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

『説明可能 AI（XAI）とは？』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：
「1．説明可能 AI（XAI）とは？～深層学習の説明性向上と XAI の今後の展望～」
情報処理, 63(8)， e1-e7 (2022-07-15)

■記事内容の要約とコメント（→★）

【機械学習と深層学習】

・深層学習：
＊学習用のデータを用意すれば、入出力変換を作成可能。
＊様々な分野に適用可能
＊必要な学習用のデータが膨大
＊処理内容が理解困難 → 企業利用の障害

・図１：現状の機械学習の説明性と精度の関係
＊深層学習は、精度は高いが、説明性が低い
＊決定木は、精度が劣るが、判断基準を理解しやすい
＊説明性と精度の間にトレードオフの関係あり　→★一般論としては疑問
＊本文では、深層学習の説明性向上について述べる

【機械学習の説明性について】

・ＡＩの説明性向上の具体項目：
① 入力から出力を得る「判断根拠の可視化」
② 入力から出力への「機序（処理手順）の提示」
③ 未学習の新たなデータに対する応答・出力

・説明性の考え方（納得しやすさ）には下記の個人差あり：
１）理論：法則・定理による証明　→ＣＡＥ(computer aided Engineering)の分野で必要
２）論理：演繹推論（例：Ａ→Ｂ, Ｂ→Ｃならば、Ａ→Ｃ）
３）数式：複雑でない、単純な式
４）図・グラフ：視覚表現　→従来から利用されている
５）事例：類似する学習済みデータの提示　→未来予測での説明のよりどころ
６）言葉：キーワード、文章による説明　→多くの人に受け入れられそう
→★そもそも深層学習の方式が前提の説明要求には無理があるのでは？

・ＸＡＩの設計：説明の対象者が期待する方法が必須
・内閣府「人間中心のＡＩ社会原則検討会議」の「公平性、**説明責任**及び透明性の原則」では、
ＡＩ利用において、説明性を適切に確保することを挙げている。
・特に、人の生命や財産にかかわる場面で必要：
＊自動運転の人検知処理が深層回路の場合の事故発生時
＊医学的な診断・治療支援
・企業でのＸＡＩ利用事例は、特集後半の解説にあり。

【深層学習の説明手法】

判断根拠の可視化例１：中間層・特徴量の可視化

・初期のＸＡＩ：学習済みの深層回路の内部信号の可視化
例：画像のクラス分類で、特定のクラスに強い信号を出力する際に、
深層回路の中間層の出力を画像として提示することで、
エッジなどの原始的な特徴量が組み合わさってパーツとなり、
クラスの特徴が形成される様子が見える。
しかし、直感的に理解しづらいという意見も多い。

・深層回路は、前半で入力から特徴量を抽出し、後半で最終判断するので、
中間層の信号を特徴量とみなせる。
・低次元に削減された特徴量を、２～３次元の図で理解しやすく可視化する。
・クラス分類の場合は、正解ラベルごとにデータを着色して分布状態を見やすくする。
・教師なし学習として、クラス分類や異常検知に用いられる。
ただし、特徴の意味が人の感覚・印象と会わないこともある。

→★中間層の状態表示が説明機能の役割を果たすためには、
中間層の各ノードが示す具体的な特徴を表現する必要あり。

判断根拠の可視化例２：影響度のヒートマップ表示

・特定の出力信号に対する入力信号の影響度を数値や色で、
ヒートマップ／アテンション表示して、深層回路の判断根拠を提示する手法。
・画像認識や自然言語処理でよく用いられる。例えば、犬と猫の分類アプリにおいて、
「犬」の判断に大きな影響度を持つ部分を入力画像の中に示すと、利用者は安心する。
なお、本手法は、入力から出力に至る判定プロセスは示していない。
・産業応用において、製品の製造工程の画像による欠陥検査処理に深層学習を適用する際、
判断根拠となる画像領域を提示することで、処理が正常であると確認できる。

→★この手法では、出力層から入力層へのバックトレースが必要と思われるが、
影響度の計算方法が、詳細不明。
（入力から出力までのニューロン間の信号の流れをすべて記録しておく？）

機序の提示例１：多次元空間の局所線形化

・機械学習によるクラス分類に用いる特徴空間は、数十～数百次元か、それ以上なので、

クラス間の境界面（識別面）は、非常に複雑な多次元曲面になることが多いため、
指定された一つの事例の判断根拠をわかりやすく説明するのは難しい。
・そこで、ＬＩＭＥでは、対象事例の近傍をサブサンプリングして、局所的に線形分離し、
その識別面を判断根拠として提示するので、説明がシンプルになり、納得感が高まるが、
事例ごとに説明が異なり、「判断根拠の一貫性が低い」ともいわれる。

