

高精度教育ビッグデータをベースとした教育支援の公教育への導入推進

国立大学法人岡山大学

- マイクロステップ・スケジューリング技術によって収集される、高精度教育ビッグデータを利用する新型 e-Learning システムのメリットを最大化することで、マイクロステップ・スタディ (MSS) の社会実装を推進した。
- 六つの開発テーマに取り組み、そのうち四つのテーマについて当初の目標が達成できた。
- データを共有する仕組みやデータの管理方法、共有ルール、関連する教育データの共有ルールなど、検討しなければならないことを明確化した。

1 研究の目的

本研究開発は、個々の学習者の学習成績の上昇を可視化してフィードバックできる新型 e ラーニング「マイクロステップ・スタディ (MSS)」のメリットを最大化し、社会実装を推進することまでを目標とした。これは、マイクロステップ・スケジューリング技術で収集される、「いつ」という時間軸上の条件がそろった大量の縦断的行動データ (高精度教育ビッグデータ) の利用によって実現されるものである。また、学力上位層の英語力を向上させることだけを目的とせず、MSS のサービスを受ける学習者の意欲向上と学力向上に寄与することを最大の目的とした。学習効率を向上させる「個別最適化」の機能の実装や、文部科学省の GIGA スクール構想で配備された端末を学校現場と自宅で有効活用する上で生じるネットトラブル、および通信費や経費的な問題などを解決できる端末ソフトの開発から、ゲーミフィケーションの要素を学習に導入する方法の検討とモデル実践による明確な成果の取得まで、さまざまな目的を設定した。

2 実施期間と方法

(1) 実施期間

2018 年度～2020 年度

(2) 実施方法

六つの開発テーマを設定し研究開発を実施した。なお、SIP 事業は 2020 年度で終了したが、2021 年度以降も本開発テーマの研究開発を継続している。

1) 開発テーマ1: 高度な学習効果を生むための学習データ・情報プラットフォームおよび学習支援システムの構築と運用

MSS において、学習内容をスケジューリングすると同様に学習や指導の効果を可視化できるよう、学習者の学習意欲や自己効力感などの心理尺度を同時にスケジューリングし、日々の学習が終了するごとに図 1 のような心理尺度の質問項目に数個回答を求める仕組みを年間を通じて運用した。

なお、MSS の導入による教育支援の社会実装を全国に広げるため、導入費用の一部を受託事業費として大学に納める仕組みを、本研究開発以前の 2014 年度から自治体に対して採用してきた。

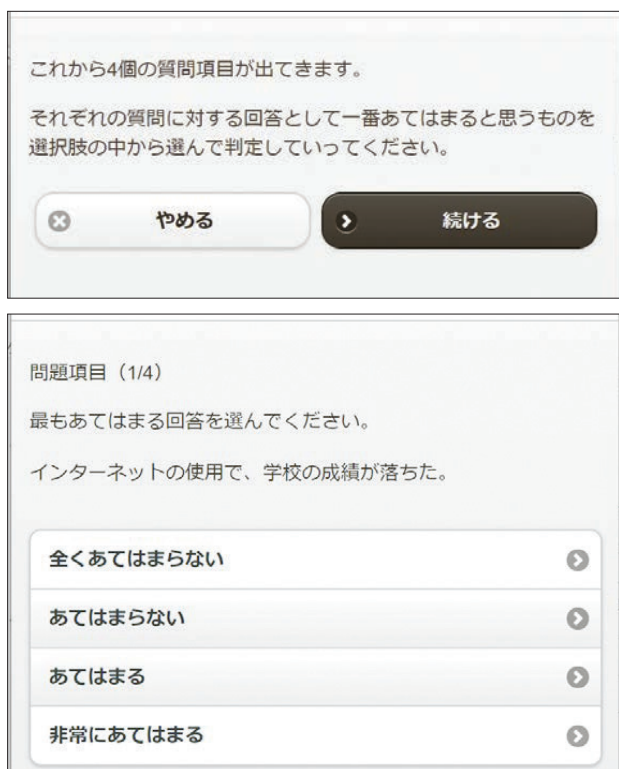


図1 スケジューリングされた心理尺度に対して回答するインターフェースの例

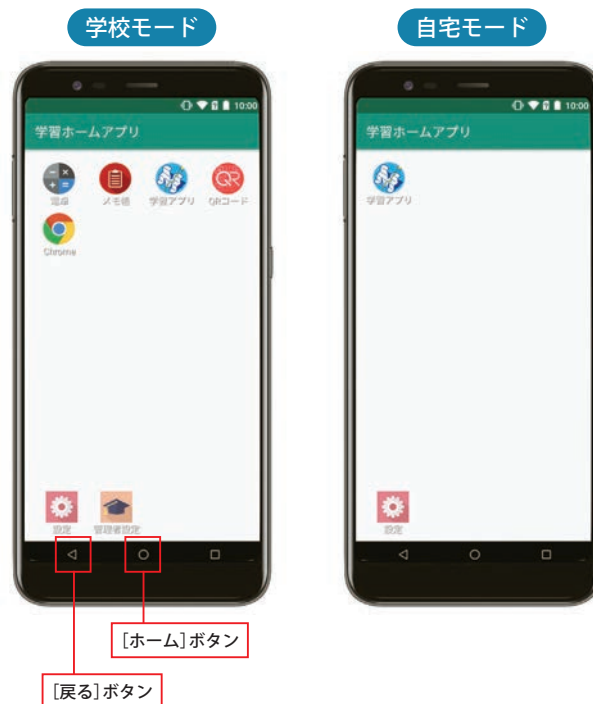


図2 学校と学校外でネット環境が切り替わるエリア依存処理を実装した端末の画面例

2) 開発テーマ2：プライバシー保護を考慮したセキュアな学習データ・情報プラットフォームの構築と運用

どこでも使えるインターネットの特徴に制限をかけ、場所という情報に応じてセキュリティポリシーを変更したり、利用できるアプリを制限したりするなど、さまざまな処理が自動的に実行される仕組みを端末に導入した(図2)。それにより、教師や保護者が利用を許可したアプリのみが端末画面に表示され、学習以外に端末を利用できない仕組みが作り出されたほか、学習だけで通信を利用する状況が構築できたことで、非常に安価な料金プランをメーカーが設け、実際に提供できた。学習者が操作できるインターネットを安全に、安価に教育利用できる環境を構築した。

3) 開発テーマ3：テラーメイド学習を促進する高度な学習支援システムの実現

マイクロステップ・スケジューリング技術により、学習とテストのタイミングを制御しつつ、学習効果を連続測定することで、従来のテストでは不可能であった実力レベルの習得度の正確な推定が可能になる。その推定を用いて、完全に習得できたと判断された内容(英単語など)を学習から排除していく、個別最適化の処理を可能にした。

4) 開発テーマ4：教育現場や多様な学習の場などと認知科学、心理学などの研究開発を融合させた実践に基づく新たな学習モデルの構築と社会への展開

1か月に1回ずつの学習を7か月継続(計7回学習)して1か月後にその効果を測定した場合と、1日に7回学習を繰り返し、1か月後にその効果を測定した場合を比較したところ、前者の方がおよそ4倍学習効果が大きいことが明らかになった。英単語の場合、1日に6回以上の反復学習は実力に影響しない(積み上がっていない)ことも明らかになっている。従来から、集中学習に比べて分散学習の効果が大きいといわれているが、実際の教育コンテンツを用い、半年以上の学習を対象にして、学習回数の影響を具体的に検討できるようになったことを意味しており、その意義は大きいと考えられる(図3)。

5) 開発テーマ5：コンテンツ企業が情報企業に変わるモデルの明示および人材育成

実社会で実際に用いられている学習コンテンツのバリエーション増大を意図し、複数の教材会社から提供してもらった学習コンテンツにスケジューリングを施しながら、多数の学習者にMSSを提供し、それぞれの学習内容ごとに大量の学習データを収集することを目指した。

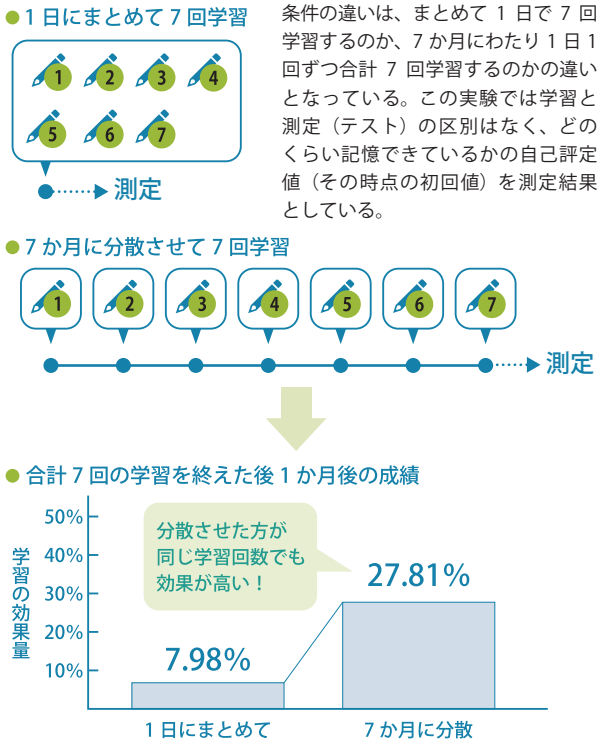


図3 計7回の学習を1日で終わらせた場合と7か月に分散させた場合における、学習の1か月後に測定された成績の比較

6) 開発テーマ6：高精度教育ビッグデータの運用を管理する仕組みの確立

小中高校大学などで、7,000人規模の実証実験を行って開発した機能を検証する仕組み、および教師など指導者がクラスや学校など、学習者集団の学習状況を把握できる仕組みを設計した。ゲーム要素を実装した学習・フィードバックシステムのモックを完成させ、ネットを利用しないデモデータを使ったデモなどを実施した。

3 本研究の成果

(1) 開発テーマ1の評価

さまざまな学習データを、大規模に取得・バックアップできるストレージシステム、および学習データベースを2020年度までに構築し、さらに意識調査を一元的に実施。縦断データを完全な匿名データとして提供できる情報プラットフォームを、2022年度までに構築した。

また、学習データを全て回収・記録するシステムを完成させ、2022年度には、1万2,000人を超える学習者を対象にMSSを提供。実際の運用には至らなかったが、収集されるデータをさらに集約して匿名化処理を実施したデータを生成し、そのデータから必要なデータを抽出

できるAPIの試作版を開発した。

基盤システムの開発はほぼ計画通り実施でき、目標以上の成果を達成できたと評価する。

(2) 開発テーマ2の評価

Tコード通信原理という新たな特許技術を利用し、フィードバックデータとメールアドレスなどを学習・フィードバックシステムに置かずに、学習者が個別にフィードバックデータを受け取れる仕組みを2021年度までに構築する計画であった。

Tコードの特許利用については権利企業から利用の内諾が得られたが、Tコード通信原理が導入できなかったため、当初の目的は十分達成できなかったと評価する。

(3) 開発テーマ3の評価

学習者の反応を定点観測し、実力を正確に測定して、その時間変動データから学習問題ごとの到達度を正確に推定する方法はすでに構築しており、その方法をベースに、実力レベルになり学習がもう必要ないと推定された問題を特定し、学習対象から外していく仕組み（個別最適化処理）を本プロジェクトで試行した。

個別最適化の仕組みを2019年度までにシステムに実装し、2020年度は約7,000人の学習者に対して導入したことで、当初の目的は達成できたと評価する。

(4) 開発テーマ4の評価

潜在記憶に関する最新の知見を広く周知するため、論文や書籍の公刊、シンポジウムの開催などを行った。記憶の長期持続性に関する新たな知見と、MSSの導入によって得られた知見を広く周知するため、2021年3月3日に成果報告会を兼ねたフォーラムをオンラインで開催し、多くの参加者を得た。それらの一般の方々への周知を目的として書籍を出版した。フォーラム参加などをきっかけにして、MSSの導入依頼が2021年度に数件増えて、当初の目的は達成できたと評価する。

(5) 開発テーマ5の評価

現在のロイヤリティのビジネスモデルから、情報サービスを提供する企業に生まれ変わることをモデルとして示すため、2020年度には企業から提供されるコンテンツを用い、新型eラーニングを開始する計画で、提供されるコンテンツを使ってスケジューリングを開始したが、それを活用した実証実験はできなかった。

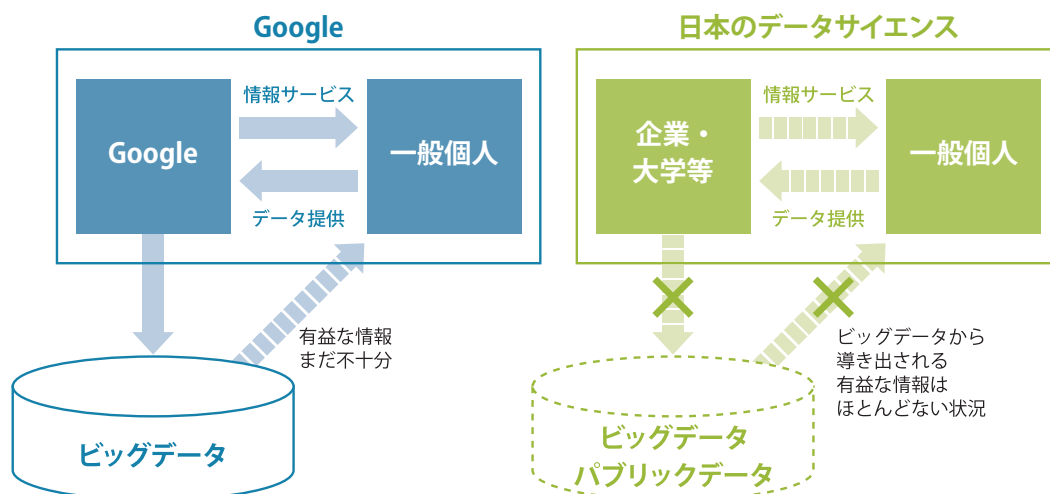


図4 一般個人が必要とするサービスがなければ行動データは収集できない

一方、コンテンツ企業ではなく、私立学校において学習システムの改良を独自財源で実施し、小学校低学年向けのルビが表示される学習システムを開発。2021年度からそのシステムを使ったMSSを提供しているほか、2022年度には公立の小学校でのシステム利用が始まっており、目的はおおむね達成できたと評価する。

(6) 開発テーマ6の評価

収集される高精度教育ビッグデータの利用によって生み出されるメリットを、教育サービスなどに還元する仕組みとして、サービスを提供するベンチャーを2020年度に設立し、データ利用に関するステークホルダーが明確になると考える研究開発の後半（2021年度）に、データ利用のルールや制約などを検討する組織設立の準備組織を立ち上げる計画であった。ベンチャーは設立に至らなかったが、データの共有に関しては、当初想定していた以上の課題に気づくことができ、データの管理・運用を担う組織のあり方の検討に有益な情報が得られた。ベンチャー設立自体は容易であるが、当初想定していなかったデータ活用の制約などを考慮して慎重に設立していく必要があることは間違いなく、データの管理・運用のあり方を深く検討できた点で、重要な一歩を踏み出したと評価する。

理と難易度調整による学習効率の向上」、「一般向けの広報、成果報告会（フォーラム）の実施による社会実装の推進」などを通じて、2019年度には、第16回日本e-Learning大賞「文部科学大臣賞」を受賞した。2021年度末の時点で、約1万2,000人の学習者にMSSを提供し、高精度教育ビッグデータの集約を実現した。

AIやデータサイエンスに対して、本研究開発およびこれまでに実現してきたe-learningプラットフォームは大きなメリットを提供することになる。GAF（Google、Amazon、Facebook、Apple）に代表される企業が、世界のICTを牽引している力の源泉は、それぞれの企業が、一般個人に大きなメリットを提供できるサービスを持っていることにある（図4）。特に、21世紀に入って急激に性能を上げたAI技術は、人間と同様、さまざまな情報を学習（経験）することによって成長していく原理に従っている。すなわち、学習（経験）させるデータがなければAIはつくれる。本プロジェクトで収集された膨大な縦断的データを分析することで、学習法や指導法に関わる新しい知見は多数手に入ると考えられる。

今後、手に入れられた成果、そしてすでに蓄積されている大量の高精度教育ビッグデータの解析結果を発表していくことで、社会実装はさらに大きく広がると考えている。一方で、収集されるデータは個々の学習者の生活に直結しており、個人が特定される危険性も高く、データ活用による新たな格差や問題を生み出す可能性がある。データを共有する仕組みやデータの管理方法、共有ルール、関連する教育データの共有ルールなど、検討しなければならないことが多数あることも指摘しておきたい。

4 まとめと今後の展望

本研究開発では、「ネット利用を制御できる安全な端末アプリの開発」、「ゲーム要素の実装」、「個別最適化処

AIを活用したヒューマン・インタラクションによる取り組み

認知症の本人と家族の視点を重視するマルチモーダルなヒューマン・インタラクション技術による自立共生支援 AI の研究開発と社会実装

株式会社エクサウィザーズ、国立大学法人静岡大学

- 介護サービスのバリューチェーンにおける、AI 利活用の機会に関わる状況、およびベネフィットを整理し、AI 技術が介護分野において受け入れられるための感情的基盤の構築を目指して「科学的介護」モデルの部分的な深化を行った。
- 市町村などの自治体が有する介護レセプトや、介護認定情報などを活用した「時系列予測 AI」による将来の要介護度などに関する予測モデルを開発した。
- 国内各メディアにおける研究・開発成果に関わる記事の掲載や、開発した予測モデルの海外論文プラットフォームへの投稿など、国内外へ積極的に普及活動を行った。

1 研究の目的

近年の高齢化の進展に伴う社会保障費の増大を抑制するためにも、介護領域における AI 利活用の重要性が高まっている。AI を利活用した介護サービス・ソリューションを提供する事業者は増加しているが、「政府・自治体」、「病院・介護施設」、「AI 企業・研究機関」の協働をトップダウンで支援するアプローチはまだ十分ではない。本研究開発では、医療・介護に関わるビッグデータを AI と組み合わせ、認知症のある人の自立を支援するため、「政府・自治体」、「病院・介護施設」、「AI 企業・研究機関」、「認知症の人を中心とした本人・家族・介護者」の協働による成功事例を創出した。さらに、その成果をオープン化・横展開することで、社会システム全体で認知症当事者の QOL 向上と社会保障費の低減を目指した。

2 実施期間と方法

(1) 実施期間

2018～2020 年度

(2) 実施方法

1) 基盤系のテーマ

「戦略基盤」については、2018 年度に介護サービスのバリューチェーン（未病～ケア～看取り）における AI の利活用を整理した。その結果、バリューチェーン横断での取り組み、およびヒューマン・インタラクションの AI 化については、十分検討されていない状況であった。そこで、「政府・自治体」、「病院・介護施設」、「AI 企業・研究機関」が連携・協力し、エビデンスに基づく科学的介護を介護者に提供することを実現する方向性を定めた（図 1）。2019 年度は、AI 技術が介護分野で受け入れられる感情的基盤の構築を目指し、「文脈独立」、「文脈依存」の概念のケア領域への応用を指針として「科学的介護」モデルの部分的な深化を行った。

「データ基盤」については、2018 年度に基礎自治体*1 と実証実験の実施に向けて意見交換・交渉を開始したが、「将来の介護分野における AI の活用促進にあたり、AI 技術の開発に向けたデータの集積が何よりも重要である」との声が多かったため、「介護に関するデータ活用」の先進的事例の構築を通じた、国内の介護データ基盤整備に関わる機運醸成を目指すことにした。2019 年度は神奈川県下の自治体と連携し、当該自治体を持つ実際の介護レセプトデータなどを活用し、将来の要介護度に関

わる予測モデルを開発して予測を実施した。

「評価基盤」については、「各取り組みが人の身体的に与える効果」や「各取り組みの実施がもたらす経済的なインパクト」などをより科学的な見地から明確にすることで、科学的に効果が実証された介護手法やアプローチの実現が必要であるという視点に立ち、関係者・有識者との意見交換を深めた。その中で、取り組み効果の評価指標の一つ「歩容」に着目することが妥当と判断し、高齢者の安心・安全な「歩行」を支援する歩容解析AIの開発を決定した。2019年度は介護事業者などと連携し、介護サービスの現場において実際に高齢者の歩容を評価している、理学療養士や有識者が監修する「歩容解析AI（技術）」を開発した。

「制度基盤」について、2018年度および2019年度は、個人の情報・データの保護および利活用に関する世界的な規制の策定・施行が進められる中、世界基準の一つであるGDPR（General Data Protection Regulation：一般データ保護規則）において、「（個人情報が含まれる）動画情報の共有」などを行う際に確認・注意すべき事項を整理した。また、2019年度においては、今後の日本の介護分野において、さらなるAI技術の開発・活用に重要な役割を果たすことが想定される、データベースに関わる検討およびヒアリングを実施。その結果、介護事業者などのケア現場へのヒアリングを通じて、「当該データベースへの入力に関わる負担が制度的な義務として介護事業者に課されること」へ懸念の声を上げる事業者が少なくなかった。

2) 実証系のテーマ

「実証実験に関わるプラットフォームの役割定義・実装」については、2020年11月に静岡大学とみんなの認知症情報学会が石川県加賀市と連携協定を締結し、自治体を巻き込んだ社会実装体制を構築した。静岡大学は「研究開発機関」として、ケア情報学研究所を主体にAIシステムとデータ基盤技術を開発。みんなの認知症情報学会は「社会実装母体」として、自治体・施設などを含む約30の賛助会員と「みんなの認知症見立て塾」の開催・評価などを担当。加賀市に実験場を提供いただき、持続的にデータを蓄積する仕組みの開発と社会実験を推進する「社会実験の拠点」として協力いただいた。

「本人支援AIに関わる開発・実証」については、認知症に関わる人たちが認知症当事者の状態を適切に評価するため、「見立て」や「ケア」力を学べるプログラム「みんなの認知症見立て塾」を開発・実践した。より多くの人に、継続的に安定して学べる環境を提供するため、オンライン協調学習システムを開発した。システムを実践することで、ケアの工夫の前に根本的原因の改善に向けて適切な医療につなげられる環境を構築することを目指した。

「介護者支援AIに関わる開発・実証」について、「せん妄」は、突然発生して変動する精神機能障害であり、認知症の症状を含有するが、通常は適切な治療とケアにより回復可能である。しかし、一般市民はせん妄に関する知識がないため、せん妄の診断のために必要となる情報を適切に医師に提供できず、認知症と誤診される場合

