

〔論 文〕

AI・データサイエンス技術の芸術分析への適用

——江戸小唄の特徴パターンの分析——

前 田 利 之

矢 島 ますみ

要 旨

AI・データサイエンス技術の芸能分野への適用例として、まず江戸小唄のテキストマイニングによる分析について議論する。本研究では歌詞の作家による傾向の抽出・分類の可能性の検証として、西條八十と平山盧江が作詞した小唄の歌詞について、潜在的意味インデキシングとクラスター分析を行なった。これにより、作家別の分類が可能であることが示唆された。AI・データサイエンス技術の芸能分野への適用は大きな可能性を持っていると確信しており、今後さらに発展していくことを期待している。

キーワード：AI・データサイエンス技術, 小唄, テキストマイニング, 潜在的意味インデキシング, クラスター分析

1. 序

近年、AI・データサイエンス技術の発展は圧倒的なものがあるのは論をまたない。特に生成AIの分野に関してはさまざまな分野での可能性が考えられており適用事例も無数にわたり、研究開発も世界中で進められている。

しかし、生成AI以外のAI・データサイエンス技術の社会適用も重要であろう。その中には、現在注目されている、膨大な量の計算機資源を必要とする深層学習をベースとしたものだけでなく、旧来の技術の応用により新たな知見を得られる可能性も十分にあると考えており、実際に適用例もある。

本論文では、これまで我々が取り組んできたAI・データサイエンス技術の芸能分野への適用例として、まず江戸小唄のテキストマイニングによる分析について議論し、それをうけて今後の展望、課題について論じる。

2. 関連研究

近年、自然言語解析の手法を用いて、定型化されていない文章の集まりから有用な情報を抽出するテキストマイニング技術が活用されている。文学分野では、竹田正幸や福田智子らによる「文学作品からのテキストマイニング：文学における発見を支援する(2001)」や「古典和歌からの知識発見：モビルスーツを着た国文学者(〈特集〉人文科学における情報処理)(2002)」があり、また、松田卓久と飯島淳一(東

京工業大学大学院)は、「テキストマイニング技術の音楽情報への適用(経営情報学会 2002)」で、楽曲と文書の対応や音楽のプロの感覚分析を試みている。

さらに、小林佳織、狩野恵里奈、鈴木崇史(東洋大学社会学部)らは、1977年から2011年までの過去35年間にわたる女性グループによる楽曲(34組439曲)を対象に、「女性グループの歌詞の計量テキスト分析(言語処理学会 2013)」を行ない、売れる歌詞やヒット曲のエッセンスを抽出している。

3. 江戸小唄のテキストマイニングによる分析

江戸小唄は、安政年間に清元のお葉によって創始され、幕末から現代に至るまで伝承されてきた日本独自の表現分野である。小唄は、短い歌詞と三味線の音で構成される音楽であり、庶民的な文化として発展してきた大衆芸能であり、これまで学術的な研究の対象として捉えにくかった分野である。

江戸小唄は、文化文政の頃から徐々にその姿が固定され、清元のお葉によって成立するが、そのルーツは、それまでに唄われていた狂言小唄、隆龍節、弄斎節、投げ節、やり歌、田舎歌、上方歌などに影響を受けて発展した。特徴としては、「粋(いき、すい)」の要素が多分にあり、「いき」は「意気(こころの張り)」「意気地」でもある。また、描かれる風姿も、洗い髪、素足といった洗練された色っぽさや清潔感が伴って小唄の「粋」を構成し、詞型も簡潔かつ、節調の多様性が魅力である。

このように、江戸小唄は、ある意味感覚的なクオリティを持つ表現であり、「粋」を根底とすると言える。この捉えづらい表現感覚は、小唄以外でも文学や詩歌に多くみられ、従来それら进行分析するにあたっては、主観的な解釈や指標を免れない傾向が見受けられ、客観的に説明しづらいものとされていた。しかし、近年の「テキストマイニング」手法による分析が多くの手がかりを示すようになってきている。

「江戸小唄」は、江戸の情緒を取り入れた三味線のテンポの良い小唄であり、小唄の舞台公演や発表会では、唄に舞踊が振り付けられた「小唄振り」が最後に披露され、華やかに幕を閉じるのが恒例である。「江戸小唄」は約700曲あり、また、舞踊の振付けを前提とした「舞踊小唄」として約150曲程度が大正後期以降に生まれたと考えられている。

