

Journal of Japan Society of Information and Knowledge

情報知識学会誌

Vol.31 No.2 (May 2021)

~~~~~目 次~~~~~

## 特集 第29回 (2021年度) 年次大会 (研究報告会&総会)

|                                      |     |
|--------------------------------------|-----|
| プログラム                                | 157 |
| 新聞中の学術記事の自動抽出の試み - 機械学習モデルを用いて -     |     |
| ..... 西澤正己, 孫媛                       | 159 |
| 物語会話文での自称詞と対称詞の分類                    | 167 |
| ..... 村井源, 石川一稀                      |     |
| モバイルゲームにおける女性キャラクターの魅力的要素の分析         |     |
| ..... 宇田朗子, 村井源                      | 173 |
| 共起性に着目した中小企業プロジェクトの失敗リスク要因の関係性分析     |     |
| ..... 保田洋, 川向肇, 西村治彦                 | 181 |
| 患者紹介時の診療情報提供書における記載内容のテキスト分析         |     |
| ..... 石崎潤, 吉岡正昭, 西村治彦                | 186 |
| マンガの内容に基づく探索のための NDC-LD とのリンクング      |     |
| ..... 後藤佑斗, 三原鉄也, 永森光晴               | 192 |
| 新型コロナウイルス感染症による日本の研究活動への影響           | 198 |
| ..... 西川開                            |     |
| COVID-19 における Twitter の利用傾向に関する探索的研究 |     |
| ..... 渡邊憲二, 箕輪弘嗣                     | 204 |
| 官民データ活用推進基本法の施行と自治体における              |     |
| オープンデータ推進の関係                         | 211 |
| ..... 本田正美                           |     |
| 地域資料のデジタル化・オープン化の促進への一提案             |     |
| - 神奈川県事例から -                         | 224 |
| ..... 長塚隆                            |     |
| Web システム上の AI の学習システムを使った授業の実践       | 230 |
| ..... 山島一浩                           |     |
| 弱視生徒を対象とした授業におけるノートのデジタル化            |     |
| ..... 小川雄太, 宮本行庸                     | 234 |

< 裏表紙に続く >

~~~~~



情報知識学会

<http://www.jsik.jp/>



すべてを突破する。

印刷の会社だと思ってる？

ないっすよ。

TOPPAN!!! TOPPAN

情報知識学会第29回（2021年度）年次大会（研究報告会および総会）

実行委員長 村川猛彦（和歌山大学）
 実行委員 村井源（はこだて未来大学）
 実行委員 河瀬彰宏（同志社大学）
 実行委員 本田正美（関東学院大学）

1日目：5月22日（土）

10:15-10:25	オープニング	
	セッション1A（一般）	セッション1B（一般）
10:30-11:00	新聞中の学術記事の自動抽出の試み —機械学習モデルを用いて— ○西澤正己, 孫媛（国立情報学研究所）	共起性に着目した中小企業プロジェクト の失敗リスク要因の関係性分析 ○保田洋（甲子園短期大学, 兵庫県立大学）, 川向肇, 西村治彦（兵庫県立大学）
11:00-11:30	物語会話文での自称詞と対称詞の分類 ○村井源, 石川一稀（はこだて未来大学）	患者紹介時の診療情報提供書における記 載内容のテキスト分析 ○石崎潤（西宮協立リハビリテーション 病院, 兵庫県立大学）, 吉岡正昭（大阪府 済生会吹田病院）, 西村治彦（兵庫県立大 学）
11:30-12:00	モバイルゲームにおける女性キャラク ターの魅力的要素の分析 ○宇田朗子（株式会社エイチ・アイ・ ディ）, 村井源（はこだて未来大学）	マンガの内容に基づく探索のためのNDC- LDとのリンク ○後藤佑斗, 三原鉄也, 永森光晴（筑波 大学）
13:00-14:00	総会	
	セッション2A（一般）	セッション2B（一般）
14:20-14:50	新型コロナウイルス感染症による日本の 研究活動への影響 ○西川開（科学技術・学術政策研究所）	官民データ活用推進基本法の施行と自治 体におけるオープンデータ推進の関係 ○本田正美（関東学院大学）
14:50-15:20	COVID-19におけるTwitterの利用傾向に関 する探索的研究 ○渡邊憲二, 箕輪弘嗣（岡山商科大学）	地域資料のデジタル化・オープン化の促 進への一提案 —神奈川県事例から— ○長塚隆（鶴見大学）
	セッション3A（一般）	セッション3B（一般）
15:40-16:10	Webシステム上のAIの学習システムを使っ た授業の実践 ○山島一浩（筑波学院大学）	タイピングによるプログラミング学習の ためのソースコード提示に関する一検討 ○村川猛彦（和歌山大学）
16:10-16:40	弱視生徒を対象とした授業におけるノー トのデジタル化 ○小川雄太（兵庫県立視覚特別支援学 校）, 宮本行庸（神戸学院大学）	専門的データベース間の相互協力は可能 か 専門特化型データベースの存在意義 を問う ○青柳明佳（国文学研究資料館）

