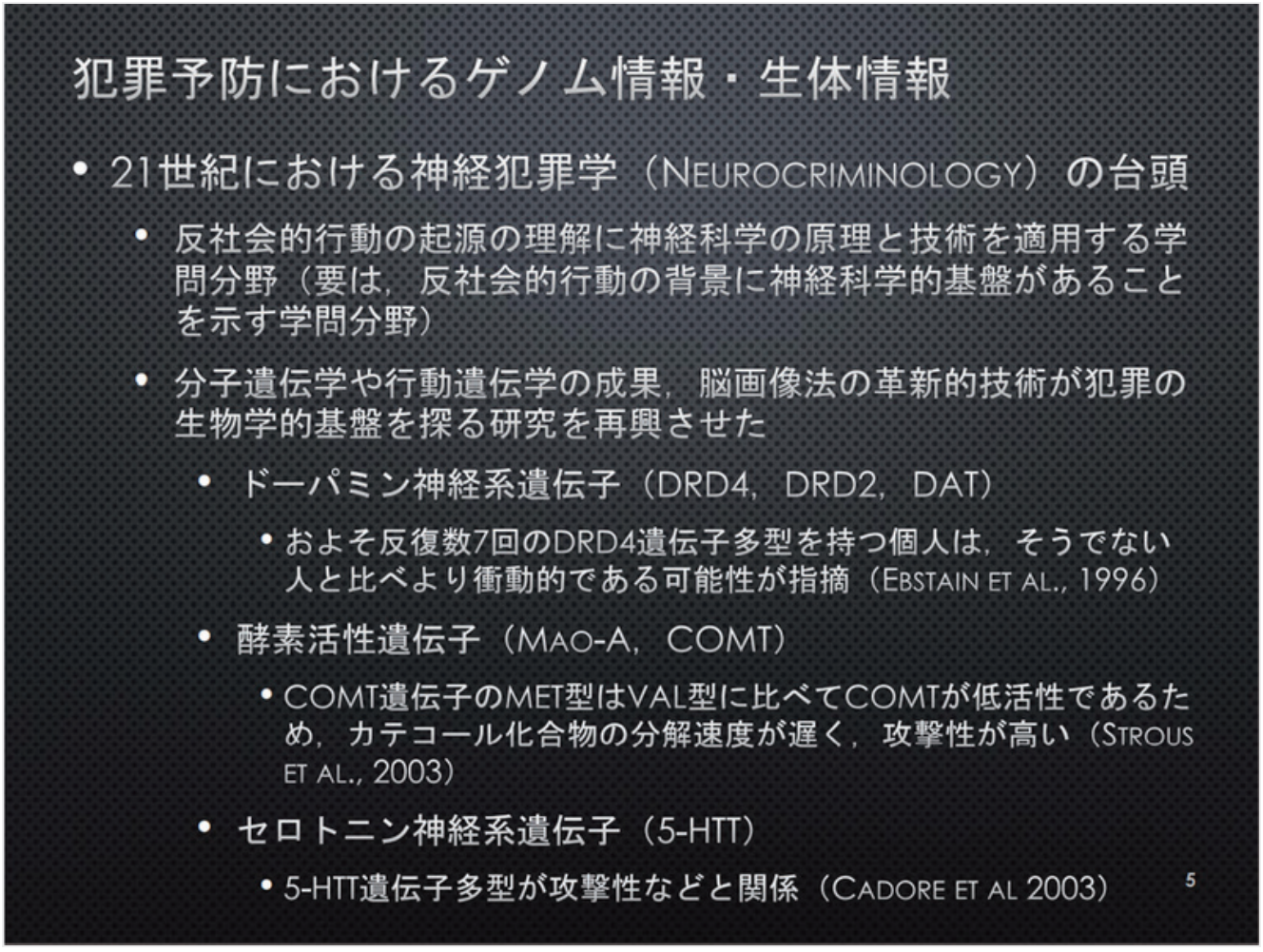


図-3

東北大学 荒井「犯罪×ゲノム・生体情報」のスライドから

後半では，日本の犯罪捜査におけるゲノム・生体情報の利用について紹介された．犯罪捜査におけるゲノム利用の問題点として，法的根拠が必ずしも明確ではない点や，警察がDNA型のデータベースを作成すること自体に法的・倫理的問題がないのかという点が提示された．

～データ保護に関する国際的な議論から～

KDDI総合研究所 加藤氏からは，ゲノムを始めとした生体情報に関するデータ保護法制の際的な議論を解説いただいた．

「日本では平成27年の改正における個人識別符号にDNAが含まれ，研究倫理指針においてゲノム情報における遺伝情報の定義なども示された（**図-4**）．

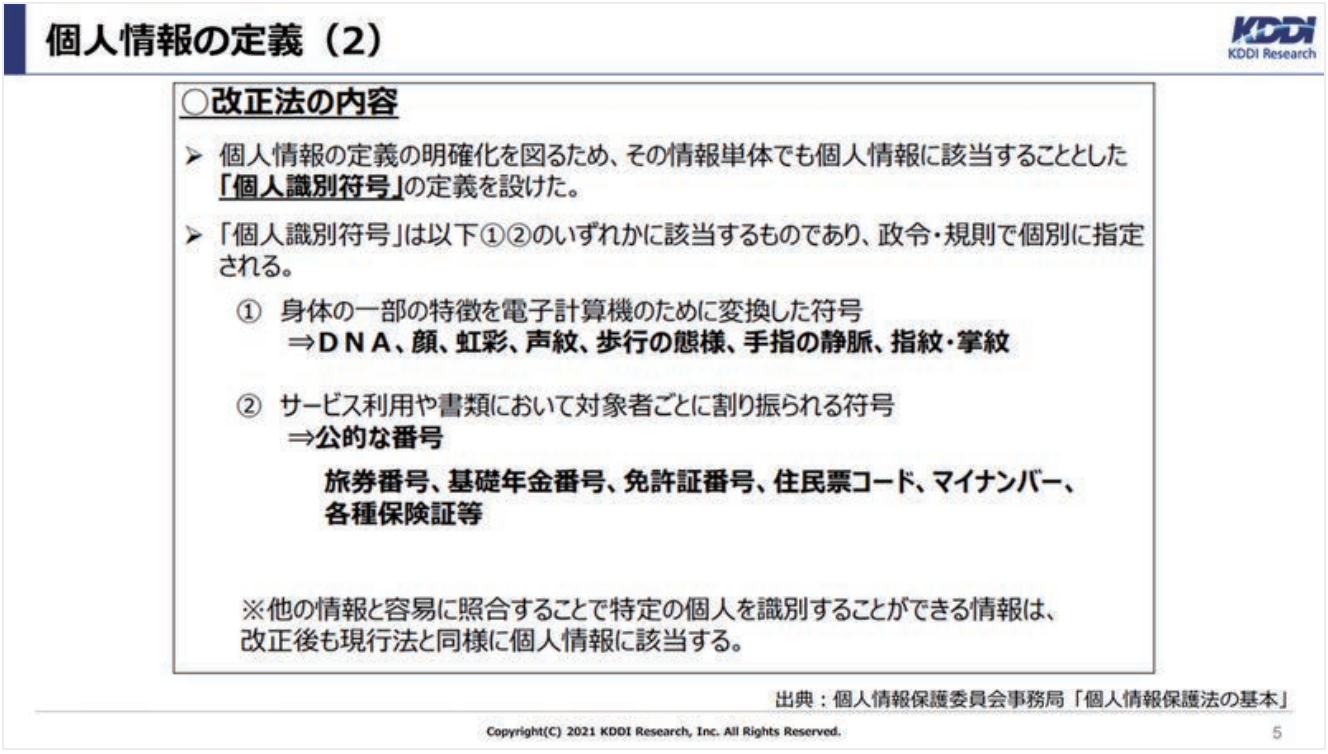


図-4

KDDI総合研究所 加藤 「～データ保護に関する国際的な議論から～」のスライドから

平成27年改正個人情報保護法

欧州委員会からは，2021年2月，GDPR（General Data Protection Regulation; EU一般データ保護規則）の下での健康情報に関する各加盟国の調査結果である，Assessment of the EU Member States' rules on health data in the light of GDPRが公開された．ここではCOVID-19の流行も踏まえて，国際連携の重要性が

認識されている」

また，欧州における医療データ共有の仕組みであるEHDS（European Health Data Space）も紹介され，ここではGDPRに沿った個人の基本的権利を尊重することなどが報告された.

パネルディスカッション

このあと，須川氏，湯田氏を司会として，いくつかの論点についてディスカッションを行った.

特異例（めずらしい症状，病気など）にかかわる多型と一般的な症状にかかわる多型の有用性について

小金渕：高地（低酸素）への適応や，イヌイットの肉食への適応などがあり，それが疾患の原因にもなることがあるので，こうした特徴的な多型が医学的に有用な場合がある．一般的な相関が有用な場合もある．特異例が個人特定につながらないかという点については，個人特定にかかわる多型は病気の多型とは別（病気の多型だけでは個人特定は困難）なので，個人特定の多型は分離，保護できると思う.

犯罪捜査のために本人の許可なくDNAが使われるか

荒井：日本では，令状がない限り，警察がDNAサンプルを本人の断りなく強制的に

取得することはできない．任意でDNAを採取し，それが違法だとして法律論争になった事例がある．

須川：米国において，DNAを登録することで家系図を作るサービスの情報を警察が犯罪捜査に使った事例がある．日本にはまだそのようなサービスがないが，そうしたサービスが将来できる場合には法整備が必要かもしれない．

DNAの大規模データベースについて

荒井：全日本人のDNAのデータベースが作られることがあり得るのか．

金子： 詳しくは知らないが，調べて目にするのはせいぜい10万人．

小金渕：企業で数万人のコホートを作る，政府がガンのゲノムデータベースを作るという話はあるが全員ということはない．

金子：フルシーケンス100円といった時代になった場合，他人に勝手にDNA分析，記録されることを阻止することが難しい．本人特定の多型を登録することで，DNA分析処理を行ったら必ず（本人特定を行い）本人に通知するという方が保護されるかもしれない．

犯罪分析について

湯田：不敬罪のような犯罪は時代により変わると思われる．それにどう対処するか．

荒井：そこまで正確には分からない．犯罪が遺伝で決まるわけではなく，遺伝的に攻撃しやすくなるというだけで，遺伝情報から犯罪がおきることは確定できない．

加藤：かつては不敬罪があったが今はない.

大規模なDNAデータベースを利用することについて

加藤：ホロコーストにおいてパンチカードシステムが利用された．大量のデータを情報処理技術で処理できたことがホロコーストの悲劇につながった，という反省が欧米においては基本的な考え方としてある．遺伝子情報をはじめとして，大規模なデータベースを用いることには慎重であるべきと考えられる.

須川：情報がデジタルデータになったときに，さまざまな問題が起こり得る.

DNAが変更不可能であることについて

加藤：ID番号は漏洩したとしても変更する手段が残されている．ゲノムは漏洩した場合に書き換えられない．その点はどう考えるか.

金子：分析が容易になると，登録を慎重にしても保護しきれない．むしろ本人に断りなく他人が利用することを制限する方が実際的と思う.

湯田：メリットよりもデメリットが多いとなると登録しないということが最適解になってしまうので，メリットを出す制度設計が必要だ.

加藤：登録と利用をわける意味合いはなんだろう？　データ保護では取得・保管・提供はセットで考える．利用だけを取り出す意味というのはなにかあるのか.

金子：映画の著作権の場合，コピー保護は完全には阻止できない．その場合，コピー保護をしつつ，配布制限を併用する必要があるのでは.

このあと，会場の一般参加者からの質疑も行われ，本イベントを終了した.

☆1　イベントの概要　https://www.ipsj.or.jp/event/fit/fit2021/FIT2021_program/data/html/event/event_A2.html

（2021年10月15日受付）
（2021年11月15日note公開）

■金子　格（正会員）
アスキーの技術開発部，大学教員を経て，2021年から東北大学データ駆動科学・ai教育研究センター技術補佐員．情報技術，オーディオ，ゲノムなど幅広い関心を持つ．ACM，IEEE，日本音響学会，惑星協会会員.

今月の会員の広場では，11 月号へのご意見・ご感想を紹介いたします．

巻頭コラム「Changing the World」

■パソコンの研究での活用の歴史とそれに対する人の意識の変化についての知見が興味深い．（祖父江真一）

■コンピュータ普及以前は，使用を否定されていたというエピソードが印象的だった．現状の有機 EL についても解説してほしかった．（鈴木広人）

■夢のある話が現実のものとなるということが具体的に実感できて明るい未来を見ることができるようで少し幸せな気分になりよかったです．（匿名希望）

■研究と研究者の在り方を教えられます．有機 EL の開発時は研究者が白い目で見られたとは知りませんでした．（滝口　亨）

■コンピュータに対する夢や可能性を想起させられ，読後感の良いコラムでした．（金子雄介）

特別解説「暗号資産の現在と将来」

■資産的価値は最終的に大衆が行っていくことになると考えているため，将来性を不安視することも重要であるが，どういった形で今後価値が見出されると考えられるか，知見がほしかった．（印部太智）

■暗号資産への投資やビットコインに関係した企業の株価がアップするなど，話題の多い暗号資産の現状だが，その成り立ちに関しての情報は多いものの，今後どのように進んでいくかについてはなかなか方向性が見えない．そのような中，交渉のメカニズムに関して論じてくださったのはとてもありがたかった．（濱　久人）

■本誌に投資物件のような暗号資産の記事が掲載されること自体，非常に興味深く感じます．著者の「根源的価値は値上がり益」という見解は私も同感です．（滝口　亨）

特集「観光情報学」

「0．編集にあたって」

■ COVID-19 でオンライン観光が注目されるようになったが，観光情報学はほかにも病気などで観光できない人のためにもなる研究だと思いました．（匿名希望）

「1．ポストコロナにおける観光」

■前半でスマートツーリズムの 3 つの基本要素が挙げられており，そこに ICT が貢献できると書いてあり，良いことだと思った．後半で，スマートツーリズムへの ICT 利用の具体例が出てくることを期待したが，その具体例があまり出てこないのでガッカリした．（中島秀之）

「2．観光情報のオープンデータ化」

■各自治体が公開しているデータはバラバラのため統合して提案してくれるアプリがあると観光客にとって有益なサポートである．チェックイン機能による訪問ログや所在地から空き時間が〇分あるならこの場所がおすすめなどもサポートしてくれるとアプリに育ってくれると嬉しい．（匿名希望）

「3．UGC を利用した観光資源の発見と推薦」

■観光スポットの質，ルートの質，ユーザの嗜好，およびリアルタイムの人流を考慮した観光ルートの生成・推奨が可能となり，観光地巡りがより充実したものになることを期待する．（山下昭裕）

「4．参加型観光情報の収集」

■本イベントで投稿された情報を一般公開することでより観光の活性化につながると感じた．ゲーム感覚で楽しめて，参加得点があるため参加しやすい．本文で言及されている通り，アプリへのハードルがあるのではと思ったが幅広い年齢から参加しているのは意外だった．（匿名希望）

「5．人流クラスタリング解析」

■観光と人流予測を合わせることで，道路の渋滞予測と同じく今からその観光地に向かうと混んでいる可能性があるなど予測してくれて面白いと感じた．ただ，展望にあるように個人情報の保護とのバランスがハードルになると思う．（匿名希望）

「6．観光ナビゲーション」

■訪問する観光地が決まってからの推薦機能が充実してきているが，観光地が数ある中から選ぶのを助ける機能も，観光活動を再開していく段階では有用になるのではないか．（佐藤章博）

「7．観光のための動画キュレーション」

■メモリアル動画にドライブレコーダの映像を用いるという発想が面白いと思った（匿名希望／ジュニア会員）

「8．観光とチャットボット」

■ほとんどがチャットボットの解説になっており，観光特有の課題，実験を紹介していないように感じる．（匿名希望）

「9．観光客の心理状態推定」

■観光における心理状態推定はこれから重要な技術になって

くると感じました．どのように心理状態を収集するか，課題もありますが発展が期待できます．（岡本克也）

DP コーナー「DX のプラクティス」

「0. 編集にあたって」

■「ニューノーマル時代を生き延びる」という副題は，さまざまな苦労を抱えながら日々 DX 推進に臨んでいる現場担当者への熱いエールのようだ．（広野淳之）

「1.［解説論文］DX 先進企業から見る DX の現在地，構造，方向」

■日本の最優先課題の 1 つである DX の推進状況を，実際の調査結果をベースに分かりやすく整理されていました．DX 推進に関するさまざまな提案がされており，これらが各企業に伝わることにより，実効性のある DX 事例が増えればよいと思います．（後藤正宏）

「3.［招待論文］顔認証と Digital ID を活用したサービス社会の実現に向けて」

■ JR 東日本が昨年 7 月に顔認証機能付きの防犯カメラを導入して物議を醸したばかりであり，多数の市民にとって関心のあるテーマだった点や分かりやすい解説が素晴らしかった．（大塚敬義）

「4.［招待論文］事例から見る RPA 導入の課題とその解決」

■ RPA 導入における失敗（会社の恥？）を包み隠さず公表し，それを踏まえた考察なので，頷く点が多く，秀逸な論文だと思った．業務を知る現場社員の「カイゼン活動」として RPA を適用し，スモールスタートで始め，推進側が適用範囲をうまく制御するのが「コツ」だと読み取った．（小橋喜嗣）

「5. インタビュー：DX のプラクティス」

■この特集全体が DX に対する考え方で自分の中でぼやっとしていた部分をすっきりとさせてくれた印象ですが，その中で特にこのインタビューは私の中で曖昧だった点を直接説明してくれているようで，非常にためになる記事でした．（山本一公）

「［ユニシス研究会］新しい生活様式に適したセキュアなリッチクライアントの実装」

■ Windows10 および Microsoft365 の標準機能を備えたノート PC を使い，情報漏洩やセキュリティ対策を標準機能の活用で充足し，テレワークでのログオン認証を可能にする事例は，一般企業に有益と思われる．（匿名希望）

「［日立 IT ユーザ会］建設現場のデジタルシフト」

■建設現場で行われている DX 化の数々の事例について，具体的に紹介されているのが大変よかった．このように既存の要素技術を組み合わせて業界に特化したシステムを構築することが本当の意味での DX だと感じる．（匿名希望）

教育コーナー「ぺた語義」

「データサイエンスカリキュラム標準（専門教育レベル）の公開について」

■データサイエンスカリキュラム標準（専門教育レベル）を見ました．知識，スキル，態度別に目標が具体的で分かりやすいと感じました．情報処理学会で策定中のデータサイエンティスト資格の公表にも期待しています．（匿名希望）

「大学入学共通テスト「情報」サンプル問題を題材とした研究協議」

■大学入学共通テスト「情報」のサンプル問題や実施に向けての課題などを知ることができた．実際の問題点として，現場の情報科担当教員の負担をいかに軽減化するかは大切な問題であることが理解できた．（小西敏雄）

「オンライン授業導入の舞台裏」

■急遽，オンライン授業対応が迫られる中，迅速に準備を進め，さらに学生とインタラクティブに意見交換を行うことにより，さまざまな教育機関や企業が戸惑う中，高い満足度が得られるように設計されている点が興味深い．規模は違うが，今後同じような事態になったとき，あるいは別の事態であっても緊急で何かに対応する際の参考にしたい．（髙田崚介）

連載「情報の授業をしよう！：中学校技術科における双方向通信ネットワークおよび計測・制御の授業実践」

■「実践 1 と実践 2 ともに「学習のくくり」でどのような視点・考え方で学習し何を学習の獲得目標にするかが明確になっていて，その後の「ガイダンス」，「つかむ学習」，「追及する学習」，「つなげる学習」の流れと有機的につながっていると感じました．（松浦満夫）

連載「ビブリオ・トーク：ソフトウェア工学から学ぶ機械学習の品質問題」

■正解・不正解の判断だけではないということがわかりました．（くろやなぎゆうた／ジュニア会員）

■機械学習をしたことがない人たちが AI を簡単に操作したり，○○の業務はすべて AI にとって代わると信じ込んだり，AI を妖術と混同している人たちが散見されます．この書籍で，AI にかかわる人や運用する人に，機械学習の品質問題を学んでほしいと思いました．（匿名希望）

連載「5 分で分かる !? 有名論文ナナメ読み：Simeone, A.; Substitutional Reality: Using the Physical…」

■ SR という，これから来るかもしれないトレンドおよびそ

の流れについて分かりやすく解説されており，紹介される論文内容も自身の研究分野に合致したものであったため興味深く感じた．（髙田峻介）

「先生，質問です！」には以下の質問をいただきました．

■資金と向き合っている回答者に感銘を受けました．今回のテーマについて，資金を得ることが目的とならないことを願う次第です．（伊藤治夫）

■研究資金獲得に関する（企業ではうかがいしれない）苦労が垣間見えた．（上田晴康）

■自分が勤務先学内で競争的研究資金の獲得に成功しても，なぜ学外において科研費の獲得で苦労を重ねるのかその原因の一端を認知できた．（大塚敬義）

会議レポート「ACM CHI 2021 会議報告（2）」

■オンライン会議で課題とされている問題について ACM CHI 2021 ではどのように対応していったのか，それに対する反応がどうだったのかがまとまっており非常に参考になるのではないかと考える．（柴田　晃）

■オンラインポスター発表の待ちぼうけ問題という言いにくい話題を文章化した著者の勇気と知性に敬服した．（大塚敬義）

オンライン化について，以下のようなご意見やご要望をお寄せいただきました．今後の参考にいたします．

■ EPUB だけなら反対（祖父江真一）

■ YouTube 感覚で内容が分かるとよいかと思っています．YouTube に記事の紹介動画を UP すればよいなと思っています．または第三者が記事内容を YouTube で講評すればよいかと．（伊藤治夫）

■ PC で PDF 版（全体版）を読んでいるが，オンライン記事はどれもページの両端や行間が妙に広く，図の上下なども変に空白が多く非常に読みにくい．もう少し学会誌としての体裁を整えた編集をしてほしい．（上田晴康）

■ PDF 化の恩恵により，視力の低下した市民でも文字を拡大して読解できる利点があるのでオンライン化に賛成です．（大塚敬義）

【本欄担当　山本祐輔，水上雅博／会員サービス分野】

これらのコメントは Web 版会員の広場「読者からの声」< URL : https://www.ipsj.or.jp/magazine/dokusha.html > にも掲載しています．Web 版では，紙面の制限などのため掲載できなかったコメントも掲載していますので，ぜひ，こちらもご参照ください．会誌や掲載記事に関するご意見・ご感想は学会 Web ページでも受け付けております．今後もより良い会誌を作るため，ぜひ皆様のお声をお寄せください．

「情報処理」アンケート回答フォーム▶
https://www.ipsj.or.jp/magazine/enquete.html

IPSJ MAGAZINE Vol.63 No.2 (Feb. 2022)

CONTENTS

Preface
46 **Forgetful Body**
Yuya KIKUKAWA (ORPHE Inc.)

Special Article
48 **The Establishment of The "Digital Division" in The Examination for Comprehensive Service - The Overview and Sample Questions -**
Takeshi SATOU (National Personnel Authority)

Special Freatures
What Will Smart Factory Change in Factories?
54 **Foreword**
Mikiko SODE TANAKA (International College of Technology) and Koichi TANAKA (Mitsubishi Electric Corp.)
56 **Outline**

Digital Practice Corner
Data Science in Big Data : Big Data in the New Normal
58 **Foreword**
Yohei SATO (Village AI / nat / Lupinus) and Kazuo ISHII (Suwa Univ. of Science)
60 **Outline**

Let's Learn Informatics
64 **Data Analysis Lessons Using Smartphone's Built-in Sensors in a High School Technical Studies Course**
Tomonari KISHIMOTO (Osaka Electro-Communication Univ. High School)

"Peta-gogy" for Future
71 **How Can You Take Online Classes Comfortably?**
Toru OCHI (Osaka Institute of Technology)
72 **Report of The Symposium of FIT2021- The Future of "Informatics" on The Common Test for University Admissions –**
Rieko INABA (Tsuda Univ.)
77 **Comments on The Introduction of The Subject "Information" in The Common Test for University Admissions**
Tatsuya KAWAHARA (Kyoto Univ.)
79 **Compulsory Subject "Informatics" in The Entrance Examination of National Universities in Japan**
Yasuichi NAKAYAMA (The Univ. of Electro-Communications)

70 **Questions for Experts**
84 **Biblio Talk**
86 **Skimming a Famous Paper in Five Minutes**
88 **Committee Reports**
90 **IT Travelog Manga**

Online Only

Special Features
What Will Smart Factory Change in Factories?
e1 **Latest Factory Automation Technologies and Activities for Achieving "Smart Factory"**
Kazuhiro KUSUNOKI (Mitsubishi Electric Corp.)
e7 **Real-time AI Technologies for Digital Transformation in Manufacturing**
Yasushi SAKURAI (Osaka Univ.)
e13 **IoT Platform Now and Future**
Satoshi SUZUKI (NTT DATA Corp.)
e19 **Private 5G to Support The Realization of Smart Factories - Understanding The Systems, Technologies and Considerations for Implementing Private 5G -**
Hiroaki KAKIMOTO (NTT Communications Corp.)
e26 **Future Industrial Infrastructure and Production Systems for Sustinable Society - From The Roadmap Toward 2050 in The Consortium for Human-Centric Manufacturing Innovation -**
Masayo IWAI and Tamio TANIKAWA (AIST Consortium for Human-Centric Manufacturing Innovation)

Let's Share Working Styles! <by Info-WorkPlace Committee>
e50 **CASE 3 : Our Office Has Moved to Online**
Kaori ITO (Yahoo Japan Corp.)

e33 **What Kind of Exam Questions on Informatics Will Appear in University Entrance Exams?**
e57 **Conference Report**

読後のご意見をお送りください

本誌では，現在約 200 名の方々に毎号のモニタをお願いしておりますが，より多くの読者の皆さんからのご意見，ご提案をおうかがいし，誌面の充実に役立てていきたいと考えておりますので，以下 Web ページから奮って事務局までお寄せください．

「情報処理」アンケートページ　**https://www.ipsj.or.jp/magazine/enquete.html**

一般社団法人 情報処理学会　会誌編集部門　E-mail: editj@ipsj.or.jp

人材募集
（有料会告）

有料会告

申込方法：任意の用紙に件名，申込者氏名，勤務先，職名，住所，電話番号および請求書に記載する「宛名」，Web 掲載の有無などを記載し，掲載希望原稿（[募集職種，募集人員，（所属），専門分野，（担当科目），応募資格，着任時期，提出書類，応募締切，送付先，照会先]）を添えて下記の申込先へ，E-mail，Fax または郵送にてお申し込みください．
＊都合により編集させていただく場合がありますので，ご了承ください．

申込期限：毎月 15 日を締切日とし翌月号（15 日発行）に掲載します．

掲載料金：国公私立教育機関，国公立研究機関　22,000 円（税 10％込）
賛助会員（企業）　33,000 円（税 10％込）
賛助会員以外の企業　55,000 円（税 10％込）
＊本会誌へ掲載依頼いただいた場合に限り，追加料金 4,400 円（税 10％込）で同一内容を本会 Web ページに掲載できます．

申 込 先：情報処理学会　会誌編集部門（有料会告係）　E-mail: editj@ipsj.or.jp　Fax(03)3518-8375
＊原稿受付の際には必ず原稿受領のお知らせを差し上げています．もし 3 日以内（土日祝日除く）に返信がない場合は念のため確認のご連絡をください．

＊特に指定がないかぎり履歴書には写真を貼付のこと

■神奈川工科大学創造工学部 自動車システム開発工学科

募集人員　教授　1 名
職務内容　創造工学部自動車システム開発工学科における教育，研究および運営
専門分野　自動車の情報通信，自動運転，交通システム
担当科目　情報通信・制御プログラミング・データ解析と人工知能・自動運転などに関する講義・実習，および卒業研究指導
応募資格　(1) 博士の学位を有するか，あるいはそれに相当する実績を有すること，(2) 自動車の情報通信・自動運転に関する研究実績または実務経験を有することが望ましい，(3) 教育・研究・学生指導に熱意を有し，そのために必要な学識を有すること，(4) 大学・学科などの組織運営に協調的に参加できること
着任時期　2022 年 4 月 1 日以降なるべく早い時期
提出書類　(1) 履歴書（学歴，職歴，学会活動，社会活動など，連絡先と E-mail を明記のこと．書式は本学所定の履歴書・業績書（教員用）を使用．http://www.kait.jp/recruit/），(2) 研究・開発業績リスト（論文等，著書，登録特許，研究開発プロジェクト名など）．なお論文等の著者は全員記入してください．書式については (1) と同様，(3) 主要論文別刷またはそのコピー（2 ～ 3 編程度），(4) これまでのご自身の研究・開発の実績，これからの具体的な抱負について A4 用紙 4 枚以内にまとめたもの，(5) 可能な方は，推薦書または所見を求めることができる方 2 名以内の氏名・所属・連絡先
※(2) の学協会印刷発表論文は履歴書（B-1）に，査読付き国際会議発表論文は履歴書（B-2）に，その他の発表論文・特許等は分類して履歴書（B-3）に，それぞれ新しい順に番号を付けて記入
応募期間　2021 年 12 月 1 日～ 2022 年 2 月 11 日
なお，応募書類受付次第，順次，面接～決定を実施する場合がある
送 付 先　〒 243-0292 神奈川県厚木市下荻野 1030
神奈川工科大学　庶務担当部長　保坂精一宛
※応募書類は，封筒の表に「自動車システム開発工学科教員　応募」と朱書し，簡易書留または書留でお送りください
照 会 先　創造工学部自動車システム開発工学科
教授　脇田敏裕　E-mail: wakita@cco.kanagawa-it.ac.jp
Tel(046)291-3091
そ の 他　【選考方法】書類選考による 1 次選考後，面接による 2 次選考を行います．2 次選考時，15 分程度の模擬授業を実施していただきます．なお，選考に要する旅費などは支給いたしません
応募書類により取得した個人情報は，選考および任用の手続きを行う目的で利用するものであり，この目的以外で利用または提供することはありません．なお，採用に至らなかった方の応募書類は，当該採用選考業務終了後，適切な方法にて破棄いたします．
給与等は本学の規程によります
詳細情報および書式，書類提出先等は，本学 Web ページ「教職員採用情報」を必ず参照してください
http://www.kait.jp/recruit/
https://jrecin.jst.go.jp/seek/SeekJorDetail?id＝D121120223

■青森大学ソフトウェア情報学部

募集人員　准教授または講師または助教　1 名（任期なし，試用期間 1 年）
専門分野　情報工学（主に EdTech の周辺分野の研究領域が望ましい）
※また上記にかかわらず，広く情報処理技術に関連する分野も対象とします
担当科目　卒業研究，ゼミを含む情報工学分野の複数科目（講義，演習）およびプログラミング演習
応募資格　博士の学位を有するか着任までに取得見込み，または同等の業績や実務経験を有し，教育・研究に熱意のある方
着任時期　2022 年 4 月 1 日
提出書類　(1) 履歴書（連絡先として E-mail も明記），(2) 研究業績リスト（著書，査読付き論文，国際会議，特許等に区分），(3) 主要論文別刷またはコピー（3 編程度），(4) これまでの研究概要（A4 用紙 2 ページ程度），(5) 学会および社会における活動（学会活動，社会貢献，地域貢献，実務経験等の実績），(6) 教育・研究に関する抱負（A4 用紙 1 ～ 2 ページ程度），(7) 本人に関する所見を求め得る人（2 名）の氏名と連絡先（所属，住所，電話，E-mail）
※書類はすべてスキャンデータ等でメール添付での送付を推奨します
※紙面での送付も受け付けます
応募締切　2022 年 2 月 14 日（必着）
※応募から順次選考を実施，適任者の採用が確定次第，募集を締め切ります
送付先／照会先　〒 030-0943 青森県青森市幸畑 2-3-1
青森大学ソフトウェア情報学部長　角田　均
E-mail: tsunoda@aomori-u.ac.jp　Tel(017)738-2001（代表）
※メールで提出の場合，データサイズが大きい場合（10Mbyte 以上）は分割して送信をお願いします
※ 紙面で提出の場合，「ソフトウェア情報学部教員応募書類在中」と朱書し，（簡易）書留でお願いします（応募書類は返却しませんのでご了承ください）
※メール／紙面とも受領通知をメールでお送りします
そ の 他　【勤務地】青森大学青森キャンパス（青森県青森市幸畑 2-3-1）
【選考方法】書類選考を経て面接（模擬授業を含む）を実施します
※面接および模擬授業について，可能な場合はオンライン（Zoom 等利用）で実施します
※オフライン実施の場合は，旅費・滞在費のうち一定額を支給します

FIT2022 第 21 回情報科学技術フォーラム
選奨論文・一般論文 講演募集予告

会　期：2022年9月13日（火）～15日（木）
会　場：慶應義塾大学　矢上キャンパス

FIT2022 Web ページ　https://www.ipsj.or.jp/event/fit/fit2022/

受付期間(予定)：2022 年 3 月 29 日（火）～5 月 11 日（水）

◆論文ページ数：2～8ページ程度　　◆講 演 時 間：20分
◆3ページ目以降は追加ページ代（4,000円／ページ）が必要です

電子情報通信学会 情報・システムソサイエティ（ISS）並びにヒューマンコミュニケーショングループ（HCG）と情報処理学会（IPSJ）は、2002年から毎年秋季に合同で「情報科学技術フォーラム(FIT：Forum on Information Technology)」を開催しています。2022年9月には、第21回目を慶應義塾大学　矢上キャンパスで開催します。FITは、両学会の大会の流れをくむものであると同時に、従来の大会の形式にとらわれずに新しい発表形式を導入し、タイムリーな情報発信、活気ある議論・討論、多彩な企画、他分野研究者との交流を実現してきております。皆様の研究成果発表の場として、標記のとおり論文発表を募集致しますので奮ってお申込み下さい。

●申込主要日程（予定）

登録申込／投稿受付期間：2022年3月29日（火）から　2022年5月11日（水）まで
最終掲載原稿締切：2022年6月24日（金）
※ FIT2017 より、査読付き論文は廃止とし、選奨論文制度を取り入れました。
※ 登録申込と原稿投稿は上記のFIT2022 Webページよりお願い致します。詳細は決定次第 Webページでお知らせ致します。

●表彰

FITには、以下の表彰制度がありますので是非ともチャレンジして下さい。
いずれの賞も、電子情報通信学会又は情報処理学会の会員であることが受賞条件となりますのでこの機会に是非御入会下さい。

船井ベストペーパー賞	選奨論文の中から、FIT 学術賞選定委員会で審査の上 3 件選定。賞金は船井情報科学振興財団より 20 万円贈呈。
FIT 論文賞	選奨論文の中から、FIT 学術賞選定委員会で審査の上 7 件程度選定。賞金は FIT 運営委員会より 5 万円贈呈。
FIT ヤングリサーチャー賞	2022 年 12 月 31 日現在で 33 歳未満の講演者（選奨論文および一般論文）の中から、発表件数の 1.5%を上限として選定。賞金は FIT 運営委員会より 3 万円贈呈。本賞受賞は本人に対し一回のみ。
FIT 奨励賞	一般発表のセッション毎に座長の裁量で優秀な発表を 1 件その場で選定（該当なしもあり）。FIT 終了後に賞状を贈呈。

●選奨論文（4～8 ページ程度）

投稿された論文の担当研究会を決定していただきます。FIT2022 Web ページに掲載の研究会取り扱い分野をよく御確認のうえ御自身の論文内容と一致した研究会を、申込者御自身の責任において投稿時に適切に選択して下さい。
船井ベストペーパー賞、FIT 論文賞への審査を希望する場合は、Web からの講演申込みの際に必ず論文形式で『選奨論文』を選択して下さい。但し、賞を前提とした論文形式となりますので、電子情報通信学会又は情報処理学会の会員であることが投稿条件となります。非会員の方は御入会手続きをお済ませの上御投稿下さい。選奨論文は FIT 初日の選奨セッションに組み込まれ、各セッションにて選奨委員 2 名による 1 次審査を行います。1 次審査の結果は当日の夕方までに大会会場に掲示されます。2 次審査は FIT 終了後実施され、上位 3 件が船井ベストペーパー賞、次点 7 件程度が FIT 論文賞の受賞となります。
※4 ページ以上の投稿が必須ですが、3 ページ目からは追加ページ代（4,000 円／ページ）が発生します。例えば 6 ページ投稿の場合、4 ページ分の追加ページ代が発生しますので、講演参加費のほかに「4,000 円×4＝16,000 円」の追加費用が必要となります。

●一般論文（2～8 ページ程度）

FIT2022 Web ページに掲載の研究会取り扱い分野をよく御確認のうえ御自身の論文内容と一致した研究会を、申込者御自身の責任において適切に選択して下さい。
※3 ページ以上の投稿される場合は、3 ページ目からは追加ページ代（4,000 円／ページ）が発生します。例えば 4 ページ投稿の場合、2 ページ分の追加ページ代が発生しますので、講演参加費のほかに「4,000 円×2＝8,000 円」の追加費用が必要となります。

●論文誌推薦制度

選奨論文の中から船井ベストペーパー賞の審査を通して優秀な論文と判断されたものを、FIT プログラム委員会が電子情報通信学会または情報処理学会（FIT 講演申込フォームの講演応募分野（研究会）で選択した研究会が属する学会）の論文誌へ推薦します。掲載の採否は、それぞれの学会の論文誌編集委員会が決定します。論文誌への投稿の際には、投稿先論文誌編集委員会の評価基準を満足しうる、完成度の高い論文に仕上げて頂くことをお薦めします。なお、推薦を辞退することも可能です。

●問合せ先（FIT2022事務局）

〒101-0062 千代田区神田駿河台 1-5　化学会館 4 階
情報処理学会　事業部門 TEL. 03-3518-8373　FAX. 03-3518-8375　E-mail:ipsjfit@ipsj.or.jp

情報処理学会　第 84 回全国大会　聴講事前申込受付中
イベント企画のみの聴講参加は「無料」!!　ハイブリッド開催
申込はこちらから⇒ https://www.ipsj.or.jp/event/taikai/84/
事前申込がお得です！　ぜひ皆様お誘い合わせの上、奮ってご参加ください

『変わる社会と情報処理』

大 会 会 期：2022 年 3 月 3 日（木）～ 5 日（土）
大 会 会 場：愛媛大学　城北キャンパス（愛媛県松山市文京町 3）　ハイブリッド開催
共　　催：愛媛大学
後　　援：愛媛県　愛媛県教育委員会　全国高等学校情報教育研究会

情報処理学会第 84 回全国大会の「大会聴講参加」の申込を受付中です.
■ イベント会場・特別会場において開催される「特別講演／招待講演／イベント企画／各種展示」を聴講・ご覧になる場合
→「大会イベント企画限定聴講参加」（無料）
■ 上記に加え、「一般セッション／学生セッション」を聴講する場合
→「大会共通聴講参加」（有料）

イベント企画のみ聴講希望の方は，大会 Web ページから申込みをする際，「**大会イベント企画限定聴講参加**」にお申し込みください.
通常の一般セッション・学生セッションも聴講希望の場合は，「**大会共通聴講参加**」にお申し込みください（聴講参加費は有料となります）
事前申込受付期間を過ぎると当日価格となりますのでお申し込みはお早めに！.

事前申込受付期間：2021 年 12 月 6 日（月）～ 2022 年 2 月 15 日（火）

招待講演・特別講演・公開講演企画【聴講参加無料】：招待講演 4 件、特別講演 3 件、公開講演 1 件を予定しております.

招待講演 –1	3 日（木）16：20 ～ 16：35　未定（The Korean Institute of Information Scientists and Engineers）
招待講演 –2	3 日（木）16：35 ～ 16：50　未定（China Computer Federation）
招待講演 –3	3 日（木）16：50 ～ 17：05　未定（IEEE Computer Society）
招待講演 –4	3 日（木）17：05 ～ 17：20 「Intelligence? Smartness? Emotion? What do we expect from future computing machinery?」（Association for Computing Machinery）
特別講演	4 日（金）15：20 ～ 16：20 「『ポスト量子』暗号 --- 量子計算機に対して安全な暗号の最前線」
	4 日（金）16:30 ～ 17:30 「スパコン富岳による飛沫エアロゾル感染リスク評価のデジタルトランスフォーメーション」
	5 日（土）15：30 ～ 17：30　IPSJ-ONE
公開講演	5 日（土）13：20 ～ 15：20 「デジタルが地域に変革をもたらす　－愛媛から始める DX －」

イベント企画【聴講参加無料】：各イベント企画では，その分野の最前線で活躍されておられる方をお招きし，講演・パネル討論等の開催を予定しております.

