

画像認識 AI における評価技術の動向と紹介

城所 龍 Ryu Kidokoro, カランダン ローレンス Lawrence Earl Carandang

[要 旨]

人工知能(AI)の急速な進化と、医療や品質検査など精度が重要な分野への応用拡大に伴い、AI モデルおよびその開発プロセスの包括的な評価が必要である。本稿では、画像認識 AI の評価における実践的手法を概観し、精度だけでなく解釈性や信頼性を含む評価フレームワークの中で、説明可能 AI(XAI)の役割を示す。パイプラインレベルでの評価の必要性を示し、Local Interpretable Model-Agnostic Explanations(LIME)、Class Activation Maps(CAM)などの代表的な XAI 手法による推論解釈やモデル改善への応用を紹介する。また、テスト時拡張やベイジアンネットワークを用いた不確実性の定量化によるロバスト性評価についても論じる。さらに、評価手法の多様なタスクや実データへの適用、評価指標と開発目標の統合に向けた課題について展望する。

1 はじめに

近年、ディープラーニングの進展により AI はビジネスや医療など幅広い分野で急速に普及し、特に生成 AI や LLM は企業 IT 戦略の中核を担っている。一方、AI の意思決定の不透明性は信頼性確保において課題であり、公平性や説明責任を担保する評価技術の確立が急務である。こうした背景を踏まえ、説明可能 AI(XAI)を中心とした包括的評価フレームワークの重要性が高まっている。

XAI は、AI のブラックボックス問題を解消し、判断根拠を理解可能にする技術として注目されている。医療や金融など重要領域では、AI の判断根拠の可視化が法規制や倫理面で不可欠である。しかし、現状の AI モデルの評価は精度や速度に偏り、誤判断やバイアスの原因特定は困難である。LIME (Local Interpretable Model-agnostic Explanations)¹⁾や SHAP (SHapley Additive exPlanations)²⁾などの XAI 手法が寄与度分析を可能にしているが、説明の忠実度や安定性評価は発展途上である。信頼できる AI には、性能だけでなく説明の質や不確実性を統合した包括的評価技術が必要である。

本稿では、XAI による解釈性、公平性、説明の信頼性を定量化する新たな評価指標と手法を紹介する。

検出モデルの学習データとして Microsoft COCO データセット (© 2014 Microsoft. Licensed under the Creative Commons Attribution 4.0 International License.)を使用した³⁾。

2 AI モデルの評価

第 1 章で述べた AI の信頼性の課題を踏まえ、本章では AI モデルの評価技術を紹介する。

評価には量的および定性的な多様な観点が存在する。例えば、精度は分類等のあるタスクに対するモデルの性能を量的に評価する。学習時間は、AI モデルをどれだけ迅速に作成できるかを示し、時間制約のある業務において重要な要素となる。また、判断根拠の妥当性は AI が出力する結果の信頼性や説明可能性を定性的に示す観点である。これらの観点は、AI の性能を総合的に判断するために不可欠である。

2.1 AI 開発パイプラインの評価

第 2 章では「AI モデル」の評価について述べたが、評価対象はモデル自体に限られるわけではない。本節では、評価対象として「AI 開発パイプライン」を紹介する。AI 開発パイプライン(単に「AI パイプライン」)とは、AI モデルの開発から運用直前の一連の流れを指し、開発者、データセット、AI モデル、推論結果の 4 つの構成要素と、データ収集、前処理・後処理、学習、推論、性能計算の 6 つのプロセスから成り立っている。図 1 では、それぞれの要素とプロセスがどのように連携しているかを示している。

従来の機械学習では、モデルを構築してから精度を評価し、その結果に応じて改良を加える後追い型の開発が一般的だった。しかしこの方法は、特にデータ準備の段階で必要な特徴量の選定が不十分な場合、無駄なデータ収集や前処理が発生し、開発効率が低下する。この課題に対応するには、モデル評価を通じて、性能に貢献しない、あるいは悪影響を与える特徴量やノイズの多いデータを特定することが重要である。これにより、特徴量の取捨選択を含むデータ設計を最適化でき、限られたデータでも高精度なモデル構築が可能になる。

