

2022.8 ブログ：『A I 判断の根拠を説明するX A I』を読んで、の詳細
(→ <http://www.1968start.com/M/blog/index2.html#2208>)

『産業利用における説明可能A I の使いどころ』を読んで

中所武司

■この本の読書のきっかけ

情報処理学会誌の最新号に掲載されていた下記の解説に興味を持った。

特集：AI 判断の根拠を説明する XAI を使いこなす：

「2. 産業利用における説明可能 AI の使いどころ」情報処理, 63 (8), e8-e13 (2022-07-15)

■記事内容の要約とコメント (→★)

【説明可能A I の実現要件】

- ・X A I の代表的な技術 LIME (2016 年) 以降、多くの提案があったが、A I の説明性の課題が解決されてはいない。
- ・求められる要件と技術のギャップは小さくない。要件に応じた技術選定の方法論が必要。

図1 (説明の実現) の説明要件リスト：

- * 誰に説明するか？
- * どんな説明表現が必要か？
- * データ形式は？
- * システムとして考慮すべきことは？

→★そもそも説明要求には無理があるのでは？ (ないものねだり？)

通常のアプリケーションプログラムでも、バグがない保証はしていない。

開発側のシステムテストや発注側の受け入れテスト実施後に実用化されている。

→★第一次A I ブーム (1960 年代) のとき、手書き文字認識の研究が盛んだったが、

人間に読めない文字がシステムで読めるのはおかしくないかという話があった。

「あ」か「お」かの判断不可の文字を「あ」と判定した理由を説明されても・・・

【A I の説明方式】

ホワイトボックスとブラックボックス

- ・A I モデルには、以下の2種類あり：
 - * ホワイトボックスモデル：説明力は強いが、精度は劣る。→線形回帰モデル、決定木モデル
 - * ブラックボックスモデル：説明力は弱いが、精度あり。
 - アンサンブルモデル (ランダムフォレスト、ブースティング)、深層学習
- ・ホワイトボックスで十分な精度が得られないときは、ブラックボックス利用となる。
ブラックボックスモデルには2種類の説明方法がある：
 - * 大局的説明：A I モデル自体を説明
 - * 局所的説明：個々の予測結果を説明

大局的説明

- ・全体としてどの特徴量が重要かの計算や、解釈可能なモデルでの表現がある。
- ・特徴量に注目し、重要度を算出する方式は簡単でよく利用される。
A I モデルに依存しないで算出でき、木構造モデルの内部構造で算出できる場合あり。
さらに、特徴量の値と予測値の平均的な関係を表現するものもある。
- ・複雑なモデルをホワイトボックスモデルで置き換えた代理モデルで説明する方式あり。

局所的説明

- ・特定の入力に対するブラックボックスモデルの予測結果の根拠を提示する方法。
予測に寄与した特徴量を算出する方式が代表的なもの。
説明対象モデルに依存しない技術や、深層学習モデルがある。
- ・データに基づく方式では、予測に寄与した学習データ、予測根拠が類似するデータ、予測結果が変わるデータなどを用いて説明する。

【データ形式とX A I】

- ・この章では、表形式データ、テキストデータ、画像データを取り上げる。

表形式データの説明

- ・特徴量の一つひとつが意味をもち、大局的説明も局所的説明も可能。
精度向上のため、特徴量同士の比の変数や交互作用項を追加した場合、
人間にとっての解釈性が問題となることがあり、注意が必要。

→★具体例での説明がほしい。

特徴の種類がリスト化されているようだが、それらの値を入力するということ？

- ・スパースデータ（ゼロ要素が多い高次元データ）では、非ゼロ要素のみで説明すべき。
（例）購買履歴データからの顧客分類では、購入商品だけでの説明が直感的となる。

→★当たり前の話では？ このアプリのどこがA Iなの？

テキストデータの説明

- ・単語が特徴量となり、意味をもつので、大局的説明も局所的説明も可能。
テキストデータでは、文脈が重要な意味をもつので、説明の可視化では、
もとのテキストデータにハイライトを重ねるべき。

→★具体例での説明がほしい。

単語の出現頻度が特徴量なら、それを入力とするA Iアプリとはどのようなもの？

画像データの説明

- ・画像データでは、入力は各画素（画像上の座標）のR G B値で、
画像ごとに対象物の写る位置が異なるので、画素単位の大局的説明は不可。

そこで、画像のどの領域に判断根拠があるかを示す局所的説明を用いる。

＊画像処理によって輪郭を抽出したスーパーピクセルを用いる方式

＊単純に格子状に画像を分割する方式

- ・画像データに相性の良い深層学習では、その内部構造を活用した手法が用いられる。
- ・音声データやセンサデータでは、時系列上の区間を特徴量としてハイライトする方式

→★具体例での説明がほしい。

たとえば、猫を認識するA Iでは、「猫」と「猫でない」の判断根拠として、画像の一部分の領域を提示するようだが、多層のニューラルネットの出力層から入力層への逆トレースができるということ？

- ・どのデータ形式においても、データで説明する方式は有効な手段となる。

→★具体例での説明がないのでよくわからない。

【A Iシステムから見たX A I】

- ・X A Iの観点から、A Iシステム構築における留意点を述べる。

・X A IのA Iモデルへの依存性

クラウドサービスのAPI使用のA Iモデルでは、モデル依存のA I使用に制限あり。
モデルに依存しないサービスでは、モデルの構造を利用した効率的な計算法使用不可。
説明したいモデルと説明方法が、システムとして整合性が取れるかは検討事項。

