



SDSCの要素技術を集約し価値化を加速するAI価値基盤

私たちNTTスマートデータサイエンスセンタ（SDSC）が取り組んできたさまざまな産業ドメインの成果を集約し、そのコアパーツを利用して価値化を加速するためのAI（人工知能）価値基盤を開発しています。分析事例をカタログ化することで問題設計を支援するほか、集約した実データの加工技術を自動的に試行し、データ分析の初期の試行錯誤を短期化することで、価値化を加速します。本稿では、AI価値基盤の実現に向けた取り組みを紹介します。

キーワード：#AI価値基盤、#データ分析、#数値最適化

はた 秦 藤島 福田
 たかひろ 崇洋 美保 健一
 こだま /児玉 /安達
 みどり 翠 悠 夕貴

NTTコンピュータ&データサイエンス研究所/
 NTTスマートデータサイエンスセンタ

はじめに

私たちNTTスマートデータサイエンスセンタ（SDSC）では、街づくりDTCの実現のため特に生活・消費・移動・物流・資源などの分野において、お客さまの課題を引き出し、実データを用い、課題解決してきました（図1）。

こうした実績を顧みると、実世界（Physical Space）における価値の源泉は現実に課される制約条件の下で取得できる実データに基づき、Cyber Space上のデジタルツイン（DT）で最適な状態を計算し、その結果を実世界の制御やレコメンドとして返却するCPS（Cyber Physical System）（図2）の仕組みで説明できることが分かってきました。ここで、もっとも重要な役割

を果たす最適化は複雑で巨大なモデルも研究されていますが、AI（人工知能）・機械学習の技術が発展した現在では、豊富で整ったデータを十分な計算時間で学習すれば解けるようになってきました。一方で、実世界の価値化では計測条件・運用条件・コストなどの現場ごとのさまざまな制約条件を満たす必要があり、計測したままの生のデータ（raw data）はそのままでは利用できないことがよくあります。

こうした状況の下、PhysicalとCyberのギャップを埋めて街づくりの領域でデータに基づいて実世界へ価値を還元するデータ価値化を加速することを目的に、SDSCでは自ら解決してきた街づくりDTCの各種ドメインにおけるソリューションから、コアとなる技術や知見を結集したAI価値

基盤を開発しています。

AI価値基盤は、実産業ドメインにおけるAIによる新たな価値を定義、具現化し、1つひとつの価値を実産業ドメインの業務やお客さま価値として実感いただけるサービスとして具現化しています。その中には、実データ加工や予測・最適化機能といったデータ分析機能と、実際にデータ連携に必要な機能、価値を業務における活用やお客さま個人に体験いただけるインタフェースも含まれます。AI価値基盤では、コアとなるAI機能を含むパーツを複数の産業ドメインで提供されるサービスの中でも有機的に利用できるようにしています。実現価値の特性に合わせて、それらのパーツを組み合わせて使えるようにすることで、新規のサービスに対しても、サービスプロ



図1 これまで取り組んできたドメイン例

ダクトに必要な機能の選択を容易にしたり、初期に必要なデータ加工を自動化したり、試行錯誤を最小化することが可能となり、価値化を加速することができます。

本稿では、このAI価値基盤機能の中で価値化を加速するさまざまな工夫のうち、(a)これまでの実績から、実社会で運用できるように工夫した問題設計を集約しリファレンスとして提供する、(b)これまでの実績から、最適化の前段階であるデータのクレンジングや未来予測などの前処理を共通分析機能として切り出し、他のデータへ適用できるようにする、の2点について主に紹介します。

AI価値基盤が提供する効果

ここでは、前述の(a)、(b)が価値化のプロセスにどのような点で貢献をするか説明します。

SDSCの価値化プロセスは、図3に示す典型的なデータ分析の手順に沿って進むものがほとんどです。①の課題設計フェーズでお客様の抱える問題と、その問題を取り巻く周辺知識を分解し、最適化問題へと帰着させます。②のデータ分析フェーズでは、実環境でデータを計測し最適化手法をDTで実現します。これらの手順で効果が見込めると判断すれば、③の実証フェーズにおいて実際の機器を操作し運用するフィールド実証へと進みます。(a)と(b)の工夫は、特に課題設計フェーズとデータ分析フェーズの加速をねらっています。これは価値化を加速するだけでなく、お客様の協力が必要不可欠な実データを用いるSDSCの取り組みにおいて、机上の結果を示す①と②を加速しいち早く結果を示すことにも貢献します。なお、AI価値基盤はそれ自身をDTとして利用することも想定