従って、モデル評価は単なる性能確認ではなく、データ設計とモデル改善のための重要な指針となる。

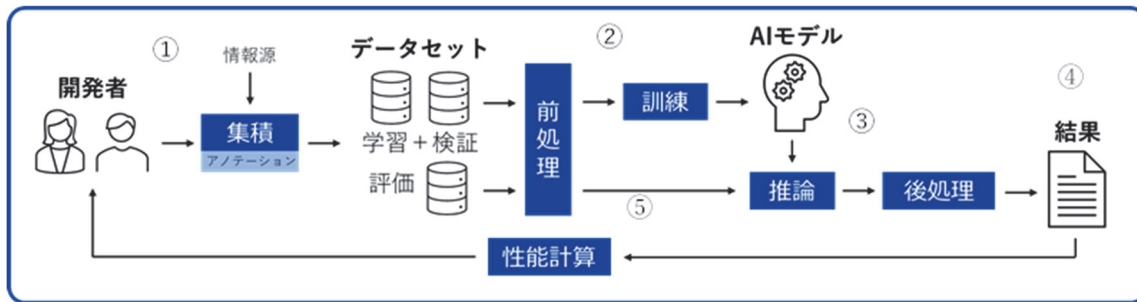


図1 AIパイプライン。① 開発者がさまざまな情報源からデータを取得し、必要に応じてデータにラベル（アノテーション）を付けてデータセットを作成する。データセットを学習用、検証用、または評価用データの3つのサブセットに分ける。② 学習用データと検証用データを前処理にかけ、あるAIモデルを訓練する（AI学習）。③ 評価用データを前処理にかけ、②で訓練したモデルに推論させる。④ 必要に応じて③の結果を後処理にかけ、最終結果を求めてモデル性能を測定する。⑤ 必要に応じて、モデル性能に基づいて条件を変えて実験をやり直す。

2.2 説明可能 AI「XAI」

近年、モデルの信頼性と性能の向上を目的に、ブラックボックスのAIモデルを解釈するための研究が世界中で数多く行われている。こうした研究の成果の1つが、いわゆる「Explainable Artificial Intelligence (XAI)」や「説明可能 AI」である。XAIとは、AIの判断や予測の根拠を人間が理解できる形で提示する技術や手法を指す。AIが注目する特徴と人間の視点を比較することでAIの評価手法の1つになる。

本稿では、広く用いられているXAI技術をいくつか紹介する。

2.2.1 大局説明と局所説明

XAIは、その説明範囲に基づいて大局説明と局所説明の2つに分類される¹⁾。大局説明は、AIモデルの全体的な振る舞いや内部の支配的な傾向を分析するものであり、複数のテストデータを用いてモデルの挙動を評価する。代表的なものとして決定木が挙げられる。

局所説明は、個々の予測結果の判断理由を理解し、入力の中の成分が最も貢献したかを把握するものである。大局説明と違って局所説明では、1つのテストデータのみを用いてモデルの挙動を評価する。代表的なものとして顕著性マップ¹⁾が挙げられる。本稿では、局所説明に焦点を当てるとともに、その代表的手法である顕著性マップについて、第3.1節「顕著性マップとは」にて詳述する。

2.2.2 モデル依存手法と非依存手法

XAI手法は、モデルの内部パラメータ（重み、勾配など）にアクセスできるかで分類できる。

モデル依存(model-specific)手法は内部パラメータを利用してネットワーク層間の出力を解析するため、特定のモデル構造にしか適用できない。例えば、リカレントニューラルネットワーク(RNN)の出力は

