

ABCI 3.0開発加速利用（2025年度）成果概要（公開用）

課題名：
研究データ連携基盤構築プロジェクト

実施時期：2025/4/1~
所属機関名：産業技術総合研究所
代表者氏名：濱崎雅弘、阪野貴彦

成果概要：研究所内の分野横断データを連携して活用し、自己教師あり学習（SSL）と段階的転移学習を組み合わせた学習プロセスを整備した。複数タスクでモデルを試作し、所内データ連携を前提にした学習・評価・活用を検討した。

成果のポイント：本取り組みでは、所内に分散する研究データを分野横断で持ち寄り、データ連携を前提にした学習プロセス（事前学習、所内データを用いたSSL、段階的転移、下流タスクへの適用）を、再現可能な手順として整備した。個別テーマごとに学習設計を一から組むのではなく、データ連携を前提に学習を進めるための手順として、一連の流れをまとめた。

この学習プロセスを、材料・バイオ・地質を含む複数領域の画像データセットに適用し、複数の分類タスクでモデル（CNNおよびViT）を作成した。あわせて、事前学習条件、SSLの有無、段階的転移の有無、モデル構造の違いを組み合わせ、タスク横断で比較できる形に整理した。その結果、データ連携に基づくSSLや段階的転移が通常の転移学習に対して上振れするケースがあること、挙動がタスク・モデル・事前学習条件によって変わることを確認し、別テーマに展開する際に比較の枠組みを揃えるための整理を行った。

研究者のAIモデル作成のハードル低減という観点では、ラベル不足、データ量の制約、分野ごとのデータ分散、評価の観点のばらつきがボトルネックになりやすい。SSLを挟んだ段階的転移という学習設計を共通の手順として用意することで、ラベル付きデータが少ないテーマでも学習を開始しやすくし、データの追加取り込みによる効果を確認するための手順を揃えた。また、タスク横断比較の形式をそろえることで、試行錯誤の観点を共有しやすくし、設定探索や比較の知見を別テーマに持ち込みやすい形に整理した。

以上のように、データ連携に基づく学習プロセスを、学習（SSL・段階的転移）、横断比較、可視化、データ選定の観点まで含めた一体的な枠組みとして整備した。今後はデータセットやタスクを拡げ、連携範囲を広げた条件でも同様の手順で検証を継続する。並行して、利用促進のため、GUI等から推論可能な形での整備も進めている。

成果についてより詳細な情報を提供しているWebページ、発表論文などの情報：
特になし