

令和 6 年 6 月 18 日現在

機関番号：34437

研究種目：若手研究

研究期間：2020～2023

課題番号：20K20133

研究課題名（和文）文章の数量分析による西鶴を中心とした浮世草子の著者および成立年代の解明

研究課題名（英文）Quantitative Analysis of Texts to Investigate the Authorship and Period of Ihara Saikaku's Ukiyo-zoushi

研究代表者

上阪 彩香（Uesaka, Ayaka）

大阪成蹊大学・データサイエンス学部・講師

研究者番号：60780252

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：本研究課題では、日本文学に影響を与えた重要な古典籍として位置づけられている作家である井原西鶴（1642～1693）の浮世草子における種々の疑問を対象に、近年急速に発展しているデータサイエンスの手法を用いることで、新たな視点からこれらの問題へのアプローチを試みた。西鶴の浮世草子24作品をLDAトピックモデリングで分析し、個々の作品および全体のトピックの変遷について検討した。さらに、多変量解析および機械学習により分析することで、数理的観点から著者の「特徴・クセ」を把握することを試みた。

研究成果の学術的意義や社会的意義

西鶴弟子の北条団水作とされる8作品（『武道一覧』『好色破邪顕』『正月揃』等）、西村市郎右衛門作とされる6作品（『花の名残』『諸国心中女』『浅草拾遺物語』等）のデジタル化を行った。西鶴の浮世草子の分析を通じて、初期作品は好色物を中心に展開していたが、後期作品では幅広いテーマが扱われるようになったことを定量的に示した。また、作品内におけるトピックの変化について検討した。本研究は、浮世草子のトピックの変遷と、作家の創作活動におけるトピックの選択の経年変化を捉えるための基盤を築き、これらの研究成果を国内外での学会や書籍等で公表した。

研究成果の概要（英文）：In this research project, we aimed to address questions regarding Ihara Saikaku's (1642-1693) Ukiyo-zoushi, which is regarded as an important classical work that influenced Japanese literature. By utilizing rapidly developing data science techniques, we attempted to approach these issues from a new perspective. We analyzed 24 of Saikaku's Ukiyo-zoushi works using LDA topic modeling to investigate the changes in topics across individual works and the entire corpus. Our analysis quantitatively demonstrated that while early works primarily focused on romance, later works encompassed a broader range of topics. Furthermore, through multivariate analysis and machine learning, we sought to understand the author's characteristics from a mathematical perspective.

研究分野：応用統計学，データサイエンス，人文情報学，計量情報学，デジタルヒューマニティーズ

キーワード：トピックモデリング 近世文学作品 統計科学 多変量解析 デジタル化 浮世草子

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

1. 研究開始当初の背景

井原西鶴(1642~1694)は近世文学を語る上で欠かすことのできない俳諧師・作家であり、その作品は日本文学史において重要な位置を占めている。西鶴の代表作である浮世草子は、江戸時代の風俗や生活を描写し、現代に至るまで多くの読者に親しまれている。しかし、西鶴の浮世草子については、長い間いくつかの疑問や論争が存在してきた。これまでの研究では、文献学的アプローチや歴史的背景の分析を通じて、これらの疑問に対する解明が試みられてきた。本研究課題では、近年急速に発展しているデータサイエンスの手法を用いることで、新たな視点からこれらの問題にアプローチする。

2. 研究の目的

本研究の目的は、データサイエンスの手法を用いて西鶴の浮世草子に関する疑問にアプローチし、作品の内容(トピック)と文書の数量的特徴との関連性を明らかにすることである。本研究では、特にLDA(Latent Dirichlet Allocation)トピックモデリングおよび多変量解析を用いて、各作品の文書データを解析した。このことにより、直感的・主観的な分析では捉えきれなかった情報を抽出し、各作品の特徴やテーマの変遷を客観的かつ定量的に把握することを試みた。

3. 研究の方法

LDAトピックモデリングは文書集合からトピックを抽出し、各文書のトピック分布を推定する確率モデルである。本研究では、西鶴の浮世草子24作品を対象にトピックモデリングを適用し、作品の内容に関連する主要なトピックを抽出した。解析に際しては、事前処理として出現頻度の低い単語やストップワード(文書の内容に影響を与えない単語)の除去を行った。トピックモデリングを用いて分析することにより、各作品がどのようなトピックで構成されているのか、またそれらのトピックが作品ごとにどのように変化しているのかを明確に示すことが可能となる。

また、多変量解析により、各作品間の類似点や相違点を視覚的に把握し、帰属不明の作品がどの作家の特徴に近いかを定量的に評価した。特に、主成分分析や階層的クラスタリングを用いて、作品群を視覚化し、帰属に疑問が提出されている作品の位置づけを明らかにした。

