IMES DISCUSSION PAPER SERIES

情報セキュリティ・シンポジウム（第20回）の模様：
金融分野における機械学習システムの適切な活用に向けて

Discussion Paper No. 2019-J-11

IMES

INSTITUTE FOR MONETARY AND ECONOMIC STUDIES

BANK OF JAPAN

日本銀行金融研究所

〒103-8660 東京都中央区日本橋本石町 2-1-1

日本銀行金融研究所が刊行している論文等はホームページからダウンロードできます。
https://www.imes.boj.or.jp

無断での転載・複製はご遠慮下さい。

備考： 日本銀行金融研究所ディスカッション・ペーパー・シリーズは、金融研究所スタッフおよび外部研究者による研究成果をとりまとめたもので、学界、研究機関等、関連する方々から幅広くコメントを頂戴することを意図している。ただし、ディスカッション・ペーパーの内容や意見は、執筆者個人に属し、日本銀行あるいは金融研究所の公式見解を示すものではない。

１．はじめに

日本銀行金融研究所情報技術研究センター（Center for Information Technology Studies：CITECS）は、2019 年 3 月 27 日、「金融分野における機械学習システムの適切な活用に向けて」をテーマとして、第 20 回情報セキュリティ・シンポジウムを開催した。

近年、金融分野では、AI（artificial intelligence）を利用して既存業務の効率化や新しいサービスの提供を検討する動きが活発化しており、そのコアとなる技術として機械学習（machine learning）が注目を集めている。もっとも、機械学習を実装した情報システム（機械学習システム）には、機械学習に特有の脆弱性が存在するほか、品質（要件の充足度合い）の評価に際して従来のソフトウェア工学に基づく手法だけでは十分に対応できないケースが存在する。こうした状況を踏まえると、今後、金融業界において機械学習システムを活用していくに当たっては、セキュリティ対策や品質保証をどのように実施していくかが重要な課題となる。

今回のシンポジウムでは、そうした観点を踏まえ、機械学習システムのセキュリティと品質にかかる最新の研究動向を講演で紹介するとともに、金融機関が機械学習システムを金融サービスに活用する際の留意点や課題についてパネル・ディスカッションを行った。情報セキュリティ技術にかかわる金融機関の実務者、大学等の研究者、システムの開発・運用に携わる実務者等、約 100 名が参加した。本稿では、下記のプログラムに沿って、シンポジウムにおける議論の概要を紹介する（以下、敬称略、文責：日本銀行金融研究所）[1]。

【第 20 回情報セキュリティ・シンポジウムのプログラム】

- キーノート・スピーチ「金融分野における機械学習システムの適切な活用に向けて」
  横浜国立大学　教授　松本勉
- 講演 1「機械学習システムのリスクとセキュリティ対策」
  日本銀行金融研究所　井上紫織
- 講演 2「機械学習システムの品質評価」
  日本銀行金融研究所　清藤武暢
- 講演 3「機械学習システムの品質保証ガイドラインの動向」
  国立情報学研究所　准教授　石川冬樹
- パネル・ディスカッション「金融機関が機械学習システムを金融サービスで効果的に活用するための留意点や課題」
  モデレータ：横浜国立大学　教授　松本勉
  パネリスト：国立情報学研究所　准教授　石川冬樹
  日本 IBM 東京基礎研究所　部長　細川宣啓
  シティグループ証券株式会社／シティバンク、エヌ・エイ東京支店
  マネージング・ディレクター　オペレーション・テクノロジー　ヘッド
  日高寛公

[1] 文中の各参加者の所属と肩書きはシンポジウム開催時点のものである。また、本稿に示された

２．キーノート・スピーチ「金融分野における機械学習システムの適切な活用に向けて」

松本は、金融分野における AI の活用状況を説明するとともに、機械学習システムにおけるセキュリティと品質保証にかかる課題について、以下のとおり発表した。

（１）金融分野における AI の活用状況

近年、金融分野における AI の活用が広がりをみせている。金融情報システムセンターのアンケート調査によると、AI を「導入中」、「準備段階」または「検討中」と回答した金融機関の割合は、2015 年度は 10%程度であったが、2017 年度には 50%程度にまで増加している（金融情報システムセンター［2018］）[2]。これらの金融機関における AI 活用の目的をみると、社内の過去情報の有効活用、チャットボットによる顧客応対の向上、マーケティング分析の高度化、不正取引の検知、審査業務の支援等、多岐にわたっている。主要なクラウド・ベンダーが MLaaS の提供を開始するなど[3]、金融機関が AI を活用しやすい環境が整いつつあることも、こうした動向の背景の 1 つになっているとみられる。

