

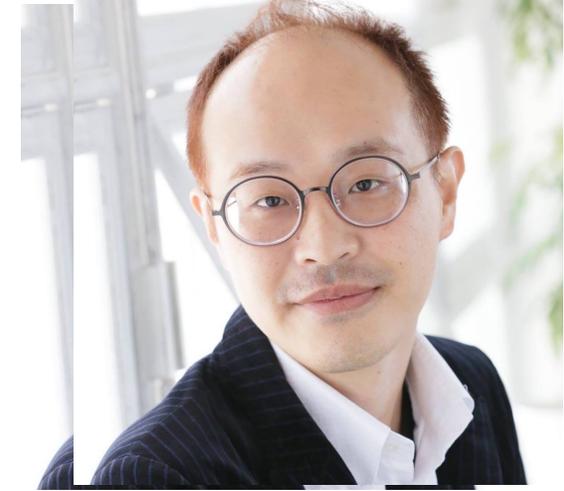
＼ データドリブンを目指す方へ、データサイエンスの第一人者が語る！ ／
データ・AI活用によるデータドリブンビジネス創出のための実践
2022.05.15

データ・AI活用による データドリブンビジネス創出 のための実践

武蔵野大学 データサイエンス学部 データサイエンス学科長 准教授
アジアAI研究所研究員
国際大学GLOCOM 主任研究員
デジタルハリウッド大学大学院客員教授

中西 崇文

プロフィール



- 中西崇文(Takafumi Nakanishi)
- 武蔵野大学 データサイエンス学部 データサイエンス学科長 准教授
- 国際大学グローバル・コミュニケーション・センター(GLOCOM) 主任研究員
- デジタルハリウッド大学大学院 客員教授
- 博士(工学)。

- 1978年12月28日生まれ、三重県伊勢市出身。
- 2006年3月、筑波大学大学院システム情報工学研究科にて博士(工学)の学位取得。
- 2006年4月、情報通信研究機構(NICT) 研究員。
- 2014年4月、国際大学グローバル・コミュニケーション・センター(GLOCOM) 准教授・主任研究員
- 2018年4月、武蔵野大学 工学部 数理工学科 准教授
- 2019年4月、現職

専門

- データマイニング、データ分析システム、統合データベース、感性情報処理、メディアコンテンツ分析。
近年は、ビッグデータ分析手法を通じたデータ分析工学分野の創出に興味を持つ。

- 総務省「AIネットワーク社会推進会議」構成員
経済産業省「流通・物流分野における情報の利活用に関する研究会」委員
総務省「ICTインテリジェント化影響評価検討会議」構成員
等歴任。

- Line ID: piano_lullaby
- Twitter: @piano_lullaby
- Facebook: <http://facebook.com/pianolullaby>

著書



第6回 データサイエンス学 データサイエンス学部 データサイエンス学科 中西崇文 准教授

人間は何を感じ、どう表現するかをデータを通じて解明



データサイエンス学部
准教授
中西 崇文
Takafumi Nakanishi

Profile

筑波大学大学院システム情報工学研究科にて博士（工学）の学位取得。2006年より情報通信研究機構 研究員、2014年より国際大学グローバル・コミュニケーション・センター(GLOCOM)准教授・主任研究員、2018年より武蔵野大学工学部数理工学科准教授を経て、2019年現職。現在、国際大学グローバル・コミュニケーション・センター(GLOCOM)主任研究員、デジタルハリウッド大学大学院客員教授でもある。音楽ユニット「タイアップ」のメンバーとしても活動中。

私たちは、買い物をする時には「いいな」と感じたものを選び、悲しい映画を見ると深く心を動かされます。人間の行動にさまざまな影響を与える“感性”。人間の感性に沿ってコンピュータが動くようになれば、社会にはこれまでにない新たな価値が生まれる可能性があります。人間の感性とデータサイエンスを結び、AIを駆使して言葉や音楽のデータから感性を抽出、表現する手法を生み出そうとする、データサイエンス学部の中西崇文准教授の研究を紹介します。

「感性」をキーワードに取り組む3つの研究



「データサイエンス」とは何か

ーデータで分かった「稼ぎの良いタクシー」の特徴ー

突然ですが、みなさんに質問です。街中を走りながら乗客を探す「流し」のタクシーの運転手さんは、どんな乗客が多いと、売上がアップすると思いま



大学案内

入試情報

教育

研究

国際交流・留学

学生生活・就職



NEWS & EVENTS

Musashino TIMES

HOME > 教育 > 学部 > データサイエンス学部 > データサイエンス学部

FACULTY OF DATA SCIENCE

データサイエンス学部



学部紹介 ▼ データサイエンス学科 ▼ 高校生データサイエンスONLINEセミナー ▼ MUDSオープンキャンパス ▼ インターンシップ受入れのお願い ▼

https://www.musashino-u.ac.jp/academics/faculty/data_science/

データサイエンス専攻 [修士課程]

