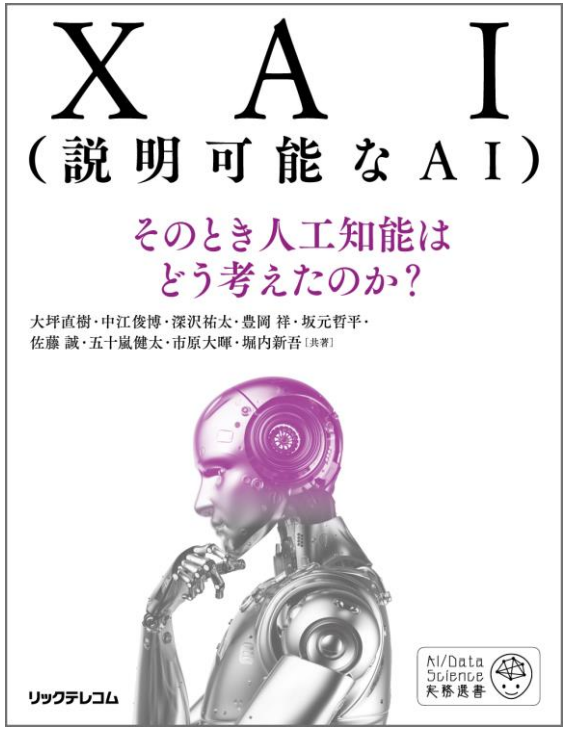
NTT DATA
Trusted Global Innovator

NTTデータ数理システム　ユーザーコンファレンス2021
ＸＡＩ（説明可能なＡＩ）
～そのとき人工知能はどう考えたのか？～

2021年11月17日･18日
株式会社NTTデータ

本の紹介

タイトル	XAI（説明可能なAI） ――そのとき人工知能はどう考えたのか？
著者	◎大坪直樹、中江俊博、深沢祐太、◎豊岡 祥、◎坂元哲平、佐藤 誠、五十嵐健太、◎市原大暉、堀内新吾［共著］
出版社	リックテレコム
発売日	2021/7/14
規格・ページ数	B5変型判・248ページ
定価（税別）	2,600円
公式URL	https://www.ric.co.jp/book/new-publication/detail/1843



出版の経緯

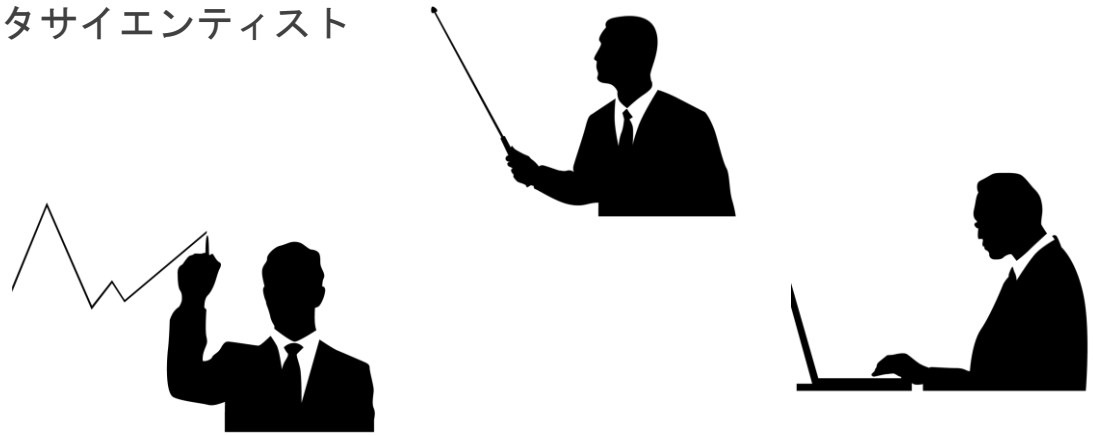
書籍化のニーズと、業務における経験の蓄積の２つがタイミングよくマッチしたことから出版に至った。