（２）金融サービス等における機械学習システムの活用事例

機械学習システムにはさまざまな構成が想定される。今次シンポジウムでは、比較的シンプルな構成のシステムを前提に議論を行う。まず、訓練データを学習アルゴリズムに適用して訓練を実施し、判定・予測を行うソフトウェア（判定・予測モデル）を生成する。機械学習システムを利用する際には、判定等を行うデータをシステムに提示する。そうすると、そのデータが判定・予測モデルに入力され、そのモデルの出力に基づいて判定・予測結果が示される。

金融機関における顧客応対向けのチャットボットの利便性を向上させるために機械学習システムを活用する事例では、照会対応のノウハウや金融商品に関する情報等を訓練データとするほか、顧客からの照会情報を判定・予測モデルに入力し、その出力に基づいて顧客への回答を提示する。審査業務の効率化を企図して機械学習システムを活用する事例では、金融機関が有する顧客の信用度にかかる情報等を訓練データとして判定・予測モデルを生成したうえで、顧客が提

意見はすべて発言者たち個人に属し、その所属する組織の公式見解を示すものではない。

[2] 金融情報システムセンターのアンケート調査の対象期間は 2017 年 4 月 1 日～2018 年 3 月 31 日、有効回答機関数は 679 社（調査対象機関数：683 社）。このアンケートにおける「AI」は、「自然言語処理、機械学習、ロボティクスといった要素技術を用いるもの」と定義されている。

[3] MLaaS（Machine Learning as a Service）は、AI／機械学習による判定・予測等をクラウドによって提供するサービスの総称。

示した審査に必要なデータを判定・予測モデルに入力し、その出力に基づいて審査結果を生成する。また、不正な金融取引の検知に機械学習システムを用いる事例では、まず、金融取引にかかる既存のデータを訓練データとして判定・予測モデルを生成する。そのうえで、不正か否かの判定の対象となる取引のデータを判定・予測モデルに入力し、その出力に基づいて当該取引が不正か否かを判断する。

（３）機械学習システムにおけるセキュリティと品質保証の課題

新しい技術を金融分野で活用する際には、予めそのセキュリティ上のリスクを考慮し、対策を適切に講じることが求められる。この点、機械学習システムには、機械学習特有のリスクが存在することに留意が必要である。

機械学習システムの品質をどう確保するかも重要な課題である。従来の IT システムでは、その振舞いを概ね把握可能であり、一定の品質を保証することができた。一方、機械学習システムでは、その振舞いを事前に把握することが困難であり、従来のソフトウェア工学による手法のみでは対応が難しい場合がある。その結果、有用な技術であっても品質を保証できず、技術の差別化を図ることが困難となりうる。また、想定外の事故が発生した際に無過失であることを説明できないリスクも発生しうる。

こうした課題に対応するための取組みが各所でなされている。例えば、産業技術総合研究所サイバーフィジカルセキュリティ研究センターでは、産業界と連携しつつ、機械学習システムの品質にかかる基準の策定と品質を確認・検査する手法等の研究開発を進めている。

今次シンポジウムでは、機械学習システムのセキュリティ対策や品質保証にかかる研究開発動向を把握するとともに、金融サービスにおいて機械学習システムを適切に活用するうえでの留意点や課題等を議論していきたい。

３．講演１「機械学習システムのリスクとセキュリティ対策」

井上は、井上・宇根［2019］に基づき、機械学習システムに特有の脆弱性を説明するとともに、金融サービスで活用される機械学習システムの事例を対象に、想定されるリスクとセキュリティ対策のあり方について、以下のとおり説明した。