第 1 イベント会場	3 日　9：30 ～ 11：30 「一次産業と ICT」 4 日　9：30 ～ 11：30 「ヘルスケア情報の利活用に資する匿名加工技術の実現に向けて ～ 匿名加工コンテスト PWS Cup 2021 ～」 4 日　12：40 ～ 15：10 「知能と計算とアーキテクチャの新しい関係を目指して」 5 日　9：30 ～ 12：00 「情報入試ー共通テストと個別試験（仮題）」 5 日　13：20 ～ 15：20 「① IPSJ KIDS，②大学共通テスト解説」
第 2 イベント会場	3 日　9：30 ～ 11：30 「2021 年サイバー事件回顧録 ～技術と法制度の両面から～」 3 日　12：40 ～ 15：10 「～コンピュータパイオニアが語る～『私の詩と真実』」（オンライン） 4 日　9：30 ～ 11：30 「一般情報教育と数理・データサイエンス・AI」 4 日　12：40 ～ 15：10 「新世代委員会企画」 5 日　9：30 ～ 15：20 「第 14 回情報システム教育コンテスト」
第 3 イベント会場	3 日　9：30 ～ 11：30 「革新的アルゴリズム基盤の構築に向けて」 3 日　12：40 ～ 15：10 「IoT が拓く未来：～ 2030 年の未来予想図～」 4 日　9：30 ～ 11：30 「アジャイル開発の契約上の問題点と対策」 4 日　12：40 ～ 15：10 「日本機械学会／情報処理学会 合同企画 モノづくりと情報処理における人材育成について」 5 日　9：30 ～ 12：00 「初等中等教員研究発表セッション」

第 4 イベント会場	3 日　9：30 ～ 11：30　「IT 情報系キャリア研究セッション」 3 日　12：30 ～ 13：30　「AI TECH TALK」 3 日　15：20 ～ 17：20　「IT 情報系キャリア研究セッション」 4 日　15：20 ～ 16：20　「インダストリアルセッション」 5 日　9：30 ～ 12：00　「情報科学の達人」 5 日　13：20 ～ 15：20　「中高生情報学研究コンテスト」（オンライン）
第 5 イベント会場	3 日　9：30 ～ 11：30　「8 周年を迎えた認定情報技術者制度 CITP（Certified IT Professional）の現状と今後の方向性」（オンライン） 4 日　12：40 ～ 15：10　「論文必勝法」（オンライン） 5 日　13：20 ～ 15：20　「切迫する社会課題の克服に向けた A I/ ビッグデータビジネスの新展開と人材育成」（オンライン）
第 6 イベント会場	5 日　9：30 ～ 11：30　「Exciting Coding! Junior2022@Ehime」

一般セッション・学生セッション【聴講参加 有料】：

約 1,500 件の研究成果発表があります．大会 3 日間でおよそ 30 会場を使用して，190 あまりのセッションが組まれ，活発な発表，議論・討論が行われます．

■聴講参加費・講演論文集代（税込）

現地参加、オンライン参加ともに同価格です。学生の大会共通聴講参加費は「無料」です．

申込種別	事前価格（2/15 まで）	価格（2/16 以降 ~ 最終日）
大会イベント企画限定聴講参加	無料	無料
大会共通聴講参加（正会員）＊全論文の PDF アクセス権付	9,000 円	10,000 円
大会共通聴講参加（一般非会員）＊全論文の PDF アクセス権付	15,000 円	17,000 円
大会共通聴講参加（学生会員・ジュニア会員・学生非会員）	無料	無料

◇留意事項

※「大会イベント企画限定聴講参加」は，特別講演，招待講演，イベント企画，IT 情報系キャリアセッションのみ聴講参加可能です．一般セッション・学生セッションの聴講はできませんのでご注意ください．

一般セッション・学生セッションも聴講参加希望の場合には，大会共通聴講参加（有料）にお申し込みください．学生の方は大会共通聴講参加費が「無料」です．

※「大会共通聴講参加」は，一般セッション・学生セッションを含む大会すべてのセッションの聴講参加が可能です．

※ 講演参加申込の方，座長，イベント企画者および登壇者は聴講参加申込は不要です．座長には別途ご請求の案内をいたします．

◇ハイブリッド開催について

オンラインミーティングツール Zoom を併用しながら現地でイベント企画・各発表セッションを開催致します。インターネット・オーディオ機器に接続できる PC とヘッドセットを各自で必ずご準備願います。

イベントによっては、オンラインのみのものがあります。現地ではパブリックビューイング会場でご覧いただけます。

■懇親会（有料）

大会参加者の皆様の親睦をぜひ深めてください．

開催日時：2022 年 3 月 3 日（木）18：00 ～ 20：00（予定）

開催会場：ホテルメルパルク松山（松山市道後姫塚 123-2）

■講演論文集代（税込・送料込）

残部のある限り販売を行いますが，確実に御入手いただくには 2022 年 2 月 3 日（木）までのお申し込みをお勧めいたします．受け取りは大会終了後の郵送となります．

申込種別	予約価格（2/3 迄）	価　格
講演論文集分冊（個人・法人問わず）	13,000 円	14,000 円
講演論文集セット ＊DVD-ROM 1 枚付き（個人・法人問わず）	60,000 円	66,000 円
講演論文集 DVD-ROM（個人）	10,000 円	
講演論文集 DVD-ROM（法人）	60,000 円	

■聴講参加および講演論文集の予約申込，詳細は，以下のサイトからお願いいたします．

第 84 回全国大会公式 Web サイト　https://www.ipsj.or.jp/event/taikai/84/

■問合先

一般社団法人情報処理学会 事業部門

〒 101-0062 東京都千代田区神田駿河台 1-5 化学会館 4F　　電話（03）3518-8373　　E-mail: ipsjtaikai@ipsj.or.jp

■ 編集室 ■

半導体不足が続いている．産業のコメとも呼ばれる半導体は，家電や携帯電話，自動車，ゲーム，飛行機など，大小さまざまなモノに使われており，私たちの生活に欠かせない存在だ．教育現場でも Raspberry Pi や Jetson など学生の教育に欠かせない部品の入手が困難となっており，非常に困っている．

かつて日本は半導体の分野で世界を牽引していた．現在弱体化はしてきているが日本の半導体ウェーハ処理工場の生産能力はまだ世界の 16% を保持している．経済産業省の試算では半導体市場は，2030 年までに約 100 兆円規模に到達すると見込まれている成長産業である．デジタルの世界では，半導体チップは生産の生命線であり，IoT の世界ではサイズの小さな低消費電力の半導体が必須であり，高速処理が可能な新たな AI チップは必要である．半導体製造は正に工場の強さが鍵を握る分野である．加えて，その他の製造業の要である．コロナ禍の不景気下でも売れ行きが良い商品にもかかわらず，半導体不足のため生産できない製品が多く存在する現状を鑑みると日本半導体の底力を示せればと願わざるを得ない．

本特集を企画し，工場運営の考え方には日本独自の考え方があり，製造業が日本のお家芸と言われているゆえんであることを再認識した．製品，そこで働く人を大切に思い，心を尽くす姿勢が日本らしさなのだと思う．その結果として品質が高く，壊れにくい優れた製品が製造できる．私はこの企画の編集に携わり日本の工場の将来は明るいと感じた．

（袖美樹子／本特集エディタ）

次号（3 月号）予定目次

編集の都合により変更になる場合がありますのでご了承ください．

※はオンライン版のみの掲載となります

「特集」知能コンピューティング― AI とハードウェアの出会い―※

AI は新しいハードウェアを欲しているか？―知能と計算とアーキテクチャの新しい関係／確率的コンピューティングの再開拓―その場学習が可能な極低電力エッジ AI に向けて―／画像の解像度と知的処理の関係を見つめ直す―知的な高解像度リアルタイム処理に向けて―／機械学習に適したハードウェア・ハードウェアに適した機械学習アルゴリズム／ランダム・スパース・ストカスティック―新しい計算の形を目指して―

委員会から：<Info-WorkPlace 委員会企画 > お届け Info：今年度もやります！全国大会の“デリバリー”

学会活動報告：IFIP 近況報告―情報処理国際連合―

教育コーナー：ぺた語義

連　載：5 分で分かる ⁉ 有名論文ナナメ読み／情報の授業をしよう！／先生，質問です！／ビブリオ・トーク

コラム：巻頭コラム

掲載広告カタログ・資料請求用紙

広告のお申込み

■広告料金表（価格は税 10%込）

掲載場所	4 色	1 色
表 2	363,000 円	—
表 3	302,500 円	—
表 4	423,500 円	—
表 2 対向	330,000 円	—
表 3 対向	291,500 円	170,500 円
前付 1 頁	275,000 円	148,500 円
前付 1/2 頁	—	88,000 円
前付最終	—	162,800 円
目次前	—	162,800 円
差込 （A4 変形判 70.5kg 未満　1 枚）	302,500 円	
差込 （A4 変形判 70.5kg ～ 86.5kg　1 枚）	385,000 円	
同封 （A4 変形判　1 枚）	385,000 円	

■「情報処理」

発　　行　一般社団法人 情報処理学会
発行部数　20,000 部
体　　裁　A4 変形判
発 行 日　毎当月 15 日
申込締切　前月 10 日
原稿締切　前月 20 日
広告原稿　完全版下データ
原稿寸法　1 頁　天地 250mm ×左右 180mm
　　1/2 頁　天地 120mm ×左右 180mm
雑誌寸法　天地 280mm ×左右 210mm

■問合せ・お申込み先

〒 169-0073　東京都新宿区百人町 2-21-27
アドコム・メディア（株）（Tel/Fax/E-mail は下に記載）

＊原稿制作が必要な場合には別途実費申し受けます.
＊同封のサイズ・割引の詳細についてはお問合せください.

掲載広告の資料請求

掲載広告の詳しい資料をご希望の方は，ご希望の会社名にチェック ☑ を入れ，送付希望先をご記入の上，Fax にて（または E-mail にて必要事項を記入の上）アドコム・メディア（株）宛にご請求ください.

■「情報処理」 63 巻 2 号　掲載広告（五十音順）

□オーム社…… 表 2 対向　　□すべての会社を希望
□とめ研究所…… 目次前上
□フォーラムエイト…… 表 2

■資料送付先

フリガナ
お 名 前 ＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

勤 務 先 ＿＿＿＿＿＿＿＿＿＿＿＿＿＿ 所属部署 ＿＿＿＿＿＿＿＿＿＿

所 在 地 （〒　　-　　）＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

TEL（　　）　　-　　FAX（　　）　　-

ご専門の分野 ＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

お問合せ・お申込み・資料請求は
広告総代理店 アドコム・メディア（株）
Tel.03-3367-0571　Fax.03-3368-1519　E-mail: sales@adcom-media.co.jp

賛助会員のご紹介

本会をご支援いただいております賛助会員をご紹介いたします．
Web サイト（https://www.ipsj.or.jp/annai/aboutipsj/sanjo.html）「賛助会員一覧」のページからも
各社へリンクサービスを行っておりますので，ぜひご覧ください．

照会先　情報処理学会　会員サービス部門　E-mail: mem@ipsj.or.jp　Tel.(03)3518-8370

●●● 賛助会員（20 ～ 50口）

（株）日立製作所

富士通（株）

日本電気（株）

三菱電機（株）

（株）サイバーエージェント

日本アイ・ビー・エム（株）

●●● 賛助会員（10 ～ 19口）

（株）リクルート

グーグル合同会社

（株）NTT ドコモ

（株）東芝

日本電信電話（株）

日本マイクロソフト（株）

（株）フォーラムエイト

●●● 賛助会員（3 ～ 9口）

（一社）情報通信技術委員会

（株）NTT データ

グリー（株）

（一財）インターネット協会

（一社）情報サービス産業協会

トレンドマイクロ（株）

（株）BFT

NTT コムウェア（株）

NTT テクノクロス（株）

（株）うえじま企画

エッジテクノロジー（株）

沖電気工業（株）

コアマイクロシステムズ（株）

三美印刷（株）

ソニー（株）

（株）テクノプロ
テクノプロ・デザイン社

みずほリサーチ&テクノロジーズ（株）

一般社団法人
情報処理学会
Information Processing Society of Japan

https://www.ipsj.or.jp/event/taikai/84/
情報処理学会　全国大会　検索

情報処理学会
第 84 回全国大会

大会テーマ：変わる社会と情報処理

開催日　2022.3.3(木)～5(土)

会　場　愛媛大学　城北キャンパス(ハイブリッド開催)

事前申込受付期間　2021.12. 6(月)～ 2022.2.15(火)

聴講参加費(税込)　現地参加、オンライン参加共に同価格です。

申　込　種　別	事前価格(2/15まで)	価格(2/16～最終日)
大会イベント企画限定聴講参加	無 料	無 料
大会共通聴講参加(正会員)*全論文のPDFアクセス権付	9,000円	10,000円
大会共通聴講参加(一般非会員)*全論文のPDFアクセス権付	15,000円	17,000円
大会共通聴講参加(学生会員・ジュニア会員・学生非会員)	無 料	無 料

◆ハイブリッド開催について
オンラインミーティングツール Zoom を併用しながら現地でイベント企画・各発表セッションを開催致します。インターネット・オーディオ機器に接続できる PC とヘッドセットを各自で必ずご準備願います。
イベントによってはオンラインのものがあり、現地ではパブリックビューイング会場でご覧いただけます。

事前申込がお得です！　皆さまお誘い合わせの上、奮ってご参加ください

大会イベント企画（聴講無料）

3/3(木)

一次産業とICT
2021年サイバー事件回顧録
～技術と法制度の両面から～
革新的アルゴリズム基盤の構築に向けて
8周年を迎えた認定情報技術者制度CITP
(Certified IT Professional)の現状と
今後の方向性
～コンピュータパイオニアが語る～
「私の詩と真実」
IoTが拓く未来:～2030年の未来予想図～
AI TECH TALK
IT情報系キャリア研究セッション

3/4(金)

ヘルスケア情報の利活用に資する匿名加工技術の
実現に向けて ～ 匿名加工コンテストPWS Cup 2021 ～
アジャイル開発の契約上の問題点と対策
一般情報教育と数理・データサイエンス・AI
知能と計算とアーキテクチャの新しい関係を目指して
新世代委員会企画
日本機械学会／情報処理学会 合同企画
モノづくりと情報処理における人材育成について
論文必勝法
スパコン富岳による飛沫エアロゾル感染リスク評価の
デジタルトランスフォーメーション
「ポスト量子」暗号 --- 量子計算機に対して安全な暗号の最前線
ランチョンセッション
インダストリアルセッション

3/5(土)

IPSJ-ONE
情報入試ー共通テストと個別試験
第3回初等中等教員研究発表セッション
情報科学の達人
Exciting Coding!Junior2022@Ehime
第14回情報システム教育コンテスト
デジタルが地域に変革をもたらす
ー愛媛から始めるDXー
①IPSJ KIDS, ②大学共通テスト解説
第4回中高生情報学研究コンテスト
切迫する社会課題の克服に向けたAI／
ビッグデータビジネスの新展開と人材育成

愛媛大学 EHIME UNIVERSITY

共催：愛媛大学　後援：愛媛県　愛媛県教育委員会　全国高等学校情報教育研究会

〒101-0062 東京都千代田区神田駿河台一―五
編集人 稲見昌彦
発行所 東京都千代田区神田駿河台一―五 一般社団法人 情報処理学会
発行人 木下泰三
電話 東京（〇三）三五一八―八三七四
振替口座 〇〇一五〇―四―八三四八四
印刷所 東京都荒川区西日暮里五―十六―七 三美印刷株式会社
会員外発売所 東京都千代田区神田錦町三―一 株式会社オーム社

第4回 情処ウェビナー

https://www.ipsj.or.jp/ipsjwebinar/webinar04.html

一般社団法人 情報処理学会 Information Processing Society of Japan

無料

AIで人の表情・感情を可視化する

―表情解析 AI の理論紹介と、探究授業における感情認識 AI の活用―

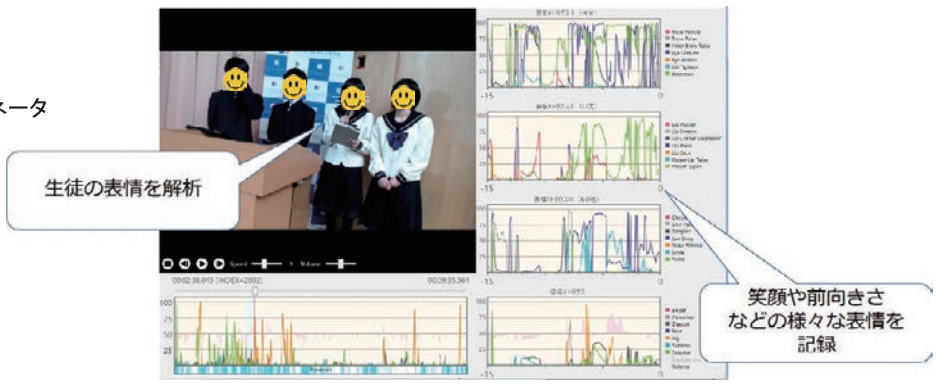
2022.1.22（土） 15:00～16:00

教育における AI 活用の可能性として感情認識 AI の理論紹介と探究学習及び探究学習におけるプレゼンテーション分析等 AI 活用について議論を行います。2015 年から「中学校デジタル化プロジェクト」により IT を活用した教育の高度化の中で、青翔開智中学校・高等学校（以下、青翔開智）の探究学習では、感情認識 AI を用いた生徒によるプレゼンテーションの実証実験を行っています。本プロジェクトにおける探究学習と IT の活用について、青翔開智における実例と IT 活用の実態を御紹介します。また、WEB 会議やリモート教育では参加者が映像を公開しないためにプレゼンテータに視聴者の表情が分からないといったリモートならではの課題が出てきています。そこで、Web 会議で利用可能な表情解析 AI の活用例を、デモを交えて実施してみます。本ウェビナーの講演中、ジュニア会員 3 名の方に表情解析 AI を利用していただき、そのモニタリング画面をウェビナーでご覧いただけます。

齋藤　学

株式会社シーエーシー
経営統括本部 経営企画部 IT コーディネータ

シーエーシーでは経営企画として IT や制度からナレッジマネジメント等まで様々な企画を行う。また、社外活動として JISA 中学校デジタル化プロジェクト座長、アクティブラーニング学会の探究など複数の部会を運営している。中学校デジタル化プロジェクトにおいて感情解析を活用して探究学習の分析等を行う。



出典）齋藤　学：感情認識 AI「心 sensor」の教育現場導入に向けた実証実験 図 9, 情報処理 デジタルプラクティスコーナー , Vol.62, No.5 (2021).

定価 1,760 円（本体 1,600 円 + 税 10%）
本誌広告一手取扱い アドコム・メディア株式会社
〒169-0073 東京都新宿区百人町 2-21-27　TEL.03-3367-0571 FAX.03-3368-1519
雑誌　05269-02
4910052690226
01600

会誌「情報処理」Vol.63 No.2（Feb. 2022）「デジタルプラクティスコーナー」

特集号招待論文

Apache ArrowによるRubyのデータ処理対応の可能性

村田賢太[1, 3]　須藤功平[2, 3]

[1]（株）Speee　[2]（株）クリアコード　[3]Red Data Tools

RubyはWebシステムの記述言語として高い生産性を発揮し，Web業界では広く浸透している．一方，分析的データ処理への対応が弱いため，データ処理分野ではほとんど利用されていない．昨今のDX推進などの流れから，Rubyで書かれた既存システムのデータ処理への対応が近い将来必要となるだろう．そのような要求に対応するためには，前もってRubyを分析的データ処理に対応させる必要がある．本稿では，Rubyを分析的データ処理に対応させる手段としてApache Arrowが有効であることを示す．Apache Arrowは，既存のデータ処理コンポーネント間のデータ連携の非効率性を解消するために提案された，データフォーマットとAPIである．RubyをApache Arrowに対応させることで，分析的データ処理に対応できるだけでなく，データ処理分野における先進的な取り組みにRubyからアクセスできるようになる．

1. Rubyを分析的データ処理に対応させる目的

1.1 データ処理分野におけるRubyの位置付けと課題

プログラミング言語Ruby[1]は，Ruby on Rails[2]によるWebアプリケーションの記述言語として高い生産性を発揮できるとして世界中で人気を集め，これまでに数多くの利用実績がある（[2]に数十万という数字が記載されている）．ここ数年のWebフロントエンド技術の進化によってWebアプリケーションの作り方が変わったことで，Ruby on Railsが新規開発で採用される機会は減ってきているものの，過去に開発されたRailsアプリケーションがいまも現役稼働している例は少なくないだろう．Railsアプリケーションの例としてよく名前が挙がるCookpad[3]，GitHub[4]，Shopify[5]のような多数のユーザを獲得している有名サービス以外でも，これまでのRuby on Railsの人気を考慮すると，無名の企業の社内システムとして稼働しているRailsアプリケーションが多数存在することが想像できる．さらに，最近になってRailsアプリケーションに対して有効なRuby用のJITコンパイラ[6]が試作されており，今後もRubyとRuby on Railsの利用が続く可能性は高い．

一方，データ処理分野においてRubyはまったくと言ってよいほど取り上げられることがない．Googleなどで“Best Programming Language for Data Science”のようなキーワードを検索して出てくる[7]のようなサイトにはRubyはまったく出てこない．このようなサイトで必ず名前が挙がるPython [8]やR [9]と比べると，Rubyは特に分析的なデータ処理への対応が弱いことに気づくだろう．第一に，Apache Spark [10]などの分析用のデータ処理コンポーネントをRubyから利用するためには，データの受け渡しに対応する必要がある．追加の開発コストを避けるためにJSON[11]のような共通フォーマットを利用すると，データのシリアライズとデシリアライズの処理コストが発生してしまい，扱うデータ量が大きい場合には無視できない問題となる．

企業のDX推進が望まれている昨今の潮流[12]により，既存の業務アプリケーションを分析的データ処理コンポーネントと連携させることで機械学習の利用やビッグデータの活用に対応していくことが必要になるだろう．この流れは，既存のRailsアプリケーションにも同様に当てはまる．そのため，Rubyを分析的データ処理に適応させていくことが急務である．

1.2 Apache ArrowによるRubyの分析的データ処理への対応

Rubyが分析的データ処理に適応していくためには次の2つの取組みの両方が必要である．

（1）Rubyで利用可能なデータ分析ツールを拡充させる
（2）Ruby以外の言語で実装されたデータ処理コンポーネントとの連携を強化する

ここで問題となるのが両者に対応するための開発リソースである．Rubyを分析的データ処理で利用する事例がある程度増えないと，恒常的に確保できる開発者の人数や費用が限られてしまうため，対応作業を少人数で進めていかなければならない．可能な限り少ない手間でRuby用のツールの拡充と多数のデータ処理コンポーネントとの連携を実現し，それらが高いパフォーマンスを発揮できることが求められる．そして，少数の開発者であっても機能開発を継続できることが望ましい．

本稿では，このような制約のもとでRubyを分析的データ処理に対応させていく手段として，RubyをApache Arrow[13]に対応させることが有効であることを述べる．本稿は次のように構成される．第2章では分析的データ処理におけるApache Arrowの役割について説明する．第3章ではRubyをApache Arrowに対応させる取り組みとしてRed Arrow[14]を紹介し，その有効性を示す．第4章で関連技術について述べ，第5章でRubyのデータ処理対応についての今後の展望を述べて論文をまとめる．

2. 分析的データ処理においてApache Arrowが果たす役割

2.1 Apache Arrowが登場した背景

一般に，データ処理システムは複数のコンポーネントを組み合わせて構成され，分散システムになる場合も多い[15]．ここでコンポーネントと呼んでいる対象は，データ処理の一部の役割を担うライブラリやミドルウェアである．たとえば，データベースは，データを検索可能な状態で保存し，問合せに対して適切なデータを取捨選択して返す役割を担うデータ処理コンポーネント

である．データベースにはトランザクション処理に向いているものと分析的な処理に向いているものとがあり，どちらか一方に特化している場合が多い．たとえばリレーショナルデータベースはトランザクション処理向けである．分析的な処理に向いているデータベースの例としてはApache HBase[16]やApache Kudu[17]がある．

データ処理コンポーネントは，自身の目的に対して適切な内部データ表現を使用して作られている．このデータ構造はほかのコンポーネントの内部データ表現とは無関係に設計され，内部データをそのままほかのコンポーネントに連携することは不可能である．

2つの異なるデータ処理コンポーネント間でデータの連携が必要になった場合，次のいずれかの方法でデータの受け渡しを実現することになる．

（1）2つのコンポーネント間で専用のデータ連携の仕組みを実装する．たとえば，多くのリレーショナルデータベースシステムは，C言語で実装されるアプリケーション用のクライアントライブラリを提供している．このクライアントライブラリとデータベースサーバの間でデータ連携のプロトコルを定め実装している．アプリケーションは，クライアントライブラリが提供するデータ構造にアクセスすることでデータを操作できる．
（2）JSONのような共通のデータフォーマットを仲介してデータ連携を実現する．この場合，各コンポーネントが内部データ表現と共通フォーマットの間のデータ変換を実装する．さらに，コンポーネント間で共通フォーマット上のスキーマを示し合わせることも必要となる．

近年，コンピュータの性能向上に伴いデータ処理システムへの要求は高度化し，複数のコンポーネントを組み合わせることで生じる問題が目立つようになった[18]．データ処理システムが複雑化することで，使用されるコンポーネントの数が増え，データ連携が必要となるコンポーネントの組合せが増大した．データ連携のパスが増えると実装しなければならないデータ変換処理も増えてしまう．専用のデータ連携パスが存在しない経路でJSONのような共通フォーマットを利用すると，今度は大きなデータ量を連携する際のデータ変換処理が大きなCPU時間を消費してしまう．

このようなデータ処理コンポーネントが抱える問題を解決する手段としてApache Arrowが提案された．

2.2　Apache Arrowが解決を目指す問題と，解決のためのアプローチ

Apache Arrowは分析的データ処理コンポーネントが持つ次の2つの問題を解決することを目指している．

（1）データ連携時のコスト
（2）コンポーネント間の機能差に起因する問題

2.2.1　データ連携時のコスト

異なる内部データ表現を採用する2つのコンポーネント間でデータを連携する方法は2つあった．すなわち（1）専用のデータ変換を実装すること，および（2）JSONのような共通フォーマットを採用する方法である．専用のデータ変換を実装する方法は，データ連携の経路が増えるに従って実装すべきデータ変換の数が増え，コードのメンテナンスコストが増大する．一方，共通

データフォーマットを仲介する場合は，シリアライズとデシリアライズでデータ変換が2回走るため連携したいデータ量が増えると実行時の処理速度が増大してしまう．このようなメンテナンスコストや実行時のコストはないほうが望ましい．

Apache Arrowは，コンポーネント間のデータ表現のハブとなる新たなデータフォーマットを定めることで，この問題の解決を目指している．**図1**はApache Arrowの公式サイト内[19]に掲載されているもので，Apache Arrowによってコンポーネント間のデータ連携が効率化される様子を示した図である．図の左側はApache Arrowがない状態であり，上述の（1）に相当する状態である．そして，図の右側の状態がApache Arrowによってデータ連携が効率化された状態である．

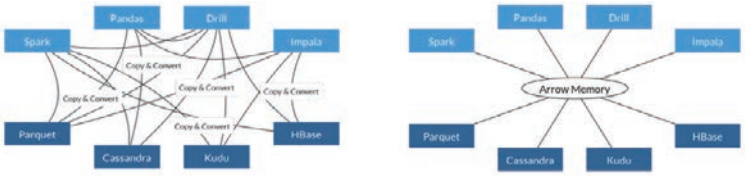


図1　Apache Arrowによってデータ連携が効率化される以前（左）と，効率化された後（右）の様子（[19]より引用）

Apache Arrowのデータフォーマットはメモリ上のデータの配置に関する仕様でありArrowフォーマット[20]と呼ぶ．Arrowフォーマットの詳細については次項で述べる．

Apache Arrowによるデータ連携では，メモリ上のArrowフォーマットをコンポーネント間でそのまま受け渡すことを想定している．そのため，データ連携のためのデータ変換処理を実装する必要はなく，実行時のデータ変換コストも発生しない．ネットワークを介したデータ連携で必要な場合にデータの圧縮をサポートするが，メモリ上で値が連続的に配置されている配列の単位で圧縮を行うため，データの受信側では受け取った圧縮データを展開してメモリ上にそのまま配列データとして配置するだけである．

JSONのようなデータフォーマットとArrowフォーマットを比較すると，Arrowフォーマットでは構文解析のコストが生じず，非常に高速にデータの受け渡しが可能となる．**図2**は，データフォーマットごとにデータの読み込み時間を計測した結果である．読み込み対象のファイルは，長さ10バイトの無作為な文字列で構成される列が1列と，倍精度浮動小数点数の値を持つ列が10列で構成されるデータフレームを各フォーマットで保存して作成した．この結果から，一般によく利用されるCSVやJSONと比較して，Arrowフォーマットが圧倒的に高速であることが分かる．

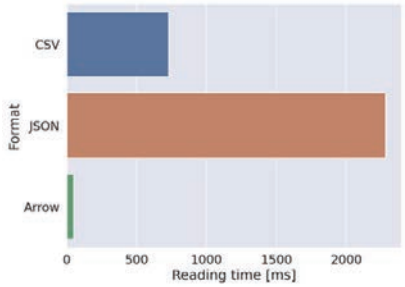


図2　データフォーマットごとの読み込み時間（シングルスレッド）

2.2.2　現代的なハードウェアの特徴を考慮したデータレイアウト

Apache Arrowが定めるArrowフォーマットは，2次元のテーブルデータと多次元数値配列データを対象とするもので，分析的データ処理で最もコンピュータの性能が発揮されることを狙って設計されている．

Arrowフォーマットでは，テーブルデータをRecordBatchと呼ばれる構造で表現する．RecordBatchは，1つのSchemaおよび列と同数の配列で構成されるオブジェクトである（**図3**）．Schemaはテーブルの論理的な構造を表すメタデータであり，テーブルを構成する各列の名前と値のデータ型の情報を持つ．列を表す配列には値が先頭から順に連続的に並んでいる．このように列単位でデータを配置する列指向のデータレイアウトを採用することで，同時にアクセスされやすいデータのキャッシュローカリティが高くなる．列指向のデータレイアウトと行指向のデータレイアウトの違いを**図4**に示す．また，配列の先頭要素はCPUのキャッシュラインやSIMD命令用レジスタのワードサイズに沿うようにアライメントされる．アライメントによって，メモリとCPU間のデータ転送効率が高くなる．

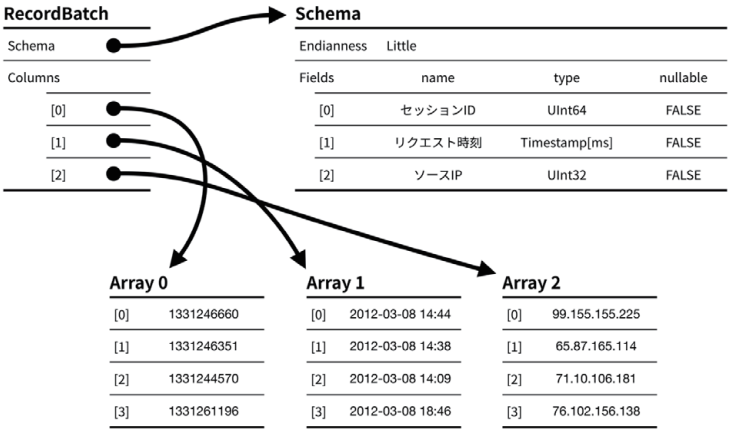


図3　RecordBatchのデータ構造

(a) 論理的なテーブル構造

	セッションID	リクエスト時刻	ソースIP
1行目	1331246660	2012-03-08 14:44	99.155.155.225
2行目	1331246351	2012-03-08 14:38	65.87.165.114
3行目	1331244570	2012-03-08 14:09	71.10.106.181
4行目	1331261196	2012-03-08 18:46	76.102.156.138

(b) 行指向のデータ配置

1行目	1331246660
	2012-03-08 14:44
	99.155.155.225
2行目	1331246351
	2012-03-08 14:38
	99.155.155.225
3行目	1331244570
	2012-03-08 14:09
	99.155.155.225
4行目	1331261196
	2012-03-08 18:46
	99.155.155.225

(c) 列指向のデータ配置

セッションID	1331246660
	1331246351
	1331244570
	1331261196
リクエスト時刻	2012-03-08 14:44
	2012-03-08 14:38
	2012-03-08 14:09
	2012-03-08 18:46
ソースIP	99.155.155.225
	65.87.165.114
	71.10.106.181
	76.102.156.138

図4　テーブルデータのメモリ上の配置方法の違い

このように，Arrowフォーマットは列の単位でメモリ上に値を連続配置する列指向レイアウトであり，キャッシュメモリとSIMD命令という現代のCPUの特徴を考慮してレイアウトが設計されているため，分析的データ処理でよく実行される列のスキャンが必要なデータ処理が効率よく実行できる[21].

Arrowフォーマットは2016年に最初の案が発表提案された．それから現在までまだ数年しか経過していないが，このフォーマットの特徴を活かした先進的なデータ処理コンポーネントがすでにいくつか存在する．たとえば，PySpark[22]，HeteroDB社のPG-Strom[23]，そしてNVIDIAのRAPIDS[24]である．また，研究段階のものとしてはFletcher[25],[26]がある．

2.2.3　コンポーネント間の機能差に起因する問題

複数のコンポーネントが類似した機能を持っている場合の機能差が問題となる場合がある．ここでは，データフレームと呼ばれるテーブルデータの分析のためのAPIを例にこの問題について説明する．

データフレームのAPIはプログラミング言語が異なっていてもおおよそ共通の機能セットを提供することになる．そのため，各プログラミング言語でデータフレームAPIを提供するライブラリ群は，多くの共通するアルゴリズムや機能を提供している．たとえば，配列の平均や標準偏差を求める機能，CSVファイルを読み込んでデータフレームオブジェクトを生成する機能などである．

既存のデータフレームライブラリは，内部データ表現と実装言語が異なるため，このような共通機能を互いに再利用できない実装として各々で独自に行っている．開発プロジェクトが分離しているため，ノウハウの共有も活発ではない[27]．

機能の実装が共有されないことで，共通機能の微妙な仕様の違いが発生している．たとえば，pandasとRでは，NaNをカテゴリ型配列のカテゴリとして扱えるかどうかの違い[28]や，データフレームのマージ処理のパフォーマンスの違い[29]が存在する．このような微妙な機能差は予測不可能なタイミングでアプリケーションにとっての問題に発展することがあり，そうなると非常に厄介である．

これらの問題は，データフレームライブラリが内部データ表現を共有し，コア機能を共通の言語で実装することで回避可能である．

Apache Arrowでは，データフレームAPI，クエリAPI，データ入出力APIについて，次の方法でこれらの問題の解決を目指している．その方法は，C++やJavaのようにハードウェアの性能を引き出せる少数のプログラミング言語でコアライブラリを実装し，それ以外のPython，R，Rubyのような言語はコアライブラリのバインディングを作成する方法である．こうすることで，異なる言語で同じ機能を実装する回数を最小限に抑えられ，どのプログラミング言語でも最高のパフォーマンスを発揮させることが可能となる．アルゴリズムで操作する対象となるデータ構造が共通であるため，実装言語が異なっていてもアプローチを共有できる可能性があり，異なるプログラミング言語の開発者同士でノウハウを共有し合い，開発の労力を減らせる可能性もある．

2.3　RubyがApache Arrowに対応するメリット

RubyがApache Arrowに対応することで，具体的にどのようなメリットがあるのだろうか．本節では，筆者がメリットであると考えている2つの要素について述べる．

2.3.1　Apache Arrowをハブとするデータ処理コンポーネントのネットワークに容易に入れるようになる

RubyがApache Arrowに対応することによって，Apache Arrowに対応したデータ処理コンポーネントのエコシステムに容易に参入できるようになる．つまり，図1の右側に示したApache Arrowによってコンポーネント間のデータ連携を効率化した後の世界にRubyが参加できるようになる．これは非常に大きなメリットであろう．

2.3.2　Apache Arrowが持つ将来性を直接Rubyの将来性に繋げられる

最近になってApache Arrowを応用した製品や研究プロジェクトが次々と登場してきている．

HeteroDB社が開発したPG-Strom[23]は，PostgreSQLから効率よくGPUを利用できるようにした拡張モジュールである．SQLクエリから自動的にGPU用のコードを生成し，クエリの実行をGPU上で非同期かつ並列に実行できる．PG-StromはもともとApache Arrowとは無関係であったが，2019年にリリースされたバージョン2.2でArrowフォーマットに対応し，SSDとGPUの間でArrowフォーマットのデータを直接転送することができるようになった．その結果，PG-Stromによって4倍以上に改善れていたクエリ実行のスループットが，Arrowフォーマットを利用することでさらに3倍以上改善された[30]．

Apache Arrowを応用する研究例としてはFletcher[25],[26]がある．Fletcherは，データ処理でFPGAアクセラレータを利用するためのフレームワークであり，FPGAにおけるデータ表現としてArrowフォーマットを採用している．Arrowフォーマットを利用したことで，演算効率が向上し，正規表現マッチで9〜49倍，文字列の書き出しで1.3倍，K-meansクラスタリングで最大2.7倍の速度向上を達成している[25]．

RubyがApache Arrowに対応するもう1つのメリットは，Apache Arrowの周囲で行われるこのような先進的な取り組みをRubyの将来性に直接繋げられることにある．

3. Red Arrow - RubyのApache Arrow対応

3.1 Rubyにとっての適切なアプローチとは

RubyをApache Arrowに対応させる方法には2つの道が存在する．すなわち，（1）C++やJavaのようにApache Arrowのデータフォーマットを採用したデータ処理ライブラリをRubyで実装する方法，そして（2）C++で実装されたライブラリのバインディングを開発する方法である．

2つの方法を比較すると，明らかに後者のバインディング開発がRubyには適している．RubyはC++やJavaと異なり遅いプログラミング言語であり，実用的な速度を発揮するには多くの機能をC++言語で拡張ライブラリとして実装しなければならないだろう．そもそも，前者の方法は前述のApache Arrowが目指している方向からブレてしまう．したがって，C++ライブラリのバインディングを開発することがRubyのApache Arrow対応の唯一の最適な道である．

次に検討すべきことはバインディングの実装方法である．バインディングは，すべてを手書きする方法，自動生成の仕組みを使う方法，そして実行時に動的に生成する方法のいずれかで作成できる．

すべてを手書きする方法は，PythonとRのバインディングで採用されている実装方法である．この方法では，C++で拡張ライブラリを手書きすることになる．その拡張ライブラリでは，C++ライブラリにおけるクラスに対応するRubyのクラスを定義し，そのクラスのインスタンスがC++ライブラリのオブジェクトをラップできるようにRubyオブジェクトの構造体を定義する．

バインディングの自動生成は，SWIG[31]が広く普及している．SWIGを採用する場合，C++ライブラリとRubyライブラリの間の対応関係を記述するSWIG独自の定義ファイルを作り，それを維持していくことになる．この定義ファイルから，SWIGが拡張ライブラリのC++コードを自動生成する．つまり，自動生成の方法でも拡張ライブラリが作られる．しかし，現行のSWIGの定義ファイルのみでは，異なる言語で共有できる部分と共有出来ない部分があり，Ruby用の定義ファイルをそのまま流用してほかの言語のバインディングを生成できるわけではない．

実行時にバインディングを動的生成するにはForeign Function Interface（FFI）を利用する．Rubyでは標準ライブラリFiddleがFFIの機能を提供しているため，Fiddleを利用して実行時に共有ライブラリを動的ロードし，オブジェクトとして取り出した関数を呼び出すようなAPIをRubyで記述する．このような仕組みは存在するが，残念ながらApache ArrowのC++ライブラリを対象とする場合は適切な方法ではない．その理由は（1）C++ライブラリの関数名を表すシンボルがマングルされてしまうこと，そして（2）C++ライブラリの一部の機能がテンプレート機能を使って実装されているため共有ライブラリではなくヘッダファイルに存在することである．C++ライブラリに対してバインディングの動的生成を行うには，C言語用のバインディングを作成し，そのCライブラリをFFIで動的ロードして利用する必要がある．

このように，バインディングの作成方法は3つあり，それぞれに一長一短の特徴がある．C++ライブラリが対象の場合，バインディングを手書きする方法か，SWIGのような仕組みで自動生成するのが現実的な選択肢として残るだろう．

しかし，Apache ArrowのRubyバインディングであるRed Arrowは，そのどちらの選択肢も採用しなかった．実際に採用された方法は，C言語用のバインディングを作成し，Rubyバインディングを動的生成する方法である．ただし，C言語用のバインディングはGLib[32]を用いて実装すること，そしてRubyバインディングの動的生成にはGObject Introspection[33]を利用する点が特徴的である．その理由は3.3節で後述するように，GLibとGObject Introspectionの利用が汎用的なバインディング開発手法になり得るからである．

3.2　GObject IntrospectionによるRubyバインディング生成の仕組み

GObject Introspectionによってバインディングが生成される際に使用されるファイルとその関係を**図5**に示す．

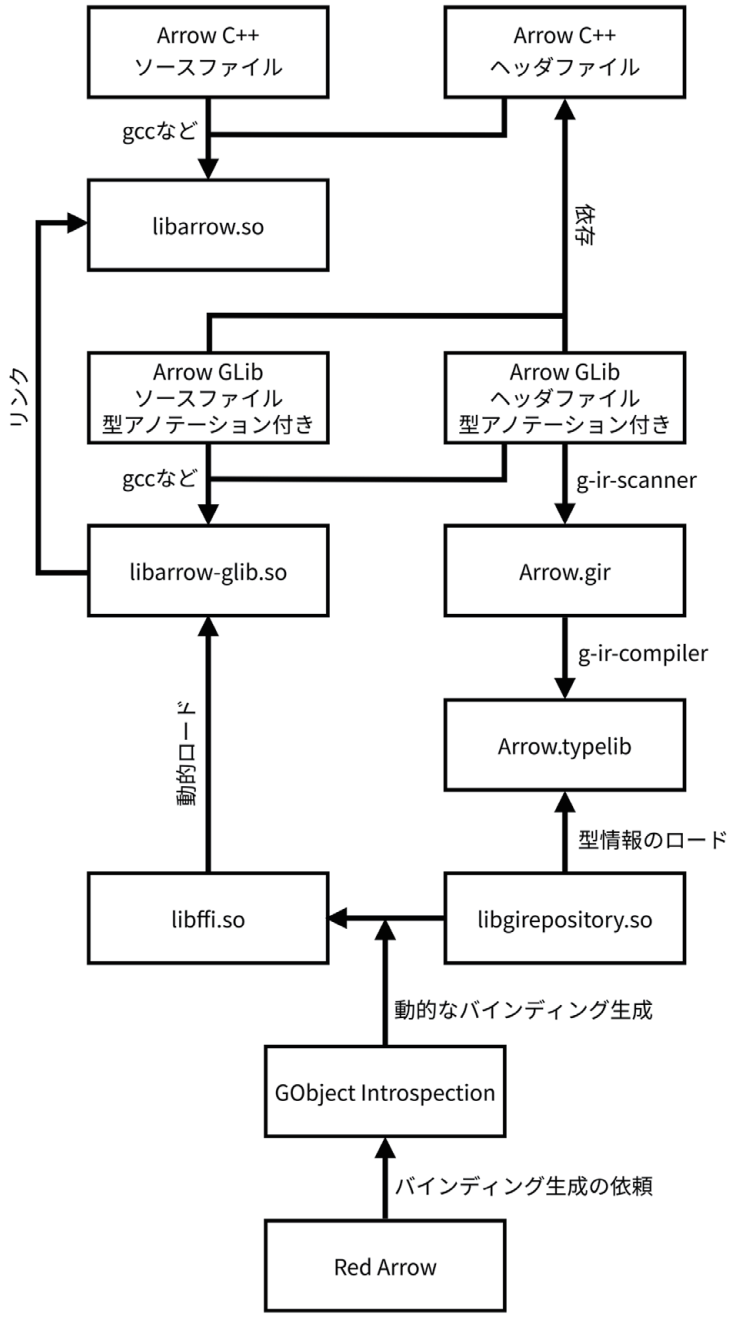


図5　GObject Introspectionによるバインディング生成

GObject Introspectionを利用するには，GObject Introspection用の型アノテーションが記述されたソースコードが必要となる．Red Arrowの場合，それはArrow GLibライブラリのソースコードである．型アノテーションつきソースコードから，GObject Introspectionのツールによって型情報データベースであるtypelibファイルArrow.typelibがArrow GLibライブラリのビルド時にlibarrow-glib.soと一緒に生成される．

