機械学習*と公平性に関する声明

2019 年 12 月 10 日
人工知能学会 倫理委員会
日本ソフトウェア科学会 機械学習工学研究会
電子情報通信学会 情報論的学習理論と機械学習研究会

私たち、機械学習の技術及び応用を研究している研究者コミュニティ（人工知能学会 倫理委員会、日本ソフトウェア科学会 機械学習工学研究会、電子情報通信学会 情報論的学習理論と機械学習研究会、以下「私たち」と表記します）は、機械学習の利用が公平性に与える影響を重く捉え、私たちがこの問題にどのように対処していくべきかをここで社会一般の皆様と共有したいと考えます。

重要なのは次の 2 点です；

(1) 機械学習は道具にすぎず人間の意思決定を補助するものであること
(2) 私たちは、公平性に寄与できる機械学習を研究し、社会に貢献できるよう取り組んでいること

背景

機械学習の不適切な利用が公平性を欠く可能性に対する懸念が高まっています。例えば 2018 年 10 月、Amazon.com は採用時に補助的に利用していた機械学習システムが、女性に対して不利益に働くことに気づき、このシステムの利用を停止したことをロイターが伝えました[1]。より広くは、機械学習が利用者の意図に応じて、あるいは意図によらず不公平を助長してしまうことがあると指摘されています（[2]をご覧ください）。

1. 機械学習は道具にすぎません

機械学習はあくまでも道具にすぎず、その使い方を定めるのは人間です。機械学習は人類社会の繁栄に大きく貢献できる可能性を秘めているとともに、不適切な利用をすれば人類社会の利益に反する可能性もあります。機械学習は過去の事例に基づいて未来を予測しますから、偏りのある過去に基づいて予測する未来は、やはり偏りのあるものになりかねません。もし、過去と異なる「あるべき未来」を求めるのであれば、機械学習による予測や判断が公平性を欠くことがないように人間が機械学習に注意深く介入する必要があります。

同時に、「何が公平か」については、科学技術や工学だけの問題ではなく、現在の人類社会が何を求めているか、という価値観の問題抜きには語れません。機械学習という「道具」を正しく使うためには、それが「公平性」という私たち人類社会の価値観に対して、どのような影響を与えるかを正しく理解し、そのリスクを評価し、方策について合意しなければならないのです。この点は、私たちだけではなく、機械学習に携わる技術者や利用者、経営者、そして組織や社会の全体が把握し向き合っていく必要があります。

2. 私たちは機械学習で公平性に寄与します

私たちは、機械学習の利用が社会の不利益になってはならないと考え、この問題を解決するために、行動指針と技術開発の双方から真摯に取り組んでいます。IEEE Ethically Aligned Design では機械学習の不適切な利用ないしは誤用、悪用を戒め、その対策を具体的に記述しています[3]。人工知能学会では、自らの社会における責任を自覚し、社会と対話するために、学会会員の倫理的な価値判断の基礎となる倫理指針を 2017 年に定めました[4]。我が国社会の様々なステークホルダ（その一部は、私たちでした）が集まって、高度な情報技術を社会でどのように使っていくべきかを議論し、その結果が、内閣府「人間中心の AI 社会原則」として 2019 年 3 月に公開されました[5]。その基本理念の 1 つは多様性と包摂であり、高度な情報技術の利用にあたっては「公平性のある意思決定とその結果に対する説明責任」を担保するように求めています。

これらに呼応して、私たちも公平性の様々な側面をいかに定量的に評価し、実現していくかについての研究を進めています。最近の主要な研究集会では必ず機械学習の公平性に関する研究発表がありますし、世界的にも公平性に関する研究論文の数は増えています。実は「公平性とは何か」を機械学習の言葉で数理的に突き詰めていくと、多数のバリエーションがあることがわかります。人々が何を公平と考えるか、様々な基準を機械学習の言葉で表現しなおすことによって、「公平」という概念をより明確なものにしていくこともできるのです。このように、私たちは、機械学習によって公平性に起きうる問題を防ぐだけでなく、機械学習をきっかけとして公平性のあり方を定義、議論することにも真摯に取り組んでいます。

今後何をすべきか

上記の 2 点を踏まえると私たちが今後何をしていくべきか、が見えてきます。公平性の問題は、技術に何ができるのかと、社会が何を求めるかの両面から粘り強く議論していかなければなりません。機械学習の公平性に関する社会の関心が高まっている今、私たちはこの問題に関して、自らの社会における責任を自覚し、従前のコミュニティの枠にとらわれず、より多くの人とともに議論を深めていきます。

* 機械学習技術を用いたシステムを「人工知能」と呼ぶことがありますが、一方で「人工知能」は人工知能研究から生まれた、あるいは将来生まれるかもしれない未知の技術やシステムを指すこともあります。本声明では議論の対象を明確にするため、「人工知能」ではなく「機械学習」を用います。

参考

[1] 焦点：アマゾンがＡＩ採用打ち切り、「女性差別」の欠陥露呈で、https://jp.reuters.com/article/amazon-jobs-ai-analysis-idJPKCN1ML0DN

[2] O’Neil, Cathy. "Weapons of Math Destruction," 2017.キャシー・オニール（著), 久保尚子（翻訳）「あなたを支配し、社会を破壊する、AI・ビッグデータの罠」インターシフト, 2018

[3] IEEE, Ethically Aligned Design -- A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, First Edition, https://ethicsinaction.ieee.org/, 2019.

[4] 人工知能学会 倫理指針, http://ai-elsi.org/archives/471

[5] 内閣府, 人間中心の AI 社会原則, https://www8.cao.go.jp/cstp/aigensoku.pdf