

**「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化」研究領域
領域活動・評価報告書
－2018年度終了研究課題－**

研究総括 喜連川 優
副研究総括 柴山 悦哉

1. 研究領域の概要

ICT の社会浸透や、実世界から情報収集するセンサーや計測・観測機器の高度化と普及に伴い、様々な分野で得られるデータは指数関数的に増大し、多様化し続けています。これらのビッグデータの高度な統合利活用により、新しい科学的発見による知的価値の創造や、それらの知識の発展による社会的・経済的価値の創造やサービスの向上・最適化などにつながる科学技術イノベーションが期待されています。

本研究領域では、ビッグデータの複数ドメインに共通する本質的課題を解決し、様々な分野のビッグデータの統合解析を可能にする次世代基盤技術の創出・高度化・体系化を目指します。

具体的には、大規模データを圧縮・転送・保管する大規模管理システムの安定的運用技術や、多種多様な情報を横断して検索・比較・可視化して真に必要な知識を効率的に取り出す技術、これらを可能にする数理的手法やアルゴリズムなどの開発を推進します。これらの研究の推進にあたり、ビッグデータから社会における価値創造に至るシステム全体の設計を視野に入れ、ICT 以外の分野との積極的な連携・融合によって社会受容性の高い次世代共通基盤技術の創出・高度化・体系化に取り組みます。

また、本研究領域では、関連領域の「科学的発見・社会的課題解決に向けた各分野のビッグデータ利活用推進のための次世代アプリケーション技術の創出・高度化」で得られる次世代アプリケーション基盤技術やデータを共有・活用するなどの連携を推進します。

2. 事後評価対象の研究課題・研究者名

件数： 6件

※研究課題名、研究者名は別紙一覧表参照

3. 事前評価の選考方針

選考の基本的な考えは下記の通り。

- 1) 選考は、「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化」領域に設けた選考委員(領域アドバイザー)13名の協力を得て、研究総括が行う。
- 2) 選考方法は、書類選考、面接選考及び総合選考とする。
- 3) 選考に当たっては、さきがけ共通の選考基準(URL: <http://www.jst.go.jp/pr/info/info1128/index.html>)の他、以下の点を重視した。本研究領域では、今後大きく展開することが予想されるビッグデータ時代の基盤的な研究開発を対象とした。基盤技術とは言え、応用を想定しない技術開発は、その評価軸の設定が必ずしも容易ではないため、ある程度の応用を想定した提案を重視した。また、様々な応用分野におけるビッグデータの新規取得とそのデータの他研究者への提供を目指すさきがけ研究の提案(データさきがけ)も募集対象とし、収集・提供するデータの有益性や希少性を重視した。

4. 事前評価の選考の経緯

一応募課題につき領域アドバイザーの内 12 名が書類審査し、書類選考会議において面接選考の対象者を選考した。続いて、面接選考および総合選考により、採用候補課題を選定した。

選考	書類選考	面接選考	採択数
対象数	35件	14件	7件

備考:

- 1) 2015年度採択課題のうち、以下は今年度事後評価対象としない。



- ・秋葉 拓哉 研究者
2016 年度に研究を中止したため。

5. 研究実施期間

2015 年 10 月～2019 年 3 月

6. 領域の活動状況

1) CREST/さきがけ ビッグデータ合同領域会議:6回

2) さきがけ合宿領域会議:2回

3) 公開シンポジウム

・2016 年 2 月 4 日 情報処理学会ソフトウェアジャパン 2016 一橋講堂

4) JST/NSF 合同国際シンポジウム/NSF PI ミーティング/DATAIA ワークショップ

・2016 年 4 月 20 日～22 日 NSF PI ミーティング 米国 ワシントン DC ジョージタウン大学

・2016 年 5 月 11 日～12 日 合同国際シンポジウム アキバホール & 伊藤国際学術研究センター

・2016 年 11 月 28 日～30 日 合同国際シンポジウム ベルサール六本木

・2017 年 3 月 15 日～17 日 NSF PI ミーティング 米国 ワシントン DC バージニア工科大学

・2017 年 12 月 20 日 JST-NSF 連携、AIP ネットワークラボ合同シンポジウム コクヨホール

・2018 年 7 月 10 日～12 日 JST-DATAIA 情報科学と AI ワークショップ 仏国 パリ Amphithéâtre Poincaré

・2019 年 3 月 11 日 JST-NSF-DATAIA 連携、AIP ネットワークラボ合同シンポジウム 京都国際会館

5) 研究総括(または副研究総括)、領域担当、事務参事の研究実施場所訪問(サイトビジット):全研究者の研究室訪問と研究者上司への挨拶を実施した。

・小野木 章雄 研究者: 東京大学 岩田洋佳准教授 訪問 2016/1/7

・酒向 重行 研究者: 東京大学 土居守教授 訪問 2016/7/27

・佐藤 彰洋 研究者: 京都大学 山下信雄専攻長・教授 訪問 2016/6/29

・島田 敬士 研究者: 九州大学 緒方広明教授 訪問 2016/7/20

・柳澤 琢史 研究者: 大阪大学 平田雅之寄付研究部門教授 訪問 2016/7/26

・山田 拓司 研究者: 東京工業大学 伊藤武彦教授 訪問 2016/7/1

7. 事後評価の手続き

年2回程度開催した領域会議、公開シンポジウムなどにおいて、研究者が口頭発表やポスター展示を実施し、その場で研究総括、アドバイザーなどから適時・適切なアドバイスを行った。また、半期毎に研究者が作成した研究報告書を研究総括、副総括が確認し、アドバイスやコメントを研究者にフィードバックした。最終年度には、事後評価会を実施し、研究者からの成果発表と総括、アドバイザーによる議論を行った。最後に、研究総括、副研究総括が研究者の終了研究報告書およびこれまでの進捗状況を総合的に判断し、最終的な事後評価結果を作成した。

(事後評価の流れ)

2018 年 9 月 25 日～26 日 事後評価会実施

2019 年 1 月 研究報告書提出

2019 年 2 月 研究総括による事後評価

2019 年 3 月 被評価者への結果通知

8. 事後評価項目

(1) 研究課題等の研究目的の達成状況

(2) 研究実施体制及び研究費執行状況

(3) 研究成果の科学技術及び社会・経済への波及効果(今後の見込みを含む)

(4) 研究の進め方(他研究者との連携、国内外研究者・産業界等との連携、研究費執行状況など)

9. 評価結果



総論

「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化」領域を今回終了した6名の研究者は、各々が主たる応用分野を想定し、ビッグデータ統合利活用という大目標に向けた挑戦を行なった。小野木氏は農業、酒向氏は天文学、島田氏は教育、山田氏は健康・医療での、そして佐藤氏と柳澤氏はそれぞれ位置情報と脳波情報でデータの関連付けを行う研究を実施した。小野木氏、柳澤氏、山田氏の研究は、ビッグデータを収集し、第三者が利用可能な形に整備することを目的とする本領域独自の「データさきがけ」として実施した。

中でも島田氏は、九州大学の開講科目において学習行動データを取得し、それらを解析して授業の改善等に役立てる研究を実施した。社会的に大きなインパクトが期待できる研究である。「データさきがけ」の成果として、山田氏は、多数の大腸がん患者のものを含む2,800名近くの腸内環境データを取得し、この分野で世界最大規模のデータベースを構築した。特に大腸がん研究での利用が期待される。同様に柳澤氏は、頭蓋内に留置した電極から、400人日分を超える高精度な皮質脳波を取得し、ライフログデータとの関連付けを行った。人間の活動と脳波の関係に関する研究での利用が期待される。

本領域は、ビッグデータ2領域のCRESTと合同領域会議を開催し、米国NSFやAIPネットワークラボの他領域と連携したシンポジウム等も開催している。また、さきがけ研究者の合宿型領域会議では企業見学なども行なっている。このような場を活用し、さきがけ研究者が異分野の研究者と長く続く関係を形成することも、さきがけ研究の成果と言えるであろう。

1. 小野木 章雄 研究者「膨大なレガシー栽培データを蘇生する(データさきがけ)」

農業は、ビッグデータ利活用の応用分野として近年注目を集めている。しかし、作物の生育には時間がかかり、利活用に十分な質と量のデータを収集するには年単位の時間を要することが多い。一方、我が国には、公設の試験場等において数十年間にわたって品種改良を試みた実績があり、その際に取得された栽培データが活用されず眠っている。これらのデータをアクセス可能とし、さらに品種改良や栽培管理に関連するデータと統合することには大きな意義がある。

本研究は「データさきがけ」として実施した。ダイズを対象に、全国460箇所以上の試験地で1961年以降に行われた栽培試験のデータを収集し、用語のばらつきが大きく紙媒体の割合が高いデータを利用可能な形に整備した。そして、別に集めた2,200品種以上のダイズのゲノムデータおよび試験場近辺の気象データとの関連付けを行った。これは世界でも稀な規模のデータベースであり、ダイズ品種改良関係者による利用が始まりつつある。

貴重なデータの収集・整備を行うことができた。今後は、整備したデータの利活用が進むことを期待したい。また、本研究で得られた知見が、ダイズ以外の作物に対するビッグデータ統合利活用にも貢献することを期待したい。

2. 酒向 重行 研究者「タイムドメイン宇宙観測用動画データの高速逐次処理法の開発」

大型望遠鏡を用いた天文観測では、従来、暗い天体をとらえるために長時間露光での撮影が一般的であったが、この方法では一過性の変動現象をとらえることが難しい。静止画ではなく動画を撮影することで、今まで観測から漏れていた現象をとらえられる可能性がある。

本研究では、天文台用広視野高速CMOSカメラ(Tomo-e Gozen)のデータ取得部とデータ処理部の開発を行ない、天文動画データを対象に、短時間変光現象や高速移動天体を検出する方式、データ解析に適した圧縮方式などを開発した。また、200CPU、900TB規模のオンサイト解析用システムの構築と運用を行なった。この規模の天文動画の取得と解析は世界初のものである。そして、取得したデータと開発した技術を用いて、スカイサーベイの試験実施、地球接近天体の連続観測、連星ブラックホールの高速観測などを実施し、学術的にも優れた成果もあげている。さらに、地球に近い宇宙での宇宙機やスペースデブリなどの観測でも成果をあげている。

宇宙で発生する突発的現象の観測について、世界をリードする研究・開発を実施することができた。この分野の研究はまだ始まったばかりであり、今後、大きな発見につながることを期待したい。

3. 佐藤 彰洋 研究者「グローバル・システムの持続可能性評価基盤に関する研究」

ビッグデータを利活用して複雑でグローバルな現象の分析を行うためには、多様なデータをいかに統合して利用するかが重要なポイントとなる。実社会の様々な関係の多くが地理的近傍に主たる影響を及ぼすことから、位置情報を用いたデータの関連付けは、データの構造化のための有望な手法である。

本研究では、世界全体を地理的にメッシュ分割して各メッシュに番号付けを行う世界メッシュコードを、我が国の地域メッシュコード JIS X0410 を拡張して定義し、様々な統計情報を関連付ける世界メッシュ統計の方式を提案した。そして、様々な情報源から得た統計データに自動的または半自動的にメッシュコードを付加し、さらに可視化等を行うための方式とシステムを開発した。公式で更新頻度の低い政府統計と更新頻度が高い民間データの統合利用も可能としている。ユースケースや新規データの開発を支えるコミュニティの形成にも務め、500 種類以上のメッシュ統計を構築した。

研究期間中に形成したコミュニティは、将来につながる貴重な財産である。今後も応用分野関係者との交流を深め、メッシュ統計の種類を増やしていくとともに、利便性の向上や普及活動を行うことを期待したい。

4. 島田 敬士 研究者「時空間粒度の異なる教育ビッグデータの非同期ストリーム処理基盤の構築」

社会的価値を創出するために知恵や知識が果たす役割は、近年次第に大きくなっている。それにともない、社会を支える基盤としての教育の重要性も増大しつつある。ビッグデータを利活用したエビデンスに基づく教育は、教育の質の向上という社会的要請に応えるための有望な考え方である。

本研究では、九州大学で実際に行われている授業を対象として、教育行動データの取得と様々な解析手法の提案・適用を行なった。教育行動データの取得には、学習管理システムとデジタル教科書システムを用いた。そして、取得したデータの解析に基づき、予習状況の集計、講義進行についていけない学生の検出、理解度に応じた教材の要約、成績予測などを行う研究を実施した。その結果、一部では予習率や成績の向上が見られており、学術的のみならず実用的な成果も得られている。

一連の研究は、教育データ利活用の事例として先進的なものである。一方、この分野には、まだ大きな発展の余地が残されている。今後、取得できるデータの種類や量を増やし、解析手法の高度化を図るとともに、他大学等にも知見を広げ、教育の質の向上に貢献することを期待したい。

5. 柳澤 琢史 研究者「思考・行動を予想する脳ビックデータ(データさきがけ)」

統合利活用による多様な応用の可能性があるビッグデータとして、脳の活動データをあげることができる。データの取得・解析方法にブレークスルーをもたらすことで、ブレイン・マシン・インタフェースの高度化、精神疾患の診断・治療などへの発展が期待できる。

本研究は「データさきがけ」として実施した。脳神経外科の治療目的で頭蓋内に電極を留置した約 50 名を対象に、のべ 400 日以上にわたる高精度な皮質脳波を取得し、視聴覚内容等のライフログのデータと関連付けた。データの作成過程では、各人の脳活動データを標準脳にマップすることで個人情報削除する方式、視聴覚データの意味ベクトルへの自動変換方式などの研究を実施した。海外を含む関連分野の研究者に作成したデータを提供し、共同研究も行なっている。また、本研究で開発した計測システムと解析パイプラインを国内の 3 施設に展開している。

貴重なデータを取得・整理し、学術的にも優れた成果をあげることができた。今後は、他の施設を含めデータの収集と加工を進め、データ利活用の促進を図ることを期待したい。また、非侵襲の計測しかできない応用分野の研究開発にも貢献することを期待したい。

6. 山田 拓司 研究者「ヒト腸内環境ビッグデータ(データさきがけ)」

人間の体は複雑なシステムであり、原因が簡単には見つからない疾病も少なくない。そのような疾病の予防・診断・治療のために、人間の健康に関連した様々なデータを収集し、ビッグデータ統合解析技術を活用することが期待される。

本研究は「データさきがけ」として実施した。当初予定していた 500 症例を大幅に上回る腸内細菌叢のメタゲノムデータ、メタボロームデータ、内視鏡所見、食生活習慣の疫学データを収集した。メタゲノムデータについては 1,000 症例以上、内視鏡所見と疫学データについては 2,800 症例近くに及んでいる。大腸がん患者のデータを 2,000 症例近く含んでいることから、大腸がんに関連した研究での利用が期待される。国際的なデータ共有も進みつつある。また、病院、企業、研究機関をメンバーとするコンソーシアムを立ち上げ、コミュニティ形成にも務めている。

この分野において、国際的にも最大級のデータベースを構築することができた。今後は、整備したデータの利活用を進め、特に大腸がんの予防・診断・治療に関連した研究やコンソーシアム活動を通じ、社会に貢

献することを期待したい。

10. 評価者

研究総括 喜連川 優 国立情報学研究所 所長／東京大学生産技術研究所 教授
副研究総括 柴山 悦哉 東京大学情報基盤センター 教授

領域アドバイザー(五十音順。所属、役職は平成31年3月末現在)