実行時には，Red ArrowがGObject Introspectionにライブラリ名"Arrow"を指定してバインディング生成を依頼する．するとGObject IntrospectionがArrow.typelibをロードして型情報と依存ライブラリの一覧を取得，依存ライブラリのロードとlibarrow-glib.soのlibffi.soによる動的ロードを行う．そして，GObject IntrospectionのRubyバインディングが型情報に基づいてArrow GLibのクラスや関数に対応するRubyのクラスやメソッドなどを定義する．

RubyからArrow C++の機能を呼び出すには，Rubyバインディングのオブジェクトに対して対応するメソッドを呼び出せばよい．すると，Rubyオブジェクトが保持しているGLibオブジェクトを取り出してArrow GLibの関数を呼び出す．Arrow GLibの関数は，GLibオブジェクトの中からArrow C++のオブジェクトを取り出し，そのオブジェクトのメンバ関数を呼び出す．**図6**に，RubyのArrow::RecordBatchオブジェクトのschemaメソッドの呼び出しがArrow C++の関数呼び出しに至る流れを示す．

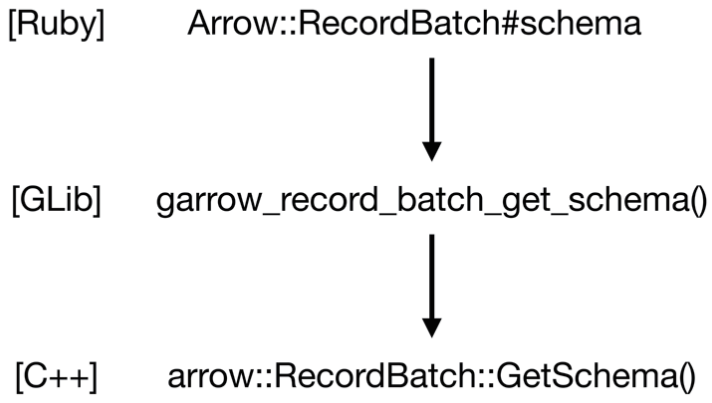


図6　RubyからArrow C++の関数を呼び出すときの流れ

3.3　GObject Introspectionを採用する利点

Red ArrowでGObject Introspectionを採用するメリットはあるのだろうか．Red Arrowが目的とするArrow C++がC++ライブラリであるため，一見するとGObject IntrospectionのためにArrow GLibをわざわざ作らねばならなくなっているように見える．つまり，バインディングをC++で手書きする，もしくはSWIGで自動生成するよりも開発コストが増えているように見える．

ここでポイントとなるのは，GObject IntrospectionがRuby専用ではないことである．つまり，Arrow GLibはRuby以外の言語でも，GObject Introspectionに対応している言語であれば理論的には利用可能である．

これは，Red Arrowにとってのメリットになっていないように見えるが，実はそうではない．Arrow GLibは現時点ではRed Arrowの開発者によって開発されている．しかし，将来的にRuby以外の言語のバインディングの開発者がArrow GLibの開発に参加してくれる可能性がある．そうなると，Red Arrowの開発者がArrow GLibの開発に割く労力が減るため，結果的にRed Arrowの開発コストが減る．これがRed ArrowでGObject Introspectionを採用するメリットである．

3.4　GObject Introspectionによるオーバーヘッド

Red Arrowを用いると必ずArrow GLibの関数を経由してArrow C++の関数が呼び出される．そのため，C++ライブラリへのバインディングを直接記述する場合には発生しない2種類のオーバーヘッドが生じる．garrow_record_batch_get_schema関数（**図7**）を例にこれらのオーバーヘッドについて説明する．

```
GArrowSchema *
Garrow_record_batch_get_schema(GArrowRecordBatch *record_batch)
{
  // 引数で受け取ったGLibオブジェクトから
  // C++オブジェクトを取り出す
  const auto arrow_record_batch =
    garrow_record_batch_get_raw(record_batch);

  // Rubyから呼び出したかった目的のC++関数の呼び出し
  auto arrow_schema = arrow_record_batch->schema();

  // C++オブジェクトをラップするGLibオブジェクトを
  // 生成して戻り値とする
  return garrow_schema_new_raw(arrow_schema);
}
```

図7　garrow_record_batch_get_schema関数の定義

1つ目のオーバーヘッドは関数呼び出しの増加である．これはArrow GLibの関数を経由することで生じる．Arrow GLib関数の内部では，第1引数で渡されたGLIbオブジェクトからC++オブジェクトを取り出し，目的のC++関数を呼び出し，そして必要な場合は戻り値をGLibオブジェクトでラップする処理も行う．garrow_record_batch_get_schema関数はこの3つの処理をすべて実行しているため，C++関数を直接呼ぶ場合と比較すると，見えている部分だけで3回の関数呼び出しが追加されている．

2つ目のオーバヘッドはメモリ割り当ての増加である．これは，C++オブジェクトをGLibオブジェクトでラップするために発生する．garrow_record_batch_get_schema関数では，戻り値となるarrow::SchemaオブジェクトをGArrowSchemaオブジェクトでラップするためにメモリの割当が発生している．

このようなオーバーヘッドは，たった数回のメソッド呼び出しで発生する程度なら問題にならない．また，メソッドが何度も呼ばれる場合であっても，メソッド内で呼ばれるC++関数の処理に時間がかかる場合はオーバーヘッド自体がほとんど気にならなくなる．

オーバーヘッドが問題となるのは，Rubyでループを回し，その中でバインディングのメソッド呼び出しを行う場合である．たとえば，RecordBatchを行方向にスキャンし，各行をRubyの配列に変換したものを集めて1つの配列にする処理をRubyで書くと**図8**のようになる．この図のraw_recordsメソッドでは，get_column_dataメソッドの呼び出しがgarrow_record_batch_get_column_data関数の呼び出しに対応し，その中でGArrowArrayオブジェクトが毎回割り当てられる．したがって，行数×列数分のGArrowArrayオブジェクトが割り当てられ，捨てられることになる．このようなオーバーヘッドを回避するには，**図9**のように各列の配列オブジェクトを事前に作成してキャッシュしておく必要がある．Red Arrowでは，RecordBatchの行と列のアクセスで無駄なオブジェクトが生成されないように，このような工夫がすでに実装されている．

```
class Arrow::RecordBatch
  def raw_records
    n_rows.times do |i|
      n_columns.times.map do |j|
        # 第 j 列目の配列オブジェクトを取得
        array = get_column_data(j)
        # 配列オブジェクトから i 番目の値を取得
        array[i]
      end
    end
  end
end
```

図8 RecordBatchを行の配列に変換するraw_recordsメソッドのRubyによる実装

```
class Arrow::RecordBatch
  def raw_records
    # 先に配列オブジェクトを作っておく
    arrays = n_columns.times.map do |j|
      get_column_data(j)  # 第 j 列の配列オブジェクト
    end
    n_rows.times do |i|
      n_columns.times.map do |j|
        # 事前にキャッシュした配列オブジェクトから値を取得
        arrays[j][i]
      end
    end
  end
end
```

図9 RecordBatchを行の配列に変換するraw_recordsメソッドのRubyによる実装

3.5　Rubyバインディング独自機能のC++による実装

図9のように工夫をしたとしても，まだ回避可能なメモリ割り当てが発生する可能性が残っている．それは，文字列型の配列から値を取り出す場合に生じるGBytesオブジェクトの割り当てである．文字列型の配列から値を取り出すときに呼び出される

garrow_binary_array_get_value関数は，**図10**に示した処理を行う．最後の行のreturnで返している値が新たに割り当てられるGBytesオブジェクトである．このGBytesオブジェクトはArrow GLibを経由しなければ必要ないものである．

```
GBytes *
garrow_binary_array_get_value(GArrowBinaryArray *array,
                              gint64 i)
{
  // GLibオブジェクトからC++オブジェクトを取り出す
  // 戻り値はarrow::Array型のポインタ
  auto arrow_array = garrow_array_get_raw(array);
  // arrow::BinaryArray型のポインタにキャストする
  auto arrow_binary_array =
    std::static_pointer_cast<arrow::BinaryArray>(arrow_array);
  // 第i番目の文字列の範囲を表すビューオブジェクトを得る
  auto view = arrow_string_array->GetView(i);
  // ビューと同じ範囲を表現するGBytesオブジェクトを返す
  return g_bytes_new_static(view.data(), view.length());
}
```

図10　garrow_string_array_get_string関数の定義を簡略化したコード

このように，Arrow GLibでは必要だがRubyでは不要となる操作は，GObject Introspectionによって生成されるバインディングを使う以上は回避できない．これを回避するには，C++で機能を実装する必要がある．実際にRed Arrowは，前節で例示したraw_recordsのC++実装を拡張ライブラリとして提供している．

4. Rubyのデータ処理に関する関連技術

Red ArrowによるApache Arrow対応は，共通データフォーマットに対応し，データ処理コンポーネントとの連携の可能性を高めることによってRubyをデータ処理に対応させるアプローチである．これとは異なるアプローチでRubyをデータ処理に対応させる事例がいくつかある．

1つはPyCall[34]である．PyCallはRubyからPython用のライブラリを利用するための言語ブリッジである．PyCallはRubyと同一プロセス下でPython処理系を動かし，Python処理系のC APIを用いてPythonとRubyの間のデータ連携を実現する．

過去に存在し，現在は開発が終了してしまったものもいくつかある．RSRuby[35]とRinRuby[36]は，R言語とのブリッジを実現するライブラリである．RSRubyは，Rubyとは別のプロセスで起動されたR処理系とソケット通信でデータ連携する．RinRubyは，PyCallのようにRubyと同一プロセスでR処理系を動かし，R言語のC APIを利用してデータ連携する．

ruby-spark[37]も開発が終了してしまったライブラリである．これは，Sparkとの連携を実現するものであった．

RSRuby，RinRuby，ruby-sparkの開発が終了してしまった最大の要因は，利用者がそこまで増えなかったことだと筆者は推測している．利用者が増えないとアクティブな開発者も増えない．その状況で，主な開発者がこれらのライブラリを必要としなくなると開発が止まり，開発を引き継ごうとするものも現れず，プロジェクトが自然消滅してしまう．

5. Rubyのデータ処理対応に関する今後の展望

論文の最後にRubyのデータ処理対応についての今後の展望についてまとめる．

5.1 Red Arrowの開発の継続

本稿では，Rubyをデータ処理に対応させていくにあたり，Apache Arrowに対応することが肝心であることを説明した．つまり，Red Arrowの存在がRubyのデータ処理対応にとって非常に重要となる．

Apache Arrowはすでに完成したライブラリではない．むしろ，最近バージョン1がリリースされたばかりで，データフォーマットとAPIの両方について発展途上である．C++ライブラリに限定しても，まだメモリ上のテーブルデータの表現と操作，基本的な演算，ファイルI/Oなどの基本機能，および高次機能に関してはデータソース抽象化の一部のみ提供されている．C++ライブラリの高次機能として計画されているクエリエンジンは本稿執筆時点で開発が始まったばかりであり，さらにデータフレームについては計画があるだけの状態である．

Apache Arrowは今後も速いスピードで開発が進行し，どんどん新しい機能を増やしていくはずである．Red Arrowが開発を止めず，Apache Arrowの進化に追従していくことが今後の展開として期待される．

5.2 Rubyのデータ処理エコシステムの拡充

Apache Arrowへの対応だけが充実していても意味がない．Rubyのデータ処理対応を成功させるには，Rubyプログラマが使用できるデータ処理ツールを拡充しなければならない．

昨今，Rubyコミュニティの一部で，Ruby用のデータ処理ツール開発が盛り上がっている．Red Arrowの開発を推進しているRed Data Tools[38]は，可視化ライブラリCharty[39]も開発している．そのほかにも，機械学習ライブラリRumale[40]や，深層学習用のライブラリであるlibtorchのバインディングTorch.rb[41]が作られたりしている．この盛り上がりを長期に渡り維持していくことも，Apache Arrowへの対応と同程度に重要なことだと筆者は考えている．

5.3 Ruby外のコンポーネントとの連携強化

データ処理システムは多くのコンポーネントを組み合わせて作らなければならない．Rubyで作られたシステムをデータ処理に対応させるには，Ruby外のコンポーネントとの連携強化が必要である．

Red ArrowによってApache Arrowフォーマットを活用できる環境が整備されているので，これを利用して他言語のコンポーネントへのアクセス手段を整備するとよいだろう．たとえば，今度こそPySparkのようなSpark連携を実現し，Sparkによる分散データ処理をRubyでも実装しやすい環境をつくる良いタイミングかもしれない．また，Apache Arrowのための分散計算プラットフォームであるApache Arrow Ballista[42]との連携を実現するのも魅力がある．

参考文献

1）Flanagan, D., まつもとゆきひろ（著），卜部昌平（監訳），長尾高弘（訳）：プログラミング言語Ruby，オライリー・ジャパン（2009）．
2）Ruby on Rails : https://rubyonrails.org
3）Cookpad : https://cookpad.com
4）GitHub : https://github.com
5）Shopify, https://shopify.com
6）Chevalier-Boisvert, M., Gibbs, N., Boussier, J., Wu, S. X., Patterson, A., Newton, K. and Hawthorn, J. : YJIT : A Basic Block Versioning JIT Compiler for CRuby, In Proceedings of the 13th ACM SIGPLAN International Workshop on Virtual Machines and Intermediate Languages, VMIL 2021, pp.25-32 (Oct. 2021).
7）Gallinelli, N. : The 10 Best Data Science Programming Language to Learn in 2021, https://flatironschool.com/blog/data-science-programming-languages (Mar. 2021)
8）Python : https://python.org
9）The R Project for Statistical Computing : https://www.r-project.org
10）Apache Spark : https://spark.apache.org
11）ECMA-404 : The JSON Data Interchange Format (2nd ed.), ECMA International (Dec. 2017).
12）経済産業省：産業界におけるデジタルトランスフォーメーションの推進，https://www.meti.go.jp/policy/it_policy/dx/dx.html
13）Apache Arrow : https://arrow.apache.org
14）Red Arrow - Apache Arrow Ruby : https://github.com/apache/arrow/tree/master/ruby/red-arrow
15）Rodriguez, S. A., Chakraborty, J., Chu, A., Jimenez, I., LeFevre, J., Maltzahn, C. and Uta, A. : Zero-Cost, Arrow-Enabled Data Interface for Apache Spark, arXiv:2106.13020 (cs.DC) (June 2021).
16）Apache HBase : https://hbase.apache.org
17）Apache Kudu - Fast Analytics on Fast Data : https://kudu.apache.org
18）Ousterhout, K., Rasti, R., Ratnasamy, S., Shenker, S. and Cun, B.-G. : Making Sense of Performance in Data Analytics Frameworks, Proceedings of the 12th USENIX Syposium on Networked Systems Design and Implementation (NSDI 15), pp.293-307 (May 2015).
19）Apache Arrow Overview : https://arrow.apache.org/overview/
20）Arrow Columnar Format Version 1.0 : https://arrow.apache.org/docs/format/Columnar.html
21）Zhang, H., Chen, G. and Tan, K.-L. : In-Memory Big Data Management and Processing : A Survey, IEEE Transactions on Knowledge and Data Engineering, Vol.27, No.7, pp.1920-1948 (July 2015).
22）PySpark Documentation : http://spark.apache.org/docs/latest/api/python/
23）PG-Strom Manual : https://heterodb.github.io/pg-strom/

24）RAPIDS - GPU-Accelerated Data Analytics & Machine Learning : https://developer.nvidia.com/rapids
25）Peltenburg, J.W., Straten, J.V., Wijtemans, L., Leeuwen, L.V., Al-Ars, Z. and Hofstee, H.P. : Fletcher - A Framework to Efficiently Integrate FPGA Accelerators with Apache Arrow, Proceedings - 29th International Conference on Field-Programmable Logic and Applications, FPL 2019, pp.270-277 (2019).
26）Peltenburg, J.W., Straten, J.V., Brobbel, M., Hofstee, H. P. and Al-Ars, Z. : Supporting Columnar In-memory Formats on FPGA : The Hardware Design of Fletcher for Apache Arrow, In Hochberger, C., Koch, A., Diniz, P., Woods, R. and Nelson, B. (Eds.), Applied Reconfigurable Computing : 15th International Symposium, ARC 2019, Proceedings, pp.32-47 (2019).
27）McKinney, W. : Apache Arrow and the Future of Data Frame with Wes McKinney, Tech Talk, ACM, https://learning.acm.org/techtalks/apache (July 2020)
28）Differences to R's factor : in pandas User Guide, https://pandas.pydata.org/pandas-docs/stable/user_guide/categorical.html#differences-to-r-s-factor
29）McKinney, W. : Some pandas Database Join (merge) Benchmarks vs. R base::merge, blog post, https://wesmckinney.com/blog/some-pandas-database-join-merge-benchmarks-vs-r-basemerge/ (Jan. 2012)
30）Kohei, K.：PostgreSQLだってビッグデータ処理したい！！〜Apache Arrowが可能にする毎秒10億レコードのデータ処理〜，講演資料 https://www.slideshare.net/kaigai/20191211apachearrowmeetuptokyo (Dec. 2019)
31）SWIG : http://www.swig.org
32）GLib - 2.0 : https://docs.gtk.org/glib/
33）GObject Introspection : https://gi.readthedocs.io/en/latest/
34）PyCall : Calling Python Functions from the Ruby Language, https://github.com/mrkn/pycall.rb
35）Ruby - R bridge : https://github.com/alexgutteridge/rsruby
36）Dahl, D. B. and Crawford, S. : RinRuby : Accessing the R Interpreter from Pure Ruby, Journal of Statistical Software. Vol.29, No.4, pp.1-18 (Nov. 2009).
37）Ruby Wrapper for Apache Spark : https://github.com/ondra-m/ruby-spark
38）Red Data Tools - Data Processing with Ruby! : https://red-data-tools.github.io/
39）Charty - Visualizing Your Data in Ruby : https://github.com/red-data-tools/charty
40）Rumale is a Machine Learning Library in Ruby : https://github.com/yoshoku/rumale
41）Deep Learning for Ruby, powered by LibTorch : https://github.com/ankane/torch.rb
42）Ballista : A Distributed Scheduler for Apache Arrow, https://arrow.apache.org/blog/2021/04/12/ballista-donation/

村田賢太（正会員）muraken@gmail.com

（株）Speeeにてフルタイム OSS開発者に従事．2010年よりRubyコミッタ，2018年よりApache Arrowコミッタも務める．Rubyをデータ処理に対応させることを目指して精力的に活動中．

須藤功平（非会員）kou@clear-code.com

（株）クリアコード代表取締役．2004年よりRubyコミッタ，2017年よりApache Arrow PMCメンバ．2017年よりRed Data Toolsプロジェクトを開始．

受付日：2021年9月1日
採録日：2021年10月27日
編集担当：青木学聡（京都大学）

会誌「情報処理」Vol.63 No.2（Feb. 2022）「デジタルプラクティスコーナー」

特集号招待論文

大阪府の特定健康診査データの因果探索

大山飛鳥[1]　古徳純一[1, 2]　土岐　博[1]

[1]大阪大学キャンパスライフ健康支援・相談センター　[2]帝京大学大学院医療技術学研究科

AI研究はさまざまな分野で応用されているが，予測に用いられるモデルは必ずしも因果関係を反映しておらず，相関関係の記述にとどまることが多い．我々は，大阪府保険者協議会および大阪府国民健康保険団体連合会の協力で大阪府民60万人規模の特定健診データの提供を受け，DirectLiNGAMと呼ばれる数理モデルを用いた健診データ間の因果関係の構築を行い，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，実体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．

1. 因果探索の必要性

相関と因果は違うというのは，統計学を習った経験があるなら，誰しも知っていることであるが，相関をデータから取り出すのは簡単でも，因果をデータから取り出すのが可能なのか疑問に思われるかもしれない．実は，ある種の妥当な仮定の下で，データだけからかなりの因果的構造を推定することができる．

因果の構造を，原因となる変数と結果となる変数を矢印でつないで変数間の原因と結果の形を図形状に表したものを因果ダイアグラムと呼び，データから因果ダイアグラムの推定を行うことを因果探索と呼ぶ．厳密にはあまり区別されての使用はされていないが，傾向スコアのように解析者が因果モデルを仮定し，はたして因果があると言ってよいかどうか推測する作業のことを因果推論と呼ぶ．たとえば，医療ビッグデータのように非常に多くの要因が絡み合っているような問題を取り扱う場合には，あらかじめ因果関係を仮定することなしにデータ間の因果関係のネットワークを抽出する因果探索を実践できれば大変有用である．

本稿では，筆者らの行った大阪府全体の特定健康診査データ（以下，特定健診データ）への因果探索の応用[1]について実体験を交えながら赤裸々に語る．数値データのみから自動的に因果ダイアグラムを構築する因果探索のモデルとして使用したDirectLiNGAM[2],[3],[4],[5]の数理的側面について次章で簡単に紹介し，その後，陥りがちなピットフォールについて注意を喚起しつつ，解析の全容についてお伝えする．

2. LiNGAMの数理

2.1 LiNGAMの構造方程式

ここから，因果探索のモデルである LiNGAM の解説に入る．LiNGAM は，考えたい変数 x_i，$(i = 1, 2, \ldots, p)$ 間の因果関係として，式 (1) のような線形の関係を仮定するモデルである．このとき，外生変数と呼ばれる項 e_i，$(i = 1, 2, \ldots, p)$ は，非ガウス分布に従うと仮定する．LiNGAM の構造方程式では，外生変数 e_i が，いわば外から与える初期値で，これらはある確率分布からのサンプルとして実現される．我々が観測する量は，これらの線形結合として内生されるというのが，LiNGAM の世界観である．ここで，b_{ij} は，変数 x_j から変数 x_i への直接効果を表現する係数で，自分自身へのループのようなフィードバックは考えないことにする．これを式で表すと，

$$x_i = \sum_{j \neq i} b_{ij} x_j + e_i, \quad (i = 1, 2, \ldots, p) \quad (1)$$

のように書ける．このままではプログラムする上で計算上の見通しが悪いので，内容は同じであるが，行列での表現に書き換えたものが式（2）または式（3）である．式（1）での $\sum_{j \neq i}$ の部分は，行列 B の対角成分が0として表現されている．

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} = \begin{pmatrix} 0 & b_{12} & \cdots & b_{1p} \\ b_{21} & 0 & & \vdots \\ \vdots & & \ddots & \\ b_{p1} & & \cdots & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{pmatrix} \quad (2)$$

あるいは，

$$\boldsymbol{x} = B\boldsymbol{x} + \boldsymbol{e} \quad (3)$$

ただし，

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}, B = \begin{pmatrix} 0 & b_{12} & \cdots & b_{1p} \\ b_{21} & 0 & & \vdots \\ \vdots & & \ddots & \\ b_{p1} & & \cdots & 0 \end{pmatrix}, \boldsymbol{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{pmatrix} \quad (4)$$

となる．

LiNGAMの目指すところは，外生変数 $\boldsymbol{e}$ が非ガウス分布に従うという仮定のもとで，与えられた観測量 $\boldsymbol{x}$ から，係数行列 B を推定することである．

2.2 回帰による因果順序の推定

因果ダイアグラムの推定法にはさまざまな方法が考案されているが，今回の解析では，「2変数の因果関係を考えた場合，説明変数が原因である場合には残差と説明変数は独立だが，説明変数を結果にしてしまった場合，説明変数と残差が従属する」という性質を用いる．

どのようなことか，例を通して見てみよう． e_1, e_2 を独立な外生変数とし，次のように確率変数 x_1, x_2 を生成するLiNGAMを考える．

$$\begin{cases} x_1 = e_1 \\ x_2 = b_{21}x_1 + e_2 \end{cases} \quad (5)$$

このモデルを， $b_{21} = 0.8$ として， e_1, e_2 は[-0.5,0.5]の一様乱数からそれぞれ10,000個生成して，シミュレーションを行ったものが**図1**である．

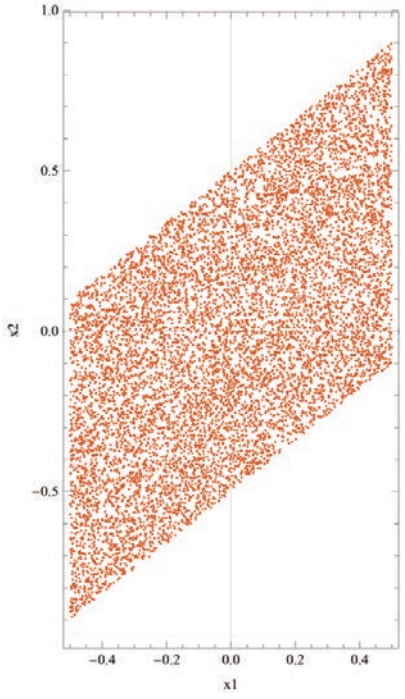


図1　x_1 と x_2 の散布図

2.2.1　原因となる変数が説明変数のとき

このとき，回帰直線は $x_2 = b_{21}x_1 + e_2$ の形になるから，残差は e_2 である．仮定から，e_1, e_2 は独立であるから， x_1, e_2 も独立になる．この様子を示したのが**図2**である．図下の散布図から分かるように， x_1 軸，残差軸のどちらに平行な直線で切っても，同じ確率密度分布が得られるから，このとき，説明変数と残差は独立であることが見てとれる．

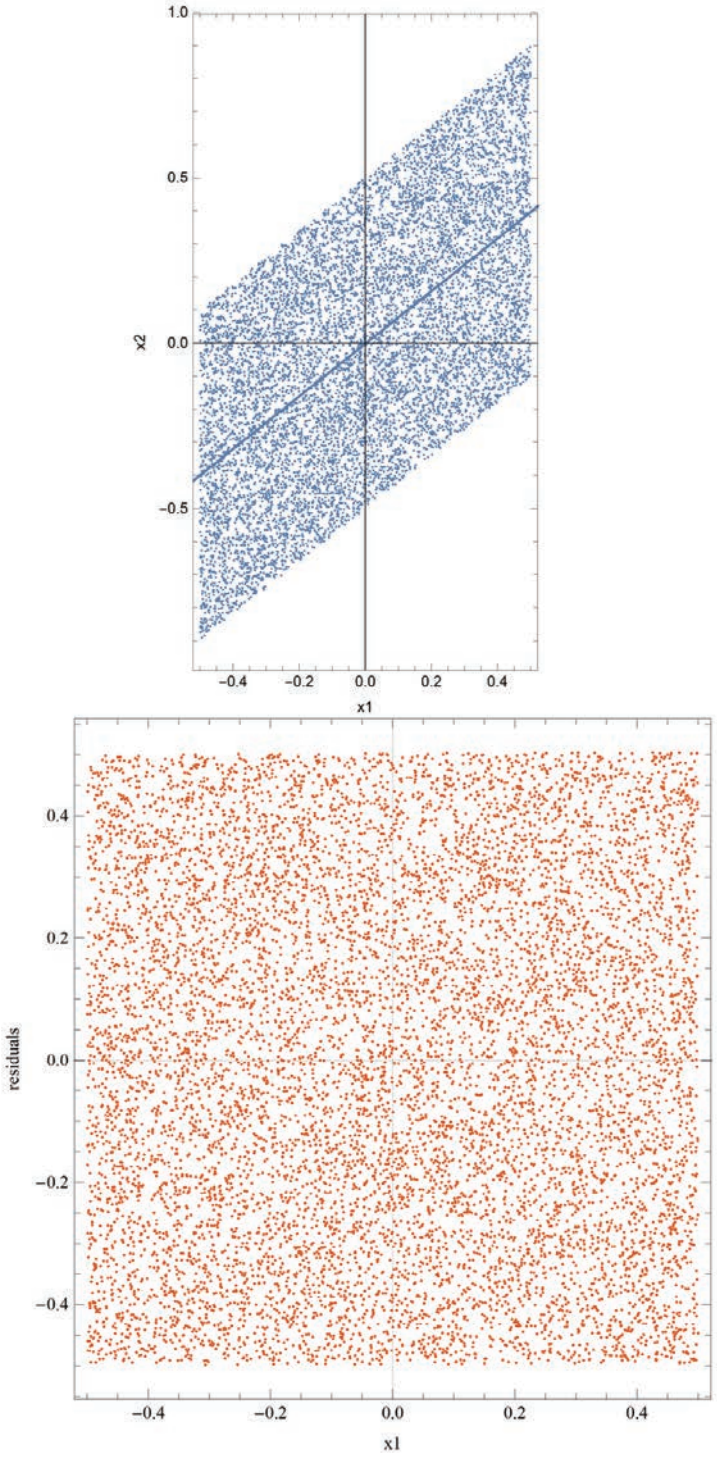


図2　x_1を説明変数にして回帰した直線（上），説明変数と残差の散布図（下）

2.2.2　結果となる変数が説明変数のとき

次は，$x_1 = b_{12}x_2 + \varepsilon$ という形に回帰してしまったときのことを考えよう．

シミュレーションしたデータに対する回帰の例で示したのが，**図3**である．特に，下の図を見ると，x_2 一定直線で切った切り口で，残差の確率密度分布が大きく変わってしまうことから，残差と説明変数が独立でないことが見てとれる．

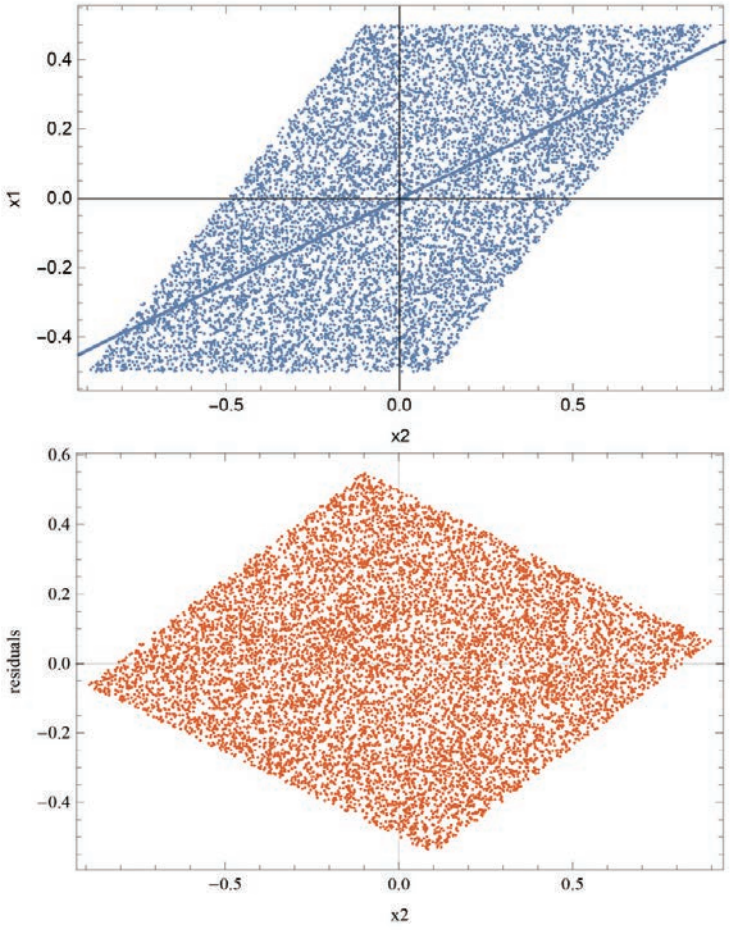


図3　x_2 を説明変数にして回帰した直線（上）と，説明変数と残差の散布図（下）

この話を一般化して，変数の数が3つ以上の場合にも，因果の最初の変数を推定することができる．

定理　LiNGAM モデル（式（1））を仮定する．観測数 n は，推定誤差を無視できるほど十分に大きいとする．説明変数を x_j ，応答変数を x_i として線形回帰したときの残差を

$$r_i^{(j)} = x_i - \frac{\mathrm{cov}\left(x_i, x_j\right)}{\mathrm{var}\left(x_j\right)} x_j \ (i, j = 1, \ldots, p,\ i \neq j) \quad (6)$$

と表す．このとき，観測変数 x_j が因果的順序において最初の変数になることができるのは，観測変数 x_j が，残差 $r_i^{(j)}\ (i = 1, \ldots, p,\ i \neq j)$ と独立であるときに限る．

この定理は，説明変数の因果的順序を決める上で，重要な役割を果たす．

2.2.3　2変数間の独立性を測る

それでは，次にLiNGAMの中で，説明変数と残差の独立性を相互情報量を使って測定することを考えてみよう．

変数 x_j と残差 $r_i^{(j)}$

$$\boldsymbol{y_j} = \begin{pmatrix} x_j \\ r_i^{(j)} \end{pmatrix} \quad (7)$$

として，相互情報量は，

$$I\left(\boldsymbol{y_j}\right) = H(x_j) + H(r_i^{(j)}) - H\left(\boldsymbol{y_j}\right) \quad (8)$$

とエントロピー$H(\cdot)$ を使って表現できる．

変数x_i と残差$r_j^{(i)}$

$$\boldsymbol{y_i} = \begin{pmatrix} x_i \\ r_j^{(i)} \end{pmatrix} \quad (9)$$

として，相互情報量は，

$$I\left(\boldsymbol{y_i}\right) = H(x_i) + H(r_j^{(i)}) - H\left(\boldsymbol{y_i}\right) \quad (10)$$

となる．

基本的には，これらの相互情報量の大小関係を調べるわけであるが，計算の前に，それぞれの変数を標準化して，標準化した量の変数の記号の上にバーをつけて表すことにする．たとえば，

$$\overline{x_i} = \frac{x_i - \mu\left(x_i\right)}{\sigma\left(x_i\right)} \quad (11)$$

などである．

相互情報量計算の前に，

$$\overline{\boldsymbol{y_j}} = \begin{pmatrix} \overline{x_j} \\ \overline{r_i^{(j)}} \end{pmatrix}, \overline{\boldsymbol{y_i}} = \begin{pmatrix} \overline{x_i} \\ \overline{r_j^{(i)}} \end{pmatrix} \quad (12)$$

と標準化したベクトルをつかって

$$m_{ji} = I\left(\overline{y_j}\right) - I\left(\overline{y_i}\right) \quad (13)$$

と定義される量 m_{ji} を導入して，$m_{ji} < 0$ なら $\overline{x_j}$ の方が $\overline{x_i}$ よりも因果的順序が早いと判断する．

2.2.4 DirectLiNGAMは因果的順序をどのように構成するか

変数間の独立性を測る方法を得たので，順序を構成することができる．複数の変数の因果的順序を求めるために，

$$M(x_i; U) = -\sum_{j \in U} [\min(0, m_{ji})]^2 \quad (14)$$

という量を導入し，$M(x_i; U)$ が最大になる変数を因果的順序の最初になることのできる変数であると推定する．ここで，U は候補となる変数の添え字の集合である．

最初の変数が決まったら，その変数の添え字を U から取り除く．そして，基本的には再び M を計算したいのだが，今，親の変数を仮に x_1 とすると，残りのすべての変数に親変数 x_1 の寄与が含まれているはずであるから，ほかの変数については，x_1 で回帰した残差を親変数の影響が取り除かれた変数と思うことにし，再び M を計算して残りの変数の親を決める．

今の例だと，ほかの変数というのは $x_i\ (i = 2, \ldots, p)$ のことだから，

$$x_i' = x_i - \frac{\mathrm{cov}(x_i, x_1)}{\mathrm{var}(x_1)} x_1,\ (i = 2, \ldots, p) \quad (15)$$

という計算で，親変数の影響を取り除き，U を $U = \{2, 3, \ldots, p\}$ と更新して，$x_i'\ (i = 2, \ldots, p)$ から $M(x_i'; U)$ を再び計算して，観測変数 x_i' の親を決める．これを繰り返していくことで，最終的に，変数 $x_i\ (i = 1, 2, \ldots, p)$ が因果的順序で $k(i)$ 番目であることが決定できる．

2.2.5 係数行列Bの推定

因果的順序 $k(i) \in \{1, 2, \ldots, p\}$ が推定できたら，今度は係数行列Bを構築する作業に入る．因果的順序で，自分より前にある変数すべてを説明変数として回帰し，そのときの偏回帰係数を係数行列Bの要素 b_{ij} とする．

すなわち，回帰モデルは

$$x_i = \sum_{j \in A_i} b_{ij} x_j + e_j \quad (16)$$

であり，ここで，A_iは

$$= \{j | k(j) < k(i)\} \quad (17)$$

で定義される，変数x_iよりも因果的に前にある変数の添え字集合である．

基本的には，上の回帰式で直線回帰を行いたいのだが，説明変数の数が多い場合には，適切な正則化を行って回帰をしなければならない．清水らの方法では，ここでAdaptiveLassoを適用している．我々は，Elastic Netを用いてスパースな偏回帰係数を求めた．

3. 大阪府民のKDBデータ

我々が今回解析に用いた国保データベース（以下，KDB）とは，国民健康保険団体連合会（以下，国保連合会）が保険者の委託を受けて行う各種業務を通じて管理する，国民健康保険（以下，国保）に加入している被保険者の健康記録や介護記録などの健康情報を保持したデータベースのことである．我々のプロジェクトは，生活習慣病等の予防と健康寿命延伸を目指した調査分析を目的としており，大阪府保険者協議会および大阪府国保連合会の協力によって，匿名化された大阪府民の国保加入者約200万人規模のKDBデータの提供を得ることができた．これらのデータを隅々まで把握し分析を行うためには，医療の専門知識を要することはもちろん，データハンドリングの能力も欠かすことはできない．それゆえ，筆者らは医療・数理統計・物理の研究者からなる研究グループを結成し，多くの分野の共同研究を行いながらプロジェクト推進を目指している．これからKDBデータを利用したデータ解析を行いたいと考える読者はもちろんのこと，筆者らにとってもこれほどまでのビッグデータを触れる機会というものはこれまでになく，合計で1テラバイトほどのテーブルデータをコンピュータでどのように扱ってきたかを医学物理を専攻してきた解析者の立場から記していきたい．

初めに大阪府民のKDBデータ（以下，大阪府KDBデータ）の概要を記述し，その後，大阪府KDBデータに触れて気付いたデータの問題点や注意点，実際に大規模な健診データに対してPythonを用いたデータ前処理や因果探索プログラムの実装について記載する．

3.1 大阪府KDBデータの構成

今回大阪府から提供を受けたKDBデータは，協力を得られた41市町村の2012年度から2018年度までの7年度分のデータが含まれており，大阪府KDBデータは以下の7種類のデータから構成されている．

- KDB被保険者台帳
- 健診結果
- 医療レセプト管理
- 医療傷病名
- 医療摘要
- 医療最大医療資源ICD別点数
- 介護給付実績

これらのファイルは大阪府の市区町村・処理年月ごとに1つのCSVファイルに分割されており，約68,000に分割されているCSVファイルをすべて合わせるとおおよそ1テラバイトの大規模なデータサイズとなる（**表1**）．特にKDB被保険者台帳，医療傷病名，医療摘要は100ギガバイトを超えるデータサイズであり，これらを利用した解析を行う場合はデータ操作に注意しなければならない．筆者らは提供を受けてからデータの分析を始める前に，匿名化された個人番号を二

重でハッシュ化した上で解析者に渡し，データファイルとヘッダ情報を別ファイルで保存しておく・使用するコンピュータを制限する・研究室からのデータ持ち出しの禁止などのルールを設けることで，セキュリティ対策にも細心の注意を払っている．

表1　大阪府KDBデータのファイル数とファイルサイズ

ファイル名	ファイル数	データサイズ (GB)
KDB 被保険者台帳	11059	113.6
健診結果	10833	19.6
医療レセプト管理	10833	24.5
医療傷病名	10833	229.3
医療摘要	10833	458.4
医療最大医療資源 ICD 別点数	10833	46.3
介護給付実績	3172	0.76
合計	68396	892.4

ちなみに，筆者はたまたま機会に恵まれて大阪府以外のKDBデータにも触れることがあったが，データベースの構造はKDBによって異なるようである．これは，KDBデータの場合，厚生労働省が全国レベルで一貫した管理が行われている「レセプト情報・特定健診等情報データベース（以下，NDB）」のデータとは異なり，市町村ごとにデータの管理や収集・構築が行われるため，作成した自治体によってはフォーマットなどが違ってくるのが理由だと考えられる．大阪府KDBデータは比較的整然とフォルダ分けされていた印象だったが，地域によっては（解析をするにあたっては）分かりづらい可能性もあるので解析前に必ず確認が必要である．

これらのファイル間は個人番号（ユニークID）で突合することが可能であり，解析に必要なファイルを適宜突合して解析に臨んでいる．本稿の解析では，これらのファイルのうち，主に被保険者の特定健診の結果が記載された健診結果ファイルを使用する．ただし，健診結果のファイルには被保険者の年齢と性別の情報が欠如していたため，これらの値は突合したKDB被保険者台帳から性別と生まれ年を抽出し，性別は被保険者台帳の値をそのまま利用，年齢については抽出した生まれ年と健診実施年月日との差分から計算し，これを利用した．

