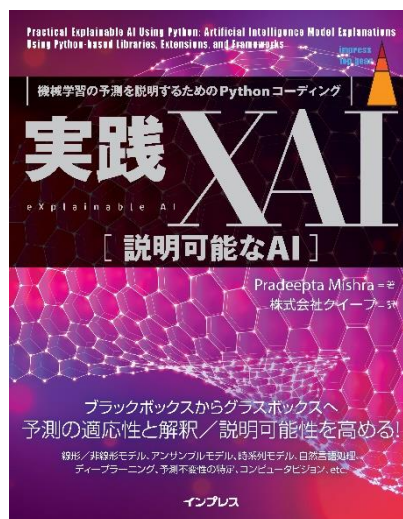


各 位

2023年6月20日
株式会社インプレス

予測を解釈・説明するための手法を網羅的に把握できる！
『実践 XAI [説明可能な AI] 機械学習の予測を説明するための Python コーディング』
を 2023年6月20日（火）に発売
発刊記念キャンペーンとして書籍の冒頭から第3章までを無料公開

インプレスグループで IT 関連メディア事業を展開する株式会社インプレス（本社：東京都千代田区、代表取締役社長：小川 亨）は、機械学習の結果を意味づける手法を解説した新刊『実践 XAI [説明可能な AI] 機械学習の予測を説明するための Python コーディング』を 2023年6月20日（火）に発売いたします。発売を記念して書籍の一部無料公開も実施いたします。



■各種モデルのブラックボックスをガラスボックスに！

急速な拡大を見せる ChatGPT サービスや、政府の有識者会議「AI 戦略会議」の開催など、AI の実用化がますます注目を集めています。実際の現場に AI を導入する際には、ユーザーが AI の予測を信頼できるかどうか重要です。そのために予測の根拠についての説明が求められています。しかし、ディープラーニングなどの機械学習のモデルはブラックボックス化されており、個々の予測の根拠が明確ではなく説明が難しい、という技術的な課題があります。本書では、各種モデル全般において予測を解釈・説明するための XAI（Explainable AI：説明可能な AI）の技術を解説。ブラックボックスをガラスボックス化するための手法を具体的に紹介します。

■XAI Python ライブラリで予測の背景を探る！

本書で取り上げる機械学習の予測モデルは、線形・非線形モデルのほか、アンサンブルモデル、時系列モデル、自然言語処理、ディープラーニング、コンピュータビジョンです。各種のモデルとデータに対して、XAI Python ライブラリの LIME、SHAP、Skater、ELI5、skope-rules などを使ったコーディングを行い、実行結果を見ながら、モデルがなぜそのように予測するのかを探っていきます。本書に掲載されたコードと結果を追っていくことで、XAI の手法を具体的に把握することができます。

<以下のような方に本書をおすすめします>

- ・ 予測結果についての説明力を上げたい方
- ・ 予測結果の根拠を追跡したい方
- ・ データサイエンティスト
- ・ 機械学習エンジニア（ソフトウェアエンジニア）
- ・ データエンジニア
- ・ 情報系の学生・研究者

■紙面イメージ

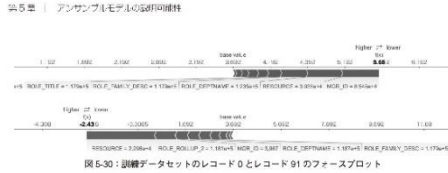


図 5-30: 訓練データセットのレコード 0 とレコード 91 のフォースプロット

図 5-30 のフォースプロットは、上のプロットに示されているプッシュ特徴量 (赤) の値と、下のプロットに示されているプル特徴量 (青) の値に基づいて、年の予測値がどのように決定されるのかを示している。上のプロットで、1.192 から 6.192 までの [] に反映されているのは年の予測値であり、クラス 1 に属する予測値の対数オッズである。対数オッズ [] の平均値は 3.692 である。プッシュ値は、MGR_ID、RESOURCE、ROLE_DEPTNAME などの特徴量を使って予測値を引き上げている (下のプロットでは、プル値が RESOURCE、ROLL_ROLLUP_2、MGR_ID などの特徴量を使って予測値を引き下げている)。レコード 0 の特徴量値は、最終的な年の予測値である。対数オッズとも呼ばれる。次に、各特徴量と SHAP 値の関係を散点図として可視化してみよう (図 5-31)。