荒川 薫 明治大学総合数理学部 学部長・教授
石塚 満 情報・システム研究機構国立情報学研究所コグニティブ・イノベーションセンター センター長／東京大学 名誉教授
上田 修功 NTT コミュニケーション科学基礎研究所 上田特別研究室長(NTT フェロー)／機械学習・データ科学センター 代表／理化学研究所革新知能統合研究センター 副センター長
田中 英彦 岩崎学園 理事／情報セキュリティ大学院大学 名誉教授／東京大学 名誉教授
辻井 潤一 産業技術総合研究所人工知能研究センター 研究センター長／英国 マンチェスター大学 教授
徳田 英幸 情報通信研究機構 理事長
徳山 豪 東北大学大学院情報科学研究科 教授
東野 輝夫 大阪大学大学院情報科学研究科 教授
堀 浩一 東京大学大学院工学研究科 教授 (2017年12月より参画)
北川 博之 筑波大学計算科学研究センター 教授
山西 健司 東京大学大学院情報理工学系研究科 教授
Calton Pu Professor, Georgia Institute of Technology
Nozha Boujemaa Director of Research, Inria (2016年11月より参画)

岩野 和生 東京工業大学大学院イノベーションマネジメント研究科 客員教授
(2013年10月～2015年10月まで参画:所属、役職は退任日現在)
室田 一雄 東京大学大学院情報理工学系研究科 教授
(2013年10月～2015年9月まで参画:所属、役職は退任日現在)

(参考)

件数はいずれも、2019年1月末現在。

(1) 外部発表件数

	国内	国際	計
論文	13	40	53
口頭	193	65	258
その他	29	1	30
合計	235	106	341

(2) 特許出願件数

国内	国際	計
5	1	6

(3) 受賞等

・島田 敬士

①第1回 IMS Japan 賞 最優秀賞 2016/10/26

②第13回日本 e-Learning 大賞 教育ビッグデータ特別部門賞 2017/10/25

・柳澤 琢史

- ①バイオインダストリー奨励賞, 2017/10/11, BioJapan2017, バイオインダストリー協会
- ②BCI award, BCI conference, 2017/9/21
- ③中谷賞奨励賞 中谷医工計測技術振興財団、2017/2/24

(4)招待講演

国際 7件

国内 16件

別紙

「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化」領域
事後評価実施 研究課題名および研究者氏名

研究者氏名 (参加形態)	研究課題名 (研究実施場所)	現職(2019年3月末現在) (応募時所属)	研究費 (百万円)
小野木 章雄 (兼任)	膨大なレガシー栽培データを蘇生する (データさきがけ) (農業・食品産業技術総合研究機構)	農業・食品産業技術総合研究機構 次世代作物開発研究センター 主任研究員 (東京大学大学院農学生命科学研究科 農学研究員)	30
酒向 重行 (兼任)	タイムドメイン宇宙観測用動画データの 高速逐次処理法の開発 (東京大学)	東京大学大学院理学系研究科 助教 (同上)	38
佐藤 彰洋 (兼任)	グローバル・システムの持続可能性評価 基盤に関する研究 (京都大学)	京都大学大学院情報学研究科 特定 准教授 (京都大学大学院情報学研究科 助教)	35
島田 敬士 (兼任)	時空間粒度の異なる教育ビッグデータの 非同期ストリーム処理基盤の構築 (九州大学)	九州大学大学院システム情報科学 研究院 准教授 (九州大学基幹教育院 准教授)	40
柳澤 琢史 (兼任)	思考・行動を予想する脳ビッグデータの (データさきがけ) (大阪大学)	大阪大学高等共創研究院 教授 (大阪大学大学院医学系研究科 助教)	33
山田 拓司 (兼任)	ヒト腸内環境ビッグデータ(データさきがけ) (東京工業大学)	東京工業大学生命理工学院 准教授 (東京工業大学生命理工学研究科 講師)	36

研究報告書

「膨大なレガシー栽培データを蘇生する(データさきがけ)」

研究タイプ: 通常型

研究期間: 2015年10月～2019年3月

研究者: 小野木 章雄

1. 研究のねらい

現在人口増加や気候変動により世界的に食料問題が深刻化しており、食料増産や生産の効率化及び安定化が危急の課題となっている。この課題解決のために、気象、圃場環境、栽培管理及びゲノムに関するビッグデータを用いたデータセントリックな品種改良・栽培管理法に大きな期待が寄せられている。特に全ゲノム情報を用いたゲノミックセレクション法(genomic selection、注)は新たな品種改良手法として導入が期待される。しかしこれら技術の実現には大きな障壁がある。それは品種改良や栽培管理の直接のターゲットとなる作物の性質、つまり収量や形態、収穫期、耐病性などに関する栽培データが不足していることである。栽培データが他のデータに比べ不足する要因は、そのデータ量が年(時間)により制約を受けるためである。作物の栽培は基本的に年に1回しか行うことができず、さらに栽培結果は同じ場所であっても年により変動するため、ある品種がその場所で平均的にどのような性質を示すかを知るためには多年に渡る栽培が必要となる。

一方で、近代的な品種改良は数十年にわたり全国の公設試験場などで様々な作物について行われており、その過程で膨大な栽培データが蓄積されている。例えばダイズ品種改良の栽培試験は一部試験場では戦前から開始され、戦後始まった国主導の事業では全国数か所にある主要育成地とほぼ全国の地方公設試験場で行われてきた。しかしこの過程で得られた膨大な栽培データは各育成地や試験地に死蔵されており、レガシーデータとなっていた。

そこで本課題では、こういった過去のダイズ品種改良の過程で収集され、その後死蔵されている栽培データを「蘇生」し、データセントリックな品種改良・栽培管理法を迅速に構築することを目標とした。ここでいう「蘇生」とは栽培データの収集及び整理と、それを活用するために必要な情報、つまりゲノムや気象データ、との関連付け、さらに活用に向けた統計学的手法や枠組みを提供することを指す。この研究により、作物の増産や生産の高効率化が迅速となり、日本及び世界の食糧問題解決に向けた大きな貢献となることが期待される。

(注)ゲノムデータと作物の性質を統計学的或いは機械学習的手法で関連付け、ゲノムデータのみから作物の性質を予測し、栽培することなく優良品種を選抜する手法。品種改良の高速化や高効率化への貢献が期待される。

2. 研究成果

(1) 概要

栽培データは主要ダイズ育成地、すなわち農業・食品産業技術総合研究機構(農研機構)の4センター(東北農業研究センター、次世代作物開発研究センター、西日本農業研究センター、九州沖縄農業研究センター)、北海道立総合研究機構の2試験場(中央農業試験場、十

勝農業試験場)、及び長野県野菜花き試験場を中心に収集した。アメダスにおける気象観測が1961年から開始されているため収集の対象を1961年以降とした。品種改良における栽培試験では、これら育成地から全国の試験地へ種子が配布され試験されるため、実際にデータに含まれる試験地は460箇所以上になる。収集した栽培データは主に以下の3種類のデータから成る。(1)ある品種をある年ある試験地で栽培した結果示した性質を記録したデータ(以下試験結果と呼ぶ)、(2)施肥や播種期など栽培条件を記録したデータ(以下耕種概要)、及び(3)試験時の生育状況を記録したデータ(以下生育概要)である。これら栽培データはデジタル化・統合を経て、農研機構内のサーバーにデータベースとして格納した。

さらに収集した栽培データを気象データと関連付けるために、過去約50年にわたる全国の公設試験場の沿革を調査して試験場所を特定し、最寄りのアメダス観測所との関連付けを行った。またゲノムデータについては育成地やジーンバンクから2,208品種(系統)について種子提供を受け、DNA抽出の後ゲノムデータを取得した。

全国規模かつ多年にわたるヒストリカルな栽培データと、気象及びゲノムデータをどのように活用するかという問題は必ずしも明らかではなく、国際的にも同様の取り組みは乏しい。そのため本課題ではまず気象及びゲノムデータの利用について、双方を同時に利用し作物の生育を高精度に予測する統計学的手法の提案とパッケージの開発を行った。またヒストリカル栽培データは不均衡な多変量データであるが、有用遺伝子の検出手法とゲノムから性質を予測する統計学的手法について不均衡多変量データにおける性能の検討を行った。さらに後者の知見に基づきゲノミックセレクション法で用いる統計学的予測モデルの構築とそのためのフローを開発し、サイズだけでなく他の作物に対しても適用できるような枠組みを構築した。

(2) 詳細

【データ収集と整理】

栽培データはどの育成地においても概ね2000年前後までは試験成績書と呼ばれる紙媒体で、その後はエクセルファイルなどのデジタル媒体で保存されていた。そのためまず専門業者委託による紙媒体のデジタル化を行った。続いてデータのフォーマットが育成地により様々であり時代による変遷もあること、さらに例えば収量であれば、「アールあたりの子実重 kg」や「子実重 kg/a」など、性質の名称も一貫性がなかったために、フォーマット及び性質名の統一を行った。性質名については性質名に加え、測定時期、測定単位をアンダーバーで結合する簡易的なオントロジーを用いて統一した。以上の作業は栽培データを構成する3つのデータ(試験結果、耕種概要、生育概要)について行い、それぞれをテーブルとして構成した後、これらテーブル間での関連付けを行った。現在確定しているテーブルサイズは83,886行×606列(試験結果)、10,376行×95列(耕種概要)、5987行×13列(生育概要)となる。以上のデータを格納するために農研機構高度解析センターにデータベース「作物育種データを管理するDBとWebシステム」を構築し、現在サイズ品種改良関係者に順次公開している。以上の取り組み及び成果は研究成果(3)の1)及び2)で報告した。

【気象及びゲノムデータとの関連付け】

公設試験場は移転することがあり、同じ名前でも過去と現在で所在地が異なることも珍しくな

い。また所在地が同じでも組織改編により名称が変更することも多い。所在地の明確化はその後の統計解析や気象データとの関連付けにおいて重要であることから、沖縄県を除いた関係する 46 都道府県の公設試験場についてその沿革を調査し、試験場の所在地の変遷を可能な限り特定した。これにより各試験場について最寄りのアメダスから気象データを得ることや、公開されている気象及び土壌データベースから情報を得ることも可能となった。ゲノム情報については 2,208 品種(系統)の種子を発芽させ本葉から DNA を抽出した。ゲノムデータは全ゲノムシーケンス法、アレイ(Thermo Fisher Scientific 社)法、または RAD-seq 法により得た。これらの手法は得られるゲノムデータの量に差があり、全ゲノムシーケンス法では全塩基配列(約 1.1G 塩基)、アレイ法では約 150K 塩基、RAD-seq 法では約 1K 塩基分のデータが得られる。これらの手法には品種の重要度に合わせてそれぞれ 9、575、及び 2,208 品種(全品種)を供した。これらゲノムデータは試験結果と関連付けることが可能であり、収集した栽培データを用いたゲノミックセレクション法や有用遺伝子の検出などに利用することが可能となった。

【データ活用手法に関する研究】

ヒストリカル栽培データと気象及びゲノムデータを今後の農業や品種改良にどのように役立てればよいかという問題は未だ道筋がついておらず、国際的に見ても取り組みは乏しい。ここでは特に品種改良への利用で重要となる以下の4点について研究を行った。

1) 気象及びゲノムデータ双方を考慮した作物生育予測手法

安定した収量及び品質を得るには、その土地の気候に作物が適応していることが必須であり、例えば開花や成熟がその気候の下での望ましい時期に起こる必要がある。また管理面から播種や収穫時期が限定されていることも多く、任意の気候の下でどのような生育をするかデザイン可能であると品種改良が効率化される。本課題担当者は以前に植物生理学的生育モデルとゲノム情報を組み合わせることで、高精度に開花予測が可能なことをイネにおいて例証したが(文献参照)、同様のモデルをイネの開花だけではなくダイズを含めた様々な作物の生育において構築可能な汎用 R パッケージ「GenomeBasedModel」を開発した(研究成果(3)の3))。このパッケージではユーザーが任意の生育モデルとゲノム情報を引数として与えると、その生育モデルのパラメータとゲノム情報を自動的に関連付け、任意の気候の下での生育をベイズ統計の枠組みで予測する。同パッケージは現在 Github (<https://github.com/Onogi/GenomeBasedModel>) で公開中である。

2) 多変量混合効果モデルに基づくゲノムワイド関連解析の有効性について

ゲノムワイド関連解析は広大なゲノム領域から標的の性質を調節する遺伝子を検出する解析を指すが、混合効果モデルに基づく検定は現在その解析の中心的手法である。一方ヒストリカル栽培データは性質や試験地などの面で多変量データであるため、通常用いられる単変量モデルではなく、多変量モデルを用いるという選択肢も考えられる。しかしながらゲノムワイド関連解析において単変量と多変量モデルの優劣は明らかではなかった。そのためシミュレーションを用いて様々な条件下で比較を行い、3つのパラメータ、つまり変量間の相関、遺伝子効果の相対的な大きさ、データ欠損率、によりその優劣が変化することを見出した(図1及

び研究成果(3)の4)。

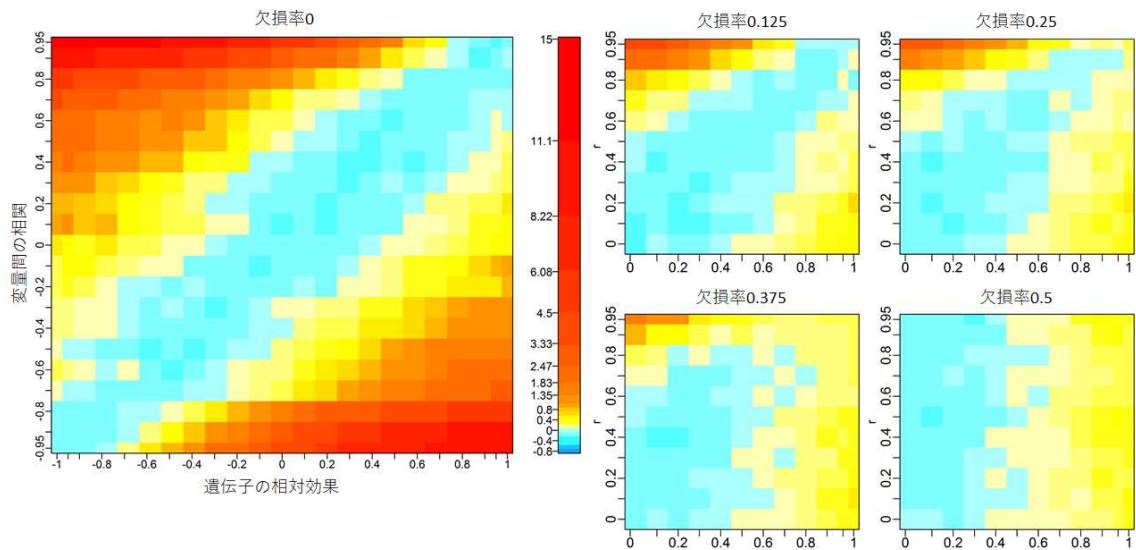


図1 多変量(変量数は2)と単変量モデルでの検定で得られた $-\log_{10}P$ 値の差。赤は多変量で青は単変量で検出力が高かったことを示す。縦軸が変量間の相関、横軸が遺伝子効果の相対的な大きさ。欠損率が0の場合は変量間相関と遺伝子相対効果は -1 から 1 の値を、欠損率が0以上の場合は0から1の値の結果を図示。

3) 多変量混合効果モデルに基づくゲノミックセレクション法の有効性について

混合効果モデルはゲノムから性質を予測するゲノミックセレクション法においても中心的な手法であるが、2)のゲノムワイド関連解析同様、多変量と単変量モデルの予測能力の優劣については十分に明らかにされていなかった。そのためシミュレーションにより両者の比較を詳細に行い、遺伝的要因により生じる相関(遺伝相関)と、それ以外の要因により生じる相関(残差相関)の差が変量間で大きいほど多変量混合効果モデルが優れていることを見出した(図2)。