（１） 機械学習システムに特有の脆弱性

機械学習システムにおいては、通常の IT システムと同様に、そこで取り扱われるデータ（訓練データ等）やシステムを構成するソフトウェア（学習アルゴリズム、判定・予測モデル）等が保護の対象となり、それらの機密性・一貫性・可用性の確保がセキュリティ目標となる。これを実現する手段として、通信路上の

データを暗号化する、各エンティティが有するデータやソフトウェアへのアクセス制御を実施するなどの一般的なセキュリティ対策を活用できる。

もっとも、機械学習システムの場合、そうした対策のみでは対応が困難な脆弱性が存在する点に留意する必要がある。例えば、判定・予測モデルの入出力から、訓練データや判定・予測モデルにかかる情報が漏洩する場合がある。また、判定・予測モデルへの入力に微小なノイズが加わると、その（ノイズ付きの）入力に対して誤った判定・予測が出力される場合があるほか、訓練データにノイズが加わると、判定・予測モデルの精度が大きく低下しうる。

各脆弱性への対応の要否は、脆弱性が顕在化した場合の影響の大きさに基づいて判断することになる。訓練データや判定・予測モデルにかかる情報の漏洩に対しては、訓練データの機密性や判定・予測モデルの資産性が判断材料となる。判定・予測の精度低下に対しては、誤った判定・予測によるリスクの多寡が判断材料となる。

（２）金融機関で活用される機械学習システムのリスク

顧客の照会にチャットボットを用いて応答する機械学習システムにおいて、訓練データとして公開情報（一般的な照会とそれに対する回答等）が用いられる場合、訓練データの機密性は低く、判定・予測モデルの資産性も高くないと想定される。一方、判定・予測モデルの精度低下により、多くの顧客が誤った回答を受信すれば、金融機関のレピュテーションが低下する可能性がある。こうしたリスクを無視できないならば、何らかの対策を講じる必要がある。

スマートフォン・アプリ等を用いて顧客の信用度を評価するシステムでは、訓練データとして、年齢や収入、勤務先等、個人情報が含まれる場合があるほか、システム自体の資産性も高いと考えられる。したがって、訓練データや判定・予測モデルの漏洩への対策を講じる必要がある。判定・予測モデルの精度低下についても、顧客の信用度の不適切な評価は本来よりも緩い条件での貸出の実行等に繋がりかねないため、対策の検討が求められる。

（３）セキュリティ対策のあり方

訓練データ等にかかる情報の漏洩に対しては、判定・予測モデルの出力を変換して提示したり、判定・予測の確からしさを示す値（確信度）を丸めて提示したりするなど、推定に必要な情報を攻撃者が入手できないようにすることが考えられる。また、個人情報が訓練データに含まれる場合、個人を特定できないよう

に加工するなど、推定時の影響を軽減することも対策方針として挙げられる。さらに、訓練データの推定が困難な学習アルゴリズムの採用も有用である[4]。

判定・予測モデルの精度低下に対しては、誤った判定・予測を誘発する入力等を事前に検知可能な判定・予測モデルを別途生成して利用するという手法が挙げられる。また、誤った判定・予測を誘発する入力の影響を低減させる学習アルゴリズムを採用する手法も対策として挙げられる[5]。

攻撃手法や対策手法は日々進化している。金融機関が機械学習システムのリスクを把握しセキュリティ対策を検討する際には、最新の研究動向を注視することが必要である。また、いったん導入したセキュリティ対策に関しても、その効果が失われていないかを定期的に確認し、必要があれば対策の内容を見直すことが肝要である。

４．講演２「機械学習システムの品質評価」

清藤は、機械学習システムの品質評価にかかる課題とそれへの対策手法の研究動向を説明したうえで、金融分野で活用される機械学習システムの事例を対象に、品質を評価し保証する際の留意点や課題について、以下のとおり説明した。