１．書籍化のニーズ

２．業務における経験の蓄積

出版社

AIエンジニア・
データサイエンティスト



1．書籍化のニーズ

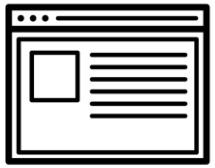
世間一般に「XAI」をターゲットとした体系だった書籍は存在していなかった。

⇒XAIそのものに対するニーズは高く、書籍化のニーズがあった。

・初学者にも理解できる技術習得
・ビジネス目線も含めた実践的な知識獲得

に適した書籍は出版されていない！

- XAIに関する関心の高まり（次ページ）

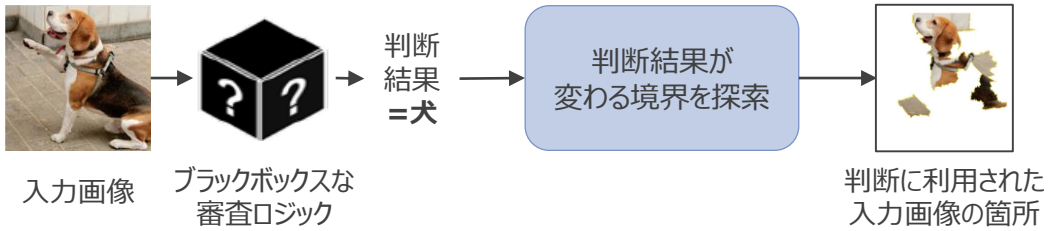
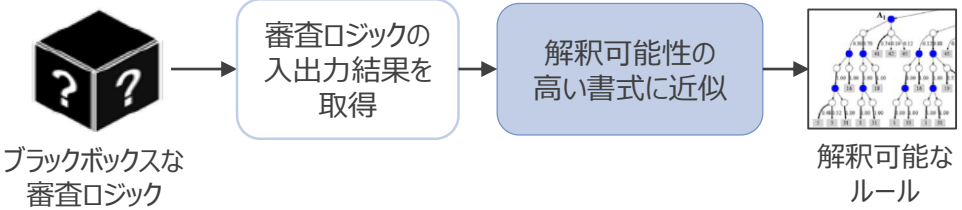


- WEB媒体：数ページ程度、簡潔な論調に留まる。
- 個人ブログ：XAIライブラリの使い方等を紹介、体系立った解説には及ばない。

- 書籍（国内）：XAIを主軸に据えて解説したものは存在しなかった。

XAIとは

XAI（eXplainable Artificial Inteligence）　：複雑化するAIの予測根拠を理解するための技術

	技術概要	用途
局所的XAI	個々の判断における入力項目と結果の関係性を把握 	個々の入力ごとの判断根拠を説明→ • モデル劣化の検知 • 誤答原因の分析
大局的XAI	複雑な審査ロジックを解釈可能なルールに変換 	モデル全体の判断傾向を可視化→ • モデル適用可否（リリース）の判断