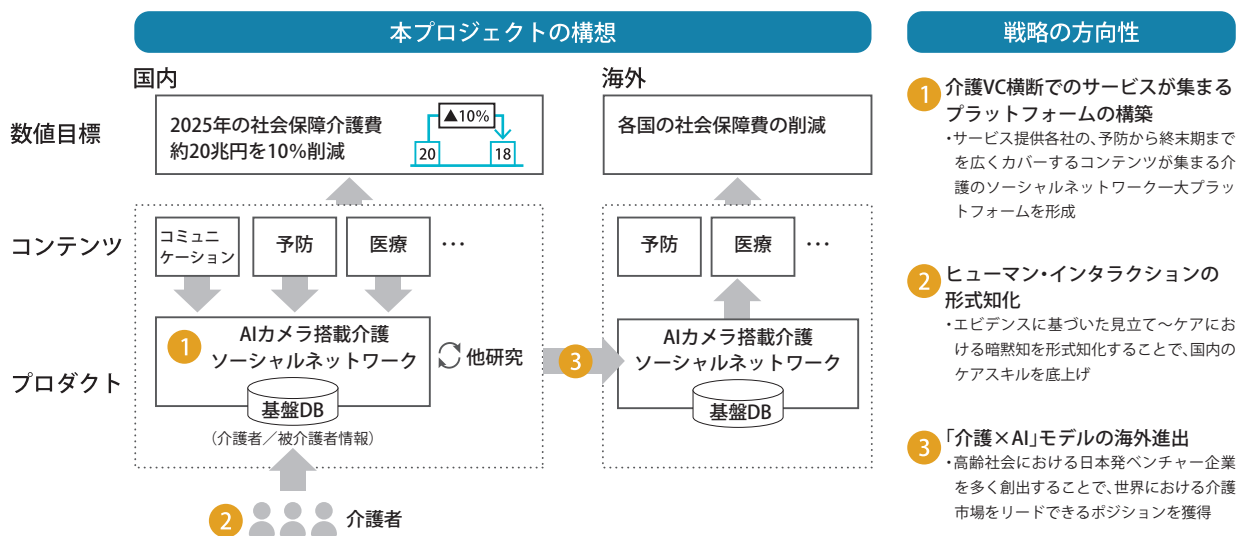


図1 戦略の方向性

*1 国の行政区画の中で最小の単位で、首長や地方議会などの自治制度がある自治体。

がある。また、認知症者の半数以上に、せん妄の合併がみられることが報告されているが、せん妄をBPSD（行動・心理症状）と勘違いして改善できないと思ひ込み、放置されていることがある。こうした問題を解決につなげるスクリーニングシステムを開発した。

「ネットワーク構造変容・社会変容（AI 利活用コミュニティ）に関わる実証」について、他者や社会集団によって個人に押し付けられた、ネガティブなレッテルと定義される「スティグマ」は、当事者以外に家族や社会の行動にも影響を与える。一方で、認知症になると「何もできなくなる」という決めつけは、認知症の方の自尊心を低下させ、生きる力を奪ってしまう。このような社会的構造を変えるためにも、認知症の方に対するスティグマの解消を目指し、Web や紙媒体雑誌などのコンテンツ開発を行い、クロスメディアの活用によってスティグマの解消や家族・社会の行動変容につなげることを目的とした。

3 本研究の成果

(1) 基盤系のテーマに関わる成果

「戦略基盤」については、2019 年度までに整理された戦略的事項を積極的に周知普及させ、2020 年度は国内での政治・行政および介護などの各分野における主体の認知を高めることを目標として掲げ、おおむね達成した。厚生労働省の関連部局との意見交換をはじめ、政府与党の関連部会・会議での発表や、日本経済新聞などのメディアにおける研究・開発成果に関わる記事の掲載などが実現した。本開発研究の全体を通じて、介護分野での AI 技術の利活用に関わる機会の全体像を整理した上で、国内連携・役割分担のあり方も整理し、日本での「介護領域における AI を利活用した課題解決の促進の方向性」を明らかにするという当初目標に対して、おおむね達成できたものと考えられる。

「データ基盤」については、2019 年度より調整を行っていた I 市の協力を得て、2020 年度に将来の要介護度、介護費および医療費予測を含めた予測モデルを開発した。当該予測モデルは、各自治体の将来の社会保障費のより正確な予測に大いに資するものであり、今後の実

装への期待も含め、I 市をはじめとした各自治体から大きな関心が寄せられた。また、当該予測モデルを主なテーマとした論文の表彰、海外の論文プラットフォームへの投稿および受理など、開発成果の国内外への周知・普及についても大きな前進があった。本研究全体を通じて、「介護領域における国内のデータを活用した AI 予測モデルの開発および普及」に一定の成果を達成できた。加えて、今後の介護領域におけるデータ整備・活用の推進にあたり「データを活用した AI 技術が介護領域にもたらすベネフィット」の一端を示すことができ、「介護領域におけるデータ利活用の促進に向けた基盤」の構築についても、一定程度達成できたと考える。

「評価基盤」については、2019 年度において「各取り組みが人の身体に与える影響」などの可視化を実現する歩容解析 AI を開発し、2020 年度においては「各取り組みの実施がもたらす経済的なインパクト」の効果検証・評価のための基盤整備を進めるため、「認知症領域における医科学的エビデンス」、「経済学的な視点」を踏まえたロジックモデルを検討・設計。認知症高齢者の BPSD（認知症の行動・心理状態）通減モデルでは、専門職に限定せず家族介護者をも対象とし、ユマニチュード*2 講習会により期待できる認知症高齢者の BPSD 予防や ADL 改善による家族介護者の負担軽減、QOL 向上による経済効果を示した。生活機能維持モデルでは、生活機能を維持する運動機能向上プログラムによる経済効果を示した。

(2) 実証系のテーマに関わる成果

「実証実験に関わるプラットフォームの役割定義・実装」については、2020 年度に「AI を使った住民の健康長寿を支援するまちづくり」に向け、フレイル*3 予防に向けたエビデンスに基づく介入提案、認知症ケアの向上のための市民向け・専門職向け学習プログラムの研究開発と実践、および健康増進・介護予防に向けた心身健康データの分析と利活用などに取り組む、自治体連携体制を構築。「みんなの認知症見立て塾」を加賀市にて 5 回実施し、学習状況のデータを蓄積した（図 2）。

「本人支援 AI に関わる開発・実証」については、設計した事業モデルに基づき、多様な地域・フィールドで提案システムを展開することで、ケアの工夫の前に根本的

*2 人間らしさを取り戻すこと。

*3 健康な状態と要介護状態の中間に位置し、身体的機能や認知機能の低下が見られる状態。

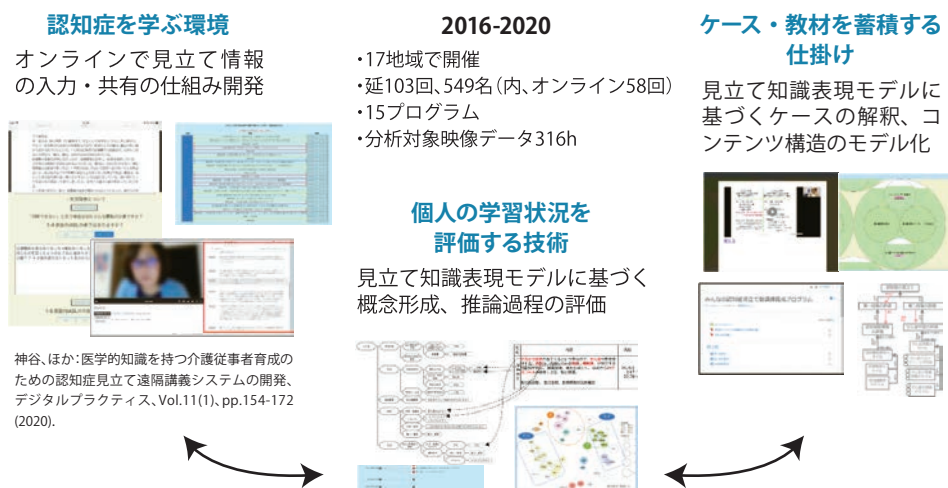


図2 みんなの認知症見立て塾における研究開発の技術概要

原因改善に向けて適切な医療につながられる環境を構築できる見通しを得た。2020年度末までに、全国17地域15プログラムにて、オンライン58回を含む延べ103回開催し、549名が参加。延べ316時間の受講者の学習場面の映像データを収録し、学習者の見立てスキル向上の効果、および学習者の属性の違いによる変容パターンの違いなどの知見を獲得した。

「介護者支援AIに関わる開発・実証」については、改善可能認知症のスクリーニングシステム開発に関して、介護職は効率的に入力できる仕組みが必要だが、家族は入力することで学習につながる事が意見として挙がり、システムの有用性が示された。生活環境デザインシステムのプロトタイプ構築に関しては、データの可視化機能の充実化が今後のアセスメントや記録の質の向上に貢献すると考えられる。

「ネットワーク構造変容・社会変容（AI利活用コミュニティ）に関わる実証」については、本開発研究によって、認知症に対するネガティブな心理社会的状態の緩和・改善に向けた、認知症リテラシー変容促進の知見を得た。さらに、国際競争力を持ちつつ海外団体との連携を実現するために、海外団体の参考事例や競合可能性、連携に向けて解決すべき課題・法規制を整理した。

(3) AI技術の開発に関わる成果

1) 将来の要介護度などに関わる予測モデルの開発

(データ基盤関連)

鎌倉市とI市から協力を得て、3年から5年後の個人レベルでの要介護度を予測するモデルの作成や、将来の要介護度別認定人数の予測などを実施。その結果、

2018年度の要介護度予測モデル開発においては、従来までの一定の集団の特徴に基づいた予測ではなく、個人の特徴に基づいた予測を行うことに成功した。

2) 歩容解析AI（評価基盤関連）

特別なセンシングデバイスを用いず、スマートフォンのカメラ機能などの活用のみによって、理学療法士的な観点からの歩容の解析や評価を実現する、動画解析のアルゴリズム・歩容解析AIを開発した。2020年度の神戸市や堺市との実証事業において、本歩容解析AIを活用した事例が開発でき、本事業の目的である開発技術の社会実装について確たる実績が残せた。

4 まとめと今後の展望

基盤系のテーマに関しては、今後も策定された国内外での戦略の実現を推進していく。また、本研究で開発された予測モデルは各自治体から高い関心が寄せられており、今後も各自治体などでの活用促進を図っていく。さらに、新たな機能拡張を加えた歩容解析AIを活用したビジネスサービスの運営だけでなく、当該AI技術やロジックモデルなどを活用した成果連動型民間委託契約、および民間からの外部資金調達を伴う成果連動型民間委託契約の実装を呼びかける。

実証系のテーマに関しては、今後は地域やコミュニティごとにカスタマイズしたプログラム設計と実証評価を進め、連携先企業の発掘など実証評価に向けた方策を検討する予定だ。

遠隔医療AIが連携した 日本式ICT地域包括ケアモデルの研究開発

株式会社アルム、学校法人慈恵大学東京慈恵会医科大学、
日本テクトシステムズ株式会社、データセクション株式会社

- 地域包括ケア現場における ICT インフラを活用し、各種ウェアラブル IoT、医療機器・福祉機器のデータを連携・収集してモニタリング AI と遠隔医療 AI の研究開発を行った。
- オープンなプラットフォームを構築し、慢性的な人手不足である介護・看護現場を最新の介護 AI、医療 AI や専門医による遠隔診療が支援する基盤を開発した。
- SIP コンソーシアム間、府省の公的研究開発事業、民間主導の研究開発などとも連携し、新たな研究開発・社会実装を推進して介護現場の社会課題解決につなげた。

1 研究の目的

最先端の介護・看護・専門医療 ICT 連携 AI の日本式モデルによって介護現場の負担軽減につなげるために、地域包括ケア現場における ICT インフラを活用し、各種ウェアラブル IoT、医療機器のデータを連携・収集し、モニタリング AI と遠隔医療 AI の研究開発を行った(図 1)。介護士・看護師が携帯する業務用スマートデバイスを通じて業務支援を行うとともに、慢性的な人手不足である介護・看護現場の業務データから、申し送りなどのタイムリーな気づきデータを活用し、ヒューマン・インタラクション AI 開発による生産性向上と医療への橋渡しによる高度化を実現した。

2 実施期間と方法

(1) 実施期間

2018 年度～ 2022 年度

(2) 実施方法

1) スタンドアロンで利用される医療機器のネットワーク化の研究開発 (2018 年度～2022 年度事業)

在宅・施設介護における患者の容態を把握するために、アルムの地域包括ケア推進システム「Team」を基盤に、医療・看護・介護従事者がプラットフォームを通じて、リアルタイムに要介護者をモニタリングするための機能開発を行った。介護分野におけるオープンプラットフォームとして、府省プロジェクトや医療機器、福祉機器、AI 技術とネットワークを介して連携した。

2) 急性期疾患を検知するモニタリングAIの研究開発 (2018 年度～2022 年度事業)

深層学習モデルを使用し、12 誘導心電図およびホルター心電図の特定疾患検知のためのアルゴリズムを開発した。特定疾患としては、脳卒中への関連が大きい心房細動と期外収縮を選択し、検知精度 95% 以上を達成することを目指した。アルゴリズムは、アプリケーション化と、基盤となる介護プラットフォームである Team と連携可能な機能を実装した。アプリケーションは、現場に機器を提供し、医療機器として開発するにあたってのデータを取得した。

地域包括ケア現場におけるICTインフラを活用し、各種ウェアラブルIoT、医療機器のデータを連携・収集。モニタリングAIと遠隔医療AIの研究開発を行った

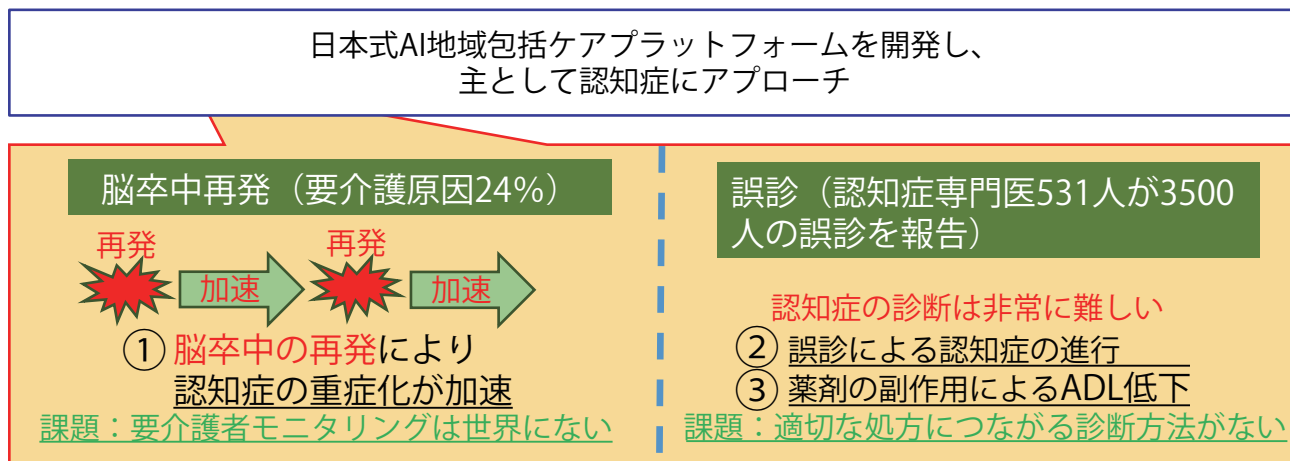


図1 研究の目的

3) 慢性期疾患の重症化予防やスクリーニングの要介護者認知症サポートAIの研究開発（2018年度～2022年度事業）

全体の2/3を占める認知症を呈す要介護者に対して、重症化予防（介護の負担軽減）を目的とした認知症サポートAIを開発した。認知症は適切な診断が難しい例もあり、適切ではない診断や投薬されるケースも報告されていることから、客観的データとなるMRI画像と神経心理検査の結果を使って、医師の認知症診断の補助となるアウトプットの提示機能の開発を目指した。開発にあたり課題となる、MRIの撮像誤差を吸収するAIアルゴリズム機能も開発し搭載した。社会実装に向けては、UXを重視し、医療機関でヒアリングした内容を製品のUI、アウトプットおよびアルゴリズムに反映した。

4) 多職種連携における介護者サポートAIの研究開発（2018年度～2022年度事業）

介護者サポートAIでは、日々蓄積される申し送り事項および入力される記録紙データ、指示情報から転倒リスクおよびポリファーマシーのリスクを検出して介護者に知らせるため、自然言語解析を行った。また、モニタリングAIから出力されるアラートや認知症サポートAIから出力されるレポートを受け取り、Teamを利用する介護士や看護師のインターフェース機能を開発し、実装した。加えて、有識者会議および府省連携に応じて、認知症の重症化評価につながるBPSD（行動・心理症状）の入力・フィードバックインターフェースを開発し、介護者・医師がBPSDのスコアリングを参照できるように

し、製品版の提供を開始した。

5) インタラクション指標解析プラットフォームの研究開発（2018年度～2022年度事業）

モニタリングAI、認知症サポートAI、介護者サポートAIのインプットをもとに、インタラクション系や医学系、および経済系指標の関係を整理するモデルを構築した。モデルの構築にあたっては、介護分野において重要となる指標について、アンケートおよびヒアリングを行った。介護原因の上位となる、脳卒中、認知症、骨折については、本研究で開発したソリューションが介護負担軽減と費用に与える影響について検討するためのシナリオを作成し、モデル化した。また、介護士の直接的な業務負荷、および精神的負担を軽減する手段として、日々の介護業務で発生する介護記録紙や利用者の状態を関係者で共有する申し送りなどのデータ（業務データ）、および機器のネットワーク化により新しく介護データと連携する医療機器などからのデータや医療データを活用し、現場にとって必要な研修などの体制が整っていない部分や、介護士が業務で連携する医師・看護師などとのコミュニケーションのうち、改善の余地がある部分は円滑なコミュニケーションを支援するために検討した。

6) SIPコンソーシアム間の連携研究開発（2021年度～2022年度事業）

SIPとして研究開発が進められている、ほかのコンソーシアムの成果物と連携した。連携にあたっては、介護現場で基幹システムとして実際に使用されているソリュ

ーションである Team の実績を考慮し、データの受信と提供の両方の機能を提供した。

7) 介護クラスター感染AIの研究開発 (2021 年度事業)

介護施設では、新型コロナウイルス感染症のクラスターが発生し、日々のケアを行う介護士の離脱のほか、感染した要介護者の搬送先医療機関が見つからないなど、介護施設におけるクラスターは重大な脅威となっている。PHR (Personal Health Record) アプリを介して介護施設に従事する職員に対して、日々の問診とバイタルのモニタリングを行い、新型コロナウイルス感染症のクラスター発生と重症化予測に関して多変量解析を行った。

8) 音声によるMCIスクリーニングAIの研究開発 (2021 年度～2022 年度事業)

早期に MCI (軽度認知障害) を発見するため、多忙な介護現場のスタッフによる検査実施も可能な、音声による MCI スクリーニング AI を開発した。スマホなどから音声で神経心理検査課題 (設問) を提示し、受検者が音声で回答したものをテキスト変換し、AI が数分で判別する。スマホアプリや電話 (自動音声) での実施にも応用することで、対面や非対面 (遠隔、無人) での運用も可能とする。MCI の判定以外に、健常、認知症の判別機能も持たせた。

9) 脳波AIを活用した認知症の遠隔鑑別スクリーニングの研究開発 (2021 年度～2022 年度事業)

脳波データから 90% 以上の精度で認知症とてんかんの鑑別が行える、遠隔鑑別スクリーニング AI の研究開発を行った。社会実装に向けて、遠隔鑑別スクリーニングの薬事認可および DPC (Doctor to Patient with Caregiver: 専門医によるオンライン診療を介護士・看護師がサポート) モデルの保険適応を狙う。学習用の脳波データの収集とラベリング作業を行い、てんかん検出のための鑑別アルゴリズムの設計・開発を行い、専門医による治療の適正化 (治療方針変更の有意性) の実証をした。

3 研究の成果

1) スタンドアロンで利用される医療機器のネットワーク化の研究開発

心電図、生体情報モニター、血圧測定機器、パルスオキシメーター、脳波センサー、体組成計について、ネットワーク化と介護プラットフォームである Team との連携機能の開発を行い、実証を実施した。また、SIP 第 2 期「フィジカル空間デジタルデータ処理基盤」開発のデバイスとも連携を行い、介護現場において使用される機器からのデータを、ネットワークを通じて介護プラットフォームに蓄積するソリューションを実現した。さらに、ネットワーク化に関する開発を行ったことにより、後述のオープンプラットフォームの基盤を形成した。

最終年度では、3,000 施設、6 万人以上の介護士・看護師に商用提供を開始した。

2) 急性期疾患を検知するモニタリングAIの研究開発

2021 年度は、心電図による脳卒中可能性検知エンジンで、心房細動の検知精度 95% 以上を実現するモニタリング AI の開発を進め、正常・異常の検知精度は 12 誘導にて 94.7% を達成し、心房細動の検知精度はホルター心電図にて 85% を達成した (図 2)。また、モニタリング AI のアプリケーション化および Team との連携を開始し、第二種医療機器製造販売業許可の取得を完了した。

最終年度では、モニタリング AI において心房細動の検知精度 95% 以上を実現した。論文を 1 本発表し、医療機器の規制当局である PMDA から得られた助言を参考としながら、認証取得して製品化する予定である。

3) 慢性期疾患の重症化予防やスクリーニングの要介護者認知症サポートAIの研究開発

日本人 4,600 人分の脳画像を 30 クラスターに分類した上で、専門家の用手的作業によって、解剖学的に分割して作成した脳アトラスを使用し、高精度の脳体積測定ツール (MRI 画像解析ツール) を開発した。認知症の診療においては、全脳各部位の体積変化速度の測定が重要視されており、変化速度を精度高く測定する機能を用いて認知症の経過を追うことで有益な情報が得られる。本機能は規制要件をクリアし、新たに医療機器として認証を取得した。

求められる性能要件と技術的な達成度	社会実装の状況	
<p>ホルター心電図</p> <p>A.モニタリングAI</p> <p>心電波形から脳卒中の確率を予測</p>	<p>モニタリングAIの心電図による脳卒中可能性検知エンジンでの心房細動・期外収縮の検知精度95%以上</p> <p>【達成度】 正常・異常の検知精度は12誘導にて94.7%を達成し、心房細動の検知精度はホルター心電図にて85%を達成した</p>	<p>Teamクラウドサーバーとの連携に向けた解析モデルのアプリケーション化、およびデータ連携プラットフォームとの接続は実装</p> <p>市販向け、規制対応手続き中</p>
<p>MRI</p> <p>B.認知症サポートAI</p> <p>神経心理検査の結果、MRI画像等をAI解析し、治療指標を作成</p>	<p>熟練の専門医が示す処方判断に対し、95%以上の正答率を導くMRI脳画像による要介護者認知症サポートAI</p> <p>【達成度】 熟練の専門医が示した処方判断の93%以上を達成した</p>	<p>要介護者認知症サポートAIのβ版の開発 脳画像解析ツールの管理医療機器（クラスII）で認証取得</p>
<p>Team介護PF</p> <p>C.介護サポートAI</p> <p>PFに記録される病状等と服薬情報を分析</p>	<p>地域包括ケアの申し送りテキストデータに対し、80%以上の精度のリスク検出を行う介護者サポートAI</p> <p>【達成度】 半年以内の骨折予測で感度74%を達成した。精度は80%以上であることを確認した</p>	<p>介護者・看護者向けIFの設計・Teamへの実装</p> <p>介護士および看護師に対するアンケートでは、80%が参考になったと回答</p>

