山本レポート：システム安全性向上―世界の最前線（4）

AIシステムの境界

国立大学法人 名古屋大学
大学院 情報学研究科
教授　山本 修一郎
博士（工学）

AIシステムも人間が作る人工物であるから、AIシステムには外部との境界がある。以下ではAIシステムの境界がもたらす限界について説明する。

AIシステムのコンテクスト

図1に示すようにAIシステムにもAIシステムが動作するコンテクストがある。AIシステムがコンテクストと相互作用する境界がAIシステム境界である。AIシステムが導入されるあらゆるコンテクストを想定できないからある開発環境のコンテクストでAIシステムが開発される。したがってAIシステム開発時のコンテクストと、AIシステム導入時のコンテクストの不適合があれば、AIシステムは故障する。新たなコンテクストに適応するためには、そのコンテクストに適合する機能拡張がAIシステムに求められる。

学習の限界

アマゾンがAIを活用した人材採用システムの運用を断念した[1]。アマゾンでは、これまでの10年間に提出された履歴書のパターンを学習させた結果、AIシステムが男性を採用するのが好ましいと認識することになった。この理由は技術職のほとんどが男性からの応募だったことにある。このため、AIを活用した人材採用システムの採用をアマゾンは断念した。

うまく動かないAIのことを「ポンコツAI」と呼ぶそうだ[2]。「ポンコツAI」の症状として、定番のわずかな種類しか予測できない需要予測システムや、新しい製品が出ると価格が当たらなくなる価格予測システムが紹介されている。「ポンコツAI」が生まれる原因は、適切で十分な学習データがないからだ。AIだから精度の高い需要予測ができるはずだというのは幻想でしかない。

これらの例はいずれも、過去のデータに基づいて将来を予測する機械学習型AIの本質的な限界だ。十分な学習データが用意できないのであれば、最初からそのような業務分野にAIシステムを導入すべきではないことに気づくべきだろう。

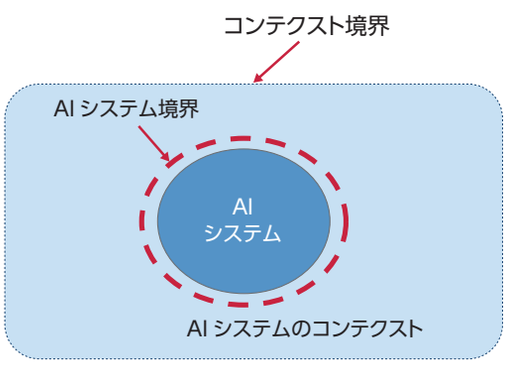

図1　AIシステムのコンテクスト

適応の限界

ヒト型ロボット「Pepper（ペッパー）」の法人向け導入が2015年10月にはじまってから3年が経過した。日経xTECHによれば、ペッパーのレンタル契約（3年）の更改を予定する企業が15%にとどまるとのことだ[3]。更新契約を解約する理由は、コスト（月額約5万5千円）、効果（機能）、故障だそうだ。コストと効果は表裏一体だ。85%の企業にとって、ペッパーには、費用に見合うだけの機能がなかったということだ。

企業によるロボット導入では多様な利用環境に個別に適応する必要がある。そのためには、新たな機能開発やカスタマイズがロボットに必要になるのは当然である。あらゆる環境に適応できる機能を備えた万能なロボットなど、最初からいるはずがない。

また、故障の原因については、カメラや赤外線センサー、ジャイロセンサー、タッチセンサー、ソナー、マイクなどの多くの装置を内蔵していることが考えられる。多くの装置を持つほど、故障発生確率が高くなる。さらに、これらのセンサー信号を制御するソフトウェアが現場環境

に対応できていない可能性もある。たとえば、赤外線センサーは太陽光による温度変化などの影響で誤作動する可能性がある。センサーがどのような範囲であればペッパーが適切に動作できるかという境界条件を確認しておく必要がある。センサーの境界条件が明示されたとしても、次の問題は、動作環境に応じて境界条件を拡張できるように制御ソフトウェアを拡充するための新たな開発が必要になることだ。