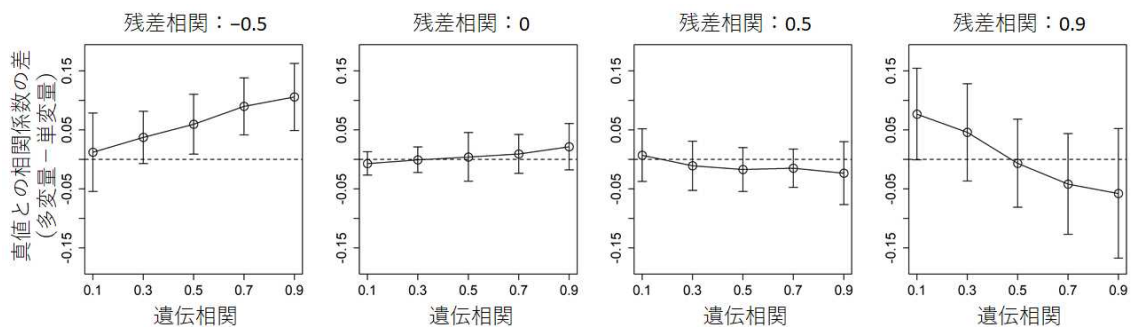


図2 多変量(変量数は2)と単変量による予測の差。予測能力は真値とのピアソン相関係数で測定し、多変量から単変量による相関係数を引いた値を図示した。100回のシミュレーションの平均値(バーは標準偏差)。

4) ヒストリカル栽培データからのゲノミックセレクション法のための統計学的予測モデル構築

前述の3)からヒストリカル栽培データを多変量としてモデリングすることは必ずしも良い予測

をもたらさないことが明らかとなった。そのため各試験地の各性質を単変量として扱うモデルを基本とし、3)で明らかとなった条件に合致するものだけ統合し多変量として扱う方法が適していることが示唆された。試験地によっては様々な栽培条件(施肥、栽植密度、播種期など)が試みられているため、栽培条件と遺伝の相互作用を考慮しながら主要 15 地域の 18 試験地において予測モデルの構築を行った(図3及び研究成果(3)の5)。図3の例から、同じ品種であっても地域及び播種日により期待される収量が異なることがわかる。これらモデルを品種候補の選抜や交配計画策定に用いることにより今後各試験地での品種改良が高効率化できると考えられる。さらに一連の予測モデル構築を他の作物に対しても適用可能なようにモデル構築フローを開発した(研究成果(3)の5)。

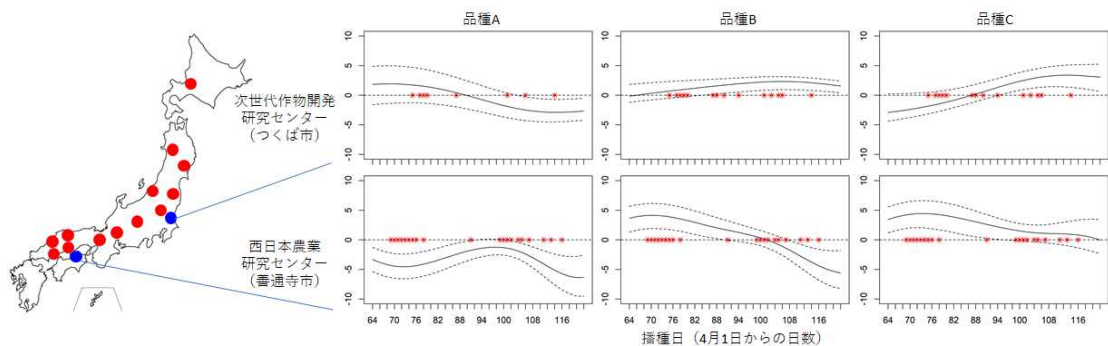


図3 ゲノミックセレクション法のモデルを構築した 15 地域と解析例。例は農研機構次世代作物開発研究センター及び西日本農業研究センターの結果であり、代表的な3品種(A から C)について収量(子実重)の結果を示した。遺伝と播種日の相互作用をスプライン関数で表現しており、縦軸は期待される収量の偏差、横軸は播種日(4月1日からの日数)を示す。図中の赤いアスタリスクは栽培データが存在する点、実線は事後平均、破線は 50% highest density interval を示す。

以上1)から4)の成果は、今回収集したダイズのヒストリカル栽培データのみならず、今後収集が進むと考えられる他作物の栽培データについても適用可能であり、今後のヒストリカル栽培データを利用した品種改良における重要な貢献であると考えられる。

文献) Onogi et al. (2016) Theor. Appl. Genet., 129: 805-817

3. 今後の展開

【データ収集及びデータベースの拡張】

本課題ではヒストリカル栽培データを格納するデータベースを構築したが、品種改良のための栽培試験は毎年行われているため、データは今後も蓄積していくことになる。そのためそれらデータを再び死蔵データにしないために、定期的にデータベースへアップロードしていく仕組みが必須となる。そこで短期的(今後1~2年間)には新たなデータをデータベース管理者である本課題担当者が集積し一括アップロードする仕組みを、より長期的(今後2~3年以降)にはダイズ品種改良担当者が直接栽培データをアップロードする仕組みを構築する予定である。また当該データベースは検索・表示・ダウンロードが主な機能となるが、それをさらに拡張し、図

表によるサマリーなどの出力や簡単な統計処理、または気象や土壌データベースとの連携なども可能なように拡張していく予定である(今後3年以内を予定)。

【ゲノムデータとヒストリカル栽培データを用いた品種改良手法の実装】

ゲノミックセレクション法の予測モデルをより広域的に多くの試験地において構築し、国内全体をカバーできるように拡大させる予定である。また多変量化を含めモデルのブラッシュアップを行う。いずれも今後1～2年程度の短期的な目標と考えている。またさらにモデルの予測能力を実証するために、実際に栽培試験を行う。これから3～5年程度かけモデルのブラッシュアップと実証試験を繰り返し、ゲノムとデータ解析に基づくよりデータセントリックな作物品種改良手法を実装していきたい。

4. 自己評価

本課題の柱であるレガシー化していた栽培データの蘇生については、主たるダイズ育成地からこの分野では世界的にも稀な規模のデータを収集し、データベース構築まで行えた点で目的を果たしたと考える。また合わせてゲノムデータの取得も行ったことにより、栽培データの利用価値が格段に向上し、ゲノミックセレクション法に代表されるデータセントリックな品種改良手法近の社会実装へ大きな貢献ができたと言える。一方で情報・数理系分野など農業分野以外の研究者との連携は希薄であったこと、論文発表が遅れたことは反省すべき点であり今後の課題としたい。

5. 主な研究成果リスト

(1)論文(原著論文)発表

該当成果なし

(2)特許出願

該当成果なし

(3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

1)小野木(2017)育種と統計とデータさきがけ, オペレーションズ・リサーチ, 62: 233-238

2)小野木(2017)ダイズ育種におけるヒストリカルデータ, 日本育種学会第132回講演会ワークショップ「ダイズにおける育種の新しい流れ」, 岩手大学

3)小野木(2018) GenomeBasedModel: 任意の生物学的統計モデルのパラメータとそのパラメータに対するゲノムワイド SNP の効果を同時推定するための R パッケージ, 日本育種学会第134回講演会口頭発表, 岡山大学

4)Onogi(2019) Comparison of F-tests for Univariate and Multivariate Mixed-Effect Models in Genome-Wide Association Mapping. Front. Genet. 10:30. doi: 10.3389/fgene.2019.00030

5)小野木(2019)全国規模のヒストリカルデータをいかにゲノミックセレクションへ利用するか:ダイズにおける研究例, 日本育種学会第135回講演会口頭発表, 千葉大学

研究報告書

「タイムドメイン宇宙観測用動画データの高速度逐次処理法の開発」

研究タイプ: 通常型

研究期間: 2015年10月～2019年3月

研究者: 酒向 重行

1. 研究のねらい

次世代天文学として注目される動的宇宙の探査では天文学が経験したことのない大量の動画データ(タイムドメインデータ)が生成される。そして、その中に稀に発生する微弱で一過性の変動現象を確実に検出し、早期に追観測へ繋げることが求められている。本研究では世界初の天文観測用広視野動画カメラである Tomo-e Gozen のデータ取得部とデータ処理部の開発を通じて実際の宇宙動画ビッグデータから微弱で一過性の変動現象を検出する技術の創出をめざす。Tomo-e Gozen カメラは東京大学木曾観測所の口径 1m シュミット望遠鏡に搭載され 1 晩に 30 TByte におよぶ大量の動画データを生成する。観測所はネットワーク帯域幅が細い僻地に位置するため、データの処理はオンサイトにて行う必要がある。得られる一過性の変動現象は S/N が低い信号ほど価値が高いが偽信号の確率も高くなる。このような特徴をもつデータと環境の制約に対して、本研究では並列計算、機械学習、動画認識、データ圧縮、データ可視化の技術を導入することで、宇宙の大量動画データから微弱な変動現象を逐次に検出・分類する技術の創出とオンサイト動画データ処理システムの構築を行う。

2. 研究成果

(1) 概要

本研究は、「サンプルデータの取得と評価」、「逐次処理システムの開発」、「一過性イベントの検出法と分類法の開発」、「データの圧縮と配布システムの開発」、「新しい宇宙動画ビッグデータの取得」の5つのテーマからなり、最終年度にこれらが融合して今後10年間にわたり実施する宇宙の広視野動画サーベイ観測につなぐ計画である。

以下では5つのテーマに対応した研究成果を順に述べる。①世界初の天文観測用広視野動画カメラ Tomo-e Gozen の開発に成功した(Sako et al. 2018)。②東京大学木曾観測所(長野県木曾郡)に 200 CPU、900 TByte ストレージから構成されるオンサイト逐次処理システムを構築した。このシステムを用いて 1 晩(8 時間)連続で 200 Mbyte/sec の宇宙動画ビッグデータの取得に成功した。③動画データの中から突発現象を検出・分類する手法を開発した。代表的なものに、ハフ変換と機械学習を用いた微光流星の検出法の開発(Ohsawa, Sako et al. 2018)、機械学習を用いた高速移動する地球接近小惑星の検出法の開発(Kojima, Sako et al. in prep.)がある。④GoDec 法を用いた天文動画データの圧縮法の開発に成功した(Morii et al. 2017)。また、国立天文台天文データセンターのアーカイブシステムを通じて観測データを世界に公開する準備を開始した。⑤2018年11月から Tomo-e Gozen を用いた継続的な広視野動画サーベイ観測を開始した。

このように本研究では当初計画の各テーマの目標を達成することに成功した。加えて以下の3項目に関して計画を上回る成果を挙げた。⑥「天文学の新しいフィールドの開拓」。秒以下の

短時間変動現象の探査(連星ブラックホール;Sako et al. 2018、高速電波バースト、パルサー)、高速移動天体の探査(地球接近小惑星、微光流星)に代表される新しい分野の観測的開拓に成功した。⑦「天文学以外への波及」。動画による広視野探査により、重力波天文学、ニュートリノ天文学との連携を実現した。宇宙機やスペースデブリの地上観測を通じて宇宙工学との連携を実現した。⑧「科学普及」。社会一般および教育機関に対して宇宙動画データを用いて科学の意義と楽しさを伝える活動を実施した。

(2) 詳細

上述した6つのテーマに関する研究成果の詳細を述べる。

① 研究テーマ「サンプルデータの取得と評価」

世界初となる広視野高速 CMOS カメラ Tomo-e Gozen のデータ取得部とデータ処理部の開発に成功した。Tomo-e Gozen は 84 台の 2k x 1k CMOS センサを搭載する天文用カメラであり、木曾シュミット望遠鏡の焦点に設置することで、空の 20 平方度の動画を取得できる(図 1)。高速読み出しが特長である CMOS センサを搭載した広視野カメラは、現状 Tomo-e Gozen のみであり、この装置だけが 10 秒以下のタイムスケールで宇宙を探査することができる。Tomo-e Gozen の開発は計画どおりに進み、2019 年 4 月までに 84 センサの全てが稼働する予定である。

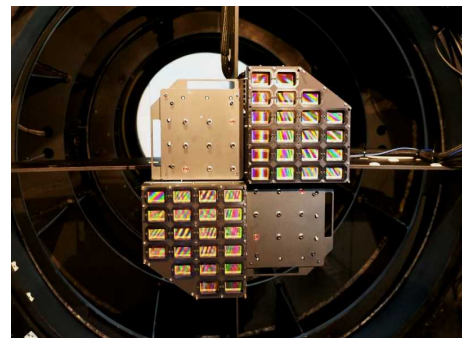


図 1. 木曾シュミット望遠鏡焦点部に設置された Tomo-e Gozen カメラ(センサを 42 チップ搭載)

② 研究テーマ「逐次処理システムの開発」

オンサイトの限られた資源で 30 TByte/夜の動画ビッグデータから一過性イベントをいかに迅速に検出し世界へ公開できるかが課題になる。本研究では取得した観測データを7日間で消去するという天文学では通常採用しない「揮発性」を導入することで、動画ビッグデータに対応できる逐次処理システムを実現した(図 2)。計算機は計 200 台の CPU と計 900 TByte のストレージからなり、望遠鏡システムと連携した自動観測を実現する。また、観測と解析の状況を把握する可視化システムも構築した。

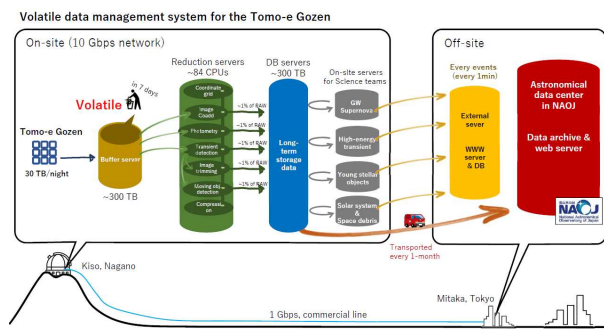


図 2. オンサイト逐次処理システムの概要

③ 研究テーマ「一過性イベントの検出法と分類法の開発」

本研究では天文動画データから短時間の変光現象や高速移動天体を検出する手法の開発を行った。(1)「高速移動する伸びる像」の検出法の開発(Ohsawa, Sako et al. 2018)。微光流星の検出がこれに相当する。各動画フレームをハフ変換した画像に対して信号のピークを検出することで微光流星を検出する。加えて、機械学習にて誤検出を除くことにより、1 晩に 1,500 イベ

ントを超える過去に無い、多数の散在流星の統計データを得ることに成功した。(2)「高速移動する点像」の検出法の開発(Kojima, Sako et al. in prep.)。高速移動する地球接近小惑星(NEO)の検出がこれに相当する。ある天域にて2Hzで6フレーム連続のデータを取得する。そして、天域を次々にスイッチしながら同様のデータを取得する。連続した6フレームの最初と最後の差分をとることで高速に移動するNEOを検出する。加えて、像形状、重心の移動速度と分散などの特徴量を使った機械学習により誤検出を低減する。結果、false-positive rateが1%のモデルを構築することに成功した。

④ 研究テーマ「データの圧縮と配布システムの開発」

本研究では膨大な動画データを科学的な情報価値をできるだけ残したまま圧縮する手法を開発した(Morii et al. 2017)。動画を行列で表現した後、GoDec 主成分分析法を使って、高速にLow-rankとスパースの行列に分解する。これによりRankを限定することでサイズを約10%に削減できると同時に、スパース行列から突発変動イベントの検出が可能となる。再構成したフレームの天体像に対して物理測定が可能な定量性が保持されていることも確認した。

⑤ 研究テーマ「新しい宇宙動画ビッグデータの取得」

2018年11月からTomo-e Gozenを用いた継続的な広視野動画サーベイ観測を開始した。評価の結果、設計性能を満たすデータが得られていることが確認された。天の川方向では、2Hzの観測で1フレームあたり約10万天体が検出される。この動画観測により1晩あたり約23億個の天体の情報が記録される。

⑥ 研究テーマ「天文学の新しいフィールドの開拓」

超新星サーベイや重力波追観測を想定した広域動画サーベイの試験を実施した。この試験により、2時間で高度40度以上の空を全てサーベイする能力があることを実証した。これは高頻度広域サーベイでは世界最高の性能になる。また、地球接近天体 2012TC4 の連続観測や連星ブラックホール MAXI J1820+070 の6msec 時間分解能の高速観測にて初期科学成果を挙げた(図3)。

