

学習する機械の性能をどう測るか



マシュー・ホーランド 大阪大学産業科学研究所 助教
Matthew J. Holland

新しい技術の信頼性を考える際、その性能と挙動がユーザーの期待や予想と一致しているかどうか重要な側面です。しかし、現在の機械学習技術では「期待通りか否か」以前の問題として、性能としてそもそも何を期待するか、さらにはその性能をどう測定するかが判然としておらず、人工知能 (AI) の本質的な信頼性をめぐる議論はまだ始まっていない状況です。

本研究では、「AIの性能を如何にして測るか」を起点に、性能保証とユーザーの意思決定支援メカニズムを統合する方法論の確立を目指します (図1)。機械学習の現状は、いわば「平均重視」(学習後の性能が平均的に良ければ、十分だという考え) ですが、それでは表現できない「価値観」もあります。私は、たとえば、男女間のパフォーマンスの差を是正することや、外れ値を含むデータのロバストな処理を行うこと等、「平均」から逸脱した価値観を取り扱うことのできる機械学習の基盤創成に取り組んでいます。

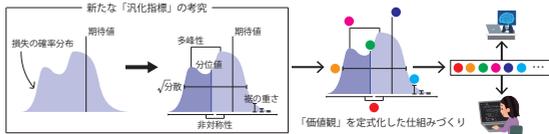


図1 汎化能力(未知のテストデータに対する識別能力)の測り方から学習則の保証とユーザーの意思決定まで体系化

キーワード

機械学習、統計的学習理論、確率的最適化

応用分野

学習する大規模情報システムの安全かつ透明な基盤



[研究の先に見据えるビジョン] 信頼に基づく機械学習と人間の共創

本研究では、性能保証を諦めることなく、機械学習の性能の測り方とフィードバック形成の刷新に取り組んでいます。この新しい技術が普及すれば、機械学習技術を活用する「AIの応用現場」のワークフローがより透明で、より確かなものになっていくことが期待されます。何より、利用者の意思が忠実に反映される学習過程を実現することで、パフォーマンスの予測と改善策をめぐる意思決定が明確になります (図2)。その上で、成功例が蓄積していけば、いつしか電卓やエレベータのように誰もが信頼する技術へと昇華します。

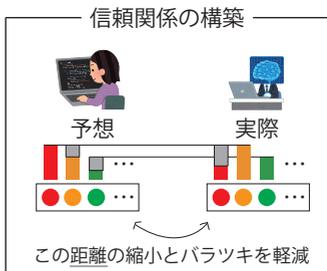


図2 予想と実際の挙動が一致してこそこの信頼