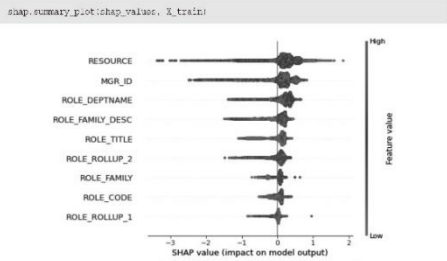


図 5-31: モデルの出力に対する影響を対数に示した SHAP 値のサマリープロット

図 5-31 は、各特徴量とそれらがモデルの出力に与える影響の大きさを示している。最も重要な特徴量が RESOURCE で、その後に MGR_ID などの特徴量が続いていることがわかる。

結果に対する各特徴量の影響度合いを確認 (第 5 章 アンサンブルモデルの説明可能性)

Weight	Feature	Weight	Feature
+0.884	excited	+0.814	excited
+0.754	refreshing	+0.754	refreshing
+0.729	perfect	+0.729	perfect
+0.710	superb	+0.710	superb
...
-0.977	more negative	-0.977	more negative
-0.711	acid	-0.711	acid
-0.714	lacks	-0.714	lacks
-0.728	forgettable	-0.728	forgettable
-0.731	forgettable	-0.731	forgettable
-0.754	lame	-0.754	lame
-0.761	horrible	-0.761	horrible
-0.768	dull	-0.768	dull
-0.780	falls	-0.780	falls
-0.780	falls	-0.780	falls
-0.780	falls	-0.780	falls
-0.811	terrible	-0.811	terrible
-0.857	poorly	-0.857	poorly
-0.857	poorly	-0.857	poorly
-0.885	lacking	-0.885	lacking
-0.892	disappointment	-0.892	disappointment
-1.000	waste	-1.000	waste
-1.331	waste	-1.331	waste
-1.372	waste	-1.372	waste

図 7-3: ポジティブクラスから抽出された、無意味な名前の特徴量 (左) と意味のある名前の特徴量 (右)

7.5 局所的な説明に対する特徴量の重みを計算する

特徴量の重みは次のパスを使って計算する。既定パスは、一連の `if/else/when` 文に従って、既定のルートから決定木の枝まで到達されたクラスを繰り返していく。木の節では、層ごとの重みを削減するためのモデルとしてロジスティック回帰を使っているため、重みはロジスティック回帰モデルの係数である。サンプルレベルでのモデルの重みは、決定木の特定の枝で使われる特定のパスである。基本形式は、重みはモデルの訓練ニューズで使う既定のパスである。

7.5.1 局所的な説明: 例 1

この例では `CountVecInfer` と線形分類器を使っているため、線形分類器の係数に依存する各単語の重みをマッピングする必要がある。1 行の `show_prediction` を表示する (局所的な説明を生成する) には、そのレビューを `show_prediction` 関数に渡す必要がある。

```
# 局所的な説明を生成
e15.show_prediction(c1f, lndb.review[15576], vec=vec,
target_names=cat(lndb.sentiment))
```

ここで使っているのは二値分類器であるため、対数オッズ (`log_odds`) を計算すると、1.726 であることがわかる。 $\exp(\log_odds) / (1 + \exp(\log_odds))$ の式を使うと、確率値 (0.849) が得られる。つまり、15577 行目のレビュー (レビュー 15576) の感情がポジティブである確率は 84.9% である (図 7-4)。なお、陣中の `lndb` はロジスティック回帰分類器の切片を表している。