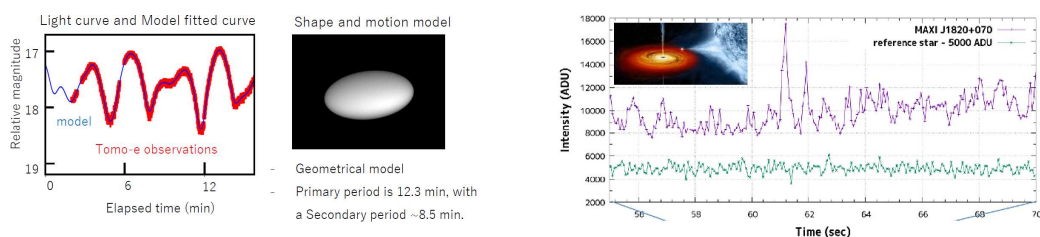


図3. 左: 地球接近天体 2012TC4 のライトカーブと再構成した小惑星の形状モデル。密な連続観測により高精度な形状モデルの構築に成功した(Urakawa et al. 2018)。右: 連星ブラックホール MAXI J1820+070 のライトカーブ。これまで可視光では得ることができなかった多くのフレアをとらえている(Sako et al. 2018)。

3. 今後の展開

本研究により宇宙動画ビッグデータを取得しオンサイトの逐次解析により突発現象を迅速かつ確実に検出する技術を創出することができた。今後は、本研究で得た要素技術を観測システ

ムに本格的に実装することで観測・解析・配信の完全自動化を進める。また、宇宙動画ビッグデータの利活用へと研究範囲を拡大する。宇宙動画ビッグデータの獲得と利用は未だ他国に例を見ない。国際連携を進めつつも日本がイニシアティブをとることができる分野へと成長させていきたい。

4. 自己評価

本研究では当初計画で掲げた各テーマの目標を達成することができた。本研究を通して天文学だけでなく、物理学、惑星科学、数理学、情報学の専門家からなる宇宙動画データの研究チームを形成できたことは今後の発展につながる大きな価値を持つ。また、天文学のために取得した宇宙動画ビッグデータと天文学のために開発したデータ解析技術が、他の自然科学研究分野だけでなく、宇宙産業、環境問題対策、教育普及活動に多大に貢献できることが実証されたことも重要な成果である。今後は宇宙動画ビッグデータの利活用を進めるとともに、広義のネイチャービッグデータを活かした Society5.0 時代の新しい多分野間連携を推進していきたい。

5. 主な研究成果リスト

(1) 論文(原著論文)発表

- | |
|---|
| 1. R. Ohsawa, S. Sako et al. "Imaging Observations of Faint Sporadic Meteors with a Wide-Field CMOS mosaic camera Tomo-e PM", Planetary and Space Science accepted, 2018 |
| 2. M. Morii, S. Ikeda, S. Sako, and R. Ohsawa "Slow-scanning in Ground-based Mid-infrared Observations", The Astrophysical Journal, 2017, Volume 835, 5 pp. |
| 3. Y. Utsumi et al. "J-GEM observations of an electromagnetic counterpart to the neutron star merger GW170817", Publications of the Astronomical Society of Japan, Volume 69, 101pp |
| 4. B. P. Abbott, et al. "Multi-messenger Observations of a Binary Neutron Star Merger", The Astrophysical Journal Letters, 2017, Volume 848, 59pp |

(2) 特許出願

研究期間累積件数:0件(公開前の出願件名については件数のみ記載)

(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

国際学会

- S. Sako et al. "The Tomo-e Gozen wide field CMOS camera for the Kiso Schmidt telescope", Proceedings of the SPIE, Volume 10702, id. 107020J 17 pp. (2018), invited.
- S. Sako et al. "Development of a prototype of the Tomo-e Gozen wide-field CMOS camera", Proceedings of the SPIE, Volume 9908, id. 99083P 15 pp. (2016)

報告

- S. Sako et al. "Detection of 10-msec scale optical flares in the black-hole binary candidate MAXI J1820+070 (ASASSN-18ey)", ATel11426 (2018)

プレスリリース



- ・ “Tomo-e Gozen Q3 ユニット ファーストライト”, 2018/11/27
- ・ “東京大学木曾観測所による地球接近天体 2012 TC4 の観測について”, 2017/10/12
- ・ “はやぶさ2」地球スイングバイの観測に成功”, 2015/12/4

研究報告書

「グローバル・システムの持続可能性評価基盤に関する研究」

研究タイプ: 通常型

研究期間: 2015年10月～2019年3月

研究者: 佐藤 彰洋

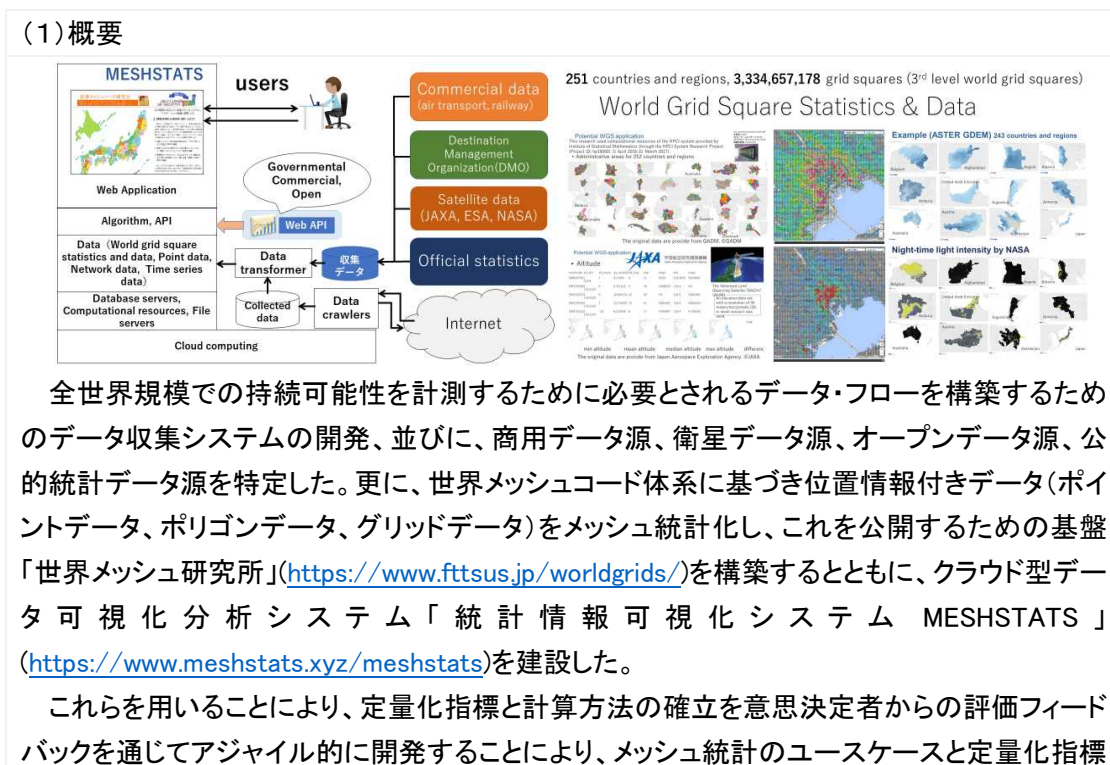
1. 研究のねらい

UNEPのイニシアティブの1つである”持続可能な観光”を定量的に評価するためのデータ駆動型支援システム開発に特化し、金融、輸送、観光に関するデータ・フローを構築し、経済性・供給と需要の計測、移動データの収集と分析、エネルギー消費量・物質投入量の算出、労働・環境の評価などを通じ、観光セクターにおける持続可能性を計量する情報基盤システムを現実のフィールドと相互作用的な評価を行いつつ段階的に建設する。特に、データ・フローの構築、指標と計算方法の開発、クラウド型のデータ分析可視化基盤の開発、現実と相互作用しながらプロジェクトを評価する方法論の開発を通じて実施し、データ、ビッグデータ分析手法、ステークホルダーとのコネクションを蓄積することでアジャイル的に実施した。

この研究により蓄積されるデータ、開発された手法とデータ基盤、ステークホルダーとの関係性により持続可能性を現実世界において定量評価することを可能とし、全世界規模で持続可能性を計測評価するための方式を開発する。このような、基盤技術は全世界規模で今後必要となる持続可能性をデータ駆動的に判断するためのシステムの基盤となるとともに、組織的な基礎を与えることを目指した。

2. 研究成果

(1) 概要



のパターン、計算方法、可視化方法を開発した。

これらのオープンソース、オープンデータと世界メッシュコードに関する技術資料を公開する専用のサイトとして世界メッシュ研究所を設置した。

世界メッシュ統計のユースケースおよび新規データの開発を支えるコミュニティとして世界メッシュ研究所サポート会員(2018年12月現在 57名)、統計情報可視化システム MESHSTATS のモニターユーザー(2018年12月現在 245名)からなる研究協力者を獲得し、定期的に世界メッシュコード研究会を開催することによりオープンイノベーションにより技術の開発と普及を行っている。

日本工業規格地域メッシュコード(JIS X0410)の上位互換性を有する世界メッシュコード体系構成し、世界メッシュコード関連オープンライブラリを 5 言語(R, PHP, Javascript, Python, JAVA)に対して開発し、世界メッシュ研究所 Web ページより公開を行っている。

世界メッシュ統計データを全世界の行政界ポリゴンデータである GADM、NASA が提供する夜間光画像、経産省が NASA と共同運用する衛星 Tera に搭載される ASTER に基づく 30m 角陸域標高データ、JAXA だいち (ALOS)で作成された 30m 陸域標高データを用いて作成し、251 の国と地域に対して 33 億メッシュ以上(3 次メッシュ)を生成し、世界メッシュ研究所からオープンデータとして公開した。

統計情報可視化システム MESHSTATS は以下の機能を有する。

- ・世界メッシュコードに基づく全世界規模でのメッシュ統計用 Data as a Service (DaaS) 各種データの収集および世界メッシュ統計データの半自動生成を可能とする技術
- ・クラウドベースでのデータ分析可視化機能(GUI)と機械判読可能な API 機能
- ・31の機能(メッシュデータ種類、集計・可視化方式)

統計情報可視化システム MESHSTATS を用いることにより、全世界を世界標準的な規格化されたメッシュ統計データとして取り扱うことが可能であり、かつ、秘匿化された状態でデータを保管、検索、結合分析することが可能である。また、MESHSTATS 上に実装された機能とメッシュ統計の特性を用いることにより、経済社会システムに内在する効率性の改善、潜在的な需要の発見、変化の検出が可能となる。ビッグデータ解析技術として、世界メッシュ統計を半自動で生成するアルゴリズム、メッシュ統計を任意の形状に再集計する機能、異なる場所の構造物配置の特徴を比較する方法、複数のメッシュ統計からパターンを検出する方法を開発した。

(2) 詳細

・研究テーマ1「データ・フローの構築」

OAG からデータを購入することにより、全世界規模での日次航空機タイムテーブルデータを5年間分取得し、全世界の航空輸送の状況を把握するとともに、我が国の国際航空輸送可能量の推計を行った。その結果、全世界で年間50億座席の航空輸送容量が存在していることを突き止めるとともに、我が国の年間国際航空容量が約4000万座席(2017年)であることが判明した。

各国の中央銀行のRSSから自動的にデータを取得し、全世界の約60通貨ペアに対する通貨交換レートを自動的に整理統合できるデータ基盤の開発を行った。この機能を用いて MESHSTATS の貨幣価値の自動変換機能を実装した。

多言語対応を行うため、日本語と英語を人間が入力することで自動的に機械翻訳を行う機能と人間によるレビュー機能を有する多言語辞書(語彙基盤)の作成を行った。この機能を用いることで、14か国語で MESHSTATS を利用できるシステムとした。

全世界規模での基礎データとして、JAXA が運用していただいち(ALOS)、経済産業省が NASA と共同運用する地球観測衛星 TERA で運用される ASTER センサー由来の衛星リモートセンシングデータ、および、NASA が公開する夜間光衛星画像をデータ源として、全世界規模での標高と夜間光の世界メッシュ統計データを作成した。更に、土地利用インデックス、欧州、日本、オーストラリアの公的統計由来メッシュ統計を世界メッシュへ変換することにより、様々な世界メッシュ統計を作成することができ、最終的に500種類以上、33億メッシュ(3次元メッシュ)以上の世界メッシュ統計の作成と、インターネット由来のホテル宿泊プランメッシュ、求人メッシュを日次で自動生成することに成功し、研究計画で記した政府統計とインターネット上のビッグデータを合成することによる、これまでになかった高頻度統計を実現した。

・研究テーマ2「情報基盤システムの構築」

構築した世界メッシュ統計を任意の場所と分野に対して Web ブラウザ経由で抜き出し、分析・可視化するためのシステム「統計情報可視化システム MESHSTATS」をクラウドサーバー上に構築した。このシステムをモニターユーザーに開放し、データ利用の方法、データ分析のユースケースや、関心とするデータ分野について VoC(Voice of Customer)の手法を利用して、聞き取り調査と意見収集を研究会(世界メッシュコード研究会)において実施することで、システム開発とデータ開発の仕様と要求の抽出を行った。

更に、グラフィカル・ユーザー・インターフェースを聞き取りした仕様をもとに開発、世界メッシュ統計を簡単に可視化し、相関パターンの検出や、外れ値の特定を行うことができるデータ分析機能を実装した。これらの仕組みはソフトウェア・データマイグレーションモデルに従い、クライアント上に送り込まれたソフトウェアとデータによりサーバーとの通信なしにインタラクティブ分析を実現している。

「統計情報可視化システム MESHSTATS」は世界メッシュ統計に特化した、Data as a Service(DaaS)であり、約250名のモニターユーザーへ提供している。この研究を通じて、世界メッシュ統計を半自動的に連続生成し、サービスとして提供する方式のワークフローとフレームワークを開発した。このようなワークフローとフレームワークの構築および世界標準に準拠したシステムの開発が可能となることは当初の予想を上回る成果となった。

・研究テーマ3「世界規模での情報基盤システムへの拡張」

14か国語、31種類(データ種別×集計種別×分野)の機能を個別ユーザー管理のもとで、提供するシステムとして、全世界規模で世界メッシュ統計データを提供することができるシステムとして、日本国内で開発した機能をその国の人が、その国の言葉で、その国の世界メッシュ統計を自身の国のために開発されたシステムであるかのように提供可能なシステムとして拡張を完了した。内部には、14か国言語(日本語、英語、イタリア語、ドイツ語、スペイン語、韓国語、ベトナム語、中国語簡体字、中国語繁体字、ポーランド語、タイ語、トルコ語、フランス語、アラビア語)、1000単位以上の多言語語彙基盤を有し、機械翻訳 API を用いた、自動翻訳機能と多言語環境でレビュー機能を有する機能が埋め込まれている。これにより、システム内部で言語的専門能力を有する専門家が Web ブラウザを通じて世界どこからでも言語テ

ブルの編集作業を行うことができるシステムとなっている。このシステムを通じて、在日ベトナム人研究補助者、在日中国人研究補助者、ドイツ在住の研究協力者、タイの在住の研究協力者、米国在住の翻訳者が語彙基盤のレビュー作業に従事することのより、多言語環境の精度向上が機械翻訳と連動させることで可能であることを実証した。更に、ドイツ在住の研究協力者は統計情報可視化システム MESHSTATS を利用することでビッグデータ教育プログラムの開発を行っている。

