

令和 6 年 6 月 5 日現在

機関番号：14401

研究種目：基盤研究(C)（一般）

研究期間：2019～2023

課題番号：19K12714

研究課題名（和文）新約聖書デジタル写本における深層学習による写字識別キュレーションシステムの構築

研究課題名（英文）Development of a Curation System for Scribe Identification in Digital New Testament Manuscripts Using Deep Learning

研究代表者

三宅 真紀 (Miyake, Maki)

大阪大学・大学院人文学研究科（言語文化学専攻）・准教授

研究者番号：80448018

交付決定額（研究期間全体）：（直接経費） 3,400,000円

研究成果の概要（和文）：本研究は、IIIFに対応した新約聖書大文字写本のデジタル画像を活用し、写字の字形特徴を分析した。バチカン写本においては写字の形状属性による識別精度の違いについて観察し、シナイ写本では字形が似ている文字と異なる文字のデータセットを作成した。異常検知手法であるOne-Class Deep SVDDを用いて深層学習を適用した結果、基準の文字と字形の似た文字については、入出力の差分と異常スコア分布の重なりが確認された。また、異常スコアが高い画像には、インクの濃さや隣接文字の一部が混入しているケースが多いため、字形とは別の要素が識別に関与している可能性が示唆された。

研究成果の学術的意義や社会的意義

人工知能分野の深層学習による写字識別のアプローチは、聖書学の伝統的正文批判研究や神学的解釈の固定概念からとは異なる観点を可能にし、客観的情報による新たな特徴の発見の可能性がある。高精細画像共有規格に準拠したIIIF対応の写本のデジタル画像を活用し、既存のデジタル資料の共有・再活用の実践的研究として意義があり、デジタル・ヒューマンティーズ分野の学際的研究の進展に貢献する。

研究成果の概要（英文）：This study leveraged IIIF-compliant digital images of New Testament majuscule manuscripts to analyze the characteristics of script shapes. In the Codex Vaticanus, differences in identification accuracy based on shape attributes of the script were observed, while in the Codex Sinaiticus, datasets were created based on character shape similarities. As a result of applying the deep learning method, One-Class Deep SVDD, an anomaly detection technique, overlaps in the differences between input-output and the distribution of anomaly scores were observed for the characters similar in shape to the reference characters. Additionally, it was suggested that elements other than shape might be involved in identification, as images with high anomaly scores often had cases where ink density or parts of adjacent characters were mixed in.

研究分野：Digital Humanities

キーワード：IIIF 深層学習

様式 C - 19、F - 19 - 1 (共通)

1. 研究開始当初の背景

聖書学における正文批判研究の水準は高く、数ある古典文献研究のなかでも格段に発達した領域である。多数の写本資料と高度な正文批判技術により、新約聖書の読みは確定したかのように考えられているが、依然として読みの選択が研究者によって解釈が異なり、議論が続いている異読もある。近年の深層学習による画像認識や手書き文字認識技術の革新的な技術向上により、写本デジタル画像を活用した人文情報学的アプローチによる諸問題の解決が期待されている。

2. 研究の目的

本研究は新約聖書大文字写本のデジタル画像を活用し、正文批判研究において議論されてきた写字生の特徴、および写本間の読みの違いについて分析を行うことを目的としている。デジタル画像相互運用の枠組みで公開されている写本画像から分析データを作成し、伝統的な校訂本における異読情報のアパルトゥスでは観察することが困難な文字の形状や連りについて、深層学習技術を用いて把握する。

3. 研究の方法

新約聖書大文字写本のデジタル画像は、オンライン公開されているパチカン図書館の DigiVatLib サービス、および Codex Sinaiticus Project にアクセスして収集を行った。IIIF (International Image Interoperability Framework) に対応の写本画像から文字切り出しを行うために、IIIF Image API や OpenCV ライブラリを利用して、IIIF 対応画像一括取得スクリプトを作成した。