データサイエンス専攻 [修士課程] ▼



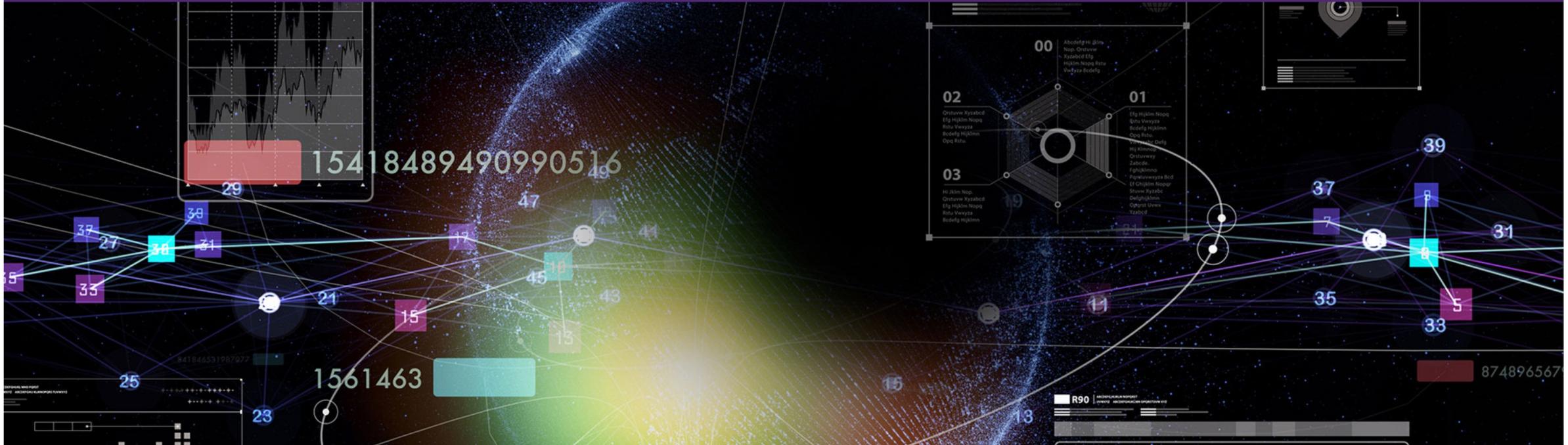
- ▼ 概要
- ▼ TOPICS
- ▼ カリキュラム
- ▼ 授業紹介
- ▼ 研究プロジェクト紹介
- ▼ 教員紹介

https://www.musashino-u.ac.jp/academics/graduate_school/course/data_science/major/

2021年4月 開設

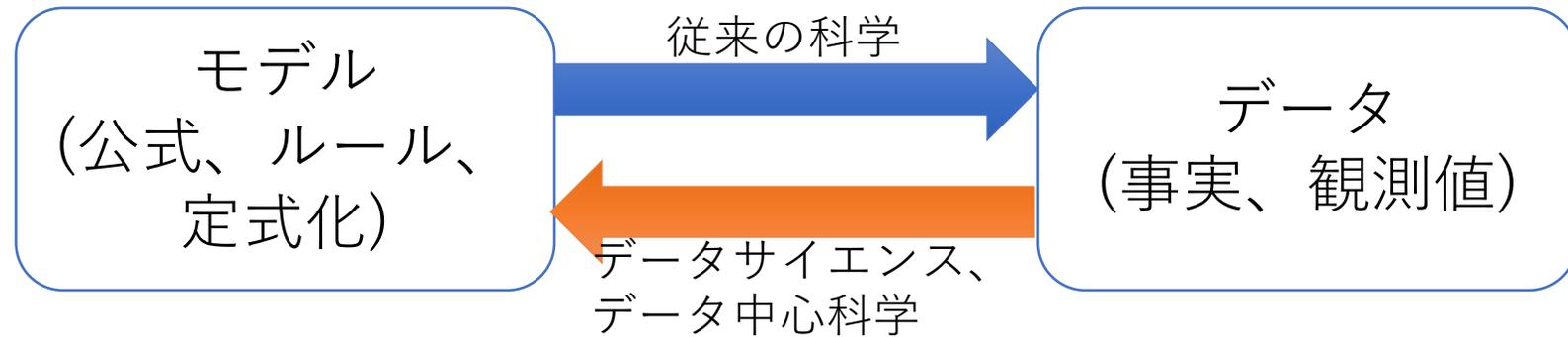
データサイエンス専攻 [博士後期課程]