畳み込みネットワーク(CNN)とは異なる。

一方、モデル非依存(model-agnostic)手法は、入力と出力の変化を解析することで、内部構造に関係なく幅広いモデルに適用できる。

モデル依存手法は内部パラメータを用いるため、AIの特定の部分を見ることでより細かい説明や評価ができる。例えば、画像認識ニューラルネットのある層を見ることで、AIが物体の形、色、サイズなどが適切に学習されているか確認できる。一方、モデル非依存手法は内部構造に関係なく、あらゆるモデルに適用可能という利点がある。そのため、モデルの仕組みや内部パラメータが取得できるかに応じて、使い分ける必要がある。なお、不確実性の具体的な定量化手法については、第4章「不確実性計算を用いたモデル堅牢性の評価」にて詳述する。

2.3 評価指標

精度は性能を測る上で最も重要で第一の評価指標だが、唯一の評価基準ではない。AIモデルとその開発パイプラインも、その透明性、一貫性、データの品質、計算効率といった観点から評価できる。これらの評価観点は、モデルの精度と直接的または間接的に関連に関連するため、考慮する価値がある。

推論論理の解釈(透明性)は、正しい予測であっても誤った情報に基づく可能性を評価する。例えば、画像分類モデルが背景の雪を見てホッキョクグマと予測するケース(Clever Hans 効果⁵⁾)。前述のとおり、XAIなどの手法を使えば、モデルが注目している特徴を特定し、人間が見ている特徴と一致しているかが確認できる。これにより、学習手法の有効性を評価し、ユーザの信頼性も向上する。

モデルの堅牢性の定量化(一貫性)はカメラ画像のような入力データにおいて、照明や角度といった不可避な要因によるばらつき

¹⁾ 顕著性マップの可視化には、ヒートマップがよく使われる。実際、多くの場合、顕著性マップは「顕著性マップ」ではなく、「ヒートマップ」と呼ばれることがある。

を評価する。人間はこうした変化にも対応できるが、AI モデルは画素値のわずかな違いにも影響を受けやすい。優れたモデルは、条件の変化に左右されずに安定した検出ができる必要がある。堅牢性(ロバスト性)はこの安定性を測る指標であり、モデルがどのような条件下で信頼できるかを評価する手がかりとなる。

学習データの帰属(Training Data Attribution, TDA)は、誤ったラベル付けや冗長なデータが与える影響を分析し、データの品質を評価する。例えば、画像分類問題で猫の写真に「犬」というラベルを誤って付け、その画像にモデルを学習させると、性能低下を招く可能性がある。こうした問題に対処するため、TDA を活用し、モデルのパラメータ更新時に各学習データの重要性(重み)を算出する。この重要性に基づき、モデル学習に寄与しない、あるいは学習を阻害する可能性のあるデータを特定し、データセットから除外することが可能となる。

モデルの計算効率の測定(効率性)は、精度を維持しつつモデルのサイズの最適化ができるかを評価する。モデルが高性能になるほどファイルサイズや必要な実行環境も大きくなる。量子化⁶⁾や蒸留⁷⁾などの手法は精度を保ったままモデルを簡素化する手段となり、これらの前後でサイズを比較することで計算効率を評価できる。

2.4 性能向上への XAI 技術の応用

本稿ではハイパーパラメータの調整に関するアイデアを開発者に提供することに加えて、評価結果を AI パイプライン内で何らかのフィードバックループとして直接活用する方法をいくつか紹介する。例えば、物体検出モデルの作成中に、特定の学習用画像に対する顕著性マップ(詳細は第 3 章)を計算し、ground truth(正解データ)と比較することができる(図 2)。この比較結果を逆伝播に使用する追加の損失値として利用することが可能であり、これは XAI 手法による損失/特徴量/勾配の拡張の一例となる。顕著性マップは推論結果の補正にも活用でき、例えば、検出されたバウンディングボックスが顕著性マップ上の高顕著性領域と一致しない画像に対しては、信頼度スコアを下げるといった処理が可能である。