説明可能性の限界

機械学習では、どんな根拠で判断にいたったのかを論理的に説明できないという「ブラックボックス問題」がある[4][5]。このため、意思決定の根拠を説明できるXAI（eXplainable AI、説明可能なAI）が必要になっている。

たとえば、社会保障制度の利用申請で、AIシステムが申請者の条件を自動判定して、審査請求を却下したとしよう。このとき、却下理由を問い合わせても、理由を説明できないのでは、申請者は納得できないだろう。一方で、説明可能AIに対して、人間でも意思決定理由を説明できないことが多いという反論もある[6]。しかし社会的に広く利用されるシステムでは説明責任が求められるのは当然だ。「AIシステムが自動判定したからわからない」ということでは社会的に許されない。

たとえば、2018年5月に施行された欧州連合（EU）の一般データ保護規則（GDPR, General Data Protection Regulation）では、ユーザの「説明を受ける権利」が確立された。GDPRでは、アルゴリズムがユーザについて下した決定について、ユーザが説明を求めることができる。したがって、GDPRの下では、AIシステムは意思決定結果をユーザに説明できなくてはならない。ユーザに関する法令を遵守する取り組みがAIシステムにも求められている。

複雑な処理を自動化するAIシステムの場合、意思決定結果について説明するためには、意思決定のプロセスを監視する方法も必要になる。たとえば、自動運転システムでは、危険な状況では自動運転を中断して、運転を代わってもらうように運転者に通知する。なぜ、自動運転を中断するのかを説明できなくては、通知された運転者が適切に対応できない可能性がある。また、自動運転を中断する原因は多様であるから、自動運転中止判断の意思決定プロセスを明らかにして逐次監視する必要がある。

相互運用性の限界

個別的に開発された異なる組織のAIシステムが同じコンテクスト環境で動作する場合、コンテクスト環境を共有するAIシステム間で相互接続性が必要である。また、異なるAIシステムが相互接続するためには意思決定の透明性が必要である。人間に対する説明だけではなく、AIシステム間での透明性も必要になる。

たとえば、自動運転システムが公道という共有コンテクスト環境で利用される場合、自動運転システムを提供する各社が自動運転の意思決定方式を公開する必要がある。たとえば、2台の自動運転車が対向するようにして同じ道路を走行しているところに、新たに緊急車両が侵入した場合、自動運転車間で緊急車両への対応についての意思決定結果が事故に繋がらないようにしなければ安全とは言えない。通常の自動車は運転者が日本の道路交通法を遵守する必要がある。自動運転システムについても、対向車の挙動を考慮して、道路交通法に基づく緊急車両を考慮した意思決定が求められる。

まとめ

AIシステムには、上述したように、学習の限界、適応性の限界、説明可能性の限界、相互運用性の限界がある。したがって、AIシステムの導入を成功させるためには、これらの限界を克服するために、AIシステムの境界を明らかにするとともに、コンテクスト（利用環境）への適応策を考慮する必要がある。

【参考】
[1]BLOGOS, 焦点：アマゾンがＡＩ採用打ち切り、「女性差別」の欠陥露呈で, http://blogos.com/article/331473/
[2]日経BP, ポンコツAI回避術, https://nkbp.jp/2RH2u6y
[3]日経xTECH、ペッパー 4年目の真実ーさらばペッパー、契約更改を見送った企業の本音, https://tech.nikkeibp.co.jp/atcl/nxt/column/18/00466/101000001/
[4]Esther Shein, The dangers of automating social programs, Communications of the ACM: Volume 61 Issue 10, pp.17-19, 2018
[5]Zachary C. Lipton, The mythos of model interpretability, Communications of the ACM: Volume 61 Issue 10, pp.36-43, 2018
[6]George Nott?Googleのリサーチ責任者、「説明可能なAI」の価値に疑問符, Computerworld, 2017/06/28, https://tech.nikkeibp.co.jp/it/atcl/idg/14/481542/062800387/