データサイエンス専攻 [修士課程] ▼ データサイエンス専攻 [博士後期課程] ▼



- 概要
- カリキュラム
- 授業紹介
- 研究プロジェクト紹介
- 教員紹介
- 教員プロフィール

データサイエンス → データ中心科学



ビッグでないといけない？
正確なデータが重要

統計、人工知能(AI)

データサイエンスの基本ツール

CM好感度予測システム

株式会社コラージュ・ゼロ、CM総合研究所共同研究

海外・国内のベンチャー系ニュースサイト | TECHABLE

TECHABLE

INTERVIEW SOCIALMEDIA TECH MOBILE MARKETING ENTERPRISE START UP RELEASE RANKING

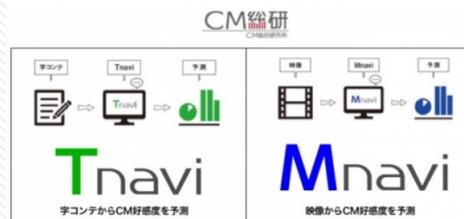
🏠 / Enterprise / AIを駆使したCM好感度予測システム「Tnavi」「Mnavi」

2019/7/6 Enterprise

AIを駆使したCM好感度予測システム「Tnavi」 「Mnavi」

毎日、テレビ、ラジオ、ネットで大量のCMが流れている。企業としては、自社の商品やサービスを広く知ってもらいたい一心であるものの、せっかく制作するCMがユーザーにどう受け止められるのかは、気になる問題だろう。・CM好感...

SHARE ON



<https://techable.jp/archives/102973>

リアルタイムで顧客の好みを学習 (アリババ)

- 淘宝 (タオバオ、Taobao) – ショッピングモール
 - 「強化学習」の一つである「半教師学習」を導入
 - ごく短期間のうちに顧客行動を学習することが可能
 - サイト訪問者それぞれにあったカスタムページをリアルタイムで作成し、サイト訪問者が興味を持ちそうな適正な価格の商品を提示
- 商品情報の作成 → AIコピーライター
 - 「ディープラーニング」、「自然言語処理」をもちいて1秒間に2万行のキャッチコピーを生成
 - どのようにすれば、検索結果の商品リンクをサイト訪問者にクリックしてもらえるか？
 - クリック率の向上

予測出荷の実現を目指す（アマゾン）

- 顧客の行動を予測して、注文すらしないうちから商品の出荷作業を進める予測出荷
 - 顧客の注文前から次の注文を予測し、予測した商品を配送先住所の指定はせず、その顧客の地域の大型倉庫に出荷する
 - 顧客の近くに運んでおいた商品が注文されなかった場合、返品輸送するよりも、無料で顧客にプレゼントしてしまった方が、費用対効果が高いかもしれない

リアルとバーチャルをつなげ顧客の傾向を読み取る(バーバリー)

- 英国ファッション小売企業バーバリー
 - 高級ブランド製品 50カ国 500以上の直営店および販売店
 - 実店舗の競争力を維持しつつ、ネット販売を展開
 - 高級品を購入する層
 - 買い物をしながら特別な場所で手厚いおもてなしを受ける
 - オンラインでの当たり前を実店舗で実施
 - 顧客に事前に許可→オンラインと実店舗の購買傾向をシームレスに予測
 - リアルの商品はRDIFタグによりどれを手にとったかを把握
 - リアルの店員は顧客の嗜好を把握→より興味をもちそうな商品を提案

在庫を最小限にして顧客の好みにあった商品を送付 (ステイッチフィックス)

- 米国 創業2011年
顧客が身につけたいアイテムを予測して自宅に届けるパーソナル・スタイリング・サービス
 - スタイリングの仕事を担うのはデータサイエンティスト
 - 顧客の好みを分析し、その顧客が着たいと思う服を予測、提供
- オンラインショッピング(食品以外の総小売支出額に占める割合)
 - 2013年：11.6%、2017年：24.4%
- ファッション小売業界→返品率が多い→在庫管理が難しい
 - 過剰な在庫を抱え込まざるを得ない
 - 顧客の嗜好を分析し期待を満たす服を提供しつつ、返品を最小限に抑える
 - 人間の目利きと機械の目利き
 - 顧客の気に入りそうなアイテムを自動的に送付
 - 倉庫のスペース、配送コスト、返品コスト、過剰な在庫を最小限にする

