

信頼の概念を考慮した AI による提案の受容予測

中原 英里¹ 石井 方邦^{1,2} 倉沢 央¹ 大槻 知明³

概要：様々な産業分野において Artificial Intelligence (AI) 技術を利用し業務改善を試みた事例が増加している。しかし、AI が導出した業務改善提案を人が必ずしも受容するとは限らず、AI 活用の課題となっている。本研究では、職場において AI 利用者が AI の提案を受容するのに、どのような要素が影響しているのかを明らかにすべく、信頼の概念や AI に対する考え方、eXplainable AI (XAI) 手法を模した図を用いて AI の提案の説明する場面について聴取する質問紙を作成した。そして、この質問紙を使用し、オンラインにてアンケート調査した。統計分析の結果、同じ XAI 手法を模した図を用いた場合、AI による提案と人による提案の受容しやすさの間に違いは無いことがわかった。また、XGBoost を用いて AI の提案の受容を予測する二値分類モデルを作成した。特徴量重要度を確認することで、AI 利用者の人への信頼しやすさが AI の提案の受容に影響することがわかった。さらに、AI へのポジティブな認識を問う 3 つの質問を入力としたモデルで、AI の提案を受容するか否かを ROC-AUC = 0.73 の精度で予測できた。この結果から、AI にポジティブな認識を持つ AI 利用者は AI による提案を受容しやすいことがわかった。これらの検討により、簡易に AI の提案の受容しやすさを予測可能になった。

Acceptance Prediction of AI Proposal Based on Trust

Eri Nakahara¹ Masakuni Ishii^{1,2} Hisashi Kurasawa¹ Tomoaki Ohtsuki³

1. 研究背景

デジタルトランスフォーメーションの一貫として、様々な産業分野において Artificial Intelligence (AI) 技術を用いてデータ分析を実施し、業務改善を試みた事例が増加している。しかしながら、2018 年時点の我が国における企業への AI 導入率は 39% に留まり、中国の 80% や米国の 51% を大きく下回っている [1]。また、同年に情報処理推進機構が実施した調査では、AI 技術を業務に導入する際の課題として、「AI についての理解が不足していること」や「導入効果が得られるか不安であること」が挙げられている [2]。つまり、AI 技術を業務に導入するためには、AI 利用者が AI の提案を受容しやすい説明をすることが重要である。

AI の提案の受容促進に向けて、2017 年頃より AI の出力を説明するため技術である eXplainable AI (XAI) が盛んに研究されている [3]。XAI 技術により多種多様な形式で AI から得た結果を可視化及び解釈することが可能になっ

た。しかし、XAI 技術はあくまでも AI の解釈を機械的に行う技術であり、AI 利用者にとってどのような説明がわかりやすいのか、提案を受容しやすいのかについて十分に検討されていない。

また、AI の安全性を保ち説明責任を果たすために、複数の機関が AI のガイドラインを策定している。2019 年には EU が信頼できる AI に関するガイドラインを策定し、信頼できる AI の実現には、AI によって人々の生活に害を及ぼすことなく実社会で AI を信頼して活用することが重要であると示している [4]。

しかし、AI の説明性や、安全性、説明責任は、AI 利用者の AI 受容性を高める要素の一部であり、AI の提案の受容への影響については十分調査されていない。また、実社会で信頼して AI を利用するためには、AI 利用者が十分に納得及び受容できるような説明を行う必要があると考えられる。先行研究 [5][6] では、信頼の概念に着目し、AI による提案の理解について調査を実施し分析しているが、これらの研究の質問紙で採用されている場面設定が AI による実刑判決の提案や食事プランの提案といったものであり、

¹ 日本電信電話株式会社 コンピュータ&データサイエンス研究所

² 慶應義塾大学大学院 理工学研究科

³ 慶應義塾大学 理工学部 情報工学科

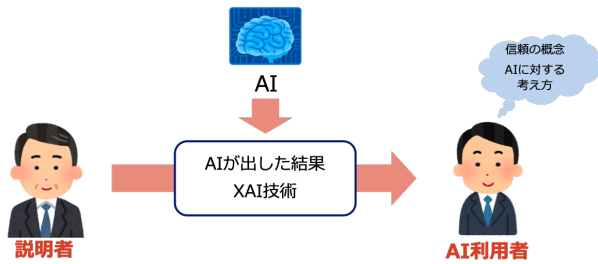


図 1 AIと「説明者」と「AI利用者」の関係

様々な産業分野での業務改善に関連した調査や検討は未だ十分でない。

本研究は、職場で AI 利用者が AI の提案を受容するために、どのような要素が影響しているのかを明らかにすることを目的とする。図 1 に示すように、AI による提案を説明する「説明者」と、その説明を受ける「AI 利用者」を定義する。また、AI 利用者が AI の提案を受容するか否かは、AI の提案によって生じる利益や損害に依存することが知られているため、職場場面に限定して聴取した [7]。