XAIが注目される背景

精度偏重に発展してきたAIの課題

- 画像解析、自然言語処理など様々なタスクで人間を超える精度を達成
- 内部を複雑に組み合わせてきた精度追求のAIを、人間が具体的に理解することは困難

世界中で「AIの説明可能性」について訴求

- 指針やガイドラインなど各国から打ち出されている

説明可能性に対する各国の見解

AI利活用原則案（総務省，2018）
透明性の原則 ・ 説明責任の原則

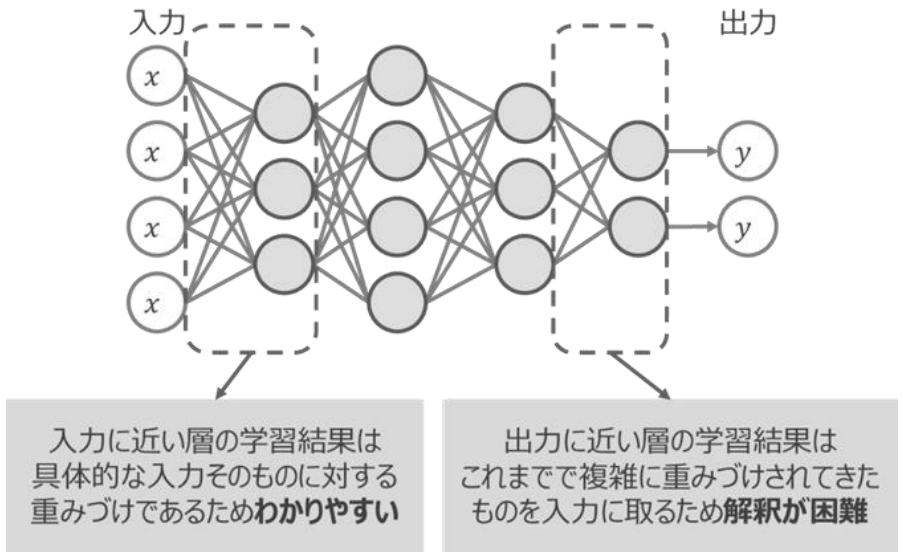
一般データ保護規則（GDPR）

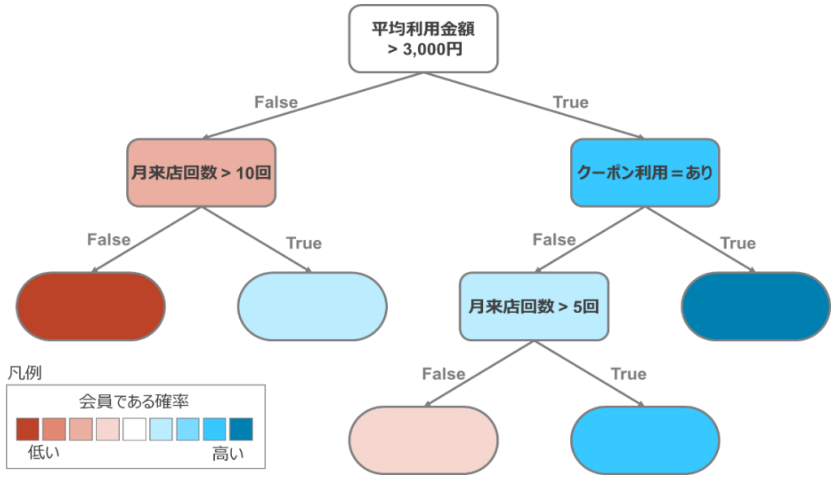
XAI，DARPAプロジェクト

実社会へAIを活用していくうえで生じる摩擦を避けるために、様々な要件を定めている

そのひとつに「説明可能性」があり、AIの予測に対して解釈できることの必要性が訴求されている

AIの解釈性





- AIの高精度化：ディープラーニング、など
- AIの内部ロジックイメージ（左上図）：内部過程を理解することは不可能に近い、実際のモデルはさらに複雑
- 古典的なAI（左下図）：理解はしやすいものの、精度は望めない
- つまり、AIの精度と解釈性はトレードオフ

- 業務的な目線からは・・・単純な精度だけでなく根拠が必要とされる場面が多い
- 高精度なAIを用いつつ、予測に対する説明性を持たせるための技術としてXAIが登場した