3.2　大阪府健診データのフォーマット

特定健診は，メタボリックシンドロームに着目して病気のリスクの有無を検査し，生活習慣病の予防を目的とした，40歳から74歳までの国民を対象に行われる健康診査である．75歳以上の場合は後期高齢者医療制度に加入することになるため，後期高齢者医療健康診査を受けることになり，特定健診の対象とは厳密には異なるが，本稿では特に区別することなくまとめて健診データと呼ぶ．なお，本稿の解析には単年の健診データのみしか利用していないが，縦断研究を行う場合は国民健康保険制度から後期高齢者医療制度への切り替えが存在することに注意しなければならない．

健診データには，身長や体重，腹囲などの身体計測，血圧を測り循環器系の状態を測定する血圧測定，中性脂肪やコレステロール値などの血中脂質検査，ALT（GPT）やAST（GOT）などの肝細胞障害の状態を測定する肝機能検査，空腹時血糖やHbA1cなど糖尿病の判定基準となる血糖検査などの項目が含まれる．また，脳卒中や心臓病などの既往歴や喫煙・飲酒，運動や食生活等の生活習慣に関する質問票データも含まれている．さらには医師の判断により実施された場合には，心電図検査や眼底検査，血清クレアチニンやeGFRといった血液検査の結果も記録されており，生活習慣病の発症リスクが高いと判断され，特定保健指導を受けた場合はそれらのデータも記録されている．そのほかに特定健診を実施した日付やメタボリックシンドローム因子の有無などの項目を含めると，全部で112の項目から構成されており，健診データを利用することで健診を受けた時点での被保険者の健康状態を取得することができる．

4. KDBデータを扱う際の注意点

KDBのような大規模データを研究に利用する最大の利点は，国保に加入した被保険者のデータを網羅した悉皆性の高いデータであり，対象者に偏りが少なく母集団の代表性に優れていることだと思われる．また，疫学研究でたびたび問題となるようなサンプルサイズも，余程稀なアウトカムを選択しない限りは問題とならないだろう．一方で，KDBデータを利用するにあたって非常に高い障壁となるのがデータハンドリングの困難さであり，研究の多くはデータの理解とクレンジングが大半の時間を占めることになると予想される．

ここでは，研究チームでの議論の中で得られた考察や，実際にデータを触ってみてようやく気付いた問題点などを筆者の経験に基づいて記載する．ここで記載した内容が初めてKDBデータを触る人たちの一助となれば幸いである．

4.1 被保険者数の数合わせ

大阪府KDBデータに何人の被保険者が存在するかを確認するため，筆者らは最初に被保険者数の数合わせを行った．まず，2012年度から2017年度までのすべての被保険者台帳のデータ（以下，台帳データ）からハッシュ化された個人番号を引き出し，重複を除いた人数をカウントした．さらに，ある特定の日に被保険者がどれくらい存在しているかをカウントした．筆者らはこれらの作業を数名の解析者が独自のプログラムで単独で行い，それぞれ算出した人数を突き合わせることでこれらが一致するかを確かめた．データをいただいた当初は膨大なデータ量と慣れないデータ形式から人数が大きく異なっていた．特に，台帳データの項目の意味を読み解く作業にほぼ1年を費やし，その間に台帳データそのものに不足分があることも判明した．初めは比較的人口の少ない市町村のみでの数合わせを行いながらKDBデータの癖とコーディング手法を学び，台帳の不備を補った後，さらに1カ月ほどの時間を費やして，最終的に解析者全員の個人番号の数が大阪府全体で9,421,332名で一致した．大阪府の人口が2021年度現在で大体900万人であるため7年間の台帳であることを考慮しても随分多いように思えるが，これはKDBデータの特性であり，国保から後期高齢制度に移った場合は追跡ができず，重複してカウントしてしまう，国保の場合は転居などで市町村を移動した場合は個人番号が変わってしまうといった理由が存在す

るためである．追跡調査を行う場合は，このような点にも注意しながら解析しなければならない．なお，台帳に含まれる人数をある一時点でカウントした場合は，どの年度でもおおよそ200万人ほどであった．

続けて，台帳と同様に2012年度から2018年度までの全データベースを対象にした個人番号のカウントも行った．両者の個人番号の数が一致することが確認できれば，被保険者台帳に含まれていないにもかかわらず，健診やレセプトデータなど，そのほかのファイルを持つ被験者が存在する，といったおかしなデータがないことが確認できるからである．この結果，両者の数値は一致し，台帳に含まれていないが健診データが存在するといった人は大阪府KDBにはいないことが確認できた（台帳には存在するが，そのほかのデータを持っていない被保険者はたくさん存在する）．

4.2　性別と生年月日の不一致

被保険者の人数が解析者間で一致したため，数合わせについてはこれで一件落着と思われた．だが，続けて個人番号と性別・生年月日の関係を調べた際に新たな問題が生じた．台帳ファイルには被保険者の性別と生まれ年の情報が記載されているが，同一の個人番号にもかかわらず，性別や生まれ年が異なる被保険者が含まれていたのである．全体の人数と比較すると大した数にはならないが，性別が一致しない被保険者は227名，生まれ年が異なる被保険者は57名存在した．台帳データの更新時に入力に誤りがあったのか，同一の個人番号がたまたま振られてしまったのか，理由を特定することは難しいが，いずれにせよ性別や生まれ年（年齢）はどのような解析を行うにしても非常に重要な因子となり得るのでこれらのデータを持つ被保険者は全解析から除き，残った9,421,051名を解析の候補とした．

4.3　保険加入期間の定義

今回提供を受けたKDBデータは国保加入者を対象としている．そのため，台帳データには被保険者ごとの「国保（後期）取得年月日」と「国保（後期）喪失年月日」が記録されており，被保険者が国保に加入していた期間を定義する必要がある場合にはこれらの項目を参照した．何らかの理由で国保の喪失があった場合でも再度取得していれば，特に問題がない限りは台帳には必ず記載されている．ただし上述したように，KDBデータでは市町村が変わると追跡できないので注意する．同様に介護資格の取得／喪失年月日も被保険者台帳には含まれており，介護認定を受けた期間などもここから取得可能である．余談ではあるが，有効期間の開始日と終了日という項目が台帳には記録されているが，これらが示す日付の定義が分からなかったため，現在はどの解析にも利用していない．

4.4　傷病名の定義

続いて目的変数や介入の有無などに傷病名を利用したい場合について言及する．KDBには診療報酬明細書である医療レセプトデータが含まれており，患者の傷病名と行われた医療行為が記録されている．そのため，疾患をアウトカムとするような研究デザインを計画する際には，医療傷病名データに含まれる傷病名やICD-10を利用したいと考えたが，研究チームでの議論の中で，医師からはこれらの利用には慎重になるべきとの指摘があった．

当時，筆者は知らなかったが，レセプトデータには疑い病名と呼ばれ，レセプトで検査や投薬を行うためには病名が確定していなくても，病名が疑われる状態で病名を記入しておくことで検査や投薬を実施するという慣習があるようだ．たとえば，鎮痛薬を使用する患者の胃炎防止のために胃薬を処方したい場合，実際に胃炎に罹患していなくてもレセプトに胃炎の病名を記載する，ということがあるそうだ．このような問題は厚生労働省が公表しているNDBを利用した解析でも議論されており，レセプトデータに記載された傷病名やICD-10の妥当性の調査を目的としたバリデーション研究なども行われている．もちろん，ICD-10を実際に利用するかどうかはどの傷病を対象にしたいかにもよるだろうし，研究チームの方針にもよるだろうが，どうしてもこの慣習が研究利用の足枷となってしまう．筆者らの研究チームでは，レセプトデータの傷病名は極力利用せず，その代わりとして医療摘要データに含まれる薬効分類（ATC分類コード）を利用し，実際に処方された薬から傷病名を定義している．

4.5　KDBデータの項目をそのまま利用すべきか

続いて，KDBデータの値をそのまま鵜呑みにすべきでない場合があることを筆者の犯したミスと経験に基づいて言及する．大阪府のメタボリックシンドロームの罹患率を調査する目的で，筆者が被保険者のBMIから肥満と定義される群の人数のカウントを行ったときのことである．BMIは健診データに記録されており，年齢と性別で層別化してそれぞれの群で肥満の割合を算出していたが，ここで筆者のミスに気が付いた．日本肥満学会の定めた基準では，「BMIが25以上」の群を肥満としているところを，「BMIが25より大きい」群を肥満としてしまっていた．しかし，よく知られているように，BMIは身長と体重から算出される連続量であることから，解析にはほとんど影響しないだろうと考えていた．だがその予想は大きく外れ，肥満群に含まれる人数が数千人単位で変化してしまったのだ．当時筆者は驚いてしまったが，聡明な読者の方々であれば簡単に気付く自然なトリックであり，その大きな原因はBMIが小数点以下第一位までしか入力されていなかったためである．つまり，実際にはBMIがピッタリ25でした，という人はほとんどいないだろうが，四捨五入が行われたせいでBMIがちょうど25になってしまった人が思いの外多かったのである．きっかけは筆者の些細なコーディングミスであったが，本稿の因果探索を始め，BMIなど別の値から目的の変数を算出できる場合は，両者を比較して適切な方を利用することを決まりとするようになった．

これは糖尿病診断基準など，上の傷病名の定義にも通ずるものであり，傷病名と検査項目の両方から定義できる疾患や，質問票・既往歴・薬効分類などから網羅的に特定できる疾病についても，どの項目を利用するか，あるいは複数の項目で1つでも当てはまれば疾患と見なすなど，医師や疫学者を交えて慎重な議論が必要となる．

4.6　健診データのクレンジング

ここでは，健診データの前処理方法について記載する．KDBデータに限らず，多くの医学データや健診データは日々の診断や治療，介護の状態を記録したものであり，データ解析に扱いやすいよう整えられているものではない．そのため，これらのデータを解析に用いるために，各種解析ソフト（Python，R，SASなど）で扱いやすい形に成形してやる必要がある．この作業をデータのクレンジング（クリーニング）と呼ぶ．

まず初めに，KDBに含まれる全健診データのうち，条件抽出で2016年度の健診データのみを呼び出す．今回，2016年度を解析対象とした理由は，各年度の健診データの中で最も被保険者数が多い年度であったためである．この時点で重複した個人番号を除いた結果，679,351名の被保険者が解析対象の候補となった．また，本解析では多重共線性の影響を考慮した上で，健診データから以下の11変数を解析に使用した．

- 身体計測（身長：height，BMI）
- 血圧関係（収縮期血圧：sBP）
- 脂質関係（LDL コレステロール：LDL，HDL コレステロール：HDL，中性脂肪：TG）
- 肝機能関係（GOT，γGT，GPT）
- 血糖値関係（空腹時血糖：fBG，HbA1c）

ただし，上でも述べたように，BMIだけは健診データの身長と体重から算出した．次に，これらのデータをLiNGAMモデルに投入できるよう健診データに欠損を含む被保険者を解析対象から削除する．ここで気を付けなければならないのは，特定健診の項目や所属する市町村によって欠損の定義が異なる，ということである．**表2**に使用した変数の基本統計量を示す．

表2 欠損値削除前の基本統計量．欠損数は各変数の欠損値（空欄）の数を示す

変数名	Min	25%	50%	75%	Max	Mean	Std	欠損数
BMI	0	20.5	22.6	24.8	999.9	22.82	3.76	0
GOT	0	19	22	27	9999	24.45	19.68	0
GPT	0	13	17	23	9999	20.48	20.22	0
HDL	0	51	61	73	999.9	123.21	32.36	0
HbA1c	0	5.3	5.5	5.8	999.9	5.64	2.56	0
LDL	0	102	122	142	9999	123.21	32.36	0
TG	0	71	97	136	9999	114.86	84.41	0
fBG	0	84	92	101	999.9	86.49	37.24	0
height	0	151	157.2	164.4	999.9	157.65	9.77	0
sBP	0	118	130	140	999.9	129.70	17.78	0
γGT	0	16	23	36	9999	35.64	84.86	0

表2は679,351名から得られた統計量である．以下の章「Daskによる分散処理」にてDaskと呼ばれるPythonのライブラリを使用した欠損値の削除方法の例を記載するが，dropnaで削除可能な欠損値はブランク（空欄）のみである．表2から見てとれるように，使用したすべての変数はブランクを含んでおらず，上記の方法で削除可能なデータは存在しない（ただし，血清クレアチニン検査値やeGFRなど，検査項目によってはブランクを含むデータが存在することに注意しなければならない）．しかしながら，各変数の最大値と最小値に着目すると，最小値はすべての値で0であり，最大値は変数によって異なるが，999.9や9999といった通常の検査範囲では取り得るはずのない値が入力されていることが見てとれる．これらの値は，健診データの入力者が便宜上何らかの理由でブランクの代わりに代入したものだと想像できる．

それゆえ，我々は欠損値を削除する代わりに，健診時に取り得るはずがなく，かつ人為的に入力されたと考えられる値を異常値と定義し，これらの値の削除を行った．加えて，異常値とは別に，入力者のタイピングミスによって想定し得る検査値から大きく外れた値を持つデータも存在した（たとえば，身長が16.8cmと入力されているデータが存在した．これは168cmのタイプミスだと容易に想像がつくが，こういったデータを意図的に修正すべきではない）．安定した解析結果を推定するために，これらの値を外れ値と見なして削除する必要があった．

5. Pythonを用いたビッグデータ処理

近年ではビッグデータが注目を集めており，KDBデータに限らず，膨大かつ多種多様なデータをどのように管理し，データの処理や分析を行うかが問題となってくる．大規模データのハンドリングに関してはエンジニアに委託する手もあるが，アカデミアの学術研究として研究者ら自身でデータの管理や成形ができるに越したことはない．

一般的な表形式の大規模データの分析手順をまとめると，Hadoopなどの分散処理技術によって分割されたビッグデータの格納・処理を行い，SQLなど収集したデータを整理・操作するためのデータベースの形式に落とし込み，必要なデータを条件抽出することによってようやく各種分析ソフトで扱えるようなデータになる．現在は多くの参考書が出回っており，ビッグデータ解析のためのセミナなども開催されているためこれら技術の習得はそこまで難しくはないように思える．しかし，各ソフトウェアの理解や習得には膨大な時間を要することと，筆者らのチームが汎用性の高いプログラミング言語であるPythonを主に使用してきたことから，これらの処理をすべてPythonのみで実装することにした．

Pythonは人工知能やデータサイエンスが注目を集める中で最も人気を集めているプログラミング言語の1つであり，汎用性の高さと無料のオープンソースである点から研究開発から商業利用まで幅広く利用されている．

ここで，PythonでCSVファイルのような表形式のデータの処理や解析を行うための代表的なライブラリの1つにPandasが挙げられる．Pandasは表形式データのデータ解析を実施するための機能を提供しており，SQLやRのようなデータフレーム処理をPython上で可能とするライブラリである．Pandasはデータ解析を行うためのメソッド（実装済みの関数）が大変充実しており，Pythonでデータ解析を行ったことがある方やPythonの入門書を読んだことがある方は一度は目にしたことがあるのではないだろうか．

このように，Pandasは表形式データにおいて非常に便利なライブラリであるが，一方でビッグデータ解析には向かない面も存在する．Pandasはオンメモリでの解析を前提としているため，Excelなどでも開くことができるCSVファイルであればなんの問題もないが，大阪府健診データのようなビッグデータをPandasで扱おうとすると，コンピュータの処理が非常に重くなるか，動作中にメモリエラーとなり，そもそもデータを扱うことができなくなる可能性がある．このような問題はPandasに限った話ではなく，RやSAS，SPSSといった解析ソフトでも同様のエラーが生じ得る．解決策として，大容量のメモリを搭載したPCやワークステーションの購入を検討すべきであろうが，予算の都合もあるだろうし，今後膨大に増え続ける可能性が高い健診ビッグデータに対してマシンパワーにコストを費やし続けるのは最善策とは言い難い．

そこで我々は，Pandasでは扱いきれないような大規模なCSVファイルに対して，並列処理や分散処理を容易に行うことのできるDaskというPythonライブラリを紹介する．Pythonを使ってKDBデータ処理を検討されている読者の参考になれば幸いである．

5.1 Daskによる分散処理

Daskは並列処理や分散処理を用いて，一般的なコンピュータではメモリ不足となるようなデータサイズの大きなデータセットに対しても処理や解析を実施するためのライブラリである．ビッグデータを扱うためのPythonライブラリはほかにもいくつか候補が挙げられるが，DaskはPandasのコーディング表記に非常に類似しており，またDaskとPandasのデータは相互に変換可能であることからデータ解析に不慣れなPythonユーザでも取っ付きやすいだろう．Python上でDaskを使用するためには，pipコマンドでインストールするか，筆者のようにAnacondaでPython環境を構築した場合は，最新のAnacondaをインストールすればすでにインストールされているはずである．

たとえば，Pandasを用いてCSVファイルを読み込む場合は下記のように記載するだろう．

```
import pandas as pd
df = pd.read_csv( 'data.csv' )
```

一方でdaskを用いてデータを読み込む場合は下記のように書き換えればよい．

```
import dask.dataframe as dd
ddf = dd.read_csv( ' data.csv ' ,blocksize=10e6 )
```

上記のように，ほとんどPandasと同じような表記でCSVデータを読み込むことができることが分かる．ここで，blocksizeは1つのパーティションのデータサイズを指定しており，上の場合は1パーティションあたりの最大データサイズが100メガバイトとなるように自動で分割してくれる．Daskでの処理はこのパーティション単位で行われ，適切なblocksizeを指定することで並列処理を自動で行ってくれる．分割したいパーティション数を直接指定することも可能で，この場合はrepartitionのメソッドでnpartitionsに分割数を指定すればよい．

```
ddf = ddf.repartition(npartitions=8)
```

上の場合は，パーティションが8つになるように自動で分割してくれる．特別，並列処理を行うためのコードを記述せずとも，CPUが許す限りは並列で実行してくれる．

次にDask上でデータ処理を行いたい場合，たとえば欠損のあるデータの削除を行いたいとする．Pandasではdf.dropnaと記載するが，Daskでも同様の記法で実行できる．

```
ddf = ddf.dropna( )
ddf = ddf.compute( )
```

このように，Pandasで実行可能なメソッドの大半はDaskでもほぼ実行可能だと思って差し支えない．また，Pandasの利用に慣れている場合は，メソッドチェインを用いて複数の処理を1行で書くことも可能である．一方でDaskでは，dropnaと記述したコードを実行した時点で即座に欠損値削除が行われるわけではない．Daskを用いて記載した処理は内部で記録され，最後にcomputeと入力することで初めて処理が実行される．なお，最終的にcomputeで渡される変数はPandasのデータフレームとなるため，compute以降はそのままPandasを用いたデータ解析や可視化手法等を行うことができる．

ちなみに，Dask内でどのような並列処理が行われているか確認したい場合は，computeの代わりにvisualizeと入力することで**図4**のように内部の処理過程を可視化することもできる．

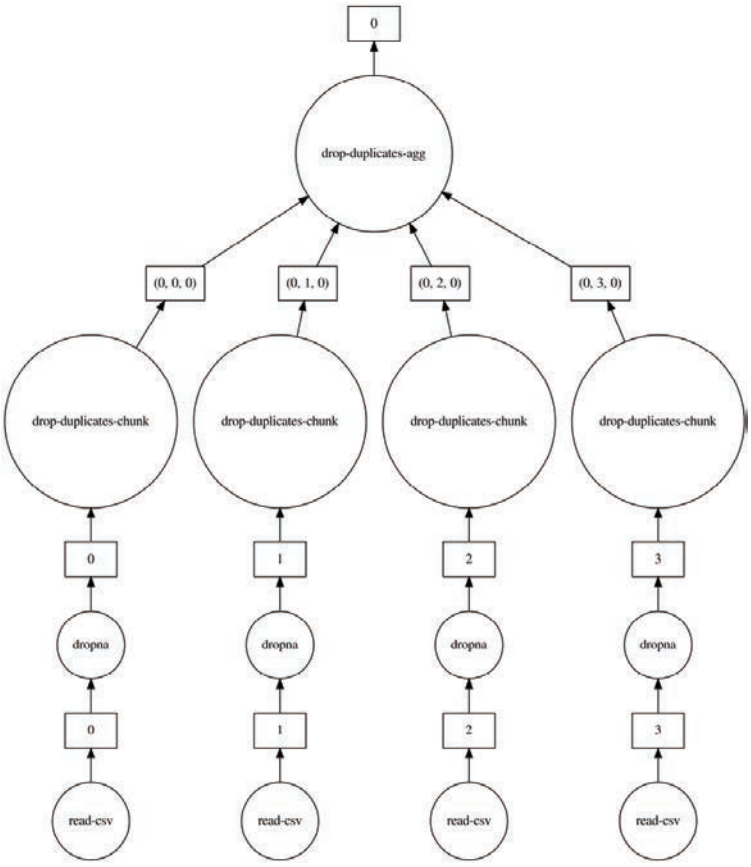


図4　Daskでの処理の可視化例．データを4つのパーティションに分割し，欠損値を削除した後，重複を削除して結合した様子をグラフで示している

ほかにもDaskを利用するメリットとして，複数ファイルを読み込む際にもPandasと比べて簡潔に記載することができる．下記のように，Pandasのread_csvは1つのファイルパスしか指定できないため，複数のファイルを読み込むためにはfor文などで1つずつ呼び出す必要がある．

```
filelist = [ 'data1.csv ',
             'data2.csv ' ,
             'data3.csv ' ]
dflist = [ ]
for   filename in filelist :
      df = pd.read_csv( filename )
      dflist.append( df )
df_all = pd.concat( dflist )
```

一方でDaskではファイルパスのリストやglob string（ワイルドカード）を含んだファイルパスを使用することが可能である．

```
filelist = ['data1.csv ',
            'data2.csv ',
            'data3.csv ']
ddf_all = dd.read_csv ( filelist ,blocksize=100e6 ) . compute ( )
```

Pandasで作成したdf_allと，Daskで作成したddf_allは，reset_indexなどのメソッドでインデックス名を振り直せば同じ出力結果が得られる．年月や市区町村で細かく分割された健診データをfor文を使わずに読み込める利便さを考えると，大きな利点である．詳しい操作方法については Dask の公式リファレンスを参照していただきたい（https://docs.dask.org/en/latest/）．

ただし，DaskはNumpyやPandas，Scikit-learnのようなAPIを完全にサポートしているわけではないため，メモリ不足となりやすいデータの突合やデータクリーニングにはDaskを使用し，中小規模のPCでも扱えるほどのデータサイズまで落とした後はNumpyやScikit-learnを利用するのがよいだろう．また，データの突合の際など，パーティションを分割しすぎると計算時間がかかりすぎてしまい，かえって不便になってしまう場合がある．その際にはパーティションの数を見直したり，ある程度の処理ごとにcomputeして中間ファイルを作成するなどの対策を取るとよい．

5.2　異常値と外れ値の処理

本解析データに欠損値（ブランク）が存在しなかったため，データの前処理として異常値の削除と外れ値の削除を行う．ここで，異常値とは，前述の通り実際の検査範囲では取り得るはずがなく，かつ人為的に入力されたと考えられるデータのことである．これは表2の0や999.9，9999などが該当する．まずはこれらの値を定義しなければならないが，どの変数にどのような異常値が含まれているか把握できるまでは，表2のように変数ごとの最小値と最大値を記録し，異常値が含めれていたら抜く，という作業を繰り返しながら目視で確認する必要がある．とはいえ，普段から健診データを扱う機会の多い医師や保健師であれば数値を見ただけで異常値の定義は明らかであろうが，筆者のように自身の健康診断の結果しか見ることのない解析者にとって

は，ただでさえ扱いの難しい大規模データであるのに加えてなかなか骨の折れる作業であった．最終的には医師などを交えて議論し，現在はデータの更新が行われるまでは記録しておいた異常値のリストを解析のたびに引用することにしている．

今回解析に使用した11の変数において異常値の割合を調べたところ，各変数の異常値の割合はfBGを除けば0.004〜0.51%と非常に小さく，HDLとGOTの異常値の割合が0.004%と最も低かった．一方でfBGだけは13.3%とほかの変数と比べて異常値の割合が随分と高かったが，これは，fBGは空腹時の血糖値を測定するため，食事をしてきた健診受診者の値は入力されないからだと考えられる．

異常値をすべて削除した後は，検査範囲から大きく外れた値（外れ値）を削除する必要があるが，外れ値の削除方法はすでにさまざまな方法が提案されている．筆者らは非正規分布の変数に対して利用されることの多い外れ値の削除方法として，各変数分布の上下数%のデータを外れ値として除くことにした．ここでは上下0.05%を外れ値の閾値と見なした．これらの処理をした後の基本統計量を**表3**に，各変数のペアプロットを**図5**に示す．

表3　欠損値削除後の基本統計量

変数名	Min	25%	50%	75%	Max	Mean	Std	欠損数
BMI	13.8	20.5	22.6	24.8	40.9	22.82	3.35	0
GOT	11	19	22	27	177	24.33	9.32	0
GPT	5	13	17	23	186	20.29	12.18	0
HDL	25	52	62	74	144	63.74	16.65	0
HbA1c	4.5	5.4	5.6	5.9	13.1	5.72	0.63	0
LDL	32	102	121	142	260	122.67	30.51	0
TG	26	70	95	131	1009	110.45	66.37	0
fBG	62	88	94	103	310	98.53	19.28	0
height	129.1	151	157.3	164.5	186.1	157.87	9.17	0
sBP	81	118	130	140	207	129.64	17.49	0
γGT	8	16	22	36	804	33.97	40.70	0

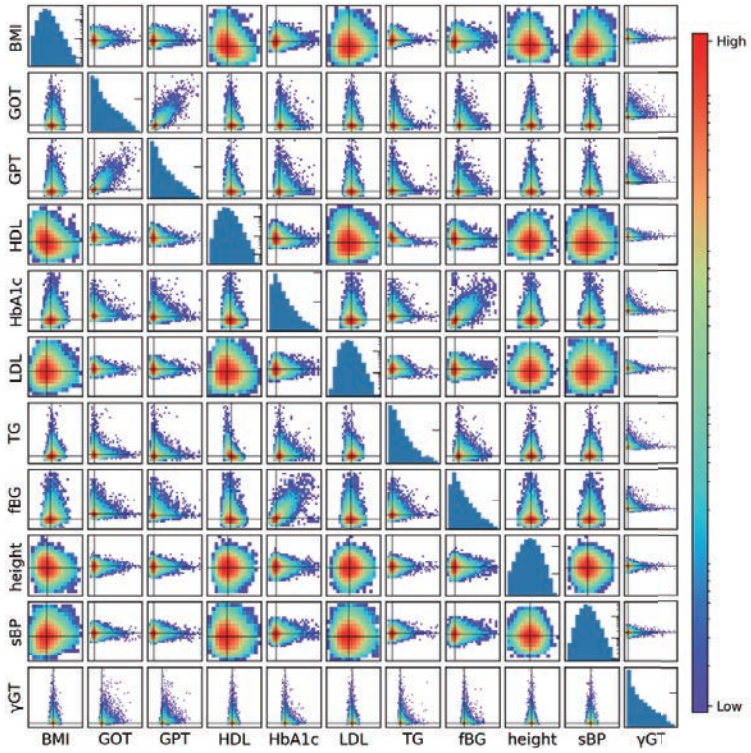


図5　解析に使用した変数のペアプロット．対角には変数の対数ヒストグラムをプロットし，対角以外はある2変数から生成された2次元密度分布をプロットしている

表3と図5は最も人数の多かった70代女性（N=131,036）の基本統計量と変数間のペアプロットである．表3を確認すると，異常値と外れ値を削除することによってもっともらしい統計量を得ることができた．同様に図5のペアプロットからも明らかな異常値や外れ値が含まれていないことが見てとれる．以降では，データ数の最も多い70代女性の解析データに焦点をあてて因果ダイアグラムを推定結果を示す．

6. 大阪健診データを用いた因果ダイアグラムの推定

データクリーニングを終えた検診データを用いて，いよいよ因果ダイアグラムを推定する．前述の通り，DirectLiNGAMによる因果ダイアグラムの推定は，1. 因果順序の推定，2. 係数行列の推定の順に行う．

なお，性別と年代によって健康に対する指標は大きく異なるため，因果ダイアグラムを推定する際には男女で分割し，さらに年齢も10歳刻みで分割してから推定を行った（**表4**）．

表4　性別と年齢層で分けた解析対象者の人数

年齢	男性	女性
30–39	374	429
40–49	22333	23539
50–59	20316	26654
60–69	69892	109529
70–79	97327	131036
80–89	32594	46906
90–99	2147	4881
100–109	17	86

6.1　健診データを用いた因果的順序の推定

初めに，2変数間の独立性の測定を繰り返すことで，因果的順序を一意に推定する．例として，HDLとBMIの関係を調べてみる．**図6**の左上の図はHDLを横軸に取り，BMIを縦軸に取った密度プロットである．この図を見ると，HDLとBMIは負の相関を持っていることが見てとれる．これに対してHDLを説明変数，BMIを目的変数として線形回帰を行い，HDLに対して残差を2次元密度プロットで表現したものが右上の図となる．この図上で，線を引いた位置で断面図を取ったものが左下の図であり，それをそれぞれ規格化して条件付き確率の分布としたものが右下となる．これを見ると，3カ所で確認した例ではあるが，HDLとBMIは，かなり独立性が高いことが見てとれる．つまり，HDLはBMIの原因となっていると推論することができる．

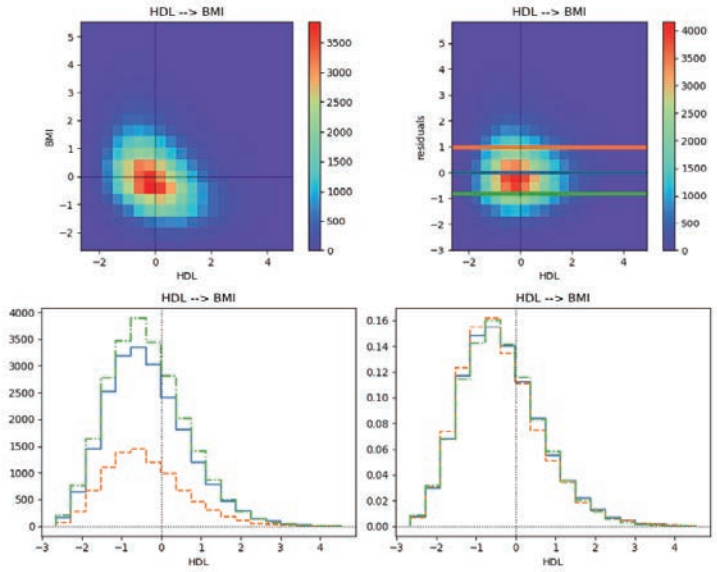


図6　70代女性についてHDLとBMIの2次元密度プロット（左上）を，HDL を説明変数，BMI を目的変数として直線回帰したときのHDLに対する残差の密度分布（右上）．この密度分布をx軸に平行な直線で切った断面（左下）と，それらを規格化して条件付き分布として表現したグラフ（右下）

上記の手順をすべての変数の組合せで行うことで，初めに因果的順序が最も早い変数を決定する．ここで，我々は解析を進める中で，因果的順序の推定に使用するサンプル数が比較的少ない場合，順序の推定にばらつきが生じることを確認した．そこで，復元抽出ブートストラップ法を用いて1,000個のブートサンプルを作成し，1,000回の因果的順序の推定を試みた．**図7**に最も早い因果的順序を決定するために計算した式のMの推定結果を示す．

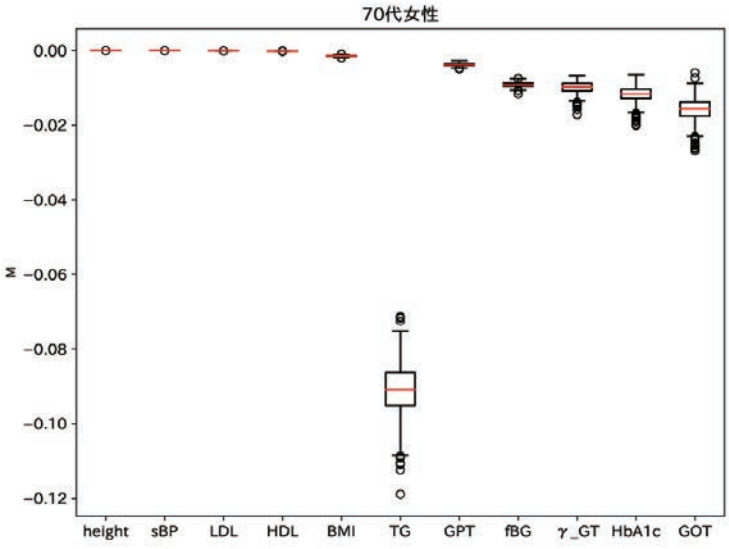


図7　最も早い因果的順序を決定するために推定したMの箱ひげ図

図7は1,000回のMの推定結果を箱ひげ図でプロットしている．この図を見ると，height，sBP，LDL，HDLはほとんど違いがなく，どれも独立性が高いことが見てとれる．次点でBMIのMの推定値が高く，それ以降はMの値が小さくなっている様子が見てとれる．このような独立性の評価を繰り返し，最終的にブートストラップ法を用いて得られた1,000個の因果的順序を**表5**に示す．

表5　ブートストラップ法によって得られた1,000個の因果的順序

出現回数	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th	11th
958	height	sBP	LDL	HDL	BMI	TG	GPT	fBG	γGT	HbA1c	GOT
16	height	sBP	LDL	HDL	BMI	TG	fBG	GPT	γGT	HbA1c	GOT
10	height	sBP	LDL	HDL	BMI	TG	fBG	γGT	HbA1c	GPT	GOT
6	height	sBP	LDL	HDL	BMI	TG	fBG	HbA1c	GPT	γGT	GOT

得られた1,000個の因果的順序はいずれも非常に似たような順序を示していることが見てとれる．因果的順序の早いものから順に，height，sBP，LDL，HDL，BMI，TGはどのパターンでも一致しており，身長や収縮期血圧，コレステロール値といった項目はどの変数からも影響を受けにくいことを表している．一方でHbA1cやfBGの糖尿病の指標となる変数や，GPTやGOT，γGTといった肝機能障害の指標となる変数は因果的順序が遅く，これらの変数はほかの変数から影響を受けやすいことが見てとれる．

なお，サンプル数の多い70代女性の場合，表5が示すように非常に頑健な因果的順序の推定を行うことができたが，サンプル数が少ないほかの年齢性別群で因果的順序の推定を行った場合は望ましい推定結果を得ることができない群もあった．望ましい因果的順序の推定を行うために必

要となるサンプル数を調べるため，10,000から100,000人の間でランダムに解析対象者を抽出し，1,000回の因果的順序の推定の中で，最も多い順序のパターンが何回出現したかを記録した（**図8**）．

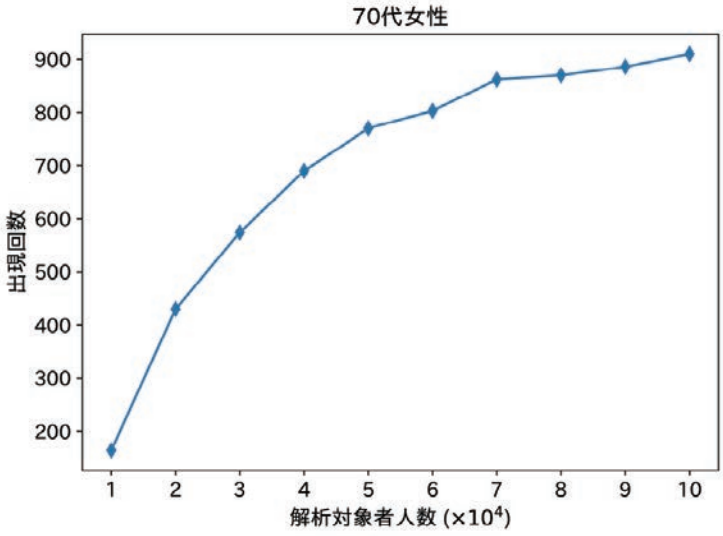


図8　解析人数ごとの最も出現頻度の高かった因果的順序の回数のプロット

70代女性のように，もともと100,000を超える対象者を用いて因果的順序を推定した場合は，表5で示したように最も多いパターンで9割を超えていた．一方でサンプル数を小さくしていくと最頻のパターンはどんどん少なくなっていき，10,000名での解析では，最頻パターンでも2割に満たない結果となった．本解析に使用した変数項目の場合，おおよそ30,000を超えるサンプル数を用意することで，最頻パターンは5割を超え，満足な推定結果を得ることができた．

6.2　係数行列の推定

上で推定された因果的順序に基づいて係数行列Bを推定する．推定された因果的順序から回帰モデルを作成する流れは**図9**のとおりである．因果的順序のうち，自分（目的変数）より前にある変数すべてを説明変数とした回帰モデルを作成し，このモデルから得られた偏回帰係数を係数行列Bの要素とする．これを最も因果的順序が早い変数を除いたすべての変数が目的変数となるまで繰り返し，係数行列Bを作成する．

図9　推定された因果的順序に基づいて回帰モデルを作成する流れ

本稿の解析では，回帰モデルにElastic Netによって正則化を行った回帰モデルを使用した．さらに，推定された偏回帰係数のうち，因果関係の結びつきが弱い因果効果を削除（枝刈り）するため，因果的順序の推定時と同様に，係数行列Bを1,000回推定し，各係数の95%信頼区間を算出した．この信頼区間の中に0が入っていれば，その因果関係の結びつきは弱いと判断し，0を推定値とした．それ以外の場合は1,000回の推定値のうち，中央値をその因果関係の推定値とした．70代女性の健診データから推定された係数行列B（**表6**）と，枝刈り後の係数行列Bから作成した因果ダイアグラム（**図10**）を示す．

表6　因果的順序に基づき得られた係数行列B

変数名	height	sBP	LDL	HDL	BMI	TG	GPT	fBG	γGT	HbA1c	GOT
height	0	0	0	0	0	0	0	0	0	0	0
sBP	-0.030	0	0	0	0	0	0	0	0	0	0
LDL	0.025	0.045	0	0	0	0	0	0	0	0	0
HDL	-0.008	-0.030	-0.019	0	0	0	0	0	0	0	0
BMI	-0.126	0.151	-0.015	-0.291	0	0	0	0	0	0	0
TG	0.020	0.054	0.085	-0.421	0.107	0	0	0	0	0	0
GPT	0.027	0.006	-0.060	0.020	0.175	0.098	0	0	0	0	0
fBG	0.028	0.066	-0.038	-0.031	0.140	0.089	0.109	0	0	0	0
γGT	-0.002	0.008	-0.018	0.061	0.016	0.107	0.381	0.046	0	0	0
HbA1c	-0.014	-0.028	-0.002	-0.044	0.033	0.012	0.041	0.687	-0.021	0	0
GOT	-0.037	0.009	-0.026	0.022	-0.089	-0.036	0.801	-0.031	0.064	-0.043	0

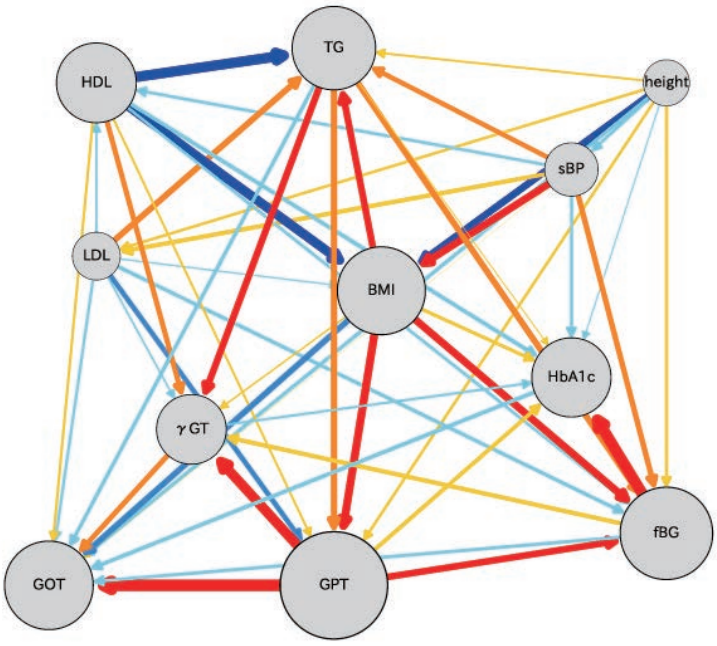


図10　70代女性の健診データから推定された因果ダイアグラム

表6は推定された因果的順序のとおりに並んだ変数を行名と列名にとり，列方向の変数を原因，行方向の変数を結果として回帰モデルを作成した場合の偏回帰係数である．ここでも因果的順序の推定時と同様に，最も出現数の多かった因果的順序をもとにブートストラップ法を用いて係数行列Bを1,000回推定し，その中央値を入力した．

図10がLiNGAMモデルによって最終的に得られた因果ダイアグラムである．図中で，各項目同士が矢印で結ばれているところは，矢印の根本（原因）から先端（結果）に因果関係が推定されたことを示している．原因の変数が1標準偏差変化すると，結果の変数が何標準偏差変化するかを矢印の太さと色で示している（暖色は正の寄与，寒色は負の寄与を表し，矢印の太さは因果効果の大きさを表す）．たとえば，HDLが増加すれば，BMIや中性脂肪，血糖値を改善（低下）させ，一方でBMIが増加することで，血糖値や肝機能の指標に悪影響をおよぼす（増加させる）ことを示している．同様に解析対象人数が30,000人を超えていた60代・70代男性の検診データから推定された因果ダイアグラム（**図11**）と，60代・80代女性の健診データから推定された因果ダイアグラム（**図12**）を示す．多少の違いはあれど，多くの因果関係が一致していることが見てとれるだろう．

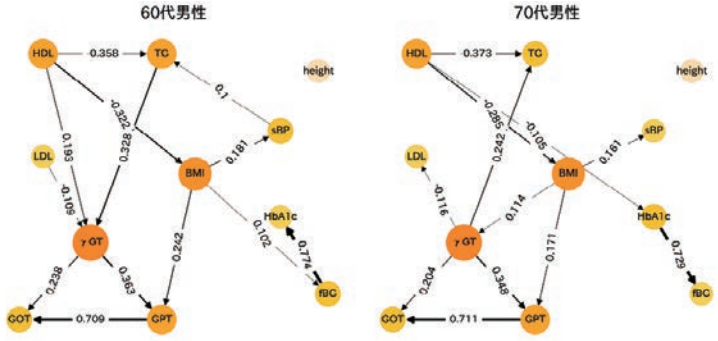


図11　60代男性の健診データから推定された因果ダイアグラム（左）と，70代男性の健診データから推定された因果ダイアグラム（右）

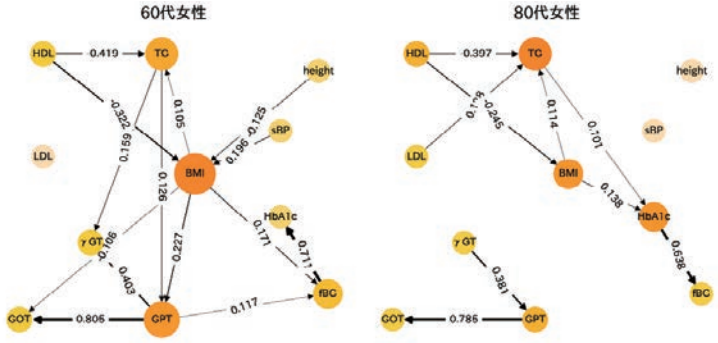


図12　60代女性の健診データから推定された因果ダイアグラム（左）と，80代女性の健診データから推定された因果ダイアグラム（右）

7. これまでの総括と今後の展望

本研究ではDirectLiNGAMと呼ばれる因果関係を自動的に構築するアルゴリズムを用いることで，大阪府が持つ膨大な量の健診データから，生活習慣病因子間に存在する因果関係を明らかにした．本研究成果により，これまで経験に基づいた保健指導において，AIに基づいた解析を行うことによって保健指導による健康指標改善が具体性を持って可視化することが可能となった．これにより，保健指導の実施がエビデンスによって支えられ，指導を受ける側にも説得力を持って受け入れられやすくなるというメリットが期待できる．

LiNGAMモデルを用いることで，医学的な事前知識などに頼らずとも，原因と結果の関係を推定することが可能となった．一方，今回の結果から推定された因果効果と，これまでの医学的見地が一致していない関係や，エビデンスが不十分な因果関係を示すこともあった．たとえば，本解析ではHDLがBMIや中性脂肪に強い因果効果をもたらすことが明らかとなったが，医学の立場からこれらの因果関係を裏付ける十分なエビデンスが得られていないというのが実情である．このような結果が得られた理由としては，これらの変数は因果的順序で隣り合った近い順番で並んでおり，両者の因果的順序を決める指標に大きな差がなかったことや，本解析で取り入れることのできなかった潜在的な交絡因子の影響が観測できていなかったなどの理由が考えられる．たとえば，運動や食事はHDLとBMIの両方の変数に影響をおよぼす可能性があり，これらの変数を取り入れることができなかったため擬似的な因果関係が観測された可能性を否定できない．本研究で用いたLiNGAMモデルは非ガウス分布を仮定した連続分布のみを対象としているため，本解析に運動習慣や食習慣といった項目を含めることができなかった．今後は投薬や治療介入の有無，あるいは各疾患の有無といった離散変数まで取り込むようなモデルの開発を目指す．また，今日では傾向スコアなどを用いることで擬似ランダム化を仮定した因果効果の推定も盛んに行われており，本解析結果との比較検討にも着手したい．

参考文献

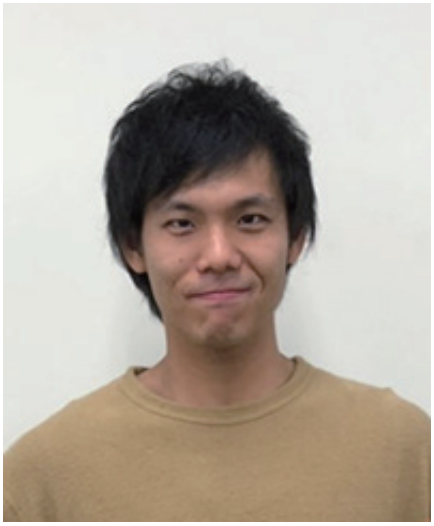
1）Kotoku, J., Oyama, A., Kitazumi, K., Toki, H., Haga, A., Yamamoto, R., Shinzawa, M., Yamakawa, M., Fukui, S., Yamamoto, K. and Moriyama, T. : Causal Relations of Health Indices Inferred Statistically Using the DirectLiNGAM algorithm from Big Data of Osaka Prefecture Health Checkups. *PLoS ONE*, Vol.15, No.12, pp.1-19 (Dec. 2021).