本稿の構成は以下の通りである。まず、第 2 章では関連研究を示し、第 3 章では研究目的及び仮説について述べる。次に第 4 章では質問紙の構成、第 5 章では調査方法について示す。さらに、第 6 章では統計検定について示し、第 7 章で機械学習モデル構築による分析について述べる。最後に第 8 章にてまとめを行う。

2. 関連研究

様々な産業分野において AI 技術を利用し業務改善を試みた事例が増加し、人間の意思決定を支援するようになっている。このように実社会で AI を安全に利用するためには、AI 技術に対する信頼を構築することは必要不可欠である。信頼できる AI の実現に向けて、ガイドラインの開発 [4] や、XAI 技術、Human Computer Interaction (HCI) の研究に注目が集まっている。

XAI 技術は、AI の大部分のモデルがブラックボックスであるという課題に着目し、AI から得た結果を多種多様な形式で可視化することを可能にした。代表的な手法として LIME[9] や SHAP[10] が挙げられる。これらの手法を用いると、どのような種類のモデルであっても重要度を可視化できるため、広く用いられている。重要度とは、AI から得た結果に影響を与えた度合いのことを示し、特徴量ごとに算出される。重要度が大きい特徴量ほど導出結果に影響を与えた特徴量と判断できる。他の可視化形式として、モデルの判断ルールを可視化する手法 [12]、AI の提案の効果を仮想的に検証し、AI から得た結果の妥当性を示す反実仮想を用いた手法 [13]、AI から得た結果と類似するデータを示すこと説明する手法 [11] がある。

しかしながら、これらの XAI 技術は手法ごとに独立に評価されており、個々の事例でどの技術が有用かわからな

表 1 質問紙の構成

項目	質問内容
基本属性	性別、年齢、出身地、未婚婚、子供の有無、世帯年収、最終学歴、最終学歴での専攻、業界、職種、雇用形態、役職、勤続年数
性格診断	BigFive による性格診断 [17]
AI への考え方	AI に対する知識の有無、業務で AI 活用の程度、AI に対する認識、AI への情報提供意思の有無
信頼への考え方	一般的信頼尺度 [14]、対人信頼尺度 [18]、上司や同僚への信頼 [15]、職務肯定感 [16]、企業への信頼 [19]
AI の提案の受容	場面 1: 営業部長によるお客様リストの提案、 場面 2: AI によるお客様リストの提案

いという課題がある。その課題に対し HCI の文脈においては、AI の提案を利用する事例を設定したユーザ実験により検証している。先行研究 [5][6] では、事例に用いた AI の提案を説明するために複数の XAI 技術によつ可視化を行い、ユーザにどの XAI 手法による可視化がモデルの理解や信頼に効果的だったかを評価している。このような HCI における研究では、AI の提案を説明する役割をコンピュータが担っている研究が大部分を占める。そのため、AI の提案を人が提示し、説明する場合の信頼について検討が不十分である。

人の信頼についての研究は、主に人文及び社会科学の分野にて取り組まれてきた。山岸 [14] は、信頼の概念を『相手の行動によって自分の「身」が危険にさらされる状態で、相手がそのような行動をとらないだろうと期待すること』と定義した。具体的には、相手の人間性や行動特性の評価に基づく相手の意図に対する期待としての信頼に、相手にとっての損得感情に基づく相手の行動に対する期待としての信頼を含めることで、信頼の概念に安心の考え方を導入した。この定義に沿って山岸が作成した一般的信頼尺度は広く用いられている。また、山岸の信頼の概念を踏まえ、吉田 [7] が部下の人事評価の納得感と上司との信頼関係の関係性について調査及び分析を行った。その結果、部下の人事評価への納得感には、直属の上司への信頼感や人事評価そのものの公平感が影響を与えていることを明らかにしている。先行研究 [15][16] では、職場場面での信頼感、労働者の職場の上司や同僚における対人的信頼感や、労働者自身の職務肯定感が関連すると報告されている。これらの先行研究より、結果や提案への納得及び受容には、結果や提案を説明する上司との信頼関係が重要であるとわかっている。

