

課題名(タイトル):

環境資源科学研究センターのクラウド AI 分析基盤構築に関する調査研究

利用者氏名:○神沼英里(1)、蒔田由布子(1)、有田正規(2)、松井南(1)

理研における所属研究室名:

(1)環境資源科学研究センター 合成ゲノミクス研究グループ

(2)環境資源科学研究センター メタボローム情報研究チーム

1. 本課題の研究の背景、目的、関係するプロジェクトとの関係
環境資源科学研究センター(CSRs)での AI・機械学習の研究基盤環境の構築に向けて、商用クラウド環境と HOKUSAI スーパーコンピュータ環境を比較する。近年、AI の研究発表が増えており、CSRs でもクラウド AI 分析基盤の構築が望まれている。本研究では、商用クラウド環境を利用して AI モデルを構築し、HOKUSAI スーパーコンピュータ環境と比較を行った。AI モデルは、CSRs の画像計測データを使い、シロイヌナズナ表現型の画像分類モデルを構築した。この AI モデルの試作で得られた知見に基づいて、CSRs に最適なクラウド AI 基盤環境を設計する。

2. 具体的な利用内容、計算方法
現在利用可能な 18 種類の AI 用画像注釈機能を持つ商用クラウドの機能を調査した。試験用データには、理研 CSRs で藤田博士らが開発している植物鉢搬送ロボットシステム(RIPPS)の植物画像計測データを用いた。18 種類の商用クラウドでは、課金・機能・動作状況などを確認した。CSRs の実験研究員等が AI モデルを構築する場合を想定して、商用クラウドを選抜した。

3. 結果
クラウド AI 分析基盤で用いる為の商用クラウドとして、「V7Labs」と「Amazon SageMaker Ground Truth」の 2 種類のサービスを選抜した。両者はデータアノテーションの自動注釈機能を持つ。両環境で RIPPS 計測の植物画像データの注釈ラベルを作成して、AI モデルを試作した。また処理の流れは、実験研究者でも再現できるように作業をプロトコル化してマニュアルとして明文化した。Ground Truth ではデータラベリング

処理と AI モデル学習処理は、同じ AWS (Amazon Web Service)内で作業が完結した。しかし、推論モデルの運用はコストが高いと判明した。また V7Labs の場合は、データラベリングの自動注釈機能が優れていたが、クラウドでの AI モデル学習機能は未整備だった。一方、HOKUSAI 内では、AI データラベリング機能や AI モデル学習機能は自分達で実装する必要がある。以上の調査結果から、CSRs 内でクラウド AI 分析基盤を構築する場合は、下記表の様に商用クラウド環境と HOKUSAI 環境の両者を、処理段階で切り替える方式が良いと考えられる。

	AI モデル構築処理	基盤環境
1	データラベリング	商用クラウド
2	AI モデル学習	商用クラウド
3	AI モデル推論(運用)	HOKUSAI

4. まとめ
CSRs のクラウド AI 分析環境として、商用クラウドサービスで、データラベリング処理と AI モデル学習・運用処理のテスト実験を行った。結果として商用クラウドと HOKUSAI の両環境併用方式が、CSRs の AI 分析基盤に適していると考えられた。

5. 今後の計画・展望
現課題とは別の課題として、HOKUSAI SS (Sailing Ship)スパコンを用いて AI モデルの推論運用テストを実施していく予定である。

6. 利用がなかった場合の理由
商用クラウド AI 環境でのテストが主体だった為、HOKUSAI BW スパコンを利用しなかった。

2020年度 利用研究成果リスト

【雑誌に受理された論文】

なし

【会議の予稿集】

なし

【口頭発表】

- 神沼英里. 植物表現型画像解析のための手作業注釈加速化手法とActive Learning. 第62回日本植物生理学会年会データベース講習会. 2021年3月14日. オンライン開催.
- 神沼英里. 深層学習を用いた植物表現型画像解析とデータアノテーション環境. 理研 CSRS インフォマティクス・データ科学推進プログラム成果報告会. 2021年3月8日. オンライン開催.

【ポスター発表】

なし

【その他(著書、プレスリリースなど)】

なし