2日目：5月23日（日）

	セッション4A (学生)	セッション4B (学生)
10:30-11:00	炎上の早期解決を促す情報解析システムの 実現のため、発言の立場ごとに発言内容 を分類する手法の研究 ○横山拓也, 箕輪弘嗣, 渡邊憲二 (岡山 商科大学)	
11:00-11:30	Narration Adder : 自動ナレーション追加 システムの開発 ○岩田和樹, 箕輪弘嗣 (岡山商科大学)	因子分析を用いた恋愛小説における文体 的特徴の抽出 ○白鳥孝幸, 村井源 (はこだて未来大学)
11:30-12:00	画像認識AIを利用した危険診断システム の研究 ○田中蓮, 箕輪弘嗣 (岡山商科大学)	機能語に基づく物語会話文の文体分類と 推定 ○逢坂駿也, 村井源 (はこだて未来大学)
13:00-14:00	論文賞授賞式・記念講演、永年会員発表	
	セッション5A (学生)	セッション5B (学生)
14:20-14:50	多言語分析のための感情辞書の検討 ○大川慎 (名古屋市立大学)	二次創作作品を対象とした物語特徴分析 ○石川一稀, 村井源 (はこだて未来大学)
14:50-15:20	女性就業者数促進のための計量分析アプ ローチの研究 ○植杉京佳, 大槻明 (日本大学)	神話物語と神話を原型にした現代物語の 構造比較 ○吉田拓海, 村井源 (はこだて未来大学)
15:20-15:50	PAC分析を利用したオタクの「布教」に対 する意識や態度についての研究 ○細川怜椰, 松村敦, 宇陀則彦, 堤智昭 (筑波大学)	クエスト構造に注目したロールプレイン グゲームの物語構造と物語内容分析手法 の提案 ○中村祥吾, 村井源 (はこだて未来大学)
	セッション6A (学生)	セッション6B (学生)
16:10-16:40	情報収集行動の振り返り支援を目的とす る閲覧履歴提示手法 ○佐藤千尋, 高久雅生 (筑波大学)	N-gramを用いた日記文章の文章構造の分 析 ○松田雛乃, 村井源 (はこだて未来大学)
16:40-17:10	時系列を考慮したクエリ満足度の推定 ○楊之卓, 高久雅生 (筑波大学)	エッセイに使用されているレトリックの 分類と計量分析 ○根本さくら, 村井源 (はこだて未来大 学)
17:20-17:30	クロージング	

第29回年次大会予稿

新聞中の学術記事の自動抽出の試み

—機械学習モデルを用いて—

An Attempt to automatically extract academic articles in newspapers - Using machine learning models -