2）Shimizu, S., Hoyer, P. O., Hyvärinen, A. and Kerminen, A. : A Linear Non-gaussian Acyclic Model for Causal Discovery, *Journal of Machine Learning Research*, Vol.7, No.72, pp.2003-2030 (2006).

3）Shimizu, S., Inazumi, T., Sogawa, Y., Hyvärinen, A., Kawahara, Y., Washio, T., Hoyer, P. O. and Bollen, K. : DirectLiNGAM : A Direct Method for Learning a Linear Non-gaussian Structural Equation Model, *Journal of Machine Learning Research*, Vol.12, No.33, pp.1225-1248 (2011).

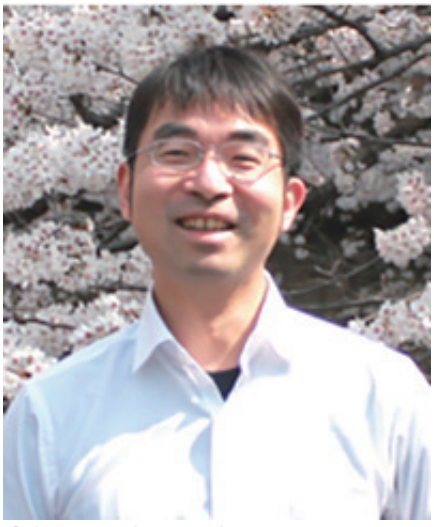
4）Thamvitayakul, K., Shimizu, S., Ueno, T., Washio, T. and Tashiro, T. : Bootstrap Confidence Intervals in DirectLiNGAM, In *Proceedings - 12th IEEE International Conference on Data Mining Workshops, ICDMW 2012*, pp.659-668 (2012).

5）Hyvärinen, A. and Smith, S. M. : Pairwise Likelihood ratios for Estimation of Non-gaussian Structural Equation Models, *Journal of Machine Learning Research*, Vol.14, No.1, pp.111-152 (Jan. 2013).



大山飛鳥（非会員）ooyama@hacc.osaka-u.ac.jp

2021年帝京大学大学院医療技術学研究科診療放射線学専攻博士課程修了．博士（診療放射線学）．現在，大阪大学キャンパスライフ健康支援・相談センター特任助教．機械学習技術を用いた生活習慣病等の疾病予測モデルの開発研究に従事．医学物理学会会員．



古徳純一（非会員）kotoku@med.teikyo-u.ac.jp

2004年東京大学大学院理学系研究科物理学専攻博士課程修了．博士（理学）．現在，帝京大学大学院医療技術学研究科教授．大阪大学キャンパスライフ健康支援・相談センター招へい教授．第110回，第115回日本医学物理学会大会長賞受賞．数理的手法を武器とした医用システム開発の研究に従事．IEEE，AAPM，医学物理学会，日本物理学会，各会員．

土岐　博（非会員）toki@rcnp.osaka-u.ac.jp

1974年大阪大学大学院理学研究科物理学専攻博士課程修了．博士（理学）．2010年大阪大学教授を退職．現在，大阪大学名誉教授・大阪大学キャンパスライフ健康支援・相談センター保健管理部門特任教授．専門分野は原子核理論物理，電磁ノイズの理論研究，健康に関するビッグデータの理論解析．日本物理学会論文賞受賞（1993，2011年）．ドイツフンボルト賞受賞（2001）．日本物理学会会員．

受付日：2022年9月1日
採録日：2022年10月26日
編集担当：澤邉知子（日本大学）

会誌「情報処理」Vol.63 No.2（Feb. 2022）「デジタルプラクティスコーナー」

特集号招待論文

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[1]　北内　啓[2]　高柳慎一[1]　早川敦士[1]　林　樹永[1]　長田怜士[1]

[1]（株）ユーザベース　[2]（株）ニューズピックス

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マーケティングとしてABM（Account-Based Marketing，アカウントベースドマーケティング）の活用が広がっている．ABMをソフトウェアによって実践する方法の1つとして，既存企業のデータを教師データとし，受注確度が高いと考えられる既存企業との類似度の高いターゲット企業を推薦するモデルを構築する方法がある．最終的に人間がどの企業に対してマーケティング施策を行うのかという意思決定を行う都合から，ターゲット企業の推薦モデルにはモデルの解釈性，すなわち，特徴量が企業類似度に与える影響を解釈しやすいことが要件として望まれる．（株）ユーザベースが提供するB2B事業向け顧客戦略プラットフォームFORCASでは企業の所属業界，企業の推定利用サービス等を特徴量としたナイーブベイズを独自拡張したモデルを従来使用していたが，ナイーブベイズにおいて仮定される特徴量の独立性は現実には満たされていない．このため特徴量間の正相関の効果によりアカウントスコアの値が嵩上げされるという問題が生じ，結果として推薦されるターゲット企業に偏りが出てしまっていた．この問題を解決するため，本稿ではL2正則化項付きのロジスティック回帰モデルを用いて企業類似度を算出し，この類似度に基づいた推薦アルゴリズムを提案する．FORCASが保有する1社あたり約1,500個の特徴量を持つ合計約11万社の企業データに対し，業種が異なる4種類の既存企業のデータに対して検証した結果，シンプルな手法ながらもナイーブベイズ拡張モデル，また一般的に高精度と評価されるGBDTと同等以上の精度を達成することを確認した．また，提案手法は業種の違いに対して頑健であり，ABMを実践するためのモデルとして適切であることも確認した．

1. ターゲット企業推薦モデルとは

1.1　ターゲット企業推薦モデルとは

近年，Web上の情報に効率的にアクセスするためのAPIを代表とした技術の普及により，さまざまな企業において顧客データの収集と管理が進められている．また，人工知能によるデータ処理技術の発展とその普及も日進月歩の勢いで進んでおり，結果として，多くの企業において顧客データの収集，およびその活用が進んできている．特に，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，アカウントベースドマーケティング）の活用が広がっている[1]．ABMはアメリカに本社をおくアドバイザリーファームITSMAが2003年に初めて提唱した概念である[2]．2010年頃よりアメリカで注目され始め，ABMに特化したソリューションが次々と誕生している[3]．シンフォニーマーケティング（株）は「全社の顧客情報を統合し，マーケティングと営業の連携によって，定義されたターゲットアカウントからの売上げ最大化を目指す戦略的マーケティング」とABMを定義している[4]．また，（株）ユーザベースFORCAS執行役員CEO田口槙吾は，ABMを「ターゲット企業（アカウント）を定義し，ターゲット企業（群）別に営業・マーケティング情報を集約し，ターゲット企業（群）別に営業・マーケティング組織を再編成し，ターゲット企業（群）からのLTV最大化を目指すマーケティング」と位置づけており，ABMは顧客戦略プラットフォームへとさらに進化していくと報告している[5]．従来のデマンドジェネレーション（マーケティング活動において営業部門への見込み顧客を渡す活動全般）は個人に集中して実施されていたのに対し，ABMではターゲットとなるアカウント（企業）に集中して実施する点が最も大きな違いである[3]．また，ABMを実行するためのABMプラットフォーム覇権争いも激化してきており，米国大手企業の参入も続いている．米オラクル社はABMを「マーケティングに関するカスタマイズされたアプローチであり，見込み客へのターゲティングを通じてブランドの認知や製品購入意欲を高めてもらいながら，情報提供を通じた関係構築を深めるのに役立つマーケティングである．特に，マーケティングの際に使用されるクリエイティブや送付メッセージは，顧客のある特定の問題点に関連したものである」と定義している[6]．

受注確度が高いと推測される企業（以下，ターゲット企業）の特定には，自社の営業データ，特に受注済みの顧客データと豊富な属性データを持つ企業マスタデータが必要となる．属性データとしては，従業員数，業種，上場／非上場など企業の規模や種別を表すファーモグラフィックデータと呼ばれるデータがしばしば利用されるが，企業をより特徴付けているあたかもユーザの行動ログのように見做せるデータ，たとえば企業が利用しているアクセス解析ツールやチャットツールなどのサービス，特定の国・地域に進出しているか否か，企業が求人票において募集している職種，といった企業の現在の行動に基づいた行動解析的なデータも有用である．これらのデータを組み合わせたマーケティング活動を実施することで営業活動の生産性を飛躍的に向上させる企業が増えてきており，ABMの重要性が高まっている．特に，資金力に乏しく初期顧客となり得るターゲット企業に焦点を絞ったアカウントベースの戦略を通じて急速な収益成長を促進する必要のあるスタートアップ企業向けにABMのベストプラクティスに関する実践的な手引が存在するほどである[7]．

ABMをソフトウェアによって実践する方法の1つとして，受注済みの企業（以下，既存企業）のデータを教師データとし，受注確度が高いと考えられる，企業類似度の高いターゲット企業を推薦するモデル（以下，ターゲット企業推薦モデル）を構築する方法がある．最終的に人間がどの企業に対してマーケティング施策を行うのかという意思決定を行う都合から，受注確度の高い企業をターゲット企業推薦モデルはモデルの解釈，特に企業の特徴量が企業類似度に与える影響を解釈しやすいことが要件として望まれる．

我々は，既存企業のデータを分析して既存企業以外の企業（以下，潜在企業）の中からターゲット企業を特定できるシステム（以下，本システム）を2019年に構築した[8]．本システムにおいては，ターゲット企業の推薦問題を受注するかしないかの二値分類問題として捉え，ナイーブベイズを独自に拡張させ，特徴量の重要度を算出しつつさらにスムージングによって補正することで，推薦するターゲット企業の根拠を解釈しやすい特徴量の重要度とともに提示することが可能となった．また，本システムはユーザが特徴量の重要度を変更でき，ほかの特徴量に影響を与えずに潜在企業の企業類似度を再計算することも可能となるよう構築された．なお，企業類似度は分かりやすさの観点から本システムのユーザに対してはアカウントスコアと呼称されており，本稿においても適宜使い分ける．

以下に，典型的なシステム構成を**図1**に示す．システムは大きく2つのステップからなる．（1）ユーザがアップロードした既存企業のデータを名寄せ処理によりシステム内の企業マスタデータと紐づけ，ファーモグラフィックデータおよび行動解析的なデータを特徴量としてターゲット企業推薦モデルを学習する．（2）ターゲット企業推薦モデルをもとに潜在企業の受注確度を企業類似度として推定し，出力された企業類似度の高い順に提示することでターゲット企業の推薦とする．また各企業のデータから作成した特徴量を重要度とともに提示することで，その企業が推薦された根拠をユーザが理解できる．本システムが上記の2つのステップを持つ理由を述べる．（1）については，同名の企業が存在する場合や，ユーザが管理する既存企業のデータにおける企業名と弊社システム内で管理する企業名の表記が異なる場合がある．アップロードされた各既存企業を一意に特定するために，企業名，所在地，電話番号，法人番号などの情報を用いて名寄せする必要がある．名寄せ後は企業マスタデータと既存企業のデータの対応付けができているので，ターゲット企業推薦モデルの学習に必要なデータを作り，実際に機械学習モデルを学習させる．以上のような理由がある．（2）については，ターゲット企業にビジネスの活動を集中し，優先順位をつけるために必要な推薦アルゴリズムであり，また，特徴量の重要度も算出されるため，解釈性に優れているという理由がある．

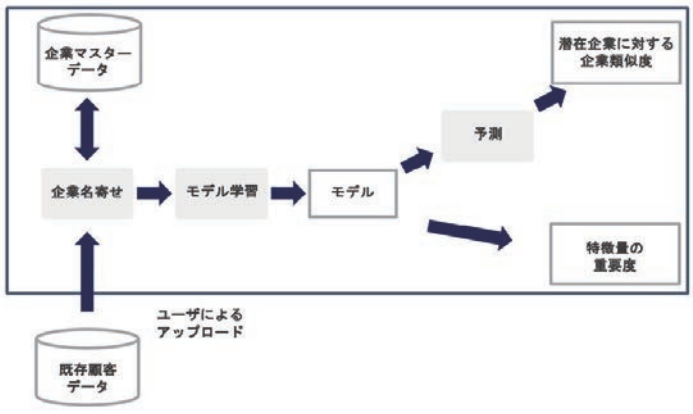


図1　ターゲット企業推薦システムの構成

1.2　ABMに関連する研究

公表されているABMに関する研究，特に機械学習モデルの理論・実践はきわめて少ない．[9], [10]においては，B2B業界におけるABM自体の進化，およびそのメリットについて論じているものの，機械学習を用いたターゲット企業推薦モデルの構築法について論じたものではない．河村ら[11]は，勘と経験に頼っていた不動産営業を置き換えるために，機械学習を用いて申込み顧客リストを作成することによって営業の効率化を支援する推薦モデルを提案した．この研究では，ベテラン営業への聞きとり調査を2度実施している．1度目の調査により作成した特徴量に基づくモデルに対しての推定結果を掲示し，フィードバック（2度目の調査）を得て特徴量を作り，営業の経験を加味したモデルを構築している．内見時のアンケート結果，物件情報の基礎データのほかに熱意と地域ポテンシャルという独自の特徴量を加え，ランダムフォレストを用いて二値分類問題として解いたところ，53.8%のPrecision（適合率）で申込み顧客を推薦することができたと報告している．中山ら[12]は法人向けの営業活動における初期訪問から受注に至るまでの一連のプロセスの各ステップおよび各ステップで選択されるアクティビティに注目し，営業日報などの非構造化データを入力情報として，受注確率の高い営業プロセスの規則性を発見する学習モデルを構築，法人営業活動を営む企業において営業活動の意思決定支援システムを構築した．実際の営業担当者による試用を通して3カ月の営業活動期間を対象にして営業活動意思決定支援システムの適用前後を比較したところ，受注率において10%程度の改善効果を確認している．これら2つの研究は営業効率向上という意味ではABMと類似しているものの，ABMそのものに主眼を当てた研究ではない．また，（株）セールスフォース・ドットコムのSalesforce Einsteinという人工知能製品においてもABM機能を搭載している旨が公表されているが，その詳細なアルゴリズムなどは不明である[13]．

2. 本研究の位置づけ

2.1 ターゲット企業推薦システムの抱える課題

本節では，本システムで使用している既存手法の紹介をするとともにその問題点を明らかにする．ターゲット企業の推薦問題は分類，回帰，協調フィルタリングなどさまざまな手法で定式化できる．本研究では二値分類問題と捉え，潜在企業の企業類似度を算出する．すわなち，企業類似度は0から1の間の値をとり，その値が1に近い値をとる企業ほど既存企業に近しく受注確率が高いと考えられるので優先的にマーケティング施策・営業攻勢をかけるべき企業として推薦される．また，企業の属性データのほとんどがあるサービスを利用しているか否かや，ある特定の条件（シナリオと呼称）に当てはまるか否かなど二値データとなるため，従業員数や売上げ等の量的データも二値データに変換することで，すべての特徴量を二値データとして扱う．すべての特徴量を二値データ化するため，特徴量の解釈方法を統一させることが容易であり，ユーザビリティの向上にもつながる．一般に，Webサイトからの問合せやオンラインセミナーの参加者情報，メールマガジンの配信などマーケティング施策によって生じる見込み客はリードと呼ばれる．マーケティング施策の結果，今後もフォローすべきであると判断されたリードは営業へと引き継がれ，そこから先の営業活動が営業担当者の仕事となる[14]．ABMにおいてはターゲット企業のリストがリードとなる．本システムではターゲット企業が持つ特徴量の重要度をアカウントスコアの根拠として提示する．推薦されたターゲット企業をもとに営業活動を行う際，その根拠が明確で理解しやすいことが重要である．ABMではマーケティング担当者と営業担当者の連携が不可欠であり，マーケティング担当者から営業担当者にターゲット企業のリストを提供したときそのリ

ストがなぜでき上がったのか？　の根拠が明確でないと，営業側が一方的にリストを押し付けられ，営業を命じられたと感じるなど信頼を得にくいからである．一方，明確な根拠があれば，営業側も適切な営業手段を選択でき，売上の増大に貢献する可能性が飛躍的に高まる．そのためには，重要度がアカウントスコアの増減に与える影響を解釈しやすいことが望ましい．また，特徴量同士の依存関係がある場合でも，特徴量単体でどの程度重要かが分かるような重要度が良い．すなわち，推定時における特徴量間の相互作用の影響がなるべく小さくなるようなモデルが良いと考えられる．

上述の要件を考慮し，既存手法として採用した手法はナイーブベイズを独自に拡張した手法（以下，NBEM）である．xを特徴量ベクトル，x_iをxのi番目の要素とすると，ナイーブベイズにおけるクラスyの条件付き事後確率は以下となる：

$$P(y|x) = P(y)\prod_i P\left(y_i\right).$$

ただしここでiは特徴量ベクトルのすべての要素に対して走る．また，既存企業，潜在企業のクラスをそれぞれy_1，y_2とすると，双方の条件付き事後確率の比$R(x)$は次式のようになる．この値をNBEMにおける潜在企業のアカウントスコアと定義する：

$$R(x) = \frac{P(y_1|x)}{P(y_2|x)} = \frac{P(y_1)\prod_i P(x_i|y_1)}{P(y_2)\prod_i P(x_i|y_2)} = \frac{P(y_1)}{P(y_2)}\prod_i \frac{P(x_i|y_1)}{P(x_i|y_2)}.$$

また$S(x_i)$を以下のように定義し，特徴量x_iの重要度とする：

$$S(x_i) = \frac{P(x_i|y_1)}{P(x_i|y_2)}.$$

ここで$S(x_i)$は，既存企業における特徴量x_iの生起確率が，潜在企業におけるそれと比べてどの程度大きいかを表す値である．$P(x_i|y_j);j$= 1,2は次式によって算出する：

$$P\left(x_i\middle|y_j\right) = \frac{N(x_i,y_j)}{N(y_j)} + \alpha.$$

ここで$N(x_i,y_j)$はクラスy_jにおいて特徴量x_iを持つ企業数，N（y_j）はクラスy_jの企業数である．右辺の第1項は多変数ベルヌーイモデルでの最尤推定によって導出される値である．α(0<α<1)はスムージングのための値であり，$N(x_i,y_j)$が小さいほど$S(x_i)$は1に近づく．また，既存企業における特徴量x_iの生起確率が潜在企業におけるそれと等しい，すなわち$N(x_i,y_1)/N(y_1)=N(x_i,y_2)/N(y_2)$のとき$S(x_i)$=1となり，スムージングしないときと同じ値となる．本システムにおいて，この特徴量の重要度は星マークによる9段階表記にてユーザに表示される．

このように，本システムでは企業の所属業界，企業の利用サービス等を特徴量としたナイーブベイズを独自拡張したモデルを使用していた．一方，この計算方式では，アカウントスコアの計算において特徴量間の正相関の効果によりアカウントスコアの値が嵩上げされるという問題があり，推薦されるターゲット企業に偏りが出てしまう場合があった．この問題はあるサービスを利用しているか否かを表現する特徴量のうち，ある特定のジャンルのサービス群に属するサービス間において強く出る傾向にある．たとえばソフトウェア開発会社においては複数のプログラミン

グ言語や開発フレームワークを用いて開発することは至ってあたり前のことであり，その結果として当該利用サービス（あるプログラミング言語やライブラリ，または開発フレームワーク等）における特徴量間の正の相関が強くなることに起因している．たとえばプログラミング言語Pythonを利用している企業は機械学習ライブラリであるPyTorchやscikit-learnも使用している傾向にあることは直感的にも理解できよう．

本問題を解決するため，アカウントスコア算出ロジックにおいて，L2正則化項付きロジスティック回帰モデルを用いた方法を提案する．

2.2 L2正則化項付きのロジスティック回帰モデルを用いた解決方法の提案

2.1節で説明した問題を解決するため，本稿ではL2正則化項付きのロジスティック回帰モデル（以下，L2LR）を用いた企業類似度を新たに定義し，この類似度に基づいた推薦アルゴリズムを提案する．正則化に関しては，スパースモデリングの文脈においてL1正則化法の一種であるLASSO（Least Absolute Shrinkage and Selection Operator）[15]が非常に有名である．モデルのパラメータを最小二乗法で推定する際，サンプルサイズがモデルのパラメータ数に比べて過小である場合や，あるいは特徴量間の相関が高いパラメータが複数存在する場合において，パラメータの最小二乗推定量が存在せず推定自体が不安定化してしまう．このとき，LASSOを用いることで，パラメータのL1ノルムがペナルティとして学習時の損失関数に加算され，パラメータの推定が安定的に行われると同時に変数選択も行うことができるという特徴がある．

一方，ターゲット企業推薦モデルの改善においてはLASSOを用いたモデルは適切ではないと判断しその採用を見送った．その理由は，LASSOを用いた場合，重要ではあるがほかの特徴量との相関が高い特徴量が選ばれずに，人が見たときに直感的に分かりにくい結果となってしまう場合があるからである．たとえば，売上高や従業員数という特徴量は，東証一部に上場しているという特徴量と強い相関を持っているため，モデル推定としてはそのどちらかが削除されてしまっても問題ない場合が多い．一方，人が直接的に各企業の特徴量をフィルタリングの条件として使用する際には売上高X円以上，従業員数Y人以上，東証一部上場企業という形でフィルタリングを行うことは一般的である．したがって，たとえ特徴量間の相関が強かったとしても，モデル推定の結果として当該特徴量に対するパラメータが有限の値を持ち，また，その特徴量の重要度が算出されることが強く望まれる．

そこで本稿においてはL2LRを用いたモデル構築を行うこととした．L2正則化のメリットとしては，ある特徴量のパラメータの値だけを極端に大きくするということをしないために過学習を防ぐことができ，たとえ相関の強い特徴量があっても，ある一方の特徴量が完全に削除されることはなく，それぞれの特徴量に対応したパラメータに対しある程度の大きさの値を割り振ってくれるよう作用する点である．このようなL2正則化の効果により，推薦されるターゲット企業に存在する偏りを解消することが期待される．またL2LRはロジスティック回帰モデル同様，標準化されたデータに対し，パラメータの大きさそのものが予測モデルへの影響度合いを表しており，解釈性が非常に高い点もABMのモデルとして好ましい．L2LRのパラメータ推定は以下のように定式化される：

$$\min_{w,c} \frac{1}{2} w^T w + C \sum_{i=1}^{n} \log(\exp(-y^{(i)}({x^{(i)}}^T w + c)) + 1).$$

ただしここで，$y^{(i)}$,$x^{(i)}$はそれぞれi番目のサンプルのクラスと特徴量ベクトル，nはサンプルサイズ，wは各特徴量に対して推定されるパラメータ，cは切片項，またCは正則化の強度の逆数を表す．アカウントスコアは，L2LRの出力する受注確度の高い企業である確率に対し，独自の確率分布の補正などを行い最終的な値を算出している．

3. 実証実験

本章では，提案手法がどの程度ターゲット企業を正しく推薦できるかを確認するため，分類性能を評価した結果を紹介する．また，その結果としてABMとして有用なモデルであるかについての評価について述べる．

3.1 データセット

FORCAS上にある既存企業データと企業マスタデータを用いる．本実証実験においては，FORCASが保有する1社あたり約1,500個の特徴量を持つ合計約150万社の企業データに対し，業種が異なる4種類の既存企業のデータを，当該業種が対象とするであろう企業群を想定して作成し検証を行う（以下，検証データ）．異なる業種に所属するデータA～Dに対して検証を行うことで，提案手法の業種に対する頑健性を確認し，ABMツールとしての本システムの妥当性を検証する．なお，本稿における業種は筆者らが独自に命名したものであり，既存ユーザの企業の業界に合うよう設定したものである．

企業マスタデータの特徴量には従業員数，業界区分，各企業が利用していると推定される利用サービス（例：広告サービス，チャットサービス），また独自に作成しているシナリオ（例：北米進出企業，増収企業）等が含まれる．利用サービスやシナリオは二値のデータであり，従業員数のような数値データもビン分割により二値データに変換し，すべての特徴量を二値のデータとして扱う．また，ある程度の規模の企業を分析したいというFORCASユーザの要望とモデル推定および予測の計算コストを勘案し，実際に推薦対象となる企業は約11万社となっている．検証データに含まれる既存企業の数を**表1**に示す．

表1　検証データの概要

ID	業種	既存企業数
A	インターネットサービス	約 200
B	電気機器	約 900
C	情報通信	約 2000
D	工業製品	約 100

3.2 実験条件

NBEM，ロジスティック回帰（以下，LR），および勾配ブースティング決定木（以下，GBDT）の3種類を提案手法であるL2LRと比較する．既存企業を正例，それ以外の企業を負例として3分割の交差検証を実施する．評価指標としてはAUC（Area Under the Curve）とPrecision@Nの2種類を採用した．ここで，Precision@Nとは推薦した上位N個の企業における既存企業の割合で定義される評価指標である．本項において，この2種類の評価指標を採用した理由は，

- 業界で標準的に使われているAUCにより大域的な精度を検証
- Precision@Nで推薦結果ランキングの上位のみという条件付きの局所的な精度を検証

という2点についての検証を行いたいためである．Precision@Nを採用した理由は，FORCASのターゲット企業推薦結果画面の最初に現れる推薦結果が正しいものであることをできるだけ高精度にするためである．これは実際に推薦された企業に対しマーケティング・営業活動を行いたいというユーザニーズの特に強い推薦結果ランキングの上位をより局所的かつ重点的に評価したいがためである．既存企業のデータにおける教師ラベルの割合はおおよそ正例：負例＝1：1000と不均衡になっている．また，L2LRにおける正則化項のパラメータはグリッドサーチを用い，最も精度が高くなる値を採用した．

3.2 実験結果

提案手法をNBEM，LR，GBDTと比較した．その結果，全体的な傾向として

- AUCについてはL2LRが最も良い
- Precision@20，100についてもL2LRが良い傾向にある

ことが確認された（**表2**）．これらの精度改善の結果を受け，我々はL2LRをシステムとして運用していくモデルとして採用することを決定した．

表2　各手法のパフォーマンス比較

ID	モデル	AUC	Precision@20	Precision@100	Precision@500
A	NBEM	0.96	0.20	0.11	0.08
	GBDT	0.93	0.40	0.31	0.19
	LR	0.81	0.25	0.28	0.14
	L2LR	0.98	0.55	0.35	0.22
B	NBEM	0.88	0.60	0.35	0.21
	GBDT	0.86	0.60	0.58	0.37
	LR	0.85	0.55	0.47	0.32
	L2LR	0.92	0.55	0.60	0.39
C	NBEM	0.90	0.10	0.24	0.27
	GBDT	0.87	1.00	0.75	0.50
	LR	0.94	0.75	0.69	0.52
	L2LR	0.96	0.70	0.73	0.61
D	NBEM	0.93	0.15	0.08	0.05
	GBDT	0.84	0.25	0.18	0.08
	LR	0.77	0.20	0.12	0.04
	L2LR	0.97	0.30	0.21	0.09

3.2　実証実験から得られた知見と考察，およびAccount-Based Marketingとしての評価

本節では実証実験から得られた知見および考察をまとめる．特に，独自にモデルを開発するL2LRがなぜ最も良い結果となったかについての考察を行う．

提案手法であるL2LRのAUCが全体的にLRと比べて高い．これはL2正則化による過学習の抑制が効果を発揮し，精度を向上させたためであると考えられる．同様にL2LRはGBDTやNBEMに比べてもAUCが高い傾向にある．推薦結果ランキングの上位のみという条件付きの局所的な精度を表すPrecision@Nにおいても同様の傾向が出ており，これにより，実際に推薦された企業に対しマーケティングや営業活動を行いたいというユーザニーズの特に強い推薦結果ランキングの上位に関しても，十分に実務に耐え得る精度であると考える．

一般に高いパフォーマンスを発揮する手法として知られているGBDTが，本実験において高いパフォーマンスを発揮しなかった原因としては特徴量間の相互作用を捉えられるGBDTの良さが発揮できなかったことに起因すると考えられる．たとえば，すべての特徴量はすべて0か1の二値で表現されており，そのため最適な分岐点を探索できず高精度の分類器として学習しづらいのではないかと考えられる．あるいは特徴量の性質から深い木構造になりやすく過学習を起こしやすいのではないかとも考えられる．詳しい原因を知るためにはさらなる研究が必要である．また，検証データCについてはLRとLRL2の評価指標の差が大きく変わらない結果となった．Cは既存企業数が約2,000とほかのデータに比べて多く，既存企業数とパフォーマンスの関係について関連性があると考えられ，この点に関するより詳細な研究が必要である．

次にABMとしての評価について述べる．検証データA〜Dまでの既存企業データは異なる業種に所属するデータであった．ABMを実践するためのモデルとして提案手法であるL2LRは業種の違いに対して頑健であり有効に動作すると考える．したがって，業種が異なる企業に属するユーザであっても，提案手法を使用することに問題はなく高いパフォーマンスを発揮することが期待される．一方，今後，本システムが想定するユーザの業種が拡大するにつれ，本研究において用いた検証データもアップデートしていかなければならない．また，アカウントスコアを算出するためのモデルとして，提案手法であるL2LRはNBEMやGBDT，そしてLRと比べても解釈性が同程度，あるいはそれ以上に高い．NBEMにおいてみられた特徴量間の正相関の効果によりアカウントスコアの値が嵩上げされる問題も表面化することはなく，業種によらずに安定的なパフォーマンスを発揮している．したがって，ABMのモデルとして本提案手法を採用することは妥当であると考える．

4. まとめと今後の展望，および課題

本稿では，既存企業データを分析し，成約確度の高い企業をターゲット企業として推薦するシステムの概要について説明した．既存手法として，ナイーブベイズを拡張した手法（NBEM）を用いて企業のアカウントスコアを算出し，スムージングにより特徴量の重要度を補正する手法を実装していたが，特徴量間の正相関の効果によりアカウントスコアの値が嵩上げされるという問題が生じ，結果として推薦されるターゲット企業に偏りが出てしまう場合があり，この問題を解決するため，L2正則化項付きのロジスティック回帰モデル（L2LR）を用いて企業類似度を算出し，この類似度に基づいた推薦アルゴリズムを提案した．提案手法の有効性を評価した結果，提案手法がNBEMや一般的に高精度と評価されるGBDTと同等以上のAUC，およびPrecision@Nを達成することを確認した．

今後は，解釈性ある，より高い予測精度のモデルを構築していく手法を検討したい．最終的なビジネス施策の実行を担うセールスやマーケティング担当者が納得感をもってビジネスに邁進できるようにするためには，機械学習モデルの出力した結果が機械学習の素人であっても理解できることが肝要である．機械学習，および人工知能における解釈性の研究は日進月歩の状況ではあるが，現状その問題が解決されているとは言いがたい[16]．そのため，我々自身が自分たち自身のニーズに応えられるような解釈性の高いモデルあるいはそのフレームワークを構築する価値があると考える．

また，現状，特段の特徴量選択のアルゴリズムを本システムに搭載していないが，今後の機能として重要な特徴量の自動抽出を検討したい．本システムには約1,500個の特徴量が搭載されている．数多くの特徴量があるおかげで各企業を特徴量空間においてより正確に表現し，その類似度をうまく計算することができている一方，L2正則化項付きモデルで高い精度が出たことからアカウントスコアや重要度に寄与しない不要な特徴量が存在することも事実である．これらの特徴量を自動的に削除するアルゴリズムを組み込み，ユーザに対し見る必要のない情報は出さないというユーザエクスペリエンスを提供していきたいと考えている．

参考文献

1）Bacon, A. : Strategic Account-Based Marketing : How to Tame This Beast, Management for Professionals, in : Uwe G. Seebacher (ed.), B2B Marketing, chapter 17, pp.419-435, Springer (2021).
2）Burgess, B. and Munn, D. : A Practitioner's Guide to Account-Based Marketing : Accelerating Growth in Strategic Accounts, Kogan Page (2017).
3）庭山一郎：究極のBtoBマーケティング ABM（アカウントベースドマーケティング），日経BP社 (2016).
4）（株）シンフォニーマーケティング：アカウント・ベースド・マーケティング（ABM）とは，https://www.symphony-marketing.co.jp/abm/about/ (2021年8月29日現在)
5）田口槙吾：データとテクノロジーの力で顧客を再定義する―B2Bマーケティングの成果を最大化する『ABM』の基本概念とその実践事例―，MarkeZine Day 2021 Spring (2021).
6）Oracle Corporation : Account-Based Marketing Handbook, https://go.oracle.com/LP=79765 (2021年8月29日現在)
7）Day, D. and Shi, S. : Automated and Scalable : Account-Based B2B Marketing for Startup Companies, Journal of Business Theory and Practice 8, p16, 10.22158/jbtp.v8n2p16 (2020).
8）早川敦士，北内　啓：Account-Based Marketingのためのターゲット企業推薦システムの構築，人工知能学会 (2019).
9）Kumar, G. P., and Rajasekhar, K. : Account-based Marketing in B2B industry, Journal of Interdisciplinary Cycle Research, Volume XII, Issue II, February/2020, ISSN NO : 0022-1945 (2020).
10）Paavola, A. : Designing an Account-based Marketing Program, Master's Thesis, Lappeenranta University of Technology (2017).
11）河村一輝，諏訪博彦，小川祐樹，荒川　豊，安本慶一，太田敏澄：飲食店向け不動産営業を支援する申込み顧客推薦モデルの提案，人工知能学会 (2017).
12）中山義人，森　雅広，斎藤　忍，成末義哲，森川博之：営業活動における意思決定システムの適用と評価，In IEICE Conferences Archives. The Institute of Electronics, Information and Communication Engineers (2019).
13）Saini, N. K. and Sharma, H. : Salesforce Einstein : Artificial Intelligence for Customer Success Platform, International Journal of Scientific Research & Engineering Trends, Vol.6, Issue 3, May-June-2020, ISSN (Online) : 2395-566X (2020).
14）福田康隆：THE MODEL，翔泳社 (2019).
15）Tibshirani, R. : Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society : Series B (Methodological) 58.1, pp.267-288 (1996).
16）Molnar, C. : Interpretable Machine Learning A Guide for Making Black Box Models Explainable, Lulu.com (2020).