5.7 SHAP を使って CatBoost ベースの多クラス分類モデルを説明する

CatBoost 分類モデルは、多クラス分類モデルの訓練にも使われる。訓練済みモデルオブジェクトを SHAP の `TreeExplainer` と変更し、SHAP 使用を促すことができる。

```
model = CatBoostClassifier(loss_function='MultiClass',
                           iterations=300,
                           learning_rate=0.1, random_seed=123)
model.fit(X_train, y_train, cat_features=cat_features, verbose=False, plot=True)

explainer = shap.TreeExplainer(model)
shap_values = explainer.shap_values(pool(X_train, y_train,
                                       cat_features=cat_features))
```

各特徴量と SHAP 値の関係を散点図として可視化してみよう (図 5-32)。

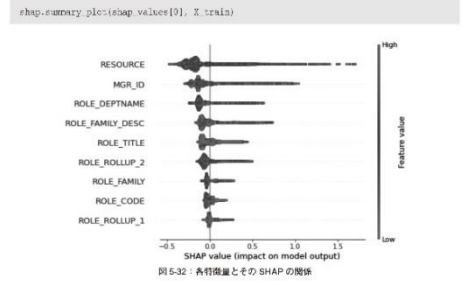


図 5-32: 各行の値と SHAP 値の関係

CatBoost ベースの分類器と回帰器には、さまざまなハイパーパラメータがある。図 5-6 に、重要なパラメータと、それらのパラメータがモデルの出力にどのような影響を与えるのかをまとめておく。

y-positive (probability 0.849, score 1.726) top features

Contribution*	Feature
+1.726	Excited (sentiment)
-0.017	-BIAS-

one of the other reviewers has mentioned that after watching just 1 oz episode you'll be hooked. they are right, as this is exactly what happened with me.

 the first thing that struck me about oz was its brutality and unflinching scenes of violence, which set it right from the word go. i just me, this is not a show for the faint hearted or timid. this show pulls no punches with regards to drugs, sex or violence. its is hardcore. in the episode use of the word

 /> is called oz as that is the nickname given to the overall maximum security state penitentiary. it focuses mainly on emerald city, an experimental section of the prison where all the cells have glass fronts and face inwards, so privacy is not high on the agenda. em city is home to many, many, many, gangs, latinos, christians, italians, irish and more... so scuffles, death stans, dodgy dealings and shady agreements are never far away.

 i would say the main appeal of the show is due to the fact that it goes where other shows wouldn't dare. forget pretty pictures painted for mainstream audiences. forget charm, forget romance, or doesn't show around. the first episode i ever saw struck me as so nasty it was surreal. i would say i was ready for it, but as i watched more, i developed a taste for oz, and got accustomed to the high levels of graphic violence. not just violence, but injustice (crooked guards who'll be sold out for a nickel, inmates who'll kill on order and get away with it, well mannered, middle class inmates being turned into prison bitches due to their lack of street skills or prison experience) watching oz, you may become comfortable with what is uncomfortable viewing... unless if you can get in touch with your darker side.

図 7-4: ポジティブクラス (y-positive) の単語に対する局所的な説明

図 7-4 において、図で表示されている単語は対数オッズスコアに対してプラスの値を示し、赤で表示されている単語はマイナスの値を示す。最終的な貢献度は、単語と他の単語の貢献度を合計したものであり、対数オッズスコアに等しい。

7.5.2 局所的な説明: 例 2

同様の方法で、レビュー 110 を説明してみよう。レビュー 110 はネガティブに分類されている (図 7-5)。

```
# 局所的な説明を生成
e15.show_prediction(c1f, lndb.review[110], vec=vec,
target_names=cat(lndb.sentiment))
```

レビュー 110 がネガティブクラスに分類される確率は 77.8% であり、観ては送られる平均は予測のネガティブクラスに発生している。

局所的な説明として各単語の貢献度を可視化 (第 7 章 自然言語処理の説明可能性)

※7 章は 4 色で掲載

■発刊記念キャンペーンとして本書の冒頭から第3章までを無料公開

本書の発刊を記念して、本書の冒頭（口絵）から第3章（線形モデルの説明可能性）までを期間限定で無料公開します。Webブラウザで下記URLのページにあるリンク先にアクセスするとお読みいただけます。