・研究テーマ4「実際のフィールドと相互作用する評価」

平成27年度(2016年3月17日)に京都大学においてキックオフワークショップ(20名参加)、平成28年度に、イタリア・マテラ(2016年7月9日)、ドイツ・ハンブルグ(2017年3月8日)、イタリア・トリノ(2017年7月6日)において国際ワークショップを共催し、国際的な参加者とともに、世界メッシュ統計の利用方法の紹介と、MESHSTATS の利用シーンを説明した。更に、国連世界観光機関技術と観光部会に Google と計画の提案を行い、各国の観光分野の専門機関へ世界メッシュ統計とMESHSTATSを紹介するとともに、自国の観光施策の実現に世界メッシュ統計が利用可能であることを、我が国の公的統計由来メッシュ統計の利活用方法を紹介することにより説明を行った。平成30年12月3日には京都大学時計台百周年記念館において最終成果報告会を開催(のべ約60名の参加)し、これまでの研究成果について説明するとともに、世界メッシュ統計に関連する招待講演12件とパネルディスカッションによる本研究活動の総括を行った。



この最終成果報告会を契機として、参加者および講演者との間の共同研究が促進され、その後も、データ利活用のユースケース開発(HR 分野、観光分野)とメッシュデータ開発(医療分野)を企業実務家、大学研究者と実施している。

3. 今後の展開

これまでの研究から、世界メッシュ統計を作成できる位置情報付きデータの種類と、世界メッシュ統計の生成方法は特定できた。インターネット上で公開されている位置情報付きデータを用い

て、10,000種類以上のメッシュ統計を作成することが可能である目途が立った。更に、衛星リモートセンシングデータを用いることにより、全世界規模で衛星データ由来世界メッシュ統計を実時間に近い頻度で生成することが可能である技術的目途が立った。これにより、全世界規模で世界メッシュ統計データを生成し、流通すること、および、複数連動する MESHSTATS に組み込むことで異なる分野で収集される位置情報を含むデータを連結分析し、組織をまたぎ相互融通することができることが分かった。

この技術を用いると、様々な分野で収集蓄積されているが、利用が未着手または、データフォーマットが異なるため相互利用が十分にすすんでいないデータを相互に結合、統合して分析することが全世界規模で可能となると見込まれる。これにより、データに基づき、持続可能性を計測する技術や、公的統計の近代化、調査無き統計などを社会実装できる基盤技術として開発した研究成果を利用、普及することができる技術的目途がたつた。今後エコシステムを構成するために、世界メッシュコードで作成されたデータを公開する企業や非営利団体と協力を深め、世界メッシュ統計データを利用し、統計情報可視化システム MESHSTATS を運用・利用する人と組織を着実に増やしていきながら、分野横断的にメッシュ統計の利活用ユースケースを開発し、市場化していくことが課題である。

4. 自己評価

本研究では、全世界規模での世界メッシュ統計基盤技術と世界メッシュ統計データの大規模な開発を目標とし、持続可能性をエビデンスより評価可能とするクラウド型基盤の開発と、データコミュニティの開発を目指した。研究開始当時 2015 年 10 月においては、世界メッシュ統計のライブラリは R 言語で部分的に地域メッシュコードの計算ができるのみであったが、研究開始後、R 言語による世界メッシュコード関連関数の開発に成功し、その後、このライブラリの移植作業とこのライブラリを用いた世界メッシュ統計データの作成に着手した。2018 年 9 月時点で、衛星データ由来の世界メッシュ生成技術を確立し、世界メッシュ統計の種類は 500 種類を超え、世界メッシュ統計は世界 251 の国と地域に対して、33 億メッシュ以上にまで到達した。データ・フローの構築は当初の予想を上回る成果となった。開発したライブラリと世界メッシュ統計はオープン化しプロジェクト Web ページである世界メッシュ研究所 (<https://www.fttsus.jp/worldgrids/>)より日本語、英語、中国語(繁体、簡体)、ベトナム語で公開するとともに、2018 年 2 月 15 日に京都大学、科学技術振興機構、一般財団法人宇宙システム開発利用推進機構、MM 総研の連名でプレスリリースを行った。

世界メッシュ統計を任意の場所と分野で抜き出し、集計方法と可視化方法で再集計するためのクラウドベースでの分析基盤「統計情報可視化システム MESHSTATS」(<https://www.meshstats.xyz/meshstats/>)をアジャイル的に開発し、31機能と14か国言語(日本語、英語、ドイツ語、イタリア語、スペイン語、タイ語、ベトナム語、中国語繁体字、中国語簡体字、フランス語、アラビア語、ポーランド語、トルコ語、韓国語)で利用し、全世界規模で、世界メッシュ統計を抜き出し、可視化、分析ができる機能を有するシステムとなってきた。本研究で目標としたデータ分析システムの国際化の目標を十分に達成することができた。

本研究の最終成果報告会(2018 年 12 月 3 日開催)において、評価委員 10 名による研究プロジェクト評価を実施した。評価委員はこれまで、過去2年以上に渡り、本プロジェクトへの参加、打ち合わせや研究会参加活動経験が3回以上ある研究者、実務家を対象とした。本研究

プロジェクトで開発された研究成果である世界メッシュ統計データのビッグデータ度、統計情報可視システム MESHSTATS の充実度、指標の有益度、研究プロジェクトへの協力・参加により得られたメリットの有無などから多面的主観評価尺度を構成し、アンケート調査方式で意見の収集を行った。その結果、ビッグデータ度は80%の評価者が豊富または利用しきれないくらい豊富と答え、プロジェクトの成果物である統計情報可視化システム MESHSTATS の充実度は70%の評価者が実用に耐えられる30%の評価者が更なる開発を求めると回答した。更に、90%の評価者は活動で何らかの恩恵を得ていたことが判明した。

世界メッシュコードを算出するライブラリ、基本的世界メッシュ統計はオープンデータとして公開しているが、これらのオープンライブラリ、オープンデータに関する問い合わせや、利用方法の照会が行われている。世界メッシュコード研究会の開催実績はプロジェクト開始より、15回を超え、引き続き研究会開催がプロジェクト終了後も継続している。世界メッシュコード研究会の登録者数は日本全国で約60名に達し、研究会開催時には常時10名以上の参加者が集い、オープンイノベーションによる世界メッシュ統計利用シーンとユースケース開発およびデータ仕様とシステム要求の開発に協力して頂いている。

このことから、研究計画の進捗度は計画通りまたはそれを上回っている。更に、本研究プロジェクトへモニター参加、共同研究、連携等による協力者数の増加傾向を鑑みると、本研究プロジェクトは当初予想していなかった恩恵を参加者へもたらすものとなっている。

5. 主な研究成果リスト

(1) 論文(原著論文)発表

1. Aki-Hiro Sato, Shoki Nishimura, Tsuyoshi Namiki, Naoki Makita, Hiroe Tsubaki, "World Grid Square Data Reference Framework and its Potential Applications", 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), (2018), pp. 398-409, DOI: 10.1109/COMPSAC.2018.00062
2. Aki-Hiro Sato, Paolo Tasca, Takashi Isogai, "Dynamic Interaction Between Asset Prices and Bank Behavior: A Systemic Risk Perspective", Comput. Econ. (2018) pp.1-31, DOI: 10.1007/s10614-018-9792-y
3. Aki-Hiro Sato, Shoki Nishimura, Hiroe Tsubaki, "World grid square codes: Definition and an example of world grid square data", 2017 IEEE International Conference on Big Data (Big Data), pp. 4238--4247 (2017), DOI: 10.1109/BigData.2017.8258450
4. Anna Carbone, Meiko Jensen, Aki-Hiro Sato, "Challenges in Data Science: a complex system perspective", Chaos, Soliton and Fractals, Vol. 90 (2016) pp. 1-7, DOI: 10.1016/j.chaos.2016.04.020
5. Aki-Hiro Sato, Hidefumi Sawai, "Risk Assessment for a Global Air Transport System Using Socioeconomic-Technological-Environmental Databases", 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC2016) (2016) pp. 572-581, DOI: 10.1109/COMPSAC.2016.179

(2) 特許出願

研究期間累積件数:0件

(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

・2018/11/19 京都大学プレスリリース「統計情報可視化システム MESHSTATS 実用化プロジェクトを開始しました」(2018年10月22日)

http://www.kyoto-u.ac.jp/ja/research/events_news/department/jyohogaku/news/2018/181022_1.html

・2018/10/6 佐藤彰洋, 椿広計, “データ利活用基盤とグローバル観光サービスデザイン”, 第9回横幹連合カンファレンス OS09「データ駆動型社会における観光サービスの展望」, 電気通信大学, 東京都調布市, https://doi.org/10.11487/oukan.2018.0_A-2-5

・2018/2/15, 科学技術振興機構プレスリリース「衛星ビッグデータと世界位置情報を統合解析する手法を世界で初めて開発～標高ビッグデータで津波リスク予測ほかデータ利活用を促進～」, <https://www.jst.go.jp/pr/announce/20180215/index.html>

・2017/11/17, Aki-Hiro Sato, Shoki Nishimura, Tsuyoshi Namiki, Naoki Makita, Hiroe Tsubaki, “World Grid Square Statistics and their application to data analytics”, Joint UNECE/UN-GGIM Workshop on Integrating Geospatial and Statistical Standards, Stockholm, Sweden

https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.58/2017/mtg3/S3_SATO_2017World_Grid_Square_Statisticsv52.pdf

・2017/6/23, Aki-Hiro Sato “Data Infrastructure for Data Applications”, 6th UNWTO International Conference on Tourism Statistics: Measuring Sustainable Tourism, 21 June 2017 - 24 June 2017, Manila, Philippines,

<https://custom.cvent.com/E5C28A0D212A415D9AD3C8B699EBC072/files/82661a63c16c4780ba2124dd68cf48bd.pdf>

研究報告書

「時空間粒度の異なる教育ビッグデータの非同期ストリーム処理基盤の構築」

研究タイプ: 通常型

研究期間: 2015年10月～2019年3月

研究者: 島田 敬士

1. 研究のねらい

都市、人流、医療、広告、購買など様々なビッグデータの利活用時代の到来により、個人やスポットのような「点」に注目していた従来の解析から、群衆や面のように「空間」的に実世界を捉えて、社会全体の最適化や個人の快適性を向上させるための解析に軸足が移りつつある。従来のビッグデータ解析においては収集したデータを様々な時空間粒度で分析して、その結果を将来や次世代にフィードバックするリレー形式の活用が多く見受けられる。一方、データの収集からフィードバックまでの時間をできるだけ短縮するためのストリーム処理が実現できれば、ビッグデータの提供元である現場や現場のユーザが直接的な恩恵を受けることができる社会を実現できる。

本研究では、時空間粒度の異なるビッグデータのストリーム処理基盤の構築を目的とし、特に教育分野への応用を考える。近年、教育現場へのICT導入が積極的に行われており、その結果として教育ビッグデータが形成されるようになりつつある。従来のICT教育は主に個人学習の解析を対象にしていたが、教育ビッグデータを活用することで他の学習者との比較やグループ・教室・科目・大学のような異なる空間粒度での解析や、リアルタイム・週単位・学期単位・年単位などの異なる時間粒度での解析が可能になる。教育ビッグデータには、電子教材や情報端末、デジタル学習環境などから学生や教師の膨大な活動ログが非同期に収集されているため、本研究で開発する処理基盤技術の有効性検証に利用できる。

本研究で具体的に扱う教育ビッグデータは、学習管理システム(e-Learningシステム)とデジタル教科書システム(e-Bookシステム)から収集される大学生の学習活動ならびに講義を行う教師の教育活動データである。各学生がシステムを利用するタイミングは各々で異なるため、データは非同期に収集される。また、同一講義、同種の講義間、学期間、大学間などで同種の情報であっても異なる時間や空間の粒度でデータが収集される。そのような教育ビッグデータを利用して、学習困難に陥っている学生を発見したり、教室全体の学習理解度を把握したり、学習到達度を予測したり、学生に理解を助長する資料を提供したり、講義進行の調整案を教師に提示したりできるフィードバックシステムを実現し、教育学習環境の最適化を目指す。

2. 研究成果

(1) 概要

研究者が所属する九州大学で学習管理システム(Moodle)、デジタル教科書システム(BookRoll)を中心としたデジタル学習環境による教育ビッグデータ収集基盤を構築し、PC必携化が行われている大学教育において時空間粒度の異なるビッグデータのストリーム処理基盤の構築を行った。研究成果は空間粒度、時間粒度の観点で分類できる。空間粒度としては、個人レベルの粒度では、デジタル学習環境で収集されるイベントストリーム(学生がシス

テムを操作した際に得られる記録)から個人の学習パターンをモデル化することで異常行動を検知する技術, 個人の学習理解度に応じて復習教材を自動生成する技術の開発を行った。また, 教室規模(100~200 名程度が受講する講義)の粒度では, デジタル教科書の閲覧パターン分析技術, 予習用要約教材の自動生成技術, 講義中のリアルタイム閲覧状況分析技術, 着席位置と学習活動の関係解析技術の開発を行った。さらに, 同種の講義を対象とした科目レベルの粒度では, コース間の類似・非類似性を把握するコース横断分析技術, 成績予測技術の開発を行った。これらの成果は, 時間粒度の観点でも整理可能であり, 以下の研究成果詳細では教育ビッグデータの分析結果をフィードバックする際の時間粒度の違いで成果を分類し, 取り組み内容を報告する。

研究テーマ A 「現場の教育・学習を支援するリアルタイム学習分析技術の開発」

研究テーマ B 「各講義の内容理解を助長する教材分析技術の開発」

研究テーマ C 「長期間データを利用した学習活動分析技術の開発」

(2) 詳細

研究テーマ A 「現場の教育・学習を支援するリアルタイム学習分析技術の開発」

本テーマでは, 現場の教育・学習をリアルタイムに支援するための教育ビッグデータ処理基盤を開発した。本成果を利用することで, 講義を行う教師は現場の学生の学習状況に応じて柔軟な講義進行を行えるようになる。

予習状況分析技術では, デジタル教科書のアクセス履歴からページ単位で講義受講者の予習状況を集計し, 未予習者の多いページを特定することが可能である[1]。また, 予習理解度を確認するための小テストを実施し, 正答率の低い設問に対応する教材内の該当ページを教師に提示することが可能である。設問と教材を接続するために, 設問文内の単語と教材の各ページに含まれる単語間の類似度を計算することで, 関連ページを特定する技術を開発した。教師は提示された教材のページの説明時間を長く確保したり, 説明を補足したりするなど柔軟な対応を行えるようになった。

閲覧状況分析技術では, 講義に参加している学生がデジタル教科書システムにアクセスして教科書を開いたり, ページを移動したりする際に記録される学習活動ログを瞬時に集計し, 1分毎の閲覧分布を教師のPC上に可視化するシステム(図 1)を開発した[1]。デジタル教科書システムへのアクセス履歴とアクセス者の履修情報を統合分析することで, 各授業に参加している学生を特定しつつ, 教室内の学習活動を分単位で統合化処理する技術を開発した。教師へのフィードバックは各ページの閲覧者数をヒートマップ風に表示することで行われ, 表示は1分ごとに更新される。ヒートマップ上では教師が説明しているページが赤枠でハイライトされる。教師が説明しているページ付近に多くの学生が閲覧している状況が確認できれば, 授業進行は概ね順調とみなせるが, 閲覧分布に偏りが見受けられれば, 学生が講義進行についていけない可能性や, ページの先々を見ている可能性が考えられる。教師はそのような状況を瞬時に把握し, 授業進行のスピードを遅くして説明を補足したり, 学生がキャッチアップするための時間を設定したりするなど, 現場の実状況に応じた柔軟な講義進行が可能になる。

学習活動変化検出技術では、講義に参加している各学生のシステムへのアクセスパターンをモデル化し、モデルから逸脱するパターンを検出することを可能にした[2]。先行研究では教室全体のアクセスパターンをモデル化する技術が提案されていたが、実際のアクセスパターンを分析すると個人による特徴が強いことが分かったため、アクセスパターンの個人化を行った。また、現場でのリアルタイム変化検出に向けて、個人のアクセスパターンをポアソン過程により逐次的にモデル化する方法を開発した。本技術により異常パターンとして検出された学生は、他の学生よりも事後テストの成績が低いことが確認されたことから、開発した技術の有用性が確認できた。