図2 研究成果のまとめ

MRI 画像解析ツールによる脳部位の体積値と ADAS (Alzheimer's Disease Assessment Scale) の検査結果の同期および有効薬剤等指標提示機能を連携した、要介護者認知症サポート AI の β 版を開発した。AI が導き出す目標正答率を、熟練の専門医が示した処方判断の 93% 以上と設定して、目標を達成した (図 2)。結果は介護プラットフォームである Team で閲覧可能なよう、レポート機能を開発しシステム連携を行った。

4) 多職種連携における介護者サポートAIの研究開発

Team に登録されたデータのうち、1,400 万レコードを用いて転倒・骨折の関連キーワードについての自然言語処理を行い、リスクモデルを作成した。半年以内の骨折予測で感度 74% を達成した (図 2)。精度は 80% 以上であることが確認された。ポリファーマシーについては、Team に登録されている要介護者の薬剤情報と、出現している症状に対する記録をもとに、服用薬に関連する副作用が出現していないか確認する介護士用ガイドスを開発した。介護士および看護師に対するアンケートでは、80% が参考になったと回答した。認知症における診断では医師との対面時だけではなく、普段の生活の様子の記録が重要であるが、認知症の周辺症状である BPSD に関する評価スケールを Team に実装し、認知症サポート AI からのインプットと合わせて解析するための基盤を整えた。

5) インタラクション指標解析プラットフォームの研究開発

Team に記録されている介護送り事項の言語解析結果を用いて、介護効率化および医療費削減に寄与可能なストーリーの精緻化を行った。介護原因のうち大きな原因である骨折について解析を進めた結果、要介護者本人のふらつきが介護記録上に出現することが重要であることが明らかとなった。アンケートおよびヒアリングからは、介護士の負担感として人手不足や肉体的な負担感のほか、多職種とのコミュニケーションへの課題が挙げられ、後者については IT による課題解決が期待されていた。これを踏まえ、多職種間のインターフェース設計を行った。介護業界は公的介護保険により収益を得ているが、基本的には時間単価が定められており、IT を使って時間を短縮すると収益が下がる構造となる。そのため、モデルの構築では介護士一人当たりが受け持つことのできる要介護者数を算出して評価を行った。作成した評価モデルでは、ソリューションの実装により特定の集団に対してであるが 10% を超える生産性の向上が見込まれると想定された。

6) SIP コンソーシアム間の連携研究開発

KDDI コンソーシアムのマルチモーダル音声対話システム「MICSUS」と連携し、対話終了後の対話履歴や対話時映像を Team に取り込み、地域包括ケアシステムの情報共有画面での表示およびリスク分析への反映も実装

した。aba コンソーシアムとは、排泄検知 AI 開発に向けた介護データ（処方・排泄情報など）の正規化・提供を行った。各コンソーシアムとは、社会実装に向けた協議も行っている。

7) 介護クラスター感染AIの研究開発

新型コロナウイルス感染症の重症化予測について、多変量解析を実施した。すでに重症化が予兆できる患者を除外した上で、発症初日の症状やバイタルから重症化予測アルゴリズムを開発し、統計的に精度が高い結果 (AUC 0.822) が得られた。発熱・体温・咳・喉の痛みといった一般的な新型コロナウイルス感染症の症状の感度が最も良いことから、これらの症状から PCR・抗原検査を推奨し、重症化予測をアラートすることで適切な医療・隔離管理・ケアへつなげる。実証に参加した介護機関が引き続きアプリを利用していることから、現場での有用性は確認された。

8) 音声によるMCIスクリーニングAIの研究開発

音声による MCI スクリーニング AI の研究開発を行った。試作開発後、臨床研究を実施し、新たに健常、MCI、認知症の方の音声データを入手した。入手したデータを用いてアルゴリズムの修正、精度向上のための改善を繰り返し、目標の正答率 85% に対して、85.3% の

結果で達成した。同時に製品版となるアプリ開発を実施し、社会実装に向けて、検査時間の短縮や精度の確認などを行い、薬事申請のための治験の準備も開始した。

9) 脳波AIを活用した認知症の遠隔鑑別スクリーニングの研究開発

脳波データから 90% の精度を目標に認知症とてんかんが鑑別できる、遠隔鑑別スクリーニング AI の研究開発を行い、93% に達成した。社会実装に向けて、遠隔鑑別スクリーニング AI の薬事申請および DPC モデルの保険適応を見据えた臨床研究を行った。学習用の脳波データの収集とラベリング作業を行い、てんかん異常波検出のための鑑別アルゴリズムを設計・開発し、遠隔鑑別スクリーニング AI の薬事申請を行った。また、70 名の認知症患者に対して臨床研究を行い、DPC モデルの有用性を検証した。認知症専門医が遠隔から面談した結果、認知症患者の処方変更につながり、認知症周辺症状の改善につながった。

4 まとめと今後の展望

「介護者への負担増加を抑制」および「被介護者への生活の質向上」（図 3）を目指し、より多くの介護が必

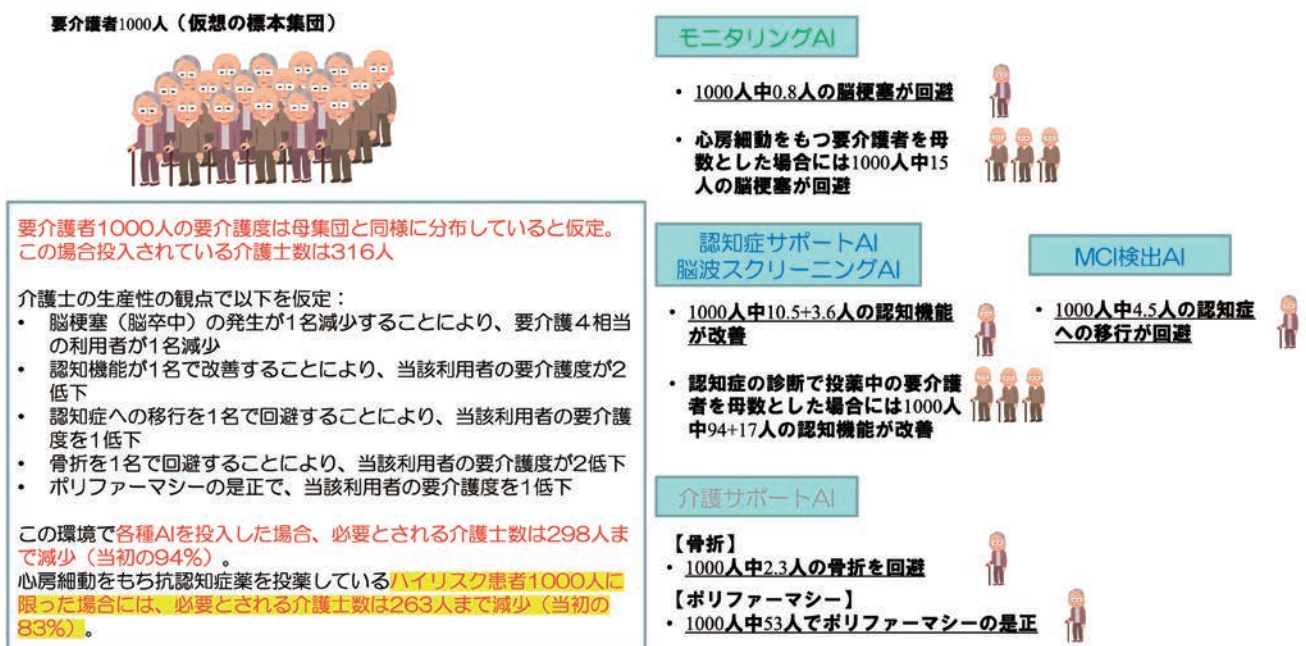


図3 「介護者への負担増加を抑制」と「被介護者への生活の質向上」

- 本研究では、「介護者への負担増加を抑制」および「被介護者への生活の質向上」を目指し、より多くの介護が必要となる状態への移行を予防するAIを開発。
- 主な介護原因として知られる、認知症、脳卒中、骨折に焦点を当て、検査機器からの情報や介護記録からの情報をインプットとし開発し、AIを地域包括ケアプラットフォームであるTeamに接続。
- これにより、介護現場へのシームレスな導入が可能となり、疾患の進展や発症を予防が期待される。

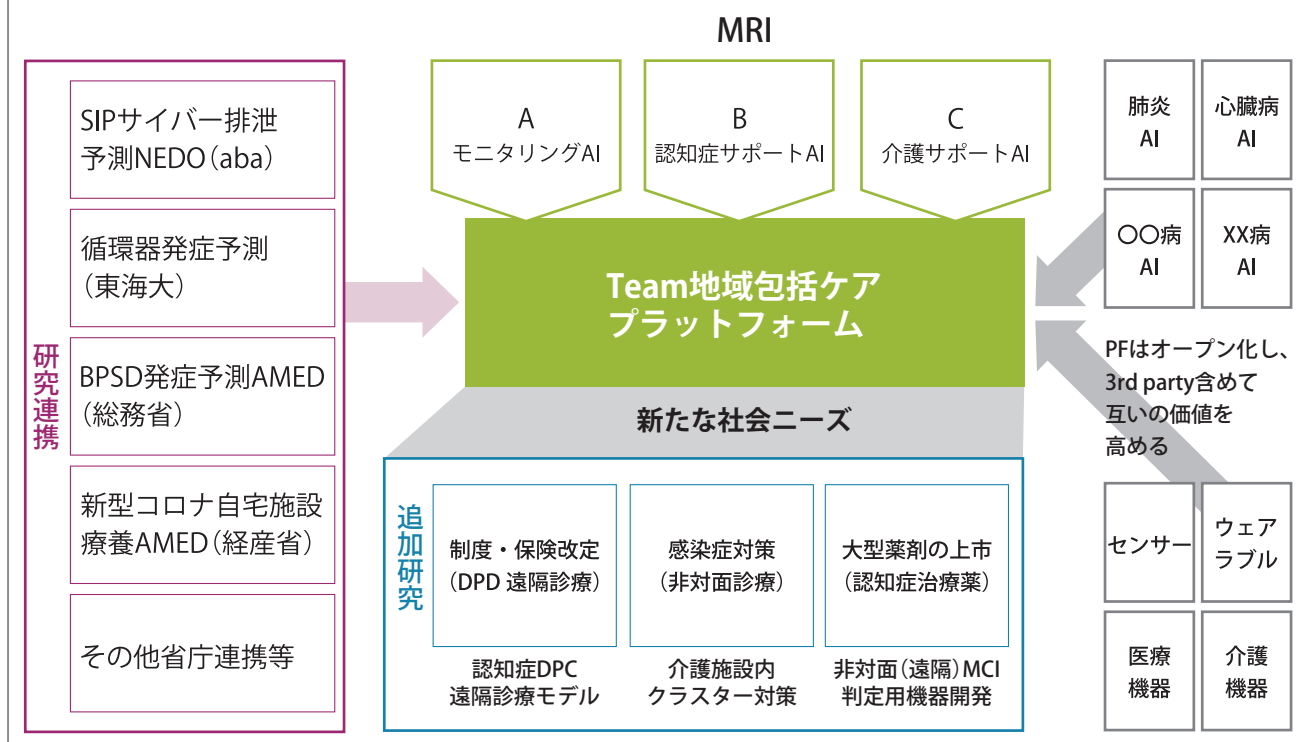


図4 今後の展望

要となる状態への移行を予防するAIを開発した結果、介護現場の課題であった労働力不足に対する解決策の一つとなった。今後、介護の効率化はますます重要になってくる。高齢化に直面する状況は日本だけでなく、海外においても同様である。国ごとに介護に関する制度は異なるものの、高齢者人口が増加し、若年人口が限られたリソースでそれを支えていく構図は変わらないと考えられる(図4)。

今後も本事業での成果を地域包括ケアのプラットフォームとして、日本と世界の介護現場における効率化と質の向上に寄与するよう尽力していく。

排泄情報を基軸とした介護業務スケジュールの最適化およびケアの質向上を実現するシステム

株式会社aba

- さまざまな状況に適した排泄デバイスを開発して各センサーの特性を考慮し、尿便識別が可能なAIモデルを作成した。
- 介護現場の中で業務効率の根幹であるにも関わらず、最適化が難しい業務スケジュールを開発することで全体的な生産性向上を図った。
- 本成果を「排泄情報を基軸とした介護業務スケジュールの最適化およびケアの質向上を実現するシステム」に導入し、その効果の定量的な可視化を実施した。
- コンソーシアム企業間連携の一つとして、プライベート空間であるため介護者の目が行き届きにくいトイレ内にて、介護者の代わりに要介護者と対話し、見守りを可能にするシステムを開発した。

1 研究の目的

本事業では「排泄情報を基軸とした介護業務スケジュールの最適化およびケアの質向上を実現する介護者支援システム」の開発を行った。介護現場を支える多くの人材は、介護の専門的知識を十分に身につけられずに介護を始めている。一方で、適時的確に適切なケアを実践すれば、要介護者のQOLを向上させることが可能である。また、要介護者が好状態であれば介護者の負担も軽減され、双方にとって好循環が生まれる。

本開発により無資格未経験者であっても適切なケアを実践できる機能をシステムに構築し、その効果の定量的な可視化を実施した。

具体的には、介護業務の中で介護者が行っている業務内容を推定するため、動作情報（介護者の身体動作をIMUセンサーや画像解析などで測定した情報）や、音声情報などを取得し、介護業務の内容推定を行った。

2 実施期間と方法

(1) 実施期間

2018年度～2022年度

(2) 実施方法

2020年度までは主に調査期間として、開発するシステムについて現場のニーズに近づけるため、指標の検討や業務分析を行った。2021年度以降は開発期間として、モジュールごとのシステムやデバイスの開発を行った。最終年度は検証期間として、実際の施設や事業所の協力をもとに実証実験を行い、その効果を可視化した。

1) 効果検証Before（業務分析による現状把握）、効果検証Afterの実施

2020年度は、2018年頃から開発していた排泄センサー用のWebアプリケーションを実証実験の現場で使用し、先行的な効果検証を行った。abaの既存サービスでは、尿量や便の形状などの詳細な排泄情報の記録はパソコンからの入力にしか対応していなかったため、新たにスマートフォンアプリケーションを開発して導入した。

2) 実証実験先協力施設の実験環境整備

実証実験や効果検証を行う上で、実験環境の整備は非常に重要である。実験が行えなければ本末転倒であり、あらかじめ実験環境が整っている機関や企業を除いて、ほとんどの企業が実験環境の整備に多くの時間を消費している。特に、研究の効果実証のために人に協力を依頼する際は、倫理的な観点から個人情報保護やプライバシー保護がひととき重要になってくる。本事業では実験環境の整備のため、同意書の作成や実験実施の際の検討事項について改めて洗い出しを行い、ノウハウ化した。

3) 排泄センサーデバイスおよび尿便識別AIの開発

排泄センサーデバイスでは、以下の三つのコンセプトに大きく分けて開発を行った（図1）。

- 小型化
- メンテナンス性

• シーンごと（日中、夜間）の利用

まず小型化の開発においては、近年では従来品より薄いにおいの濃度でも反応可能な高性能センサーが流通し始めており、2019年度および2020年度には2種類のおむつ装着型の小型デバイスを開発した。MEMS技術*1を利用した高性能センサーを搭載することで省電力化および小型化を実現し、一方はボタン電池で約2日の駆動が可能であり、もう一方はUSBの有線式による安定した稼働を実現していることが特徴となっている。

次にメンテナンス性の開発においては、abaの従来製品では排泄臭気をポンプで吸引してセンサーに当てていたが、デメリットとして尿や便が漏れ出した際に汚物も吸い込んでしまい、メンテナンスが大変であるという課題があった。そこで、前提となる方式を刷新し、空気を吸引しない無吸引を特徴としたデバイスを実現した。

最後にシーンごとでの利用においては、前記の無吸引



図1 実際に開発した実験用排泄デバイス

*1 微細加工技術によって、機械要素部品やセンサー、電子回路などを一つのシリコン基板などの上に集積化したデバイス。

デバイスを日中および夜間の二つのシーンを想定し、それぞれ開発を行った。夜間は従来製品と同様にベッド上で使用可能なシートタイプ、多くの時間を過ごすことになる椅子や車椅子といった座乗で利用できる座布団型デバイスを開発した。開発したデバイスでは、社内での基礎評価のほかに、実験に協力いただける施設においても臨床評価を行った。

また、現時点では尿便識別に特化したセンサーが存在しない。そのため、ハードウェアとして尿便の識別を可能とするための排泄センサーの開発を行い、並行してソフトウェアとして尿便識別アルゴリズムを用いた尿便識別 AI の開発を行った。

尿便識別 AI の開発では、開発した排泄センサーデバイスで取得したデータを利用し、尿便識別のための機械学習モデルの研究開発と識別実験を行った。また、テスト用の環境だけではなく、実際の現場である介護施設において精度検証を行った。精度検証は協力を得られた 2 件の施設に対して日中と比べデータの収集しやすい夜間帯にて 1 週間ほど行った。

実験は、尿便識別 AI による予測の結果を職員に対してメールやチャットアプリケーションなどを利用して通知し、排泄介助および排泄記録を実施していただいた。その後、排泄記録と通知時刻を照合した結果を集計し精度を評価した。

4) 業務スケジュール最適化 AI の開発

業務スケジュール最適化 AI の開発では、「業務分析による現状把握」で得られた結果をもとに、業務スケジュールを最適化する上での要素を洗い出した。具体的には、職員のシフト情報（勤務可能時間帯、出勤パターン）や資格情報と、要介護者ごとの 1 日のスケジュールを入力することで 1 日分の業務スケジュールが出力される仕様とした（図 2 中央）。また、介護者が扱いやすい専用画面を開発するため、従来製品などに対して実際の介護現場職員が抱く不満などのヒアリングを実施した。

実証実験は、ユニット型特別養護老人ホーム施設の協力のもと、早朝の時間帯（6:30～8:30）の 2 時間に限定して 5 日間実施した。ここで、早朝の時間帯としたのは実施の現場において最も慌ただしくスケジュールが読めない時間帯であり、業務スケジュール最適化 AI の効果を検証するのに適しているためである。協力いただいた施設の職員（介護者）は 3 名（ベテラン〔熟練者〕2 名、新人 1 名）である。また、施設に入居されている

7 名の利用者（要介護者）にも協力いただいた。

具体的には、ベテランの要介護者に対する介助時間と業務スケジュール AI が最適化したスケジュールを新人が確認しながら業務を実施した介助時間とを比較した。

5) 生活スケジュール最適化 AI の開発

排泄を基軸とした生活スケジュール最適化 AI 開発のため、排泄に関して直接的な因果関係のある以下の要素を洗い出し、現場職員へのヒアリングを行った。

- 食事摂取
- 水分摂取
- 投薬（下剤）

この中で、意見の多かった投薬物（下剤）に注目し、既存システムに対して下剤情報を管理できるユーザーインターフェースを追加した（図 2 下）。

また、将来的には夜勤帯に介護無資格・未経験者が 1 名で多数の要介護者を見なければならぬ可能性が十分に考えられることに着目し、適切な排泄確認タイミングを算出する機能を検討・改良した。これは、現場介護者からのヒアリングにより、夜間帯のおむつ交換について要介護者の睡眠障害防止のため「排泄があり、かつ眠りの浅いタイミングでおむつ交換をしたい」というニーズがあったことから、介護現場にある従来の睡眠センサーと排泄センサーを併用し、排泄パターンから算出されたタイミングにおいて要介護者がどのような睡眠状態であるかを考慮しつつ訪室する実証実験を実施した。

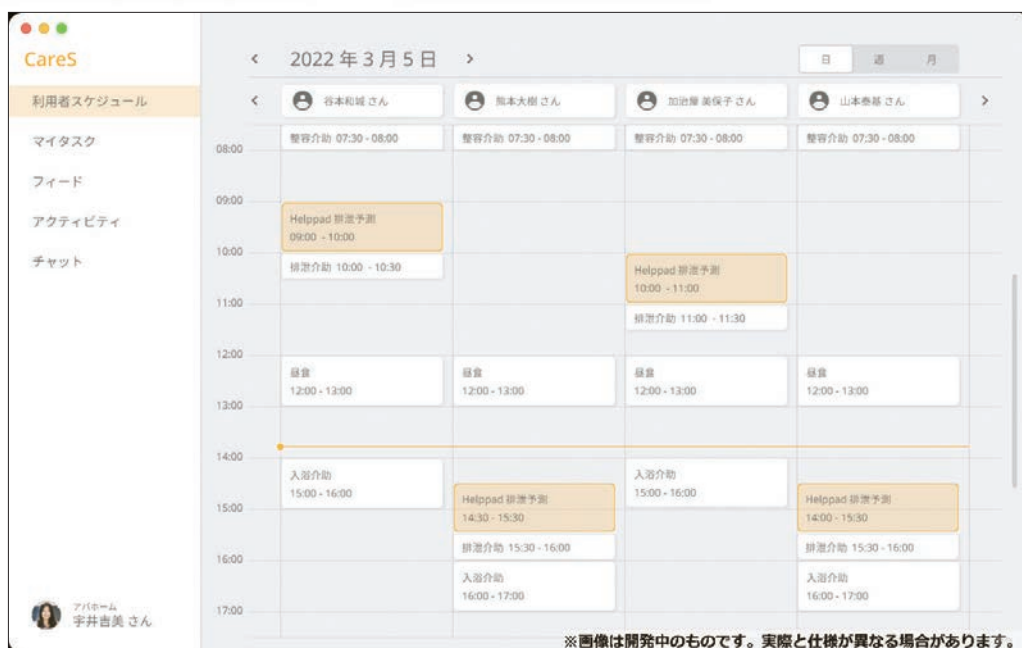
6) トイレ内対話型システムの開発

2021 年度からコンソーシアム企業間連携として aba の排泄検知技術と、NICT、KDDI、NEC ソリューションイノベータ株式会社のマルチモーダル音声対話システムである「MICSUS」の対話技術を連携し、プライベート空間でもあることで介護者の目が行き届きにくいトイレ内にて、介護者の代わりに対話しつつ見守りを可能にするシステムの開発を行った。

