3.10.5.2 連携研究室

室長　　岩爪 道昭　ほか10名

AI研究開発プラットフォームの構築と連携研究プロジェクトの推進

■概要

連携研究室では、NICTが強みとするAI技術（機械学習技術）と脳情報通信技術、セキュリティ技術、リモートセンシング技術等とそれらに由来する各種ビッグデータを利活用したオープンイノベーション型のAI研究開発に取り組んでいる。平成30年度は、AIデータテストベッドの早期公開に向けて基盤システムの設計と開発を行った。オープンイノベーション型研究プロジェクトの推進として、不均衡データを対象とした機械学習手法の研究開発及び材料物性分野への応用手法の開発、AI×脳科学、AI × セキュリティの研究開発、産業技術総合研究所（産総研）との共同研究によるニューラル翻訳、音声合成の研究開発を実施した。

■平成 30 年度の成果

平成30年度の重点的研究開発課題は次のとおり。

1．AIデータテストベッドの設計と開発

昨年度整備した『最先端AIデータテストベッド計算機設備』上に、多種多様なAIデータセットを管理・共有・公開可能とする以下の基盤システムを設計、実装を行った（図１）。

（1）管理基盤：多種多様な形式のAIデータセットを一元管理する基盤データベース。本年度は、連携プロジェクトを通じ、脳情報、セキュリティのデータベースを2019年２月末までに詳細設計を終えて実装。

（2）公開基盤：集積したAIデータセットを利用者へ提供するためのWebベースシステムを詳細設計、実装。 データ公開用Webサイトを2019年５月に公開予定（図２）。

（3）利活用基盤：AI初級者でもGPGPU環境上でAIツール等を容易に活用するためのサンドボックス機能。多様なデータセット、AIツールに対応するため、AI関連の標準的なベンチマークデータセットを調査し、基本機能の設計を実施。

2．不均衡データを対象とした深層学習法の開発と材料物性分野への応用

実問題に深層学習を適用しようとした場合、分類クラスによってデータ量の偏りがあり、学習精度が向上しない問題がある（例えば、医療画像の分類では健常者データが90％、患者データが10％の比率となってしまう等）。我々は、深層学習の中間層から得られる特徴量から疑似的にデータを生成することで分類クラスごとのデータの偏りを解消し、学習精度を向上させるCavity Filling法を新規開発した（図３）。10クラス画像分類タスクで、従来手法（アンダーサンプリング法、SMOTE法）と比較し最大で約10％程度の性能向上を確認した。また、この知見に基づき、未来ICT研究所及び東京大学と連携し、物質材料における新しい深層学習技術の研究開発に着手し、約１万件の物性データから超電導臨界温度を約70％の精度（人間の専門家で３％程度）で推定可能であることを示した。

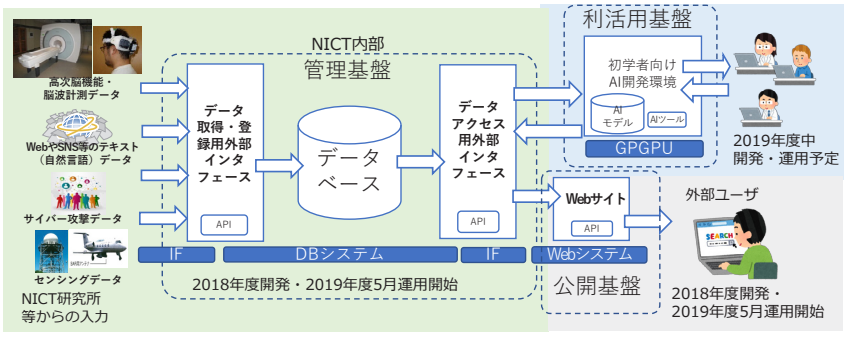

図1　ＡＩデータテストベッド基盤システム開発

図2　データ公開用Webシステムイメージ

3．AI × セキュリティの研究開発

近年、サイバー攻撃は、機械学習等による自動化、AI化が加速し新しい脅威となっており、それらに対処するセキュリティオペレーションの自動化、AI化への取組が急務である。平成30年度は、前年度立ち上げたAI×セキュリティ融合研究プロジェクトついて、研究員を増員し、研究開発体制を強化、確立した。これにより、以下の成果を得た。

（1）インシデント対応の優先順位の自動判定：セキュリティアラートの対応優先度自動判定技術を今後５年間で構築すべく、本年度は大量のアラートのクレンジングを実施。結果として200以上の特徴量を持つ13.7億件の実験用データセットを構築した。