して開発していますが、本稿で紹介する(a)、(b)とは異なる領域の議論となりますので、ここでは割愛します。

(a) 最適化の問題設計の事例をリファレンスとして提供する

実世界での課題解決は、現実のさまざまな制約条件を満たして特定の指標を最小化(または最大化)する数理最適化問題に帰着できる場合が多くあります。SDSCが解決してきた課題を例に挙げると、

- ・空調のエネルギー消費を削減するために、来館者の快適性を保ちつつ1日の空調運転の総消費エネルギーを最小化したい
- ・より多くの顧客の元に時間どおりに商品を届けたいので、複数台のロボットの商品配送をもっとも多く、より早く配達できるルート計画を立案したい
- ・緊急通報に迅速に対応するために、救急車の待機する消防署を変更して現場到着所要時間を最小化したい

など、いずれも制約条件付きの数理最適化問題でした。ところが、このように現実の困りごとを問題に帰着させることは実は難しく、価値化でもっとも重要かつ困難なプロセスといえるでしょう。

この問題設計は、お客様の抱える問題や取り巻く環境などを正しく理解して分解し、解決に許された時間内に解くことができる手法と、うまく対応付けて再設計する能力が必要です。この問題設計が実用可能であるためには、

- ・誤差や遅延、欠損などを許容したデータ計測

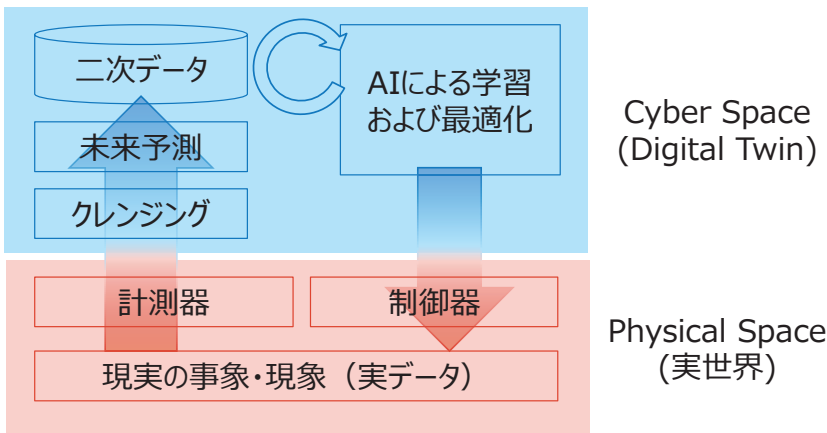


図2 Cyber Physical Systemの概念図

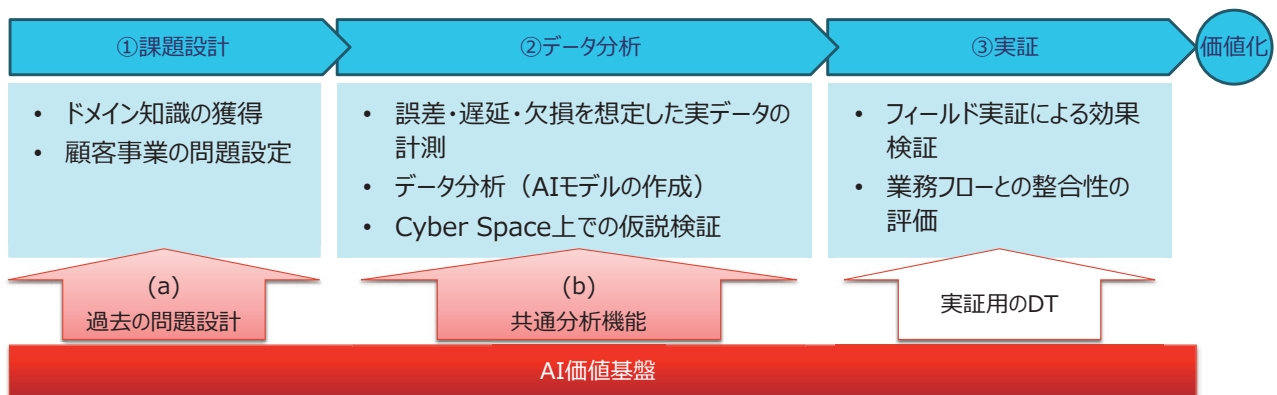


図3 価値化の流れとAI価値基盤の貢献

- ・計測データから最適解を導出する分析ロジック
- ・分析結果の実行までの間に許容できる時間内の制御
- ・その実行手順を業務フローに落とし込む運用

などの点をすべて解決する必要があります。もちろん、これらを最初から完全に見通して1度の試行ですべて解決できることは稀で、実際には何度も問題設計を行い実際に運用してみて、さまざまなフィードバックを受けて必要な箇所を見直す、というサイクルを繰り返す必要があります。

データ分析の経験を多く積んだデータサイエンティストは、計測・分析・制御・運用の各工程で起こり得る問題をさまざまなシチュエーションで経験し、成功あるいは失敗の事例を数多く知っているため、少ないサイクルで現実的な解決手段に到達できます。またこうした経験豊富なデータサイエンティストになるためには、自身が多く経験を積むことのほかに、経験者のアドバイスを受けることや過去の検討資料を洗うことなど、先人の試行錯誤を追体験する何らかの手段が有効です。そしてこれが、最適化の問題設計事例をリファレンスとして蓄積することのねらいです。