西澤正己*, 孫媛

Masaki NISHIZAWA*, Yuan SUN

国立情報学研究所

National Institute of Informatics

〒101-8430 東京都千代田区一ツ橋2-1-2

E-mail: nisizawa@nii.ac.jp, yuan@nii.ac.jp

我々は大学に関連したプレスリリースを調査しており、それは近年大幅に増加し、新聞への掲載も増加している。また、新聞への掲載率が原論文の掲載誌や分野に依存していることも分かってきた。しかし、新聞掲載の要因の分析にはプレスリリースと非対応の学術記事の抽出も必要である。ここではプレスリリースと対応した新聞報道記事を教師データとした機械学習により新聞報道中の学術ニュース記事（速報記事）、フィーチャー記事（特集記事）の抽出を試みる。

キーワード: 大学, 研究成果, 報道, 学術論文, プレスリリース, 新聞, 機械学習モデル

Keyword1, University, Research output, Press report, Scientific publication, Press release, Newspaper, machine learning models

1 はじめに

日本では、新聞報道での学術的研究成果に対するニュース記事は、大学等研究機関から出されるプレスリリースに基づくものが多い。プレスリリースからニュース記事への採用は記者の判断による部分が大きいと思わ

れる。これまでの研究で、我々はプレスリリースと新聞報道の関係を調査してきた。そこでは、近年、大学からのプレスリリースが急激に増え、研究成果を積極的に発表するようになったこと[1]、それに従って新聞報道も増えていることがわかった[2][3][4]。また、プレスリリース発表の元となる原論文情報

では、学術雑誌からの発表に関するものが2005年から2015年までの間に17.6倍の増加がみられ[5]、大学等の機関が、著名な学術誌への掲載を積極的に、組織的にプレスリリースするようになったとみることができる。また、有力オープンアクセス誌への移行も目立ってきている[6]。

これまではプレスリリースから新聞報道への経路で見てきたが、この経路だけでは研究成果の新聞報道に対する全体像は把握できず、要因の分析[7][8]も部分的なものになる。よって、ここでは新聞報道中の学術成果に関する記事全体の抽出を目的に、プレスリリースと対応した新聞報道記事を教師データとした機械学習により新聞報道中の学術ニュース記事（速報記事）およびフィーチャー記事（特集記事）の抽出を試みる。

2 データ

2.1 プレスリリース

プレスリリースは発表する機関側から新聞メディア等への能動的な発表形態であり、「大学」関連のプレスリリースにおいては、大学が成果等を発表したもの、企業との産学連携や提携、その他、企業の製品が大学などに納入されたというような記事も含まれている。

日経テレコン21[9]での検索では、見出し記事の他、記事本文の検索が行われる。また、検索は本文や見出しに対して付けられたキーワードに対して行われているので、「大学」で検索した場合は、本文には「東大」しかなくても「東京大学」がキーワードに含まれており、機関名に関しては、省略形でもほぼ検索にはヒットするようである。

今回の分析では「大学」のキーワードでヒ

ットした日経プレスリリース（現在は日経新聞プレスリリース）で報道されたプレスリリースに対して、研究成果に関連した原論文のDOI特定作業が終わっている2011年から2016年のデータを用い、対応した新聞記事データを第1段階目の機械学習正解データ（ニュース記事）としている。

2.2 新聞記事

新聞記事については、日外アソシエーツから発売されている新聞記事データ集[10]を使用した。ここでは、全国版および地方版がそろった、読売新聞および毎日新聞の新聞記事データ（2011-2016年）を第一段階の教師データとして使用し、朝日新聞（2010-2015年）および2011年から2018年の毎日および読売新聞を第二段階以降の予測データおよび教師データとして用いた。

