特別講義録フォーマット

第 9 回特別講義 レポート

日時	2022 年 1 月 7 日（金）10:00～12:00
実施形態	オンライン（Zoom）開催
テーマ	機械学習品質マネジメントガイドライン策定と標準化の取り組み
講師名・所属	大岩　寛　氏 産業技術総合研究所　デジタルアーキテクチャ研究センター
司会	石川 冬樹 氏（国立情報学研究所／本研究会 研究コース 5 主査）
アジェンダ	1．背景 1.1　なぜ機械学習 AI（に品質）が必要か？ ・周囲の状況 1.2　機械学習品質マネジメントガイドラインの策定 ・検討の経緯 ・内容 1.3　標準化等の取り組み
アブストラクト	機械学習 AI を用いたシステムが広く世の中で使われるようになっていくなかで、システムの信頼性や安全性を担保するための品質マネジメントの必要性が生じてきている。現在産学協同の取り組みで検討している機械学習品質マネジメントガイドラインの策定について、その背景となる世界の状況や品質管理の考え方、標準化の状況などについて紹介する。
講義の要約	

◆講師紹介
2005 年 3 月：東京大学大学院情報理工学系研究科博士課程修了。
同年 4 月：独立行政法人産業技術総合研究所に入所。
情報セキュリティ・ソフトウェア工学などの研究に従事し、2021 年 4 月より現職。

1. 導入・背景
- なぜ AI に品質が必要か？
  今は日々の生活・人命をソフトウェアによる制御に預けている。
  →航空機や鉄道車両、自動車、家電、インフラなどはコンピュータによって制御されている。

  最近はソフトウェア構築に、機械学習のソフトウェアが入ってきており、機械学習 AI に命を預けるという状況が目前まで来ている。

- なぜ AI が必要か？
  機械学習は現状では品質管理に不利で、不安要素を解消できないが、従来のソフトウェア構築に限界が来ているため、機械学習が必要とされている。
  ・品質管理に不利
    従来の品質管理手法が通用しない。なぜ上手くいったのか説明できない。
  ・不安要素を解消できない
    次の開発でも正しく動くことを保証できない。
  ・従来のソフトウェア構築に限界
    ・扱う問題の複雑さが増し、従来の構築手法では追従できなくなっている。
    ・勘や経験、総合的な判断が必要な場合などを記述できない。
    ・人ができないことを計算機にやらせたい。

- 社会からみた AI への恐怖と要求
  AI の「得体の知れなさ」や「社会的影響」への恐怖も顕在化している。
  最近は規制や合意をもって制約をかけていく方向になっている。
  ・人間中心の AI 社会原則
    セキュリティ（安全性・信頼性を含む）や公平性・説明責任・透明性の確保についても記述されている。
  ・OECD Principles on AI
    OECD の各国における、AI に関するルール化などの方向性についての合意文書。
    公平性と公正性の確保や、堅牢・セキュア・安全性とリスクアセスメント、開発運用者責任について記述されている。

  ＜法律・ガイド層の各国の取り組み＞
  ・アメリカ
    NIST が AI リスクマネジメントフレームワークの作成に向けて関連情報を募集。
    法律ではないが、政府調達の要件に入れることで社会に影響を与える。
  ・欧州
    欧州委員会が AI の規制法案を公表。使用方法や使用できない場所などを定義している。
    施行された場合、欧州向け市場で影響が大きい。
  ・日本
    AI ガバナンスガイドラインを発表。
    ガイドラインで共通認識を形成することで、自主的な取り組みを促進している。

- ビジネスからみた AI と品質
  ＜問題点＞
  ・顧客に安心して買ってもらえない。PoC から先の開発に進めない。
  ・誤作動時の責任問題。どこまで責任を負うのかわからない。
  ・売買契約において不利になることがある。

- AI の品質・セキュリティリスク
  AI システムも IT システムの一種
  →全く違うプロセスで行うべきではない。IT システムとして品質を説明する必要がある。

  なぜ AI システムの品質が難しいか？
  →環境の多様性が大きすぎる（サイバーフィジカルシステムの難しさ）。
  　ソフトウェアとしての構造が特異で、機械学習にすることによって生まれるリスク。

- 従来のソフトウェアの品質の考え方
  システムが構造的に構築されることに対応して、品質も構造的に作っていく考え方。
  すべてのリスクに対策を実現した＝安全であるという思想で、開発プロセスに抜けがないことを担保する。

- 機械学習ソフトウェアの特異点
  ・リスクの要因をすべて学習させても正しく判断するとは限らない。
  ・テストをしても網羅性が分からない。
  ・修正をした場合、ほかの部分への影響を避けられない。

- 高品質な機械学習ソフトウェアへ向けて
  従来のソフトウェア工学の前提が崩れている。
  →機械学習 AI に合わせて新たな「品質の作りこみ」の枠組みを作る必要がある。