新井和弥（非会員）kazuya.arai@uzabase.com
2020年名古屋工業大学大学院工学研究科博士前期課程修了．修士（工学）．2020年（株）FORCAS入社．2021年統合により（株）ユーザベースへ転籍．B2B事業向け顧客戦略プラットフォームFORCASの開発に従事．

北内　啓（非会員）akira.kitauchi@uzabase.com
1998年奈良先端科学技術大学院大学情報科学研究科博士前期課程修了．同年（株）NTTデータ入社．2014年（株）ユーザベース入社．以来，自社サービスのアルゴリズム開発に従事．2017年（株）FORCASに転籍，2021年（株）ニューズピックスに転籍．推薦システム開発に従事．

高柳慎一（正会員）shinichi.takayanagi@uzabase.com
2020年総合研究大学大学院複合科学研究科統計科学専攻博士課程修了．博士（統計科学）．2020年（株）FORCAS入社．2021年統合により（株）ユーザベースへ転籍．B2B事業向け顧客戦略プラットフォームFORCASの開発に従事．徳島大学客員准教授，情報処理学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．

早川敦士（非会員）atsushi.hayakawa@uzabase.com
2015年電気通信大学大学院情報理工学研究科博士前期課程修了．修士(工学)．2018年（株）FORCAS入社．2021年統合により（株）ユーザベースへ転籍．B2B事業向け顧客戦略プラットフォームFORCASの開発に従事．徳島大学客員准教授を兼任．

林　樹永（非会員）tatsunaga.hayashi@uzabase.com
2017年慶應義塾大学大学院理工学研究科博士前期課程修了．修士（理学）．2019年（株）FORCAS入社．2021年統合により（株）ユーザベースへ転籍．B2B事業向け顧客戦略プラットフォームFORCASの開発に従事．

長田怜士（非会員）ryoji.nagata@uzabase.com
2015年大阪電気通信大学情報通信工学部情報工学科卒業．2020年（株）FORCAS入社．2021年統合により（株）ユーザベースへ転籍．B2Bに特化した顧客戦略プラットフォームFORCASの開発に従事．

受付日：2021年8月30日
採録日：2021年10月15日
編集担当：戸田貴久（電気通信大学）

会誌「情報処理」Vol.63 No.2（Feb. 2022）「デジタルプラクティスコーナー」

特集号招待論文

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[1]　辻　智[2]　田村光太郎[3]

[1]成城大学　[2]成城大学データサイエンス教育研究センター　[3]（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，データ分析に詳しい人材の育成は教育目標の大きな柱の1つである．本学では 2015 年度より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を振り返り，その教育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の今後の展開を述べたい．

1. 人文・社会科学系大学におけるデータサイエンス教育の目的

成城大学（以下本学）は，経済学部，文芸学部，法学部，社会イノベーション学部の人文・社会学系4学部からなるいわゆる文科系大学である．本学では，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た．

なぜ，文科系大学にデータサイエンス科目を設置するのか．その経緯については次章で述べるが，設置当時，2014年には「日経ビッグデータ」☆1が創刊され，蓄積した膨大なデータを分析，活用して商品開発やマーケティング，業務の効率化や収益の向上などを達成したという多くの先進的な企業の姿が日経新聞などでも頻繁に報道されるようになった頃である．早晩，製造，販売，金融，物流，サービスなどあらゆる業種のあらゆる部署の企業活動に，ビッグデータ活用が波及することは想像できた．しかしながら，そのスキルを持つ人材は理系学生や社内教育では賄いきれず，常に不足していた．今でもその状況は大きく変わっていないように思われる．

本学2020年度卒業生の就職先を業種別に**表1**に示した．理科系学部に比べると製造業への就職比率が小さいが，どの学部も，さまざまな業種に極端な偏りなく就職している．

表1　学部ごとの就職先

業種	経済	文芸	法	社会イノベ	大学全体
サービス業など	24%	41%	22%	31%	30%
卸売・小売業	19%	22%	18%	17%	19%
金融業	18%	8%	20%	11%	14%
建設・不動産など	31%	21%	24%	29%	27%
公務員など	4%	4%	11%	4%	5%
製造業	5%	4%	4%	8%	5%

ビッグデータを活用するための科学をデータサイエンスと位置付けるならば，文科系大学とはいえ，卒業後多様な現場で働く学生にとっては必要な素養となる．本学では，カリキュラムの目標として「データに関心を持ち，データに基づき考え，行動する学生の育成」を掲げ，カリキュラム作成にあたっては，統計学的なデータ分析手法の習得に終始するのではなく，実用例を通して，データに目を止めて考える姿勢を身につけることを重視した．

このような，人文・社会科学系大学におけるデータサイエンス教育に対する考え方は，数理・データサイエンス教育強化拠点コンソーシアム☆2が推奨する数理・データサイエンス・AI（リテラシーレベル）モデルカリキュラム[1]とも合致するものであり，本学データサイエンス・カリキュラムの基礎的部分は，文部科学省より令和3年度「数理・データサイエンス・AI教育プログラム（リテラシーレベル）」として認定された[2]．

学生が社会に出てから必要になる素養というだけではなく，人文・社会科学系大学の学生がデータサイエンス教育を学ぶべき理由はまだある．

国の第5期，第6期の科学技術・イノベーション基本計画で目指すべき方向としているSociety 5.0は，ITとディープラーニングに代表されるような高度な機械学習の技術をフルに使って，スマートな情報活用社会の実現と環境問題，経済格差などの難しい課題の分野横断的な解決を目指すというものである．Society 5.0が目指す新しい社会は，理工系，情報系人材の独壇場ではない．これらの難しい課題を解決するためには，どのように社会や経済が回っているのか，環境問題はなぜ，どのようにして起こるのかといった社会・経済の問題に関する深い洞察とデータサイエンスとのリンクが必須であると考えられる．

科学リテラシーを「科学の常識や知見にしたがって意思決定し，行動できる能力」，データリテラシーを「データ（エビデンス）に基づき論理的に意思決定し，行動できる能力」と定義したならば，これらの重要性の認識は社会が大きな困難に直面するたびに高まっているように感じる☆3．これらは市民として社会全体が持たなければならないリテラシーであり，それは教育でしか成し遂げられないものである．

2. 成城大学におけるデータサイエンス教育導入の経緯

2.1 経済学部専門科目としてのデータサイエンス教育

まず，全学データサイエンス教育導入以前より行っていた，経済学部における専門科目としてのデータサイエンス教育について紹介する．基礎科目として「データ解析入門」（経済学科）あるいは「データ分析」（経営学科）習得後に，専門科目として「統計学」（経済学科）「経営統計学」（経営学科）を選択することができる．さらに，2021年度よりプログラミングや機械学習の基礎を学べる科目も導入した．また，「経営情報論」（経営学科専門科目）では，2012年度から1つのトピックとしてビッグデータの活用を取り上げ，商品企画・開発，マーケティング，収益（集客，客単価）の向上，災害時の危機管理，医療，集合知を利用した予測など幅広い分野での事例紹介を行ってきた．学生にとっては身近なテーマが多く，大変興味を持って聞いてくれている．

2.2　全学共通教育科目としてのデータサイエンス教育の導入

学校法人成城学園は，2017年に学園創立100周年を迎えたことを機に，次の100年の教育目標として「国際教育」「理数系教育」「情操・教養教育」の3つを掲げて教育改革を進めている．教育改革“3つの柱”の主な取り組みは以下のとおりである．

- 国際教育：語学的教養を通じて国際性を強化する「国際教育」の取り組み
- 理数系教育：数学的教養を通じて論理的な思考力を強化する「理数系教育」の取り組み
- 情操・教養教育：芸術的教養を通じて人間性を強化する「情操・教養教育」の取り組み

理数系教育の中核として，データサイエンス教育を行うという発想は，2013年，当時の油井学長が学外会議で同席した日本IBMの方々との情報交換の中で生まれた．2014年，成城大学は日本IBM東京基礎研究所と包括的な連携協定を締結した．その内容は「成城大学の持つ経済学，文化芸術，法学など人文・社会科学的視点と，日本IBM東京基礎研究所のデータベース，自然言語処理，機械学習，人工知能などの高度なICTとの融合を図り，より豊かな未来社会の実現と学術研究の振興に寄与すると同時に，ビックデータを活用できる人材の育成をともに目指す」というものである．

2015年度からは，日本IBM東京基礎研究所より授業科目「データサイエンス概論」の提供を受け，それを本学におけるデータサイエンスへの入門科目として，全学共通教育科目の中に，データサイエンス科目群6科目を設置した．

2.3　データサイエンス教育研究センター（CDS3）の開設

本学は，「情操・教養教育」を担うセンタとして共通教育研究センターが2007年に開設されており，「国際教育」を担うセンタとして2015年に国際交流室を国際センターに改組した．理数系教育の中核としてのデータサイエンス教育の企画・運営は，共通教育研究センターに設置された専門部会が行っていたが，その任務を担う専門部署として2019年4月にデータサイエンス教育研究センター（以下CDS3：Education and Research Center for Data-driven Social Science & Humanities of Seijo University）を開設した（**図1**）．本稿の筆者の1人増川は初代センター長を務めた☆4．

図1　成城大学データサイエンス教育研究センター（9号館2F）
（左：9号館全景，右：共創の場としてのスクエア）

CDS3は，中核ミッションとして，教育「データに関心を持ち，データに基づき考え，行動する学生の育成」と研究「データサイエンスの人文・社会科学分野への応用」の2つを掲げて，人文・社会科学の専門知識とデータサイエンスの基礎力を兼ね備えた独創的な教養人の育成を目標としている．

次章で述べる正規の教育プログラムに加え，統計検定やG検定等の資格取得の支援，企業が主催するデータ関連コンテストやSignateやKaggleコンペの参加支援も行っていきたいと考えており，2020年度からディープラーニング協会が主催するG検定のための講習会を開催している．

また，CDS3の活動を学内外に発信するためのシンポジウムを定期的に開催している．2019年度は「人文・社会科学系大学におけるデータサイエンス教育」と題するシンポジウムを開催した[3]．CDS3は，大学や企業で実際にデータサイエンティストとして活躍されている研究者や実務家の方々に，外部アドバイザリー委員として企画運営にご協力いただいている．シンポジウムでは，これらの方々と人文・社会科学系大学におけるデータサイエンス教育の在り方や今後の方向性について議論した．2021年度は「人文・社会科学研究におけるデータサイエンス」をテーマとしたシンポジウムを開催する．

そのほか，与えられたデータを使ってビジネス提案を競う学内「データサイエンス・コンテスト」[4]や，「データサイエンス・ワークショップ」（2021年度は，計量文献学へのデータサイエンスの応用を学ぶ「データサイエンス・ワークショップ2021―文学作品のテキストマイニングーデータサイエンスが解き明かす作品の痕跡ー」[5]）開催などにより，データサイエンスの実践を通してより多くの学生にその面白さを知ってもらうための企画を行っている．

3. 成城大学におけるデータサイエンス教育

ここからは具体的に成城大学におけるデータサイエンス教育の内容について述べる．

3.1 カリキュラムデザイン

本学のデータサイエンス教育においては「文系でもできる！」をスローガンとして，学習する内容は，文理融合的で実践的・実務的なものとなっている．この科目群を系統的に学ぶことで，専門以外の分野にも視野を広げ，卒業後どのような分野に進んでも活かせるデータ分析力を身につけることを目標としている．

表2に現行のカリキュラムと科目ごとの到達目標を示した☆5．現行のカリキュラムは全部で6科目（12単位）でリテラシー・基礎4科目と応用2科目の2群からなる．データサイエンスの基礎を身につけた学生には「基礎力ディプロマ」が，さらにその上の応用力を身につけた学生には「EMSディプロマ」☆6が授与される制度となっている．2020年度終了時点で26人が「基礎力ディプロマ」を5人が「EMSディプロマ」を授与されている．

表2　データサイエンス・カリキュラム（現行）

カテゴリ	科目名	科目目標
DS の基礎を学ぶ （基礎力ディプロマ ）	DS 概論 DS 入門 I, II DS スキルアップ	実践例を知る 基礎的手法を身につける 学んだ手法を実践する
DS を応用する （EMS ディプロマ ）	DS 応用 DS アドバンスト	課題の解決の方法を学ぶ 課題を見つけ，解決する

3.2　シラバス

3.2.1　リテラシー・基礎レベル教育〜基礎力ディプロマ〜

リテラシー・基礎レベル教育の科目は，AIやデジタル・トランスフォーメーションを概観する「データサイエンス概論」，記述統計学を学ぶ「データサイエンス入門I」，推測統計学を中心とした「データサイエンス入門II」，実践的スキルをさらに伸ばすための「データサイエンス・スキルアップ・プログラム」の4科目である．希望する学生全員が履修可能となるように，どの科目も「前提がないのが前提」との方針なので，履修生のレベルを揃えるような事前のテスト等は実施していない．各学部・学科の授業を優先させる形で，曜日・時限の調整を可能な限り行っており，5時限も積極的に活用している．さらに，通年授業は行わず，前期・後期で同じ授業を展開し，なるべく多くの学生が履修しやすいように工夫している．

特に本プログラムの導入科目でもある「データサイエンス概論」は，文系学生の理数系科目に対する苦手意識に留意し，授業とハンズオンを毎回組み合せている．このことにより，「学ぶ楽しさ」や「学ぶことの意義」が体感的に進むように工夫している．授業の部分では，パワーポイントによる資料投影を中心とした授業形式で行う．その際，ビデオ資料投影も多く盛り込み，映像と音声により臨場感を高め，体感的に理解が進むようにしている．ハンズオンでは，実際に卓上からWebやクラウドにアクセスして，AI系のアプリやコンテンツにより実習を行う．その際，海外のデータセンタとのやりとりでも，数秒で結果が戻ってくる．この圧倒的なスピード感で現実のITやネットワークの最新技術を体感する．また，90分という限られた授業時間の中でも，このスピードが功を奏し，実習は時間的にもかなり思いどおりに進めることができる．

「データサイエンス概論」は，6科目のデータサイエンス科目群の中で最も入り口に位置するため，この科目で失敗すると，後の科目履修に続かない恐れがあり，どのような内容にするのかがとても重要である．学生がワクワクする内容が必要であり，AIやデジタル・トランスフォーメーションを概観できるようなトピックの組合せを考える必要がある．また，“コンピュータ・サイエンス”のような大きくて硬いイメージの題目ではなく，“社会やビジネスを大きく変える”や“医療技術支援”のような身近に感じる題目を多く取り入れた．シラバス作成にあたっては，ICT業界側の視点が多く盛り込まれているのも特徴で，さらに履修生の要望を取り入れて，年々進化させている☆7．

3.2.2　応用レベル教育〜EMSディプロマ〜

データサイエンスの応用レベル科目は，データサイエンスの概観を知り，入門を終えた学生を対象とし，統計を中心に教える「データサイエンス応用」と機械学習・AIを教える「データサイエンスアドバンスドプログラム」がある．授業のねらいは，実際の現場で利用されることの多いデータサイエンスの各手法に関して広く触れてもらうことと，興味や目的に応じて自身で深めていくことができるようになってもらうことである☆7．

前半は講義形式，後半は分析プロジェクトで構成される．統計や機械学習の手法の紹介だけでなく，実際のデータサイエンスの現場を視野に入れて，与えられたデータに価値をもたらし，データ分析により課題を解決するという点も重視している．

講義形式の授業では，スライドを使った説明と，その内容をPythonやR言語で実装していくプログラミングの2つを行う．難しい数学を前提にしないよう，また，プログラミングの経験がない学生が多いため，授業の中で学生に大きな負担がかからず，興味を失わないことを念頭に進めている．プログラミングでは，授業中に十分な時間を取り，教員のサポートの基，授業内容の実装とともに簡単な課題を与え，結果を提出させることで評価を行っている．

後半ではプロジェクトベースドラーニング（以下，PBLと表記する）が行われる．学生複数人からなるグループを作り，各グループが分析のテーマを定め，分析を実施し，結果を発表することが求められる．ここでは「企画」「分析」「表現」の一連のデータ解析プロジェクトのプロセスを体験する．実務に近い状況でのデータサイエンスの適用を学ぶとともに，自身に課せられた分析に対して，分析の理解や説明能力が高まることが期待される．授業の最終回では発表会を開催し，各グループの成果を外部の実務家講師などにお願いする審査員に講評していただいている．

本授業でPBLが導入されている背景として，実務では，分析手法の深い理解だけでなく，プロジェクトをマネジメントするスキルも，データサイエンティストの中では重要なスキルとして認識されていることによる[7]☆8．データサイエンティスト協会が公表するデータサイエンティストの3スキルの中では，ビジネススキルと呼ばれる．これは，ビジネス課題に関連するデータとその分析がもたらす価値を評価したり，複雑な手法を利用者に説明したりするコミュニケーション能力である．複数のメンバで構成される実務のデータサイエンスプロジェクトでは，コミュニケーション能力は重要なスキルであり，データサイエンティストに欠かせない基礎体力と考えている．

2015年度のカリキュラム設置の初期から，このようなPBLを導入できたことは，若手教員や実務家からアドバイスを受ける形でシラバスの設計を行ってきたことが大きい．現在も構成を大きく変えずにシラバスが組まれていて，特に，人文・社会科学系の学生向けのプログラムとして適していると考えている．

3.3　履修生の理解度

3.3.1　リテラシー・基礎レベル

履修学生の構成は，学期やクラスごとにコントロールしていないので，授業初回にアンケートを実施し，これまでの経験やスキル，今後の授業への希望や不安を聞いて，クラス全体としてのレベルや雰囲気を分析し，その後の授業の進め方を決めている．また，毎回の授業の後，感想や要望をリアクションとして書いてもらい，それを基に授業の細かな軌道修正も適宜行っている．

「データサイエンス概論」の授業は全15回の構成であるが，ほとんどが初年度の学生なので，当然のことながらデータサイエンスに関して初学者が多い．開始時の第1回と終了時の第15回には同じ質問内容の認知度アンケート調査を行い，理解度のチェックを行っている．「次のデータサイエンスに関する用語の認知として，あなたに当てはまるレベルをクリックしてください」という問いかけで，あらかじめ設定した85のデータサイエンスに関する用語に関して認知度を聞いている．学生にとってよく耳にする用語から，聞いたことがないような専門用語までを意図的に選定している．これらの用語に対して，4段階の順序尺度で認知度を聞いている．また，アンケートの最後部には，第1回には「この授業への期待，要望，質問，不安など」，第15回では「この授業の良かった点，残念だった点」などを自由記述形式で書き込みできるように設定している．第2回から第14回までの途中の回でも，毎回授業の終わりに理解度や要望を50〜300字程度の自由記述形式で書いてもらっている．

データサイエンス概論の2019年度前期（教室での対面授業）の第1回授業前と第15回授業後を比較した結果を例示する．**図2**は，第1回授業前（以下，Beforeで表す）と第15回授業後（以下，Afterで表す）の学生の認知度を集計し，出力したものである．分析サンプルとしてのコメント数は，Beforeが42，Afterが37となっている．図2には，興味深い点がいくつかある．10の用語は，人工知能（AI）に関係した用語である．AIに関しては，高校において学習してきたためか，Beforeの段階でほとんど認知されており，機械学習についてもかなり認知されている結果となっている．Beforeでは，ほかの用語に関しても，認知している割合が多くなっている．Afterになると，AIおよび機械学習もよく理解していて，他者に説明できる割合が増え，授業でツールとして使用してきたIBM Watson，Microsoft Azureなどへの認知度も大幅に増えている．また，学生の積極的自己学習のおかげで，チャットボットなど授業中にトピックとして扱った内容以外でも，サポートベクタマシンなど，授業では説明していないにもかかわらず，Afterで認知度が大幅に上がっている．また，AIに関する認知度のスコアはBeforeとAfterであまり変わらないが，その質が大きく変化することが特徴として挙げられる．その理由は，途中回のコメントを追跡することで分かる．AIに関しては，授業が進むたびにその理解が深まり，サポートベクタマシンについては，自分で授業内容について調べているうちに目に触れるようである．

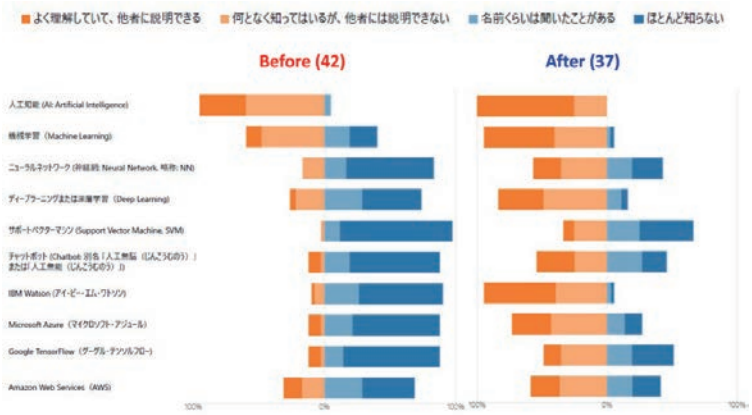


図2　データサイエンス理解度チェック（2019年度）

教室での対面授業とオンデマンド授業の効果を比較するために，授業初回と最終回に実施している認知度アンケートの結果を，2021年度前期（オンデマンド授業）のデータサイエンス概論に関して，**図3**に示す．今回示す2019と2021のデータは，学年や学部が同じような構成のクラス同士をクラス単位で比較している．図2と図3を比べてみると，人工知能関連用語10語に関する結果であるが，2019前期教室対面授業と2021前期オンデマンド授業では，Before & Afterで各語の認知度の変化傾向は似ている．この結果からは，教室対面授業からオンデマンド授業になっても，履修生の授業に対する理解度は大きく変化せず，理解度が著しく落ち込むということはなかったようである．

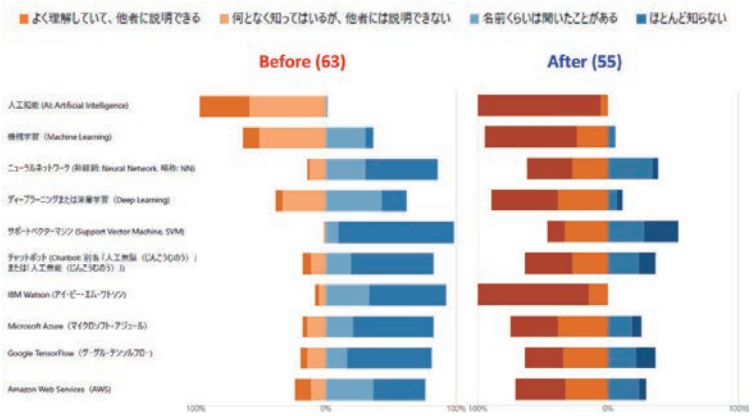


図3　データサイエンス理解度チェック（2021年度）

3.3.2　応用レベル

履修生は数学をあまり得意としない文系の学生であるため，数理的な知識を前提にしない授業形態や，プログラミングの直接指導を行っていることを先に書いた．この授業設計は，学生同士あるいは教員との議論をやりやすくするので，細部で行き詰まってしまう学生を減らすことがで

きる．

応用レベルの授業での細部についての学生の理解は，さまざまである．しかし，ライブラリ（汎用性の高いプログラムをまとめたもの）の利用やオープンソースソフトウェアによって，分析は問題なく実施できる．手法の細部まで知ることを後回しにしても，ライブラリの使用方法さえ理解できれば，わずか数行のスクリプトで分析が実施でき，結果について素早く学生と議論できるからである．これらの取り組みは，数理的な理論についていけなくなって学生が履修をあきらめてしまうことや，プログラミングに対する抵抗感をなくすことができ非常に有意義である．

最終的に，いずれの学生もデータ解析プロジェクトではメンバの一員として，授業の中で興味を持った手法やスキルを目的に応じて自身で深め，プロジェクトの中で他者に説明できるまでのレベルに達している．発表会やレポートの内容は，発表会審査員や外部講師から高い評価を得ていることからも，授業を通して学生のデータサイエンスへの理解は高まっているといえる．

4. 成城大学におけるデータサイエンス教育導入の成果

データサイエンス科目群は，開設以来順調に履修者数を伸ばしている．**図4**に2015年度から2021年度の履修者数の推移を示している．また，**図5**は2021年度を例としたデータサイエンス概論の学部別，学年別の履修者の割合である．経済学部が半数，それに次いで多いのが文芸学部，それに社会イノベーション学部，法学部と続いている．これは毎年ほぼ同じ傾向である．文芸学部にデータサイエンスに関心を持つ学生が多いことは本学の特色であろう．

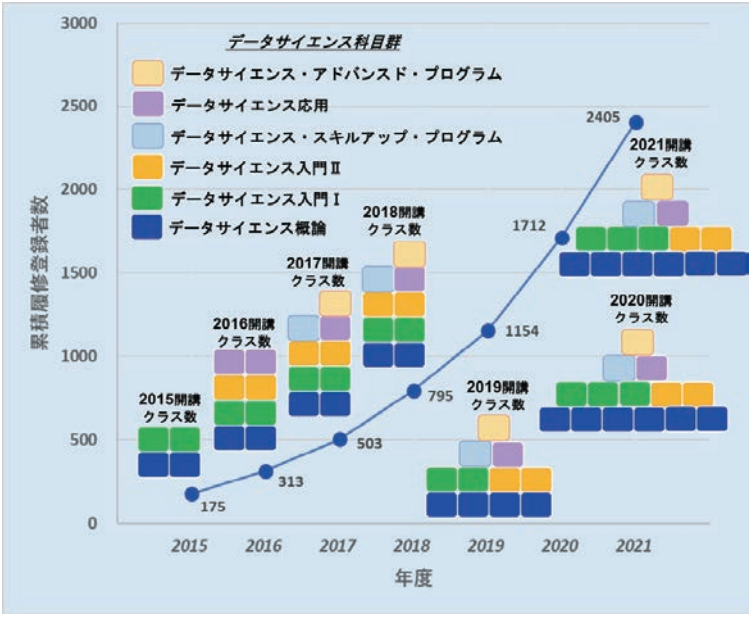


図4　データサイエンス科目群履修者数推移

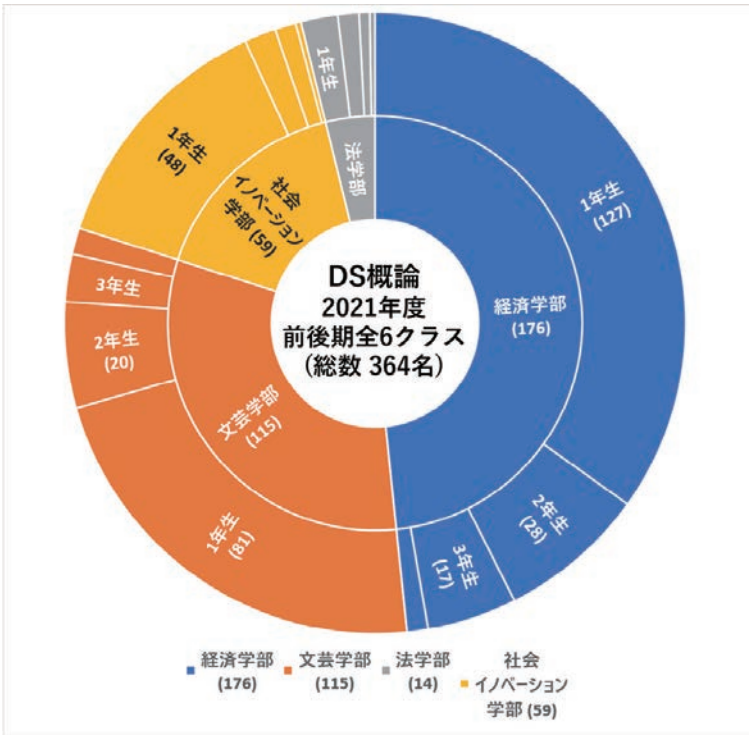


図5　データサイエンス概論の学部別，学年別の履修者の割合

4.1　学生による授業評価アンケートから

本学では，毎年前期・後期に履修する学生全員を対象にした「授業改善アンケート」を実施している．**表3**は，2020年度後期で実施されたアンケートにおける評価点（5点満点）のデータサイエンス科目群全体の平均値と大学全体のすべての授業科目の平均値の比較である，学生からの授業評価は非常に良い評価を得ている．

表3　授業改善アンケート（2021 年度後期授業）

タイプ	項目	評価値	
		データサイエンス	大学
授業のレベルと学生の取り組み	この授業のレベルはあなたにとって適切であった	4.1	4.1
	この授業の内容を理解するために努力した	4.4	4.4
学生の授業分野への評価	この授業は総合的に判断して自分にとって有意義だった	4.6	4.2
	この分野への興味・関心が引き起こされた	4.5	4.2

アンケート結果では，データサイエンス科目群は，本学Webサイトに公開している12項目すべてで大学全体の平均値を上回っていた．中でも「この授業は総合的に判断して自分にとって有意義だった」という設問で，大学全体の平均値が4.2に対して，データサイエンス科目群は4.6であり，データサイエンス科目群の履修学生の満足度と理解度の高さを示している．

応用レベルの授業に関しては，比較的難易が高いとの声もあるが，それに対する努力や授業後の興味関心の高まりは良い傾向にある．学生にとっては，なじみのない内容の授業内容であるが，授業のかなりの時間を学生のサポートに割くことで，多くのコミュニケーションが生まれることから，多様な質問や希望を授業に反映できる点が重要な要因と考える．

「授業改善アンケート」の中に，「この授業を通じて，下記の各資質・能力のうち，どの項目が身につきましたか．身についた資質・能力をすべてマークしてください」という設問があり，データサイエンス科目群履修学生は，本アンケートの大学全体の平均値よりも，以下の項目で「身についた」と回答した割合が高かった．1．この分野の知識・学力，2．数理的能力，3．構想力，4．柔軟な発想力，5．俯瞰力，6．課題発見力，7．課題解決力，8．協働力である．中でも，2．数理的能力（大学全体平均値6.5％に対して27.4％），5．俯瞰力（大学全体平均値8.9％に対して17.7％），6．課題発見力（大学全体平均値9.8％に対して17.7％）の割合が特に高かった．

しかしながら，応用レベルの授業では途中で履修を辞めてしまう学生がいることも事実である．特に，プログラミングに難しさを感じる学生はいまだに多い印象であり，履修人数が増えていく状況やコロナ禍でのオンライン／オンデマンド授業で，一人ひとりをどのようにフォローしていくかは課題が残る．現状は，ティーチングアシスタントの導入により，リソースを増やす形で，プログラミング支援を検討している．

4.2　担当教員から見た学生の変化

リテラシー・基礎レベルの授業についてはすでに述べたので，ここでは応用レベルのデータサイエンス科目を履修した学生の変化について紹介する．

履修する学生は，人文・社会科学系の学生であるため，データサイエンスの理論やプログラミングに不慣れな学生が多い．しかし，どのように学んでいくのかを知りたいというモチベーションの学生が多く，データサイエンス自体には強い興味を持つ傾向にある．

このような学生は，未経験のプログラミングであっても，授業でインターネットでの調べ方などを教えることで，自律的に学習できるようになる．特に，授業で扱うプログラミングの技術が一定程度のレベルに達すると，学外のコンペティションやコンペティションサイトへ参加したりと，自身で積極的にデータサイエンスへかかわろうとする意欲を持つ学生が増えていく．データサイエンスのスキルを客観的に示したいという意志で応募する傾向があり，データサイエンスへ参加する場を求める傾向にあると考えている．

そのため，データサイエンス応用やアドバンスト・プログラムの授業後半のPBLのグループワークにおいて，さまざまな形でデータサイエンスを体験することは非常に有意義であると感じる．データサイエンスに何らかの形で貢献したという経験により，モチベーションが高まる学生

が多く，実際に，データサイエンス系の企業に就職した者や，大学院に進学した学生が出てきている．

本データサイエンス・カリキュラムは2015年度より開講されているが，履修する学生の経年の変化も多分に感じられる．特に最近では，授業参加時点でプログラミングの経験がある学生が増えてきた点である．データサイエンスの最初の接点として，KaggleやSignateなどのデータサイエンスのコンペティションサイトへの参加を経験していたり，興味やある程度の知見を持っている学生が増えてきている．そのため，授業との関連性においても無視できない．正規のプログラム以外に，外部のイベントなどを積極的に活用していくことも今後重要と考えている．

学生同士でデータサイエンス科目群の履修を進める動きが出てきたことは，我々にとって大変嬉しいことである，たとえば，学部デーを利用して，学生有志がデータサイエンス授業を語る場があり，先輩から後輩へとその経験が受け継がれている．また，ディプロマ修了者が大学案内等の各種パンフレットにロールモデルとして登場して，データサイエンス科目群について率直な感想を述べていることも，在学生の授業履修に繋がっている．

4.3　人文・社会科学系専門科目へのフィードバック

応用レベル教育では，統計とAI・機械学習を中心に，データ分析を学ぶことになる．実際のデータサイエンスの現場では，意思決定支援において統計分析の利用も多く，データサイエンスと人文系・社会科学系の共通部分は大きく，親和性も高い．本カリキュラムでは，マーケティングデータから国家統計のようなビジネスから経済まで幅広いデータを扱う．さまざまなデータでの分析仮説を立てたり，その検証をしたりすることで，基本的なデータ操作や基礎分析を身に着けることができ，他分野にも有用と考える．

一方で，統計分析では十分に対応できないタスクも存在する．たとえば，自然言語処理，画像処理／認識である．これらは，業務の自動化が求められるプロジェクトではよく問われる領域でもあるし，人文・社会科学系でも多くの分析需要があり，機械学習やAIが得意とする領域でもあるので，本授業では発展的な内容として教えている．人文・社会科学系の授業や研究にも新たにバリエーションを与える領域とも期待されていて，データサイエンスのスキルを身につけることはこれらの領域での実務にも貢献できると期待される．

5. 今後の展望

2022年度から，表3に示した成城大学データサイエンス教育プログラムの次のバージョンがスタートする（**表4**）．

表4　データサイエンス・新カリキュラム（2022〜）

カテゴリー	科目名
リテラシーレベル （基礎力ディプロマ ）	DS 概論 DS 基礎
基礎・応用レベル （中級ディプロマ ）	データアナリティクス基礎 データアナリティクス応用 機械学習基礎 機械学習応用
発展レベル （EMS ディプロマ）	DS ワークフロー・プログラム DS アドバンスト・プログラム DS 特殊授業 I〜IV

新カリキュラムの大きな特徴は3つある．1つは，カテゴリを3つに改編したことである．リテラシーレベルのDS概論とDS基礎は基礎力ディプロマ取得のための必修科目であるが，現行カリキュラムのDS入門I，IIとDSスキルアッププログラムの統計学や機械学習といったやや数理科学的な内容の科目を分離して，データアナリティクス基礎，機械学習基礎として基礎・応用レベルに組み込んだ．これは，基礎力ディプロマを取りやすくして，数学に特に苦手意識のある学生にもデータサイエンスを学んでほしいからである．データアナリティクス基礎，機械学習基礎は中級ディプロマ取得の必修科目であるが，さらに詳しく統計学や機械学習を学びたい学生のための選択科目として，データアナリティクス応用，機械学習応用を設けた．リテラシーレベルと中級レベルのプログラムの必修科目は，それぞれ，数理・データサイエンス教育強化拠点コンソーシアムが推奨する数理・データサイエンス・AIモデルカリキュラムのリテラシーレベル[1]と応用基礎レベル[8]の内容を含むものとなっている．これも特徴である．

3つ目の特徴は，発展レベルをより実践的な内容にしたことである．EMSディプロマはDSワークフロー・プログラム，DSアドバンスト・プログラム，DS特殊授業I〜IVから2科目を選択することを必修とした．DSワークフロー・プログラムはビジネス・ファイナンス系の実践的データサイエンスを企業で実務に携わる方に担当してもらう授業・演習科目である．また，DS特殊授業I〜IVは主専攻の専門科目とデータサイエンスとの連携を目的とした科目群で，人文・社会学系の分野でデータサイエンスを応用した研究を行っている方に担当していただく予定である．先に述べた，毎年開催する予定のデータサイエンス・ワークショップ[5]はそのための調査・準備の役割がある．

2021年度からは，公立はこだて未来大学との連携を行う．その1つの試みとしてデータサイエンス科目のTA（ティーチング・アシスタント）として大学院生を派遣していただくことになっている．理系の大学院生と議論することで，人文・社会科学系の学生と発想の違いを学び，データサイエンスの広さを知ることができる．ほかにも，KaggleやSignateといった外部コンペティションサイトやハッカソンへ共同で参加を行うなど，いろいろと期待を膨らませている．

また，データサイエンス関連イベントへの参加や関連資格取得の支援は，積極的に行っていきたい．最近，データサイエンティスト協会からデータサイエンス検定が発表されるなど[9]，一定の認知がある資格やイベントなどが増えてきている．スキルを証明することが難しく悩む学生には，それが視覚化される資格やイベント参加などの実績は，就職活動などで大きな武器となると考えられる．

データサイエンスはすべての学生が「育むべき新たな力」であることは間違いない．それをどのように学生に伝えたら良いのか，CDS3は「人文・社会科学系大学におけるデータサイエンス教育」をこれからも模索していく．

参考文献

1）数理・データサイエンス教育強化拠点コンソーシアムモデルカリキュラム（リテラシーレベル）：http://www.mi.u-tokyo.ac.jp/consortium/model_literacy.html
2）数理・データサイエンス・ＡＩ教育プログラム認定制度（リテラシーレベル）：https://www.mext.go.jp/a_menu/koutou/suuri_datascience_ai/00002.htm
3）成城大学データサイエンス教育研究センタシンポジウム：人文・社会科学系大学におけるデータサイエンス教育，https://www.seijo.ac.jp/news/jtmo42000000s0r0.html
4）成城大学データサイエンス・コンテスト2021：https://www.seijo.ac.jp/education/support/cds3/contest/index.html
5）成城大学データサイエンス・ワークショップ2021：文学作品のテキストマイニングーデータサイエンスが解き明かす作品の痕跡ー，https://www.seijo.ac.jp/news/jtmo42000000zwve.html
6）成城大学シラバスについては：https://www.seijo.ac.jp
7）河本薫：データ分析と意思決定の狭間」とそれを埋める力，情報処理学会デジタルプラクティス Vol.6 No.3（通巻第23号）(2015)
8）数理・データサイエンス教育強化拠点コンソーシアムモデルカリキュラム（応用基礎レベル）：http://www.mi.u-tokyo.ac.jp/consortium/model_ouyoukiso.html
9）データサイエンティスト検定リテラシーレベル：https://www.datascientist.or.jp/dskentei/

脚注

☆1　今は「日経クロストレンド」に統合された．
☆2　本学も2021年度より連携校として参画した．
☆3　東日本大震災時の福島原発事故による放射能汚染に対する対応や，コロナ禍での行動変容などに表れるリスクに対する個人の認識には大きなばらつきがある．
☆4　2021年度より小宮路経済学部教授がセンター長に就任した．
☆5　2022年度より新カリキュラムがスタートする．
☆6　EMS : Excellently Motivated Student
☆7　シラバスの詳細は成城大学Webサイトをご覧いただきたい[6]．
☆8　数理・データサイエンス教育強化拠点コンソーシアム「モデルカリキュラム」，データサイエンティスト協会「スキルチェックリスト」，情報処理推進機構「ITSS+」など．

増川純一（非会員）maskawa@seijo.ac.jp

1958年生．1987年大阪大学工学部基礎工学研究科後期博士課程修了．工学博士．成城大学経済学部教授．専門は経済物理学．物理学会会員．

辻　智（非会員）dstsuji@seijo.ac.jp

1959年生．1986年日本アイ・ビー・エム（株）入社．1996年名古屋大学大学院工学研究科量子工学専攻博士後期課程修了．博士（工学）．2018年より成城大学特別任用教授．日本認知科学会会員．

田村光太郎（非会員）k.tamura.phd@gmail.com, k4-tamura@nri.co.jp

1988年生．2016年東京工業大学大大学院総合理工学研究科博士課程修了．博士（理学）．（株）野村総合研究所主任データサイエンティスト，成城大学データサイエンス教育研究センタ外部アドバイザリー委員，電気通信大学客員准教授．専門は複雑系物理学，データサイエンス．物理学会会員，人工知能学会会員．

受付日：2021年9月1日
採録日：2021年9月30日
編集担当：江谷典子（Peach・Aviation（株））

会誌「情報処理」Vol.63 No.2（Feb. 2022）「デジタルプラクティスコーナー」

特集号招待論文

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[1]

[1]東京農工大学

近年，付加価値の高い農作物の生育状況の把握や収量を予測するためにドローンや衛星リモートセンシングによる空撮画像が使われている．一方，収量に大きな影響を及ぼす空撮画像から得られる有効な説明変数はいまだ明確になっていない．本稿では，空撮画像からトマトの各株における草高や植生指数を計測し，各株ごとにそれらの平均値や分散などの1次元計測値，画像のきめや画素間の空間パターンを考慮できるテクスチャ情報を2次元計測値としてそれぞれ計測した．次に，得られた計測値から計5種類の変数選択法によってトマトの果実部を除く地上部バイオマス重（以下，バイオマス），果実重，果実数にそれぞれ影響の大きい計測値を選択し，選択された計測値を説明変数とし，3つの機械学習モデルを用いて，バイオマス，果実重，果実数の予測を試みた．その結果，1次元計測値だけでなく2次元計測値を説明変数候補として考慮することで予測精度が大幅に向上すること，収穫の約1カ月前の2次元計測値が果実重や果実数を予測する上で影響の大きい説明変数であることが明らかとなった．この結果は，収量予測におけるデータ取得の効率化に大きく寄与するものである．

1. ドローンの農業活用

トマト（Solanum lycoperisicum L.）は，世界で広く栽培されている野菜の1つであり，人間の健康維持に重要な役割を果たしている．生鮮トマトの世界生産量は約1億8,000万トンであり，そのうち約4分の1は加工用として栽培され，ケチャップ，サルサ，ジュースなどの形で消費されている[1]．主な生産国は中国，インド，パキスタン，トルコ，アメリカで，これらの国で世界の生産量の約60%を占めており，生産量と収穫面積は年々増加している．

近年，植物の草高や葉の形態などの形質を広範囲で測定する手段として無人航空機（以下，ドローン）が注目されている．ドローンは，操作が比較的容易で，携帯性に優れ，飛行経路や飛行速度，撮影高度や撮影間隔の制御を事前設定でき，圃場☆1における生育状況の把握等に用いられる高分解能の画像を効率よく取得できる利点がある[2],[3]．また，ドローンには，RGBカメラ[4]やマルチスペクトルカメラ[5]など，農業用途に役立つさまざまなセンサを搭載することができ，農業と情報の技術融合に新しい視点をもたらすことが期待されている．ドローンを使ったト

マトの生育状況の把握を試みた研究には，RGBおよびマルチスペクトルの時系列空撮画像を用い，機械学習モデルによって群落における葉面積や成長速度，収量，果実数の予測可能性を調べたものがある[6]．しかしながら，既往の研究は，群落単位での評価を試みたものであり，より詳細な株単位でのトマトの収量予測を試みた研究はない．

ドローン空撮画像により得られた作物の生育期間中における表現形質を用いた収量予測研究では，機械学習アプローチが用いられ，有用な知見が得られている[7]．一方，多時期の表現型データセットを収集するには，一般に時間と人的労力を要する．したがって，トマトのバイオマスや収量の予測にとって重要となる生育時期の主要な表現形質を効率よく得ることができれば，データ収集とその解析処理にかかる労力は大幅に削減可能となる．ドローン空撮撮影により得られる表現形質と機械学習を用いた作物の収量予測のために最適な説明変数の検討は比較的数多く実施されているが，トマトの株単位におけるバイオマス，果実重，果実数の予測のための重要な変数の探索・選択を試みた研究はほとんどない．したがって，これらの予測にとって重要となる変数を選択し，機械学習アルゴリズムを用いることで，トマト収量の予測を高精度で実施することを可能にする技術の確立が期待できる．また，育種サイクルの短縮に資する情報の抽出が期待できるともに，圃場におけるデータ収集やその前処理にかかる労力が軽減される．本稿では，ドローン空撮画像から得られる草高☆2や植生指数☆3マップから，バイオマス等の予測に影響の大きい変数群を変数選択法によって選択し，選択した変数群を用いて複数の機械学習アルゴリズムによってトマトのバイオマス，果実重，果実数の予測を試みた結果を報告する．

2. 圃場栽培と空撮画像取得

2.1 栽培実験圃場

栽培試験は，露地栽培に適した加工用トマト（品種：なつのしゅん）を用い，2020年5月13日から7月30日まで，東京農工大学フィールドミュージアム府中の実験圃場において**図1**に示す3反復（各区5m × 5m）で実施した．なお，移植の1カ月前に温室内で育てた苗床を圃場に移植し，トマトの苗は支柱等で固定せずに栽培を実施した．栽植密度は，畝間☆4 0.85m，株間0.40mとした．施肥は，移植前に元肥（N：P：K＝10：10：10kg $10a^{-1}$）を与え，移植後1週間は，各区画に灌漑チューブを用いて500mlの点滴灌漑を朝夕の計2回，30分ずつ実施し，以降は雨水のみによる栽培を行った．また，各プロットには農業用のプラスチックマルチフィルムを用い，定期的に手作業で除草を実施した．