具体的には、介護施設のトイレに設置するハードウェアの新規開発、および aba の排泄検知エンジンと MICSUS の連携部分を新規に開発した。また、職員が持つお知らせ受信用のスマートフォンアプリケーションについても別途開発した（図 2 上）。動作や精度検証は、弊社社員の自宅やオフィスのトイレに設置を行い、実施した。また、ユーザビリティ検証は別途実験協力の得られた 2 か所の施設や事業所で実施した施設内トイレに



※画像は開発中のものです。実際と仕様と異なる場合があります。



※画像は開発中のものです。実際と仕様と異なる場合があります。

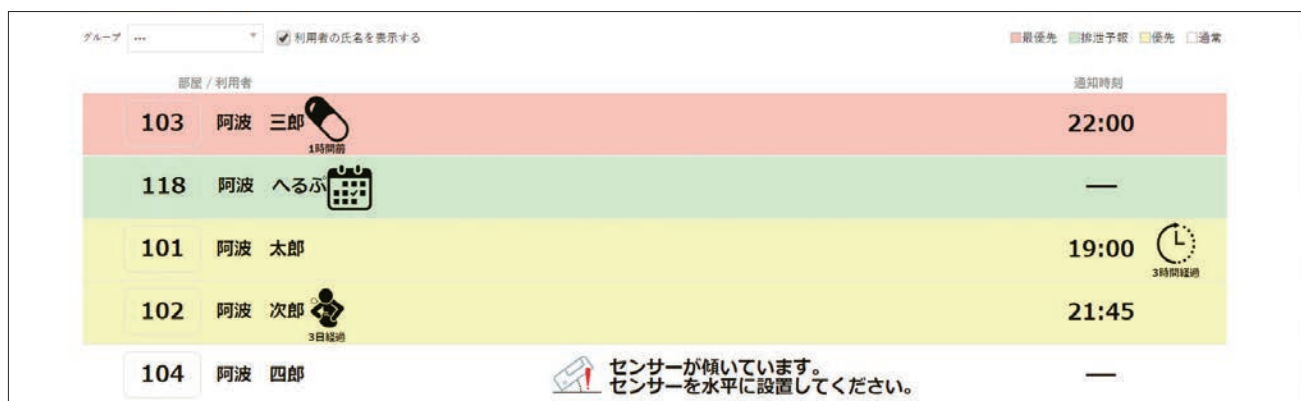


図2 業務スケジュールユーザーインターフェース画面および生活スケジュールユーザーインターフェース画面（下割通知）

設置された本システムを、介護職員と利用者を利用して
もらい、その後使用感のヒアリングを行った。

7) 介護業務活動記録技術の開発

業務スケジュールの最適化と並行して、介護者の介護
状況を自動で記録・学習するための行動解析システムに
ついては開発を進めていた。事業開始当初は、安価なウ
ェアラブルデバイスを用いた位置や姿勢の推定を行う計
画であったが、装着したウェアラブルデバイスにより要
介護者を傷つけてしまう可能性を考慮し、画像や動画か
ら人物の骨格をリアルタイムで検出できる「OpenPose
(オープンポーズ)」というソフトウェアに着目し、介護
者の介護状況を推定できるような行動解析システムの開
発を検討した。しかし、カメラを利用する場合、プライバ
シー保護の観点*2から実証が難しく、カメラに代わ
る位置情報取得デバイスを開発する必要が生じた。

再検討の結果、2021年度から Bluetooth Low Energy
(BLE) ビーコンを利用して動線データを取得し、介護
業務において誰が・いつ・何をしたかを可視化するシス
テムの開発を開始した。この介護支援システムの導入前
後で、介護職員の行動変容の可視化やシステムの評価を
定量的に行えるようになり、業務効率化につながると考

えられる。本事業では、主にコンソーシアム企業間連携
として産業技術総合研究所とともに実証実験を行い、動
線データの取得・解析を実施した。

3 研究の成果

(1) 効果検証 Before (業務分析による現状把握)、 効果検証 After の実施

排泄センサーシステムの導入前は、ケアステーション
において記録が滞っていた情報を、業務終了直前に入力
する必要があり、その結果業務時間が長くなるという問
題が発生していた。しかし、システム導入後はスマート
フォンからアプリケーションを通じてすきま時間での情
報を入力することが可能となり、効率的であるという声
が多かった。実際に入力情報が多い時で1～2時間程
度残業していたが、0.5～1時間に改善された。

さらに、ユーザーへのヒアリングの結果を受け、Web
アプリケーション画面の改善も行った。具体的には、夜
勤帯に見やすい配色・コントラスト、文字の大きさなど
視認性に関する改善を行った。現場では、片手間にチェ
ックや入力を行うため、1画面上における要素数を少な

	ベテラン		新人	
	スケジュールなし	スケジュールなし	スケジュールあり	
			コールなし	コールあり
利用者 A	03:30	07:01	05:25	06:15
利用者 B	09:24	14:02	13:56	13:56
利用者 C	11:08	18:22	16:44	18:32
利用者 D	05:23	10:53	14:04	14:04
利用者 E	07:19	08:47	17:07	17:07
利用者 F	02:52	05:48	04:21	04:21
利用者 G	07:27	13:00	18:03	18:28
合計時間	00:47:03	01:17:53	01:29:40	01:32:43
比率	0.60	1.00	1.15	1.19

表1 ベテランと新人による利用者(要介護者)への介助時間の違い

*2 骨格情報に限定した実際の映像が残らない仕組みであっても、居室内にカメラが存在することに対する精神的不安感への意見が多かった。

くする工夫を行った。そのほか、プリストルスケール表（便の性状分類）についてイラストのみでは理解しづらいという指摘があり、情報説明を追加する形で改善した。

(2) 実証実験先協力施設の実験環境整備

実験環境整備のための成果として、検討・準備・実施（運用）の三つのフェーズに大きく分けて、実験者および被験者の選定条件や、同意書などの資料準備のタイミング、実験実施時に起きうるインシデントやその対策などについて可能なものはフローなどにまとめた。

(3) 排泄センサーデバイス開発および排泄検知AIの改良

開発したデバイスを用いて実験を行った結果、小型デバイスでは以下のような課題が明らかとなった。

- 全体的に小型すぎておむつへの装着位置が定まらず、検知が行えるようなデータが取得できなかった。
- ボタン電池型では稼働時間制限による電池交換や置き忘れ問題、紛失や要介護者による誤飲のリスクがあった。
- USBによる有線型ではコードの取り回しに難があり、絡まりなどによる危険が示唆された。

また、座布団型や夜間シート型では小型デバイスでの課題をフィードバックできたため、比較的安定して検知可能なデータを取得できるようになった。また、本デバイスの開発により、実用性を考慮した以下の課題と解決策を明確にすることができた。

- 課題1：座ったり寝たりするとノイズが発生してしまう。

解決策：起床／臥床時を検出する仕組み（圧力センサー等）を組み込むことによる、ノイズの除去。

- 課題2：センサーデバイスのほかにデータ収集用スマートフォンが必要。

解決策：遠距離の通信規格（WiFi/LTE）を活用したセンサーデバイスの開発。

AIの開発では、主にセンサーデータの質の向上に組み、1回の排泄記録を1データとする取得目標数を2,400データと設定していた。しかし、新型コロナウイルス感染症（COVID-19）による実証実験現場の閉鎖（感染クラスターなど）の影響で実験実施が簡単に実施できずにいた。さらに、実験協力者の拡充を進めていたがコロナ禍での要介護者への負担増加や家族説明の難しさという問題から同意を得ることができず、おむつへの装着

型デバイスを用いた実証実験では、データ数1,198データという目標数の約半分程度の結果となった。

データの取得目標は未達であったが、取得できたデータを用いて尿便識別AIを開発した。

本事業では三つの尿便識別モデルを開発した。それらの中で最も性能の高い深層学習モデルでは、従来アルゴリズムを利用したモデルよりも「排便」の識別性能が2倍優れている結果を示した。

最後に、施設で行ったAIの精度検証実験で取得できたデータは全53件（正報：47件、誤報：5件、失報：1件）であった。

このうち排便のデータは4件のみであったが、本事業で開発したAIは実験中に排便の排泄通知ができた。チューニングなど改良の余地はあるものの便の特徴を捉えられていることが示唆される結果となった。

(4) 業務スケジュール最適化AIの開発

2021年度および2022年度に、業務スケジュール最適化AIと専用の画面を開発した。職員のシフト情報（勤務可能時間帯、出勤パターン）や資格情報と要介護者ごとの1日のスケジュールを入力することで、1日分の業務スケジュールを出力させることができた。

実証実験で得られた結果については、ベテランと新人における要介護者に対する介助時間の比較を表1に示す。なお、今回は移動時間を考慮せず、ナースコールなどの緊急性のある呼び出しは別途考慮した。

まず、業務スケジュールを用いない普段通りの介助時間をベテランと新人で比較した場合、新人の介助にかかる合計時間はベテランの合計時間に比べて最大でおよそ2倍の時間を要していることが分かった。一方で業務スケジュールを用いた場合、合計時間上では普段よりも時間を要する結果となっているが、利用者ごとの所要時間を見ると7名中4名で所要時間の減少が見られ、1名あたり平均で14.5%の所要時間が減少した。ただし、これらの結果は2の4)で前述したとおり、明確に条件をそろえるために補正を実施している。

上記の差が生まれた理由として、ベテランと新人の各業務において介護者自身の中で「迷いの有無」が目立った。ベテランは常に何時までに何を終えていなければならぬかを把握しており、ゴールからの逆算で業務を行う一方で、新人はその見通しが曖昧であるため、介助ごとに目移りする対応が散見された。業務スケジュールを使用することで、その迷いがなくなり次の業務が明確に

ビジョン実現に向けたアプローチ

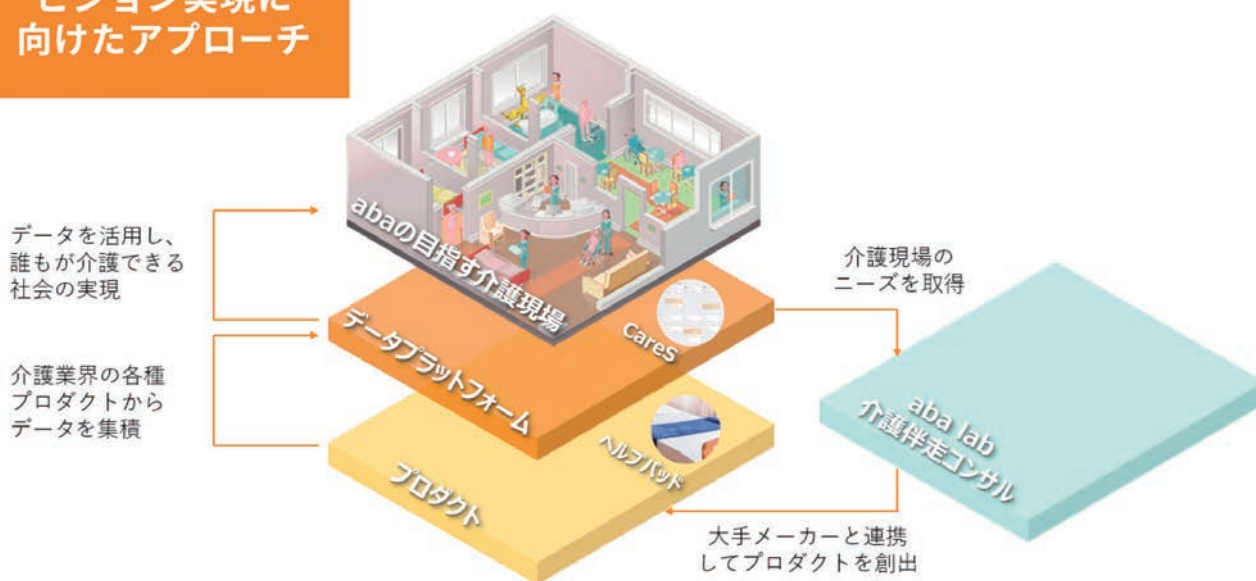


図3 今後のアプローチ

なり、介助時間の軽減につながったと考えられる。

実験後の新人へのヒアリングでも「スケジュールのおかげで迷いなく、業務に集中できた」という回答が得られた。本実験の結果では、移動時間について触れていないが、全体的な業務の見通しが立つことで業務と業務の間の移動時間が短縮される。また、ベテランと新人において業務内容の理解に差が存在しており、新人は曖昧な定義のまま理解・業務を行ったり、何を満たせば完了したことになるのかが曖昧となっていたりすることが分かった。これらは細やかなコミュニケーションが前提である介護現場において、人材教育に時間を多く割けない施設や事業所で起こり得る一種の「情報の非対称性」であるといえる。ベテランは普通だと思って認識していることが新人からすると当たり前ではないだけでなく、情報すら把握していない可能性がある。今後はベテランと新人の間に存在するノウハウや情報といった部分も加味していく必要があると考える。

(5) 生活スケジュール最適化 AI の開発

既存システムに対して下剤情報を管理できるユーザーインターフェースを追加し、排泄通知時に下剤に関連した排泄であるかなどの付加情報を提示する形へと改良した。今までは下剤による影響で漏れ出しているかなどの予想ができず、排泄介助を完了するまでに15～20分ほど必要していたが、下剤が影響している排泄である可

能性が高いという情報があることで、準備などの時間を約5分短縮することができた。

夜間の排泄介助タイミング・おむつ交換回数の適正化や睡眠センサーとの併用などの検証では、システムが算出した排泄パターンに基づいたタイミングで夜間の排泄介助を実施したことで、検証以前は夜間帯に平均約4回行っていた排泄介助を平均約3回に減少できた。また、検証以前に多かった空振りの回数も減少することで、間接的に利用者の睡眠効率の向上が見られたという結果となった。排泄センサーの社会実装という視点でも、睡眠センサーとの連携は有効性があり、「排泄×睡眠」は有効な組合せであることが分かった。

(6) トイレ内対話型システムの開発

動作や精度検証時に収集した排泄データを用いて、排泄検知エンジンを開発した。abaの排泄検知AIでは、十分な精度検知までに15分を必要とし、一般的な排泄行動としての所要時間を超えてしまうことや、対話型システムに必要な速やかなシナリオ展開が困難であることが判明したため、検出精度は劣るが1分程度で判定できるエンジンを開発した。

また、トイレ内外で生じる雑音による発話処理が遅延する事象が多発したが、発話を処理しているMICSUS側で調整いただき、実使用に堪えられるシステムに改善することができた。ユーザビリティ検証については、2

施設で実証実験を実施した結果、利用者8名中6名がトイレ内での対話が「退屈しない」、「楽しい」という好意的な意見が得られた。一方で8名中2名から、トイレ内で話しかけられて「びっくりした」、「対話リズムが噛み合わなかった」という意見があった。また、介護職員4名中4名からはトイレの見守りが人ではなくロボットやシステムだと人目を気にする必要がなくて良いという意見が得られた。

両者の意見から本システムがトイレ内で利用者に対し、抵抗感を感じさせることなく対話できる可能性を見出せたため十分なニーズも確かに存在していることが分かった。

一方で、システムからの声掛けが行われないと利用者には不快感を覚えさせてしまう場合もあると分かったため、利用者を選ばず確実に対話を始められる仕組みの検討が今後の課題である。実証実験で得られたシステムの課題として、次のような点が挙げられる。

- 対話開始判定の改善
- 対話の反応精度改善
- 利用者からの問い掛けから始まる対話応答の実装
- 排泄判定の改善

今後はこれらの課題を解決するためにさらに実証実験を重ねる予定である。

(7) 介護業務活動記録技術の開発

ビーコンを用いた実験では、位置情報や業務内容などを組み合わせて動線を可視化できた。このような業務活動記録があれば、新しい介護支援システムの導入前後で介護職員の行動がどのように変化したのかが確認でき、導入されたシステムを定量的に評価することも可能である。また、職員のスキルセットの違いや、個人のくせなどによる業務効率の低下、および熟練職員のノウハウをデータから抽出して勉強会で共有することで、人材教育にも生かすことが可能となり、職員スキルの底上げが期待できる。

さらに、2020年頃から世界中で猛威を振るっている新型コロナウイルスなどの感染症は、介護施設の利用者にとって重症化リスクが高いため、陽性者が出た際には移動経路の履歴から濃厚接触者の特定や、早い段階での健康診断や隔離といった感染拡大の抑制対策にも期待できる。

4 まとめと今後の展望

本事業では、認知的インタラクション支援技術で構築したシステムを、介護者支援事業へ活用するための一部として、以下のような内容を実施した。

- 排泄センサーデバイスおよび尿便識別 AI 開発
- 業務スケジュール最適化 AI の開発
- 生活スケジュール最適化 AI の開発
- トイレ内対話型システムの開発

今後の展開については、以下のように構想している。

- 排泄センサーに関しては、2023年秋頃を目処に、SIP内で開発した尿便識別AIを搭載した排泄センサーを発売、社会実装する。
- 今後の普及計画としては、2030年までに30万台の普及を目指している。

これは、本排泄センサーがターゲットとしている「寝たきりで意思疎通の取れない高齢者/障害者」が国内のみで160万人おり、その20%のシェア率から算出している。なお普及させていく中で、尿便識別はもちろんのこと、多様な排泄情報をさまざまなセンシング方式/AIによって検出できるように進化させていく。最終的には、排泄臭からの疾病検出についても研究開発を進めている。

次に業務スケジュールならびに生活スケジュールに関しては、2025年頃にサービスリリースを目標としている。2025年のリリース時点では、SIP内で開発を進めてきたスケジューリング機能はもちろん、介護現場の業務可視化なども目指していく。施設導入数は2030年時点で、国内のみで数千施設を目指す。

トイレ内対話型システムに関しては、今後排泄センサーの普及をさせていく中で、トイレ内での排泄検知も目指していく。その中でトイレ内の対話型システムの導入も見据えていく。

今後は、これらの内容をデータプラットフォームとして展開すべく、排泄情報を基軸とした介護業務スケジュールの最適化およびケアの質向上を実現するシステム「CareS」や一部の成果について、製品化を視野に入れている(図3)。

ヒューマン・インタラクション基盤技術の社会実装に向けたオープンイノベーション活動

国立研究開発法人産業技術総合研究所、一般財団法人デジタルコンテンツ協会

- 「ヒューマン・インタラクション基盤技術コンソーシアム（メタコンソ）」を設立した。
- メタコンソのメンバーに対してヒューマン・インタラクション基盤技術開発プロジェクトの技術カタログを公開した。
- ニーズ調査として50社に対する調査を実施し、さらに絞り込んだ10社に対して追加ヒアリングを実施した。
- 追加ヒアリングを実施した企業のうち4社と社会実装に向けた実証実験を進めている。
- 接客業務訓練支援システムの普及を目的とした「接客業務訓練支援システム開発ガイドライン」を作成・公開した。

1 はじめに

「ヒューマン・インタラクション基盤技術」の社会実装に向けたオープンイノベーション活動を実施することを目的として、「ヒューマン・インタラクション基盤技術コンソーシアム（メタコンソ）」を設立し、成果である高度人材育成システム、学習支援システム、介護支援システム等を用いてビジネスを展開する企業、ユーザーとなる企業、さらには将来に向けての改善や現場適合を実施する企業の参画を促し、サービス産業への成果の社会実装を支援・加速する仕掛け・仕組みを構築した。

本コンソーシアムの活動については、Webサイト (<https://hi-conso.org/>) において情報発信を行っている (図1)。Webサイトでは、各プロジェクトが開発している技術について、より分かりやすく伝えることを目的として「株式会社コトバデザイン」、「国立研究開発法人産業技術総合研究所」、「東京大学」による3件の接客支援技術、「NTT コミュニケーション科学基礎研究所」、「国立研究開発法人理化学研究所、株式会社フォトンラボ」による2件の教育支援技術、「株式会社アルム、東

京慈恵会医科大学、日本テクトシステムズ株式会社、データセクション株式会社」、「株式会社aba」による2件の医療・介護支援技術のPR動画を作成し、Webサイトで公開した (図2)。

本コンソーシアムの統括責任者は、持丸正明氏 (国立研究開発法人産業技術総合研究所 人間拡張研究センター 研究センター長)。事務局は、国立研究開発法人産業技術総合研究所と一般財団法人デジタルコンテンツ協会が共同で運営を行っている。

メタコンソでは、以下の3点に資する活動を実施している。

- ニーズを充足するオープンな技術開発。
- 市場ニーズの取り込み (技術開発企業の企業ニーズも含む)。
- 技術を社会実装するための社会的・ビジネス的ハードルの低減に資する活動。

メタコンソとしてNEDOと共催で、2022年10月26日 (水) ~ 10月28日 (金) に幕張メッセで開催された「第2回XR総合展秋」において、「ヒューマン・インタラクション基盤技術」の研究開発成果を紹介することを目的とした出展を行った。



図1 Web サイト (https://hi-conso.org/)

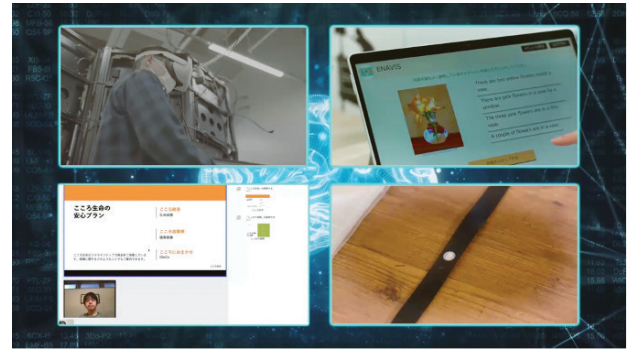


図2 PR 動画

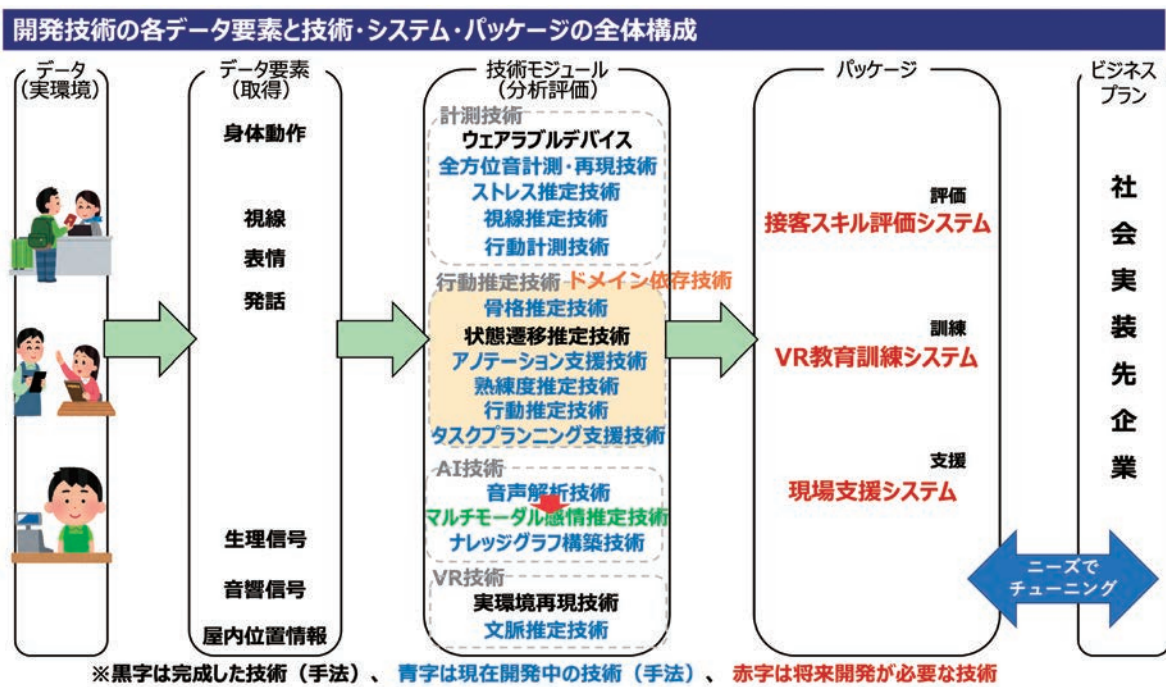


図3 技術カタログ

具体的には、「マルチモーダルデータによる自動状態記述システム」(国立研究開発法人産業技術総合研究所)、「飲食業向け気づきスキルVRトレーニングシステム」(国立研究開発法人産業技術総合研究所)、「インタラクションを支援するリモート接客システム」(株式会社コトバデザイン)、「インフラ領域における職人の技の伝承教育と機器実装」(国立研究開発法人理化学研究所)、「遠隔医療AIが連携した日本式ICT地域包括ケアモデル」(株式会社アルム、東京慈恵会医科大学、日本テクトシステムズ株式会社、データセクション株式会社)、「ヒューマン・インタラクション基盤技術コンソーシアムの紹介」(ヒューマン・インタラクション基盤技術コンソーシアム)の6事業についての展示を行った。