4. 研究成果

(1) 作品のデジタル化

西鶴の弟子である北条団水の8作品(『武道一覧』『好色破邪顕』『正月揃』等) 同時代に活躍した西村市郎右衛門の6作品(『花の名残』『諸国心中女』『浅草拾遺物語』等)をデジタル化し、文書データとして整備した。

(2) トピックモデリングによる分析

西鶴の浮世草子は、一般的に好色物、武家物、町人物、雑話物に分類され、分類に関しては諸説ある。トピックモデリングを実施した結果、好色物、武家物、町人物、雑話物に関連する単語で構成されるトピック及び各々のトピックにおいて特徴的な単語を抽出した。また、それぞれのトピックの出現確率が高い文書は、宗政(1969)¹が一般の分類として挙げた分類と対応してい

¹ 宗政五十緒.(1969).『国文学解釈と鑑賞』34(11). pp.26~35.「仮名草子から浮世草子へ」. 至文堂.

ることを明らかにした。さらに『好色一代男』の作品内においてトピックの出現確率が章ごとに変化することが示され、浮世草子を文書データとして用いたトピックモデリングは文学作品の内容を分類するのに有効な手法であることが示唆された。

また、西鶴の浮世草子 24 作品をトピックモデリングで分析し、個々の作品および全体のトピックの変遷を調査した。その結果、初期作品は好色物を中心に展開していたが、後期作品では幅広いテーマが扱われるようになったことを定量的に示した。これらの結果は、著者の作品制作における焦点の変化を示唆しており、著者の作品制作におけるテーマの経年変化を捉える基盤を築いた。

(3) 多変量解析を用いた著者に関する検討

多変量解析およびアンサンブル学習 (パギング法、ランダムフォレスト法、アダブースト法) を用い、西鶴と弟子の北条団水 (1663 ~ 1711) の文章の間に差異がみられた分析項目を用い、章単位での西鶴遺稿集の文章が、西鶴と団水のどちらの文章と類似しているのかを検討した。これらの結果をまとめ、「文学と言語コーパスのマイニング」(岩波書店、2021) にて公表した。

(4) 近世文学作品の定量分析

近世文学作品を対象に、主成分分析およびトピックモデルを用いて、カテゴリの特徴について検討した。これらの結果をまとめ、「コーパスによる日本語史研究 近世編」(ひつじ書房、2023) にて公表した。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 0件）

1. 著者名 Ayaka Uesaka	4. 巻 -
2. 論文標題 Classification of Early Modern Literature Using a Topic Modeling	5. 発行年 2022年
3. 雑誌名 Proceedings of the 6th Asia Future Conference	6. 最初と最後の頁 1336-1345
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 上阪 彩香	4. 巻 33(8)
2. 論文標題 Stefan Th. Gries (2022) Toward more careful corpus statistics: uncertainty estimates for frequencies, dispersions, association measures, and more [Research Methods in Applied Linguistics, Volume 1, Issue 1]	5. 発行年 2023年
3. 雑誌名 計量国語学	6. 最初と最後の頁 612-614
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計4件（うち招待講演 2件/うち国際学会 1件）

1. 発表者名 Ayaka Uesaka
2. 発表標題 Early Modern Japanese Literature from a Quantitative Perspective
3. 学会等名 16th International Conference of the European Association for Japanese Studies（招待講演）（国際学会）
4. 発表年 2021年

1. 発表者名 上阪彩香
2. 発表標題 データサイエンスのための統計学
3. 学会等名 データサイエンスへの期待2021 シンポジウム（招待講演）
4. 発表年 2021年

1. 発表者名 上阪彩香
2. 発表標題 日本語歴史コーパス(CHJ)を用いたトピックモデリング
3. 学会等名 「通時コーパスの構築と日本語史研究の新展開」研究発表会
4. 発表年 2021年

1. 発表者名 上阪彩香・村上征勝
2. 発表標題 役者絵における容貌の描き方の数量分析：写楽，豊国，春英の作品を対象として
3. 学会等名 日本行動計量学会第51回大会
4. 発表年 2023年

〔図書〕 計3件

1. 著者名 金 明哲、中村 靖子、上阪 彩香、土山 玄、孫 昊、劉 雪琴、李 広微、入江 さやか	4. 発行年 2021年
2. 出版社 岩波書店	5. 総ページ数 248
3. 書名 文学と言語コーパスのマイニング	

1. 著者名 数理人材育成協会	4. 発行年 2021年
2. 出版社 学術図書出版社	5. 総ページ数 272
3. 書名 データサイエンティスト教程 応用	

1. 著者名 岡部 嘉幸、橋本 行洋、小木曾 智信（編集）	4. 発行年 2023年
2. 出版社 ひつじ書房	5. 総ページ数 408
3. 書名 コーパスによる日本語史研究 近世編	

〔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