3 分析手法および結果

3.1 プレスリリースと新聞記事の相関

プレスリリースと新聞記事の内容の一致を調べるために、プレスリリース、新聞記事に対して、記事中に出現する機関名を機関名データベースと照合し機関名を抽出した。機関名が一致する記事に対して、プレスリリースの記事見出し、新聞の見出し、新聞の本文記事のそれぞれをキーワードに分解し、キーワードの一致度合い（コサイン距離）を計算した。さらに日時差を加えた指数を計算し、最終的には目視チェックで一致記事を抽出した。詳しい説明は[3]に述べており省略するが、ここでは2011年から2016年の毎日新聞および読売新聞に対して、上記の方法で一致を確認したプレスリリースとそれに対応した新聞記事の組み合わせを得ている。

表 1 : 本研究で用いた新聞データのの記事数

News	記事数	2010	2011	2012	2013	2014	2015	2016	2017	2018
朝日新聞	全体	337,530	329,475	335,071	314,503	297,466	288,799			
	機関特定	152,363	148,913	153,638	180,512	173,507	171,285			
毎日新聞	全体		287,060	296,754	290,684	286,799	276,543	271,192	262,356	242,237
	機関特定		146,705	153,802	147,472	148,249	140,850	139,337	135,319	127,121
読売新聞	全体		313,205	310,430	299,821	281,514	268,763	261,079	263,459	262,880
	機関特定		172,906	172,403	168,233	160,805	156,796	152,061	152,420	151,516

表 1 には本研究で用いた新聞データの記事数を年毎に示している。ここでは記事から機関名が抽出できた全体の半分強のデータを用いている。毎日新聞と読売新聞の2011年から2016年までは第一段階の教師データの抽出に用い、朝日新聞の2010年から2015年および毎日新聞と読売新聞の2011年から2018年のデータは学習モデルを用いた予測と2段階目以降の教師データの作成に用いている。

3.2 原論文のDOI情報

我々はこれまでに、プレスリリースの原論文情報の抽出を行ってきた[4][5]。今回の分析ではこれまでに得られた2011年から2016年のプレスリリースの原論文情報でDOIが特定できたものを研究成果の報告であるとして、これと一致した新聞記事（毎日、読売）を第1段階の学術成果に対するニュース記事の教師データとした。表2には2011年から2016年の日経プレスリリース記事数、「大学」で検索した記事数：PR(大学)、原論文DOIが

特定できた件数：PR(大学, DOI特定)、その中の新聞一致記事数：毎日-PR一致(withDOI) および読売-PR一致(withDOI) の記事数を示している。

3.3 機械学習モデル(2分類, 3分類)

機械学習モデルとしては、GoogleのTensor Flowを用いたテキスト分類のチュートリアルを参考にした[11]。ここではモデルの選択肢として、n-gramモデルとsequenceモデルが述べられており、どちらを選択するかは、S/W比 (S: Number of samples / W: Number of words per sample) が1500より小さければ前者、大きければ後者を薦めている。ここで用いる新聞記事データはWのモード値が2015年の読売新聞で553、2015年の毎日新聞で316であった。ただし、ここでのword数は本文テキストをMeCab[12]を用いて形態素分解した単語数である。ここで第一段階として学術研究の結果を報道する「ニュース記事」の正解として用いる教師データは表2の毎日、および読売新聞のPR一致(with DOI)記事であるの

表 2 : 分析に使用した日経プレスリリース (対応新聞記事) の記事数

	2011	2012	2013	2014	2015	2016	合計
PR (日経プレスリリース)	29,830	27,031	25,099	24,415	26,659	28,602	161,636
PR(大学)	1,429	1,472	1,417	1,375	1,638	1,887	9,218
PR(大学, DOI特定)	366	454	525	454	645	688	3,132
毎日-PR一致(with DOI)	29	59	74	46	37	42	287
読売-PR一致(with DOI)	54	89	75	40	46	59	363