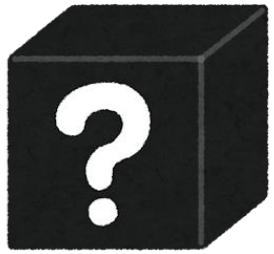
2．業務における経験の蓄積

NTTデータやNTTデータ数理システムでは、より人間の判断が重要視される領域でのAI活用に積極的に取り組んでおり、XAIに対する知見も蓄積していた。

・試行錯誤を経て培った技術の知見
・AIのビジネス適用における経験的なノウハウ

を保有

- 業務的な判断に影響するAI開発やデータ分析への取り組み多数あり。（後述）



- 業務的課題：AIのブラックボックス性

- 貢献：XAIを含めた改善・検証活動
- 強み：XAIをただ使えばよい訳でなく、事例に応じた工夫も必要。

本の紹介～目次～

XAIが求められる背景を解説

XAIライブラリの使い方を習得

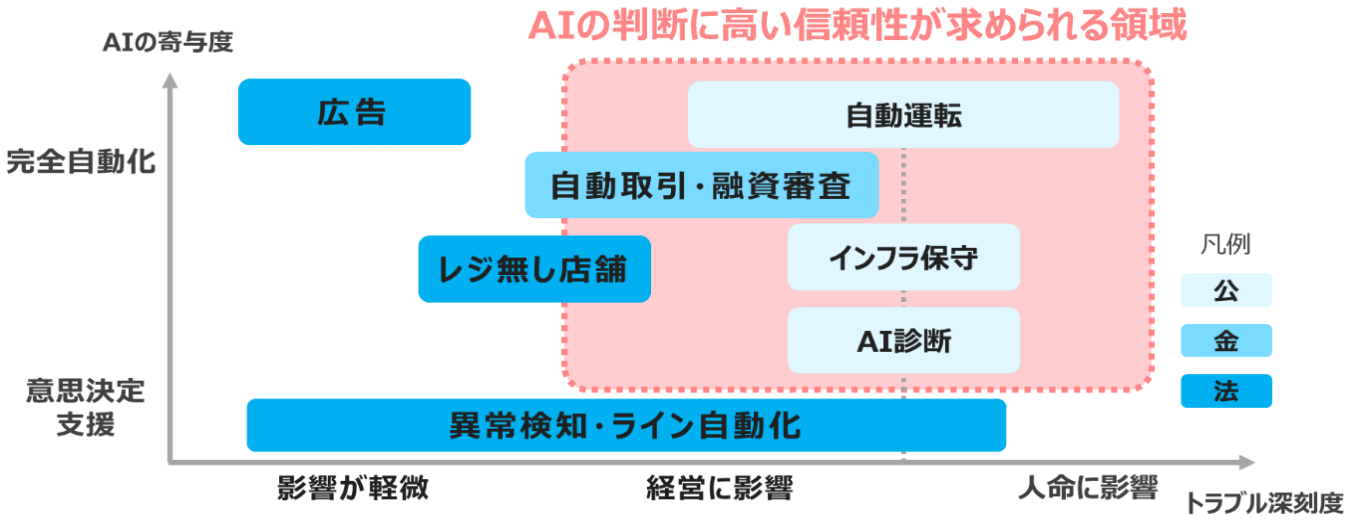
XAIに関する体系的な知識を獲得

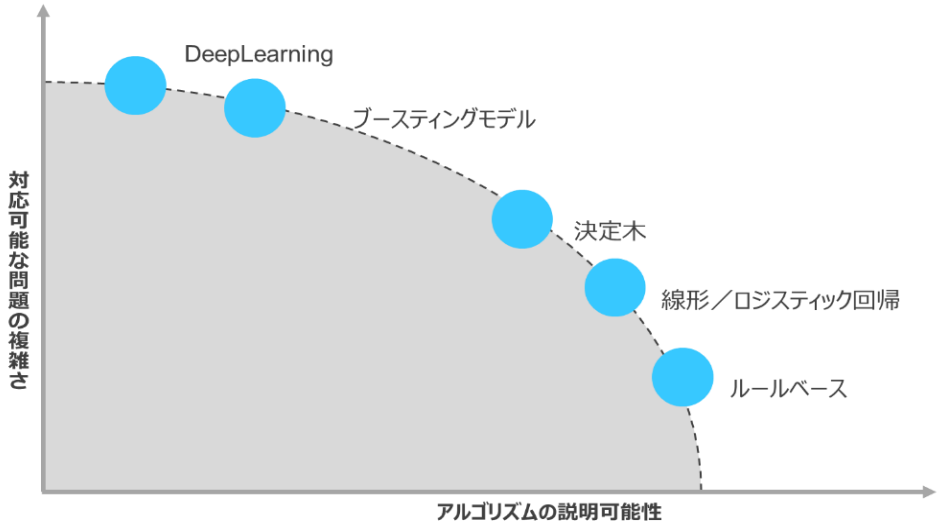
実務からみた課題や将来展望を考察

NTT DATA

本の紹介～第Ⅰ部：課題設定～

- AIの技術革新に伴い、特にビジネスにおける判断にまでAIが適用されるシーンが増加している
- AIの精度と解釈性はトレードオフ
- 高精度なAIを活用しつつ解釈性を高めたいニーズからXAIが注目されている





本の紹介～第Ⅱ部：基礎知識～

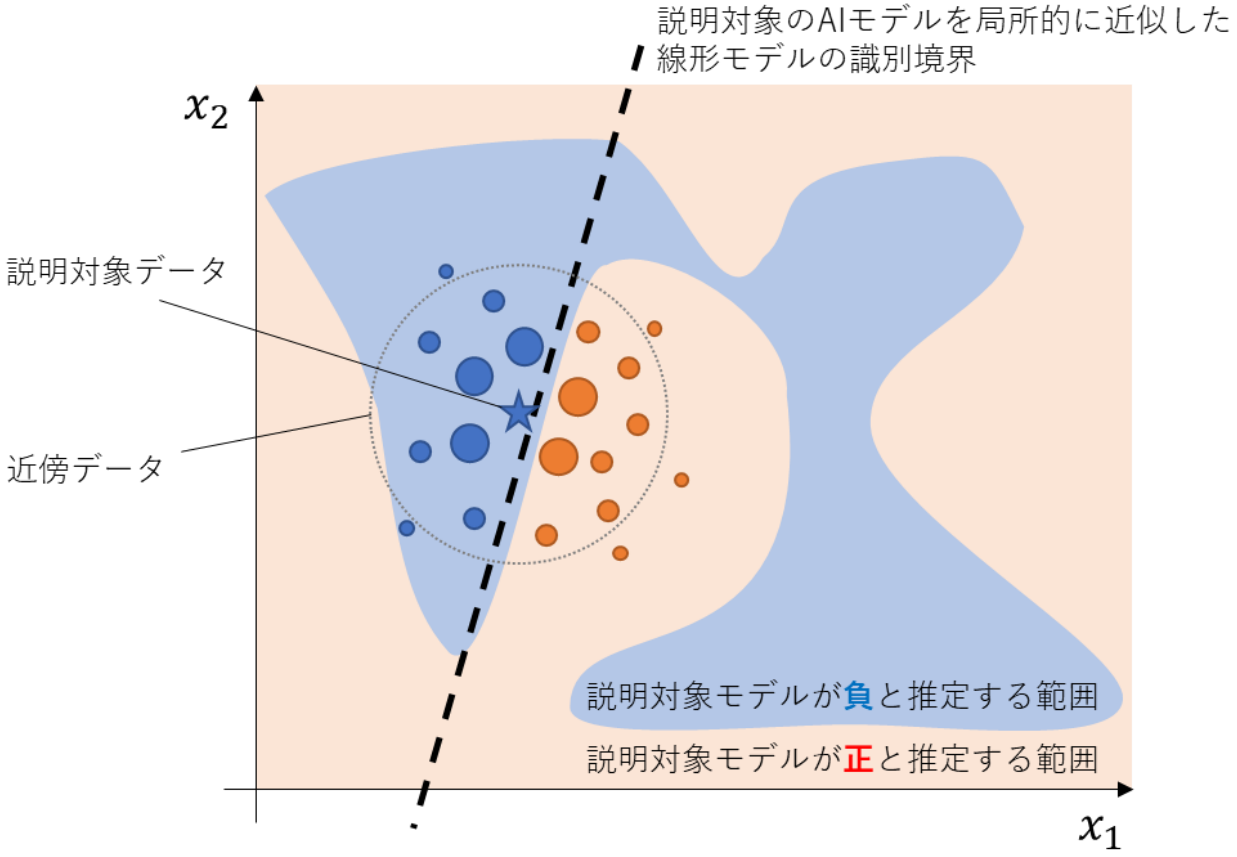
- XAIは、大局説明と局所説明がある
- 大局説明：AIの全体的な振舞いを理解するもの
- 局所説明：ある予測結果への根拠を示すもの
- 例：LIME