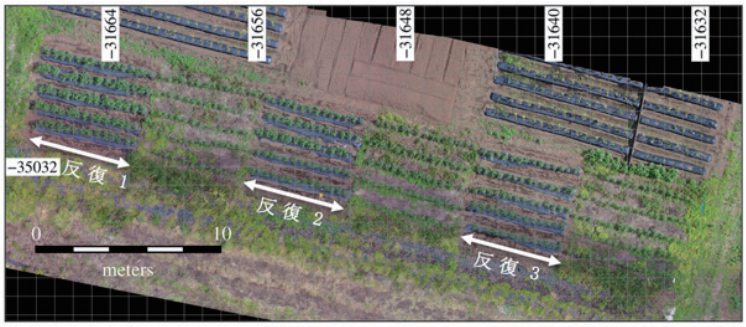


図1　栽培試験圃場（本図の撮影日は2020年6月18日）

2.2　ドローンによる圃場空撮

DJI Matrice 210V2（DJI Co.,Ltd., Shenzhen, China）に搭載したRGBカメラZenmuse X5S（DJI Co.,Ltd., Shenzhen, China）とマルチスペクトルイメージセンサカメラAltum（MicaSense Co.,Ltd., SEA, USA）を用いて，RGB画像とマルチスペクトル画像を取得した（**図2**）．各カメラの静止画解像度は，RGB画像が5,280 × 3,956 pixel，スペクトル画像（長波長赤外線を除く）は2,064 × 1,554 pixelである．

図2　ドローン本体およびセンサカメラ

トマトの株がドローン機のプロペラ回転による風の影響を受けず，かつ高解像度の画像を用いて各株の変数を選択する目的から，撮影高度は地上12mとし，また，オーバラップ率およびサイドラップ率をそれぞれ90%，70%に設定した．空撮日は2020年の5月24日，5月30日，6月5日，6月11日，6月18日，6月26日，7月2日，7月12日，7月16日，7月24日である．なお，飛

行経路を含む飛行パラメータの設定は，画像から高精度の3次元マップ点群情報の生成を可能にする写真測量ソフトウェアであるPix4Dcapture（Pix4D S.A.，Lausanne，Switzerland）を用いて行った．本研究では，RGB空撮画像および地上基準点データから数値地形モデル（DTM）および数値表層モデル（DSM）マップを作成し，スペクトル画像については，5バンド（青，緑，赤，レッドエッジ，近赤外）の反射率マップを作成した．

2.3　バイオマス・収量の計測

トマトバイオマス・収量に影響を及ぼす説明変数の選択，モデルの学習，予測精度評価のため，各株ごとにバイオマス，果実重，果実数の計測を7月29日と7月30日の両日に実施した．

2.4　草丈と植生指数の計測

各株の草高は，2.2節で得られたDSMからDTMを減算することにより算出した．また，反射率マップから，葉のクロロフィル量や成長の指標としてよく用いられる[8],[9]，3つの植生指数：緑の正規化植生指数（Green Normalized Difference Index；GNDVI）[10]，正規化植生指数（Normalized Difference Vegetation Index；NDVI）[11]，加重差植生指数（Weighted Difference Vegetation Index；WDVI）[12]を以下の式によりそれぞれ導出した．

$$GNDVI = （NIR - Green） / （NIR + Green）$$
$$NDVI = （NIR - Red） / （NIR + Red）$$
$$WDVI = NIR - a \times Red$$

ここで，NIR（Near Infrared Ray）は近赤外，Greenは緑色，Redは赤色領域の反射率を示す．aは，ソイルラインの傾きであり，NIRバンドとRedバンド間の線形関係性から得られる．なお，この3つの植生指数を選択した理由は，1）必要以上に多数の植生指数を用いると多重共線性の問題があるため，必ずしも効率的でないこと，2）NDVIとGNDVIは葉のクロロフィル量と相関があることが知られており，収量予測に広く利用されていること，3）WDVIは土壌バックグラウンド値の影響を考慮することができるため，である．

2.5　表現型計測

2.4節により得られた各株における草高および植生指数からバイオマスおよび収量に関連する可能性が高いと考えられる表現型を計測するために，以下の前処理を実施した．

（1）トマト株部分の抽出：5月24日撮影のオルソモザイク画像を用いて，NDVIの値が0.5以上の画素から作物体部分を抽出する．さらに，雑草と認識できる一部の画素については手動で除去する
（2）各トマト株の重心の決定：（1）で得られた株の閉じた輪郭ベクトルから，各株の重心位置を決定する
（3）各株の関心領域の抽出：（1）および（2）のプロセスにより導出した各株の重心位置を中心に半径20cm内に含まれる株の領域を対象領域として抽出する

以上のプロセスにより得られた撮影日ごとの各株の対象領域から，ピクセル計測値と動的成長率を抽出した．ピクセル計測値については，草高マップと植生指数マップから平均（AVE），標準偏差（SD），歪度（SKEW），範囲レンジ（RANGE），最大（MAX）の計5つを1次元計測

値として算出し，次に，グレーレベル同時生起行列GLCM（Gray-Level Co-occurrence Matrix）によりテクスチャ解析を行い，空間パターンを考慮した13種類の2次元計測値（Sum Average （SA），Entropy （Ent），Different Entropy （DE），Sum Entropy （SE），Variance （Var），Difference Variance （DV），Sum Variance （SV），Angular Second Moment （ASM），Inverse Difference Moment （IDM），Contrast （Con），Correlation （Cor），Information Measures of Correlation-1（MOC-1），Information Measures of Correlation-2 （MOC-2））をそれぞれ算出した（**表1**）．

表1　本研究で算出したGLCM計測値

GLCM 特徴量	省略形	算出式
Sum Average	SA	$\sum_{k=0}^{2(N-1)} k\, P_{x+y}(k)$
Entropy	Ent	$-\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P_d(i,j)\log\left(P_d(i,j)\right)$
Difference Entropy	DE	$-\sum_{k=0}^{N-1} P_{x-y}(k)\log\left(P_{x-y}(k)\right)$
Sum Entropy	SE	$-\sum_{k=0}^{2(N-1)} P_{x+y}(k)\log\left(P_{x+y}(k)\right)$
Variance	Var	$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(i-\mu)^2 P_d(i,j)$
Difference Variance	DV	$\sum_{k=0}^{N-1}\left(k-\sum_{k=0}^{N-1} k\, P_{x-y}(k)\right)^2 P_{x-y}(k)$
Sum Variance	SV	$-\sum_{k=0}^{2(N-1)}\left(k-\sum_{k=0}^{2(N-1)} k\, P_{x+y}(k)\right)^2 P_{x+y}(k)$
Angular second moment (Uniformity)	ASM	$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P_d(i,j)^2$
Inverse Difference Moment	IDM	$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\frac{1}{1+(i-j)^2} P_d(i,j)$
Contrast	Con	$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(i-j)^2 P_d(i,j)$
Correlation	Cor	$\sum_{i=0}^{N-1}\sum_{j=0}^{N-1} P_d(i,j)\frac{(i-\mu_x)(j-\mu_y)}{\sigma_x\sigma_y}$
Information Measure of Correlation-1	MOC-1	$\frac{HXY-HXY1}{\max(HX,HY)}$
Information Measure of Correlation-2	MOC-2	$[1-\exp[-2(HXY2-HXY)\}]^{1/2}$

ただし，$p_d(i,j)$ は，行列(i,j)の相対頻度．

$$p_x(i) = \sum_{j=0}^{N-1} p_d(i,j)$$

$$p_y(j) = \sum_{i=0}^{N-1} p_d(i,j)$$

$$p_{x+y}(k) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p_d(i,j), \qquad k = i+j = 0, 1, \ldots, 2(N-1)$$

$$p_{x-y}(k) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p_d(i,j), \qquad k = i-j = 0, 1, \ldots, N-1$$

$$HX = -\sum_{i=0}^{N-1} p_x(i) \log\left(P_x(i)\right)$$

$$HY = -\sum_{j=0}^{N-1} p_y(j) \log\left(P_y(j)\right)$$

$$HXY = -\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} P_d(i,j) \log\left(P_d(i,j)\right)$$

$$HXY1 = -\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p_d(i,j) \log\left(P_x(i) P_y(j)\right)$$

$$HXY2 = -\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} p_x(i) p_y(j) \log\left(P_x(i) P_y(j)\right)$$

$$\mu_x = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} iP_d(i,j)$$

$$\mu_y = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} iP_d(i,j)$$

$$\sigma_x = \sqrt{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (i-\mu)^2 P_d(i,j)}$$

$$\sigma_y = \sqrt{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (j-\mu)^2 P_d(i,j)}$$

次に，連続する2撮影時期から得た草高と植生指数の変化量を撮影間隔の日数で除することにより，動的成長率を算出した．以上より，合計756個の計測値（計18種の1次元・2次元計測値 × 10撮影日 × 草高・3植生指数，草高および3植生指数の生育期間中における9種の動的成長率）を変数選択のために各株ごとに抽出した．

2.6 変数選択

計算の複雑さを軽減させ，効率的なデータ解析を実施するため，バイオマス，果実重，果実数に影響を与える重要な計測値やその時期を決定するために変数選択法により変数選択を実施した．この変数選択は，機械学習アルゴリズムや回帰モデリングにおける基本的なステップになる．本研究では，2.5節で計測した計756種の計測値を候補とした．なお，1次元計測値と2次元計測値を変数候補とした理由は，1次元計測値と動的成長率を用いた場合（以下，1次元計測値）と全計測値（1次元・2次元計測値，動的成長率）を用いた場合で，予測精度に差がどの程度生じるかについて調べるためである．

トマトの生育状態とその生育時期を考慮してバイオマス，果実重，果実数に影響の大きい変数を選択するため，Boruta[13]，DALEX[14]，Genetic Algorithm（GA）[15]，LASSO[16]，Recursive Feature Elimination（RFE）[17]の5つの変数選択法を用いた．Borutaは，ランダムフォレストをベースにしたノンパラメトリックな変数選択アルゴリズムで，変数の重要度を評価することができ，統計的に有意な変数の選択に有用である．DALEXは機械学習モデルで使用する変数について，損失関数などの属性を説明するノンパラメトリックな変数分析手法である．GAは，遺伝学や生物進化の仕組みに基づいてモデル最適化を行うためのノンパラメトリックな手法である．LASSOはL1ノルムを用いたペナルティにより，特定の不要な係数を削除することで，予測誤差を最小化するための変数を選択するパラメトリックな変数選択法である．RFEは，指定した変数の数量に達するまで重要度の低い変数を削除していくノンパラメトリックな手法である．本研究では，これらの5つの変数選択法を用い，重要度スコアが最も上位の5変数をトマトのバイオマス，果実重，果実数の予測に有用な変数としてそれぞれ採用した．

変数選択により選択された各変数群は，ランダムフォレスト（RandomForest；RF）[18]，リッジ回帰（Ridge Regression；RI）[19]，サポートベクタマシン（Support Vector Machine；SVM）[20]の3種の機械学習モデルの入力に用い，バイオマス，果実重，果実数の予測を実施した．なお，全実測データの80%をモデルの学習データとして使用し，残りの20%をモデルの評価に使用した．各モデルの予測性能評価は，決定係数（R^2）と相対平均二乗誤差（rRMSE）を用いて実施した．

3. トマト収量の予測結果

3.1 草高と植生指数の時系列マップ

図3および**図4**に，生育期間における草高とGNDVIの時系列マップをそれぞれ示す．草高は開花期まで直線的に増加し，その後は葉が水平方向に広がるため，草高の増加率は非常に小さく，結実後もほとんど変化しなかった．一方，GNDVIなどの植生指数もシグモイド型の値を取り，生育初期～開花期にかけては指数関数的に植生指数の値は大きくなるが，開花期以降は葉の老化や枯死などが要因となり緩やかに植生指数は低下する．

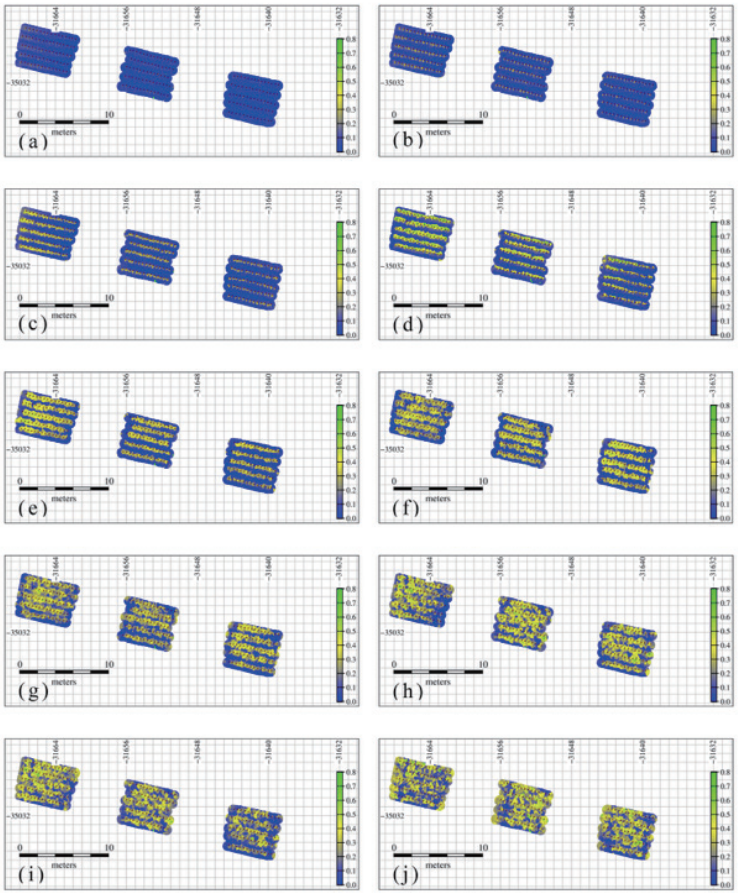


図3　草高（m）の時系列変化（（a）5月24日，（b）5月30日，（c）6月5日，（d）6月11日，（e）6月18日，（f）6月26日，（g）7月2日，（h）7月12日，（i）7月16日，（j）7月24日）

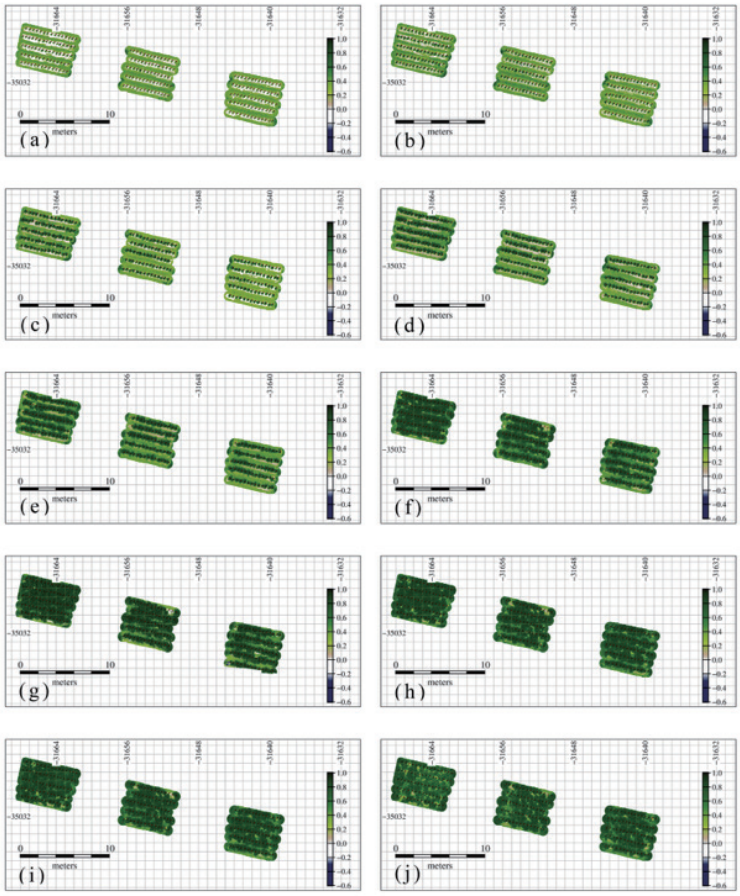


図4　GNDVI（–）の時系列変化（（a）5月24日，（b）5月30日，（c）6月5日，（d）6月11日，（e）6月18日，（f）6月26日，（g）7月2日，（h）7月12日，（i）7月16日，（j）7月24日）

3.2　変数変択の結果

5つの変数選択法を用いて重要度スコアに応じて選択されたバイオマス，果実重，果実数の予測に大きな役割を果たすと認められた上位5つの変数をそれぞれ示す（**表2**，**表3**，**表4**）．バイオマスについては，果実発育中期（6月下旬から7月中旬）の草高と植生指数に関する1次元および2次元の計測値が選択された．用いた変数選択法により結果は異なるが，1次元計測値として草高のAVEとMAX，植生指数のRANGEが複数の選択法において選択され，1次元計測値として草高のMOC-1，MOC-2，NDVIのSVとDVが多くの選択法において選択された（表2）．これらのことから，バイオマスの予測には，草高と植生指数の1次元計測値およびGLCMによるテクスチャ計測値の両方が重要であることが明らかになった．特に後述する果実重と果実数における変数選択の結果と比較し，草高に関する変数が数多く選択されていることが分かる．これらの結果は，今回候補として取り上げた変数や変数選択法によって異なる可能性があるものの，本研究では，バイオマスの予測において，草高の1次元計測値の重要性が明らかになった．バイオマスの予測は，葉の光合成による同化能力を推定する上でも重要であるため，草高とバイオマスの関係性育種家や研究者にとって興味深い情報となる．以上，バイオマスの推定には，果実発育中期以

降の草高および植生指数の1次元および2次元の計測値が相対的に重要度が高い変数であることが分かった．次に，果実重の予測において，すべての計測値から変数選択モデルで選択された変数の多くは植生指数に関するものであった（表3）．Boruta，DALEX，GA，RFEによって1次元計測値のみおよび全計測値から変数選択した結果，6月18日におけるWDVIのRANGEが，重要な変数としてランクインしていることが分かる（表3）．さらに，NDVIのAVEは，GAを除くすべての選択モデルでランクインしていることが明らかになった．以上より，収穫の約1カ月前の植生指数に関する変数が果実重の予測にとって重要であることが明らかとなった．結実開始期の植生指数が最終的な果実重を決定づける重要な因子であることを示す本結果は，圃場管理において果実重を推定する上で注目すべき変数とその栽培時期を示すものであり，栽培管理を行う上で意義深い知見である．果実重と同様に，果実数も収量を決定する上で重要な要素である．上空からの撮影画像だけでは十分に確認することができない各株の果実数を推定する技術は，栽培管理に貢献できる．本研究では，特にNDVIまたはWDVIのAVEが1次計測値および全計測値を対象とした変数選択において，果実数と関係性が高い変数として選択された（表4）．また，果実発育期間の初期から中期にかけての植生指数に関する1次および2次の計測値は果実数の推定に有用な変数となっていることが分かる．一方で，バイオマスとは異なり，草高に関連する計測値は果実数の推定にとって相対的に重要ではないことが明らかとなった．

表2　1次元計測値および全計測値から変数選択法によって選択されたバイオマス予測のための重要度の高い変数

	1次元計測値から選択			全計測値から選択		
Rank	Feature value	Statistics	Date	Feature value	Statistics	Date
Boruta						
1	Plant height	AVE	0626	Plant height	MOC-1	0712
2	GNDVI	RANGE	0716	NDVI	SV	0712
3	Plant height	MAX	0720	NDVI	DV	0712
4	GNDVI	AVE	0712	Plant height	AVE	0702
5	NDVI	SD	0712	GNDVI	DV	0724
DALEX						
1	Plant height	AVE	0626	NDVI	SV	0712
2	Plant height	MAX	0702	Plant height	SE	0712
3	Plant height	AVE	0702	NDVI	SE	0712
4	GNDVI	MAX	0530	NDVI	DV	0712
5	GNDVI	RANGE	0530	Plant height	Ent	0626
GA						
1	Plant height	RANGE	0618	Plant height	RANGE	0712
2	Plant height	AVE	0626	NDVI	SV	0712
3	GNDVI	RANGE	0716	NDVI	IDM	0702
4	WDVI	MAX	0724	NDVI	DV	0712
5	NDVI	SD	0712	GNDVI	MOC-2	0724
LASSO						
1	Plant height	AVE	0626	Plant height	AVE	0626
2	GNDVI	RANGE	0716	GNDVI	DV	0724
3	NDVI	MAX	0716	NDVI	SV	0716
4	Plant height	MAX	0702	GNDVI	Con	0618
5	Plant height	SKEW	0605	Plant height	MAX	0702
RFE						
1	Plant height	AVE	0626	NDVI	SV	0712
2	NDVI	MAX	0716	Plant height	MOC-1	0712
3	Plant height	MAX	0702	NDVI	DV	0712
4	Plant height	AVE	0702	NDVI	MAX	0716
5	NDVI	SD	0712	GNDVI	DV	0724

表3　1次元計測値および全計測値から変数選択法によって選択された果実重予測のための重要度の高い変数

	1次元計測値から選択			全計測値から選択		
Rank	Feature value	Statistics	Date	Feature value	Statistics	Date
Boruta						
1	WDVI	RANGE	0618	WDVI	RANGE	0618
2	NDVI	AVE	0618	NDVI	AVE	0618
3	WDVI	AVE	0618	WDVI	AVE	0618
4	NDVI	AVE	0626	WDVI	SA	0618
5	GNDVI	AVE	0626	NDVI	AVE	0626
DALEX						
1	WDVI	AVE	0618	WDVI	SA	0618
2	NDVI	AVE	0724	NDVI	AVE	0626
3	WDVI	RANGE	0618	NDVI	AVE	0618
4	Plant height	RANGE	0618	WDVI	RANGE	0618
5	NDVI	AVE	0618	GNDVI	IDM	0712
GA						
1	WDVI	RANGE	0618	NDVI	IDM	0716
2	NDVI	MAX	0606	WDVI	RANGE	0618
3	NDVI	AVE	0618	GNDVI	SE	0724
4	NDVI	SD	0716	Plant height	Growth Rate	0530-0605
5	NDVI	SD	0524	WDVI	MAX	0606
LASSO						
1	NDVI	AVE	0618	GNDVI	Con	0618
2	Plant height	MAX	0724	Plant height	MAX	0724
3	NDVI	RANGE	0724	WDVI	SA	0626
4	NDVI	RANGE	0524	NDVI	AVE	0626
5	Plant height	SKEW	0712	NDVI	Cor	0712
RFE						
1	NDVI	AVE	0618	NDVI	AVE	0618
2	WDVI	RANGE	0618	WDVI	RANGE	0618
3	WDVI	AVE	0618	WDVI	AVE	0618
4	–	–	–	NDVI	AVE	0626
5	–	–	–	WDVI	SA	0618

表4　1次元計測値および全計測値から変数選択法によって選択された果実数予測のための重要度の高い変数

	1次元計測値から選択			全計測値から選択		
Rank	Feature value	Statistics	Date	Feature value	Statistics	Date
Boruta						
1	NDVI	AVE	0626	WDVI	RANGE	0618
2	GNDVI	AVE	0626	NDVI	AVE	0618
3	NDVI	MAX	0618	WDVI	AVE	0618
4	GNDVI	MAX	0611	WDVI	SA	0618
5	GNDVI	SD	0626	NDVI	AVE	0626
DALEX						
1	NDVI	RANGE	0606	WDVI	IDM	0618
2	NDVI	AVE	0626	NDVI	RANGE	0626
3	NDVI	AVE	0524	NDVI	AVE	0618
4	NDVI	MAX	0618	WDVI	AVE	0618
5	NDVI	MAX	0618	GNDVI	SA	0712
GA						
1	WDVI	SD	0712	NDVI	IDM	0716
2	NDVI	SD	0524	WDVI	RANGE	0618
3	WDVI	MAX	0606	GNDVI	AVE	0724
4	GNDVI	AVE	0626	Plant height	Growth Rate	0530-0605
5	GNDVI	RANGE	0524	WDVI	MAX	0606
LASSO						
1	GNDVI	MAX	0611	GNDVI	Con	0618
2	GNDVI	AVE	0626	Plant height	MAX	0724
3	WDVI	SD	0606	WDVI	SA	0626
4	GNDVI	AVE	0712	NDVI	AVE	0626
5	NDVI	MAX	0606	NDVI	Cor	0712
RFE						
1	GNDVI	AVE	0626	NDVI	AVE	0618
2	NDVI	AVE	0626	WDVI	RANGE	0618
3	NDVI	MAX	0618	WDVI	AVE	0618
4	NDVI	MAX	0606	NDVI	AVE	0626
5	NDVI	SD	0626	WDVI	SA	0618

3.3　機械学習モデルによる予測結果

変数選択法により選択された変数群を用いて，RF，RI，SVMモデルにより予測したバイオマス，果実重，果実数の実測値とシミュレーション値の関係を**図5**，**図6**，**図7**にそれぞれ示す．また，**表5**に，変数選択により選択された変数群を用いたRF，RI，SVMモデルによる実測値と予測値間のrRMSEの値を示しており，テストデータに対する検証モデルの予測精度を反映するものである．

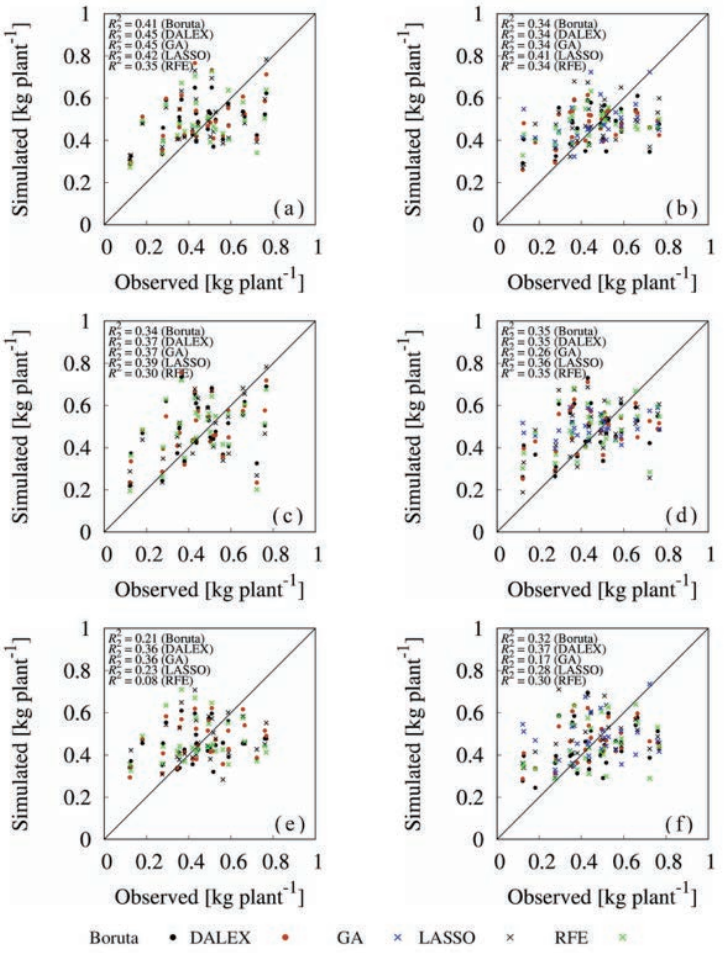


図5　バイオマスの実測値と予測値の相関図（（a）1次次元計測値から選択された変数群を用いたRFによる結果，（b）全計測値から選択された変数群を用いたRFによる結果，（c）1次次元計測値から選択された変数群を用いたRIによる結果，（d）全計測値から選択された変数群を用いたRIによる結果，（e）1次次元計測値から選択された変数群を用いたSVMによる結果，（f）全計測値から選択された変数群を用いたSVMによる結果））

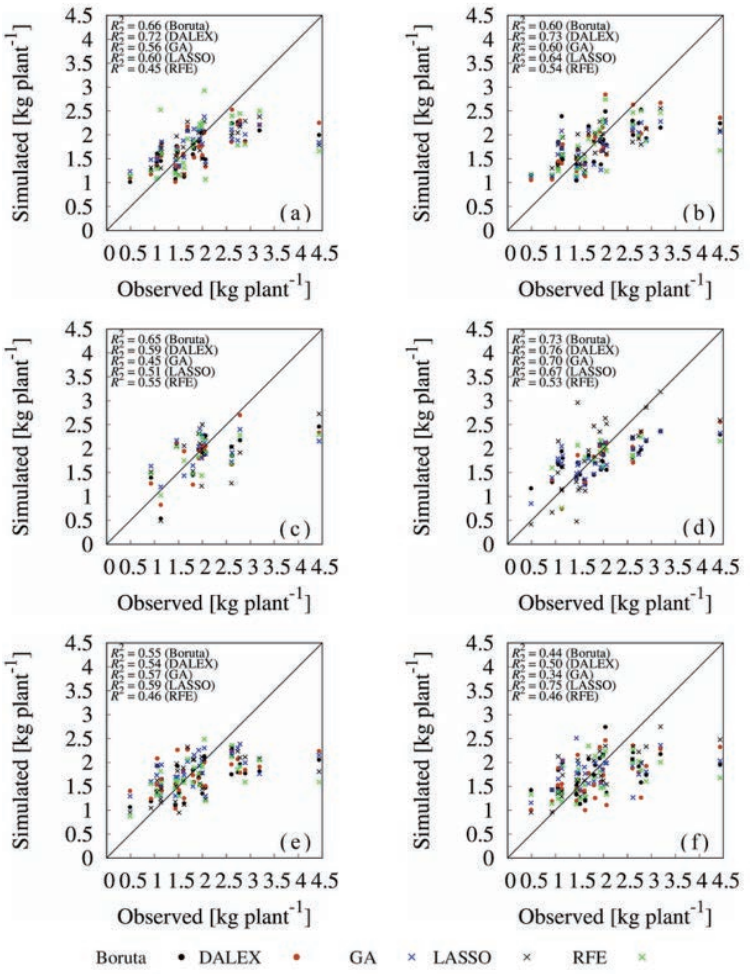


図6　果実重の実測値と予測値の相関図（（a）1次元計測値から選択された変数群を用いたRFによる結果，（b）全計測値から選択された変数群を用いたRFによる結果，（c）1次元計測値から選択された変数群を用いたRIによる結果，（d）全計測値から選択された変数群を用いたRIによる結果，（e）1次元計測値から選択された変数群を用いたSVMによる結果，（f）全計測値から選択された変数群を用いたSVMによる結果））

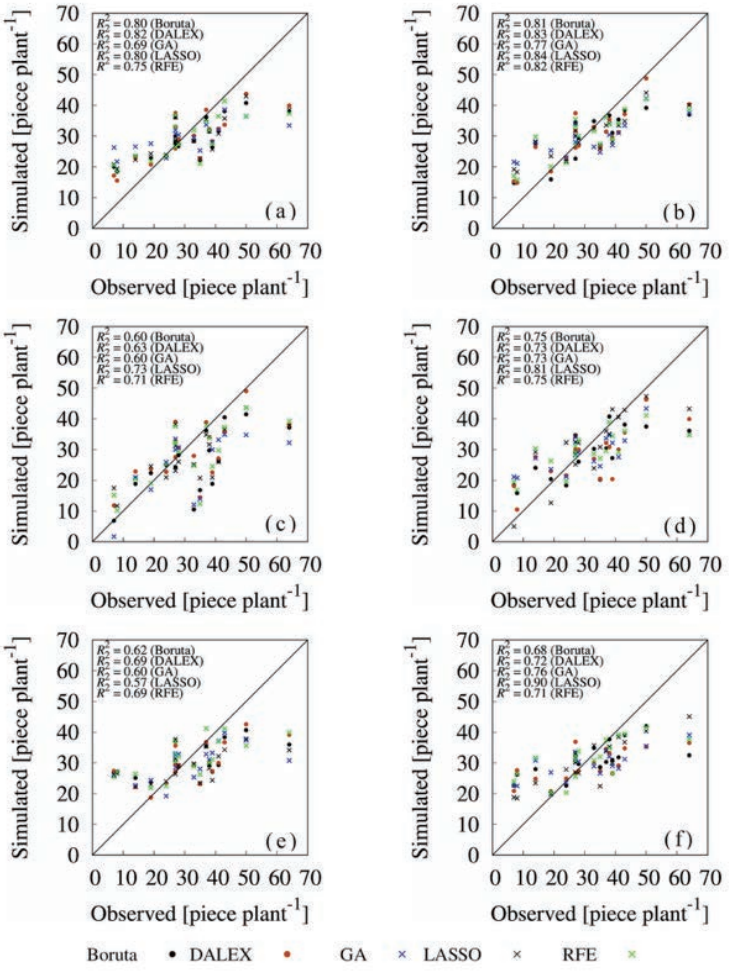


図7　果実数の実測値と予測値の相関図（（a）1次元計測値から選択された変数群を用いたRFによる結果，（b）全計測値から選択された変数群を用いたRFによる結果，（c）1次元計測値から選択された変数群を用いたRIによる結果，（d）全計測値から選択された変数群を用いたRIによる結果，（e）1次元計測値から選択された変数群を用いたSVMによる結果，（f）全計測値から選択された変数群を用いたSVMによる結果））

表5　Random forest（RF），Ridge regression（RI），Support vector machine（SVM）によるバイオマス，果実重，果実数のの予測値と実測値間のrRMSE（%）

1次元計測値から選択					
Model	Boruta	DALEX	GA	LASSO	RFE
	バイオマス [kg plant^{-1}]				
RF	17.8	22.2	16.9	22.5	22.9

RI	26.4	26.7	24.9	30.6	26.7
SVM	17.6	18.9	16.7	21.4	21.8
	果実重[kg plant^{-1}]				
RF	14.0	13.9	13.2	15.7	24.1
RI	49.6	48.5	48.5	50.1	48.0
SVM	14.3	14.5	14.6	18.7	15.9
	果実数[piece plant^{-1}]				
RF	12.6	14.2	10.0	12.4	14.2
RI	30.4	25.5	30.3	18.1	21.2
SVM	13.1	14.2	13.5	13.0	13.6
全計測値から選択					
Model	Boruta	DALEX	GA	LASSO	RFE
	バイオマス[kg plant^{-1}]				
RF	18.7	16.5	15.0	13.6	13.4
RI	22.7	20.5	11.0	25.8	17.6
SVM	21.4	18.7	11.2	21.6	20.5
	果実重[kg plant^{-1}]				
RF	18.0	15.6	12.5	13.8	15.6
RI	11.5	20.6	12.8	22.1	16.7
SVM	14.7	15.4	14.6	15.9	14.5
	果実数[piece plant^{-1}]				
RF	14.9	13.5	11.2	11.5	13.5
RI	15.5	17.7	12.7	28.1	13.4

SVM	12.8	10.9	8.8	10.6	14.1

バイオマス予測のための5つの変数群を比較すると，1次元計測値を使ってBorutaとDALEXにより選択された変数群を用い，RFモデルによって得られた予測結果は，ほかの変数選択法および予測モデルの組合せと比較して高いR^2が得られた（Boruta：R^2=0.41，DALEX：R^2=0.45）（図5a）．rRMSE指標で見た際，全計測値候補からGAによって選択された変数群を用いたRIとSVMを使って得られたバイオマスの予測結果は，ほかの組合せと比較して誤差が相対的に小さかった（表5）．一方で1次元計測値から変数を選択したケースでは，総じてRFがほかの機械学習モデルと比較してR^2の値が高く，rRMSEの指標においては，全計測値から抽出した変数群と予測モデルの組合せにおいて，RFを使った結果は総じてrRMSEの値が小さい結果となった．以上より，R^2で見た際，バイオマスの予測には，GLCMによるテクスチャ情報の重要性は低く，1次元計測値だけで高い予測精度が得られる結果となった．果実重については，全計測値からBoruta，DALEX，GAを使って選択された変数群を用いたRIモデルによる結果（R^2=0.73 for Boruta; 0.76 for DALEX; 0.70 for GA），すべての変数からLASSOを使って得られた変数群を用いたSVMモデルによる結果（R^2=0.75），1次元計測値からRFEを使って得られた変数群をRIで予測した結果（R^2=0.55）の組合せが優れた予測性能を示した．特にRIに着目すると，RFEを除き，全計測値から変数選択し，予測モデルを適用した結果は，1次元計測値のみから選択した変数群を用いたモデル予測結果と比較し，大幅に予測精度が向上した．たとえば，RIモデルの1次元計測値からGAで選択した変数群を使った結果はR^2が0.45に対し，全計測値から選択した結果ではR^2=0.70となった（図6）．また，1次元計測値，2次元計測値ともに，植生指数から得られる変数の重要度が高いことが分かる．以上の結果より，果実重を予測するには，2次元計測値から得られる特徴量が重要であることが明らかになった（図6，表5）．次に，果実数の予測では，全計測値候補から選択された変数群を使って実施した予測モデル結果は，1次元計測値から選択した変数群を用いた結果と比較して，有意に高い適合度を示した（図7）．特に，Boruta，DALEX，GA，RFEで選択された変数群をRFに適用した結果，およびLASSOで選択した変数群をSVMに適用した結果は，ほかの変数群と予測モデルの組合せと比較して高い予測精度を示した（R^2=0.81 for Boruta; R^2=0.83 for DALEX; R^2=0.82 for RFE; R^2=0.77 for GA; R^2=0.82 for RFE; R^2=0.90 for LASSO）．

本研究では，パラメトリック（ノンパラメトリック）な変数選択法とパラメトリック（ノンパラメトリック）な機械学習モデルの間に，トマトのバイオマス・果実重・果実数の予測精度に対する明確な関連性は見られなかった．一方で，果実重および果実数の予測に関して，全計測値から予測に重要となる変数を選択し，これらを用いたモデルによる予測精度は，1次元計測値のみから変数を抽出し，その変数群をモデルに適用した結果と比較して，予測精度が向上した．さらに，収穫の約1カ月前の植生指数の特徴量がトマトの果実重，果実数の予測にとって重要であることが分かった．

以上，トマトのバイオマス，果実重，果実数の予測に対して影響の大きい表現形質や生育ステージに焦点を絞ることで，効果的に表現型データの収集に貢献することができる可能性があることが明らかとなった．今後は，本研究で用いた変数選択法を多地点かつ複数年において収集した

多くのデータでテストすることにより，ロバスト性の高い予測モデルの構築に必要な特徴量を抽出することを目指していく．

参考文献

1）Ramasamy, S. and Ravishankar, M. : Integrated Pest Management Strategies for Tomato Under Protected Structures, In : Sustainable Management of Arthropod Pests of Tomato, pp.313-322 (2018).
2）Shi, Y., Thomasson, JA., Murray. SC., Pugh, NA., Rooney, WL., Shafian, S., et al. : Unmanned Aerial Vehicles for High-Throughput Phenotyping and Agronomic Research, Zhang J, editor, PLoS ONE, Vol.11 (7), e0159781 (2016).
3）Barbedo, JGA. : A Review on the Use of Unmanned Aerial Vehicles and Imaging Sensors for Monitoring and Assessing Plant Stresses, Drones, Vol.3 (2) : p.40 (2019).
4）Du, M. and Noguchi, N. : Monitoring of Wheat Growth Status and Mapping of Wheat Yield's within-Field Spatial Variations Using Color Images Acquired from UAV-camera System, Remote Sensing, Vol.9 (3) : p.289 (2017).
5）Duan, T., Chapman, S. C., Guo, Y. and Zheng, B. : Dynamic Monitoring of NDVI in Wheat Agronomy and Breeding Trials Using an Unmanned Aerial Vehicle, Field Crops Research, Vol.210, pp.71-80 (2017).
6）Johansen, K., Morton, M.J.L., Malbéteau, Y., Aragon, B., Al-Mashharawi, S., Ziliani, M., et al. : PREDICTING BIOMASS AND YIELD AT HARVEST OF SALT-STRESSED TOMATO PLANTS USING UAV IMAGERY, Int Arch Photogramm Remote Sens Spatial Inf Sci, XLII-2/W13, pp.407-411 (2019).
7）Yang, Q., Shi, L., Han, J., Zha, Y. and Zhu, P. : Deep Convolutional Neural Networks for Rice Grain Yield Estimation at the Ripening Stage Using UAV-based Remotely Sensed Images, Field Crops Research, Vol.235, pp.142-53 (2019).
8）Kyratzis, A. C., Skarlatos, D. P., Menexes, G. C., Vamvakousis, V. F. and Katsiotis, A. : Assessment of Vegetation Indices Derived by UAV Imagery for Durum Wheat Phenotyping under a Water Limited and Heat Stressed Mediterranean Environment, Front Plant Sci, Vol.8, p.1114 (2017).
9）Guan, S., Fukami, K., Matsunaka, H., Okami, M., Tanaka, R., Nakano, H., et al. : Assessing Correlation of High-Resolution NDVI with Fertilizer Application Level and Yield of Rice and Wheat Crops using Small UAVs, Remote Sensing, Vol.11 (2), p.112 (2019).
10）Rouse, J. W., Haas, R. H., Schell, J. A. and Deering, D. W. : Monitoring Vegetation Systems in the Great Plains with ERTS, Third ERTS Symposium, NASA SP-351 I, pp.309-317 (1973).
11）Qi, J., Chehbouni, A., Huete, A. R. , Kerr, Y. H. and Sorooshian, S. : A Modified Soil Adjusted Vegetation Index, Remote Sensing of Environment, Vol.48, pp.119-126 (1994).
12）Haralick, R. M., Shanmugam, K. and Dinstein, I. : Textural Features for Image Classification, IEEE Trans Syst, Man, Cybern, SMC-3 (6), pp.610-621 (1973).
13）Kursa, M. B., Jankowski, A. and Rudnicki, W. R. : Boruta - A System for Feature Selection, Fundamenta Informaticae, 101 (4), pp.271-285 (2010).
14）Biecek, P. : DALEX : Explainers for Complex Predictive Models, Journal of Machine Learning Research, Vol.19, pp.1-14 (2018).
15）Scrucca, L. : GA ：A Package for Genetic Algorithms in R, Journal of Statistical Software, Vol.53 (4), pp.1-37 (2013).