3. 研究目的とアプローチ

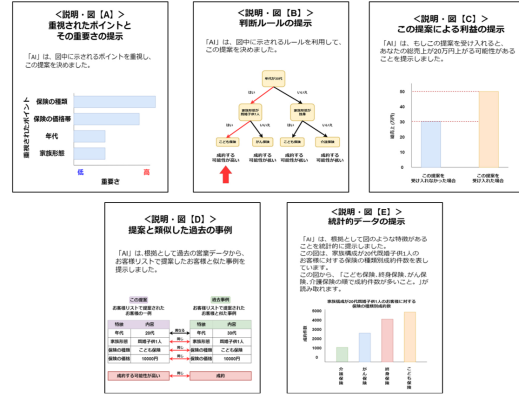
本研究では、職場で AI 利用者が AI の提案を受容するために、どのような要素が影響しているのかを明らかにすることを目的とし、以下の仮説を立てた。

- 仮説 1: AI 利用者の提案に対する受容しやすさは、AI の提案と人の提案の間で差がある。

あなたは個人向け生命保険の訪問営業をしています。
ある日、会社から、【AIが作ったお客様リスト】に従って、営業してほしいと提案されました。すでに会社からは以下のことが発表されており、あなたたち従業員もそれに納得しています。
・AIが作成したお客様リストの有効性は確認されています。また、あなたの給料は、成約件数と成約金額によって変動します。
・AIはこれまでの営業データに基づいてお客様リストを作成し、そのリストの精度は十分に高いことが確認されています。
・AIが作成したお客様リストに、人間的な操作は介入していません。
これから、会社からあなたに対してお客様リストの妥当性について説明をします。
説明を受ける立場になって以下の質問に答えてください。



(a) 説明文



(b) XAI模式図

図 2 場面 2: AI によるお客様リストの提案で使った説明文と図

- 仮説 2: 人を信頼しやすい AI 利用者は、人を信頼しづらい AI 利用者よりも、AI による提案を受容しやすい。
- 仮説 3: AI にポジティブな認識を持つ AI 利用者は、AI にネガティブな認識を持つ AI 利用者よりも AI による提案を受容しやすい。

これらの仮説の検証のため、まず、信頼の概念や AI に対する考え方について聴取する質問や、eXplainable AI (XAI) 手法を模した図を用いて AI の提案について説明する場面を含む質問紙を作成した。作成した質問紙を用いてオンラインで営業職約 7,000 人に対しアンケート調査を実施した。次に、収集したデータに対して統計検定を実施した上で、全ての質問を入力として AI を受容するかを予測する 2 種類の二値分類モデルを作成した。モデルの作成には XGBoost[8] を使用した。特徴量重要度の高い質問項目から、AI の提案を受容するのに影響を与える要素を特定した。最後に、特徴量重要度の結果を踏まえて、特徴量選択を実施し少数の質問でモデルを作成した。

本研究の結果より、説明者が AI による提案を説明する前に、AI 利用者がどの程度受容するかを予測可能となり、AI 受容度が高い AI 利用者から説明を行うことで、AI の導入に関する稼働削減の効果が期待できる。また、AI 受容度の高い利用者が多数所属する部署から AI を用いた施策を導入することが可能となり、AI を用いた施策の効果を効率的に見積もることが可能になる。

4. 質問紙の構成

3 章に示した仮説の検証のため、表 1 に示す基本属性、性格診断、AI への考え方、信頼の概念への考え方、場面想定法を用いた AI の提案の受容に関する質問による 5 項目で構成される質問紙を作成した。調査項目の性格診断や信頼への考え方、AI に対する考え方の質問の選択肢には「1. あてはまる」から「5. あてはまらない」までの 5 点のリッカート尺度を用いた。