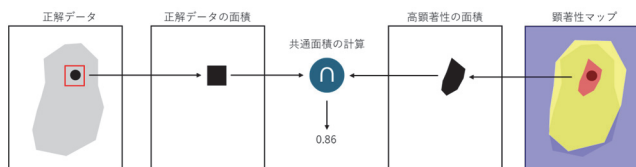


図 2 正解データと顕著性マップの比較方法の例。正解データとなるバウンディングボックスの面積と、顕著性マップによる抽出される高顕著性の面積の共通面積(Intersection over Union, IoU)を取る。

一方、前節で述べたように、TDA はデータセットの設計に活用で

きる。学習用データの選定がモデル性能に大きく影響することは広く知られており、TDA を用いたデータセット設計は、学習用サンプルの品質と情報量を定量的に評価する手段を提供する。これにより、モデルにとって最も有益なデータを明確に選定することが可能となる。

3 顕著性マップを用いた推論論理の解釈

本章では、物体検出の XAI を評価するために顕著性マップを紹介する。

3.1 顕著性マップとは

顕著性マップとは、推論時に最も注目されているデータの特徴を示す画像のことである。ここでいう「注目」とは、モデルが推論結果を出すときに寄与している特徴量のことである。図 3 は犬の画像に対して Eigen-CAM(Eigenvalue-based Class Activation Mapping)⁸⁾で可視化した顕著性マップである。このように、犬と判定した根拠は犬の中心の特徴量から判定したとわかる。

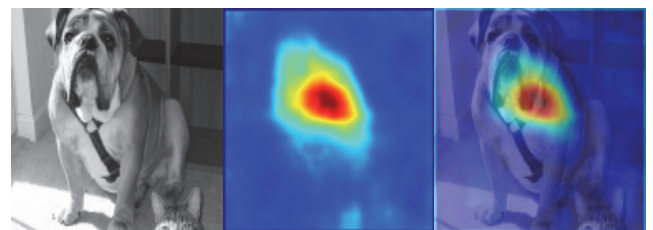


図 3 (左⇒右)対象画像、顕著性マップ、対象画像に顕著性マップを重ねた画像

3.2 顕著性マップ生成技術

顕著性マップ生成技術には多くの手法が存在するが、本稿ではモデル非依存の LIME、モデル依存の Eigen-CAM、および D-CLOSE(Detector-Cascading multiple Levels of Segments to Explain)⁴⁾の 3 つを紹介する。

3.2.1 LIME

LIME とは、入力データの一部をランダムに摂動させたサンプル群を生成し、それらに対するモデルの確信度を取得することで、局所的な説明モデルを構築する手法である。具体的には、摂動後のデータを説明変数、モデルが予測した確信度を目的変数として線形回帰を行い、各特徴の重要度を推定する。摂動箇所が予測に重要であれば確信度は低下し、重要でなければ確信度は維持される。

物体検出モデルに対して、LIME で可視化した画像作成には、摂動として対象領域の画素値をすべて 0 に置き換えたピクセルでマスクする手法が採用できる。その結果の例は図 4(b)であり、顔や頭、足から犬と判断していることが示唆されている。

3.2.2 CAM/Grad-CAM/Eigen-CAM

CAM(Class Activation Mapping)とは、CNN(畳み込みニューラルネットワーク)の構造を活用して、画像内のどの領域が分類に寄与したかを可視化する手法である⁹⁾。アルゴリズムとしては、CNNの最終畳み込み層の出力に対して、分類層の重みを乗じることで、画像内の重要領域を強調したマップ(アクティベーションマップ)を生成する。

CAMの弱点としては、モデル構造に制約がある点が挙げられる。具体的には、グローバル平均プーリング(GAP)を用いた分類層を持つCNNにしか適用できず、既存の多くのモデルに対して柔軟に利用できないという課題がある。そこで提案されたのが Grad-CAM(Gradient-weighted Class Activation Mapping)である。Grad-CAMは、モデルの出力に対する勾配を利用して、最終畳み込み層の特徴マップを重み付けし、画像内の注目領域を可視化する手法である¹⁰⁾。これにより、モデル構造に依存せず、幅広いCNNベースのモデルに対して適用可能となった。