膨大な志願者の中から絞り込む ユニリーバ

- 世界各国25万人から8000人をしぼり込む(4-6ヶ月で)
- オンライン履歴書
- オンラインゲーム
 - 勝ち負けよりも志願者の特性を明らかにする独自のゲーム
- ビデオインタビュー
 - 志願者の言葉、顔の表情、ボディランゲージを分析することで、特定のポストで成果を残せる可能性の高い人物かを判定
 - 志願者の特性を現在すでに実績のある社員と比較

説明可能なXAI

- AIのブラックボックス問題、AIの説明可能性
- 精度の高さだけを追い求めるAIではない、人間を納得させるAI
- 米国Google 2019年11月
 - クラウド型AIサービスに説明可能なAIの機能を加えた
 - 機械学習のモデルの出力について、入力のどの要素が寄与したを出力可能
 - ユーザはAIがなぜその判断をしたのか把握
- 米国マイクロソフト、米国IBMも技術開発に着手

主なAIのトラブル

年	企業	説明
2015年	Google	Photosアプリで黒人画像にゴリラのタグ付けをしたことを謝罪
2016年	Microsoft	自分で会話を学習できるAI bot Tayが暴言を吐くようになってしまったためにサービスをシャットダウン
2018年	Amazon	差別的な人材採用AIを廃止
2018年	Facebook	Facebookが収集するユーザデータの一部が英政治コンサルティング企業ケンブリッジ・アナリティカ社にわたり政治利用されたとされる
2018年	Uber	自動運転車が死亡事故を起こす
2019年	ゴールドマン・サックス	Apple Cardの利用者の信用スコアを算出する際、女性に不当に低いスコアが付されクレジットカード限度額に差が生じている問題に関して非難を受ける
2019年	Facebook	広告主の指定したターゲットとは関係なく、雇用や住宅差別を固定化する広告推奨アルゴリズムを使っていたことが判明
2019年	Changsha Shenduronghe Network Technology(中国)	ディープフェイクiOSアプリZAOのプライバシーポリシーに、自分の顔データをメーカーに無償で永続的に提供するという記述があり、さらにデータの第三者への譲渡、再ライセンスまで可能となっている問題を指摘され、同社がプライバシーポリシーを改定
2020年	Ofqual(英国の資格・試験統制機関)	アルゴリズムによる成績予測評価が労働者階級やマイノリティに属する生徒に不利な評価を下すことが判明
2020年	-	白人警察官による黒人死亡事故に端を発したBlack Lives Matter運動が加速し、顔認識AIの誤認識による危険性に注目が集まる IBMが顔認識AI事業から撤退

XAIの問題点

- ブラックボックスの機械学習モデルを解釈する方法を編み出すのは避けるべき[*]
- 「人間が直感で解釈できそうなところの一部を分析」 or 「人間が直感でも解釈できそうな別のモデルに当てはめる」
- そもそも
線形分離可能なデータ → 説明可能性の問題にはならない
線形分離不可能なデータ → 人間が解釈しやすいモデルは存在しない

[*]C. Rudin, Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead, *Nat Mach Intell* 1, pp.206–215, 2019.

GPT3

- OpenAI(<https://openai.com/api/>)が2020年6月に公開
- 文章作成AI
 - 短い文章を入力すると、それに対応する新しい文章を生成
 - ニュースのタイトルを入力すると偽ニュースを生成

- 例)

- 入力「Ukraine begins first war crimes trial of Russian soldier」
- 出力
- 「Ukraine begins first war crimes trial of Russian soldier

Russian Army soldiers are to be condemned in the Supreme Court for being in the war-caused massacre of civilians on the orders of the president of Russia.

A spokesman for the Army spokesman said the military commanders are due to begin their trials of the two soldiers.

The trial of the Russian soldiers who allegedly participated in the killings of three civilians - 12 women and six children - has been held in the Supreme Court of the Russian State Duma-Moscow.