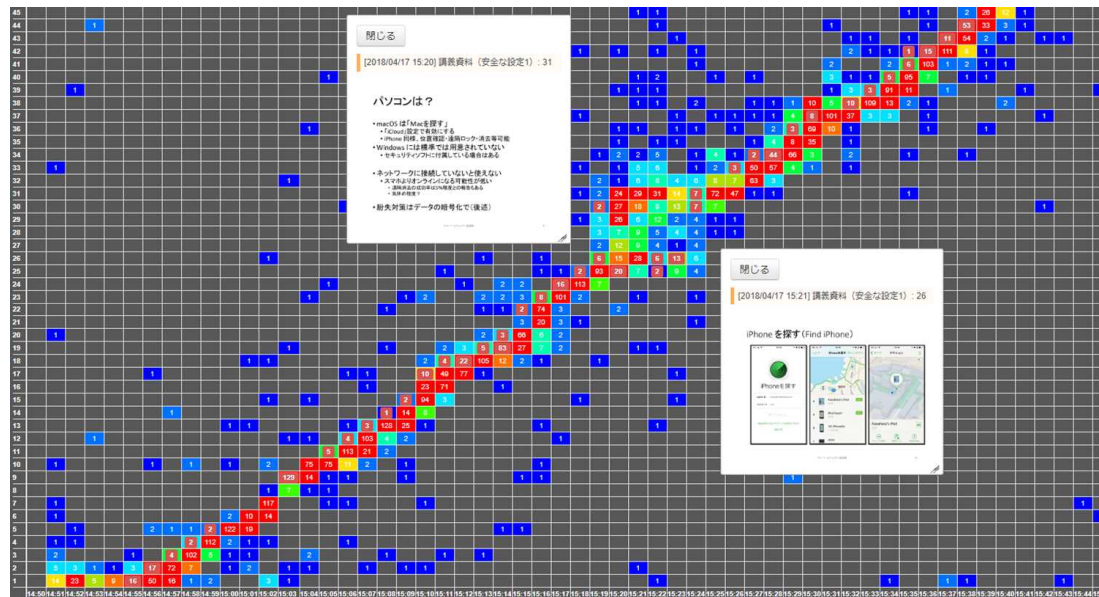


図 1 教材閲覧ヒートマップ。横軸は時刻，縦軸は教材のページ番号に対応。同時刻に同じページを閲覧している学生が多いほど赤系の色で表示される。ページ番号をクリックすると該当ページの内容がポップアップ表示される。

研究テーマB「各講義の内容理解を助長する教材分析技術の開発」

本テーマでは、デジタル教科書システムに登録されている教材を分析対象とし、学生の予習復習を支援するための教材生成技術を開発した。本成果を利用することで、予習復習の効率・効果の向上ならびに教材改善を行うことができる。

予習教材生成技術では、学生に電子教材の要約版を自動生成する技術を開発した[3]。具体的には、デジタル教科書システムに登録された電子教材に対して、画像処理とテキスト処理を適用し、視覚的な観点と単語の重要度の観点を統合し、各ページの重要度を算出する。さらに、各ページの閲覧に要する推定時間を用いて、短時間でできるだけ閲覧効果の高いページの組み合わせを自動的に抜粋し、要約版資料として再構成する技術を開発した。要約版資料を予習に利用した学生群からは、予習達成率の飛躍的な向上ならびに事前テストでの成績向上の効果が確認できた。

復習教材生成技術では、学生の講義内容の理解度に応じて復習を支援するための要約版資料を生成する技術を開発した[4]。具体的には、学習管理システムで実施される事後テストの結果を分析し、誤答した設問に対応する教材のページを自動発見するアルゴリズムを開発

した。設問文に含まれる単語を解析し、その単語に対して関連度の高い教材のページをランキング形式で出力する。また、各学習者の予習状況も同時に分析し、未予習箇所についても自動推薦することが可能である。本技術を利用して生成された復習教材を利用することで、事後テストでの理解度の向上が確認できた。

研究テーマC「長期間データを利用した学習活動分析技術の開発」

本テーマでは、週単位や学期単位で収集された学習ログを分析対象とし、学生がどのような学習活動を行っているかを定量的に分析するための学習活動分析技術を開発した。本成果は、講義を実施する教師や、授業・コースを管理する科目マネージャなどが授業改善の方策を検討したり、困難に陥っている学生に対する対応を検討したりする際に活用できる。

教材の閲覧パターン分析技術では、同一教材を閲覧した学生群の閲覧パターンを要約する技術を開発した[5]。各学生の各ページの閲覧時間を行列表現し、行列分解手法を適用することで、閲覧パターンを抽出した。本技術を1週間ごとの予習期間中の学習ログに適用したところ、代表的な閲覧パターンが複数発見され、その中には効率の良い閲覧方法で予習を実施し、結果として事前テストでよい成績を修めている学生群を発見することができた。

着席位置と学習活動の関係分析技術では、学期中に学生が教室内で着席していたエリア情報と講義内外での学習活動の関係を分析する技術を開発した[6]。学生の着席エリアを収集するために、学習管理システム上にクリッカープラグインを実装し、毎回の講義開始時に自分が着席しているエリアを学生に投稿してもらう方式を採用した。また、学習活動については、出席状況、小テストの結果、課題の提出状況、予習復習状況、授業中の閲覧状況などを数値化した。14週間に渡りデータを収集し、開発した分析技術を適用したところ、教室前方に着席傾向の強い学生の方が教室後方に着席傾向の強い学生よりも学習活動の活発さが高いことが確認された。また、着席エリアを頻繁に変更する学生は成績が低く、欠席傾向が高いことも明らかになった。

成績予測技術では、学期中に収集された学習活動ログと最終成績の関係をモデル化し、週ごとの学習活動から最終成績を予測するための成績予測モデルを開発した[7]。出席状況、小テストの結果、課題の提出状況、授業外での教材閲覧状況を週ごとに数値化し、8週(クォーター)あるいは15週(セメスター)後の成績を予測するためのリカレントニューラルネットワークモデルを生成した。ネットワークの学習には、昨年度以前のデータを利用し、実際の成績予測には現行年度のデータを利用した。全体期間の半分程度の期間情報をネットワークに入力できれば、8割以上の精度で成績を予測できることが確認できた。

3. 今後の展開

本研究を通して、様々な時空間粒度での教育ビッグデータの分析技術を開発してきたが、今後は本技術の横展開を考えていく必要がある。同様のデジタル学習環境を導入している大学への展開は比較的容易であると考えるが、学習ログの形式が異なるシステムを利用している場合は、ログフォーマットの変換あるいは分析技術の改修が必要になる。そのため、学習ログの標準化についての研究についても今後進めていく。

また、他大学とのデータ連携や学習分析結果の連携を行うためには、データの匿名化やプライバシーポリシーの策定が急務である。プライバシーポリシーについては、学内では実験倫理委員会

を中心とした関連組織での議論を進めており、ボトムアップな研究を推進するうえでのデータの取り扱いや管理方針を定める予定である。またトップダウン的には大学の枠を超えた国家レベルでの教育データの取り扱いについて今後議論が加速していくことを期待している。その際に本研究の取り組みがひとつの先行事例になるよう取り組みを続けていく。

4. 自己評価

1)教育ビッグデータを収集するためのデジタル学習環境の構築, 2)学習管理システムならびにデジタル教科書システムを利用して収集される大規模な学習活動ログを様々な時空間粒度で分析する技術, 3)現場の講義環境や学生, 教師に対して分析結果をフィードバックし, 学習改善ならびに教育改善に資する成果が得られた, という点において, 本研究が ICT を活用した教育に対して一定の成果をあげることができたと考える。

デジタル学習環境を安定運用することは, 研究成果(論文)として直接的には表れないが, 研究を実施するうえでは非常に重要な課題である。研究補助者にシステム運用を支援してもらえたことで, 研究に集中することができた。また, プロトタイプシステムを実装し, それを実運用システムに実装する際には, 専門の業者にソフトウェア開発を依頼することができたため, 開発効率が飛躍的に向上した。さらに, 国内外で研究成果発表を数多く行うために研究費を活用できた。

大学教育におけるPC必携化が進んでおり, また近い将来初等・中等教育に対して教育の情報化が実施されることを鑑みると, 本研究成果を社会展開する環境が国家レベルで整いつつあると考えられる。教育の情報化が進めば, 自ずとデータ利活用の動きが活発化すると推測されるため, 本研究成果は先行的な研究事例として注目される可能性が極めて高い。また, 開発した技術については社会展開を見据えて, 汎用性を重視した実装を進めてきており, 今後様々な教育学習システムに本研究成果が導入されることを期待する。またそうなるように研究者自身も活動の幅を広げていく。

5. 主な研究成果リスト

(1)論文(原著論文)発表

1. Atsushi Shimada, Shin'ichi Konomi, Hiroaki Ogata, Real-time learning analytics system for improvement of on-site lectures, Interactive Technology and Smart Education, Vol.15, No.4, pp.314-331, 2018.
2. Atsushi Shimada, Yuta Taniguchi, Fumiya Okubo, Shinichi Konomi, Hiroaki Ogata, Online Change Detection for Monitoring Individual Student Behavior via Clickstream Data on e-Book System, 8th International Conference on Learning Analytics & Knowledge (LAK'18), pp.446-450, 2018.
3. Atsushi Shimada, Fumiya Okubo, Chengjiu Yin, Hiroaki Ogata, Automatic Summarization of Lecture Slides for Enhanced Student Preview -Technical Report and User Study-, IEEE Transactions on Learning Technologies, Vol.11, Issue2, pp.165-178, 2017.
4. Atsushi Shimada, Fumiya Okubo, Chengjiu Yin, Hiroaki Ogata, Automatic Generation of Personalized Review Materials Based on Across-Learning-System Analysis, Cross-LAK2016, 2016.

- | |
|---|
| 5. Atsushi Shimada, Fumiya Okubo, Hiroaki Ogata, Browsing–Pattern Mining from e–Book Logs with Non–negative Matrix Factorization, the 9th International Conference on Educational Data Mining, pp.636–637, 2016. |
| 6. Atsushi Shimada, Fumiya Okubo, Yuta Taniguchi, Hiroaki Ogata, Rin–ichiro Taniguchi, Shin’ichi Konomi, Relation Analysis between Learning Activities on Digital Learning System and Seating Area in Classrooms, 11th International Conference on Educational Data Mining, 2018. |
| 7. Fumiya Okubo, Takayoshi Yamashita, Atsushi Shimada, Hiroaki Ogata, A Neural Network Approach for Students’ Performance Prediction, The 7th International Conference on Learning Analytics & Knowledge Understanding, 2017. |

(2)特許出願

研究期間累積件数:3件

1.

発明者: 島田敬士, 緒方広明

発明の名称: METHOD FOR ARRANGING SUMMARY SLIDE SET

出願人: 九州大学

出願日: 2015年9月

出願番号: 62/212608

2.

発明者: 島田敬士, 緒方広明

発明の名称: スライド要約装置、学習支援システム、スライド選択方法及びプログラム

出願人: 九州大学

出願日: 2016年4月

出願番号: 特願 2016–083102

3.

発明者: 島田敬士, 緒方広明, 清田麻寛, 毛利考佑

発明の名称: 情報表示装置、学習管理システム、情報表示方法及び情報表示プログラム

出願人: 九州大学

出願日: 2017年3月

出願番号: 特願 2017–042856

(3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

受賞

第1回 IMS Japan 賞 最優秀賞

第13回日本 e–Learning 大賞 教育ビッグデータ特別部門賞

IPJSJ/IEEE–CS Young Computer Researcher Award(授賞式は2019年6月予定)

報道発表



2017年1月25日掲載、日本経済新聞(31面)、「学習支援にビッグデータ」
2017年1月26日掲載、西日本新聞夕刊(1面)、「九大デジタル教育革命」

招待講演

日本教育工学会 第34回全国大会

「実時間学習分析による教育改革の可能性」

一般社団法人データサイエンティスト協会 5th シンポジウム

「ラーニングアナリティクスを用いた教育・学習支援の可能性」

研究報告書

「思考・行動を予想する脳ビッグデータ(データさがけ)」

研究タイプ: 通常型

研究期間: 2015年10月～2019年3月

研究者: 柳澤 琢史

1. 研究のねらい

近年、情報通信技術の発達に伴い、小型で大容量の無線通信が可能になり、詳細な脳信号など大量の情報を体内に埋め込んだ無線通信器でリアルタイムに取得し、オンラインで解析することが可能となった。このような技術は脳信号で機械を制御する Brain-machine interface (BMI) として利用されるだけでなく、てんかん発作の予知や、精神疾患の治療など医療での応用範囲が広がって行くと期待されている。実際、体内埋め込み型の脳信号計測装置は動物実験での有効性が示され、ヒトを対象とした臨床研究が進んでいる。また、Body Area Network など、生体信号をオンラインで収集しビッグデータとして活用しながら、個人へフィードバックする新たな医療システムが開発されている。近い将来には、大量の脳信号や行動を含む生体信号がオンラインで得られる状況になると予想される。また、詳細なライフログと脳信号とのビッグデータにより新しい技術開発が期待される。しかし、特に脳信号に関して、このようなデータをビッグデータとして扱う技術は十分に研究されていない。実際、頭皮脳波などの非侵襲的脳信号では、長時間安定してライフログと同期した計測をする事は困難である。一方、頭蓋内の電極から計測する脳信号は安定して計測できるが、個人情報の点から、多くの研究者が共有して解析することは困難であった。

そこで、本研究課題では、脳神経外科での治療目的で頭蓋内に留置した電極から高精度な脳信号(皮質脳波)を得て、これとライフログとを同期して計測したデータを、多くの研究者と共有可能な脳ビッグデータとすることで、新しい技術開発に繋げることを目的とした。具体的には、皮質脳波と同期して様々なライフログを連続して計測する環境を整備した。そして、得られたデータから個人情報を排除した脳ビッグデータを半自動で作成するプラットフォームを作成した。また、これを収集し、多くの研究者と共有しながら患者群としての解析ができる統合データベースを作成した。さらに、同ビッグデータを本プロジェクト内外の研究者と共有し、皮質脳波を用いた新しい技術の開発に応用した。これにより、脳ビッグデータを用いた技術開発を促進することを目指した。

2. 研究成果

(1) 概要

頭蓋内脳波とライフログの同時計測システムを開発し、各被験者につき10日程度の連続計測を49人分施行し、のべ412日分の同時計測データを取得した。各個人脳上の電極位置は、標準脳上の位置に変換し、解剖学的・機能的区分に基づいてラベルを付与することで、個人情報を排除した。被験者の視聴覚内容や運動内容は visual-semantic embedding model や deep neural network を用いることでグラウンディングを行なった。得られたデータベースを14施設20名のPIと共有し、解析を行なった。各共同研究について、成果が得られているところであるが、特に、頭蓋内脳波から視覚認知内容を推定する技術について、新たな Brain-Computer Interface の開発に繋がった。

(2) 詳細

研究テーマ A「頭蓋内脳波とライフログデータの同時計測とデータベースの作成」

本研究では、頭蓋内脳波計測と同時に様々なライフログデータの同時計測を行い、同データから個人情報情報を排除した脳ビッグデータの作成を行った。研究開始後より、既に 49 名の計測を行い、412 日分の頭蓋内脳波データを取得した。