物体検出などのタスクにおいては、Eigen-CAMが有効である。Eigen-CAMは、特徴マップに対して主成分分析(PCA)や特異値分解(SVD)を行い、固有ベクトルを抽出して可視化する手法である。これにより、クラスに依存しない形で、モデルが注目する領域を抽出することが可能となる。

CAMはさまざまな派生や改良が加えられており、例えば Cluster-CAM¹¹⁾や Expected Grad-CAM¹²⁾などが提案されている。これらの手法は、特徴量をクラスタリングする、勾配計算を改良することで処理時間や精度の向上を目指している。CAM系の手法は、説明可能AI(XAI)の技術として、モデルの判断根拠を人間に理解可能な形で提示する重要な手段であり、現在も注目され続けている技術である。

Eigen-CAMで可視化した画像は図4(c)であり、顔周の特徴から、犬であると判定している。

3.2.3 D-CLOSE

D-CLOSEは、複数の異なるサイズの摂動領域を用いて対象画像に対するモデルの予測を繰り返し取得し、それらの結果を統合することで、可視化を実現する手法である。LIMEでは、単一の分割粒度で摂動を加え、局所的な線形モデルにより説明を行うが、D-CLOSEでは、粗い分割から細かい分割まで複数のスケールで摂動を加えることで、広域的かつ詳細な特徴の影響を同時に捉えることが可能となる。D-CLOSEで可視化した画像は図4(d)であり、顔や頭から犬と判断していることが示唆されている。

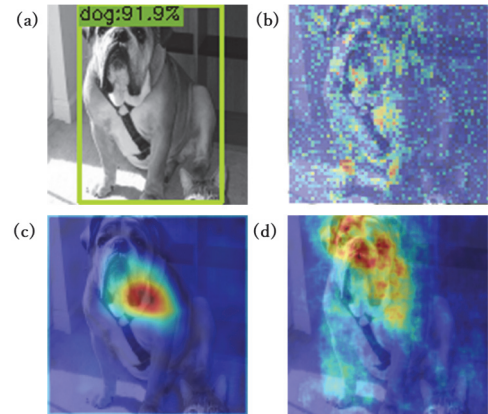


図4 (a) 推論結果付き入力画像。(b) LIMEで作成した顕著性マップ。(c) Eigen-CAMで作成した顕著性マップ。(d) D-CLOSEで作成した顕著性マップ。

4 不確実性計算を用いたモデル堅牢性の評価

本章では、モデル予測の信頼性を示す不確実性という指標の定量化手法を紹介する。

人間と異なり、AIモデルは学習していない入力に対して「わからない」とは答えられず、常に何らかの予測を返す。例えば、猫と犬のみに学習させた画像分類モデルにライオンの画像を与えると、学習していないにもかかわらず「猫」と高い確信度で誤分類する可能性がある。

予測の信頼性を評価する指標の一つに「不確実性」がある。物体検出では、確信度が物体の存在確率を示す一方、不確実性はその予測自体の信頼度を表す。前述の例にならえば、根拠(ライオン画像)のない推論(ライオンに対する「猫」という予測)は、たとえモデルの確信度が高くても、不確実性が高くなる可能性がある。

不確実性を測定する最も簡単な方法は、入力データの本質を変えない条件下で、同じ入力に対して複数回予測を行わせることである。例えば、入力画像の色を変更したり、サイズを変更したりすることで予測のばらつきを観察し、不確実性を定量化することができる。一般的な不確実性の指標には、信頼区間、変動範囲、分散、標準偏差などがある。