Moscow denied the allegation that the army commander was involved in the military killings. It said that the chief commander」

データによる意思決定を実現するための 4ステップ (AIサービスを立ち上げる4ステップ)

表出化

- 課題を設定するために、問題点だと思われる物事について、できる限り暗黙知から形式知にする。

要件化

- 課題を解決するために必要な要件をもれなく挙げる。

データ化

- 各要件がデータとして存在するか、もしくは今後取得可能かを検討する。

指標化

- 課題を解決するために、「データをどのように解釈するか」を規定すること。

表出化[1/2]

- 課題を明らかにするプロセス
- 課題を設定するためには、問題点と思われる物事についてできる限り暗黙知から形式知にする必要がある
 - 暗黙知とは言語化されていない主観的な知識
 - 他人に伝えることが難しい
 - 形式知とは言語化された客観的な知識
 - 他人に伝えることが可能
- 日々の業務の中で「ムダだな」「うまくできていないな」と思うところを特にピックアップして、それらの根本的な問題は一体何なのかをなるべく記述する

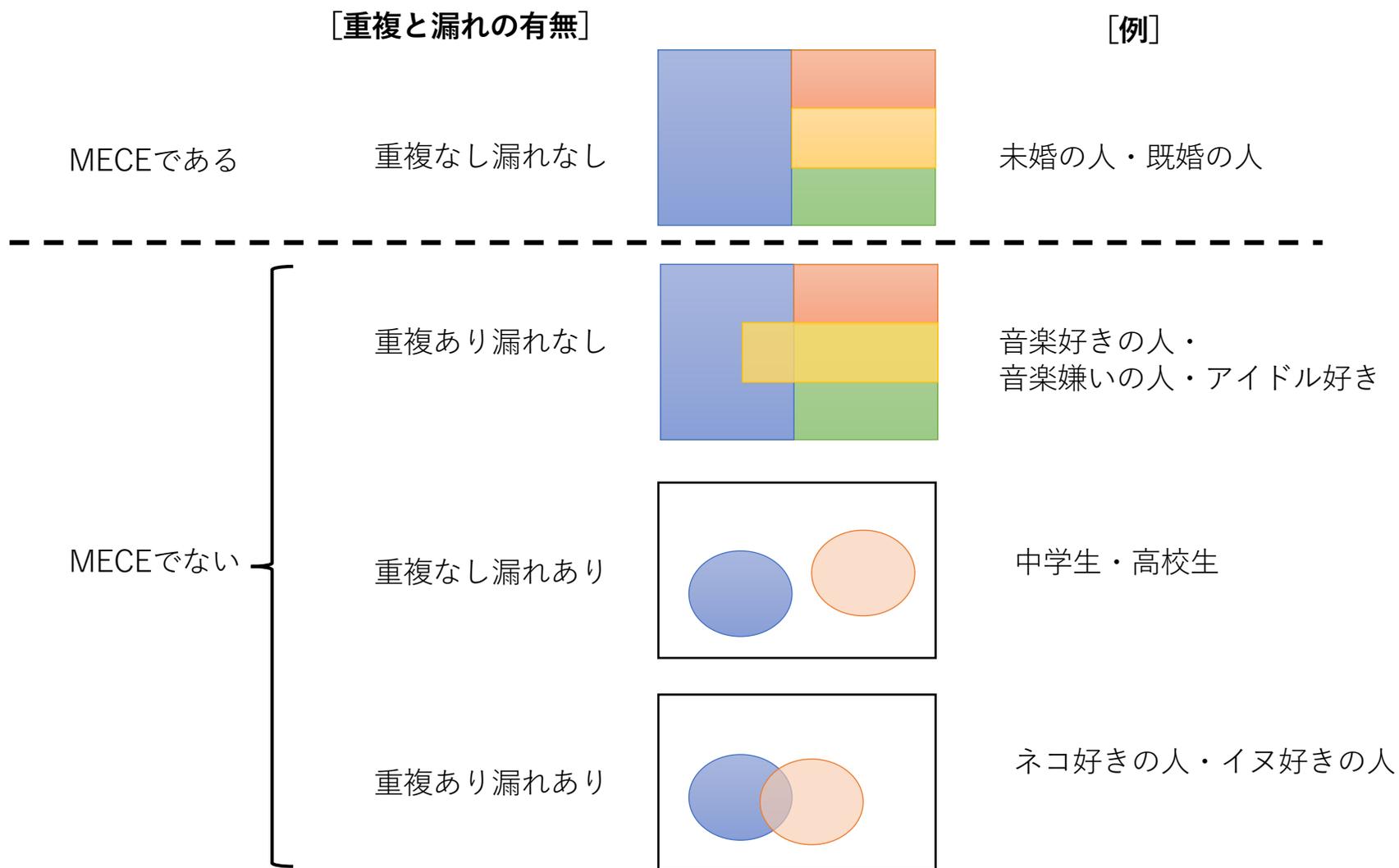
表出化[2/2]