1. 電極位置情報の変換方法の開発

頭蓋内脳波の電極位置や脳形状から個人情報情報を排除するため、標準脳上の脳活動に変換するためのプラットフォームを作成した。電極留置前の MRI 構造画像と電極留置後の頭部 CT 画像を重ね合わせ、脳表上の電極位置を推定した。さらに、これを標準脳上の脳活動に変換した。図1のように、各脳表上の電極座標に応じて球体モデル上の電極位置を計算し、これを球体モデルの標準脳に変換することで、対応する電極位置を同定し、この球体モデルから通常の3次元モデルに変換することで、標準脳上の電極位置を同定した。電極位置は個人の解剖学的特徴により不規則な配置となる。そこで、これらを統一的に解析するため、Human connectome project が開発した fMRI での脳活動領域に基づいたパーセレイションを適用し、各 ROI 内の電極を同一と扱うことで、共通の解析を実現した。これにより、電極位置の個人情報情報を排除し、匿名化してデータベースに保存した。また、被験者のライフログをグラウンディングし、同データベースへ保存した。

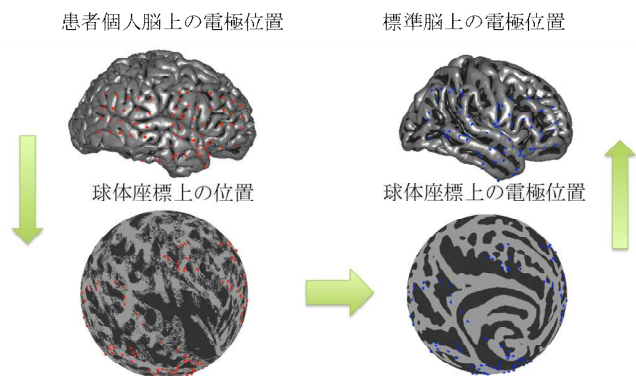


図 1: 個人脳上の電極座標を標準脳上の座標に変換する方法

2. 視聴覚データのグラウンディング

頭蓋内脳波と同期した動画視聴内容の計測を 607 時間分取得した。計測中に患者が視聴した動画を 1 秒毎の静止画に変換し、各静止画を 2 つの方法でグラウンディングした;(1)

visual-semantic embedding model を用いた方法、(2)Pre-trained DNN を用いた方法

(1)では各画像について 3-4 人の記述者(クラウド)により画像の意味を文章として記述してもらい、その文章から意味ベクトルを作成し、皮質脳波と同期したデータを形成した。作成した文章から MeCab で単語を抽出し、各単語について word2vec で 1000 次元の意味ベクトルを求め、各画像について、これらを平均したベクトルを各画像の意味ベクトルとした。

(2)の方法については、Pre-trained の DNN を用いて画像のグラウンディングを行った。ここでは、Xception 等の複数のモデルを用いた。上記の方法で用いた同じ画像について、それぞ

れの DNN で 1000 個のラベルを各画像につけた。各 DNN と人がつけたアノテーションとの間で 1000 次元のベクトルがどの程度相関しているかを Canonical correlation analysis (CCA) により評価した。その結果、幾つかのモデルについては、人手でつけたアノテーションと有意に相関することが示された。

研究テーマ B. 「脳ビッグデータを用いた Decoding 技術の開発」

動画を視聴中の脳活動について、各 1 秒間の α 、 β 、low- γ 、high- γ 帯域のパワーに変換し、これを特徴量として脳活動から意味ベクトルを推定した。Ridge regression を用いて推定を行ったところ、脳信号から推定したベクトルと元の意味ベクトルとの相関係数が、意味ベクトルの対応をシャッフルしたものと比較して有意に高い相関を得ることができた。また、各画像について脳活動から推定したベクトルと実際のベクトルとの相関を調べると、風景の画像や文字、人の顔などが映る画像では高い精度で脳活動から推定されることが明らかになった。皮質脳波が、これらの視覚的カテゴリー情報を持つことが示された (Fukuma et al., IEEE cybern. 2018)。

さらに、Ridge regression の重みを脳表にマップしたところ、視覚野を中心に主に γ 帯域のパワーに重みが強かった。多様な意味内容を予測するために必要な活動の分布をマッピングとして示すことができた。今後、てんかん手術などにおける機能マッピングとして応用できることが示された。

研究テーマ C 「他研究者とデータ共有」

得られたデータについて、14 施設 20 名の PI とデータ共有を行った。本さがけ CREST 内でも 3 名の PI と共同研究を行い、他にも国内外で共同研究を行なった。国外の研究者とは、英国、豪州等 7 名の研究者と、それぞれ別の共同研究を行なった。データ共有にあたっては、ライフログの内容やデータ長などのメタデータから、研究目的に合わせてデータを検索できる形を構築した。また、脳信号の一般的な特徴量とラベルをセットにしたデータを機械学習用のサンプルデータとして準備し、様々な手法をすぐに試せる形とした。

さらに、頭蓋内脳波データをより大規模に収集する枠組みとして、多施設共同で皮質脳波データベースを作成する研究を開始した。国内の 3 つのてんかんセンターに参加をしていただき、計測を開始した。本プロジェクトで開発した計測システムと解析パイプラインを共有し、3 施設で同じ計測が行える体制とした。脳ビッグデータを更に拡充し、新しい BMI 開発に貢献できると期待される。

3. 今後の展開

本研究成果より人の皮質脳波とライフログのビッグデータを作成するパイプラインが作成された。また、同データを用いることで、新しい脳情報解読技術の開発が可能になることが示された。今後は、多施設共同で脳ビッグデータを形成し、新しい脳情報解読技術の開発と、Brain-Computer Interface への応用、および、それらの医療応用を進めていくことが期待される。

4. 自己評価

本研究期間内において、目標としていた脳ビッグデータの作成とデータ共有を達成することができた。本研究はデータさがけとして本研究分野に参加することで、神経疾患領域でのビッグデータ利活用のための基盤技術の創出のために必要な、新規ビッグデータの作成と共有を推進した。特に、大阪大学医学部内に専用のクラスタ計算機を導入し、多施設共同でデータを集める体制を構築した。また、国内外の研究者と共同して、ライフログデータのグラウンディングなど、脳情報を抽出するために必要な技術開発を進めた。これにより、これまでは、個人情報の問題などで進んでいなかった脳ビッグデータを科学的に有用なハイスループットなデータとして作成し、安全に運用するための基盤が形成された。本研究領域が目標とするビッグデータ利活用の基盤技術開発において、当初の目標を達成したと考える。

また、本研究を通して、様々な研究者とのデータ共有を進め、幾つかの新規技術が創出された。特に、ビッグデータから脳情報を解読する技術については、JST CRESTの研究課題として開発が継続される。これは、筋萎縮性側索硬化症などを対象とした皮質脳波による意思伝達技術などに医療応用が期待される。また、脳波や脳磁図などの非侵襲的脳信号に応用することで、神経疾患の診断にも応用される。これらの成果から、脳ビッグデータが医療・社会における新たな価値を創造することが示された。本研究領域の目的でもある、次世代基盤技術の創出・高度化・体系化にも貢献したと考える。今後、これらの技術を発展させ医療応用することで、社会・経済に大きな波及効果を持つと期待される。

5. 主な研究成果リスト

(1) 論文(原著論文)発表

著者, 発表論文タイトル, 掲載誌名, 発行年, 巻号, 始頁-終頁, その他
1. Fukuma R*, <u>Yanagisawa T</u> **#, Tanaka M, Yoshida F, Hosomi K, Oshino S, Tani N, Kishima H, Real-time neurofeedback to modulate β -band power in the subthalamic nucleus in Parkinson's disease patients, <i>eNeuro</i> , 2018, in press, *Contributed equally to this work, #Corresponding to this work
2. Nakanishi Y*, <u>Yanagisawa T</u> *, Shin D, Kambara H, Yoshimura N, Tanaka M, Fukuma R, Kishima H, Hirata M, Koike Y, Mapping EGoG channel contributions to trajectory and muscle activity prediction in human sensorimotor cortex., <i>Scientific Reports</i> . 2017, 7, 45486, *Contributed equally to this work
3. <u>T. Yanagisawa</u> #, R. Fukuma, B. Seymour, K. Hosomi, H. Kishima, T. Shimizu, H. Yokoi, M. Hirata, T. Yoshimine, Y. Kamitani, Y. Saitoh, Induced sensorimotor brain plasticity controls pain in phantom limb patients, <i>Nature Communications</i> . 2016, 7, 13209, #Corresponding to this work
4. Edakawa K*, <u>Yanagisawa T</u> *, Kishima H, Fukuma R, Oshino S, Khoo HM, Kobayashi M, Tanaka M, Yoshimine T., Detection of Epileptic Seizures Using Phase-Amplitude Coupling in Intracranial Electroencephalography., <i>Scientific Reports</i> . 2016, 6, 25422, *Contributed equally to this work
5. Fukuma R, <u>Yanagisawa T</u> *, Saitoh Y, Hosomi K, Kishima H, Shimizu T, Sugata H, Yokoi H,

Hirata M, Kamitani Y, Yoshimine T, Real-Time Control of a Neuroprosthetic Hand by Magnetoencephalographic Signals from Paralysed Patients, *Scientific Reports*. 2016, 6:21781
*Corresponding to this work

(2)特許出願

研究期間累積件数:2件(公開前の出願件名については件数のみ記載)

1.

発明者: 貴島晴彦、吉峰俊樹、柳澤琢史、枝川光太郎、福間良平

発明の名称: Cross-frequency coupling によるてんかん発作検出

出願人: 国立大学法人大阪大学・国立研究開発法人情報通信研究機構

出願日: 2016/1/19

出願番号: 特願 2016-008301

(3)その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

受賞(期間内に4つ)

1. バイオインダストリー奨励賞, 2017/10/11, BioJapan2017, バイオインダストリー協会、
2. BCI award, BCI conference, 2017/9/21
3. 中谷賞奨励賞 中谷医工計測技術振興財団、2017/2/24

プレスリリース(全35件)

1. 「幻肢痛」脳研究で解き明かす, NHK ニュースほっと関西, 2016年10月27日
2. 幻の手の痛み軽減に成功、阪大など 治療法開発に期待, 朝日新聞, 2016年10月

招待講演(国内14、国際6)

研究報告書

「ヒト腸内環境ビッグデータ」

研究タイプ: 通常型

研究期間: 平成 27 年 10 月～平成 31 年 3 月

研究者: 山田 拓司

1. 研究のねらい

本研究の目的は、大規模なヒト腸内環境データの蓄積と公開である。これまでの疫学研究において食事や生活習慣から直接類推されてきた発がん性や疾病のリスクを、ヒト腸内細菌叢の動物的変動という観点から定量化することを目的として、500名以上の日本人健常人および大腸がん罹患患者に対する腸内環境のメタゲノム、メタボローム、臨床データ、生活習慣データ、食習慣データを蓄積しているのを、これを公開する。これらのデータは個人毎にその腸内環境に対して数値データ、カテゴリカルデータ、順序変数など質的に異なるデータ群であり、公開による多方面からの活用方法の開発により、更なる利用を期待することができる。ヒト臨床情報を含むため、公開は条件付き公開としている。

腸内細菌叢はもう一つの臓器と言われ、その乱れが重篤な疾患の原因になることが示唆されている[Guarner F. *et al.* Lancet. 361(9356):512-9 2003]。近年、糞便中に含まれる微生物由来のDNAを抽出し、その遺伝子配列を網羅的に明らかにしていくメタゲノム解析法が開発され、腸内細菌を網羅的に解析可能となった。これにより、関連研究が飛躍的に進み、大腸がん、肥満、炎症性腸疾患などの疾患に対して、腸内細菌が影響を与えているという報告がなされている。ピロリ菌が胃がんの直接的な原因菌であったという発見から、様々な腸疾患に対しても特定の原因菌の存在が示唆され、その探索が世界各国の研究機関でおこなわれている。また、大規模な疫学研究(コホート研究)も近年盛んに行われており、数万人規模での生活習慣やゲノムデータに至る大規模なデータが蓄積されつつある。国立がん研究センターによる多目的コホート研究では発がん性や疾病のリスクについて追跡コホート研究がなされており、タバコと肺がん、紫外線と皮膚がんなど、明確な関連性がある様々なものがリスクファクターとして提示されている。しかしながら、大腸がんをはじめとする多くの腸疾病については、食生活がリスクファクターとして重要であり、疾病との関連性が示唆されているものの、食事及び生活習慣と疾病の関係性の原因については不明確な例がほとんどである。

本研究では上記のような腸疾患に対する環境リスクを定量化するため、多くの病院、企業、研究機関の協力のもと、Japanese Consortium for Human microbiome (JCHM, <http://www.jchm.jp>)というコンソーシアムを立ち上げ、日本人の健常者及び腸疾病患者の腸内メタゲノムデータ、付随する疫学データ及び臨床所見を収集している。本データはどのような食事や生活習慣が、どのように腸内細菌を変動させ、その結果、疾病へと導いていくのかをモデル化することを目指すものであり、本研究ではこれを広く公開し、更なる利用を促進するものである。

2. 研究成果

(1) 概要

本研究の研究成果を以下の3つに大別する。すなわち、1)データベースエントリーとメタデータオントロジーの策定、2)疫学コホートデータベースシステムの構築、3)継続的なデータ取得とデータベースの拡充、である。

1)のデータベースエントリーとメタデータオントロジーの策定に関しては、ヒト腸内環境に付随するデータが数値データとカテゴリカルデータが混在している複雑な構造を持つため、それを記述するためのフォーマットの策定を行った。2)のデータベースシステムの構築に関しては、公開用データの作成とデータアクセスのためのwebベースUIの構築を行った。3)のデータベースの拡充については、外部期間との連携によるデータベースコンテンツを拡大していくことに相当する。以下、それぞれのテーマについての結果詳細を報告する。

(2) 詳細

【研究テーマ1:データベースエントリーとメタデータオントロジーの策定】

本研究で扱うヒト腸内環境—検体分に含まれるデータは糞便由来メタゲノムデータ、メタボロームデータ、内視鏡検査による医師の所見、そして食生活習慣の疫学データである。そこには細菌種(約1000種)、遺伝子(約1000万)、低分子化合物(約200種)、内視鏡所見(8つの大腸がんステージ)、そして食生活習慣(475項目)が含まれる。これらのデータはそれぞれが特定個人に紐付いている。これらを全てテキストデータ、csv形式、json形式、リレーショナルデータベース(RDB)のデータテーブルとして保持するためのデータフォーマットを策定した。メタゲノムデータ、メタボロームではこれまでの先行研究により、ある程度標準の表記方法があるため、それを利用している。内視鏡所見とアンケートによる疫学データは非常に複雑であったため、各項目を数値データ、カテゴリカルデータとして分割し、それぞれを別に構造化を行った。

【研究テーマ2:コホートデータベースの構築】

様々な研究で本データを利用することを想定し、本研究では様々な提供データフォーマットを準備した。リレーショナルデータベース(RDB)のデータテーブルの構造は、テーマ1)におけるデータ型定義に従い構築した。また、RDBを基盤としてwebベースのUIを構築し、データへの容易かつ安全なアクセスを実現している。また、データベースシステム自体も移管可能な構造をもっており、別の環境に容易に移管することができる。その場合はデータユーザーが独自データを追加することも可能であるため、秘匿性の高い独自データを本データと統合して解析に用いることも可能である。

【研究テーマ3:継続的なデータ取得とデータベースの拡充】

本研究で当初公開予定であった500症例に加え、全期間を通じて国立がん研究センターの協力を得て、糞便サンプルの収集とアンケートによる疫学データの収集を進行してきた。その結果、2000症例以上を新規に追加することができ、内視鏡所見、疫学データに関しては2783症例がデータ化されている。メタゲノムデータは1051症例分、メタボロームデータは517症例に達している。961名においては経時的に検体を収集しており、前向きコホート研究としても利用

することが可能である。

3. 今後の展開

本データは世界的に見ても非常に巨大なデータベースとなっている。特に、臨床初見、メタボロームデータが紐づいているヒト腸内環境オミクスデータは独自のデータであり、今後はこのデータの利活用を進める予定である。具体的には、大腸がん発症に関わるメカニズムの解明や薬剤動態に影響を及ぼす腸内環境の推定を予定している。また、まだデータベースには格納していないカルテ情報がある。どの薬剤を飲んでいるか、血液データなども病院には保持されているため、今後はそのデータもデータベースに入れ込む予定である。