2. 機械学習品質マネジメントガイドライン
   機械学習 AI の品質を「作りこみ」「確認し」「説明する」ためのガイドライン。
   ・主な想定読者
   　機械学習の製品やサービスの提供者や機械学習のソフトウェアの開発者
   ・2 次的な想定読者
   　サービスの利用者、第 3 者評価機関

   - 取り組みの狙い
     ・社会全体で受容性が高まり、安全性が向上すること。
     ・競争力強化、品質を見えるようにする。
     →ユーザに納得して使用してもらえるようにする。

   - ガイドラインの位置づけ
     人間中心の社会原則の下に入る位置づけで、どのように作るかを定義している。
     将来的には業種別にガイドラインを作り、各企業のガイドラインのベースにしてもらう。

   - 検討の主体は機械学習品質マネジメント検討委員会。

- 品質管理の対象と考え方
  ガイドライン内で着目する品質は、利用時品質、外部品質、内部品質。

- 外部品質：3 項目×レベル
  機械学習のコンポーネントが達成すべきことを 3 つに整理
  ① リスク回避性
  危険につながる判断をしない（する確率を一定以下に抑える）。目標レベルは 7 段階。
  従来の安全性基準（IEC 61508 SIL）準拠の 4 レベルと、従来 SIL0 に対応するものとして 3 レベル。高い安全性が必要なものは SIL 評価が必要とされる。
  ② AI パフォーマンス
  全体として平均性能を高く保つ性質。目標レベルは 3 段階。
  ③ 公平性
  入力の属性に依存して、統計的に望まない偏りがないこと。目標レベルは 3 段階。

- 内部品質
  ＜整理の仕方＞
  ・ボトムアップ・アプローチ
  AI 学会が注目している技術に着目する。
  →技術的に実現できること、何をやるべきと思っているかが判る。
  →積み上げても十分かはわからない。
  ・トップダウン・アプローチ
  AI が誤動作したときの原因を仮想的に分析。
  →対処できるかわからないが、方向性としては正しいはず。
  →ガイドラインでは、トップダウン・アプローチの中に、ボトムアップ・アプローチを組み込んで整理している。

- 内部品質 9 項目
  A) 問題の分析に基づくあるべきデータセットの設計
  A-1 問題領域分析の十分性、A-2 問題に対する被覆性
  ・従来開発の要求分析・テスト設計に対応。
  ・どういう観点で品質を問うかを決める。
  B) 設計に合致する良いデータセットの確保
  B-1 データセットの網羅性、B-2 データセットの均一性、B-3 データセットの妥当性
  ・十分なデータが均一にあること。
  ・レアケースに対しても設計すること。
  ・データセットの個別のデータが妥当か。
  C) 良いデータセットから得られる良い機械学習モデル
  C-1 機械学習モデルの正確性、C-2 機械学習モデルの安定性
  ・データセットから良い訓練済み機械学習モデルを作る工程。
  ・正確性・安定性は分けて考える必要がある。
  D) 信頼できるソフトウェア
  D-1 プログラムの健全性
  E) 品質を維持する運用
  E-1 運用時品質の維持性
  ・テストまでうまくいっていたモデルが運用中に品質劣化する可能性がある。

・更新方法には、オフライン更新、オンライン更新がある。
・ソフトウェア・レグレッション…あらかじめ運用方針を決めておく必要がある。また、テストでの確認も必要である。

- システムライフサイクルプロセス
  規格段階から運用・利用終了までの総合的な品質マネジメントを想定して整理。
  品質のチェックポイントを明確化させる意図がある。

  ガイドラインの提案するプロセスは「抽象的モデル」
  →PoC と最終開発は頭を切り替える必要がある。1 度は仕様を固定する必要がある。

- 従来の規格との関連性
  安全性が重要な分野では SIL および分野別詳細規格への対応はほぼ必須。
  →AI バックグラウンドの人は規格についてほとんど知らないことが多いため、
  ガイドラインでは両立できるような記述となっている。

- 機械学習品質マネジメントガイドライン
  ・日本語第 2 版：2021 年 7 月公開
  ・英語版第 1 版：2021 年 2 月公開→第 2 版は近日公開予定。

3. 国際標準化
   - 国際標準化の流れ
     ・ISO/IEC JTC 1/SC 42（Artificial Intelligence）
     WG1 で基本用語集の検討、WG3（Trustworthiness）で基本的な議論、データ品質、AI 開発ライフサイクル、品質指標モデルなどの議論も開始された。
     ・ISO/IEC TR 29119-11、IEEE P7003、NIST 等での活動

   - TR 5469
     IEC と合同で進めており、テクニカルレポートの公開に向けて検討中。

   - 国際標準化の現状方針
     当面は、TR 5469 に沿った実装をする際のガイドラインとなるよう準備している。

4. 今後の取り組み
   - 活動の全体像
     ・機械学習 AI 品質管理ガイドライン
     ・産業分野別 AI 品質リファレンスガイド
     ・品質管理テストベッド（Qunomon）
     ・評価ツール