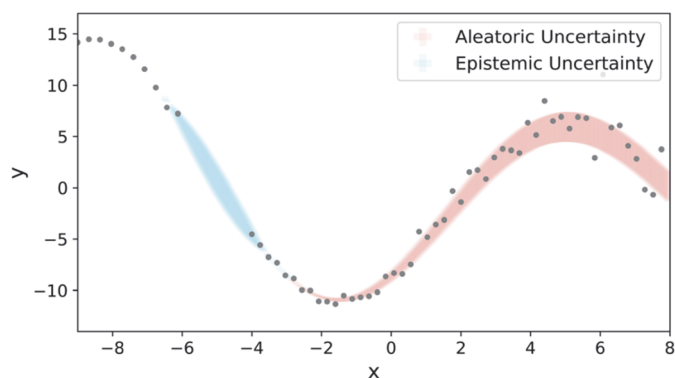


図5 認識論的不確実性(epistemic)と偶然的不確実性(aleatoric)¹³⁾。データ不足で-6~-4の間の振る舞いがモデル化できなくなり、認識論的不確実性が発生する。一方、ばらつきのため、実際の振る舞いがわからなくなる。これにより、モデルの予測は決めた値ではなく範囲になり、偶然的不確実性が発生する。

不確実性は、認識論的不確実性(epistemic)と偶然的不確実性(aleatoric)の2つに大別される¹⁴⁾。前者は、学習用以外のデータや不適切なモデル選択によって生じ、データの追加やモデル改善で軽減可能である。後者は、環境に内在するノイズやランダム性に起因し、環境を変えない限り軽減は困難である。図5は、回帰課題における両者の可視化を示している。

不確実性を定量化する一般的な方法は、同じテスト入力に対して複数の推論結果を得ることである。以下に、ネットワーク内部を変更する必要があるかどうか(モデル依存/モデル非依存)で分類して、いくつかの手法を紹介する。

4.1 モデル非依存手法

不確実性を評価する一般的な方法は、同じデータセットに対して初期値やハイパーパラメータ、アーキテクチャを変えて複数のモデルを作成する手法で、Deep Ensembles(深層アンサンブル)と呼ばれる¹⁵⁾。学習条件を変えることで、各モデルが異なる学習経路をたどり、収束点が異なれば推論結果も変わる可能性がある。学習用データに矛盾がない場合、未知データに対する高い不確実性は、モデルが重要な特徴を正しく捉えられていないこと、すなわち認識論的不確実性(epistemic uncertainty)を示す。

関連手法としてブートストラップ集約(bagging)があり、これはモデル構成を変えずに、学習用データから復元抽出した異なるサブセット(ブートストラップセット)を使って複数のモデルを作成する方法である¹⁶⁾。

ただし、複数のモデルを作成することが常に現実的とは限らない。単一モデル手法としては、テスト時拡張(Test-Time Augmentation, TTA)があり、追加の学習過程を必要とせずに既存モデルの評価が可能となる¹⁷⁾。TTAでは、評価対象の入力に対して複数の

変換(移動や回転などのアフィン変換、またはコントラスト拡張などの再着色)を施し、それぞれの変換結果をモデルに通すことで複数の推論結果を得る。図6には、そのような変換のいくつかの例が示されている。

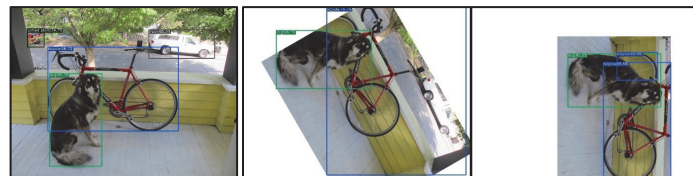


図6 元画像, 変換画像#1, 変換画像#2

変換パターンの範囲を探索し、検出信頼度を2次元配列として記録することで、モデルの不確実性の広がりグラフとして可視化することができる。これにより、モデルの弱点を特定し、学習方法を適切に調整することが可能となる。