また国立大学法人東京大学が開発した「サービス産業従事者のためのVR技術を用いたトレーニングシステム」と国立研究開発法人産業技術総合研究所が開発した「飲食業の接客担当者向けのVR訓練システム」のデモ展示を行い、800名近くの方にSIP事業の成果を体験いただいた。

2 ニーズを充足するオープンな技術開発

ニーズを充足するオープンな技術開発として、「ヒューマン・インタラクション基盤技術」開発プロジェクトの技術カタログ(図3)を、メタコンソのメンバーに対

して公開している。技術カタログでは、計測技術、行動推定技術、AI技術、VR技術などの技術モジュールについて、完成した技術、現在開発中の技術、将来開発が必要な技術に分類して紹介している。

技術カタログでは、各技術モジュールについて、概要、特徴、提供方法、利用についての紹介を行い、それら技術モジュールに興味を持った方が、事務局を通じて各担当者に連絡を取ることができる仕組みを構築している。

3 市場ニーズの取り込み (技術開発企業の企業ニーズも含む)

開発技術の社会実装に向け、想定される導入企業を対象に接客における研修・教育訓練の課題とニーズを捉えることを目的に、開発中の各システムに対する企業の受

容性について2021年度にヒアリング調査を実施。

流通・メーカー、物流、自動車、外食、娯楽、観光、金融、そのほか（生活関連）の業界区分における約25種の業種を対象に、上場企業を中心とした50社に対し、オンラインまたは対面によるヒアリング調査を行った。具体的なイメージができるよう、「接客スキル評価システム」、「VR教育訓練システム」、「現場支援システム」という三つのシステムについてのヒアリングを行った。

「接客スキル評価システム」（図4）はAIを活用して、実際の接客現場や研修用の模擬的な設備で、従業員の接客スキルを客観的に評価することができるシステムである。データに基づく客観的な接客スキルの評価が可能となっている。このシステムで測定・評価したデータは、ほかの二つのシステムでも活用できる。「接客スキル評価システム」のヒアリング結果は、接客スキルが客観的



図4 接客スキル評価システム



図5 VR教育訓練システム

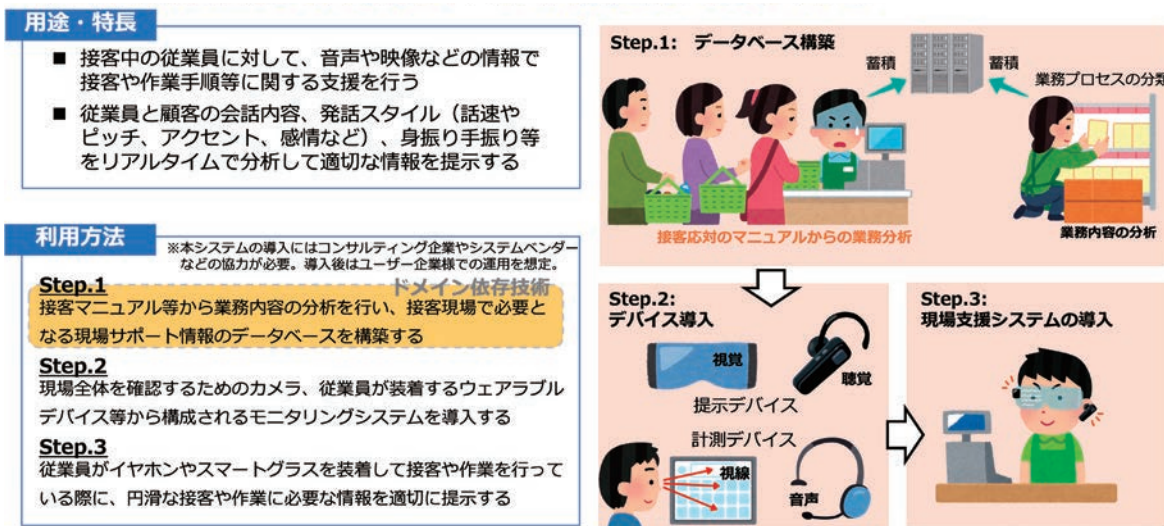


図6 現場支援システム

に評価できることでスキルの可視化、課題点の抽出ができ、社員の教育レベルが見える化できることに期待が大きい。さらに、自身のスキルレベルが見えることで、社員のモチベーション維持につながることを期待された。一方、接客の多様な業務への対応、納得できる評価基準が設定できるかという点でハードルの高さを感じられた。

「VR教育訓練システム」(図5)は、各企業の接客マニュアルや接客手順に合わせた訓練が可能なシステムである。ヘッドマウントディスプレイと呼ばれるVR用のヘッドセットを活用することで、臨場感の高いトレーニングを実現している。訓練者のスキル度合いやストレス度合いなどに応じて、訓練内容をインタラクティブに変更することができる。「VR教育訓練システム」のヒアリング結果は、新入社員や臨時スタッフなど経験の浅いスタッフへの基礎教育への活用が期待され、自習スタイルや複数同時実施などで教育のための人員、時間削減する

ことへの期待が大きい。一方、個々の顧客に応じた対応が難しく、VR環境を構築することへの費用対効果が疑問視され、VRである必要性を感じられないという意見が見受けられた。

「現場支援システム」(図6)は、実際に接客している従業員に、接客相手からの質問や苦情に対して、必要な情報やアドバイスを提示するシステムである。スマートグラスなどを着用することを想定している。「現場支援システム」のヒアリング結果は、複雑な対応、難易度の高い対応などに、迅速かつ的確に反応するサポートシステムとして期待された。特に、顧客の目に触れない業務分野での活用がイメージされ、顧客の感情が可視化される、多言語対応、ジェスチャーでの指示出しなどができれば面白いとの意見が寄せられた。しかしながら総じて、顧客の前での装着は不向きとの評価となった。

開発システムに対する評価の全体集計結果としては、「接客スキル評価技術」、「VR教育訓練システム」に対し

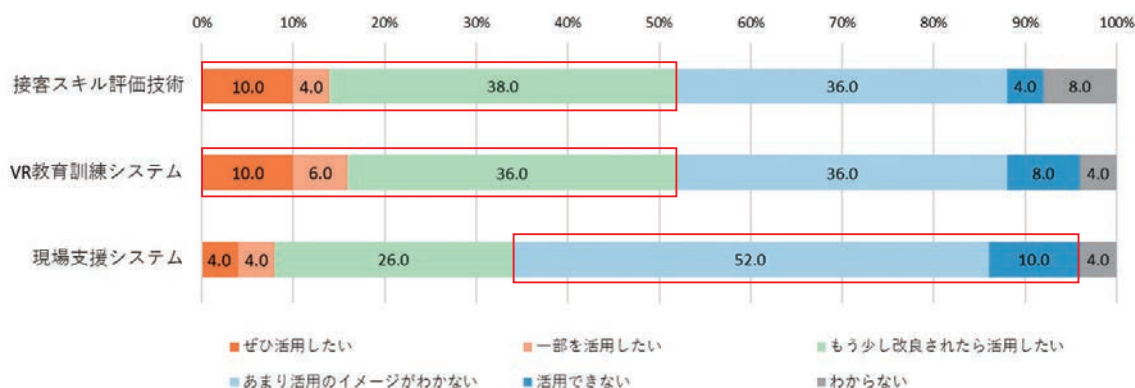


図7 開発技術に対する評価 (50件)



図8 提供システムの全体イメージ

では「ぜひ活用したい」、「一部を活用したい」、「もう少し改良されたら活用したい」との回答が全体の5割を超え、期待の高さがうかがえた。一方で「現場支援システム」は「あまりイメージがわからない」、「活用できない」との回答が6割以上を占める結果となった(図7)。

4 技術を社会実装するための社会的・ビジネス的ハードルの低減に資する活動

前項で実施したヒアリング調査結果をもとに、特に導入の可能性の高い企業10社に対する詳細ニーズヒアリングを実施するとともに、個別のニーズに合わせたシステム概要を設計する追加調査を2021年度に実施した。調査方法としては、前回の調査よりもより具体的なシステムをイメージしていただくために、動画を用いたシステム説明を行い、それに対する意見をいただいた。

追加調査の結果を踏まえ、大日本印刷株式会社の協力のもと、社会実装に向けた以下の4件の実証事業を実施している。

- 保育業務訓練：幼稚園や保育園の午睡時に行われる幼児の呼吸確認の見守り行動(午睡チェック)に係る訓練への、本研究開発成果の活用可能性に関する実証実験
- 小売業務訓練(2件)：スーパーマーケットの(セルフ)レジ・コーナーにおける顧客への挨拶やお声掛け等の顧客対応業務に係る訓練に対する、本研究開発成果の活用可能性に関する実証実験
- 宿泊業務訓練：ラグジュアリーホテルのカウンターにおける宿泊客のお困り事への対応等の顧客対応業務へ

の、本研究開発成果の活用可能性に関する実証実験

これら実証実験の結果、訓練の効率化・コスト削減、教育の最適化、スキルの可視化・評価などの効果を明らかにすることで、今後の事業化に向けた取り組みが進められる(図8)。

5 今後の予定

接客業務訓練支援システムの普及のため、「『XR接客トレーニングシステム』ガイドライン」を取りまとめている(図9)。2022年度末に完成予定で、完成後はヒューマン・インタラクション基盤技術コンソーシアムのWEBサイトでの公開を予定している。

ヒューマン・インタラクション基盤技術コンソーシアムは、今後会費を有償化することで継続して技術のマッチングや研究会を進めることを予定している。運営については、国立研究開発法人産業技術総合研究所のコンソーシアム活動とも連携を図り推進することを予定している。

テクノロジーを活用した 接客トレーニングシステムの開発に向けて

「XR接客トレーニングシステム」ガイドライン ダイジェスト版

社会の変化が激しいなかであって、サービス産業界は新たな課題に直面している。コロナ禍の影響から、接客サービスの自動化やリモート化が加速する一方で、十分なスタッフの確保も難しくなっている。

日本の強みである「おもてなし」を活かし、人と人がふれあう接客の現場に、テクノロジーがどのように貢献できるかについて検討を重ね、『業務訓練・支援システム開発のためガイドラン』を作成した。

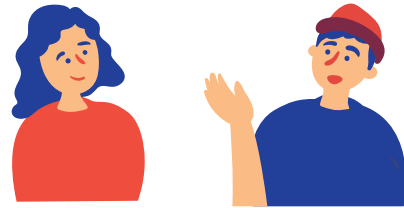
これは、そのダイジェスト版である。このガイドラインでは、VR(人工現実)・AR(拡張現実)・MR(複合現実)・SR(代替現実)を含めたXRを活用した接客トレーニングシステムにより、サービス産業界の生産性向上とさらなる発展を目指している。

サービス産業に携わる方々には、テクノロジー活用の可能性について理解を深めていただき、また開発者の方々には接客サービスについて、さらに深く知っていただくことを目的としている。

各項目について、ガイドライン本編の対応箇所を参照していただければ、詳細な情報と解説が得られるようになっている。

「おもてなし」は日本の強み!
海外からも評価が高いよ!

テクノロジーで、
その強みをサポート
できるのかな?



VRなどのXR接客トレーニングシステムのメリットや課題

Q1

VRって何?

バーチャルな空間で、
あたかも現実のような
体験ができること!

Q2

VR接客トレーニング システムを利用すると 何がよいの?

トレーナーがいなくても、自分のペースで訓練を受けることができたり、普段起こりにくいクレーム対応が体験としてトレーニングできたり、また人によって伝える内容にバラつきが出るのを防ぐこともできるよ。

目録 第6章 6.13(p.52)

Q3

VR接客トレーニング システムの課題は?

クオリティの高いシステムを作るには高額な費用が必要だったり、トレーニングを受ける人が多い場合、HMD(ヘッドマウントディスプレイ)などの機器の購入費もかかる。また、訓練するシナリオに変化がないと、慣れてしまって成果が出づらいこともあるよ。

目録 第6章 6.13(p.53)

Q4

どんなトレーニングが VRに向いているの?

たとえば火災が起こった時の対応や、難しいクレーム対応など、発生する頻度が少なかったり、経験することが難しかったりするものなどは、VRでのトレーニングに向いているよ。

目録 第8章 8.3.1(p.70)

図9 「XR接客トレーニングシステム」ガイドラインダイジェスト版

スマートシティなど未来社会に不可欠な分散型のデータ連携基盤と日本の戦略

国立大学法人東京大学大学院情報学環教授
越塚登

データ連携の現状、日本は進んでいるのか、遅れているのか

—— 越塚先生がサブプログラムディレクターを務める「分野間データ連携基盤技術」ですが、なぜ今、データ連携が重要な課題となっているのでしょうか。

越塚 実は今、日本ではどこに何のデータがあるのかが分からない状況となっています。例えば、鉄道やバス、道路などの交通に関するデータも、それぞれ個々の分野の中だけで管理されているので、それらの分野以外の人には、どこから必要なデータを取得すれば良いのかが分かりません。日本はヨーロッパなどと比べ、各分野の中ではデータ基盤もしっかりできているし、データ流通の仕組みも充実しているので、データの利活用が進んでいます。それなのに、日本の縦割り社会の中では、分野を超えたデータ収集ができないことで、いろいろな弊害が生まれています。一方、ヨーロッパではIDSやGAIA-Xなど、データ連携の取り組みが進んでいます。

例えば、都市全体を網羅するスマートシティのサービス開発を考えた場合、そこには住民サービスや教育、福祉、医療、交通、防災など、さまざまな分野が関わってきます。これらの分野を連携させて、都市全体でサービスを提供しようと考えても、現状では困難なことが多い。また、防災だけを捉えても、災害が起きる自然現象には雨もあれば、雪、地震、津波、土砂崩れ、火山噴火などいろいろな要因があります。ですが、それぞれの記録を管理している機関に蓄積されているデータをつなげて関連づけることができません。

これにより今後、重要な弊害が出てくる可能性が指摘されています。産業分野でいえば、サプライチェーンマネジメントの課題です。今後、ヨーロッパに工業製品を輸出する際には、製造時にどれくらいCO₂を排出したのかについて、データを提出することになりそうだからです。例えば、電気自動車のバッテリーを製造する場合、材料から薬品、化学物質に至るまでさまざまな分野の企業から部品を調達し



越塚登(こしづか・のぼる)

1994年東京大学大学院理学系研究科情報科学専攻博士課程修了、博士(理学)。1994年東京工業大学理学部情報科学科助手。1999年東京大学情報基盤センター助教授。2006年東京大学大学院情報学環助教授。2009年東京大学大学院情報学環教授。2015年東京大学総合分析情報学コース長。2017年東京大学大学院情報学環ユビキタス情報社会基盤研究センター長。2019～2021年東京大学大学院情報学環長・学際情報学府長。専門は計算機科学。特に、IoT(モノのインターネット)やオープンデータ、スマートシティなどの研究に取り組んでいる。

ますが、その製品をヨーロッパに輸出する際には、各社が排出したCO₂のデータを全て集める必要があります。そうして集められたCO₂排出量に、関税をかけようとする動きがあるのです。

こうした世界情勢から、ものづくりの現場においても分野間データ連携基盤技術が必要になり、サプライチェーン全体でCO₂排出に関する情報が集められない会社は、国外に製品が輸出できなくなり、投資機関からの融資が受けられなくなるかもしれません。

そこで、本研究テーマでは、産学官に散在するデータ基盤を分野横断で連携させ、AIなどで活用可能なビッグデータを供給するための技術開発を進めてきました。そのシステムのことを、CADDE(ジャッデ)と呼んでいます。そして2020年に設立された「一般社団法人データ社会推進協議会(DSA)」に、CADDEの成果をインプットし、分野を超えたデータ連携を目指すプラットフォーム「DATA-EX」の提供を2023年から開始する予定です(p.162～165参照)。これによって、日本でも本格的に分野間データ連携を社会に普及させていきます。また、分野間だけではなく、サプライチェーンなどにもデータ連携で貢献できる範囲を広げることができると考えています。

国際的なイニシアチブが求められる「スマートシティ・アーキテクチャ構築」

—— もう一つの研究テーマとなっている「スマートシティ・アーキテクチャ構築」についても教えてください。

越塚 今、世界的にスマートシティ 2.0の流れが起きていますが、日本はまだ対応が遅れているといわれています。一方で、日本にはすでにデジタルを使ったさまざまな都市サービスがつくられており、そこでの課題も分かってきました。次のポイントは、何をつくるかではなく、どうつくるかです。すなわち、「What to make」から「How to make」になります。それぞれのサービスは充実していても、つくり方をきちんと考えないとコストが下がらないし、サービスの信頼性も上がりません。日本の場合は都市全体でスマートシティに取り組む必要があり、民間企業が個別に進めていたら、いつまでたっても構築できない分野もあります。ですから、「How to make」を考える必要があるのです。

そして、「How to make」を考えるには、プラットフォームが必要になってきます。例えば、企業活動の中でも、プラットフォームなしにデータの利活用を優先して進め、いろいろなサービスをつかって

しまった結果、気がついたらデータベースが 50 個くらいあるけど、全然つながっていないという状況が起きるわけです。スマートシティにおいても、このような状況に陥って、これ以上の拡張は無理ですという状態になるかもしれません。

例えば、東京のある都市でスマートシティのサービスを構築したので、同じサービスを今度は大阪の都市で展開しようと思っても、プラットフォームがないと、もともと作り直しになります。なぜそうなるのかというと、東京と大阪にプラットフォームがないから連携ができないのです。これからこういうところに限界がくるでしょう。

スマートシティで利用されるプラットフォームは、「都市 OS」と呼ばれています。「How to make」を考えるには、まず都市 OS をどうつくっていくのかをよく考えなければいけません。とはいえ、都市 OS は単なるプラットフォームとなるシステムにしかすぎないので、それ以外にも上位階層でのルールづくりが必要になります。また、サービスもつくる必要があります。ハードウェアの面でも、さまざまな場所にセンサーを埋め込んだり、アンテナを立てたりしないといけない。やらなければならないことが数多くあります。そうしたプラットフォームをどのように作り込み、どのように連携させるのか。この How to make をまとめた、スマートシティ構築のための設計書が「スマートシティ・アーキテクチャ」になります。

2022 年の「スマートシティ・アーキテクチャの構築」というテーマにおいては、その設計書を完成させる取り組みを進めています。設計書自体はすでに 2 年前につくられていたので、今はそれを改定するためにさまざまな要素をまとめています。さらに、今後はスマートシティ・アーキテクチャの世界展開も進める必要があります。今、世界の都市のほとんどがスマートシティになっていて、自動車ですらスマートシティのノードの一つになってきました。こうなってくると、スマートシティ全体の技術的な覇権を取っていないければ、自動車の輸出すらできなくなってしまいます。

—— 日本のスマートシティ政策は、SIP の成果であるスマートシティ・アーキテクチャに基づいて進められていると考えていいのですか。

越塚 例えば、首相官邸が進めている国家戦略特区のスーパーシティ構想や、内閣官房が進めているデジタル田園都市国家構想、そして各省庁が進めているスマートシティ政策などもこのスマートシティ・アーキテクチャを参照しています。

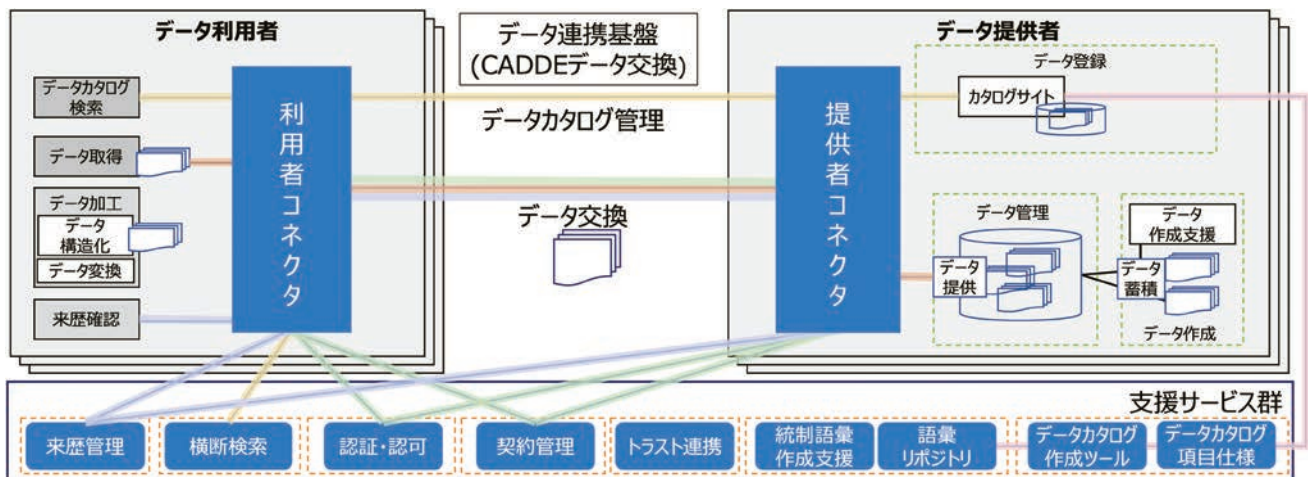
一方で、スマートシティには国自治体が関わることが多いので、多くの調達に公共調達になります。公共調達になると、工業製品などの各国の規格、および規格への適合性評価手続き（規格・基準認証制度）が貿易障害とならないよう、国際規格を基礎とした国内規格策定の原則、規格作成の透明性の確保を規定した「TBT 協定（Agreement on Technical Barriers to Trade）」に準拠する必要があります。今、日本はスマートシティの国際標準にはあまり積極的に関わっていません。このままでは、スマートシティの国際標準は全て海外でつくられてしまいます。これから、日本がどこまで国際標準に貢献できるかは分かりませんが、少なくとも、国際標準よりも優れたアーキテクチャをつくれれば認めてもらえるかもしれません。