（１）機械学習システムの品質評価にかかる課題

機械学習システムの主たる機能は判定や予測の実行であり、判定等を行うデータに対して、期待される判定・予測結果が一定以上の確率で得られることが求められる。期待される判定・予測結果が一定以上の確率で得られるという特性は、機能正確性と呼ばれ、従来のソフトウェアにおける品質特性の 1 つとして知られている。機械学習システムの実用に供する際にも、こうしたビジネス要件の充足度合いを事前に確認しておくことが望ましい。ただし、機械学習システムについて、どのような品質特性を設定するか、設定した品質特性をどう評価するかが課題となっている。

従来のソフトウェアは、通常、人間が期待する入力と出力の関係を定式化したうえで、処理の流れを明確化して生成される。したがって、その品質の評価は、実際の入出力が定式化した関係と整合的か否かを確認することによって行われ

---

[4] 例えば、訓練データにノイズを付加したうえで判定・予測モデルを生成する手法や、訓練データを複数のデータセットに分割し、各データセットを用いてそれぞれ判定・予測モデルを生成した後、それらの（複数の）判定・予測モデルから最終的な判定・予測モデルを生成する手法が挙げられる。これらの手法については、井上・宇根［2019］を参照されたい。

[5] 具体的な手法として、防御的蒸留（defensive distillation）や敵対的学習（adversarial training）が知られている。防御的蒸留は、いったん生成した判定・予測モデルに訓練データを入力してその出力を得た後、訓練データとその出力のペアを新たな訓練データとして用いて判定・予測モデルを生成する手法である。敵対的学習は、誤った判定・予測を誘発する入力を収集し、それらを訓練データとして判定・予測モデルを生成する手法である。

る。こうした評価の方法論がソフトウェア工学として長年研究されているほか、JIS X 25010 等の標準規格が策定されている。一方、判定・予測モデルについては、訓練データと学習アルゴリズムを用いて直接生成され、入力と出力の関係が必ずしも明確ではないため、従来の方法論による評価が難しい場合がある。

（２）判定・予測モデルの品質評価手法の研究動向

近年、こうした課題に対応するための研究が活発化している。入力と出力の関係が把握できない場合であっても、判定・予測モデルの特性を利用することによって、品質評価に用いるデータ（テストデータ）を生成する手法が複数提案されている。例えば、判定・予測モデルに入力するテストデータの一部を変化させた場合の、出力の変化の方向性が明確であるケースがある。こうしたケースでは、入力を変化させたときの実際の出力における変化が予想した方向か否かを確認することで、機能正確性を評価することが考えられる（こうした手法はメタモルフィック手法と呼ばれる）。

（３）金融分野で活用される機械学習システムの品質評価

顧客の照会にチャットボットを用いて応答する機械学習システムにおいては、音声やテキストによる照会に対して、従来のシステムと同程度に正確な情報をより効率的に回答することが主たる目的となる。例えば、正確な照会結果を提示するというビジネス要件に関しては、顧客からの照会内容が一部変化した際にチャットボットの回答がどう変化するかをメタモルフィック手法等によって評価することが考えられる。また、照会結果の根拠を顧客に提示するというビジネス要件を設定する場合には、その根拠を人間が解釈しやすい形式で出力する学習アルゴリズム等を活用することが有用である。根拠の提示が困難であるならば、金融機関の職員への直接の照会・問合せを推奨することが考えられる。

スマートフォン・アプリ等を介して顧客の信用度を評価する機械学習システムにおいては、顧客の属性等に応じて、従来のシステムと同程度に正確かつ公平な信用度をより効率的に評価することが求められる。特に、信用度評価の結果が公平性を満たすこともビジネス要件の 1 つとして設定されうる。特定の顧客が不利益となる（公平性を損なう）データが訓練データとして使用されていないことを確認する、あるいは、判定・予測モデルの出力の偏りを検知・排除する手法（フェアネス・アウェア・データマイニングと呼ばれる）を活用するなどの対応が挙げられる。

今後、機械学習システムの品質保証を実現していくうえで、判定・予測モデルの品質評価にかかる新しい手法を取り入れることを視野に入れつつ、運用面での対応を検討していくことが重要である。また、最近では、機械学習システムの