表3：テキスト2分類，3分類の機械学習モデルのパラメータ

環境	Mac OSX 10.15.7, python 3.6.7	
ライブラリ	Tensorflow(2.0.1), Keras:Pythonの深層学習ライブラリ(2.2.4-tf), keras(2.3.1)	
モデル	MLP(Multi-layer Perceptron)モデル (2層)	
	入力層 (損失関数)	2分類: binary cross-entropy (二値交差エントロピー) 3分類: categorical cross-entropy (カテゴリカル交差エントロピー)
	隠れ層	2層のdropout層: 正則化のため(訓練サンプルへの overfitting を防ぐため)
	出力層 (活性化関数)	2分類: sigmoid関数 3分類: softmax 出力は(0-1)の確率スコア (1に近いほどPositive)
テキスト処理	n-gram:	1,2-gram
Count mode	Tfidf	(Tf = Term Frequency, Idf = Inverse Document Frequency)
Scoring Method	f_classif	
Select top_k features	score	top 20000

で、この数は2011-2016年の650件となる。よって、最初の試みとして負荷の少ないn-gramを用いたMLP(Multi-layer Perceptron)モデルを用いることとした。表3に第1, 2段階の2分類(ニュース記事: POS, それ以外: NEG), 第3段階の3分類(ニュース記事: Jnl, フィーチャー記事: ftr, それ以外: NEG)とするモデルのパラメータを示す。

3.4 ニュース記事の分類

第1段階では、毎日、読売新聞に対するプレスリリースとの一致記事の内、原論文のDOI特定ができていない2011年から2016年の一致記事(650件)の内640件を正解データ(POS)とし、それ以外を不正解(NEG)とした。このデータで上記機械学習プログラムを用い、正解データの半分を訓練用(320件)、残り半分を検証用(320件)として学習・検証し、学習モデルを作成した。この時のライブラリに渡すパラメータはチュートリアル[11]の値

表4：ハイパーパラメータの値

パラメータ	値
learning_rate	1.00E-03
epochs	50
batch_size	128
layers	2
units	64
dropout_rate	0.2

をそのまま使用した。この時の値を表4に示す。また、この時の学習の精度はaccuracyの値として得られるが、Validation accuracy: 0.975, loss: 0.0906であった。その後、朝日新聞(2010-2015)、毎日新聞(2011-2018)、読売新聞(2011-2018)の新聞記事データに対して予測(Predict)を行った。読売新聞と朝日新聞に対してのPositive確率の分布を図1に示す。どちらも殆どがNegativeの判定となるが、確率0.25-1.0部分を拡大した右図では確率が1付近にピークがあることが確認できる。読売新聞に関しては訓練データが含まれるので、1付近のピークは鋭くなっている。この図からはNegativeからの分布の裾が0.9付近まで伸びているように思われる。

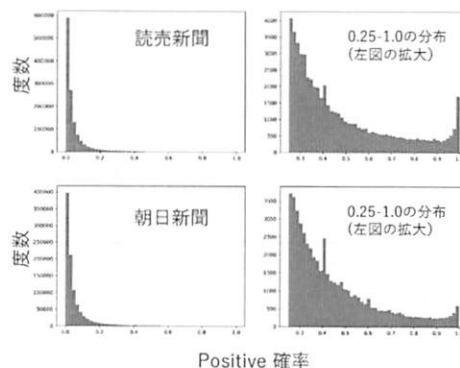


図1：予測確率分布(読売, 朝日)

表 5 : 研究成果ニュース記事選別用キーワードリスト

	キーワードリスト
group1: 動詞系	'掲載', '再録', '発表', '報告'
group1: 名詞系	'科学誌', '科学雑誌', '専門誌', '医学誌', '学会誌', '学会で', '学術誌', '機関誌', '速報誌', '電子ジャーナル', '電子版', 'オンライン版', 'オンライン速報版', '国際誌'
group2	'捏造', '改ざん', '不正', '過誤', '過失', '不適切', '虚偽', '撤回', '悪用', '懲戒'

図 1 の分布と目視の確認より, 研究結果由来のニュース記事は確率が0.75以上に殆どが分布していることが確認できた. しかし, 0.75から0.9付近まではフィーチャー記事も含まれている事, さらには研究不正に関する記事も含まれている事が確認された. よって, この確率0.75以上の朝日, 毎日, 読売新聞のテキストデータを用い, 表 5 に示すキーワードのgroup1: 動詞系と名詞系の共起かつgroup2を含まないものをニュース記事として抽出し, 第 2 段階目のPositive訓練データとして 2 分類のモデルを作成したが, 今回はこの結果については報告しない.