→★線形分離に関して、第１次ＡＩブームのときのパーセプトロンを思い出した。
線形分離可能なものしか扱えないが、私の修士論文の参考文献[34]として、
本文「1.4節 思考に関連した工学的諸研究について」で引用している。
詳細は、以下の過去ブログ／エッセイで言及している。
・2016.3 「Marvin Minsky の逝去を悼む」
▼エッセイ http://www.1968start.com/M/blog/1603Minsky.htm
▼ブログ http://www.1968start.com/M/blog/old.html#1603d

機序の提示例２：構造の単純化・最適化・自動構築
・深層回路は一般に多層の複雑な回路なので、回路網の規模を小さくして理解しやすくする。
・回路の構造単純化：
＊学習後の結合荷重が小さい信号線を削除（結合荷重：０）し、
精度劣化を補うために再度学習する。
＊結合荷重のビット数削減により、最終的に論理回路まで圧縮する量子化ベース手法
・しかし、精度を維持して回路規模を劇的に激減するのは難しい。

→★階層の削減や結合荷重のビット数削減での精度劣化は、当たり前。本末転倒。

・2010年代後半の手法：Neural Architecture Search（Auto-machine learningの一つ）
＊勾配降下法、進化計算法などの最適化法により、
結合荷重の学習中に深層回路の構造も最適化する方法
＊回路網を目的に合わせて全自動で構築する方法
・Auto-machine learningでは、以下の項目を可能な限り自動化して、
機械学習の利用を容易化する研究が盛んである。
＊使用する機械モデルの選択
＊モデルの学習係数のハイパーパラメータの調整
＊データの前処理
＊アルゴリズム・プログラム開発
・深層回路の構造最適化によって、回路網への入力信号を目的に応じて取捨選択でき、
入力信号の重要度を求めることも可能。
・筆者は、深層回路を線形回路に変換して、レーダーチャートや言葉で、
簡潔に説明する手法を開発

→★現在の研究状況は、特定のアプリケーションのために、
試行錯誤しながら良い結果の出るものを探しているように思える。
50年前の私の修士論文の研究と似ていて、・・・（苦笑い）

当時、思考モデル（ニューラルネットワーク）を Fortran プログラムで作成し、
幾つかのパラメータの値を少しずつ変更しながら良い結果の出るものを探した。
大型計算機センターに依頼したバッチジョブは 3 週間後に結果が出る状況だったので、
パンチカードをコピーし、前の結果を確認する前に次の依頼をしていた。

（参考：学会発表）中所、齋藤 ： 「思考過程のシミュレ-ション」、
電子通信学会オ-トマトン研究会資料、A70-76(Dec.1970)
▼http://www.1968start.com/M/bio/olduniv/gakkai7012.html
（参考：修士論文）中所：「思考過程の数学的表現と模擬実験」（1971 年）
▼http://www.1968start.com/M/bio/olduniv/shuuron.htm

他の手法：転移学習・浸透学習

・転移学習とは、既知の深層回路モデルの知識（構造や結合荷重）を転用するもの
・類似の目的の学習済みの深層回路の前段部分（特徴量形成部）の構造・信号の
結合荷重をそのまま利用し、後段の特徴の組合せ部を新たなデータで学習することで
特徴量の知識を利用する。十分な精度がでないときは、初期化部分を入力側に
近づけることで、説明性を担保する。
・扱う問題によっては、学習時に利用できても運用時に利用できない情報（潜在変数）と
両方で利用できる情報（顕在情報）がある。通常は健在変数だけ利用しているが、
説明性の観点では潜在変数も利用したい。
・そこで、筆者の開発した浸透学習法では、初めに最終出力に対する勾配降下で
全結合荷重を決定後、図６の浸透データが変化しないように赤色の結合荷重を
調整しながら桃色の・・・

→★このあたりの話は理解できない　(^^;;　説明性の説明が説明不足では（^^）

【ＸＡＩの今後の展望】

ＩＢＬからＥＢＬへ

・現在主流の、多数の事例を闇雲にコンピュータに与えるだけの事例に基づく学習
（Instance Based Learning）から脱却し、少数の事例を用いた説明に基づく学習
（Explanation Based Learning）へと、機械学習の質的変換が期待される。

ＸＡＩから共進化型ＡＩへ

・共進化型ＡＩでは、人と機械の相互作用により、お互いに相補的に知識を高め合う。

→★実用性の観点では、全自動は無理なので、半自動がいいということか？
もともと深層学習の説明機能は無理というなら、心情的には理解できるが・・・

（参考：関連する過去ブログ）
▼2019.1　「ＡＩ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