（2）マルウェア機能分析自動化：Androidアプリ分析技術のアルゴリズムを発展させ、検知精度を大幅に向上した（ニューラルネットワークを活用した検知率の向上（図４））。

（3）攻撃の検知・予測：潜在的脅威を早期検知する技術として、協調動作検知技術のプロトタイプ構築を行った。

さらに、前年度概念設計を行ったセキュリティデータセット公開プラットフォームのプロトタイプを構築した。また、公開用データセットの整備も進め、今後同プラットフォーム上で公開予定。

4．AI×脳科学の研究開発

AI×脳科学の研究開発では、脳情報通信融合研究センターとの連携により、fMRI/MEGによる脳活動データの大規模集積化や、機械学習等のAI技術を適用することで、脳の働きを解析・模倣する次世代のAIシステムにつながる研究開発に着手した。

（1）AIデータテストベッドのコンセプトに基づき脳ビッグデータベース基盤システムのプロトタイプを設計、実装し、2019年度公開に向けた準備を行った（図５）。

（2）AI×脳科学研究開発データセット構築のための脳活動データ収集の支援を行い、下記のデータを収集、作成した。

・動画像長時間視聴時の脳反応計測を実施するための、アノテーション付の動画像データ作成

・行動下の大規模脳波計測のためのVR環境整備と日本語に関する脳波データ収集

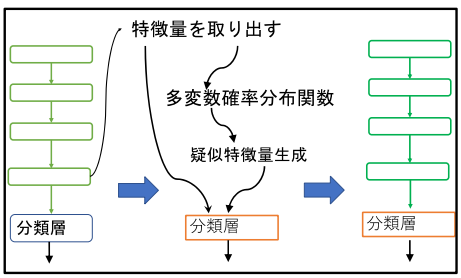

図3　不均衡データの深層学習法の開発

5．ニューラル応用研究の加速（産総研との共同研究）

産業技術総合研究所 情報人間工学領域との「情報通信分野における連携・協力の推進に関する協定」に基づき先進的音声翻訳研究開発推進センターと連携し、下記の共同研究を実施した。

（1）特許文献を対象として、ハイブリッドパラレルによるNMT訓練の高速化と翻訳精度の向上を実現ベースライン（１ GPU）に対し、5.08倍（４ GPUs）の高速化、BLEU +3.83向上を実現した。

（2）音声合成における高品質リアルタイムニューラルボコーダの構築を実現

従来手法は200秒を要した１秒間の音声合成を僅か0.07秒で高速生成可能となった。

6．他分野との連携等

リモートセンシング研究室と連携し、ゲリラ豪雨予測のための３次元降雨レーダデータへの深層学習適用に向けて、降雨エコーデータ（２次元データ）を対象にした深層学習適用の支援を行った。また、AIの最先端の技術動向に関する情報共有、意見交換を目的としてAIS AIセミナーを企画し、外部の専門家を講師に招き計２回開催した（第１回「AI自動設計」（平成30年７月23日）、第２回「構造学習理論と代数幾何学」（平成30年９月10日））。

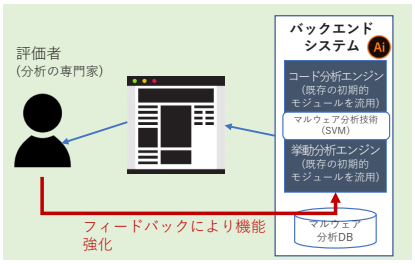

図4　セキュリティオペレーション自動化システム概念図

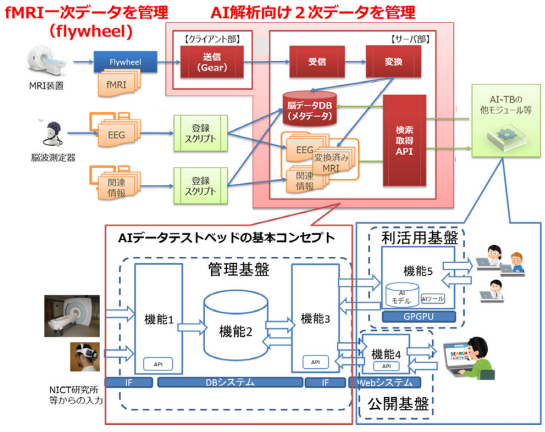

図5　脳ビッグデータベースの設計・実装