- 選ばれた課題を人工知能で適用できるようにするためには「〇〇を自動化」という形で分解する必要がある
- 分解の方法
 - Who：誰のために自動化すべきか
 - What：何を自動化すべきか
 - When：いつ自動化すべきか
 - Where：どこで自動化すべきか
 - How：どのように自動化すべきか
- どうしても人工知能で無理やりカバーしようとしなないこと
 - 技術が進展した今でも、人間がやった方がコストがかからないものも存在する
- 人工知能もツール
 - 人工知能はさまざまなシーンで適用できるような潜在能力が高いツールであるが、どのツールにするのかの選択肢の一つに過ぎない
 - どのツールを適用するのかを考える

要件化

- 表出化で抽出した課題を解決するための要件が漏れなく規定すること
- 要件を定義できない場合は人工知能を使った自動化に適さない
 - 要件がよくわからない
 - 無限にある
 - 時と場合によって異なる場合
- 表出化によって選ばれた課題を解決するために必要な要件を具体的にあげるだけ全部あげることが必要
 - この段階で必要な条件を全て挙げられるか
 - →MECE(ミーシー)
 - Mutually Exclusive and Collectively Exhaustive
 - お互いに重複せず、全体に漏れがない

図表10 「MECEである」とは？



データ化

- 要件化で挙げられた各要件がデータとして存在するか、もしくは今後取得可能か
- 現実世界で起こっていること(要件化で挙げた事象)がデータとして取得される必要がある
 - 要件化で挙げた要件についてデータとして取得できるか
 - データとして取り組むことが不可能/可能だとしてもコストが大きすぎるか
- 直接データとして取得できなかったとしても、別の関連するものがデータ化可能であれば、そこから推定できるかもしれない
 - →フェルミ推定
 - →IoTデバイス

フェルミ推定

- 実際に導出することが難しいような値をいくつかの他の取得可能な値から推論し、概算する手法
- 例) アメリカのシカゴには何人のピアノ調律師がいるのか
 - ピアノの調律師の人数を直接調査しなくても下記の値から推定可能
 - シカゴの人口
 - 一世帯あたりの人数
 - ピアノを持っている世帯数の割合
 - 一人の調律師が1日にピアノを調律できる台数
- 直接データを取得できなかったとしても、他のデータで取得できるものから推定可能かを検討する

IoT

- 現実世界の事象や現象をセンサーやデバイスなどでデータとして取得する仕組みを構築することで、さまざまなモノがインターネットで繋がっている環境を創造すること
- データ化を検討する際にIoTデバイスにどんなものがあるのかを調査すると思わぬ事象や現象が低コストでデータとして取得可能