図 2 に、AI の提案の受容に関する質問で使った場面 2 の説明文と図を示す。本研究では、筆者らが過去に取り組んだ案件を参考に、お客様リストの提案の事例を場面として設定した。仮説 1 の検証のため、お客様リストの提案者が営業部長である場面（場面 1）、お客様リストの提案者が AI である場面（場面 2）の 2 種類を設定した。場面 1 の説明文と図は、提案者を営業部長に置換しただけであり、他の説明文や図は図 2 に示した場面 2 と同様のものを使用した。また、これらの場面で、図 2 の (b) に示した各 XAI 手法の可視化を模した図である XAI 模式図を作成し、提案を説明した。XAI 模式図として、特徴量重要度、ルール可視化、反実仮想、類似事例、データ可視化の代表的な XAI 手法である 5 種類を採用した。そして、AI 技術に詳しくない回答者が理解しやすいよう、質問紙では、以下に示す名称をそれぞれ利用した。回答の変化が出ないよう、場面 1・場面 2 ともに同じ XAI 模式図を使用した。

- 特徴量重要度: 重視されたポイントとその重要さの提示
- ルール可視化: 判断ルールの提示
- 反実仮想: この提案による利益の提示
- 類似事例: 提案と類似した過去の事例
- データ可視化: 統計的データの提示

場面それぞれの各 XAI 模式図による説明に対して、XAI に提案された内容を理解できるか、受容できるかについて、以下の 5 段階で聴取した。場面は、営業部長による提案、AI による提案の順に提示した。

- (1) 理解できて、提案を受け入れられる
- (2) 理解はできたが、提案を受け入れられない
- (3) 理解できないが、提案を受け入れられる
- (4) 一部、理解できない箇所があり提案を受け入れられない
- (5) 全く理解できないし、提案を受け入れられない

5. 調査方法

職場における AI による提案の受容に関する質問紙を作成し、オンラインにて営業職約 7,000 人に対しアンケート調査を実施した。本研究で実施したオンラインによるアンケート調査は、モニター登録されている被験者へ回答を依頼し、ウェブを通じて回答を収集したため、無作為抽出調査ではない。調査期間は、2021 年 11 月 17 日（水）から 11 月 25 日（木）の 9 日間であった。

調査対象者は、最終学歴が中学卒業以上の日本語を母国語とする 20 歳から 69 歳までの営業職とした。日本における営業職の性年代構成比に合わせるため、平成 27 年国勢調査より算出した就業者の性年代構成比と営業職の市場構成比を掛け合わせ、男女別各年代に必要なサンプル数を算出した。なお、就業者中の営業職割合は調査会社の過去調査を参考にし、オンラインによるアンケート調査を通じて 7,216 件の回答を収集した。調査対象者の平均年齢は 43.5 歳で、男女比は男性が 5,690 人（78.9%）で女性が 1,512 人（21.1%）であった。雇用形態は、経営者役員が 882 人（12.2%）、正社員が 4,732 人（65.6%）、派遣及び契約社員・アルバイト・パートが 1,434 人（19.9%）、その他が 111 人（1.5%）であった。

6. 統計検定: 仮説 1 の検証

仮説 1 の検証のため、統計検定を用いた分析を実施した。各場面において各 XAI 模式図が理解できるか、提案を受け入れられるかを聴取した質問の回答に対して、「提案を受け入れられる」を含む選択肢を 1（受容）、それ以外の選択肢を 0（非受容）として集計した。その後、営業部長による提案と AI による提案によって受容度合いの間に差があるかを検証するため、ノンパラメトリック検定の一種である対応ありのウィルコクソンの符号付順位検定を用いて両側検定を実施した。本研究では、p 値が 0.05 以上であると有意差があると判定した。検定の結果を表 2 に示す。検定の結果、ルール可視化のみに有意差があり、他の XAI 模式図には無かった。よって、仮説 1「AI 利用者の提案に対する受容しやすさは、AI と人では差がある。」に反し、AI と人が同じ形式の可視化を使用する場合においては、提案者が AI・人どちらであっても受容しやすさは大きく変わらないことが判明した。この結果を受けて、我々は説明内容さえ工夫すれば、AI 利用者が AI の提案を受容するのではないかと考え、AI の提案の受容に影響する要素を明らかにすべく仮説 2 及び仮説 3 の検証を実施した。

7. 機械学習モデル構築による分析

機械学習のアルゴリズムである XGBoost を用いて、AI の提案を受容するか否かを二値分類するモデルを構築し

表 2 AI による提案と営業部長による提案の受容度合いの差の比較

XAI 模式図	営業部長		p 値
	平均値 (標準偏差)	AI 平均値 (標準偏差)	
特徴量重要度	0.554 (0.497)	0.493 (0.500)	0.019
ルール可視化	0.568 (0.495)	0.480 (0.499)	0.131*
反実仮想	0.562 (0.496)	0.500 (0.500)	0.003
類似事例	0.579 (0.497)	0.501 (0.499)	0.004
データ可視化	0.583 (0.493)	0.510 (0.499)	0.002