この研究では、「舞踊小唄」と「江戸小唄」という異なる文化的背景を持つ二つの小唄の言語的特徴を数理的分析によって比較検討する。具体的には、テキストマイニング技術を用いて自立語や語の連続などを分析し、日本独自の小唄の言語世界と舞踊の身体世界の結びつきを客観的に解明することを目指す。特に、小唄のように語数が少ない場合でも、テキストマイニング分析が可能かどうかを検討する。この研究は、日本の伝統芸能における言語と身体の関係を理解し、その独自性を明らかにする準備段階の研究と位置付けられる。ここでは昭和12年から15年にコロムビアから発行されたSPレコードに収録された舞踊小唄名曲集96曲の歌詞から、作詞者による傾向を抽出・分類することを試みる。これは同時に、テキストマイニング技術を用いて、語数や対象数が少ない場合でも分析が有効であるかを確認することでもある。

小唄のテキストマイニングによる分析として、歌詞の作家による傾向の抽出・分類の可能性を検証する。舞踊小唄名曲集第1集～第4集(96曲中)から、西條八十と平山盧江が作詞した小唄の歌詞をそれぞれ35曲、34曲ずつ選び、RMeCabによる分析を行なった。RMeCabは、日本語形態素解析器MeCabを統計解析環境であるRから利用できるパッケージである。これは形態素解析と統計解析処理を統合的に行なえるテキストマイニングツールとして利用できる。

3.1. 潜在的意味インデキシング

最初に、潜在的意味インデキシングによる分析について説明する。潜在的意味インデキシングは、ター

Aug. 2024

AI・データサイエンス技術の芸術分析への適用

ム・文書行列をまず作成し、それを圧縮することで、分類を効果的に行なう技術である。この技術は、高次元のターム・文書行列を低次元で近似的に表現し、特異値分解を用いてもとのターム・文書行列を圧縮する。本分析では、前節で用意したデータファイルを使用し、それらに対して特異値分解を行なった。その結果を示した図を図1に示す。

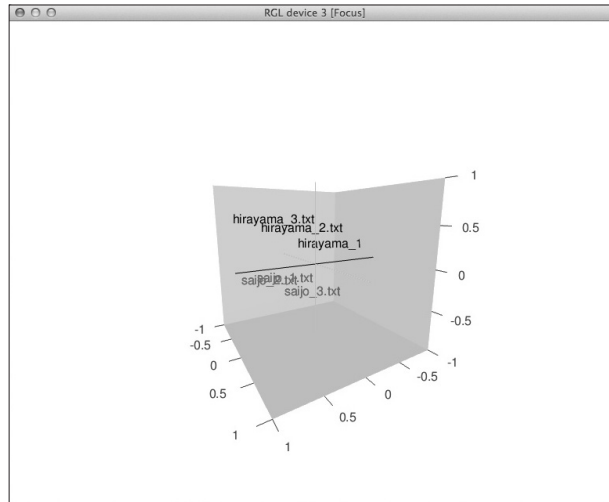


図1：潜在的意味インデキシング分析出力

図1より、hirayamaとsaijoのそれぞれの文書がクラスタリング同様近くに集まっていることがわかる。これにより、作家による歌詞の分類が可能であることが改めて示されている。ここで西條と平山が使用している頻出度の差が大きい名詞をあげると、西條は「君、恋、色、空、夢」といった名詞句を平山より多用しており、逆に平山は「花、人、尻」といった名詞を多用していることが見てとれる(表1)。

表1：西條および平山の舞踊小唄における頻出度の差が大きい名詞

名詞	西條八十の舞踊小唄における頻出度(回)	平山蘆江の舞踊小唄における頻出度(回)
君	13	0
恋	14	1
色	11	1
空	8	0
夢	8	1
花	5	11
人	2	8
尻	0	6

3.2. クラスタ分析

次に、クラスタ分析について説明する。クラスタ分析は、データの各個体間の内容の類似度を距離とみなして、その距離を基準にグループ分けする手法である。本分析では、各作詞家の曲を3つのグルー

ブに分け、それぞれのグループに対応する文書データ (sai_1.txt, sai_2.txt, sai_3.txt, hira_1.txt, hira_2.txt, hira_3.txt) と、対照として、その他の作家による舞踊なし小唄の文書ファイル (oth_1.txt, oth_2.txt, oth_3.txt) を構築したのち RMeCab により、これらの文書データから名詞、動詞、形容詞を抽出し、クラスタリング処理を行なった。データ間の距離についてはキャンベラ距離を使用した。また、クラスター作成方法については、分散を最小にするよう併合するウォード法を用いた。