  AIが用いる変数を空間的に表現

  AIはこの空間内で複雑な予測の違いを見せる

  説明対象データ（★）付近の予測より近似

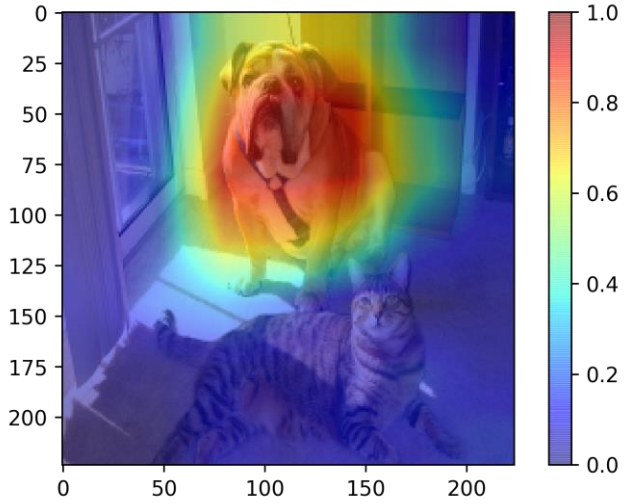
  近似式より説明対象データを簡潔に解釈可能



本の紹介～第Ⅲ部：実践指南～

- PythonのXAIライブラリ（LIME, SHAP, etc.）
- サンプルコードを用いて使い方を例示
- 対象：表形式、画像、テキストデータ
- 実際のデータに対してどのような説明が得られ、
- どういった解釈が出来るかを解説

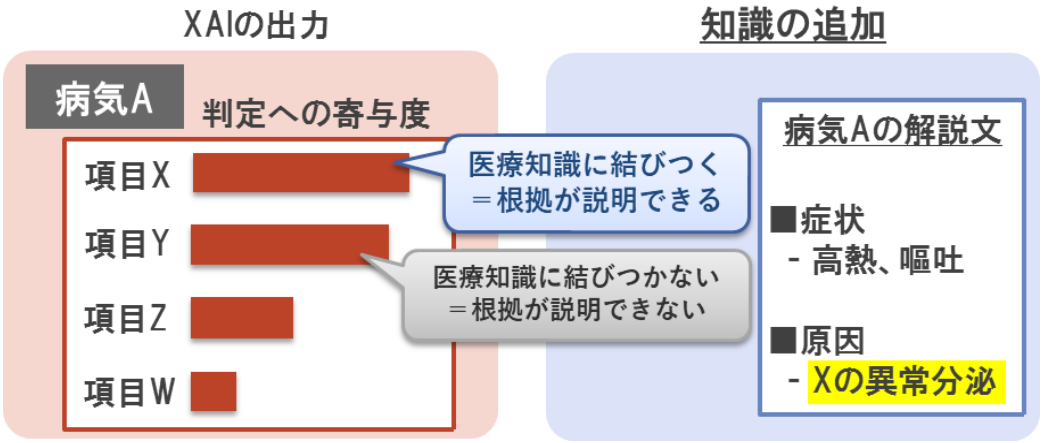
```
# 分析用テーブルを作成
pred = predict_fn(test_X)
result = pd.DataFrame({"pred":pred[:,0], "Sex":test["Sex"]})

# 男性なのに Survive と予測する例を取得
target_idx = result[result["Sex"]==1].sort_values("pred").index

# LIME の実行
i = target_idx[0]
exp = explainer.explain_instance(test_X.values[i], predict_fn, num_
features=5)
exp.show_in_notebook()
```



NTT DATA

本の紹介～第Ⅳ部：将来展望～

- 流れを整理し、AIに求められる「説明」を再度俯瞰
- 業務的に本質的に求められる説明とは何か？
- 本当の納得につながる進化へ期待



誰のため?(対象者)	何のため?(目的)	どのように?(主な方法)
モデルの開発者	モデルの振る舞いを知り改善につなげる	XAI
モデルを活用する人（エンドユーザ）	モデルが確からしいことの納得	XAI 実験・精度レポートの提示
モデルを活用する人	施策立案およびそのためのヒント	XAI 人の判断やモデル外での実装
モデルを活用する人	意思決定のためのエビデンス	過去の判断との比較 モデル外での検証（シミュレーション）
エンドユーザ社会	説明責任	実験・精度レポートの提示 当初想定とトラブル発生時の差異の提示