本データは 1000 名以上の健常人を含むが、2000 名近くの大腸がん患者データである。大腸がんに関わる腸内環境研究は世界的な課題であるため、各国の研究グループと連携してデータを共有し大腸がん発がんの機構を明らかにするための国際連携を進めている。また、本データは創薬基盤を含む産業応用可能性が非常に高い。学術分野のみならず一般の企業体も本データを利用できる形とそのしくみ作りを早急に進めていきたい。

4. 自己評価

本研究の目的はこれまで蓄積してきたデータの公開であり、その結果は十分に実現ができた。さらに、500 名を予定していた検体数は 2700 名分となり、当初予定の5倍を超えている。2019 年の時点において腸内環境データとしては国内では最大規模であり、国外と比較しても大腸がん患者に関するコホートとしては世界最大である。データを公開するためのデータベースシステムも構築することができ、安全にデータを公開することが実現している。

蓄積データを利用したデータ解析において、大腸がん発がんマーカーに関する特許を出願も行うことができ、データベースとして十分に活用することが可能であることも示すことができた。

また、蓄積されたデータを国外の大腸がん研究グループと一部のデータを共有することで国際連携も進めている。本プロジェクトで構築した大規模データの価値はその質、量ともに国際的にも評価されており、今後も国際連携を継続いく予定である。

5. 主な研究成果リスト

(1) 論文(原著論文)発表

1. Hayashi T, Yamashita T, Watanabe H, Kami K, Yoshida N, Tabata T, Emoto T, Sasaki N, Mizoguchi T, Irino Y, Toh R, Shinohara M, Okada Y, Ogawa W, Yamada T, Hirata KI. Gut Microbiome and Plasma Microbiome Related Metabolites in Patients With Decompensated and Compensated Heart Failure *Circ J*. 2018 Dec 25;83(1):182-192.
2. Yoshida N, Emoto T, Yamashita T, Watanabe H, Hayashi T, Tabata T, Hoshi N, Hatano N, Ozawa G, Sasaki N, Mizoguchi T, Amin HZ, Hirota Y, Ogawa W, Yamada T, Hirata KI. *Bacteroides vulgatus* and *Bacteroides dorei* Reduce Gut Microbial Lipopolysaccharide Production and Inhibit Atherosclerosis. *Circulation*. 2018 Nov 27;138(22):2486-2498

3. Watanabe H, Nakamura I, Mizutani S, Kurokawa Y, Mori H, Kurokawa K, Yamada T. Minor taxa in human skin microbiome contribute to the personal identification. PLoS One. 2018 Jul 25;13(7):e0199947
4. Darzi Y, Letunic I, Bork P, Yamada T. iPath3.0: interactive pathways explorer v3. Nucleic Acids Res. 2018 Jul 2;46(W1):W510–W513.
5. Yuichiro Nishimoto, Sayaka Mizutani, Takeshi Nakajima, Fumie Hosoda, Hikaru Watanabe, Yutaka Saito, Tatsuhiro Shibata, Shinichi Yachida, Takuji Yamada. High stability of faecal microbiome composition in guanidine thiocyanate solution at room temperature and robustness during colonoscopy Gut, 2016 Sep;65(9):1574–5.

(2) 特許出願

研究期間累積件数: 1件(公開前の出願件名については件数のみ記載)

(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

1. Takuji Yamada, Naoki Tsukuda, Tomonori Takahashi, Hiroshi Mori, Ken Kurokawa, Yuki Moriya, Shujiro Okuda, Susumu Goto Metabolic pathway database for Human gut microbiome 第 67 回日本生物工学会大会、城山観光ホテル(鹿児島) 2015/10/26
2. Takuji Yamada, Fecal metagenomic analysis of colorectal cancer cohort 第 74 回日本癌学会学術総会、名古屋国際会議場 2015/10/08
3. Takuji Yamada, ヒト腸内細菌叢メタゲノム解析と疾病との関連 脂質栄養学会第 25 回大会(日本、秋田) 2016/09/17
4. 山田 拓司, Human gut microbiome associated with colorectal cancer 第 76 回日本癌学会学術総会(日本、横浜) 2017/9/28–30
5. 細菌が人をつくる(書籍、TED ブックス) 単行本、ロブ・ナイト(著)、ブレンダン・ビューラー(著)、山田拓司(翻訳)

さがけ「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化」研究領域 研究総括：喜連川 優（国立情報学研究所 所長／東京大学 生産技術研究所 教授）、副研究総括：柴山 悦哉（東京大学情報基盤センター 教授）

学習データ分析で教育が変わる

島田 敬士（九州大学 大学院システム情報科学研究院・准教授）

研究課題名：「時空間粒度の異なる教育ビッグデータの非同期ストリーム処理基盤の構築」

研究期間：2015.10～2019.3

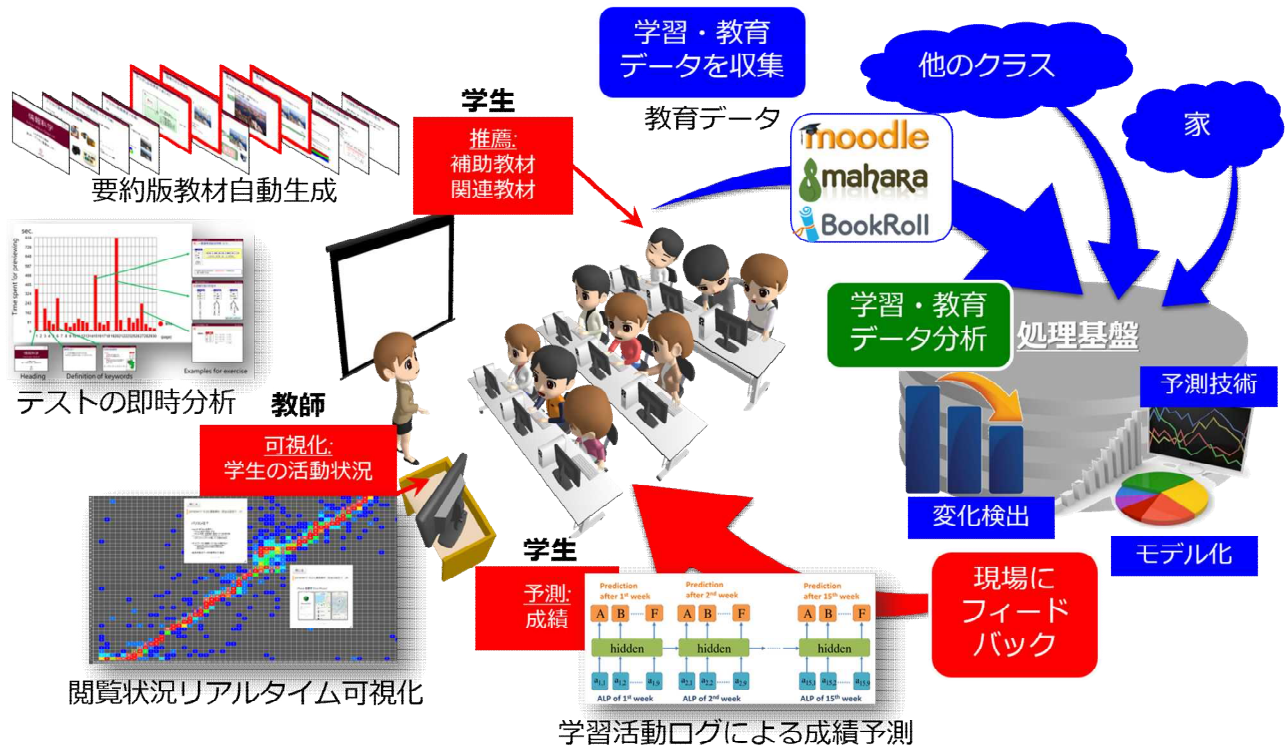


図 教育の情報化

デジタル学習環境を利用して教育現場から出席状況、小テストや課題の結果、教科書の閲覧状況などの様々な学習・教育活動の記録を収集することができます。その記録データを分析することで、学生の成績を予測したり、学生に講義内容の理解を助ける補助教材を提供したり、教師に学生の状況をレポートしたりすることができるようになります。

教育の現場に ICT(情報通信技術)の導入が広がっています。本研究では、PC 必携化(全学生が個人の PC を講義に持ち込み講義を受けるスタイル)の大学で、教育現場から学生の学習活動や教師の教育活動の記録(学習ログ)を大学規模で収集しています。学習ログを細かく記録するために、出席状況を収集したり小テストを実施したりできる学習管理システムや、教科書を電子的に配信できるデジタル教科書システムを導入しています。これらのシステムで収集される学習ログを活用して、学習困難に陥っている学生を早期に発見したり、学習を支援するための教材を自動生成したりすることができる技術を開発しています。また、講義中の学生全体の状況をリアルタイムに分析し

てその結果を教師にフィードバックすることで、教師が現場の状況に応じて柔軟な講義進行を行えるシステムの開発も行っています。今後は、初等中等教育の現場への展開を進めていきます。

>>参考情報

➤ 論文

Atsushi Shimada et al., Real-time learning analytics system for improvement of on-site lectures, ITSE, Vol.15, No.4(2018)

➤ プレスリリース

「九大、学習支援にビッグデータ、理解度把握、きめ細かく」(2017年1月)

https://www.nikkei.com/article/DGXLASJC24H3N_U7A120C1LX0000/

さがけ「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化」研究領域 研究総括：喜連川 優（国立情報学研究所 所長／東京大学 生産技術研究所 教授）、副研究総括：柴山 悦哉（東京大学情報基盤センター 教授）

人の脳活動の詳細なビッグデータを作成

研究者 柳澤 琢史（大阪大学 高等共創研究院・教授）

研究課題名：「思考・行動を予想する脳ビッグデータ（データさがけ）」 研究期間：2015.10～2019.3



図 脳ビッグデータの概念図

人の普段の行動や思考内容、また見たり聞いたりした内容などを詳細な脳信号とともに計測しビッグデータを形成します。このデータから個人情報排除して脳ビッグデータを作成します。これを様々な手法で解析することで、人の行動や思考、知覚情報などを脳信号から推定する技術や、精神神経疾患の診断治療技術を開発します。

への応用が期待されます。

情報通信技術の発展に伴って、人が普段生活する際の詳細な脳活動状態やライフログデータを大量に得ることが可能となりつつあります。これを活用することで、様々な精神神経疾患を診断したり、新しい治療法を開発できると期待されます。本研究では、そのようなビッグデータを、多くの研究者が共有して新しい技術開発に応用するために、脳ビッグデータを形成する技術を開発しました。特に、脳神経外科の治療目的で脳表面から詳細な脳信号を計測される患者さんの協力を得て、様々な運動内容や視聴覚内容、思考内容などのライフログデータと同期して高精度な脳波を得ることに成功しました。さらに、これらのデータを人工知能技術で解析することで、人の行動や知覚内容などを脳信号から解釈したり、脳波からてんかん発作を正確に同定する技術などを開発しました(参考情報)。今後、新しい情報通信技術や診断・治療法など

>>参考情報

> 論文

1. Edakawa K*, Yanagisawa T*, et al., *Scientific Reports*. 2016, 6, 25422

> 特許出願

1. Cross-frequency coupling によるてんかん発作検出, 2016-008301 (2016)

> 受賞

1. 「バイオインダストリー奨励賞」(2017)
2. 「BCI award」(2017)
3. 「中谷賞奨励賞」(2017)

> プレスリリース

1. 「てんかん発作時の特徴的な脳波を検出」(2016年5月)

<https://resou.osaka-u.ac.jp/ja/research/2016/20160513>

さがけ「ビッグデータ統合利活用のための次世代基盤技術の創出・体系化」研究領域 研究総括：喜連川 優（国立情報学研究所 所長／東京大学 生産技術研究所 教授）、副研究総括：柴山 悦哉（東京大学情報基盤センター 教授）

腸内細菌で変える未来 （注：さがけ研究課題名ではなく、研究内容がわかるタイトル）

研究者山田 拓司（東京工業大学 生命理工学院・准教授）

研究課題名：「腸内環境ビッグデータ」 研究期間：2017.10～2019.3

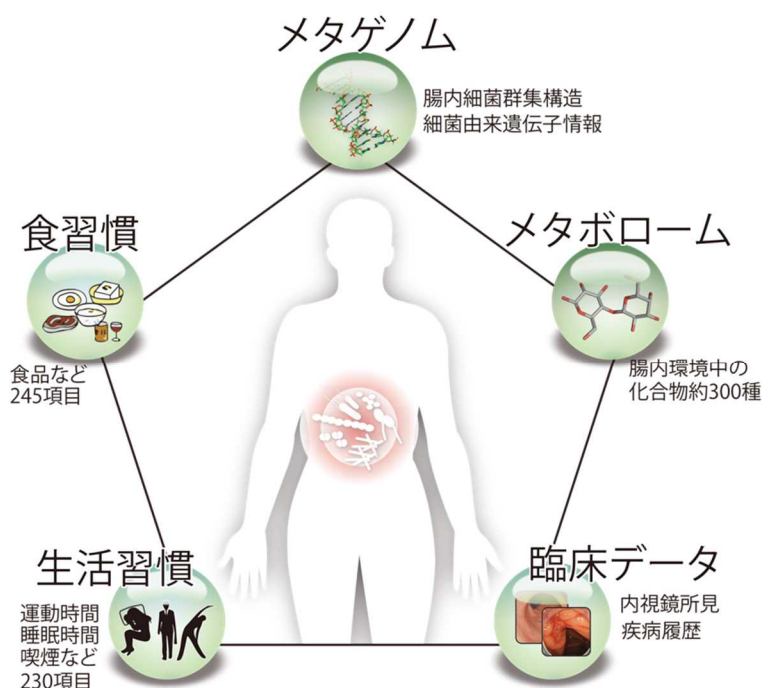


図 腸内環境をとりまくビッグデータ

ヒトの腸内環境に関わる様々な形のデータ。それぞれの個人に対して膨大なデータが存在している。

ヒトの腸内には数多くの細菌が我々とともに共生しています。これらの細菌は「ヒト常在菌」と呼ばれており、それらが形作る群集がヒト自体に様々な影響を与えることから、「もう一つの臓器」と言われています。その群集の乱れが重篤な疾患の原因になることもわかってきました。近年、便中から細菌由来のDNAを取り出し、そのDNA配列を網羅的に解析するメタゲノム解析法が開発され、腸内にどのような細菌がどの程度共生しているかを網羅的に明らかにできるようになりました。これにより、関連研究が飛躍的に進み、大腸がん、肥満、炎症性腸疾患などの疾病に対して、腸内細菌が与

える影響がわかってきました。

ピロリ菌が胃がんの直接的な原因菌であったという発見から、それ以外のヒトに常在する様々な細菌が、別の病気に対しても原因菌になっているのではないかと示唆されています。ヒト常在菌研究を基盤に、そのような原因菌探索が世界各国の研究機関でおこなわれています。

本研究では上記のような腸内環境を対象にした研究の基盤となるデータを作り、そして広く公開することを目的としています。多くの病院、企業、研究機関の協力のもと、日本人の健常者及び腸疾患患者の腸内細菌メタゲノムデータや腸内代謝物質、アンケート情報、付随する疫学

データ及び内視鏡所見のような臨床データを網羅的に収集しています。特に大腸がん患者由来のデータを広く集めています。国立がん研究センターによる研究では、タバコと肺がん、紫外線と皮膚がんなど、明確な関連性がある様々なものがリスクファクターとして提示されています。しかしながら、大腸がんについては、食生活がリスクファクターとして重要であり、疾病との関連性が示唆されているものの、食事及び生活習慣と疾病の関係性の原因については不明確な例がほとんどです。本データはどのような食事や生活習慣が、どのように腸内細菌を変動させ、その結果、病気へと導いていくのか明らかにすることを旨とするものです。