このように、「分野間データ連携基盤技術」や「スマートシティ・アーキテクチャ構築」のテーマに関しては、国内だけの問題ではなく、国際的なイニシアチブを取れるくらいの実力を日本が持つべきだと思います。

日本の良いところを生かし、強くしていく

—— 人材育成なども含め、日本は今後どうしていけば良いのでしょうか。先生のお考えを教えてください。

越塚 分野間データ連携に関する人材育成は、まだこれからになります。大学でも、データサイエンス



CADDE のコンポーネント・アーキテクチャ

やデータエンジニアリングなどデータの名前がついている教育が始まっているのですが、問題なのは、これらの学部では、実はデータを使う部分にばかりフォーカスしていることです。本来データサイエンスには、データをつかって流通し、それを使って活用するというライフサイクルがあるのに、どちらかといえば AI や統計など、ライフサイクルの下流でデータを使う場面での教育が中心になっています。より上流でのデータをつくることやデータ流通に必要な人材育成に関してはあまり積極的ではありません。

データ連携については、今はヨーロッパがリーダーシップをとっています。アメリカはメガプラットフォームなど民間に任せようという立場、中国とインドは外国と連携する気配はありません。インドでは企業はデータを政府に提出し、政府がそれを積極的に使っていくという、政府主導の完全なトップダウン型となっています。どのやり方が良いか国際競争になっています。日本のやり方が正しいと思っていますが、日本のやり方で産業を成功させないと、ほかの国のルールに従わざるを得ないかもしれません。

とはいえ、日本の場合、分野単位でのデータ基盤は結構できあがっているので、データ活用に関して結構進んでいる部分もあります。したがって、人材育成に関してはヨーロッパだけを見て合わせようとするのではなく、日本の良いところも生かせるように国際調整する必要があるでしょう。

—— データ利活用における日本の良いところとは、具体的にはどういった部分なのでしょう。

越塚 日本人は、すぐに個々の問題解決のために手が動きます。ヨーロッパの人は、すぐに手を動かさずに、まずは物事の本質を見極めようと頭を動かします。汎用的な理論だけで勝負するところで覇権が決められると日本は弱いのです。国際会議の席においても、書類ベースだけで話をされると、日本人は負けてしまいます。今、国際標準組織は多かれ少なかれそういう世界になっており、国際標準化団体の中で日本はなかなか大変です。

しかし、日本人の強さは手が動かすことなので、すぐにシステムをつくったり、製品をつくったり、適切なサービスをつくったりするのが得意です。データの分野でもそういう力を利用して、世界覇権にうまく関わっていかなければなりません。そのためにも、できるだけ良いところを強くしていきたいと思っています。

データ連携基盤を活用した取り組み

分野間データ連携基盤技術(コネクタ)の機能ツール開発

株式会社日立製作所、日本電気株式会社、富士通株式会社、
大学共同利用機関法人情報・システム研究機構国立情報学研究所

- さまざまに分散されているデータ基盤を相互に連携させ、ビッグデータとして利用するためのデータ連携基盤に必要な技術として「CADDE (Connector Architecture for Decentralized Data Exchange)」を開発した。「CADDE」は「ジャッデ」と読み、トルコ語で「道」を意味する。
- 分野間データ連携基盤技術のコネクタ機能や支援サービス群機能、支援ツール群機能の開発を進めた。

1 研究の目的

国や地方公共団体、民間などで散在するデータ基盤を連携させてビッグデータとして扱い、分野や組織を超えたデータ活用とサービス提供を可能にする。そのために、関係府省庁で整備が進められている分野ごとのデータ連携基盤や、そのほかのさまざまなデータ基盤を相互に連携させる、分野間データ連携基盤技術(コネクタ)を用いた分散型分野間データ連携を実現した。

分野共通のコア語彙や分野固有のドメイン語彙、ならびにデータ構造などを整備する関係府省庁の動きと連携し、コネクタの開発および語彙の整備を通して、分野を横断したデータのインターオペラビリティ(相互運用性)を実現した。また、プラグイン構造などにより、さまざまなデータ形式への対応を目指した。

データ連携の原本性や編集履歴を保障する来歴機能や、契約および認証認可機能、カタログ構築の支援や構造変換技術などの運用支援環境も実現。開発した技術成果の国際標準化に取り組み、技術成果を国益に還元した。さらに、具体的な分野間データ連携のユースケースについて、フィールドでの有効性検証を行い、その結果をコネクタの安定版に取り込んでいく。

最終的には、産官学に散在するデータ基盤を連携させ、AIの学習データなどとして活用可能なビッグデータを供給する、分散型分野間データ連携が持続的に自立運用

できるエコシステムの形成を目指している。

2 実施期間と方法

(1) 実施期間

2018年度～2022年度

(2) 実施方法

1) コネクタ機能開発

① コネクタアーキテクチャ

Society 5.0の実現に向けた課題の一つとして、さまざまなデータを活用した、新たなビジネス創出のための仕組みが必要というものがある。それぞれの分野ごとにデータ連携の基盤が整備されつつあるものの、データの活用といったものは、それぞれの分野内に閉じてしまっている、というのが現状である。また、どこかの分野のデータ基盤が、他の分野とつなぐための仕組み、特に、共通的な仕組みをつくることは難しい状況であり、分野ごとのデータ基盤の間をつなぎ、異なる分野のさまざまなデータを相互に連携させる、分野間データ連携のためのアーキテクチャとして、データ交換にコネクタという新しい技術要素を取り入れる。

本研究開発においては、これらのデータの発見、契約、交換、来歴管理という機能を、コネクタを基盤としたネットワークで一元的に実施できるというものを目

指す。

世の中でデータ連携基盤と呼ばれているものは、大きく二つに分けると、巨大なデータベースをつくるのか、データを交換するネットワークをつくるのか、という違いがある。

中央に巨大なデータベースをつくる、というアプローチは、誰がそのデータベースを運用・管理するのかなどの問題が発生し、なかなか実現が困難であると考える。

そこで、データ提供者とデータ利用者それぞれにコネクタを置いてもらい、コネクタが間を取り持つ形でデータを交換するネットワークをつくるというアプローチを

採用した。

安心・安全なデータ交換を実現する上で求められる、認証やアクセス制御、交換履歴の記録等はデータ交換に紐づく形で一貫通貫に実現しつつ、異なる基盤の間をコネクタが取り持つ形でデータ連携ができる、というのが本研究のメリットであり特徴でもある(図1)。

本研究の実施にあたっては、まずはオープンデータに対応するためのデータ交換機能を開発し、その後、アクセスコントロールや契約や許諾が必要なデータへの対応のため、認証、認可、データ連携契約機能を開発するという、徐々に機能を追加するアプローチを採用した(図2)。

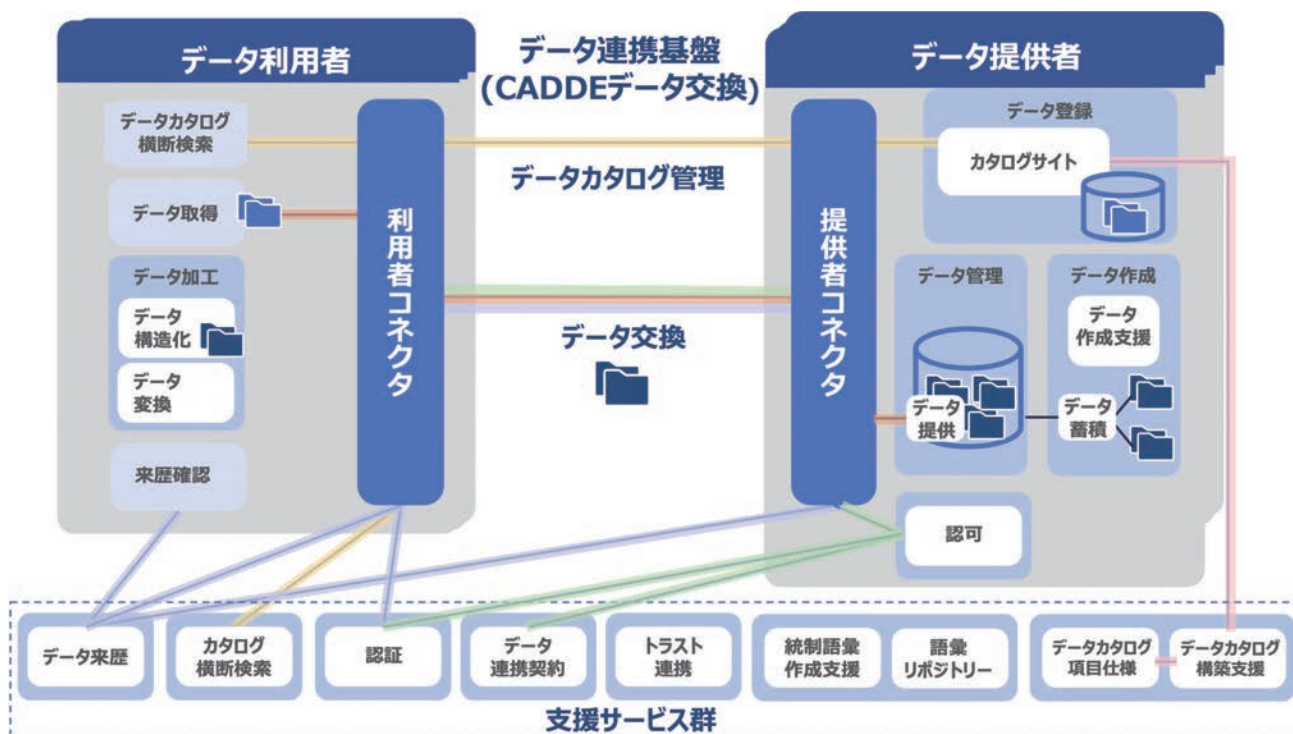


図1 開発したコネクタアーキテクチャ

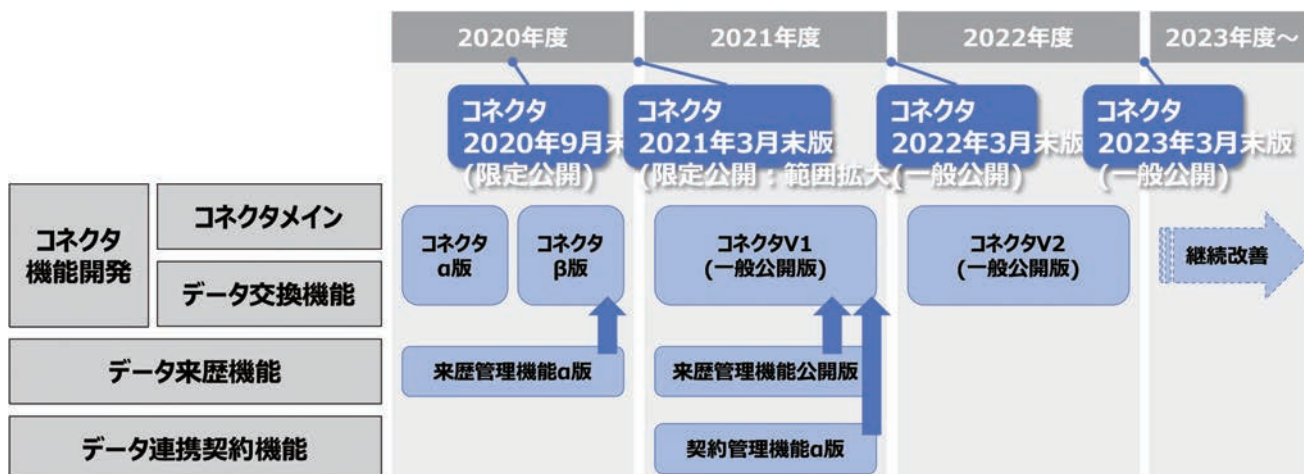


図2 開発アプローチ

α版は、開発コミュニティの立ち上げと特定環境での動作検証を目的とし、他 SIP と実証参加者と開発希望者へ限定公開とする。また、リリース時の動作確認状態は、コネクタ機能（正常系）動作確認済みとした。

β版は、開発コミュニティの拡大と動作検証の環境拡大を目的とし、他 SIP と実証参加者と「データ社会推進協議会」(Data Society Alliance。以降、DSA。p.163 参照)の会員と開発希望者へ限定公開とした。また、リリース時の動作確認状態は、コネクタ機能（正常系 / 異常系）動作確認済みであり、特定環境での動作確認済みとした。

V1以降の一般公開版は、普及促進を目的とし、一般公開を行う。リリース時の動作確認状態は、コネクタ機能（正常系 / 異常系）動作確認済みであり、数多くの利用環境での動作確認済みとした。

β版と一般公開版では、Web アプリケーションの脆弱性確認およびファジングツールでの API 脆弱性確認を実施済みとした。

②データ連携契約機能

分野によっては、契約が必要なデータなどもあり、そのようなデータには、コネクタの外に存在する機能と連携して対応できるようにする。データ取引市場と連携して契約処理が可能となるようにした。

③データ交換機能

各分野が持っているデータにはさまざまな提供形態があり、その中でも、ファイルでのデータ交換、HTTP、FTP、NGSI への対応を優先的に対応すべきものとした。

④データ来歴機能

分野を横断するデータ連携において、データ利用者は、入手したデータを自身が所有するデータと組み合わせたり、加工したりして、新たなデータを生成し、今度はデータ提供者として、生成したデータを提供することが想定される。これにより、多様なデータが生成され交換される一方、データの生成過程が複雑化していく。そのような中で、データ利用者は、入手しようとするデータについて、大元の原本が何か、現状のデータに至るまでにどのような加工が施されたのか、どのように流通されてきたのか、といったデータにまつわる履歴を確認できないと、安心して利用することができない。

従来、組織内のデータについて、原本からデータに至るまでの加工履歴を管理して提供するデータの来歴管理の機能は存在していたが、組織間を流通するデータに対する来歴管理の機能は開発されていなかった。本開発で

は、分野を横断するデータ連携において、異なる複数の組織をまたいで生成されるデータにおいて、大元の原本から、組織間でのデータ交換履歴、組織内のデータ加工履歴を一気通貫で確認可能な新しいデータ来歴機能の実現を目指した。また、来歴の信頼性の観点では、改ざんへの耐性が必要である。分野間をまたぐデータの連携においては、特定の運用者に依存して中央サーバを置くことは想定できない。このような分散環境でのデータ来歴の機能を実現するため、耐改ざん性が高く、透明性を確保したままデータ共有を可能とするブロックチェーンの分散台帳を活用した。

2) 支援サービス群機能開発

①カタログ横断検索

分野間でデータ交換を実現するには、データが見つげられる (findable な) 状態でなければならない。このデータの発見機能を、本研究開発ではカタログ横断検索として実現した。

各分野、組織などで公開されるデータは、データカタログに収録され、必要な範囲で公開される。しかし、この個別のデータカタログがあるだけでは、必要なデータがどこにあるかを探すのは容易ではない。カタログ横断検索では、これらのデータカタログのデータを収集して、1 か所で検索と内容が確認できるようにした。

②語彙リポジトリ

分野間データ連携基盤において、語彙はデータ提供者が提供データを準備する段階で用いることができる。語彙とは、データで用いられている用語やデータの記述様式のことを指し、単に用語を集めたものから、統制語彙と呼ばれる一定のルールで集めた用語・概念の集合、分類体系、あるいは、概念の実例を記述するとき用いる項目の集まりである概念スキーマといったものを含む。この語彙を、データ提供者の間で共有することによってデータで使われている語彙を共通化することで、データを理解しやすくし、データ利用の促進を図る。この共有化のためのシステムが語彙リポジトリである。

データ提供者は、自らが持つデータを公開するにあたり、なるべく多くの人々が利用しやすいデータとして提供することが望まれる。データ提供者は、語彙リポジトリにある語彙を参照することで、共通的な語彙を適用して自らのデータをより利用しやすい形に変換してから提供することが可能となる。

③認証認可機能

データ流通において、安心・安全を確保するために、ユーザーを認証する必要がある、また、オープンデータ以外に特定の利用者を識別して提供されるデータに対応する必要があるため、認証機能および認可機能を開発した。

認証機能によって、認証要求者の正当性（正当なユーザー）と真正性（authenticity*1）を検証する機能を提供。さらに、認可機能により、認証によって識別されたエンティティ（主体）にサービス利用やデータ取得の許可を与える機能を提供することで、特定の利用者を識別して提供するデータの交換が可能となり、安心・安全なデータ流通が実現できる。

なお、それぞれの分野では、ID プロバイダーを保有しており、それらとの連携が必要となるため、認証機能は、ID プロバイダーとの連携も可能とした。

3) 支援ツール群機能開発

① データカタログ構築支援機能

データカタログとは、データを取引する際にデータの概要情報をリスト化したもので、データ利用者は、このデータカタログから目的のデータを探ることができる。

データカタログはメタデータを管理しており、メタデータにはそのデータを表す属性や関連する情報などのデータの概要情報が記載されているため、メタデータを見ることで、どのようなデータであるのかを端的に把握できる。

なお、データカタログは各分野でそれぞれ整備されているものの、データカタログ項目は共通化されておらず、各分野で独自の構成となっている。そのため、他分野のデータカタログを参照する際は個別に解読する必要があり、同じ内容を含む項目であっても名称が異なっているため、分野を横断して必要なデータを検索しながらの利用ができないという課題がある。

この課題を解決するため、国際的な標準（W3C DCAT）に基づき、分野をまたいだ項目内容の理解が可能で、分野を横断的に検索することが可能な項目を策定した。

なお、項目が多数存在し、データ提供者にとって、データカタログの多くの項目に入力することは、非常に大きな負担になることが想定される。この課題を解決するため、支援機能（データカタログ作成ツール）を開発した。

② 統制語彙作成支援機能

分野をまたがって他組織でデータを使用するには、同じ事柄を表す用語がデータソースごとに異なる表記であると、機械的に使用できない。IMI 共通語彙基盤では、分野横断でのデータ利活用を目的とし、コア語彙とドメイン語彙を連携させ、異なる分野のデータの相互運用性を高めることを目指すフレームワークが提供されている。

本活動は当初、IMI の考えを踏襲しつつ、同じ意味であるが分野ごとに異なる表記の用語を推定する、統計学的機械学習を用いた自然言語処理エンジンの開発を目指した。

そのためには、分野ごとの自然言語処理に必要な学習用データ（コーパス）が必要である。コーパス構築のために、「現代日本語書き言葉均衡コーパス」などの既存のコーパス構築方法や著作権などの関連法案について調査した。あわせて、コーパス構築ではなく実際に作成されている語彙の調査として、農業や金融、ものづくり分野の語彙の作成過程と作成結果を調査した。その結果を用いて、どのような分野間の語彙の連携を行えば良いか検討した。

3 研究の成果

(1) コネクタ機能開発

1) コネクタアーキテクチャ

分野間のデータ交換の実現に向けた、データ利用者とデータ提供者間のデータ交換を可能とするコネクタのメイン機能の設計および開発をコネクタ α 版（2020 年 9 月末版）として完了した。本コネクタメインは、複数のデータ取得インターフェースを選択してデータ取得を可能にするアーキテクチャとした。

また、必要に応じてデータ来歴機能とも連携し、データの受け渡し記録を取得することを可能としコネクタ β 版（2021 年 3 月末版）の設計と開発を完了した。

2021 年 3 月末には、登録制ではあるが、一般の方に利用していただくため GitHub でのコネクタ OSS 公開（限定公開）を行った。

2021 年度は、コネクタへの追加機能として、データ来歴機能と連携した組織間のデータ交換履歴、データ交

*1 主体がそれを主張する本人であること。

換機能としてNGSI (API) へ対応した。また、商用データの利用を可能とするための認証認可機能とデータ連携契約機能との連携機能、および組織間のデータ加工履歴と組織間のデータ交換履歴機能の設計・開発を実施し、2022年3月末にコネクタV1をOSS公開(一般公開)した。

2022年度は、社会実装を見据えたCADDEの高セキュリティ化として、多要素認証や認可の分散化を行ったことに伴い、コネクタと認証機能、コネクタと認可機能間のシーケンスの見直しを実施した。

また、NGSIインターフェースで提供するデータの来歴管理対応ができるようにデータ来歴機能とのインターフェースの見直しを実施した。さらに、コネクタ全体のエラーメッセージの強化を行い、2023年3月末にコネクタV2をOSS公開予定である。

コネクタはCADDEの特徴の一つであり、分野ごとにプロトコルなどがバラバラであっても、コネクタを導入することで、データ交換相手の仕様を意識することなく、お互いにデータ交換が可能となった。また、認証機能や認可機能、データ連携契約機能を必要に応じて組み合わせることで、安心・安全なデータ交換を実現する上で求められる、認証やアクセス制御、交換履歴の記録等について、データ交換に紐づく形で一貫通貫に実現しつつ、異なる基盤の間をコネクタが取り持ってデータ連携ができる、というのがCADDEのメリットである。

2) データ連携契約機能

2021年度に、有償・非公開データを契約に基づいて利用者に提供するための、データ利用権取引市場との連携インターフェースおよびシーケンス策定をもとに、契約を要するデータ交換をデータ取引市場と連携して実現した。また、一つのデータ取引市場(EverySense, Inc.のサービス)と連携実装を行った。その際、複数のデータ取引市場との連携も可能にするため、CADDEとデータ取引市場間のAPIを規定した。

2022年度は、社会実装を見据えたCADDEの高セキュリティ化に伴い、データ取引市場との連携においても認証を実施するAPIに改定した。

分野によっては、アクセスコントロールや契約や許諾が必要なデータもあり、そのようなデータに対して、コネクタの外に存在する機能と連携して対応できるようになった。また、アクセスコントロールが必要な場合、認証機能と連携した利用者の認証や、契約が必要な場合、

データ取引市場と連携して契約処理を行うことも可能となった。

3) データ交換機能

2020年度は、コネクタ内の各機能は独立性の高いモジュールでの構成(ビルディングブロック)とし、新たな要件等への対応が必要になった場合でも簡単に対応できるように設計した。