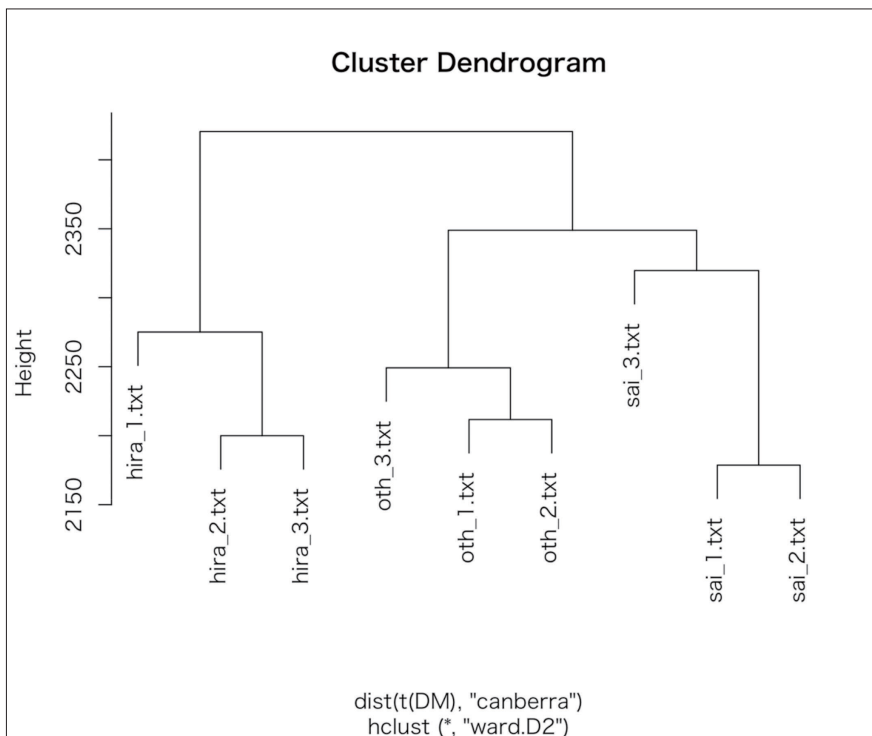


図2：クラスター分析出力

分析を行なった結果、図2のように西條八十と平山盧江と、それ以外の夫々の作家の3ファイルがクラスターを構成し、2人の作家の歌詞が自動的に分類されることが確認された。これにより、テキストマイニング技術による自動分類の可能性が認められた。

4. 結語・今後の展望

以上の議論により、テキストマイニング技術により歌詞だけの分析で作家による分類が可能であることが示唆された。今回の分析は非常に単純なものであり、今後、唄うことを主とした「江戸小唄」と、舞踊を前提にした「舞踊小唄」の両者の言語的特徴を比較検討するに際しては、さらに詳細な分析が必要となることが考えられる。例えば、助詞、助動詞、形容動詞など他の品詞についても考慮の対象とするとともに、単なる形態素の頻度情報だけでなく統語情報までも分析の範囲とすることが考えられ、さらには深層学習による分析器の構築も検討に値する。

さらに現在、我々の研究グループでは、上方舞の吉村流の流派を象徴する芸風を映像分析から検証す

Aug. 2024

AI・データサイエンス技術の芸術分析への適用

る試みとして、映像分析ソフト dartfish の機能を利用し、歌詞に対する動作と動作軌跡等の視点から共通性・類似性等を明らかにすることによって、流派としてこだわる吉村流が伝える伝承の芸(運動の特徴)を客観的に明らかにすることを推進している。分析の方法として、当初は動き部分をトレースした後、データ解析を行なう予定にしていたが、前処理の手作業でのトレースには時間が掛りすぎることが明らかになった。そのため改善策として、部分部分を手作業でトレースをする分析手法と並行して、動画像データそのものを時系列データとして解析する方法を検討する。例として、OpenCV を用いて動画像解析、特に物体トラッキングの適用可能性について検討を進めている。

このように、AI・データサイエンス技術の芸能分野への適用は大きな可能性を持っていると確信しており、今後さらに発展していくことを期待している。

謝 辞

本研究の一部は「江戸小唄の中の舞踏的身体-テキストマイニングによる小唄と舞踏小唄の特徴比較研究」基盤研究(C)(一般24520173 H24~H26)および「上方舞吉村流の粋の伝承芸~人間国宝四世「吉村雄輝」と六世「輝章」の舞の同調性分析」基盤研究(C)(一般20K00254 R2~R6)の助成をうけた。関係各署に深謝する。

参考文献

- 石田基広(2008), Rによるテキストマイニング入門, 森北出版
上村幸以(2002), 小唄ことば事典, 江戸趣味保存会
上村幸以(1997), 小唄とはなし, 江戸趣味保存会
大野恵造(1967), 江戸小唄総覧, 邦楽と舞踊出版部
大野恵造(1961), 小唄細見, 新邦楽社
英十三(1952), 新古江戸小唄全集, 文川堂書房
館野善二, 波多一索監修(2005), 小唄ベスト120, 日本伝統文化振興財団
吉田忠 編(1963), 小唄ことばの手引, 東京邦楽出版社
矢島ますみ, 前田利之(2013), 舞踏小唄の特徴パターンの分析:テキストマイニング技術を用いた歌詞による傾向の抽出・分類の検証, 明海大学教養論文集:自然と文化(24), pp. 39-43