* $p > 0.05$

た。XGBoost は、勾配ブースティングを用いたアンサンブル学習と決定木を組み合わせた手法で、高い汎化性能を持つ手法である。また特徴量重要度より、予測に影響を与えた特徴量を明らかにすることができる。

7.1 モデルの構築方法と使用した特徴量

まず、モデルの構築方法について述べる。本研究では、各 XAI 模式図を受容するかを予測するモデル（モデル 1）及び AI の提案を受容するかを予測するモデル（モデル 2）の 2 種類のモデルを用いて精度を検証した。図 3 に 2 種類のモデル構築方法を示す。モデル 1 では、各図に対する理解と受容に関する質問の「提案を受け入れられる」を含む選択肢である 1 と 3 を選んだ回答者に受容ラベルを付与し、二値分類を行った。本研究では、XAI 模式図を 5 種類用意したため、5 種類のモデル 1 が構築される。モデル 2 では、モデル 1 で付与したラベルを集約し、XAI 模式図のいずれかを受容していた場合に受容ラベルを付与し、二値分類を行った。5 種類の XAI 模式図の回答を集約するため、1 種類のモデル 2 が構築される。

次に、学習に使用した特徴量について述べる。短縮版 BigFive による性格診断 [17] で得た回答に関しては、先行研究 [17] で提案されている集計方法に準拠し、人の性格を表す 5 つの要素（外向性、情緒安定性（情緒不安定性）、誠実性、協調性（調和性）、開放性）の値を算出した。また、尺度を用いた質問（一般的信頼尺度 [14]、対人信頼尺度 [18]、上司や同僚への信頼 [15]、職務肯定感 [16]、企業への信頼 [19]）は、回答者ごとにリッカート尺度による回答の合計値を算出した。合計値に基づきサンプル全体の中央値を算出し、中央値以上の回答者にフラグをつけた。その他の質問は回答結果をそのまま特徴量として用いた。

7.2 使用したデータとハイパーパラメータ

表 3 に実験に用いた各データ件数と受容ラベル件数及び割合を示す。収集したデータを学習データ:検証データ:評価データ=6:1:3 に分割して実験を行った。モデル 1 では、どの図も受容ラベルと非受容ラベルが同程度の件数であるが、モデル 2 では受容ラベルが 8 割を占める不均衡データであった。また、施行ごとに乱数シードを変更し、学習・検証・評価データに割り当てられるサンプルを

表 3 場面 2 の AI によるお客様リストにおけるデータの内訳（括弧内は受容ラベル件数）

	XAI 模式図の種類	学習データ	検証データ	テストデータ	受容ラベルの割合
モデル 1	特徴量重要度	4,545 件 (2,520 件)	506 件 (268 件)	2,165 件 (1,200 件)	平均 55%
	ルール可視化	4,545 件 (2,582 件)	506 件 (268 件)	2,165 件 (1,230 件)	平均 57%
	反実仮想	4,545 件 (2,552 件)	506 件 (284 件)	2,165 件 (1,216 件)	平均 56%
	類似事例	4,545 件 (2,630 件)	506 件 (293 件)	2,165 件 (1,253 件)	平均 57%
	データ可視化	4,545 件 (2,650 件)	506 件 (295 件)	2,165 件 (1,263 件)	平均 58%
モデル 2	-	4,545 件 (3,463 件)	506 件 (384 件)	2,165 件 (1,644 件)	平均 76%

