連載 ビブリオ・トーク ―書評―

…→ 石井一夫（久留米大学）

# ソフトウェア工学から学ぶ機械学習の品質問題

中島　震 著
丸善出版（2020），3,080 円（税込），192p., ISBN：978-4621305737

## 機械学習の品質問題とは

機械学習の品質問題は，ソフトウェア工学における機械学習の研究開発で話題となっているホットなテーマである．当初この話題を耳にしたとき，機械学習の識別性能はデータのバイアスや分析者の分析能力などに依存する面が大きく，機械学習の品質という言葉に違和感を持った．機械学習モデルそのものについては，ソフトウェア開発段階において新たにモデルを開発するということはなく，ソフトウェアレベルでのモデルの開発者という概念はないと感じられる．現に，意図的にバイアスのかかったデータセットで学習させることで，誤った予測に誘導することも可能である．AI によるチャットボットが不適切な学習の末，ヘイトスピーチを発するようになった事件（2016 年）はそれにあたると思われる．これらは「AI の公平性（Fairness）」の問題として知られている．

一方で，機械学習の識別・予測性能の評価については，交差検定や，混同行列による真陽性 (True Positive)，真陰性 (True Negative)，偽陽性 (False Positive)，偽陰性 (False Negative) などの評価法が確立している．また，ニューラルネットワークにおける重みを学習する際に，重みが形成される仕組みの理由付けを説明することが困難であるといういわゆる「ブラックボックス問題」がある．このブラックボックス問題は，特に医療分野や，金融，自動運転などのミッションクリティカルな分野へ応用するときには，「AI のアカウンタビリティ（Accountability；説明責任）」の問題としてクローズアップされることが多い．

もう 1 つ機械学習の研究開発や社会実装を考える上で問題となる点は，「AI の民主化」の問題である．これは，AI の大衆化を意味する．すなわち，機械学習を実装する上で，さまざまなライブラリやフレームワークが開発され，機械学習の中身を知らなくても簡単なアプリを操作するだけで気軽に AI を操作できる時代になってきている．むしろ，機械学習の中身を学ぶことなく，気軽にアプリ感覚で AI を使っている人の方が多いのではと考える．こうなると，一種の占いや妖術と大して変わらなくなる．もう 1 つの「ブラックボックス問題」である．いくら手軽に使えるようになってきているとはいえ，占いや妖術感覚で，使われたのではその結果に影響が出てくる立場の人たちにとっては人権問題にもなり得る．これは，「AI の透明性（Transparency）」の問題と言われるものである．たとえ，「AI の透明性」が担保されていても使用者がそれを見透かすことができなければ，救いようがない．

このような，機械学習の品質問題に関するいろいろな課題がある中で，本書を手にすることは機械学習の研究開発や社会実装を考える上で非常に意義深いと思われる．

## 本書の構成と内容

本書はこのようなソフトウェアの品質問題について，特に「教師あり学習」に関してまとめている．

本書は，以下の 7 章構成である．

第 1 章「データ利活用の時代」は，機械学習をめぐる社会的背景と，機械学習の概要を簡単に説明している．

第 2 章「機械学習ソフトウェアとその品質」は，機械学習の中でもニューラルネットをめぐる過学習などの問題点と，機械学習の品質について，国際規格 SQuaRE で規定された品質モデル（製品品質，利用時の品質，データ品質）の点から説明している．特に，「毒入れ攻撃」や，「敵対錯乱」，「妨害攻撃」などが品質の劣化要因となると説明している．ここでいう品質の劣化とは，予測・推論にかかわる正確性とロバスト性が影響を受けることをいう．

第 3 章「ソフトウェアテスティングの方法」は，プログラム検査の基本的事項について述べたあと，機械学習プログラミングの標準的検査法になっているメタモルフィックテスティングについて述べている．機械学習システムでは，訓練アルゴリズムが正しくモデルを構築できたかは，モデルのパラメータやモデルの精度で判断するのが難しく，テストオラクル（テストの成否判定メカニズム）を直接用意する以外の方法でテスト成否の判断をしなくてはならない．このようなシステムあるいはプログラムはテスト不可能プログラムと呼ばれる．メタモルフィックテスティングは，テスト不可能プログラムに対して疑似的にテストオラクルを用意する手法の 1 つで，「入力に対してある一定の変化を与えたときに，出力の変化が理論上予想できる」という関係（メタモルフィック関係（Metamorphic Relation））を利用して，複数の入力を用意しそれらの出力間の関係を比較することでテストの成否を決める方法である．

第 4 章「データセット多様性」は，機械学習ソフトウェアの品質はデータセットに依存するが，データセットの品質を考えることは想像以上に困難であることを，偏りのある MNIST（Mixed National Institute of Standards and Technology database）データを用いた分類予測実験を基に示している．つまり，訓練データセットに偏りがあると，期待と「異なる」学習をすることを示している．

第 5 章「深層ニューラル・ネットワーク検査の実際」は，第 3 章のメタモルフィックテスティングおよび，第 4 章のデータセット多様性の具体的な適用事例について自動運転を用いて，深層ニューラル・ネットワーク検査手法の実際例を紹介している．

第 6 章「品質からみた敵対データ」は，機械学習ソフトウェアのロバスト性に影響を与える敵対データの問題を取り上げている．敵対データとは，機械学習ソフトウェアの分類に影響を与えるフェイクデータである．敵対データにより，機械学習ソフトウェアに誤予測を誘導することを敵対攻撃という．敵対攻撃には，敵対錯乱，目標明示攻撃，目標非明示攻撃があり，データ補完や学習方式の工夫，入力フィルタなどにより防御・検知できる．

第 7 章「機械学習ビジネス・エコシステム」は，機械学習ビジネス・システム開発業務の流れとその取り決めなどのビジネスを円滑に進めるための課題について述べている．さらに，価値共創プラットフォームとしての機械学習ビジネス・プラットフォームの例として，アメリカ，ヨーロッパ，台湾の例を紹介している．

## 本書をどのような人たちに推薦したいか

機械学習の開発や実装，運用にかかわる，あるいは，将来かかわりたい技術者や学生が，ぜひ目を通しておくべき重要な話題である．

（2020 年 12 月 19 日受付）

**石井一夫（正会員）**
kishii@rs.sus.ac.jp
専門分野：ビッグデータ分析，計算機統計学，データマイニング，数理モデリング，機械学習，人工知能．2015 年度本会優秀教育賞受賞．2019 年度本会シニア会員．2020 年（株）エヌ・ティー・エス学術顧問．日本技術士会フェロー，APEC エンジニア，IPEA 国際エンジニア，現在の所属は公立諏訪東京理科大学．