品質保証にかかるガイドラインの策定が進められており、それらを活用することも有用であろう。

５．講演３「機械学習システムの品質保証ガイドラインの動向」

石川は、機械学習システムの開発や品質保証の現状を説明したうえで、機械学習システムの品質を評価するためのガイドライン策定に向けた取組みについて、以下のとおり説明した。

（１）機械学習システムの特性と品質評価

通常の IT システムは、計算や判断を行うための知識・規則を人間が決定し、それを実現するプログラム（判定・予測モデル）を作成するという流れで開発される（演繹的システム開発）。一方、機械学習システムは、知識・規則を人間が決定するのではなく、訓練データから獲得してプログラムを作成するという流れで開発される（帰納的システム開発）。そのため、開発された機械学習システムの特性や限界をエンジニアですら把握困難な場合がある。

こうした課題に対処するためには、工学的な視点が不可欠である。日本ソフトウェア科学会では、2018 年に「機械学習工学研究会」を設置し、技術者や研究者による研究発表や議論の場を提供している。先般、機械学習を業務に用いている技術者等を対象にアンケートを実施したところ、機械学習システムの開発における「顧客との意思決定」や「テスト、品質の評価・保証」において、通常の IT システムの開発とは異なる対応が必要であるとの回答が多かった。機械学習システムの場合、どれだけテストを実施すれば十分かが不明確な場合が多く、不確実性が高いシステムを発注者やユーザーがどの程度受け入れられるかが大きな課題となっている。上記のアンケート結果はこうした問題を反映しているといえる。

（２）品質保証にかかるガイドライン

機械学習システムの主要ベンダーでは、品質保証にかかる原則や指針を独自に策定している。そうした指針においては、例えば、機械学習システムの入出力の傾向を分析し、問題として顕在化する前に不適切な入出力を検知することや、入出力の範囲や分布が予想と合致しているかを評価すること等が盛り込まれている。これらは、機械学習システムの振舞いが事前に想定したとおりとなっているか否かをチェックするという考え方である。わが国においては、AI プロダクト品質保証コンソーシアム（QA4AI）が、機械学習システムの品質保証に関する技術の調査研究やその体系化、機械学習システムの活用の支援等を行っており、品質保証に関するガイドラインの策定も進めている。

QA4AI のガイドラインでは、機械学習システムの品質を評価するうえで重要となる 5 つの概念を定める予定である。まず、①訓練データと判定・予測モデルの入力の整合性等の「データの品質」や、②正解率等の性能、学習アルゴリズムの妥当性等の「判定・予測モデルの品質」が挙げられる。また、③機械学習システムの価値、インシデントの発生度合いとリスク、説明可能性等の「システム全体の品質」、④品質向上のためのチェックを行う周期の短さ等の「プロセスの迅速さ」が挙げられる。さらに、⑤顧客が機械学習システムへ期待する程度や品質・リスクに関する理解の度合い等の「顧客による期待の高さ」も重要な要素であることから、ガイドラインに含まれる見込みである。

（３）今後の品質保証にかかる研究の見通し

機械学習システムの品質保証にかかる研究開発のスピードは非常に速くなっており、新しい品質評価の手法も次々と提案されている。ガイドラインの策定に向けた取組みは、QA4AI 以外の組織や団体においても活発化しているが、こうした状況を踏まえると、当面、これらガイドラインは頻繁に更新されるであろう。今後、機械学習システムの活用を検討していく際には、こうしたガイドライン等を参照しつつ、機械学習システムの特性や限界を十分理解するとともに、活用の目的についてビジョンを明確にすることが重要である。

６．パネル・ディスカッション「金融機関が機械学習システムを金融サービスで効果的に活用するための留意点や課題」

パネル・ディスカッションでは、機械学習システムにおける品質保証、機械学習システムの活用の範囲や開発形態、機械学習システムに関連する国際標準化の動向等について、以下のとおり議論を行った。

（１） 機械学習システムにおける品質保証

モデレータの**松本**は、まず、金融分野で活用される機械学習システムにおける品質の考え方について、パネリストに意見を求めた。**細川**は、金融サービスに関連する IT システムの品質として、長期間の安定稼動の実現が重視されるケースが多いとの見方を示したうえで、機械学習システムにおいても、重大なインシデントが発生しないこと等を品質として評価していくことになるのではないかと説明した。