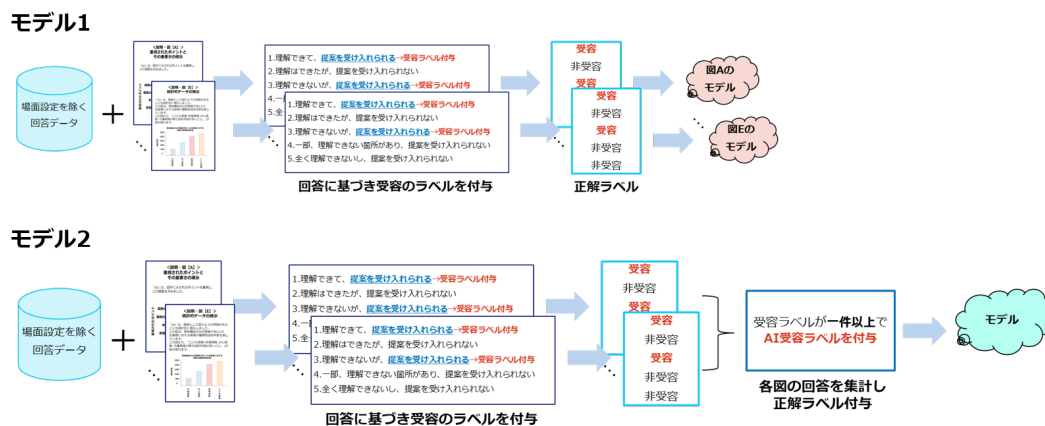


図 3 モデル 1 及びモデル 2 の構築方法

表 4 実験 1 の結果

	XAI 模式図の種類	AUC	正解率	適合率	再現率
モデル 1	特徴量重要度	0.703	0.652	0.642	0.650
	ルール可視化	0.743	0.681	0.674	0.695
	反実仮想	0.716	0.660	0.657	0.668
	例示	0.726	0.669	0.676	0.710
	データ可視化	0.720	0.665	0.683	0.712
モデル 2	-	0.737	0.716	0.746	0.863

表 5 実験 1 特徴量重要度上位 10 項目

順位	質問	重要度平均
1	年齢	6,154
2	世帯年収	2,130
3	勤続年数	1,368
4	性別	1,299
5	最終学歴	1,153
6	自己価値合計値	1,057
7	個人年収	1,034
8	対人信頼尺度合計値	1,032
9	AI に対する知識の有無	972
10	役職	919

変更しながらホールドアウト検証を行った。XGBoost の損失関数は負の対数尤度とし、ハイパーパラメータはグリッドサーチを行い、モデル 1 では $\eta=0.1$, $\gamma=0$, $\max_depth=3$, $\min_child_weight=1$, $scale_pos_weight=1$ とし、モデル 2 では、 $\eta=0.2$, $\gamma=0$, $\max_depth=6$, $\min_child_weight=1$, $scale_pos_weight=1$ とした。

7.3 評価設定

各モデルが出力する分類確率が 0.5 以上であれば「AI の提案を受容する」、そうでなければ「AI の提案を受容しない」と判定した。評価指標は、モデル 1, 2 ともに ROC-AUC (AUC), 正解率, 受容ラベルの適合率, 再現率とした。各評価指標は 10 回施行平均で算出した。先行研究 [20] より、AUC が 0.7 以上であれば有効とした。

7.4 実験設定

本研究では 2 通りの実験で評価を行い、仮説を検証した。

7.4.1 実験 1

実験 1 は仮説 2 の検証のために実施した。全ての特徴量を利用しモデル 1 及びモデル 2 を構築、精度を評価し、各モデルから得られた特徴量重要度を分析した。これにより、AI の提案の受容に影響する要素を検証した。

7.4.2 実験 2

実験 2 は仮説 3 の検証のために実施した。まず、表 6, 7 の特徴量重要度の結果より、AI へのポジティブな認識についての質問がモデルの改善に寄与していると仮定した。次に、質問紙に含まれる質問を以下の 3 カテゴリに分け、各カテゴリの質問の回答を入力としてモデルを構築し精度を比較した。最後に、最も精度が高かったカテゴリの質問の回答に対して相関を用いた特徴量選択を実施し、少数の質問で高精度に AI の提案の受容を予測できるか検証した。また、実験 1 のモデル 1 よりもモデル 2 のほうが精度が高

かったという結果を踏まえて、実験2ではモデル2を使用した。

- (1) AIへのポジティブな認識についての質問
- (2) AIへのネガティブな認識についての質問
- (3) AIに無関係な質問

7.5 実験結果と考察

7.5.1 実験1: 仮説2の検証

表4に実験1によって得られたモデル1及びモデル2の精度を示す。モデル1では全ての図のモデルにおいて、AUCが0.7以上となり、AIの提案を予測するのに有効なモデルであることがわかった。最もAUCが高かったのはルール可視化のモデルであり、0.743を示した。また、正解率や適合率については、どのXAI模式図のモデルも0.65程度を示し、AIを受容する人を正確に判別できないことがわかった。再現率についても同様であり、AIの提案を受容する人を見逃す傾向を持つモデルであることがわかった。