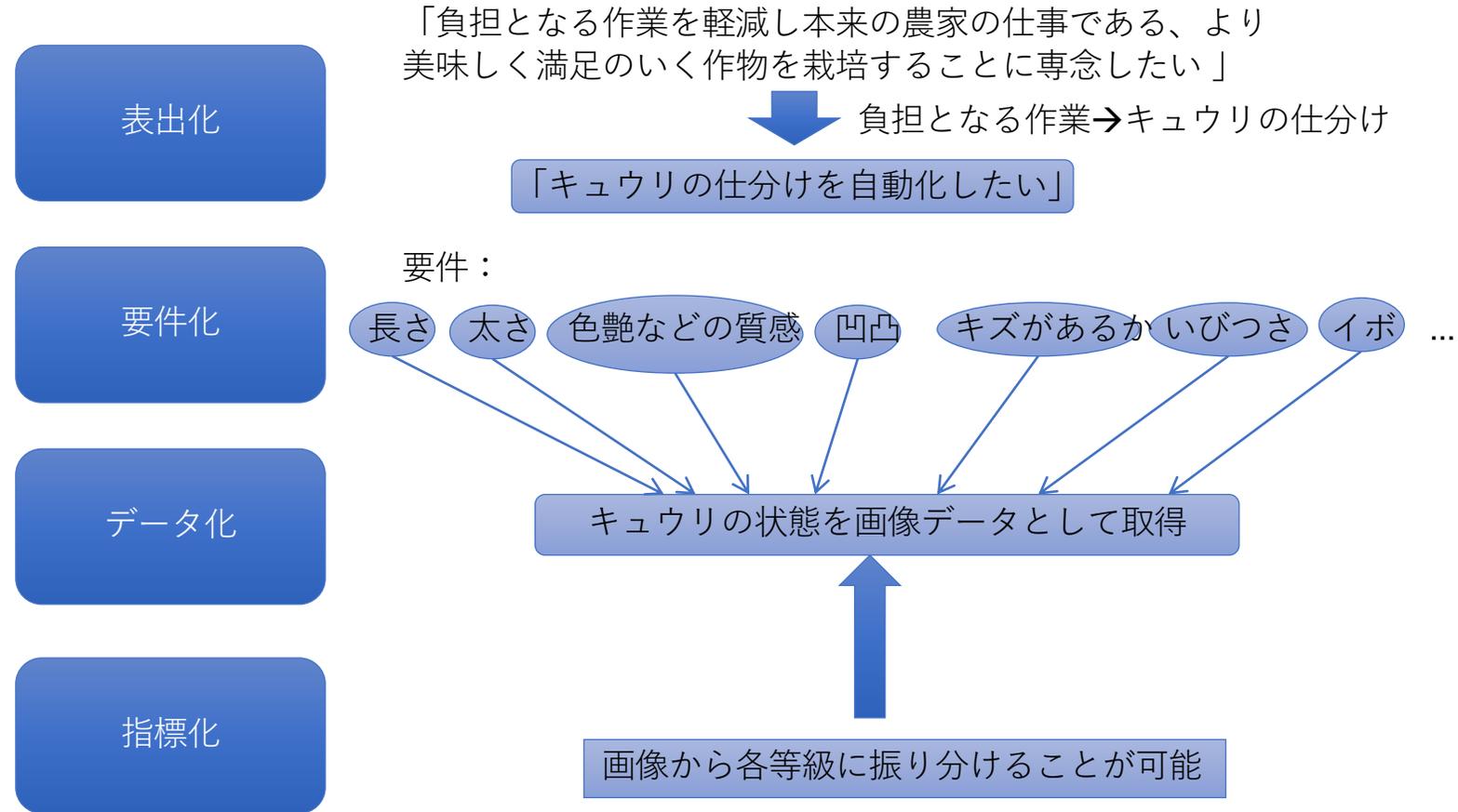
指標化

- 各要件をデータとして取得したとして、そのデータから課題を解決するために、どのように判別するのかを規定
- データをどのように解釈するかを決める
- 例)顔認識技術の場合
 - Aさんの顔写真、Bさんの顔写真
 - どういうデータがAさんなのかBさんなのか
 - 各要件に対応するデータを課題に沿ってどのように解釈できるか(ラベルづけできるか)を検討する

キュウリ農家による人工知能の活用例

- 2016年ごろに静岡県できゅうり栽培農家を営む元システム開発者が独力で、ディープラーニングによる画像認識を使って自動きゅうり仕分機を作った
- きゅうりの仕分けは、長さだけでなく、太さ、色艶などの質感、キズ、形の歪さなど9段階で等級が決められている
 - 作業に慣れた人にとっても集中力が必要となる作業
 - 負担の重い作業
- この作業を自動化することにより、本来の農家の仕事であるより美味しく満足のいく作物を栽培することに専念できる

キュウリ農家による人工知能の活用例



一人で人工知能システムを実現できる

- 人工知能と聞くと、膨大なデータが必要であったり、特別な環境下で専門の複数のプログラマーが特別な能力を使って開発をしたりしないといけないというイメージはないか？
- きゅうり栽培農家の方は一人でデータの準備から実装までを行なっている
- ディープラーニングを実現するための「TensorFlow」ライブラリ公開
- Raspberry Piと呼ばれる非常に安価なデバイス
- 単純な実装、小規模なデータ量から(スモールデータ)でもその試行結果を検証しながら進めていければ、段階を踏んで課題を解決する人工知能を含んだシステムを構築していくことが可能

分類のAIの自動化例

- 三重県伊勢市にある創業100年を超える老舗食堂ゑびや
- 顧客の来店率をリアルタイムに計測するのに分類によるAIの自動化を実現
- 人手不足に悩む小売飲食店においても人工知能を用いた自動化による効率化を実現している例