事例紹介

- 事例１：文書からのリスクチェック
- 事例２：画像検索AIシステム
- 事例３：大規模スパースデータの分類
- 事例４：公平性判断のための説明

事例１：文書からのリスクチェック

タスク概要	● 各種プロジェクトの進捗や関連文書から、プロジェクト進行上のリスク有無を判定
AIの課題	● 記述されるプロジェクトの種類や件数が多く、記述内容だけでリスクの有無を一見して判断することが難しい （どういった表現が高リスクと判断されるのか見分けるには、そのプロジェクトに対する専門性が求められる）
適用したXAI	• LIME
XAIの効果	● AIによって高リスクと見なされるプロジェクトの文書記述について、単語などのレベルでリスクへの影響度合いを可視化することで、専門性を有していなくてもなぜリスクが高いか分かる （例．事業部門によって記述された文書の表現に対して、コーポレート部門がリスクの有無を具体的に判断可能）

事例２：画像検索AIシステム

タスク概要	• 膨大な画像データの中から、検索対象の画像に対して類似性の高いものを検索するAIシステム
AIの課題	• 部分的に見ても類似性が高ければ検索対象としては上位である必要があるが、ヒットする膨大な画像の中で領域としての類似性を一見して判断することは難しい
適用したXAI	• LIME
XAIの効果	• 検索結果の画像について類似性が高い部分領域をハイライトすることによって、類似度スコアのみで検索順位をつけて結果表示する場合よりも類似性の判断が容易 （例．全体的な色味が似ている２位の画像よりも、部分的に形式が酷似している６位の画像の方が類似性は高い）

事例３：大規模スパースデータの分類

タスク概要	• 表形式データを入力として問題の有無を判別する業務のAI化
AIの課題	• データ量が大規模（総レコード数：数億件） • 入力項目が多次元スパース（数万種類から数十個前後のみが記録される） ● 通常、表形式データのXAIを用いると、数万種類の入力項目をすべて検査して説明するため、内容が冗長となり解釈が非常に困難
適用したXAI	• SHAP
XAIの効果	• 個々のデータに記録された項目の中に絞って、影響度が大きな要素を示すことで解釈が容易 （例．10000品目から10個選ばれたデータについて、10000品目すべての重要度は計算せず、10個の中から重要な要素を提示する)

事例４：公平性判断のための説明

タスク概要	• AIに求められる要件のひとつである「公平性（Faireness）」 ※例：性別を理由に合否判定が変わっている • 倫理上の観点からAIによる差別的な予測が懸念されており、作り上げたAIモデルに差別的な要素が無いか精査する必要がある
AIの課題	• 公正性が保たれたAIであることを判断するため、AIがどの項目を重視しているか示す必要があるが、個々のデータについて差別的な要素を判断に用いていないか確認することが求められる
適用したXAI	• LIME
XAIの効果	• AIの予測において重視している要素を明示することで、そのAIが性別や人種などの特徴に基づいた差別的な判断を行っていないか判断する

NTT DATA

Trusted Global Innovator

本資料に記載されている会社名、商品名、又はサービス名は、各社の登録商標又は商標です。

NTTデータ数理システム　ユーザーコンファレンス2021
ＸＡＩ（説明可能なＡＩ）
～そのとき人工知能はどう考えたのか？～

シミュレーション＆マイニング部
豊岡　祥

XAIの例: SHAP による説明変数の影響度算出

XAIのやりたいこと:

学習/テストデータ中の

- 説明変数の値
- 説明変数の有無
- 各学習データの有無

が

- モデルの性能
- モデルの挙動(予測値)

に与える影響を定量化する

XAIの例: SHAP による説明変数の影響度算出

SHAP = **SH**apley **A**dditive ex**P**lanations:
シャープレイ値という指標でモデルの出力を各説明変数の寄与の足し合わせに分解する

- 説明変数の値
- 説明変数の有無
- 各学習データの有無

が

- モデルの性能
- モデルの挙動(予測値)

に与える影響を定量化する

XAIの例: SHAP による説明変数の影響度算出

シャープレイ値

協力ゲームにおいて、各参加者 i の貢献度 $\phi(i)$ を定量化する方法のひとつ
「その参加者がいた場合といなかった場合の差分」
をベースに貢献度を計算する

$\phi(\text{player } 1) = 4$?
$\phi(\text{player } 2) = 16$?