取得するデータに合わせたインターフェースもビルディングブロック構成で設計し、3種のインターフェース(①FTPインターフェース、②HTTPファイルダウンロードインターフェース、③NGSIインターフェース)の開発を完了した。

2021年度は、取得したデータを加工した後での来歴が取得できる機能を開発した。

2022年度は、データ来歴機能とのインターフェースにもたせるパラメーターの見直しを実施し、NGSIにおけるデータ提供時の来歴が取得できるようにした(図3)。

今後、上記以上に必要となるインターフェースがある場合、必要なインターフェースのみ追加することで、新たなインターフェースに対応したデータ交換が可能となり、CADDEの適用範囲が広がるものとする。

4) データ交換機能(NGSI)

スマートシティプラットフォームで広く普及している、「FIWARE」で使用されるネットワークAPIであるNGSI(NGSI v2)に対応し、データ提供者とデータ利用者の間においてNGSI v2形式のコンテキストデータの授受を行う機能を開発した。

また、認証機能や認可機能、データ連携契約機能、およびデータ来歴管理機能との連携機能を開発。コンテキストデータにおいてもファイルデータの授受と同様に、限定された利用者とのデータの授受や、データ取引市場などの契約に基づく有償データの授受、およびファイルの授受に関する履歴を登録することを可能にした。

5) データ来歴機能

従来、組織内で生成されるデータの原本情報やデータの改変や結合といったデータの加工の履歴、組織間で受け渡されるデータ交換の履歴を一元的に管理する機能は存在しなかったが、カタログ作成ツールやデータ交換機能、データ加工ツールといったデータ連携における機能

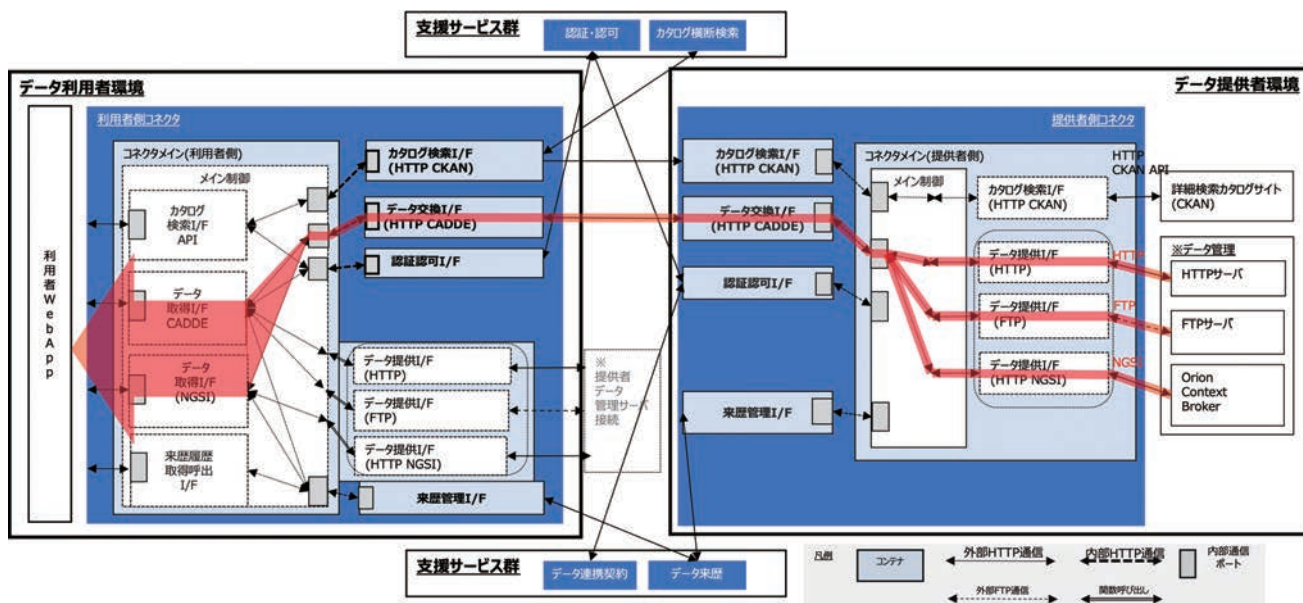


図3 データ交換機能

との協調および履歴登録のインターフェースや履歴のデータモデルの統一化を進め、組織をまたいだエンドツーエンドの来歴の管理および参加者への提供を可能とした。また、複数の異なる組織が参加するデータ連携においては、来歴情報の改ざんへの耐性が重要となるが、特定の運用者が集中して管理するのではなく、複数の運用者ならびに参加者が分散的に管理するブロックチェーンの分散台帳の活用により、どの参加者や運用者であっても途中での改ざんが困難な来歴情報の管理を可能とした。

また、分野間データ連携基盤技術(コネクタ)の有効性検証の行政・交通分野のデータ連携実証において、本データ来歴機能を適用し、データ提供者である自動車メーカーから提供されるデータの流通先の追跡を実現するとともに、データ提供者視点での有用性を確認した。本実証実験については、SBテクノロジーと富士通共同で2022年3月31日にプレスリリースを公開した。

(2) 支援サービス群機能開発

1) カタログ横断検索

カタログ横断検索は、インターネット上に分散して存在するカタログサイトから情報を収集して、検索および内容が確認できるようにするものである。実装においては、カタログサイトで世界的なデファクトとして使われている「CKAN」を基本として設計した。

収集機能ではCKAN APIを用い、収集したカタログ一覧あるいはカタログ更新一覧に基づいて、個別カタログ

のデータを収集している。公開機能としては、基本的には個別のカタログの情報を公開するが、いくつかの項目(idなど)については、横断検索内でのデータの一貫性管理のため、個別のカタログの情報から変換して生成している。検索機能では、基本的には全てのカタログのデータ項目が検索可能である。また、一部のデータ項目においては、データ項目とその値の組という形で検索を指定できる。横断検索自体も、CKAN APIに準拠したAPIを公開している。このAPIは、CADDEのコネクタなどのコンポーネントから利用される。さらにこのAPIを用いて、横断検索自身を複数設置しカスケードして運用することもできる。

CKAN以外のカタログサイトにも、一部対応している。また、国立情報学研究所が収集する学術データ(IRDB)も掲載している。2022年8月現在、62のカタログサイトから9万5,000件以上のカタログを収集公開している。学術データを含むと16万件以上になる(図4)。

2) 語彙リポジトリ

語彙リポジトリでは、語彙の登録、更新、検索に加え、CADDEのほかのコンポーネントとのAPI連携が可能である。

構築した語彙リポジトリでは、登録可能な語彙として、まずSKOS語彙と呼ばれる統制語彙がある。これはRDFスキーマあるいは簡易フォーマットで登録できる。簡易フォーマットとは、項目が規定されたCSV形式でSKOS語彙を記述するものである。SKOS語彙は統

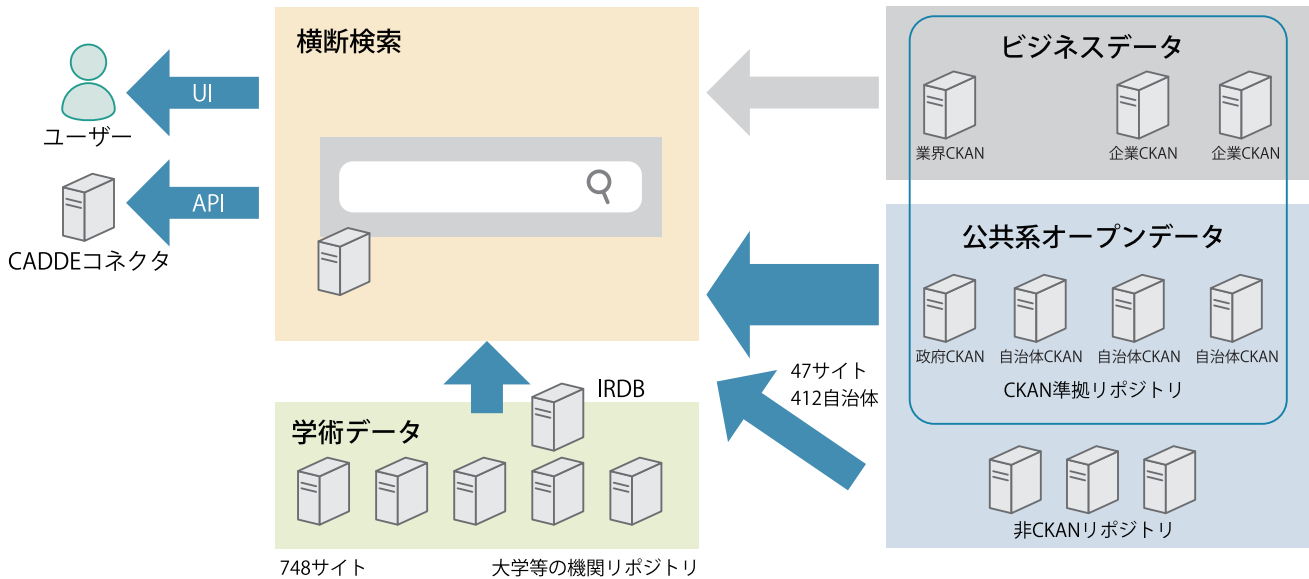


図4 横断検索の概念図

制語彙作成支援ツールでもサポートされており、この形式で2つのシステムはAPI連携する。このほか、RDFSやOWLで書かれた語彙全般、JSONスキーマも登録可能である。また、語彙共有を促進するため、ダブリンコアなど汎用のスキーマや統計LODの語彙なども事前に登録されている。

語彙リポジトリの基本インターフェースは検索画面であり、ここでキーワードなどを入力して登録されている語彙を発見する。その語彙で内容を確認し、必要であればダウンロードできる。語彙の登録においては、まずユーザー登録を行い、その上で自ら語彙データをその語彙に関する説明（メタデータ）と共に登録する。また、必要であればその語彙のprefix（接頭辞）を新たに取得できる。

3) 認証認可機能

有償・非公開データを契約に基づいて利用者に提供するための、標準仕様に基づく認証認可の連携仕様を策定し、アクセス制御を必要とするデータ交換に対応する認証機能および認可機能を2021年度に開発した。

2022年度は、社会実装を見据えたCADDEの高セキュリティ化として、多要素認証や認可の分散化を行った。

(3) 支援ツール群機能開発

1) データカタログ構築支援機能

分野ごとにデータカタログの項目やルールについては異なっているのが現状であり、異なる分野のデータカタ

ログの内容が理解しにくい課題への対応ため、国際的な標準（W3C DCAT）に基づいて、部門を横断しても理解しやすく共通的に利用できるデータカタログ項目を策定した。また、単なる共通化だけではなく、分野をまたがって検索する際に、必要となる項目として精査を実施した。

なお、精査を実施しても、必要なデータカタログ項目は89項目と多い状態であり、データを提供する人がデータカタログを作成する負担が多いため、機械学習を利用し、入力値を推測して提示する等の負担軽減のための仕組みを備えた支援ツールとして、カタログ作成ツールを開発した。これにより、データカタログ作成の工数を最大62%削減する効果が見られた。

2) 統制語彙作成支援機能

コーパス構築の調査の結果、著作権者からの利用の許諾といった著作権処理を非常に注意深く行う必要があり、また収集するコーパスに偏りがないよう注意する必要があることが分かった。著作権処理に関しては人間系で行わざるを得ず、技術的な解決は難しい。よく行われる手法としてはオープンソースソフトウェアの開発で用いられているコントリビューターライセンス同意があり、多人数の協力によるデータ生成の場合は権利譲渡のルールを事前に明記し、それに合意した上で作業を実施する方法がある。そのため、著作権処理は本研究の対象外とした。さらに、各分野で作成されている語彙の調査の結果、「語彙」として作成された用語の集合の構造が

それぞれ異なり、用語の一覧やその説明まで含めた辞書、または表形式データのスキーマに相当するものやそのデータ型まで規定されているものなど、オントロジースペクトラムという光の波長分解のようにさまざまな粒度を持つことが分かった。

語彙を構築する過程では、その分野の専門家と知能情報処理の専門家が共同で数か月かけて構築作業を行っており、その作業の基礎となるデータに、テキストデータのみならず、表形式データやデータベースのスキーマなど、統計学的機械学習の学習に適さないものも多く見られた。各用語やスキーマと IMI 共通語彙基盤のコア語彙との対応関係であるマッピングを行っている語彙は、調査した中では存在しなかった。さらに、時間が経つと新しい用語が出てくるため、語彙を一度作成して終わりではなく、語彙の更新が必要なことも分かった。

そこで我々は、各分野での語彙と共通語彙基盤のコア語彙のマッピングが実現可能であるかどうかを検証するため、ものづくり分野を例にとり、分野に紐づかないコーパスで学習した類似用語検索エンジンにより、定量的にマッピング精度を計算した。正解となるマッピングは、事前に手作業で作成した。その結果、検索結果の最上位は事前に手作業で設定した用語が出ることはほとんどなく、完全な機械処理ではマッピングできないことが分かった。

上記の調査および検証結果から、我々は機械処理のみによる語彙マッピングの方針を転換し、各分野での専門家の語彙作成を支援する機能の開発を目指した。オントロジースペクトラムの中でも最も構造が単純であり、すべての語彙作成の入口となる、同義語や上位語を定義した統制語彙の作成を支援することとした。

我々は新語に追従するために、分野の専門家のみでも統制語彙が作成できることを要件とし、語彙作成の作業負担を軽減する GUI ソフトウェアを作成した。開発したソフトウェア「Controlled Vocabulary Designer (CVD)」は、マウス操作のみで同義関係・上下関係が設定できる。AI による同義語や上位語の推定結果や、既存語彙の情報を参考にしながら操作が可能である。また、CVD では、作成した統制語彙を機械可読性の高い SKOS という、RDF 形式で出力できる。RDF を用いれば、全体部分関係 (part-of 関係) や属性関係 (attribute-of 関係) といった、より詳細な関係性を記述していく際の橋渡しを担うことが可能である。一方で、SKOS は人が見た際の一覧性に乏しく、分野の専門家が作業結果を確認した

り、ほかの人がレビューをしたりすることが難しい。そこで我々は、SKOS のサブセットとなる CSV フォーマットを定義した。同フォーマットは、CVD の入出力のみならず、語彙リポジトリの入出力にも対応しているため、作成した統制語彙を広く他分野に公開でき、一方で同一分野内においても複数人で、共同で更新作成作業を行うことが可能である。

CVD は GUI アプリであるため、機能性だけでなく、操作性は非常に重要で、我々は複数の被験者に対して UI/UX の評価を行いながらソフトウェアの改善を重ねてきた。我々が実施した評価実験では、CVD の操作に慣れれば表形式ソフトのみを用いて統制語彙を作成することに比べ、素早く不整合なく作成できることが確認されている。我々は本成果を、オープンソースソフトウェアとして公開した。

上記オープンソースソフトウェアと語彙レポジトリを用いて、いろいろな分野の持つ語彙を公開・共有することで、各分野の持つデータの意味的連携性を高めることが期待できる。我々は、ほかの SIP にご協力いただき、各プロジェクトで保持する語彙の中から、語彙を作成する作業を支援し、語彙を公開していただいた。ご協力いただいた方々に作業実施後にアンケートを実施し、RDF 化された語彙の公開により、データの表記ゆれ抑制やデータの理解に役に立つという意見を得ることができた。

4 まとめと今後の展望

(1) コネクタ機能開発

1) コネクタアーキテクチャ

アーキテクチャとして、各分野におけるデータ流通の仕組みを最大限尊重し、それらを各分野の特性に合わせて分散的に連邦化する、ビルディングブロック型とした。

コネクタのネットワークを通じてデータ交換が実現されることとなり、分散的に存在するデータ提供者とデータ利用者は、それぞれの窓口となるコネクタを用意することでネットワークに参加し、データ交換が可能となる。コネクタ間のデータ交換においては、必要に応じて認証認可や契約管理、来歴管理などの機能が呼び出されて利用される。

また、データ交換のプロセスだけではなく、データ記述語彙の共有からデータ発見、データ変換といった、デ

一タ利活用の一連のフェーズで必要となる機能を支援するツールやサービスも提供しており、幅広い活用が想定される。

その結果、Society 5.0の目指す、さまざまなデータを活用した、新たなビジネスの創出が期待される。

今後については、DATA-EXを推進する団体(p.165参照)による普及促進や、本研究開発に携わった企業・団体による機能拡充、およびサービス提供による社会実装が想定される。

2) データ連携契約機能

本機能の提供により、契約を要するデータ交換にも対応できるため、より適用範囲が広がることとなった。今後、データ取引市場の拡大が想定されるため、より広い範囲でのコネクタの普及を想定している。

また、本機能については、GitHubでオープンソースとして一般的に公開しており、さまざまな分野での活用が想定される。

3) データ交換機能

一般的なファイル交換プロトコルであるHTTPやFTPへの対応や、認証機能、認可機能との連携により、契約有無、有償無償データなど、幅広いデータ交換が可能となり、安心安全なデータ流通の促進への寄与が想定される。

また、本機能については、GitHubでオープンソースとして一般的に公開しており、さまざまな分野での活用が想定される。

今後については、競争領域において、本機能の提供のみにとどまらず、データの収集・分析等も含めたサービス提供などにより、さらなる社会実装への普及拡大を期待している。

4) データ交換機能(NGSI)

スマートシティプラットフォームで広く普及されているFIWAREで使用されるネットワークAPIであるNGSI(NGSI v2)への対応についても、FTPとHTTPと同様に認証機能、認可機能との連携により、契約有無、有償無償データなど、幅広いデータ交換が可能となり、安心安全なデータ流通の促進への寄与が想定される。

また、GitHubでオープンソースとして一般的に公開する点も同様であり、さまざまな分野での活用が想定される。

今後については、機能拡充などを行った上で、サービス提供などによる社会実装を予定している。

5) データ来歴機能

分野間における安心なデータ流通・利用の実現に向けて、データを基点として、大元のデータの原本情報およびデータ交換履歴、加工履歴を確認可能とするデータ来歴機能を実現した。本データ来歴機能は、今後の分野間や異なる組織間において転々と流通され改変されていくデータにおいて、データにまつわる各種履歴を、データ利用者から確認可能とするものであり、安心なデータ利用として必須となる機能である。

現在、富士通では本技術を発展させ、新サービス「Fujitsu Computing as a Service Data e-TRUST」として2022年度末より提供開始する予定である。

これにより安心安全で自由なデータ流通の両立を実現し、金融や製造、流通、医療分野などにおけるさまざまな課題解決に取り組む。

(2) 支援サービス群機能開発

1) カタログ横断検索

カタログ横断検索においては、デファクトで使われているカタログサイトのソフトウェアであるCKANに準拠する形でシステムを構築し、その上でいくつか拡張した。結果として、サイト数としては60程度であるが、自治体(市区町村)であれば400弱、数でいえば20%強、人口比でいえば30%強をカバーできた。ただ、これ以上のカバー率を目指すのであれば、CKAN準拠以外のサイトをどう取り込むかを検討する必要がある。

カタログ横断検索の運用においては、基本的には大きな問題はなく、運用そのものは軽微なコストである。しかし、長期運用にあたって運用体制を組む必要があり、これが今後の課題となる。

開発したソフトウェアはオープンソースとして公開している。今後、長期運用にあたっては、ソフトウェア自身の更新も必要となってくるため、その体制も検討する必要がある。

なお、自治体などのオープンデータに加え、学術データを取り込む総合サイトは世界的にも稀有であり、データ利用の可能性を広げたという点で意味がある。

2) 語彙リポジトリ

語彙リポジトリ開発においては、語彙の共有と再利

用を目指して、語彙リポジトリソフトウェアの開発と試験運用を行った。語彙を共有するメリットとしては、データの理解の向上、データ作成の簡略化、データの再利用性の向上があり、語彙リポジトリはそのハブ的な役割に期待される。

今回開発されたリポジトリは、語彙記述のデファクトである RDFS/OWL 言語をサポートすること以外に、CSV による簡易フォーマットもサポートしている。このことにより、語彙登録と利用のハードルが下がることが期待される。また、統制語彙作成支援ツール (CVD) と API 連携をすることで、ユーザーが語彙データに直接触ることなく、語彙の登録や利用ができるようになっている。この点も利用促進に貢献すると考える。

語彙リポジトリの運用においては、基本的には大きな問題はなかった。しかし、長期運用にあたっては運用体制を組む必要があり、開発したソフトウェアはオープンソースとして公開している。今後、長期運用にあたってはソフトウェア自身の更新も必要となってくるため、その体制も検討する必要がある。

3) 認証認可機能

本機能の提供により、オープンデータ以外にも特定の利用者を識別して提供されるデータの交換にも対応できるため、より適用範囲が広がり、広範囲でのコネクタの普及が想定される。また、本機能については、GitHub でオープンソースとして一般に公開しており、さまざまな分野での活用が想定される。

本機能により、分野間における安心安全なデータ交換が可能となった。

また、各分野における ID プロバイターとのつながりが容易となり、分野間のデータ交換が活性することが期待される。

(3) 支援ツール群機能開発

1) データカタログ構築支援機能

本機能の提供により、データ提供者のデータカタログ作成負荷の軽減が可能となるため、データ提供者の増加が想定される。

また、分野間をまたがった標準的なカタログ項目によるデータカタログサイトが増加することによって、より目的のデータを見つけやすくなり、データ流通そのものの活性化も想定される。

2) 統制語彙作成支援機能

オープンソースソフトウェアである CVD を用いることで、分野の専門家がいつでも簡単に統制語彙を作成できるようになった。今回の開発成果は OSS での公開であるため、実際には分野の専門家を使用するにはソフトウェアをインストールする作業が必要となる。より簡易的に使うために SaaS での提供が考えられる。また、今回の実装では一人で作業することを想定していたが、複数の分野の専門家は遠隔地にいることも想定され、遠隔地にいる専門家同士が共同で編集を実施できる語彙リポジトリサービスとのシームレスな連携が考えられる。

今回我々が定義した SKOS のサブセットとなる CSV フォーマットを用いることで、既存の表形式データや文章、PDF で保持され開示してある既存の語彙を同 CSV にコピーし、その CSV を CVD へ読み込むことで、同義語関係や上下関係を GUI 上で可視化し簡易に把握したり定義したりすることが可能になる。また、それを RDF で保存するだけで、ほかの分野の人が理解しやすくなる。このようなデータの解釈可能性を高める作業のハードルを下げることで、今後、さまざまな分野で統制語彙が作成され、機械処理可能なデータが増えていき、データ社会の実現に寄与することが期待される。

データ連携基盤を活用した取り組み

分野間データ連携基盤技術(コネクタ)の運用支援技術開発

大学共同利用機関法人情報・システム研究機構

- データの利活用に適した表形式データの検索、構造変換、意味解析技術を研究開発し、国際的な評価コンペティションに参加。この評価を通じて世界トップレベルの性能を有することを確認した。
- オープンソースのツールを開発するとともに、これらの要素技術を連携して使用するためのプロトタイプシステムを開発し、開発技術の利用促進を図った。