老舗食堂による人工知能の活用例

表出化

要件化

データ化

指標化

「店舗での施策が本当に効果的、効率的なのを知りたい。
例えば、店頭ディスプレイの効果を知りたい」



店頭ディスプレイの効果 → 入店率の変化

「入店率の測定を自動化したい」

要件：

店の前を通った人の数

入店した人の数

店頭カメラで画像データとして取得

店の前を通った人、および入店した人を
画像認識で検出

画像から店の前を通った人、および入店した人を認識し
その数から入店率を導出

さらに

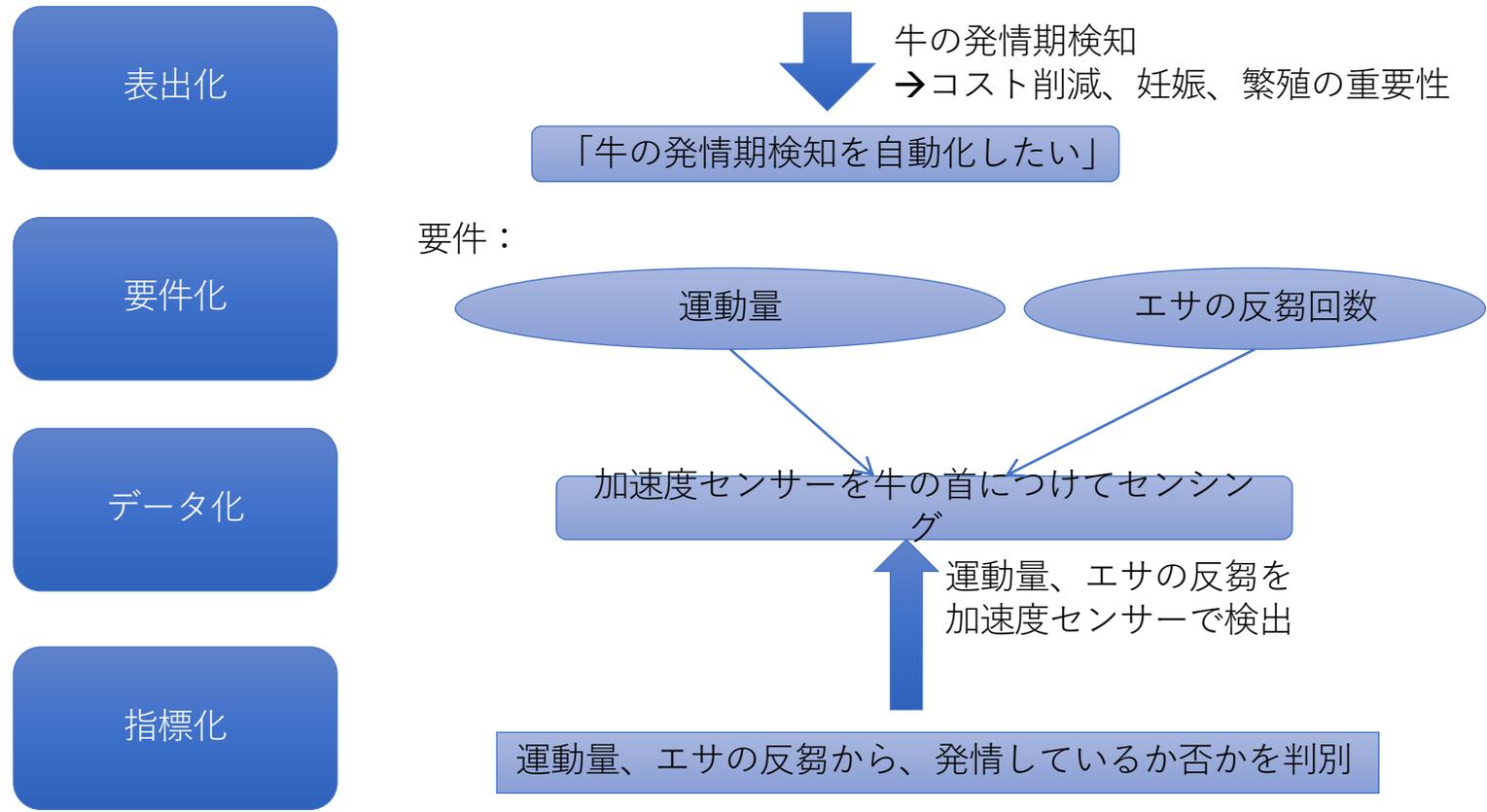
- 次に紹介する「回帰」のAIの自動化を行うことによって、入店率の予測をおこなっている
- 実際90%以上の確率で予測できる

- 食材の排気ロス
- 対人による接客を強化する取り組み
- →売上が従来の3倍に伸びる

牛の行動データから発情期の牛を分類

- 畜産農家が行なっている作業の中でも、牛の発情期を見極めが重要
- 畜産農家の利益を最大化するためには、牛の発情期をなるべく正確に機会を逃すことなく検知する必要
 - 発情しているかどうかは1頭ずつ随時観察が必要
 - 牛は3週間に1度、数時間しかない発情期を正確に判断

畜産業における人工知能の活用例



表出化

要件化

データ化

指標化

データ基盤を作ったもののユーザ部門が使えていない

- どの業務で使えるのかユーザ部門がイメージできていない
- 情シス部門とユーザ部門がそれぞれ寄り添い、その部門の課題を見つけ出し、その課題を解決するデータ分析を共に作り上げていくことが重要
- 社内事例の創出
- ユーザ部門のデータリテラシ向上