図7は、図6左の画像における「犬」の検出確信度が、画像の回転角(-150°~150°)によって変化の様子を示している。0°(元画像)では確信度は0.95だが、78°回転で0.64に低下し、~100°では検出不能になる。これは、モデルが回転に対して不変でなく、性能に影響を与える可能性を示している。

このような変動自体は不確実性として扱えるが、単一の指標が必要な場合は、標準偏差や最大傾斜などの「平坦さ」指標が有効である。本研究では、標準偏差を用いる。確信度は区間[0,1]に有界であるため、その標準偏差は区間[0,0.5]に制限される。そこで、不確実性をパーセンテージとして表すための式(1)を次のように定義する。

$$\text{不確実性} = 2\sigma \times 100\% \quad (1)$$

図7の確信度の標準偏差を求めると、 $\sigma=0.2689$ となり、式(1)より不確実性は53.78%となる。従って、図6左の画像における「犬」の確信度0.95に対しては、約53.8%の不確実性が存在すると評価できる。

不確実性はモデルの性能ではなく、安定性や一貫性を示す指標である。正確な評価には両者の考慮が要る。高性能でも不確実性が高いモデルは、本質的特徴を捉えられず背景に依存している可能性があり、その原因としては、データの多様性不足をはじめ、さまざまな要因が考えられる。一方、性能・不確実性ともに低い場合、モデルが単純であることが示唆され、構造の見直しが重要である。最良のモデルは、高性能かつ低不確実性を両立している。

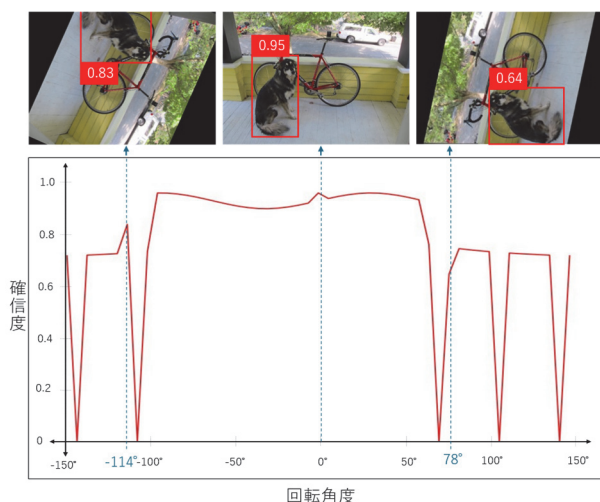


図7 画像回転度対「犬」の検出確信度グラフ

4.2 モデル依存手法

モデル非依存の手法と比較して、モデル依存の手法は、モデルのソースコードへのアクセスが必要となるため、実装が高度かつ困難である。しかし、これらの手法はより精緻な評価を可能にし、開発者がモデルの改善すべき具体的な箇所を特定することを可能にする。以下では、そのような手法のうち2つについて簡単に説明する。

従来のニューラルネットワークでは、各ニューロンまたはノードに対して正確な重みが計算され、学習過程で調整される。一方、ベイジアンネットワーク¹⁸⁾では、ネットワークの重みは確率変数であると仮定され、学習の目的は、これらの変数を記述する確率分布(通常はガウス分布)の平均と分散を予測することである。データを入力する際の順伝播では、これらの確率分布からサンプリングを行い、特定の重み値を取得する。ある入力データに対する不確実性を計算する際には、複数回の順伝播を実行し、それぞれ異なる重み値の組み合わせがネットワーク内で使用される。ベイジアンニューロンの数式を式(2)に示す。

$$y = f(\sum_i w_i x_i + b), \text{ where } w_i \sim N(\mu_i, \sigma_i^2) \quad (2)$$

y : ニューロン出力

$f(\cdot)$: 活性化関数(activation function)

w_i : 重み

x_i : 入力特徴量

b : バイアス値

μ_i, σ_i^2 : 分布 N の平均, 分散

モデルに Dropout 層²⁾が含まれている場合、それを用いて不確実性を計算する手法として Monte Carlo Dropout¹⁹⁾が存在する。Dropout は本来、訓練時に過学習を防ぐために使用されるが、こ