また、モデル2ではAUCが0.737であり、モデル1と同様にAIの提案を予測するのに有効なモデルであることがわかった。また、正解率は0.716、適合率は0.746となり、モデル1と比較して高くなった。そして、再現率は0.863となり、正解ラベルの作り方を変更することにより、AIの提案を受容する人を見逃す傾向が改善された。

構築した全てのモデルの特徴量重要度を確認したが、上位10項目が同一であったため、一例として表5にモデル2の特徴量重要度上位10項目を示す。これによると、年齢や世帯年収、勤続年数、性別、最終学歴、個人年収や役職といった個人の属性に関する特徴量が上位に入っていた。また、対人信頼尺度合計値といった他人に対する信頼しやすさの程度や、自己価値合計値、AIに対する知識の有無がAIの提案の受容に強く影響することがわかった。対人信頼尺度の合計値が上位に入っていることから、仮説2の「人を信頼しやすいAI利用者は、人を信頼しづらいAI利用者よりも、AIによる提案を受容しやすい。」は正しいことがわかった。

さらに、表6, 7に、モデル2から算出したAIへのポジティブな認識についての質問及びAIへのネガティブな認識についての質問の特徴量重要度上位5件を示す。これによると、AIへのポジティブな認識について問う質問のほうが、ネガティブな認識について問う質問よりも順位が高く、平均重要度が高いことがわかった。

7.5.2 実験2: 仮説3の検証

実験1の結果を受けて、我々はAIへのポジティブな認識についての質問がモデルの改善に寄与していると仮定した。この仮定が正しいかを確かめるため、質問紙に含まれる質問を以下の3カテゴリに分けた。

- (1) AIへのポジティブな認識についての質問
- (2) AIへのネガティブな認識についての質問

表6 AIへのポジティブな認識についての質問
特徴量重要度 上位5項目

順位	質問	重要度平均
9	AIに対する知識の有無	972
13	AIに対して費用を抑えられるという認識を持っているか	758
15	自身の業務にAIを導入したいという認識を持っているか	710
19	生活支援AIに秘匿性の高い個人情報を提供するか	656
20	AIに対して多くの情報を持つという認識を持っているか	647

表7 AIへのネガティブな認識についての質問
特徴量重要度 上位5項目

順位	質問	重要度平均
21	AIによって、仕事がなくなるという認識を持っているか	646
29	AIに対して何となく怖いという認識を持っているか	556
32	AIに対して不安であるという認識を持っているか	535
126	AIから得られる情報に対して不安に思っているか	250
128	AIにより人間の能力が退化するという認識を持っているか	249

表8 実験2の結果

質問カテゴリ	質問数	AUC	正解率	適合率	再現率
全質問	342	0.737	0.752	0.823	0.859
AIポジティブ認識	52	0.732	0.735	0.846	0.796
AIネガティブ認識	23	0.593	0.606	0.788	0.658
AIと無関係	263	0.622	0.716	0.793	0.846

(3) AIに無関係な質問

これらのカテゴリに対応する質問を入力としてモデルを構築し、評価した。各カテゴリの質問数と評価指標の精度を表8に示す。全質問を使用した場合は、AUCが0.732であった。また、AIへのポジティブな認識についての質問を使用した場合は、質問数が52件と全質問の約15%であるにも関わらずAUCが0.732であり、全質問を使用したモデルと同程度の精度であった。また、AIへのネガティブな認識についての質問を使用した場合は、AUCが0.593、AIとは無関係の質問を使用した場合は0.622となり、全質問を使用した場合よりAUCが低くなった。以上より、AIの提案の受容を予測することに、AIへのポジティブな認識についての質問が有効であることがわかった。

これら結果より、AIへのポジティブな認識についての質問52項目に対し、特徴量間の相関係数が0.7以下になるように特徴量選択を実施した。その結果、「自身の業務にAIを導入したいか」、「社内のAIに個人情報を共有するかどうか」、「AIに対して、既存の仕事の質を向上させる認識を持っているか」の3問が残った。この3問を入力としてモデルを構築し評価した結果、AUCが0.714、適合率0.656、再現率が0.854で予測できた。以上より、仮説3の「AIにポジティブな認識を持つAI利用者は、AIにネガティブな認識を持つAI利用者よりもAIによる提案を受容しやすい。」は正しいことがわかった。また、精度を維持しつつ質問数を削減でき、簡易にAI利用者のAIの提案の受容しやすさを推定できるようになった。