1 研究の目的

本研究開発は、Society 5.0 の時代においてサイバー空間に取り込まれた、さまざまな分野のデータ基盤を連携して活用するデータ編集支援ツール群を研究開発することで分野間のデータ連携を容易にし、データ活用を促進することを目的としている。

Society 5.0 で扱われるデータ空間には多様なデータ提供者とデータ利用者がアクセスするため、やりとりされるデータは、多様な形式や構造、表現が混在することになる。これらのデータを組み合わせて活用するには、データの形式や構造、表現の変換が必要になる。また、これらのデータをコンピューターで効果的に分析する場合には、データに意味を付与することが望ましい。本研究開発では、データ形式の変換技術や類似表現の処理技術、データの意味付与に関する要素技術を研究開発するとともに、各要素技術を連携して活用するデータ編集支援ツールを開発することでデータ活用コストの低減を図った。

2 実施期間と方法

(1) 実施期間

2018 年度～2022 年度

(2) 実施方法

本研究では、代表的なデータ構造である表形式データを核とし、「データ構造の発見による構造変換支援」、「意味と構造を利用したデータ発見支援」、「語彙知識とデータの関係の学習によるデータの意味同定」の三つの要素技術の研究開発に取り組む。また、これらの研究開発の成果を連携させる「データ編集支援ツール TableLinker」のプロトタイプシステムを開発する。さらに、自治体を実証フィールドとした「実データを用いた研究開発技術の検証」を行い、その結果をフィードバックして、要素技術および編集支援ツールを改良する。図 1 に、表形式データを核としたデータ編集機能の構成を示す。

1) データ構造の発見による構造変換支援

明示的な構造を持たない文書などを含む、多様なデータソースを解析して表形式データに変換することにより、データソースとして想定できるデータの範囲を拡大する。オープンデータにはエクセルや CSV*1 といった

*1 各項目間がカンマ(,)で区切られたデータ。

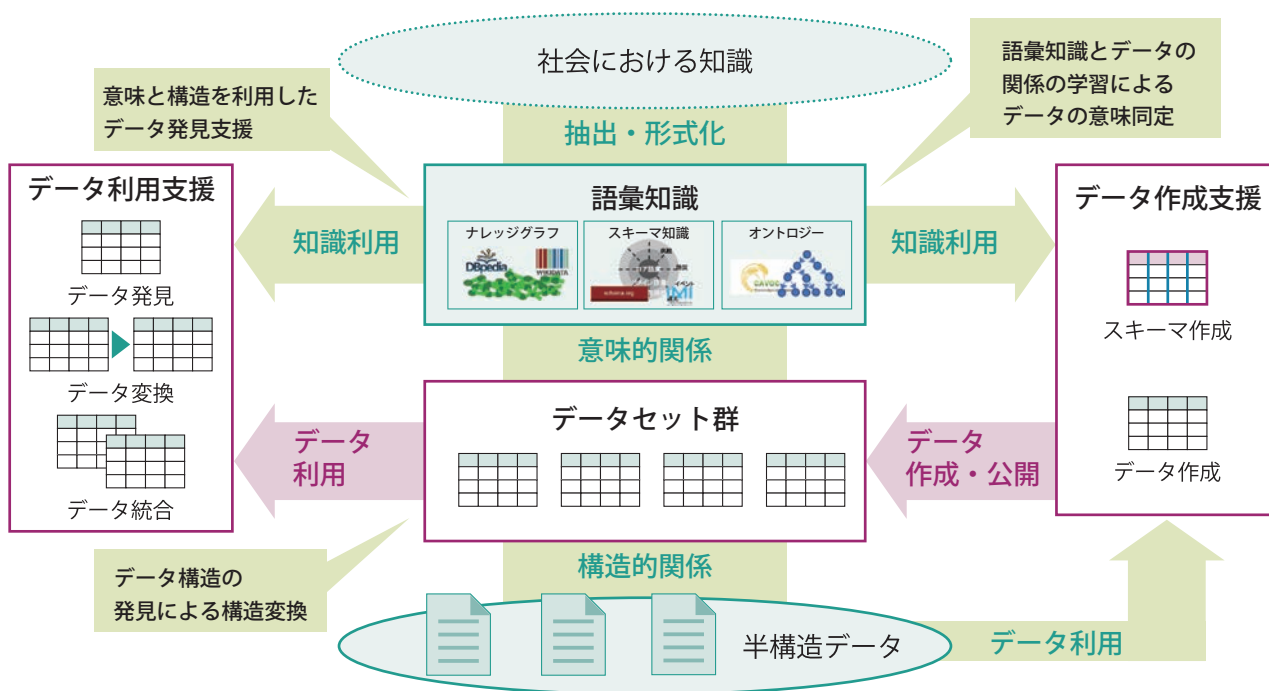


図1 データ編集支援技術群の概要

表形式データに加え、ワードなどの文書にデータが埋め込まれていることも多い。本研究では、オープンデータに含まれる割合の多いPDFから、表形式データを抽出する技術に取り組んだ。

まず、PDFファイルの各ページを分析し、タイトル、パラグラフ、図、表といった文書の構成要素に対応する矩形領域を抽出し、表に対応する領域を特定するページレイアウト分析、および表領域抽出法の研究開発を行った。次に、ページレイアウト分析によって得られた表領域から、表を構成するセルや見出しを抽出する表解析技術や、表の埋込み表現技術、および取り込み精度の高い表解析技術を研究した。

精度の高い表領域抽出および表解析技術を実現するには、大規模な訓練データが必要になる。そこでLaTeX*2などで記述された学術論文のソースファイルと、対応するPDFファイルとの組み合わせによって大規模な評価データセットを作成し、表埋込み表現の学習に用いるとともに、その精度評価を行った。

2) 意味と構造を利用したデータ発見支援

ユーザーの入力から検索要求を推測して、データカタログの語彙やメタデータ構造に対応付けすることで適切

な検索コマンドを生成し、データの利用や登録を支援することを目的とする。このために、自然言語解析技術を利用した、類似データ検索機能の高度化に関する研究開発に取り組む。2019年度までの研究開発において、自然言語処理分野の深層学習の適用によって、データ発見の機能が高度化できることを確認した。これを踏まえ、2020年度以降では、まず、ユーザーのクエリから表形式データを検索するコンペティション型の研究に参加し、「3) 語彙知識とデータの関係の学習によるデータの意味同定」と連携して検索機能の高度化モジュールを開発し有用性を検証した。また、類似構造化データの発見と自動マッピング機能を開発して、「1) データ構造の発見による構造変換支援」と連携して公的機関のオープンデータを用いた実証を行った。さらに、自然言語文によるユーザーの問い合わせを解析して、データベースクエリに自動変換する意味解析技術の研究に取り組んだ。特に、これまでの技術では扱うことが難しかった、表形式データ中のデータ値に関する条件を考慮した解析技術を開発し、国際的なチャレンジデータセットを用いて有効性を検証した。

*2 テキストベースの組版処理システム。

3) 語彙知識とデータの関係の学習によるデータの意味同定

データベースのデータや表計算ソフトウェアのデータなど、表形式データは分野横断型データの利用において重要なデータである。ここでは、多様な表形式データの意味を同定することで、分野横断型データの利用を促進する仕組みを確立した。

表形式データの意味同定の原理としては、ナレッジグラフ*3に格納された語彙知識と個別のデータの関係を学習させることで、新規データの意味が語彙知識を用いて説明できるようにした。語彙に含まれる用語ごとに、用語を用いて記述されているデータ（文字列、数値）を収集して機械学習を行うことで、新規データの各部分ごとの用語に対応しているものかを判定した。この結果、類似するデータの発見や重複する項目を用いたデータの統合、拡張が可能になる。

まず、表形式データの意味的アノテーションの基礎技術を確立し、コンペティション型の研究企画への参加を通して技術の評価した。この基礎技術に基づき、ユーザーおよびシステムから利用可能な表形式のアノテーションを行うプロトタイプシステムを開発した。

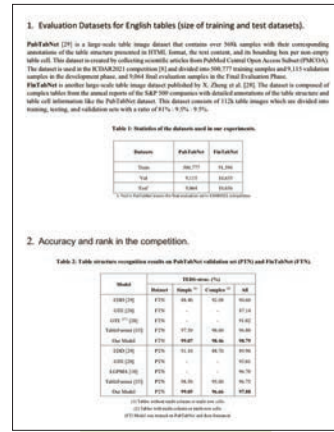
4) データ編集支援機能の実装

「データ構造変換」、「データ発見」、「意味同定」の各支援機能を組み合わせ、利用者が持つデータをコネクタに提供可能なデータに変換する過程、およびカタログ横断検索結果をまとめるデータ加工過程を、表形式データを介して支援する「TableLinker」を実装した。

まず、利用者が表形式データをアップロードして加工するために必要となるユーザー管理、データ検索、データのインポート/エクスポート機能を設計・実装し、外部からアクセス可能なクラウド上にプロトタイプシステムを設置した。次に、データ作成・加工を支援する共通テンプレート、および類似データ検索機能やデータ変換・加工機能の研究開発を行って実装した。

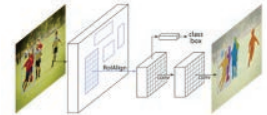
5) 実データを用いた研究開発技術の検証

2020年度は、2019年度にデータ連携の試行を実施した札幌市を中心とする地域での実証環境を用いて、異分野データの新たな連携の発見とデータ統合を対象に、1) から 2) での開発技術の検証を実施した。ここでは、

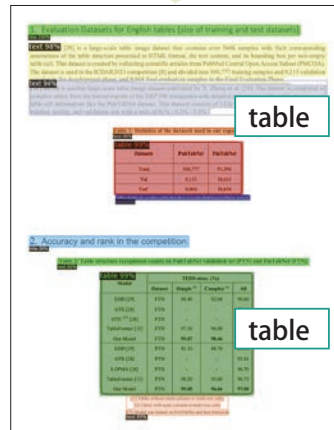


- ・ ページ分割
- ・ 領域分類

Mask R-CNN



K. He et al.



- ・ セル抽出
- ・ セル文字認識
- ・ 表構造抽出

Datasets	PubTabNet	FinTabNet
Train	500,777	91,596
Val	9,115	10,635
Test ¹	9,064	10,656

Table image with detected cell bounding boxes

Model	TEDS-struct. (%)			
	Dataset	Simple ⁽¹⁾	Complex ⁽²⁾	All
EDD [29]	FTN	88.40	92.08	90.60
GTE [28]	FTN	-	-	87.14
GTE ^(FTN) [28]	FTN	-	-	91.02
TableFormer [15]	FTN	97.50	96.00	96.80
Our Model	FTN	99.07	98.46	98.79
EDD [29]	PTN	91.10	88.70	89.90
GTE [28]	PTN	-	-	93.01
LGPM [18]	PTN	-	-	96.70
TableFormer [15]	PTN	98.50	95.00	96.75
Our Model	PTN	99.05	96.66	97.88

Predicted Table HTML (view on Web browser)

図2 表領域抽出と表解析

*3 さまざまな知識を体系的に連結し、グラフ構造で表した知識ベース。

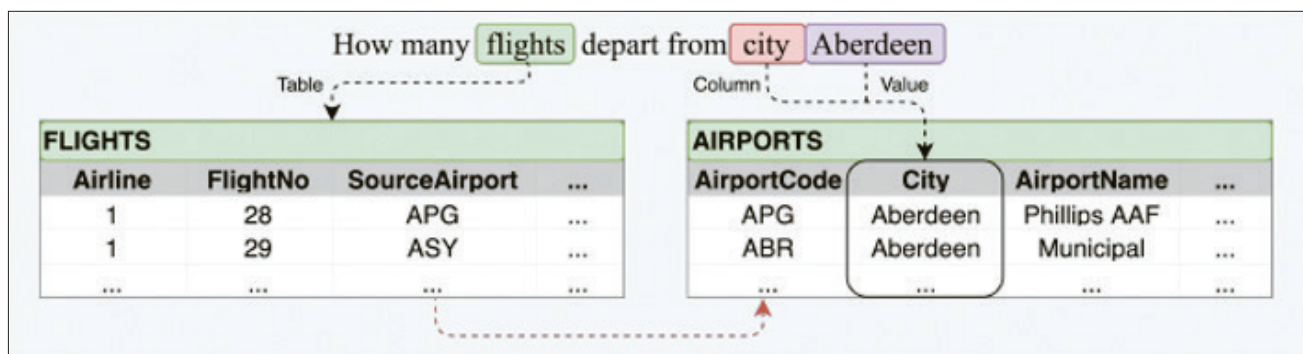


図3 ユーザークエリの言語解析

2019年度に用いた札幌市オープンデータを中心とする公的データと、流通分野（コンビニエンスストアやドラッグストアなどを想定）のデータ、および収集した道路・交通状況などによる交通流データと連携し、行政や事業者などに有用なデータ統合を各者の協力のもとで検証した。また、これらの分野間データ統合の他地域への展開、および地域間連携の検証のための実装を行う。

2021年度は、以上の活動を継続するとともに、地理空間上での情報統合において、各分野での業務や分野間での連携に重要な「路線」に着目し、分野間連携の基盤データとしての路線網データ標準を作成した。また、市民や来訪者に対して各種情報を提供するスマートシティアプリケーションにおいて、自治体が有する各種情報の統合に1)から3)での開発技術を適用した。さらに、以上の検証を反映させ、公開用に路線網データ標準を作成するとともに、スマートシティアプリケーション向けの複数分野データ統合ソフトウェアをパッケージ実装した。

3 研究の成果

(1) データ構造の発見による構造変換支援

PDF ファイル中に現れる表の抽出および抽出された表から、セルや行、列を抽出することを目的とした深層学習モデルを構築した。このモデルは、「①表構造全体の構造解析」、「②表に含まれるセルの抽出」、「③セル中の文字認識」を行うネットワークから構成される、マルチタスク型の深層学習モデルとなっている（図2）。

この分野の代表的な評価データセットであるICDAR2021 評価用データセットを用いて分析性能を評価し、ICDAR2021 コンペティションにて最も性能の良

かった手法と同程度の96.2%の精度を達成した。また、日本語表形式データの解析に適したモデルを構築するために、Wikipedia から表形式データを抽出し、評価用データセットを構築した。このデータを用いて日本語用のモデルを学習し、92.6%の精度を達成した。

(2) 意味と構造を利用したデータ発見支援

単純なキーワード検索では見つけられないデータの発見を容易にするため、埋込み表現を用いた意味検索手法を開発した。国際的なコンペティションで、7 評価指標のうち6つで参加チーム中トップレベルの性能を達成した。表形式データ連携のため、埋込み表現を用いた項目マッピング機能を開発・実装して自治体オープンデータに適用し、医療機関のオープンデータ116件に対してマッピング結果を検証したところ、96.4%の項目が人手によるマッピング結果と一致することを確認した。

また、自然言語によるユーザー問い合わせの意味解析の研究開発では、データ値を考慮した解析モジュール（図3）を実装して、この分野の代表的な評価データセットであるSPIDER データセットを用いて、従来モデルの性能が提案追加モジュールにより2%向上することを示した。

(3) 語彙知識とデータの関係の学習によるデータの意味同定

異なる情報源から得られたデータを連携して機械可読性を高める、表形式データのアノテーション技術は、人手による登録・検索作業コストの低減に重要な技術である。この技術に関しては、数値データ項目理解に関する独自のアルゴリズムと、表理解の独自の再帰的統合処理を研究開発してきた（図4）。Semantic Web Challenge on Tabular Data to Knowledge Graph

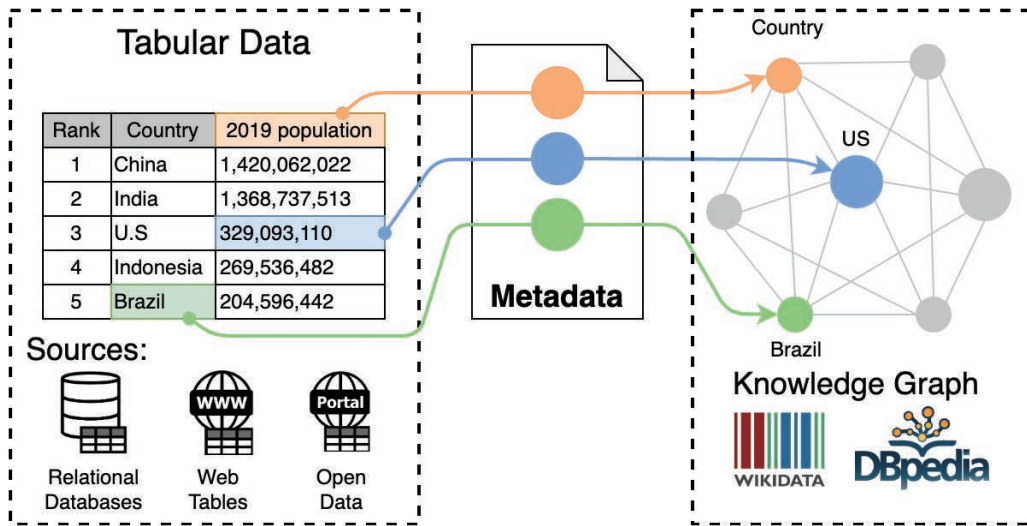


図4 表形式データの意味的アノテーションの役割

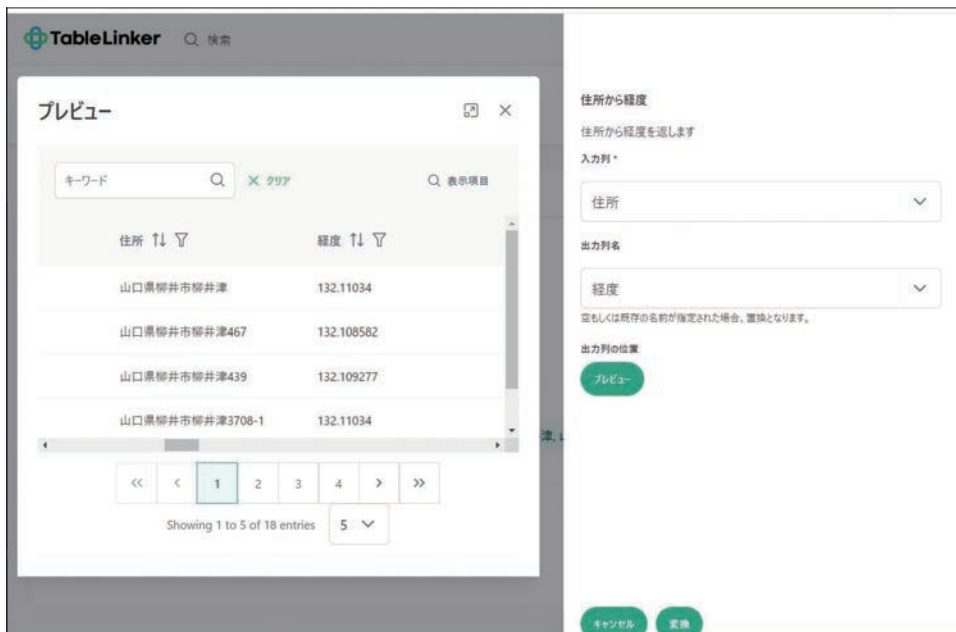


図5 開発した TableLinker

Matching (ISWC2019、2020、2021) のチャレンジに参加して、2019年と2020年には総合1位、2021年は usability 部門1位、accuracy 部門2位と評価され、世界トップレベルの性能を継続して達成した。

このアルゴリズムを用いた表形式データのアノテーションシステムを開発して試験公開を行い、ユーザー向けインターフェースとシステム連携のためのAPIを提供した。

(4) データ編集支援機能の実装

データ構造変換、データ発見、意味同定の各支援機能を組み合わせ、利用者が持つデータをコネクタに提供可能なデータに変換する過程、およびカタログ横断検索結果をまとめるデータ加工過程を、表形式データを介して支援する TableLinker を設計した。表形式データのデータ形式・文字コードのエラー検出機能の強化や、表形式データの意味型の登録機能、基本フィルタを用いたデータ変換などを円滑に利用するためのUIの設計および実装を行うとともに、(1)、(2)で開発した類似構造データの発見と自動マッピング機能および(3)で開発した表形式データの同定技術との連携モジュールを開発実装した。

(5) 実データを用いた研究開発技術の検証

コネクタ機能ツールおよび運用支援技術の実データ、実問題における技術検証を想定し、流通分野、モバイル分野、および公共分野での各データの統合に基づくマーケティング分析事例を踏まえた、分析システムのプロトタイプングを行った。

まず、流通分野におけるコンビニエンスストアのID-POSデータ、モバイル分野のスマートフォンアプリによって広く収集されるポイント型の流動データ(緯度経度で位置が取得されたデータ)を導入し、その地域の自治体オープンデータと連携させた。次に、これらの事例を踏まえ、データ統合によるサービス効果を評価するための地理情報を考慮した、時系列データの分析プロトタイプシステムを作成した。

オープンイノベーションの創出と社会実装の推進を目指した、新規サービス構想に資するユースケース整備は、2020年度からこの「実データを用いた研究開発技術の検証」と一体で進めてきたが、2021年度は、複数の自治体にてそれぞれ作成されるコンテンツの統合タスクにおいて、TableLinkerを用いた統合支援技術とシステム

の導入検討を進め、システム開発に着手した(図5)。

また、さいたま市のスマートシティ基盤を活用したアプリケーションにて、関連自治体などのコンテンツのCADDE(Connector Architecture for Decentralized Data Exchange)を適用した統合利用のためのシステム構成を検討した。

4 まとめと今後の展望

表形式データの構造変換支援技術として、PDF文書から表形式データを抽出する技術を研究開発し、国際会議のコンペティションで使用された評価用データセットを用いた実験でトップレベルの表解析性能を有することを確認した。オープンソースのライブラリとして公開し、普及を図るとともに、利用者からのフィードバックを得て技術の改良を継続する。

意味と構造を利用したデータ発見支援においては、当初計画通りに機械学習を用いた類語推定技術に基づく表形式データの検索を実現し、国際的なコンペティションでトップクラスの性能であることを示すとともに、表形式データ構造へのマッピングにも適用し、代表的なベンチマークセットを用いて有効性を示した。ただし、自然言語文による複雑な問い合わせの解析を実用レベルにするには、さらなる研究開発が必要である。

語彙知識とデータの関係の学習によるデータの意味同定においては、当初計画通りに大規模知識ベース(Wikipedia、DBpedia、Wikidata)を用いた表形式データへのアノテーションを実現し、世界トップクラスの性能であることを示した。

また、UI付きのシステムも開発し、任意のデータが動作することを確認した。ただし、ベースとなる知識ベースは、日本国内に関する情報が必ずしも潤沢とはいえないため、こういった情報を知識ベースに追加で入れていくことが必要となる。