2:8に分配

プレイヤー 1	プレイヤー 2	報酬
×	×	0
○	×	2
×	○	8
○	○	**20**

○: 参加
×: 不参加

XAIの例: SHAP による説明変数の影響度算出

シャープレイ値

協力ゲームにおいて、各参加者 i の貢献度 $\phi(i)$ を定量化する方法のひとつ
「その参加者がいた場合といなかった場合の差分」
をベースに貢献度を計算する

$\phi(\text{player } 1) = 4$?
$\phi(\text{player } 2) = 16$?

$$\phi(\text{player } 1) = \frac{((2-0)+(20-8))}{2} = 7$$
$$\phi(\text{player } 2) = \frac{((8-0)+(20-2))}{2} = 13$$

2:8に分配

シャープレイ値: $\sum_{i \in N} d(S)(v(S|\{i\}) - v(S))$
i の有無で生じる差分

プレイヤー 1	プレイヤー 2	報酬
×	×	0
○	×	2
×	○	8
○	○	**20**

○: 参加
×: 不参加

XAIの例: SHAP による説明変数の影響度算出

シャープレイ値

協力ゲームにおいて、報酬を各参加者 i の貢献度 $\phi(i)$ に分解する
「その参加者がいた場合といなかった場合の報酬の差分」
をベースに貢献度を計算する

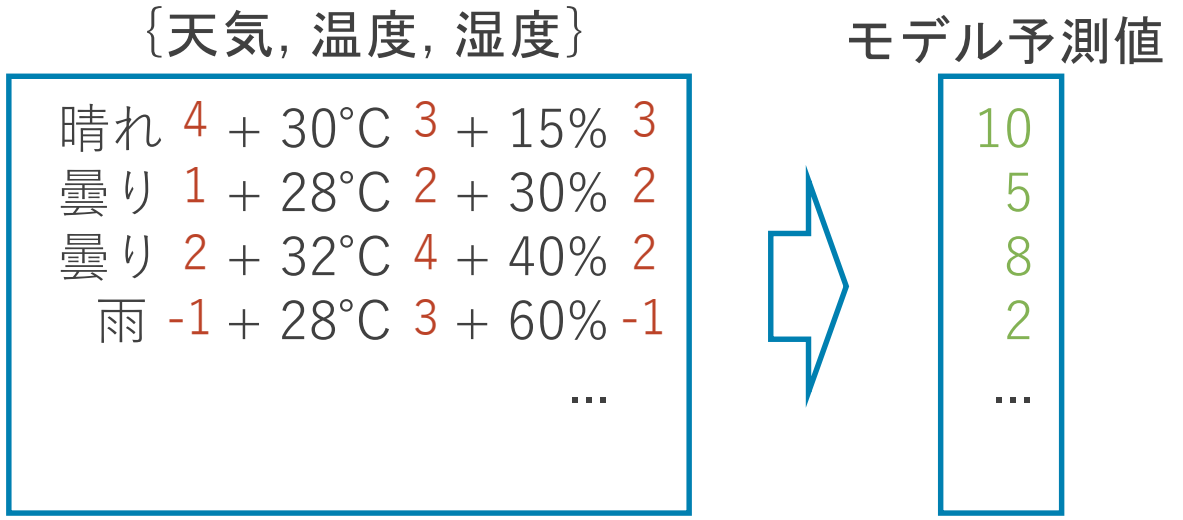
SHAP

機械学習モデルの予測値を各説明変数 i の影響度 $\phi(i)$ に分解する
「その説明変数を使った場合と使わなかった場合の予測値の差分」
をベースに貢献度を計算する