AI価値基盤は、これまでデータサイエンティストが実際に解いてきた問題設計を蓄積し、新たな問題に直面した際にリファレンスとして参照できるようにすることをめざしています。これは、新たな問題に直面したデータサイエンティストが過去の実例を追体験する手段を提供します。たとえ異なるドメインであっても、問題の分解の仕方や数理最適化問題として帰着させる工程を自身の直面した問題の価値化プロセスの中で参照することができれば、実用的な問題設計を迅速にできるようになると期待しています。

(b) 最適化の前処理を共通分析機能化し他のデータに適用できるようにする
実世界の温度や明るさ、その場所の空調の設定温度、その場所にいる人の数などは、計測器を通じて数値化されデータとなります。このデータを基に実世界の現象を理解あるいは予測し、それに対して最適な機器

の制御や行動のレコメンドを行うことで、現実で起こるさまざまな現象を最適化します。しかしセンサが故障したり、センサが移動されたり、センサの前に荷物を置かれたり、イベントにより普段は来ない数の人が殺到したり、といった非常にさまざまな原因によって、データには本来の現象との誤差や、値が取得できないという欠損、二度と起こらない異常値などが入り込みます。こうした誤差・欠損・異常値は、実世界の現象の理解を阻害し予測精度を低下させるため、最適化の前にこれらを取り除くデータクレンジングにより、最適化に適したデータへ加工します。

また、データとして計測できるのは現在より前、すなわち過去の現象です。そのため、過去のデータから機器制御や行動レコメンドをしても、機器や人が動くときには最適ではないこともあります。例えば、ビル空調では空調が管理する空間が広いために、設定温度や風量の変化が室温の変化として空間に反映されるには、数10分~数時間の時間を要し、その間に混雑時間帯が終わり過剰な空調となってしまうといったことがよくあります。これを防ぐために、最適化の結果が反映されるときに状況を予測することも、前処理の一種といえるでしょう。

さらに、SDSCでは最適化手法として大量のデータを必要とするAIを利用することがあります。しかし実世界では並列実験や加速実験ができず、大量データの収集が難しい領域でもあります。SDSCの実例によれば、こうした場合は次のような機械学習に特有のデータ加工技術を利用していることが分かりました。

- ・少ないデータ量を補うための、さまざまなデータ拡張手法
 - ・正例・負例の極端な頻度の差を補うリバランス
 - ・現象が少数多種でデータ不足に陥る場合の、事象の類型化(クラスタリング)
- これらの処理技術も、事前に計測したデータを基にして二次データを生成するという意味で最適化の前処理といえるでしょう。