- ・無料公開へのリンクがあるページのURL：

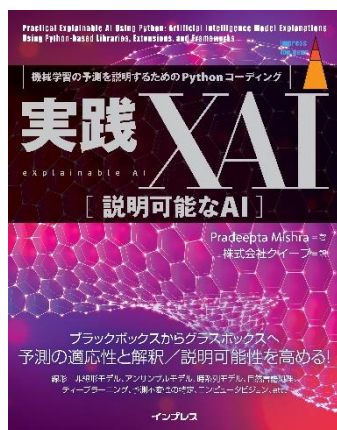
<https://book.impress.co.jp/books/1121101133>

- ・実施期間：2023/6/20（火）0:00～2023/6/26（月）23:59

■目次

- 第1章 モデルの説明可能性と解釈可能性
- 第2章 AIの倫理、偏見、信頼性
- 第3章 線形モデルの説明可能性
- 第4章 非線形モデルの説明可能性
- 第5章 アンサンブルモデルの説明可能性
- 第6章 時系列モデルの説明可能性
- 第7章 自然言語処理の説明可能性
- 第8章 What-If シナリオを使ったモデルの公平性
- 第9章 ディープラーニングモデルの説明可能性
- 第10章 XAIモデルの反実仮想説明
- 第11章 機械学習での対比的説明
- 第12章 予測不変性の特定によるモデル不可知の説明
- 第13章 ルールベースのエキスパートシステムでのモデルの説明可能性
- 第14章 コンピュータビジョンでのモデルの説明可能性

■書誌情報



書名：実践 XAI [説明可能な AI] 機械学習の予測を説明するための Python コーディング (impress top gear)

著者：Pradeepta Mishra

訳者：株式会社クイープ

発売日：2023年6月20日（火）

ページ数：320 ページ

サイズ：B5 変型

定価：3,960 円（本体 3,600 円＋税 10%）

電子版価格：3,960 円（本体 3,600 円＋税 10%）※インプレス直販価格

ISBN：978-4-295-01655-7

◇Amazon の書籍情報ページ：<https://www.amazon.co.jp/dp/4295016551/>

◇インプレスの書籍情報ページ：<https://book.impress.co.jp/books/1121101133>

■著者プロフィール

Pradeepta Mishra（プラディープタ・ミシュラ）

インドを拠点とする多国籍企業の IT サービス兼コンサルティング会社 L&T Infotech の AI データプロダクト部門上席。データサイエンティスト、計算言語学エキスパート、機械学習・深層学習のエキスパートからなる大規模なグループを率いる。以前には、Analytics India Magazine の「India's Top - 40 Under 40 DataScientists」に選出された。また、データサイエンスや AI に関する 500 以上の技術講演を、さまざまな大学や技術機関、コミュニティなどで行う。

【株式会社インプレス】 <https://www.impress.co.jp/>

シリーズ累計 7,500 万部突破のパソコン解説書「できる」シリーズ、「デジタルカメラマガジン」等の定期雑誌、IT 関連の専門メディアとして国内最大級のアクセスを誇るデジタル総合ニュースサービス「Impress Watch シリーズ」等のコンシューマ向けメディア、「IT Leaders」をはじめとする企業向け IT 関連メディアなどを総合的に展開・運営する事業会社です。IT 関連出版メディア事業、およびデジタルメディア&サービス事業を幅広く展開しています。

【インプレスグループ】 <https://www.impressholdings.com/>

株式会社インプレスホールディングス（本社：東京都千代田区、代表取締役：松本大輔、証券コード：東証スタンダード市場 9479）を持株会社とするメディアグループ。「IT」「音楽」「デザイン」「山岳・自然」「航空・鉄道」「モバイルサービス」「学術・理工学」を主要テーマに専門性の高いメディア&サービスおよびソリューション事業を展開しています。さらに、コンテンツビジネスのプラットフォーム開発・運営も手がけています。

【本件に関するお問合せ先】

株式会社インプレス 広報担当：丸山

E-mail: pr-info@impress.co.jp URL : <https://www.impress.co.jp/>

※弊社はテレワーク推奨中のため電話でのお問い合わせを停止しております。メールまたは Web サイトからお問い合わせください。