   - 品質管理テストベッド（Qunomon）
     高品質 AI を実現するための品質開発環境で、テストを評価レポートとしてまとめる、再利用性を高めることのできるツール。

   - 応用別のリファレンスガイド
     ガイドラインを実際に製品に適用するためのこうすればできる、という事例集。

民間企業の具体的事例を共有知にする。

・対応内容
データラベルの設計やラベル付けの精度・ツール支援
コーナーケースの分析・モデルの安定性などの検証
・品質アセスメントシート
製品全体の安全性管理と連携させるためのチェックシート

- 具体的な品質管理技術の研究開発
  QA/QC のための具体的な技術の追求。新しい機械学習技術の研究。

- ガイドラインの社会展開：これまでの反響
  ・ビジネスへの適用
  ・ほかのガイドラインへの埋め込み
  ・問い合わせや情報交換

5. まとめ
機械学習の品質は作りこむだけでなく、確認し、説明することも大事である。それを目指すために機械学習品質マネジメントガイドラインを提供している。ガイドラインは、ソフトウェアを作る人だけに使われるのではなく、社会全体で AI を安心して使ってもらうための基準や認証の基盤としての利用も期待している。

6. 質疑応答
＜質問＞
ガイドラインに出てくる言葉が難しい場合など、ガイドラインを読む前段階に読んでおくべき資料、おすすめの資料はあるか。
＜回答＞
現状、そこまで対応できていない。必要性を感じており、問題意識は共有している。

＜質問＞
AI を作るための本は数多くあるが、AI の品質保証に携わるうえで、一通り学んでおく必要があるのか、もしくは最低限の理解でもよいのか。
＜回答＞
ショートカットは可能だと考えている。品質管理の担当者が AI の品質管理をやることになるケースも多い。AI そのものの理解も必要となるが、品質マネジメントの構造を理解していれば、品質管理の培ってきた経験を活かせると思う。今の AI の品質にはそれが大事だと感じている。

＜質問＞
AI のリスクマネジメントは、従来のリスクマネジメントとどのような点が違うのか。どのようなリスクマネジメントプロセスになるのか。リスク分析や対策のプロセスは変わらないのか。
＜回答＞
リスクマネジメントで培ってきたプロセスは出来る限り変えるべきではないというコンセプト。従来のリスクマネジメントプロセスの外側には大きなエコシステムがあるため、モノづくりの違いによる変更だけが必要と考えて、ガイドラインを作成している。一番の変更として、テストのやり方については適応させる必要があるが、それ以外は出来る限りプロセスを変えていない。PoC でリスクを特定し潰せるようになるなど、効率的にリスク特定もできるようになればよい

が、地道なリスクマネジメントプロセスが不足している場合は社会に受け入れられないと思われる。また、ネガティブリスクが重要な分野はしばらくの間変わることが難しいと考えている。ポジティブリスクの分野に関しては、新たな方法があってもよいと思うが、現時点で方法は考えられていない。

＜質問＞
データの均一性とはどういうことか。コーナーケースを考えるとすると、問題のところは局所的だと感じる。
＜回答＞
機械学習のデータの準備の難しい点として、コーナーケースを言語化できる場合とできない場合がある。リスク管理としては言語化できるところは言語化して押さえ、言語化できない場合はコーナーケースをデータ量でカバーしたり、未知のケースにも対処したりする必要がある場合もある。被覆でとってきて点が入っているか確認するだけでなく、均一にデータが入っていることがセーフティからみたときの均一性への要求である。もう一つ、機械学習はいいサンプリングになっていないとうまく学習しない話もあるが、リスク管理からすると言語化できていないリスクを抑えたいというのがメインである。均一性を満たしていても、コーナーケースのデータが 0 ではデータとして不十分のため、両方を満たしていることをチェックすることがポイント。

＜質問＞
AI 開発案件の契約段階の留意点として品管や事業部メンバーが読んでおくべき図書としておすすめはあるか。経産省の契約ガイドラインやディープラーニング協会のハンドブックよりさらに一段ブレークダウンした資料（IPA の非機能要求グレードのような一覧）があれば具体的理解に役立つと考えている。
＜回答＞
ガイドラインの今後の取り組みで拡充していきたい。ガイドラインは、受託の構造に関わらずに共通にいえることを抽出しているため、発注者・受注者で分けて考えた時に間に何があるのか、などの観点では読みづらくなっている認識がある。非機能要求グレードのような形という点では、システムとレベルの対応については機械学習品質マネジメントガイドラインで記述されている。契約段階の切り口では提供できていないが、全体で何をすべきか俯瞰する品質アセスメントシートは役に立つと思う。使う際には、必要な箇所を切り出す必要はある。今後、参考にできるデータを出していきたい。

以上