前処理で行うべき内容は、(a)で紹介した問題設計にも依存しますが、計測器自体や

計測環境、そして計測したデータ(現象)の特性にも依存します。これはすなわち、問題設計が異なる場合でも効果が期待できることを意味しており、AI価値基盤では前処理部分を共通分析機能として切り出すことにした理由です。共通分析機能を同じ形式で実行できるようにパーツ化しておくことで、新たなデータを分析する際にさまざまな手法を一度に試行して結果を比較することができ、前処理に適した手法を素早く選択することができます。

このようにして、(a)、(b)により、SDSCの価値化プロセスを支援し加速します。

従来技術との違い

さまざまな実世界の現象をデジタル化できるようになった現在、データ分析は多くの人や企業において取り組まれる課題です。そのため、たくさんの分析技術、ライブラリやそれらを集約したプラットフォームが提供されています。本稿では、一般的なデータ分析プラットフォームや各種ライブラリ等とAI価値基盤の違いについて説明します。

ここでは代表的な分析環境の例として、R⁽¹⁾、Python⁽²⁾を例に、その特徴を解説します。

Rは、もともとAT&Tで開発されたSという言語環境を参考に開発された統計分析に優れた言語およびソフトウェア環境です。対話的に処理できるインタプリタ環境でありながら、大量のデータを高速に処理することができるほか、多くの統計的手法が実装されているため、導入してすぐにデータ分析を始めることができます。また、外部のパッケージを導入することで標準では実装されていない高度な処理もできます。

Pythonは手続型の汎用プログラミング言語です。Pythonのプログラミング言語としての側面は本筋ではないので割愛しますが、Rと同様に標準ライブラリへ多くの統計処理やベクトル・行列計算が実装されており、インタプリタでありながら高速な数値計算ライブラリが評価されているように思います。汎用プログラミング言語であるため複雑な処理の実装にも適しており、またGPU(Graphics Processing Unit)

への対応も充実していることから、ニューラルネットワークを利用した深層学習など高度な機械学習の実装にも非常によく利用されており、外部のライブラリを導入することで最先端のアルゴリズムを利用することができます。そのため、データ分析の実装を考える際には必ず選択肢に挙げられます。

代表例として2つの環境を紹介しましたが、このほかにも多数のプログラミング言語や、それらに実装されたアルゴリズムを集約した分析プラットフォームがあります。さらにデータをさまざまな側面から可視化するBI (Business Intelligence) ツールや、データを蓄積しつつ集計機能も備えたりレシーショナルデータベースなども、広い意味では分析環境と呼べるでしょう。

これらの環境についてここで注目したい共通点は、データ形式やドメインを限定しない汎用的な分析機能を備えているところです。これは分析ソリューションの提供という目的では当然のことです。しかし、こうした汎用分析環境を利用して実世界の課題を解決したい場合、問題設計および前処理の知見を有するデータサイエンティストが時間・工数を費やして多数の分析手法から適切な手法を選択する必要があります。CyberとPhysicalのギャップを解決し実用可能なかたちで最適化するためには、問題設計と前処理は価値化のプロセスに必要であり、試行錯誤による時間と労力の大きくかかる部分であることは変わりません。そこでAI価値基盤では、現実の課題を解いた実績のある前処理をパーツ化しておくことで、この試行錯誤を加速します。このように、汎用的な分析環境では支援の難しいこれらのプロセスに対して支援することを目的に開発されているという点が、従来技術と大きく異なる点です。

AI価値基盤の設計方針

ここまで紹介してきた考え方を踏まえたAI価値基盤の設計方針(図4)を紹介します。

(a) 過去の問題設計事例のリファレンス提供

ここまでで述べてきたように、現実の問題を数理最適化問題に帰着させるためには、

多くの制約を解消し問題設計してきたという経験が必要です。そしてAI価値基盤は分析者への支援として、過去の問題設計を追体験させることで、経験に代えて問題設計を支援します。つまり、この機能で提供したい知見は、数理最適化問題を歪めないかたちで実世界の制約条件をどのように解消するか、というものです。具体的にはAI価値基盤では、過去の実例を数理最適化問題として分析し、

- ・ 目標→目標に対して設定したKPI→最適化問題としての定式化
- ・ 計測データ→実施した前処理
- ・ 最適化問題→最適化アルゴリズム

という実世界の制約と採用した手法を組として一覧し、参照や検索を可能とすることをめざしています。

このとき、特に重要なのは失敗事例よりも成功事例だと考えています。失敗から教訓を得るということもありますが、価値化の加速を目的としたAI価値基盤の場合は、うまくいった事例を真似て同じように試行するのが一番の近道です。SDSCの取り組みには一見すると全く異なるさまざまなドメインの問題を解いているようにみえますが、本質的には実世界は人の社会活動により成り立っているという共通点があり、この共通点が計測データの特性や解決手法の