これを受けて、**松本**は、インシデントが発生しないことを保証するための具体的な対策についてパネリストに問うた。**石川**は、チャットボットを用いた金融取引のサービスのケースを取り上げ、例えば、顧客に無断でチャットボットが資金移動を行わないように、資金移動の際には必ず顧客が最終確認を行う設計とす

るなどの対応が考えられると述べた。日高は、機械学習システムによる判断の結果が一定の基準を超えた場合に、そのシステムが自動的に停止する機能を実装することや、人間がシステムを強制的に停止できる仕組みを導入することが考えられると説明した。これを補足して、細川は、機械学習システムでは、自分の動作を自律的に停止させる機能を実装することは難しいとの研究結果もあり、機械学習システムを停止させるためには人間の介入が必要であるとの見方を示した。また、現在の機械学習の技術では、すべての判断の基礎となる「方針」をも適切かつ自律的に学習することが困難であると指摘した。

石川は、誤った判断によるインシデントの発生は、人間が判断する際にも起こりうるにもかかわらず、機械学習システムにおける対策の必要性ばかりが過剰に意識されてはいないかとの見方を示した。これに対し、日高は、金融機関では、人間の判断についても誤りがないかを慎重に確認し、問題がある場合には業務のプロセスを停止する体制を整備していると説明した。そのうえで、金融機関における業務のプロセスをすべて機械学習システムによって実現するとすれば、既存のプロセスと同程度の安全性や信頼性を維持するための対策を講じる必要があると述べた。

（２） 機械学習システムの活用の範囲

イ．既存の業務の代替可能性

松本は、機械学習システムの動作によってインシデントが発生した場合、それを検知して停止させる別のシステムを実現することができれば、すべての業務を機械学習システムで代替することができるかという点について、パネリストに見解を求めた。

細川は、そうした代替可能性を考えるうえで、機械学習システムの機能だけでなく、そのシステムを用いたサービスを顧客がどう受け止めるかという視点も重要になるとの見方を示した。例えば、銀行の窓口業務において、すべての応対を機械学習システムが行う場合よりも、銀行員が笑顔で接した場合の方が顧客満足度が高く、結果として金融商品の売上が増えるというケースが考えられると説明した。日高は、こうした意見に賛意を示したうえで、現時点では、人間が判断する際の補助機能として機械学習システムを活用することが現実的であると述べた。さらに、すべての業務を機械学習システムに代替させる場合、トラブル発生時の責任の所在が不明確になる可能性があると指摘した。

ロ．誤った判断にかかる責任の所在

フロア参加者から、機械学習システムを人間の判断の補助機能として活用する場合であっても、その出力が引き金となって誤った判断が発生するケースが

想定されるが、この場合の責任の所在をどう考えればよいかとの質問がパネリストに寄せられた。

これに対して、日高は、金融機関内部における責任の所在については、機械学習システムの導入を判断した部門を中心に、システム部門やコンプライアンス部門等、社内の複数の部署が責任を負うことになる可能性があると回答した。石川は、発注者と開発者の間の責任分担について、機械学習システムを開発する際には、発注者と開発者が緊密に連携しながら共同で作業を進めていくケースが多いとしたうえで、発注者と開発者が共同で責任を負うことになる場合があるとの見方を示した。

また、細川は、機械学習システムのリリースを優先した結果、最終的にビジネス要件が満たされなかったり、リリース後に性能劣化が発生したりしたケースでは、発注者と開発者の責任分界点を定めることが難しくなるとの見解を述べた。

ハ．機械学習システムを採用するか否かの判断の基準

フロア参加者から、機械学習システムを採用した結果、誤った判断によって損害が発生したという事例が生じうる一方で、社会全体として捉えればトータルの損害よりも利益が上回るという状況がありうるが、こうした場合に機械学習システムを活用すべきか否かをどのように判断すればよいかとの質問がパネリストに寄せられた。