$$\sum_{i \in N} d(S)(v(S|\{i\}) - v(S))$$

XAIの例: SHAP による説明変数の影響度算出

SHAP計算結果のイメージ
アイスクリームの売上予測

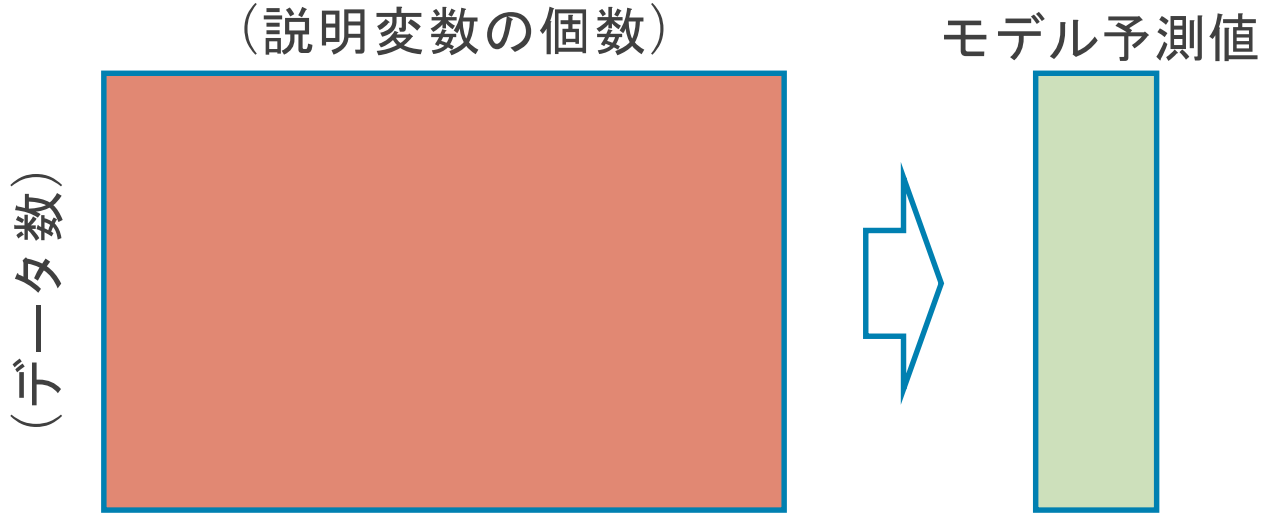


- SHAP値の総和が予測値に一致する(マイナスの値も取る)
  → 予測値の内訳として解釈できる
- 真の目的変数(売上)や、モデルの性能とは無関係
- 同じ説明変数の値でもSHAP値が同じとは限らない
- 元データと同じサイズのSHAP値行列が得られる

XAIの例: SHAP による説明変数の影響度算出

SHAP値行列の活用

元データと同じサイズの実数値行列
→ これ自体がデータ分析の対象になりうる

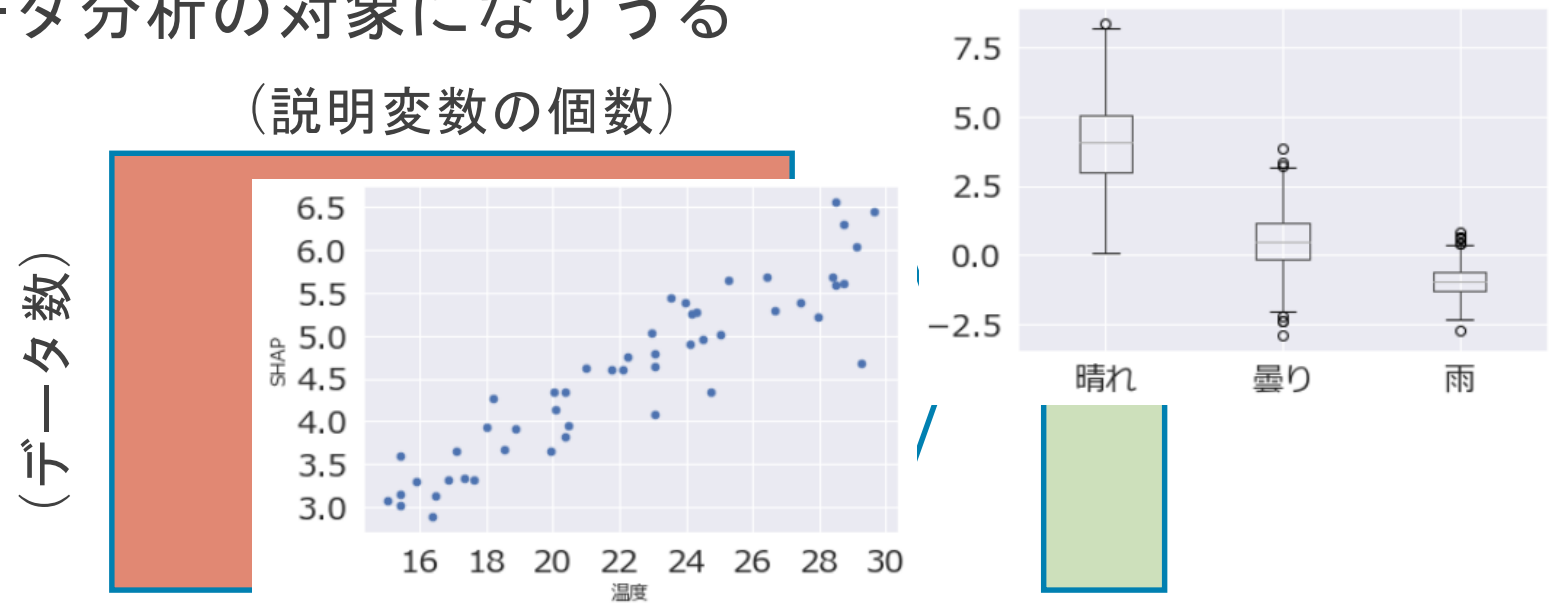


- SHAP値の分布集計
- データのクラスタリング
- 類似データの検索

XAIの例: SHAP による説明変数の影響度算出

SHAP値行列の活用

元データと同じサイズの実数値行列
→ これ自体がデータ分析の対象になりうる



- SHAP値の分布集計
- データのクラスタリング
- 類似データの検索

XAIの例: SHAP による説明変数の影響度算出

サマリ

SHAP: モデルの出力を各説明変数の寄与に分解

全データのSHAP値を俯瞰して見ることでより多くの情報を得ることもできる

※ あくまでモデルの入力・出力の対応のみを見ており、真のラベルやモデルの性能は全く反映しない
→ 目的に応じた手法の選定が必要

NTT DATA

NTT DATA Mathematical Systems Inc.