満たすべき運用条件を形成するために、異なるドメインであっても過去の事例を活かすことができると考えています。

(b) 前処理の共通分析機能化

先述の共通点は、実空間における制約条件を解消する前処理でも同様のことがいえます。

SDSCで取り組むデータ価値化は、いずれも実空間の現象を時系列数値データとして取り扱い、時刻・日・曜日・週など人間社会の周期でとらえ、周期性や類似性を利用し、平滑化や欠損補間、データ拡張などを実現しています。また、温度や明るさ、消費電力、人数など物理的現象を計測する手段は、ドメインによらず物理現象に依存した同様の手法を利用します。こうした共通点が利用できる前処理は、SDSCが解いてきた実例を分解することで抽出することができます。AI価値基盤は、このように実効性のある前処理を同じ形式で動作するようパーツ化することで、新たなデータに対して一斉適用が可能となり前処理の試行錯誤を加速します。

事業化への取り組み

前述のとおり、SDSCではお客さまの具体的課題に対してデータの前処理・未来予

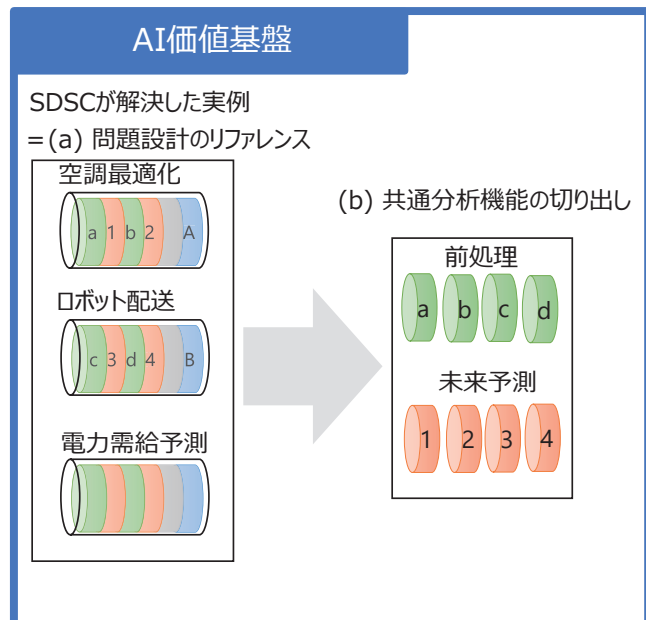


図4 AI価値基盤の構築イメージ

測・最適化と、可視化・最適制御を行うことにより新しい価値を提供してきました。AI価値基盤ではサンプルデータあるいは自身のデータを実際に分析し、有用性を確認できる図5のような環境を提供します。サービスごとにDTを設け、お客さま（テナント）それぞれに独立した実行環境を用意します。データ価値化の結果やサービスが満足できる結果になれば、AI価値基盤上の環境をそのまま利用することや、お客さ

まのシステム・設備と連携して動かすこと、あるいは接続先のシステム・設備に応じてカスタマイズをすることが可能です。

また、前述した共通分析機能を用いることができ、さまざまなデータ分析手法のアルゴリズムを試行、比較することにより、最適な手法を選択することが可能です。

具体的なサービスについて紹介します。

例えば、店舗におけるフードロス削減を目的とした取り組みにおいては、お客さま

の店舗のデータを用いて来店者数予測やメニューごとの販売数予測などの未来予測を行い、可視化することで過剰な仕入れや仕込みを抑えることにより、フードロス削減します（図6）。

この取り組みをお客さまの施設で実施する場合は、来店者数を計測するカメラやセンサ、POSデータ（メニューごと販売数）などを用いることで、上記のような結果を素早く得ることができます。サービスに必