・推論のパイプライン構築

図2②のように、前処理や特徴量エンジニアリングを含めて、A Iとして、X A Iに入力する際には、推論パイプラインとして扱いやすい設計が必要のため、説明要件は早期に検討して、実装の手戻りを防ぐべき。

→★従来のソフトウェア生産技術と同様。要求定義の手抜き厳禁。

・システムの応答時間

計算コストの高いX A I方式は、リアルタイム性が求められる場面での利用不可。

・画面・U I／U X

X A Iを必要とするシステムでは、画面の検討・開発の工数大。
視覚的情報が重要なので、ユースケースに合わせて工夫が必要。

→★従来のソフトウェア生産技術と同様。要求定義段階で具体化すべし。

【誰にとっての説明か】

- ・A Iシステムの周辺に登場する、次の5つのペルソナ（典型的なユーザ像：図3）のうち、誰に対する説明が必要かの要件定義をして、技術選定すべき。

＜A I開発者＞

＊A I実装者

＊A I開発責任者

< A I 利用者 >

- * 業務担当者（有識者）
- * 業務担当者（非有識者）
- * エンドユーザ

・ A I 実装者：

モデル改善に、A I の振る舞いを知る必要があるので、X A I がヒントになる。

- * 特徴量の重要度の算出結果が、業務知識と照らして違和感がないか、確認できる。
違和感があれば、データの取得方法や加工方法を疑うことができる。
- * 誤った予測データに対して、局所的説明により誤り分析ができる。
- * A I の苦手なデータの特徴を分析し、データ加工やモデリングの見直しができる。

・ A I 開発責任者：

A I 構築方針で、極めて高い説明性要求には、ホワイトボックスモデルの使用を検討。

リリース判断のための実験・精度評価レポートに大局的説明や局所的説明を利用。

公平性の要求がある場合、たとえば性別が予測値に影響しすぎているか確認できる。

なお、X A I を用いず、多様なテストケースを設計すれば、十分な精度確認ができるので、リリース判断に足る説明責任については、A I 品質管理技術も援用すべき。

→★従来のソフトウェア生産技術と同様。品質はテストで判定する。

・ 業務担当者（有識者）

たとえば、融資審査A I を利用する融資担当者は、業務上での個々のデータの予測の根拠を知りたいので、局所的説明を選択すべき。

業務担当者に十分な知識や経験があれば、端的な説明で十分。予測に寄与した特徴量を算出する方式で、重要項目のハイライトだけで予測の根拠を推測できると期待できる。
同様に、類似データに基づく方式も有効。

・ 業務担当者（非有識者）

融資審査A I を利用する新規配属の融資担当者は、重要項目のハイライトでは理解できない場合があるので、別の方式が必要。

- * 過去の類似事例を幾つか提示する方法がある。（ある種の前例主義）
- * 予測に寄与した特徴量に関する知識の提示も有効。ある財務指標の意味を提示し、その指標の高低と貸し倒れリスクの関係を記載した文献提示で、理解支援になる。
この知識は、検索技術でも知識グラフ利用でもよい。

→★第2次A I ブームのときの知識工学（プロダクションルールなど）に類似？
専門家の知識をルール化しておくなど。

・ エンドユーザ

融資審査の例では、借入申請する顧客。

局所的説明技術の活用が必要だが、提供する情報の粒度に留意し、

特徴量を適切な単位でまとめ、可読性を増し、納得感を醸成する。
予測結果を変えるヒントを欲する場合、
特徴量に含まれていた「業種」の寄与で融資不可と判定された場合は、
業種の変更は難しいので、制御可能な変数のみで説明することが望ましい。
たとえば、口座残高が少ないという説明なら、顧客は残高を増やす行動を採れる。
さらに、反実仮想を用いた説明で、口座残高の必要額がわかれば有益な情報となる。
X A I の出力を提示せず、慣例的に説明してきたパターンに当てはめることも重要。
たとえば、あたかも担当者から説明されるように、自然文で出力することで
システムへの安心感を高める。テンプレート文にX A I 結果をあわせるだけでもよい。

- ・以上は、あくまで代表例であり、ユースケースによっては、その限りではない。
重要なことは、誰にとって納得できる説明かを定義し、X A I を活用すべき。

→★融資審査A I を例にした説明が詳しいが、X A I の例として適切かは疑問。
この分野は、十分な量の専門家のノウハウが存在すると思うので、
第2次A I ブームのときの知識工学（エキスパートシステム）でいいのでは？

【X A I の限界と期待】

- ・現在のX A I は、A I の計算過程や予測の根拠を完全に説明するものではない。
そもそも複雑な処理を簡潔かつ完全に説明することは不可能である。
人間に理解容易な出力が可能とは限らない。提示される情報の解釈は人間の仕事である。
- ・X A I は適切に使えば有用なので、説明要件に整合した技術選定・技術開発が重要。
そのための勘所・方法論の確立が期待される。

→★X A I の特集号を出すには時期尚早だったということですね。

（参考：関連する過去ブログ）

▼2019.1 「A I システム検証へのニューロンカバレッジの有用性について」

<http://www.1968start.com/M/blog/index.html#1901>

以上