の手法では推論時にも Dropout 層を有効化し、ランダムな重みにゼロを割り当てることで、重みパターンを変化させる。これは事実上、該当する重みを無効化することになり、ベイズネットワークと同様の効果を生み出す。これにより、複数の推論結果を取得することが可能となる。

5 まとめ

AI モデル開発においては、モデル構築後に精度を評価し、その結果に基づいて改良を行うことが重要である。しかし従来は、モデルの改良には大量のデータ追加が必要となり、多大なコストがかかっていた。これは、AI の判断根拠がブラックボックスであり、効果的なデータの特徴が人間には分かりにくいことが一因である。本稿では、こうした課題に対し、XAI 技術の紹介と実施例を通じて、効果的な追加データの選別方法の可能性を示した。さらに、これらの XAI 技術を AI パイプラインの性能評価に組み込むことで、過不足なデータの存在を可視化し、理解を促進できることを示した。なお、本稿の研究では物体検出モデルの評価を行ったため、他のタスクに対しては別の評価手法が必要となる。今後は、異なる AI モデルや実データへの適用、不確実性を求める関数によるさらなる XAI 技術の開発・評価を進める予定である。また、これらの評価技術を AI パイプラインに組み込む際には、目的に応じた評価指標の設計を行う必要があり、これらについても今後の研究課題となっている。

参考文献

- 1) M. T. Ribeiro *et al.*, ACM SIGKDD **22**, 1135–1144 (2016).
- 2) S. Lundberg *et al.*, Neural Inf. Process. Syst. **31**, 4768–4777 (2017).
- 3) T.-Y. Lin *et al.*, Eur. Conf. Comput. Vis. **13**, 740–755 (2014).
- 4) V. B. Truong *et al.*, Proc. Mach. Learn. Res. **222**, 1385–1400 (2023).
- 5) J. Kauffmann *et al.*, arXiv:2006.10609 (2020).
- 6) N. Wang *et al.*, Adv. Neural Inf. Process. Syst. **32**, 7686–7695 (2018).
- 7) G. Hinton *et al.*, NIPS Workshop (2014).
<https://doi.org/10.48550/arXiv.1503.02531>
- 8) M. B. Muhammad *et al.*, arXiv:2008.00299 (2020).
- 9) B. Zhou *et al.*, IEEE Conf. Comput. Vis. Pattern Recognit., 2921–2929 (2016).
- 10) R. R. Selvaraju *et al.*, Int. J. Comput. Vis. **128**, 336–359 (2019).

²⁾ Dropout 層：学習中に、一部のニューロン（計算ユニット）をランダムに無効化する層。

- 11) Z. Feng *et al.*, arXiv:2302.01642 (2023).
 - 12) V. Buono *et al.*, arXiv:2406.01274 (2024).
 - 13) C. Yang *et al.*, J. Cheminform. **15**, 13 (2023).
 - 14) A. Kendall *et al.*, Adv. Neural Inf. Process. Syst. **30**, 5574–5584 (2017).
 - 15) B. Lakshminarayanan *et al.*, Adv. Neural Inf. Process. Syst. **30**, 6402–6413 (2017).
 - 16) L. Breiman, Mach. Learn. **24**, 123–140 (1996).
 - 17) G. Wang *et al.*, Med. Image Anal. **54**, 154–162 (2019).
 - 18) D.J.C. MacKay, Neural Computation, **4**, 448–472 (1992).
 - 19) Y. Gal *et al.*, Proc. Mach. Learn. Res. **48**, 1050–1059 (2016).
-

執筆者



城所 龍
先端技術研究所
第三研究室



カランダン ローレンス
先端技術研究所
第三研究室