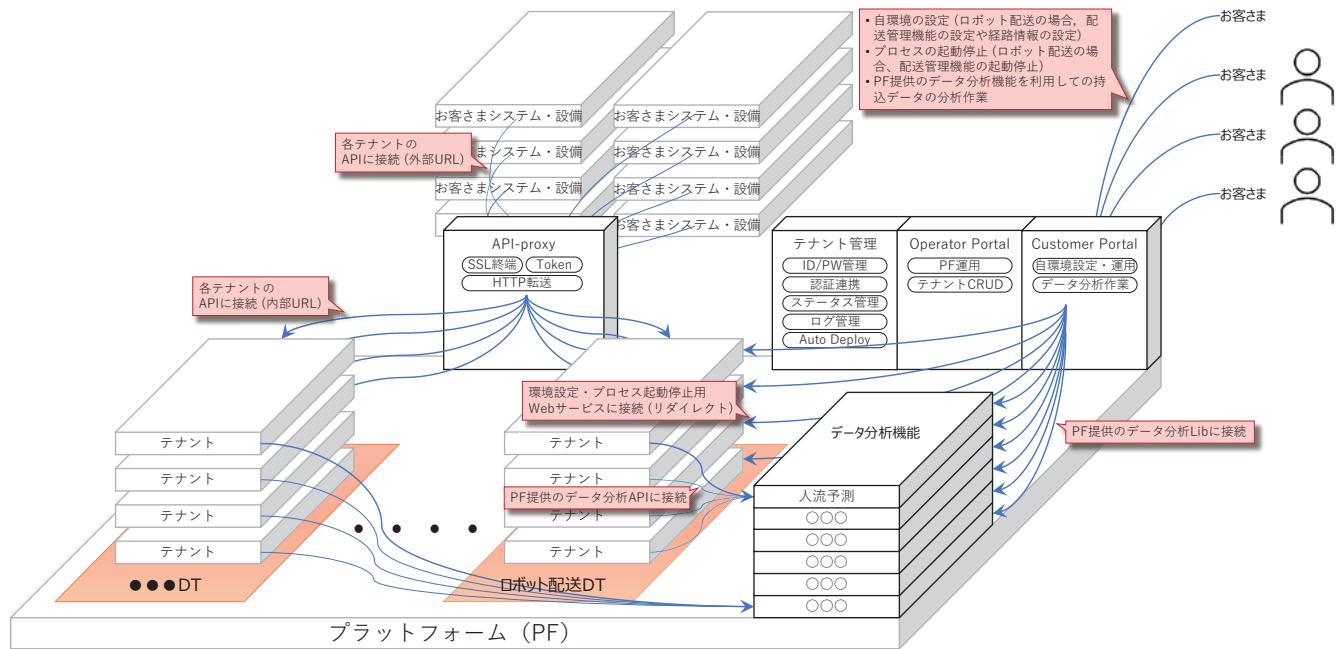


図5 AI価値基盤サービスのグランドデザイン



図6 フードロスゼロの取り組み：店舗の来店者数予測，メニューごとの販売数予測

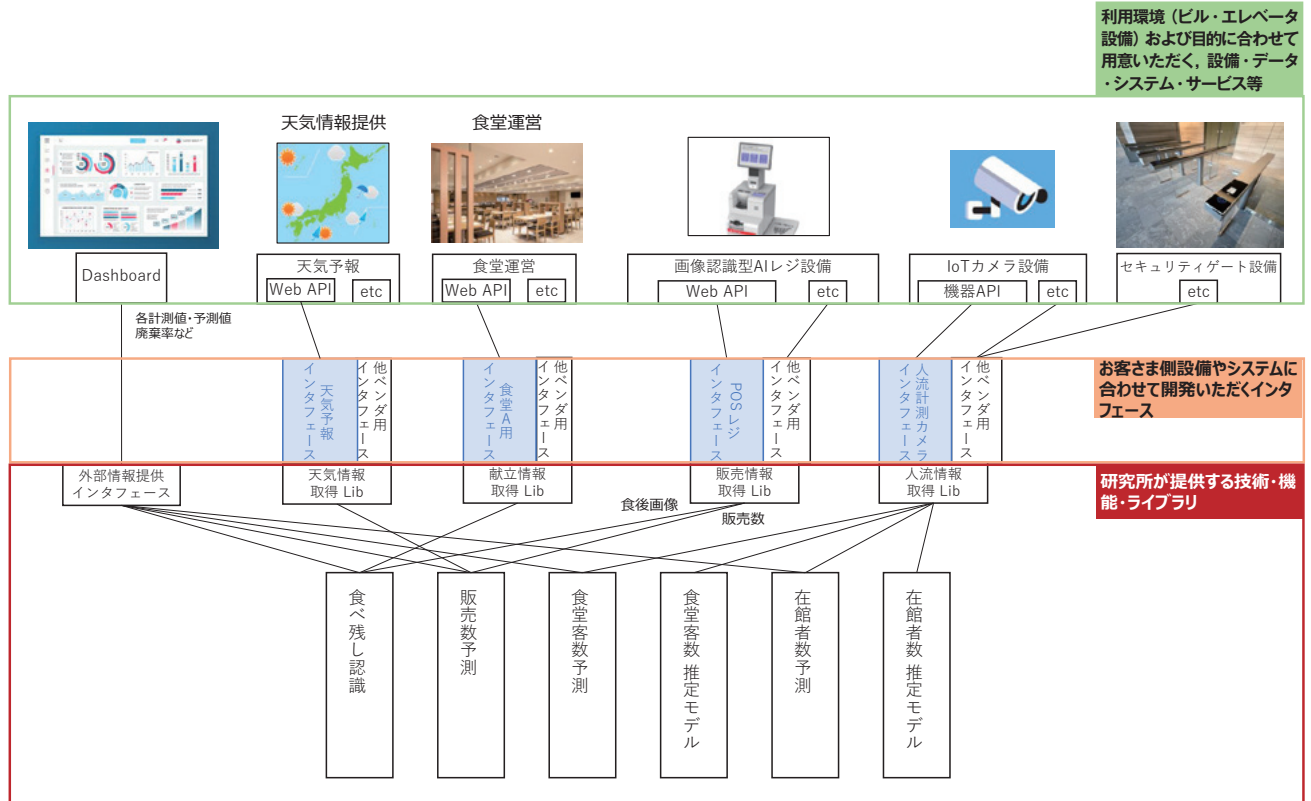


図7 フードロスゼロの施策を実施するシステムイメージ

要なデータや分析機能についてまとめたものが図7です。

SDSCではこれまで、図1のように省エネと快適性を両立する空調制御、オフィス内のパーソナル空調、飲食店のフードロスゼロ、顧客行動解析を用いた店舗運営の効率化、効率的な配送を実現するロボット配送、救急隊の配置最適化、高精度な発電量予測や需給マッチング、交通渋滞の検知と予測、農産物流通の全国規模での最適化、行動先読みによるスマートホームや街中におけるおもてなしサービスなど、データから価値を創出するデータ価値化に取り組んできました。価値化を支援するAI価値基盤により、人々の快適な暮らしを飛躍的に向上させた、未来社会の実現に貢献します。

AI価値基盤の課題と展望

本稿では、解決してきた事例の知見をAI価値基盤へと集約し、新たなデータ価値化を支援しようとしていることを紹介しました。しかしAI価値基盤へ知見を集約する

にはいくつか課題があります。

1つは、実際の課題で用いる機能は、その事例の中でもっとも効率的なかたちで実装されており、共通機能として切り出すには1つひとつの機能を他ドメインに適用した場合を想定し、同じように実行できるように再設計する必要があります。もう1つは、共通分析機能が増えてくれば、適用する分析機能を選別する必要があり、そのためには、データの特性や意味などを理解できる一種のセマンティクスのようなメタ情報が必要です。またこのメタ情報は、新規ドメインが加わるごとに見直す必要があるかもしれません。2023年度はまず最小限の実装を完了し、実際に使用した方のフィードバックを基に改良し、これらの課題にも対応していきたいと考えています。

■参考文献

- (1) <https://www.r-project.org>
- (2) <https://www.python.org>



(上段左から) 秦 崇洋 / 横畑 夕貴 / 児玉 翠
(下段左から) 藤島 美保 / 安達 悠 / 福田 健一

AI価値基盤では実際のお客さま課題から得た知見をかたちにすることにこだわって取り組んでいます。本取り組みにより、少しでも多くのお客さまへ具体的な価値を提供することをめざしていきます。

◆問い合わせ先

NTTコンピュータ&データサイエンス研究所/
NTTスマートデータサイエンスセンタ
E-mail sdsc@ntt.com