計算知能研究室

工学部　知能情報システムプログラム
http://www.eng.niigata-u.ac.jp/~yujun

自然科学系　助教
余　俊　YU Jun

専門分野	知能情報学、ソフトコンピューティング

情報通信

計算知能を用いた安全なAI社会を目指す

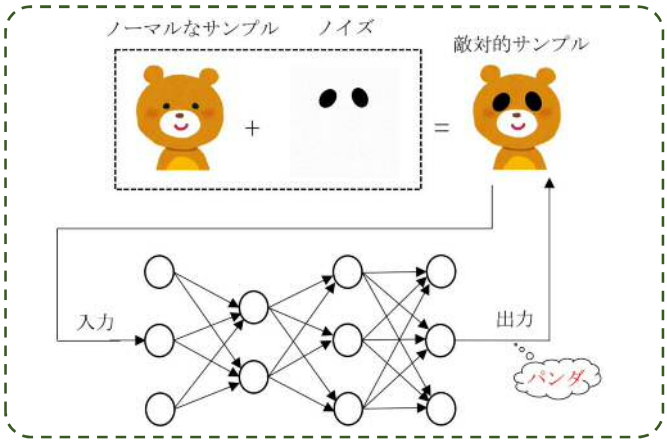
キーワード	人工知能、機械学習、深層学習、計算知能、最適化

研究の目的、概要、期待される効果

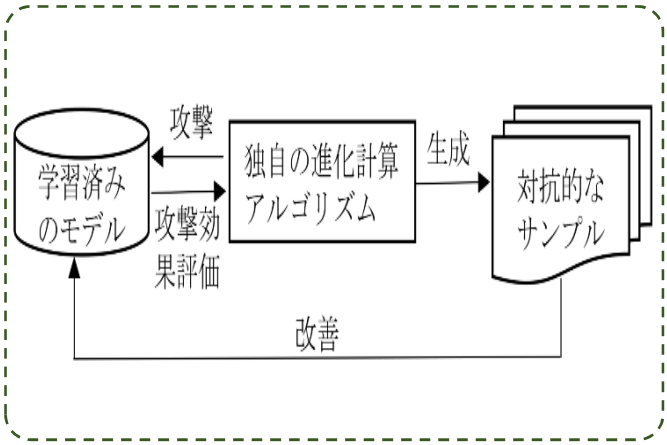
人工知能（Artificial Intelligence: AI）は多くの分野に応用されており、一部の実問題では性能が人間を凌駕するまでになっていますが、まだ脆弱性やプライバシー侵害など、安全性に課題が残されています。本研究の目的は、計算知能技術（進化計算、ファジィシステム、ニューラルネットワーク等）を用いることによって、安全・安定・安心のAIモデルを構築することを目指しています。

具体的には、モデル構築・学習の段階で、進化計算を利用して（学習サンプルに人間が知覚できないような小さなノイズを追加した）敵対的サンプルを生成し、これらのサンプルを使用して高い頑強性を持つAIモデルを構築します。同時にAIモデルの性能を劣化させずに構造を簡素化し、演算量を低減する軽量化モデルを最適化します。また、AIモデルが導入後の段階では、学習済みAIモデルに適応できる脆弱性自動検出と再学習を行い、汎用的な枠組みを開発し、自己進化能力を持たせ、様々な新しいシナリオに対応できるようします。

本研究の研究成果により、多くの実問題で使われているAIモデルにおける潜在的な危険の除去し、より安全で信頼性の高い人工知能を実現します。



誤判断を行うモデル例



モデルの脆弱性の検出と改善

関連する知的財産論文　等	Jun YU, ″Vegetation Evolution: An Optimization Algorithm Inspired by the Life Cycle of Plants,″ vol. 21, no.2, article no. 2250010 (2022). 相座悠寿，張潮，余俊「可変多集団を用いた遺伝的アルゴリズムに基づく標的型の敵対的攻撃」pp.76-80 (2022).

アピールポイント

人工知能（AI）は製造業、情報通信、金融、医療など、多くの分野で利用可能な汎用性のある技術です。AIモデル性能を向上しながら安全なAI利用の社会を追求します。

つながりたい分野（産業界、自治体等）

・画像処理、醸造管理、金属加工・機械工業など、制御・予測が必要な多岐にわたる業種。

※お問い合わせは　新潟大学社会連携推進機構ワンストップカウンター　まで　onestop@adm.niigata-u.ac.jp