これに対して、細川は、機械学習システムの活用を社会的な潮流と捉え、リスクも含めて積極的に活用するという考え方がある一方、誤った判断によるリスクを重視して活用しないという考え方もあると述べたうえで、個人がそれぞれのリスク選好に基づき判断していくことになるとの見解を示した。日高は、機械学習システムを活用することによるリスクを踏まえつつ、どのような用途で活用していくかを予め明確にしておくことが必要であり、それに基づいて判断することになると述べた。

また、石川は、機械学習システムが少子化に伴う人手不足等の社会問題を解決する手段の 1 つとして注目されており、その積極的な活用が社会のトレンドとなっているとの見方を示した。また、誤った判断を下すリスクは人間の場合でも存在するとしたうえで、機械学習システムにおいてそうしたリスクをいかにして低下させるかを考えることが重要であると述べた。

ニ．誤った判断のリスクを低下させる手法

松本は、機械学習システムが誤った判断を下すリスクを低下させるうえで、発生しうるインシデントを予め網羅的に洗い出すことができれば有用であるとの見方を示し、そうした洗出しの実現可能性についてパネリストに意見を求めた。

石川は、インシデントの洗出しは可能であるとしたうえで、実際に自動運転の分野においては、そうした対応の必要性を指摘する声が聞かれていると説明した。

日高は、複数の機械学習システムを用いて、異なる観点から総合的に判断することが有用ではないかとの意見を述べた。これについて、細川は、そうした手法の 1 つとしてアンサンブルと呼ばれる手法を紹介し、判断にかかる責任の所在が不明確になるという課題があると補足した。そのうえで、融資判断を行う機械学習システムを例として取り上げ、システムが、融資可能な金額だけでなく、融資に伴うリスクに関する情報を出力するなど、判断に関連する付加的な情報を出力する機能を実現する方が有用ではないかとの見方を示した。石川は、金融分野では、判断を行う際に合理性や論理性が重視される傾向にあることを指摘したうえで、判断の根拠が明確でない機械学習システムを複数用いたとしても、判断にかかる合理性や論理性を明確にすることは困難であるとの見方を示した。そのうえで、合理性があると認められる判断の候補を機械学習システムが複数提示し、最終的な判断を人間が選択するというアプローチがありうると述べた。

また、細川は、機械学習における重要な要素として、現時点では（学習の対象となる）データが中心的な位置を占めているが、今後、人間とのインタフェースも重要になってくるとの見解を示し、例えば、曖昧な回答や（二者択一ではない）中間的な回答等、より高度な意思決定を支援する機械学習システムをどう実現するかといった点も重要な課題であると述べた。

（３） 機械学習システムに関する標準化等の動向

松本は、米国の国防高等研究計画局（DARPA）におけるプロジェクトや、ISO/IEC JTC1/SC42 における AI や機械学習にかかる国際標準化の動きについて紹介し、こうした国際的な議論の動向についてパネリストに発言を求めた[6,7]。

細川は、近年、機械学習システムの品質評価において、人間が不快に感じるか否かといった人間の感性に関わる要素が重要であるとの認識が広がりつつあるとしたうえで、公平性（fairness）、説明可能性（accountability）、透明性（transparency）を充足する機械学習システムの開発や評価に関する研究開発が活発化しているとの見方を示した。国際標準化については、ISO/IEC JTC1/SC42 において、AI の信頼（trustworthiness）に関する検討を行う作業部会が設置され、同作業部会では、AI の信頼に関連する問題の抽出・整理や事例の収集等を行っており、収集した

[6] 米国防総省の国防高等研究計画局（Defense Advanced Research Projects Agency：DARPA）は、判断理由を説明可能な AI である「eXplainable AI（XAI）」の開発に向けて、2017 年から研究所や企業、大学が複数参画するプロジェクトを推進している。

[7] ISO/IEC JTC1/SC42 は、情報技術に関わる国際標準化を担う ISO/IEC JTC 1 に設置された人工知能にかかる分科委員会。

事例をもとに国際標準にかかる審議を進めていると説明した。

松本は、QA4AI において策定が進められている品質保証ガイドラインを、今後、国際的にどう紹介していくかについて**石川**に問うた。**石川**は、同ガイドラインを可能であれば国際的な場で発表していきたいとの考えを示した。また、中国や米国等では、運用開始時点で十分な品質の確保を要求されるわが国と異なり、問題が発生する都度、改善しながら機械学習システムを運用するケースが多いことを紹介したうえで、ガイドラインの策定等を通じて、わが国で通用する品質保証を確立することができれば、他の産業分野と同様に、高い品質が機械学習システムの分野における国際競争力の面でのわが国の強みとなるのではないかとの見解を示した。