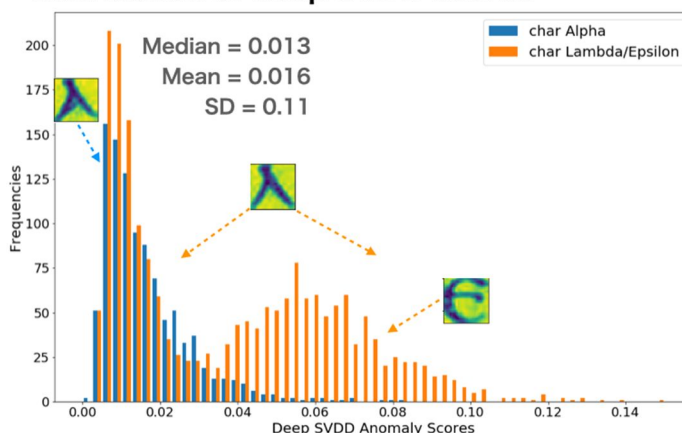
分析データには、パチカン写本の福音書の範囲からデータセットを作成し、字形の特徴調査を行った。Codex Sinaiticus Project のライプツィヒ大学図書館が所蔵している範囲に限定し、字形の特徴について分析した。字形が似ている文字と異なる文字のデータセットを作成し、字形のゆれの程度や、字形の差を測定するために、Ruff ら(2018)が提唱した One-Class Deep SVDD による異常検知手法を適用した。

4. 研究成果

IIIF 対応の画像の情報収集するためのスクリプトを作成し、URL の情報から、個々の手書きの文字の画像の位置情報を取得した。機械的に文字箇所を切り抜きを行った後、手作業で整形を行った。Jupyter notebook 上で、抽出された画像を確認しながら、文字の切り出し範囲が適切でない箇所については、手作業で範囲の再設定を行い、文字の画像位置情報の更新を行った。

当初の研究計画では、Milne & Skeat (1938) で扱われているパチカン写本とシナイ写本の字体比較をする予定であったが、対象箇所のシナイ写本の一部が IIIF に対応していなかったことから、研究計画の一部を変更した。パチカン写本では、文字の形状属性による識別精度の違いについて観察した。シナイ写本においては、字形が似ている文字と異なる文字のデータセットを作成し、基準文字からの逸脱距離を計算した。アルファ文字と字形が似ている文字ラムダ、異なる文字イプシロンに注目し、それぞれ約 7000 文字の画像データ用意し、アルファ文字を基準に文字のゆれの程度を測った。異常検知手法である One-Class Deep SVDD を用いて深層学習をさせた結果、テストデータのアルファの識別の精度については、70%の精度でアルファ文字の判別ができた。AutoEncoder と Deep SVDD とともに字形の似たアルファとラムダの入出力の差分と異常スコア分布の重なりが確認され、イプシロンについては、異常スコア分布のばらつきが大きいことが分かった。

Distribution of deep SVDD Scores



さらに、異常スコアが高い画像には、インクの濃淡や隣接文字の一部が混入しているケースが多いことから、字形とは別の要素が識別に参与している可能性が示唆された。
また、将来的な写字生の判別分析の準備として、Codex Sinaiticus Project における写字生の区分をもとにして、写字生 A と写字生 B1 の文字が混在したデータセットを作成した。

【参考文献】

DigiVatLib, <https://digi.vatlib.it/>.

Codex Sinaiticus Project, <https://codexsinaiticus.org/en/>.

Milne H. J. M. & Skeat, T. C. (1938) Scribes and Correctors of the Codex Sinaiticus, Trustees of the British Museum.

Lukas Ruff et al. (2018) "Deep One-Class Classification", in proceedings of the 35th International Conference on Machine Learning, PMLR 80:4393-4402.

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計1件（うち招待講演 0件 / うち国際学会 1件）

1. 発表者名 MIYAKE Maki
2. 発表標題 Applying Measures of Lexical Diversity to Classification of the Greek New Testament Editions
3. 学会等名 DH2019（国際学会）
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------