16）Tibshirani, R. : Regression Shrinkage and Selection Via the Lasso, Journal of the Royal Statistical Society, Series B (Methodological), Vol.58 (1), pp.267-288 (1996).
17）Guyon, I., Weston, J., Barnhill, S. and Vapnik, V. : Gene Selection for Cancer Classification Using Support Vector Machines. Machine Learning, Vol.46 (1/3) pp.389-422 (2002).
18）Breiman, L. : Random Forests, Machine Learning, Vol.45 (1), pp.5-32 (2001).
19）Hoerl, A. E. and Kennard, R. W. : Ridge Regression, Applications to Nonorthogonal Problems, Technometrics, Vol.12 (1) ,pp.69-82 (1970).
20）Cortes, C. : Vapnik V. Support-vector Networks, Machine Learning, Vol.20 (3), pp.273-297 (1995).

脚注

☆1　圃場（ほじょう）：農作物を育てる場所
☆2　草高（そうこう）：地面からあるがままの状態の作物最上部までの高さ
☆3　植生指数（しょくせいしすう）：植物による光の反射特性を使って得られる植物の量や活力を表す指標
☆4　畝間（うねま）：畝と畝との間の中心から中心までの距離

辰己賢一（非会員）tatsumi@go.tuat.ac.jp
東京農工大学農学研究院農業環境工学部門准教授，2002年京都大学大学院工学研究科環境専攻修了．専門は農業情報気象学，作物生育シミュレーション．気候変動が農業生産に与える影響の定量的分析に関する研究に従事．

受付日：2021年9月20日

採録日：2021年10月31日
編集担当：飯村 結香子（NTT）

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～」座談会

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．

里　洋平（正会員）（（株） Village AI／nat（株）／（株） Lupinus）
R 言語の東京コミュニティTokyo.R 創立者．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．

高柳慎一（正会員）（（株）ユーザベース）
2020年総合研究大学大学院複合科学研究科統計科学専攻博士課程修了．博士（統計科学）．2020年（株）FORCAS入社．2021年統合により（株）ユーザベースへ転籍．B2B事業向け顧客戦略プラットフォームFORCASの開発に従事．徳島大学客員准教授．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．

安部晃生（非会員）（（株）コネクトデータ）
（株）コネクトデータ代表取締役．企業におけるデータ利活用のためのコンサルティング，分析，開発，教育に従事．

飯尾　淳（正会員）（中央大学）
中央大学国際情報学部教授．人間と情報システムのインタラクションに関する研究に従事．特定非営利活動法人人間中心設計推進機構理事．（一社）ことばのまなび工房理事．博士（工学），技術士（情報工学部門），人間中心設計専門家．

牧山幸史（非会員）（ヤフー（株））
ヤフー（株）にてデータサイエンス業務に従事するかたわら，（株）ホクソエム代表取締役社長と徳島大学客員准教授を兼任する．

石井一夫（正会員）（公立諏訪東京理科大学）
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．

里：こんにちは，本企画でゲストエディタをやらせていただいている里です．本日は，よろしくお願いいたします．最初に，自己紹介を皆さんにお願いしたいと思います．高柳さんからお願いしてもよろしいでしょうか．

高柳：（株）ユーザベースでデータサイエンティストをしている高柳です．よろしくお願いします．

今メインでやっているデータサイエンスの業務は，営業支援システム開発です．本特集の論文にも書かせていただいたのですが，企業が営業するときにアタックリストを機械的にAIで作るようなシステムを作っています．よろしくお願いします．

里：よろしくお願いします．

次は，安部さん，お願いいたします．

安部：（株）コネクトデータ，代表取締役の安部晃生です．

私は，普段はクライアント企業のデータ利活用を支援するためにコンサルティング，分析，開発，教育等々をやっております．最近だとオープンデータ活用みたいなところに非常に興味がありまして，世の中のオープンデータの流通を活発にしたり，それをベースに何か技術発展したりのようなことを狙って，delikaというオープンデータのプラットフォームを開発，提供をしております．よろしくお願いします．

里：よろしくお願いします．

次，飯尾先生，お願いいたします．

飯尾：飯尾でございます．どうぞよろしくお願いいたします．

私は中央大学の国際情報学部というところで教鞭を取っていまして，その国際情報学部というのはほかにあまりない学部名なのですけれども，情報系と，あと法律の先生たちがいます．今の情報社会を技術的に支えるのが情報系の学者の分担であって，それを社会実装していくときに，社会のルールに合わないといけないよねということで法律系の先生方が，自動運転の車が事故を

起こしたらどうなるのみたいな，そういう話は今後，重要になってくるので，そのようなことを学生に教えているという，そういう建て付けの学部なのですね．私は法律ではなくて，情報系のほうで，いろいろ教えています．

データサイエンス関連でいうと，今日の座談会の中でお話できると思うのですけれども，本学は今すごくデータサイエンス教育に力を入れているというか，政府の，ちゃんとやれよというのに乗っかってやっていますので，そのあたりのお話なんかをできればいいかなと思っています．

里：ありがとうございます．

牧山さん，お願いします．

牧山：ヤフー（株）の牧山と申します．仕事ではデータ分析全般をやっていて，何でも屋みたいな感じです．よろしくお願いします．

里：はい．よろしくお願いいたします．

ありがとうございます．

では，最後に石井先生，お願いします．

石井：公立諏訪東京理科大学の石井と申します．今年の4月から現職に異動してきました．それ以前は久留米大学にいて，そこで医療ビッグデータを中心に教育研究をやっていたのですが，現在は本学工学部の情報応用工学科に勤めています．本学は全学を上げて，AIとか，機械学習とか，ビッグデータとかに力を入れていて，私もそれに乗っかっていろいろやるということで，相変わらず医療ビッグデータを中心に教育研究をやっています．

最近は地球温暖化にも興味があって，近ごろすごい豪雨とか，熱波とか，大変なことになっていますけれども，そういう関係の分析も含めいろいろやっています．

今日は本企画の，コーディネータとして本座談会をバックアップさせていただきたいと思います．よろしくお願いします．

里：お願いします．

データと法律

里：どなたか，これをぜひ話したいという強い思いがあれば，その話からやっていければなと思うのですが．

安部：さっき飯尾先生から法律と情報みたいな話があって，私も結構，興味がある分野で，先ほどの当社のサービスでdelikaを作ったきっかけの1つとして，著作権改正みたいなところがあるのです．日本の著作権法というのはいわゆる情報処理の用途で非常に自由に利用できるような形に改正されて，日本は機械学習天国だ，みたいなことがいわれることもあると思っています．

そういうデータの使いやすさというものが法として整備されている一方で，そのデータの流通みたいなところでいうと，こういうデータを勝手に使っていいのだろうか，そういう自制心みたいなものが働いて，データの活用みたいなのが活発になりづらいなと，サービスを運営していて思っているところがあります．皆さんそのデータの活用というところで，恐らく自分たちが獲得しているデータに関しては問題なく使っていると思うのですが，オープンデータにしろ，いわゆる売買されているデータにしろ，外部のデータに関して，どういうふうな考え方をお持ちなのかなというのに，少し興味があって，お聞きしたいなと思います．

飯尾：では，それを受けてちょっとしたエピソードというか，学生に伝えたいことという観点から，グレーゾーンというわけでもないのですけれども，外部からデータを取ってくるときに微妙なところを，超えてはいけないところはどこにあるのかというようなところが少し曖昧ではっきりしていないケースというのは結構ありますよね．そこがやはり気になりますね．

有名なところだと，昔，Librahack事件というのがありましたよね．岡崎市の図書館に1秒ごとに，あれは確か1秒ごとにアクセスしていたと思いますけれども，図書館のシステムのほうが，作りが不十分でなんかどんどんリクエストが溜まっていって，それでダウンしてしまったみたいな．1秒ごとぐらいのアクセスだったら全然オーケーなのではないかと普通は思いますけれども，図書館側のシステムがそれに耐えられないようなシステムだった．そういうことは，普通は技術屋としては想像できないですが，実際に，ああいうことが起こってしまうと，それは技術面からの問題もあるけれども，法律も整備しきれていないところですし，社会の文化的なところも，あれは大変良い教訓を残したとは思うのです．そのあたりを学生にどう伝えていくのかというのは少し気になっていますね．

今，私も，Twitterのデータを取ってきて，それで毎日分析というか，20分おきにデータを取ってきて，それで分析しているシステムを運用して，いろいろやっています．TwitterのデータはAPIを叩いて取得していますが，それ以外のデータの収集がなかなか難しい．そのあたりもデータを取り扱う技術というよりも，社会的にどこまでやってよくて，そこから先はアウトという話で，我々は経験上，ここまでだったら大丈夫だろうなという感覚は持ってはいますが，そのあたりを，学生にどう教えていくのかというところはかなり気になっています．

里：ほかの方はどうでしょうか．

石井：気になっているのは，私は医療ビッグデータを使っている関係で，この分野は個人情報保護の規制がかなり厳しくて，使うデータが個人を特定できないようにするとか，データ分析というのは個人情報との戦いみたいなところがあり，どこまで個人情報を暴くかというところで，個人を特定できるぎりぎりまで攻め込んでいくというのは結構やるのですが，そのあたりが法律と，あと倫理的なところとの綱引きというのが結構大変だなというのはありますね．

飯尾：今，倫理という言葉が石井先生からありましたけれども，最近やはり，研究倫理の問題がすごく，ややこしくなってきています．バイオの研究とか，石井先生は割とそちらのほうもずっとやられてきたので医療の土地勘はあるかと思いますけれども，人造人間を作ってはいけないとかね（笑），試験管ベビーがどうだとか，そういうような話は，これはあかんやろという感じ

で，すぐ分かるのです．けれども，だんだん最近は周辺領域というか，つまり社会科学のところまでそういうのが求められるようになってきて，昨今，とうとう我々も白旗を上げて，その波に飲まれているかなというところがあります．

私の周辺でいうと，文学部の心理学の先生たちがだいぶ戦ってくれたのですけれども，世間の波に飲まれてしまって，少しやりすぎなのではないかなというふうには思っています．要するに，人にかかわる研究というのはどこまでの倫理を求めるんだと．個人情報を扱っていれば，それは人にかかわる研究だってされてしまうのですよね，今の文脈だと．そうすると，研究倫理委員会を通さないといけないとか，面倒くさいことばかり増えて，それはしょうがないのかなとは思いつつ，困っていますね．

石井：新型コロナウイルスのデータ分析とか，まさにそれですよね，新型コロナウイルス感染症患者のデータ分析とか．

飯尾：おっしゃる通りだと思います．だから，誰かが反旗を翻してくれないかなと思っているのですけれども（笑），残念ながら私はそこに，先頭に立つ勇気がないので（笑）．まあ，そんなところですかね．

データの流通

里：ほかの方，どうですか．

高柳：だいぶ視点が違っていて最近よく考えている会社でも話している内容ですけれども，外部データとか，オープンデータに依存してしまうと，それは経営リスクだよね，みたいな話はよくしていますね．たとえば，データ分析というのは要するにデータの加工産業みたいなものじゃないですか．データを仕入れてきて，それを適切に調理して，料理として出すのが分析レポートですし，それをシステム化したりしているわけです．こういう状況で，たとえば今年だと特に野菜が値上がっているかと思うんですが，同様に外部データとして買っているものが値上がってしまったとか，データそのものの供給を止めたとかとなると，我々のそのデータ加工ビジネスが，ぽしゃってしまうのでそれは経営リスクだよね，じゃあ，内製化するか，どうしようかな，みたいな視点で話していることが結構多いですね．

オープンデータの値上げというよりも，どちらかというと，供給がいきなり止まって，代替先を探さなければというので，慌てふためくのがちょっと嫌かもみたいな話はよくしていますね．

安部：気になるところとしては，オープンデータに近い概念でオープンソースというのがあると思うのですが，世の中のシステムとかというのは結構オープンソースのコードとかを使って，まわっていたりするではないですか．オープンソースとオープンデータの性質の違いというのはどういったところにありますか．1つの違いは，継続的に更新されていくみたいなことだと思いますけれども．

高柳：我々のビジネスの話をしてしまうと完全にそれで，オープンソースだと適当にバージョンというか，GitHub，Gitだと，コミットIDとかタグで固めてしまったものさえあれば，しばらくはまわるのですが，オープンデータはイメージとして常に新鮮なデータが入ってきていないと

ヤバいみたいな話なのですね．たとえば，今，まさに我々，Zoomで会議していますけれども，Zoomを使っている企業というのをデータの中から抽出したいみたいな案件があって，それは毎月毎月，毎時毎時，新しくZoomを使うという企業があったり，やめてしまったりみたいなのがあるので，そういう意味で，継続的に見続けてなければいけないので，ちょっと違うみたいな．ストックというよりもフローに近い感じなのですよね．そこに差がある感じですね．

安部：面白いですね．確かにおっしゃる通りかなというふうに思います．それは結局オープンデータというものの外部依存というところが，あるいは，外部から一定に供給されることが止まるリスクというものが上手くコントロールできれば，まわるという形になる．

高柳：そうです．おっしゃる通りです．

安部：それは各々のビジネス主体がもう独立に動いているから，その供給するという関係性を意識せずにまわすので止まるリスクになるという感じですかね．結構，世の中というのは，どこかが止まるとほかに影響が出るから，それを止めることに対してのリスクを，公共機関とかが守ってくれたりするではないですか．あるいはビジネスサイドでも，うちのこのビジネスは赤字だからやめたいけれども，これをやめると影響が出るから困るよね，みたいなところがあると思うのですよ．

でも，データだと結構それがリスクになり得るということは，あまりデータの流通というものというのが世の中に意識されていないのかなというふうに思いましたね．

高柳：そうですね，流通が意識されていない，ありそうですね．

データの質

飯尾：オープンデータに関してはね，質の問題も結構あるのではないかなと思うのです．そのあたりは皆さん，いかがですか．つまり使い勝手のいいというか．いわゆるTimothy "Tim" John Berners-Leeの5 Starオープンデータでいえば，レベル1とか，レベル2のところでとどまっている．少なくとも私の経験では，使い勝手の悪いデータばかり流通していて，そこをなんとか加工して，クリーニングして使っているというような状況なのですけれども，皆さん，いかがでしょうか．

安部：私もそう思いますね．あの5 Starでいうと，機械判読が可能みたいなレベルが確かレベル2か，3ぐらいにあったと思うのですけれども，まずはそのレベルに到達するところが最初かなというのは思っています．今，当社が提供しているdelikaというプラットフォームも実は，5 StarだとRDFを使って，自由にデータとデータが繋がるよ，みたいなところがあると思うのですけれども，それよりも実際に利用できる部分をまず目指そうよみたいなところで，機械判読のデータというものを流通させたいなという思いがありますね．

データが使いにくくて，データサイエンティストがみんな同じデータの前処理をしているみたいなところがあるので，まずそういうところをなくすことによってデータの価値というものが社会的に認知されるようになってみたいなところを目指していきたいなと思っています．

飯尾：世界中のデータサイエンティストが圧倒的に前処理に時間を費やしているというのはものすごい時間の無駄というか，生産性を下げていますよね．そこは改善したいですね．

安部：カレンダーのデータとか，たとえば，Googleカレンダーから取ってくるだとか，気象庁のあの汚いデータを引っ張ってくるとか，表形式になっていないCSVとして扱いづらい政府データを使って，みんなされていると思います．日本の生産性を国が積極的に落としにいっているという，ひどい状態になっています．

飯尾：いや，みんな同じ思いなのだなと思って，共感しました（笑）．

里：牧山さんは何かありますか．

牧山：ヤフーはいろいろなサービスを展開しているのですが，社内のデータに関してはかなり整備されていて，どのサービスのデータがどこにあって誰にアクセス権を申請すればいいかなどが一覧で分かるようになっています．

しかし，たとえば，同じグループ会社のPayPayとかのデータに関しては，複雑な手続きを経ないとアクセスできないという問題があって，かなり苦労してデータを手に入れないといけないので，そこが障壁になっています．

それで，私が最近注目しているのは，Federated Learning（連合学習）と言って，それぞれの組織はプライベートなデータを公開せずに機械学習モデルを作るという手法なんですが，それに注目しています．たとえば，WeBankという中国のデジタル銀行が多重債務者を判定するのに機械学習を使っていて，それはFederated Learningを使って，ほかの銀行のデータと照らし合わせて多重債務者を判定するのですが，そのほかの銀行のデータというのはそのWeBankがもらっているわけではないのです．共同で多重債務者を判定する機械学習モデルを作っているという感じです．そういう仕組みに今ちょっと興味を持っています．

飯尾：それというのは何か業界団体みたいなものがあって，そこに加盟している各社が自分たちのデータは全部には公開しないけれども，共通するモデルを，自分たちが持っているデータで，それぞれが上手いこと協調させて，学習させて，1つのモデルを作るとか，そんなようなイメージなのですか．

牧山：業界団体があるかどうかはちょっと分からないのですが，イメージとしては，自分のデータだけで学習したときのモデルのウェイトだけを共有しましょうと．そしてグローバルなコンセンサスを持ったモデルを作りましょうというようなイメージで考えていただけると分かりやすいかなと．

高柳：今，Zoomのほうにリンク[☆1]を貼ったのですけれども，まあ，Googleを筆頭に，おっしゃったような個人情報のターゲティングをもうやめようよという技術として，牧山さんが言っていたFederated Learningは，今，送ったリンクだと，FLOCとかと略されてしまっているのですけれども，これ，Federated Learningの頭がFLですね，という技術が台頭してきている感じは，確かに印象は受けます．

飯尾：このFLOCというやつ，あまり評判がよくないみたいですよね（笑）．私も，よく知らないのですけれども，学生がFLOCは今後スタンダードになるんですかとか聞いてきて，ちょっとだけ調べたことがあって，なんか評判悪いらしいよというような話をした覚えはあります（笑）．

安部：そのFLOCの問題の1つとして，学習データの偏りみたいなところで，Federatedする相手先のその属性によってモデルというのが改善されていくので，偏った集団に対して学習してしまうと，公平ではないAIができあがるみたいなことは1つ問題として挙げられているかなと思っていますね．

企業の持っている集団で学習するという考え方でいうと，ビジネス上はまわりそうな気がしますけれども，たとえばGoogleみたいな大きな企業になってくると，それの偏り自体が問題視されるという話があるかもしれません．

牧山：公平性の問題は普通の機械学習でもあるような気がします．

飯尾：それこそ，技術的な話題というよりは，参加している企業の間の調整をどうするかみたいな，社会のルールがまだ未整備的な，そんな話題というふうに捉えられますよね，この問題は．

牧山：そうですね．それはあると思います．ちゃんとしたガイドラインを作らないとなかなか実現するのは，難しいかなと思っています．

人材育成

里：では，ちょっと話題を変えていきたいのですが，冒頭であったのは，人材育成だったり，あとニューノーマルなデータサイエンスだったりとか，そのあたりのお話ができればなと思っているのですが，人材育成で最近の状況や，思っていること，考えていることなど，何かあれば，皆さん，お願いいたします．

飯尾：若干ちょっと宣伝めいて恐縮なのですが，文部科学省が主導しているのでしたっけ，今，全国の大学でAIとか，データサイエンス教育をばんばんやれみたいな，そういうプログラムが動いていまして，そこに私どもも参加しています．認定[☆2]を受けるのを目標にしてどうのこうのなんていう話をしているのですけれども，全学でAIデータサイエンス教育をしようと．

それは，なかなか挑戦的なことで，中央大学というのは割と文系寄りの大学で，全学対象でAIだとか，データサイエンスのリテラシー教育をやると．そういうようなことをやっています．

それで，全学対象で，うち，8学部あるのですけれども，私どもは学際ということで理系と文系が融合したところで，理工学部以外はほぼ文系なのですが，法学部とか，商学部，経済ですね．なので，学生の8割方であるそういう学生に向けてAIとか，データサイエンスの教育をするんだと．もちろん，通りいっぺんのお話で終わってしまうかもしれない．ほとんどの学生はね，

今の社会はAIに支えられているのだよみたいな，そんなようなお話で終わってしまうのですけれども．そんな中で私どもが，それと理工学部にも情報工学科がありますので，そんなようなところで少し突っ込んだ教育をするのかなという，そういう建て付けでやっていますね．

これが，困ってしまうのは，AIデータサイエンスリテラシーレベルというのが，今，国からの，なんていうのですかね，プログラムの指針として出されて，さらにその上に，応用基礎の認定プログラムというのが，今，検討中らしいのですけれども，知り合いの先生もそこの委員会に入って，がちゃがちゃやっているというふうに聞きました．

石井：応用基礎のカリキュラムそのものは，2021年の3月に公開されています☆3．

2〜3月にパブコメ☆4があって，その直後にリリースされたと思います．

飯尾：応用基礎の認定プログラムというのが細かいところがまだ決まっていないので，どうのこうのなんていう話を1カ月前か，そのぐらいにやっていました．

それで，それが全学の50％以上が履修しないといけないとか縛りがあるらしくて，さすがにそれは難しいだろうと（笑）．うち，3万人からいますからね，そのうちの8割ぐらいが文系のはずなので，8割，9割ぐらい？　まあ，8割かそこらですよね．なので，さすがにちょっとそれを目指すのは厳しいだろうと，私は個人的には思っているのですが，志を高く持てなのか知りませんが，そんなようなところを最終的なゴールとして，このAIデータサイエンス全学プログラムにかかわっているというか，旗振り役の先生は頑張ってやっています，というご紹介でした．

これらについて質問があれば，できるだけお答えします．いかがでしょうか．

石井：企業とのかかわりとか，そういうのはありますか．

飯尾：ありがとうございます．

全学教育なのですけれども，バックエンドにAI・データサイエンスセンターという部署がありまして，私もメンバの1人なのですけれども，そこは企業とタイアップしていろいろ共同研究をやりましょうという話は，別途進んでいます．もちろんそこで得られた知見なんかも，その全学プログラムのほうにフィードバックしていくとか．

あとは，何だろうな，AIデータサイエンス総合という科目ですね，企業からの先生をお迎えして，それで最先端の話をしてもらうコマとかもあったと思いますね．やはり大学だけだと難しいですよね，企業と連携してやらないと．先ほどのオープンデータの話もありましたけれども，やはり大学だけだとリアルなデータというのはなかなか持っていないので，企業さんも，出せるデータと出せないデータ，当然，あると思いますけれども，出せる範囲でリアルなデータとか，事例とかを出していただくと，学生もいい刺激を受けますので，そんなようなことも入れていますね，プログラムの中に．

里：実際，企業の中での人材育成みたいな話を，高柳さんとか，何かありますでしょうか．

高柳：今だとすべてがリモート前提になってしまって，いろいろすごい教えにくいなというのはたぶん皆さん同意なところだと思うのですが，まずそこが1点あります．企業自体ではなく，企業と大学，また学生さんとかかわりあいの話ですと，私もちょうど学生さんに講義をさせてもらう機会があるのですが，学生に何を教えるといったときに，意外と学生さんは皆さん野望がないというか，何だろう，学生感がなくて，みなさんとても達観されているというか大人で（笑）我々がデータサイエンスをやるとこんなに楽しいですよみたいな話をしても，なんか別世界のような話に感じてしまうのがなんか問題だよねという話をしていて，じゃあ一体何を教えたらいいんだろうみたいな（笑），データサイエンスがあたかも自分とはまったく関係のないマンガかアニメの世界の話を聞いているような印象になってしまっているっぽいのでそこをなんとかしたいなと，そこのギャップをどう埋めようかなというのが，学生さんと企業の間でかかわっていこうと思ったときの課題ですね．

実務面での人材育成はやはりリモートが多くて，エンジニアリングの話だとOSとか，オープンソースとかの話もあったように，GitHubを使って開発するやり方をすれば，大体みんな非同期に開発できているから，まあ，そこはいいのだけれども，実際にデータを扱ってうんぬんとか，細かい個別の相談まわりの話になると急にやりにくさのギャップが出てきて，いろいろZoom的なツールを使って，いつでも気軽に声をかけられる状態にしつつ進めていますみたいな状態にもしているのですが，やはりまだ対面でやっていたときに比べるとギャップがあって，いかがなものかなと思っているというのが正直な感想ですね．

里：ほかの方，どうですかね．

牧山：AI人材と言ったときに，AIを研究できる人なのか，AIを実際にサービスに利用して運営させることができる人なのか，ちょっと定義が曖昧だなと思っていて，企業で必要になっている，需要が高いのは，研究する人よりは，プロジェクトを率いてサービスに機械学習を入れ込む人たちです．そこら辺がすごく今のところ人材が少ないので，ぜひ大学とかで教育していただけるのであれば非常に助かるなという感じですね．

石井：前処理をする人という意味ですか．

牧山：前処理も含むという感じですかね．実際の機械学習のプロジェクトというのは，割と工数が読めなかったり，どれぐらい成果が上がるのかというのが分からなかったりする．そこら辺を上手く進める，ちゃんと技術選定とかもして，どういうモデルを作るのか，どういう指標を見て成功を判断するか，サービスに入れるかどうか，入れることによってコストがペイできるのかとかを判断して，それで導入するかどうかを決めるとかも必要だし，いろいろな能力が必要になってきていて，そこら辺の人たちがいるとすごく助かるなという印象です．

飯尾：おっしゃることはよく分かるというか，企業さんはそういう人材を欲しがるだろうなというのは，私も昔，企業におりましたので（笑），分かるのですけれども，結構，今，求められた能力というのは，経験によるのではないかなという気がします．大学で，短い4年間しかない大学生活の中で，しかも1年生なんていうのはもう高校から上がってきて，本当に右も左も分からないと言ったら怒られますけれどもそういうようなので，実際に，AIとか，深い学習ができるというと，本当に3年，4年生になってからだと思います．そうすると2年間でどこまで経験を積めるかというと，なかなか難しいですよね．

難しいであろうということは分かっていつつ，先ほどの全学のあの教育の枠組みの中で，別途，私が今，持っているゼミとは別に，全学対象のそういうデータサイエンス系のゼミを来年から担当するのですけれども，そこで実際のリアルなデータを扱って何とか少しでもそういう経験を積ませるような教育ができればいいかなという挑戦はしようとは思っています．ただ，どこまでできるかなというとなかなか厳しい（笑）．厳しいリクエストです（笑）．

牧山：なるほど．少しでもAIプロジェクト推進の経験があれば，さらにそのAIの仕組みとかについて知識があれば，十分かなと思っています．

里：安部さんは外から人材育成を支援するということはあったりするのですか．

安部：今まさにとある企業に半年ほど毎週1時間ぐらい講義形式でやっていたりするのですが，それをやっていて思うところは，企業にいる方というのはいろんな立場の方がいて，当然，文系の方もいれば，理系の方もいらっしゃるというので，やはり知識レベルがばらばらかなというところはありますね．そういう意味では，先ほど飯尾先生もおっしゃっていたみたいな，勉学教養レベルでミニマムラインみたいなものを定義していただいて，それでそれを修めたぐらいの人材としての前提で話せるとまた話せることが違うことが多いのかなと思っています．

それこそ，さっき牧山さんがおっしゃっていた，AIを運用するとき，プロジェクトのまわし方，精度がどのくらいになるかといった読みづらいところもあるし，経験でしか積めないところではあります．なので，企業でしかできないことは企業にやらせるという前提で，大学における位置づけとしては，大学ならではの教養，基本的なところでいうと線形代数だとか，微積だとか，そういったところのミニマムの感覚みたいなものを育んでいただき企業に来ていただけると非常にやりやすいかなと思いますね．

まさにデータサイエンスというのは，サイエンスの領域に限らず，ビジネス領域も含めて，広範の領域の総合格闘技なので，あらゆるスキルを身に付けている人というのはいないと思いますが，とは言え，最低限のコミュニケーションできるレベルのというのはあるかなと思います．データサイエンティスト協会のスキルチェックリスト[☆5]をベースにしていて，その中に★1，2，3というのがあるのですけれども，その★1をベースに教育するだけでも，結構，苦労しているので，企業に来る人間が当たり前のレベルになっていると，企業の中でそのデータ活用人材というのがどういうふうに活躍していくかというのは，まさにビジネスに特化した形で進められるのでいいかなというふうに思いますね．

里：ありがとうございます．この話題に関して，何かほかにご意見などあれば，お願いいたします．

飯尾：今おっしゃった中で，線形代数は重要だよね，みたいな話は，まったくその通りで，今，プログラムで，ライブラリとかを使えば，簡単にAIみたいなものというのは作れるようになっているので，では特徴量は何にするのみたいな話をしたときに，これは笑い話で，もう卒業してしまったうちの学生ですけれども，なんか3つぐらいパラメータを入れていて，どうしても性能が上がらないんですよとか言っていて，何を特徴量に入れているのと言ったら，1つ目は，これこれで，もう1つ目は，これこれでと，似たような特徴量を入れているので，それ，いいのかなと思いながら聞いていたら，それで，3つ目は何かと言ったら，その平均値を入れています

と．おいおいおい，そんなね（笑），従属変数を入れてどうするよみたいな，笑い話がありましたけれども，そういうところをしっかり大学としては教えていかないといけないなとは常々思っています．

安部：最近，ディープラーニングをちょうど教える機会があって，その線形代数の知識が課程になっていれば，活性化関数みたいなところで非線形関数を食わせないと全部で見たときにもう線形結合の結合だから，線形結合になってしまうよねみたいなことを説明するだけでも，結構，苦労するので，そのくらいの常識感があるといいかなという感じがします．

用語として，エンジニアはエンジニアが使う用語があるし，サイエンスの人はサイエンスで使う用語があるし，ビジネスの人はビジネスが使う用語があります．そのあたりのベースラインみたいなものが業界の中で統率が取れるとたぶんお互いにコミュニケーションが楽になります．得意分野は得意な人がやればいいのですけれども，それをコミュニケーションするための最低限の教養があるといいかなというふうには思っています．

里：高柳さんのところとかは，研修というのはどんなふうにやられているのですか．

高柳：OJT（On the Jpb Training）一択という感じでしょうか．そもそも新卒採用を積極的にやっているわけではなくて，ほぼ中途採用だけなので，そういう課題はないです．逆に，私が副業でやっている会社でデーサイエンスに関する研修を提供するとなると，どちらかというと，さっき安部さんが言っていたような，★1個の内容を教える，みなさん大学生のときに意外と学んでいないのだけれども，今になって，やらなければなみたいな話とか，統計学や学問のありがた味が今になって分かってきたということが多いです．なので，細かい内容を教えるのではなくて，まさに非専門家に対してAIというのは，こう動いているのですよというのをふわっとまるっとお教えします．データを食わせて，学習させて，モデルができますみたいなのを，線形代数とか，確率の内容に関して似たようなたとえ話を用いて話すことがありますね．

みんな実際実務でやられている方々なので，彼らもそのAIを使ってデータ分析してという結果をお客さんに報告しなければならないので，ある程度分かった上で話さなければいけないのだけれども，すぽっと抜けている部分を，説明に困らない範囲でも，丁寧にお伝えしておくという感じですかね．

里：ありがとうございます．

参加者からの座談会での感想

里：では，最後に皆さんにひと言ずついただいて，この座談会，終わりにしたいと思います．

先ほどと逆順にしましょう．石井先生からお願いします．

石井：今回，座談会を企画させていただき，「ビッグデータのデータサイエンス」というテーマで，コロナ禍とか，地球温暖化とか，豪雨とか，最近まわりの変化がすごく激しくて，そんな中で自分がどうやって仕事を見つけていくかとか，今後どうやって食べていくかとか，そういう

ことを考えるいい機会かなと思っています．今日の企画を通して，皆さんの考える材料を提供することができたとしたら幸いだなと思っています．

里：ありがとうございます．

では，牧山さん，お願いします．

牧山：いろいろとお話が聞けて楽しかったです．ありがとうございます．特に大学教育でのAI人材の育成というのは非常に期待していて，今だと機械学習の研究をやっていましたみたいな人が企業に来るのですが，ミスマッチになってしまうこともあり，そういう人だけではなくて，サービスに興味を持っていてかつAIの基本的なところをちゃんと学んでいる人たちも来てくれると嬉しいなと思っています．オープンデータの話も面白かったです．みんな同じことに困っているのだなと（笑）．

里：ありがとうございます．

飯尾先生，お願いします．

飯尾：昔，90年代のころというのはITそれ自身の研究で飯が食えたのですよ．それがWindows 95みたいなのが，わあーっと広がってきて，それでITのコモディティ化と言われて，ITそのものの研究だと飯が食えなくなってしまったのですよね．そういうのを経験しているので，なんかそれのアナロジーではないのですけれども，最近は，そのAIのコモディティ化というのですかね，AIそのものの研究ではなくて，AIをどんなところに応用していくんだというところにどんどんシフトしているではないですか．だからそういう観点で見ていると，結構，いろんなことができて楽しいです（笑）．

最近，変わったところだと，高校生の異文化交流の教育の手伝いなんかもやっていて，本当に全然畑違いのことをITとか，AIとかの支援で参加するような，そんなプロジェクトにもかかわり始めているので，やることがどんどんどんどん広がっていって，今，本当に楽しいですね．

里：ありがとうございます．

では，安部さん，お願いします．

安部：そうですね，今の飯尾先生の話にあったみたいな，元々ITできるだけで食えるというところが，これがなくなって，今度はAIが似たような状況になっているというのはまさにビジネスサイドでも似たようなことが起こっているかなというのはすごく感じますね．そんな中で，今のこのリモートワークでしかできないような状況とかを考えると，どういう知識を持って，どういうところに応用していくかという，まさに応用力みたいなところが，人材の価値につながっているのかなという感じがしますね．

逆にいうと，それさえきちんと持っていれば，この状況であっても，生きていけるのではないかなという気はしていますね．とはいえ，私自身がそんな上手くやっているのかという話ではないのですけれども，そういうのを体現するためにも今作っているサービスを成功させて，皆さん

に使っていただきたいなという気はしていますね．データの活用みたいなところで，データの流通というのが大事だなというのを私は今思っているところなので，そういうところから何か，支援していけたらなというふうに思っています．

里：ありがとうございます．

では，高柳さん，お願いします．

高柳：私も牧山さんとかの話に近いのですが，非専門家がITを使い始めていて，それでデータ分析も意外とみんなやればできるじゃんという状況にはなってきている気がするので，もう少し大学と産業の連携を頑張る．まあ，インターンを増やしすぎると，今度は勉強する時間，学生としての学問をする時間がなくなるので，問題だなとは思うのですが，もう少し行き来の自由度を高めていければ，もうちょっと良くなっていくのかな，全体的にデータ活用とか，日本の産業とか，社会とかも全部そうですけれども，良くなっていくのかなというのを，今日話して，ふと思いましたという感想です．

里：はい．ありがとうございました．

安部：里さんから締めの言葉はないのですか，里さんの感想なりを（笑）．

里：いや，全然，考えていなかったです（笑）．

石井：ちょうど時間になりました．皆様，本日はお忙しい中座談会にご協力いただきありがとうございました．

脚注

☆1　https://xtech.nikkei.com/atcl/nxt/news/18/10691/
☆2　数理・データサイエンス・AI教育プログラム認定制度（リテラシーレベル）｜文部科学省
https://www.mext.go.jp/a_menu/koutou/suuri_datascience_ai/00002.htm
☆3　数理・データサイエンス・AI（応用基礎レベル）モデルカリキュラム～AI×データ活用の実践～ http://www.mi.u-tokyo.ac.jp/consortium/model_ouyoukiso.html
☆4　モデルカリキュラム （応用基礎レベル）案に関する意見募集 http://www.mi.u-tokyo.ac.jp/consortium/mc_ouyoukiso.html
☆5　データサイエンティスト スキルチェックリスト ver3.01
https://www.datascientist.or.jp/common/docs/skillcheck_ver3.00.pdf

グロッサリ

Glossary―グロッサリ―

ABM（Account-Based Marketing，アカウントベースドマーケティング）

ターゲット企業を定義し，ターゲット企業別に営業・マーケティング情報を集約し，ターゲット企業別に営業・マーケティング組織を再編成し，ターゲット企業からのLTV最大化を目指すマーケティングのこと．（高柳慎一）

API（Application Programming Interface）

ソフトウェアのコンポーネント同士が互いに連携するためのインターフェイスを定めたもの．APIには，データ構造の定義や関数の入出力の取り決め，エラーの定義などが含まれる．（村田賢太）

AUC（Area Under the Curve）

機械学習における評価指標の一種．0から1までの値をとり，値が1に近いほどそのモデルの識別能力が高いことを示す．識別能力がランダムであるとき，0.5となる．（高柳慎一）

B2B（Business To Business）

企業と企業の間での取引のこと．また似た意味を持つ用語としてB2C（Business to Customer）があり，これは企業と一般消費者の取引を表す．（高柳慎一）

CPU時間

プログラムがCPUを利用した時間．実行時間のうち入出力待ちなどでCPUが動いていない時間を除いた時間とほぼ等しい．複数のコアを同時に利用した場合は，それぞれのコアの稼働時間の合計となる．（村田賢太）

Foreign Function Interface（FFI）

異なるプログラミング言語で実装された関数を呼び出せるようにするための仕組み．（村田賢太）

GBDT（Grandient Boosting Decision Trees）

弱学習器（通常は決定木）を組み合わせて（アンサンブル）予測モデルを生成．予測誤差に対して逐次的に決定木を学習させていく点が特徴的である．（高柳慎一）

GLCM（Gray-Level Co-occurrence Matrix）

画像の輝度値の空間分布や局所的な関係性などから統計量を求めることによって画像のテキスチャの特徴を抽出する統計的解析法．（辰己賢一）

ICD-10（International Classification of Diseases, 10th Revision）

ICDとは，疾病，傷害及び死因の統計を国際比較するために WHOが定めた，疾病および関連保健問題の国際統計分類であり，ICD-10はその第10版である．各国で言語や呼び名が異なる場合でも，アルファベットと数字からなる統一されたコードで記載することで，各国間の統計比較を可能にしている．大阪府KDBデータの場合は，ICD-10は医療傷病名データの項目として，各被保険者の診療年月ごとに傷病名と併せて記載されている．（古徳純一）

JSON

データ記述言語の1つであり，JavaScriptのオブジェクト表記法のサブセットとなる構文で，型つきの構造化データを記述できる．（村田賢太）

NB（Naive Bayes classifier）

各特徴量に独立性を仮定し，ベイズの定理を用いて構築された分類器．特徴量が独立であると仮定されているため，効率的に学習が可能である．（高柳慎一）

Precision@N

クラス分類においてPositiveであろうと予測したN個のもののうち，実際にPositiveであったものの割合を表す．推薦問題においてしばしば用いられる．（高柳慎一）

RGB画像

Red（赤），Green（緑），Blue（青）の原色を使ってカラー表示される画像．（辰己賢一）

SIMD（Single Instruction Multiple Data）

1命令で複数のデータに対して同じ演算を行うこと．メモリ上の連続する2〜4つのデータに対して同時に命令を適用できるため，ベクトル同士の演算や，配列をスキャンする処理などが効率良く実装できる．（村田賢太）

拡張ライブラリ

たとえばRubyの場合，Rubyのみでは実現できないがほかの言語では実現できるような機能があったとする．そのような機能をRuby以外の言語で記述し，RubyのC APIを介してRubyから呼び出せるようにするライブラリが拡張ライブラリである．（村田賢太）

キャッシュメモリ

CPUとメインメモリの間に配置される高速化かつ小容量のメモリ．メインメモリ上のよく参照される領域をキャッシュメモリ上に配置することでCPUとメインメモリ間の直接のデータ伝送の頻度を抑え，CPUがデータを待つ時間を短くする効果がある．（村田賢太）

傾向スコア

観察研究において因果推論などを行う際に，ある共変量（説明変数）を持つ個体が介入を受ける確率として傾向スコアと呼ばれる量を定義すると，与えられた傾向スコアのもとで結果変数が介入と条件付き独立となることを示せる．そのため 傾向スコアを用いることで，無作為化比較試験をやりにくい状況下でもそれに近い状態の解析が可能になる．たとえば逆確率重み付き推定を利用すれば，介入前後での結果の変化の期待値（平均処置効果）を小さいバイアスで推定できる．（古徳純一）

構文解析

文章やプログラムコードなどを表現している記号列を読んで文法に従った構造を特定し，構文木と呼ばれる木構造を構築する処理．（村田賢太）

シリアライズ

プログラムが扱っているデータを，保存したり通信相手に送ったりするためにバイト列に変換すること．（村田賢太）

正則化

解が一意に定まらない不良設定問題を解決したり過学習を防ぐために，モデルの複雑さに対して罰則を科す手法の総称．機械学習においてはパラメータのノルムの大きさに罰則を付加することが一般的である．（高柳慎一）

テクスチャ情報

形状の情報ではない，特定の物体や物質，素材などの表面の質感などを表現するために使われる情報．（辰巳賢一）

デシリアライズ

プログラムがデータを読み出したり通信相手から受信したりした際に，バイト列からデータを復元すること．（村田賢太）

データフレーム

2次元の表形式のデータ構造であり，列と行のそれぞれにラベルを付けられ，個々の列は異なる型の値を持てるもの．R言語では組込みのデータ型として提供されている．Pythonではpandasライブラリによって実装されている．（村田賢太）

トランザクション処理

データベースのトランザクション処理を用いてデータの一貫性を保ちながら，小さいサイズのデータに対する更新や問合せを行うこと．（村田賢太）

バインディング

たとえばRubyの場合，Rubyで書かれたライブラリは直接利用できるが，C言語で書かれたライブラリは直接利用できない．このように，ほかの言語で実装されていて直接利用できないライブラリを利用するために必要なものがバインディングである．（村田賢太）

分析的データ処理

蓄積された大量のデータを多次元的に見て集計処理や統計処理などの複雑な処理を行うこと．（村田賢太）

マングル

C++の変数名と関数名を固有の名前に符号化して，リンカが識別できるようにすること．C++では名前空間とクラスがあり，また関数は同名で異なる引数を持つものを多重定義できるため，このような仕組みが必要となる．（村田賢太）

列指向

2次元データをメモリ上に配置するときの値の並べ方の流儀の1つで，列の方向に値を連続配置する場合を列指向という．対して，行の方向に値を連続配置する場合を行指向という．（村田賢太）