

マルチモーダルな言語生成 AI の開発

藤田晴啓¹ 近山英輔¹

要旨

研究の目的 考古学の研究基盤である実測図は、作図や読み取りには専門的な訓練が不可欠である。しかし、作図者が考古学の専門家でないことも多く、図化の表現力に左右され、同じ遺物でも型式認定にブレが生じ、型式学的研究に悪影響をおよぼしている。そこで、3D 計測情報に基づく形状、色、表面の質感データおよび土器資料報告書等の文書データを自然言語処理にて学習して型式記述を行うふたつのモデルを融合させた 3D dense captioning モデルを前段に配備し、日本語と形式分類対応の大規模言語モデルを後段に配備した、土器型式認定と自然言語での報告を自動で行うマルチモーダル AI の開発が本研究の目的である。

研究実施計画

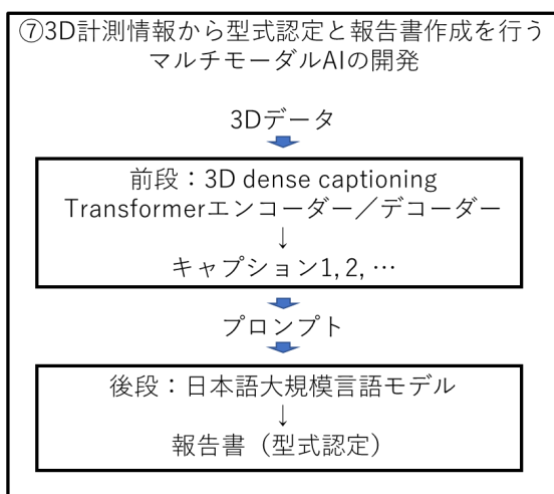
日本語対応の事前学習済大規模言語モデル、3D データおよび型式分類に関わるテキストデータを学習したモデルを結合して、3D 物体認識モデルおよび土器型式言語モデルを融合した新たなマルチモーダル型式分類 AI を開発する。最新のデータサイエンス・AI（深層学習）を活用して型式分類や年代比定といった評価を資料の形状、色、表面の質等のデータおよび文書事前学習により、新たな型式分類を行う AI 評価モデルを開発する。このような複合型 AI の開発により、考古・文化財科学分野に誰が行っても一定の結果が導きだせるデータ駆動型の研究基盤を構築し、新たな挑戦的・革新的領域を開くことが本研究の目的である。

複合型 AI によるデータ駆動型研究基盤の構築のための研究実施計画として ①須恵器坏資料、越後中期縄文土器を光学スキャナーにて計測する。②3次元座標、RGB 値、法線ベクトルの 3D 点群データ、形状、色、表面の滑らかさ、不透明度の 3D-RGBA ボクセルデータに変換する。③深層生成モデルにおける潜在空間から得られる特徴量表現をクラスターとして抽出。④潜在空間から須恵器・縄文土器の 3次元データを再構成し、大域的な情報の取得。⑤VAE および拡散モデルなどの潜在空間は連続的であるため、訓練用データとして利用した須恵器・縄文土器間の補間する。⑥最終的なクラスターが出力される。⑦3D 点群データと型式分類に関わる文書情報を統合した 3D dense captioning モデルを前段、形式分類を報告する日本語大規模言語モデルを後段とする、新たな複合型 AI 型式評価モデルを構築する。

⑦の具体的な手法としては、⑦-1 縄文土器を学習用点群データに変換する。それら点群データに対応する土器資料報告書等に存在する部分形状説明等の自然言語で記述されたキャプション用データを用意する。⑦-2 3D 点群用 Transformer エンコーダー、幾何検出・キャプション生成デコーダーより成る 3D dense captioning モデル（前段）をそれらの学習データで事前学習させる。⑦-3 前段から出力されたキャプションのリストを、形式分類・報告書作成を日本語大規模言語モデルに出力させるためのプロンプトに変換し、大規模言語モデ

ル（後段）に入力して形式分類と報告書を生成するようにファインチューニングする。⑦-4 考古資料報告書の土器に関する書誌テキスト情報，専門家の分析結果，縄文土器等の 3D データで⑦-2, 3 を繰り返しファインチューニングすることで全体の性能を向上させる。

⑧型式学と AI が融合した新たなモデルによりデータ駆動型研究基盤を構築する。



土器を説明する DL モデル案(モデル 1,2 で実現)

モデル 1

土器の 3D モデル→特徴を日本語で出力（須恵器集成の部位順の記述）

モデル 2

モデル 1 の出力（および特徴ベクトル？）を入力とし、土器の型式認定および出土地も含めた概要を説明するような文章を出力

（須恵器集成の部位順の記述を入力とし、須恵器集成外の情報も含めた出力として学習させるとよいか。）

1 新潟国際情報大学