8. まとめ

本研究ではAIによる提案を説明する「説明者」と、その説明を受ける「AI利用者」を定義し、職場においてAI利用者がAIの提案を受容するために、どのような要素が影響しているのかを明らかにした。信頼の概念やAIに対する考え方を聴取する質問や、XAI手法模式図を用いてAIの提案を説明する場面を含む質問紙を作成した。作成した質問紙を用いて、オンラインにてアンケート調査した。回答結果に対して統計検定を行い、XAI手法による可視化の形式によってAIによる提案の受容しやすさは変化しないことがわかった。また、XGBoostを用いてAIの提案を受容するか否かを予測する二値分類モデルを作成し、AUC=0.7程度で予測可能であることがわかった。モデルから出力された特徴量重要度から、年齢や世帯年収、勤続年数といった個人の属性を表す項目の他に、人への信頼しやすさがAIの提案の受容に影響することがわかった。さらに、AIへのポジティブな認識を問う質問3問を入力としたモデルにおいてAUCが0.73の精度で予測できた。これによりAIにポジティブな認識を持つAI利用者は、AIにネガティブな認識を持つAI利用者よりもAIによる提案を受容しやすかった。また、精度を維持しつつモデルに入力する質問を削減できたため、簡易にAI利用者がAIの提案を受容するか否かを予測可能になった。

今後は、AI利用者にあったAIの提案の説明を推薦できるように質問紙の構成から再検討を行い、それにあったアルゴリズムを提案したい。

参考文献

- [1] ポストンコンサルティンググループ, 企業の人工知能(AI)の導入状況に関する各国調査, 2018.
- [2] 独立行政法人情報処理通信機構, 平成29年度AI社会実装推進調査報告書, 2018.
- [3] Adadi, A. & Berrada, M. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access* 6, 52138–52160, 2018.
- [4] EU, Ethics Guidelines for Trustworthy AI, 2019.
- [5] Ashoori, M. & Weisz, J. D. In AI we trust? Factors that influence trustworthiness of AI-infused decision-making processes. *ArXiv*, 2019
- [6] Wang, X. & Yin, M. Are explanations helpful? A comparative study of the effects of explanations in AI-assisted decision-making. In *Proceedings of 26th International Conference on Intelligent User Interfaces*, 2021.
- [7] 吉田朋子, 上司との対人関係が人事評価の納得感に及ぼす影響, *産業・組織心理学研究* 30, 29–43, 2016.
- [8] Chen, T. & Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016.
- [9] Ribeiro, M. T., Singh, S. & Guestrin, C. ‘Why Should I Trust You?’ : Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data*

- Mining, 2016.
- [10] Lundberg, S. & Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017.
- [11] Hanawa, K., Yokoi, S., Hara, S. & Inui, K. Evaluation of Similarity-based Explanations. In *Proceedings of The Ninth International Conference on Learning Representations*, 2021.
- [12] Friedman, J. H. & Popescu, B. E. Predictive learning via rule ensembles, *The Annals of Applied Statistics*. JSTOR, 916–54, 2008
- [13] Mothilal, R. K., Sharma, A. & Tan, C. Explaining machine learning classifiers through diverse counterfactual explanations. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020.
- [14] 山岸俊男, 信頼の構造: ことごと社会の進化ゲーム, 東京大学出版会, (1998) .
- [15] 西村昭徳, 松尾直博, 職場場面における信頼感の構造についての検討, *東京学芸大学紀要*, 第1部門, 教育科学 53, 57–64, 2002.
- [16] 山田一子, 職場における対人信頼感および職務肯定感とストレス反応との関連: 職場の重要な他者や職務との関係性に対する感情に着目して, *東京学芸大学学校教育学研究論集* (30): 49-60 2014.
- [17] 並川努, 谷伊織, 脇田貴文, 熊谷龍一, 中根愛, 野口裕之, Big Five 尺度短縮版, *心理学研究*, 83 巻, 2 号, p. 91-99, 2021.
- [18] 堀井俊章, 対人信頼感尺度の作成および信頼性と妥当性の検討に関する研究, *日本性格心理学会発表論文集* 2, 34, 1993.
- [19] 一小路武安, 新技術受容性の高い個人とは: 革新性を中心とする個人属性と個人の組織との適合性の観点から, *組織科学* 47, 53–68, 2013.
- [20] Hosmer, D. W., *Applied logistic regression*, Wiley, 2013.