（４） 機械学習システムの開発形態

松本は、訓練データとして機微なデータを用いるケースにかかる対応について、パネリストに意見を求めた。**日高**は、金融機関が有するデータは機密性の高いデータが多く、それらを社外のベンダー等に渡して判定・予測モデルを構築することは難しいとしたうえで、自社ではクラウドを活用しつつ自ら開発していると述べた。**石川**は、判定・予測モデルの生成を外注したいというニーズは少なくないものの、セキュリティ上のリスクに配慮した結果、訓練データを十分に用意できず、生成した判定・予測モデルの精度が低く、実用に耐えないといったケースもあると述べた。

フロア参加者から**日高**に対し、自社で開発するには十分なデータを有しない企業に対して、開発済みの判定・予測モデルを提供することは可能かとの質問があった。**日高**は、パートナーシップの契約を締結し、共同事業として検討を進めることができれば、可能ではないかとの見解を示した。**石川**は、比較的少ない訓練データを活用するケースとして、製造業の不良品検出の分野では、自社の数十件の訓練データを既存の判定・予測モデルに適用することによって、自社の訓練データに適合した判定・予測モデルを生成することができたという事例があると説明した。**細川**は、欧州では、EU 一般データ保護規則によってデータ保護の動きが広がっているものの、本人の同意があれば、個人から取得したデータの二次利用や転売が認められており、自社で大量のデータを保有していない企業であっても、他社からデータを購入して活用することが可能であると説明した[8]。また、一部の企業がサーバ上で提供している学習アルゴリズムを用いることで、判定・予測モデルを自社開発することもできると述べた。

---

[8] EU 一般データ保護規則（General Data Protection Regulation：GDPR）は、欧州における個人情報の保護および個人データの自由な流通を目的として、個人データの収集や保管に関するルールを定めた EU 域内の規則であり、2018 年 5 月に施行された。

松本は、判定・予測モデルを他社に提供することに伴うセキュリティ上のリスクに関して、パネリストに見解を求めた。**石川**は、機械学習システムを提供するクラウドサービスの入出力から、その判定・予測モデルの推定に成功した事例が報告されているなど、一定のリスクが存在するとの見解を示した。もっとも、正規の判定・予測モデルに電子透かしを埋め込んでおくことで、そのモデルを推定して生成された不正な判定・予測モデルを判別する技術等、こうした攻撃に対する対策も研究されていると説明した。

（５） 機械学習システムを適切に活用するために

最後に、**松本**は、機械学習システムを適切に活用していくために重要となる事項について、各パネリストにコメントを求めた。

日高は、判定・予測モデルの仕組みや機械学習システムの特性を十分に理解したうえで、導入を検討することが重要であると述べた。**細川**は、機械学習システムを活用するうえで重要なことは、データから有益な知識をいかに抽出するかであり、そうしたスキルの蓄積が重要であるとの見方を示した。**石川**は、AI という用語に惑わされることなく、データに内在する知識を抽出する技術という本質を理解して活用していくことが重要であると述べた。

そのうえで、**松本**は、今後、金融機関が、本日のシンポジウムにおける講演やパネルでの議論を参考にしつつ、セキュリティや品質保証により一層配慮しながら、機械学習システムを適切に活用していくことが望ましいと述べて、パネルを締め括った。

以　上

【参考文献】

井上紫織・宇根正志、「金融分野で活用される機械学習システムのセキュリティ分析」、金融研究所ディスカッションペーパー2019-J-1、日本銀行金融研究所、2019 年

金融情報システムセンター、「平成 30 年度金融機関アンケート調査結果」、『金融情報システム』No.345、金融情報システムセンター、2018 年、1～200 頁