3.5 ニュース, フィーチャー記事の分類

研究成果に対するニュース記事, フィーチャー記事, それ以外の 3 分類モデルを第 3 段階とする. ここでは, 前節の第 2 段階目のPositive記事訓練データをニュース記事訓練データ, 確率0.75以上でニュース記事に分類されなかったものをフィーチャー記事訓練データ, それ以外(Neg)と 3 分類を行った.

この結果, ニュース訓練データに分類されたものが6,108件, フィーチャー記事訓練データに分類されたものが21,396件, その他(Neg)が3,378,709件得られた. この中からランダムに訓練用3,000件, 検証用3,000件をそれぞれ選び, 表 3 の 3 分類機械学習モデルを使って学習させた. この時の精度として, Validation accuracy: 0.966, loss: 0.116 を得た.

このモデルを用い, 表 1 の機関特定記事に対してモデルを適用したpredict結果を得た. ただし, 各分類に対して3,000件は訓練正解データが含まれている. この結果, ニュース記事(確率0.5以上)が9,470件, フィーチャー記事(確率0.5以上)が85,844件, それ以外(Negative)が3,310,899件であった. この分類でニュース記事の確率が0.5以上, またはフィーチャー記事の確率が0.5以上の記事に対して, ニュース記事確率分布を朝日, 毎日, 読売ごとに示したものが図 2 である. 今後独立な記事に対しての検証も必要である.

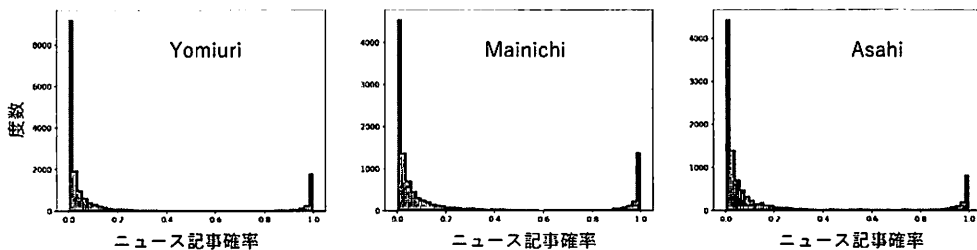


図 2 : ニュースまたはフィーチャー記事確率が 0.5 以上の記事のニュース記事確率分布

3.6 ニュース分類記事の状況

3.5章のPredict結果でニュース記事確率

表6：検証結果の分類コード

分類コード	分類内容
0	無関係(Neg)
1	論文由来の記事 (研究成果ニュース記事)
2	特集等フィーチャー記事
3	研究不正等に関連する記事
4	論説、感想等を掲載した記事 (書評を含む)
5	受賞関連、ランキング
6	用語解説
7	学会、研究会、シンポジウム等の紹介
8	論文由来ではないと思われる分析等の結果
9	人事関連
10	教育関連
11	学術誌の紹介、問題点等
12	研究を基に製品化、会社設立等
13	特許関連

が0.75以上であった記事内容について、目視で検証を行った。検証は新聞記事の見出しで判断できるものは見出しのみで判断し、確認が必要な記事に対しては本文を確認した結果である。表6に検証の結果分類されたグループ、表7に表6のグループに分類された記事数を新聞別、年別に示した。モデル作成時に研究不正に関する記事の除外を行なったが、完全には除去されていなかった。特に2014年の「STAP細胞」関連や「二重投稿」等の記事が残っていた。これらは選別のキーワード群の適正化である程度は対応できるのではないかと思われるが、今後、語の順番を考慮するsequenceモデル(sepCNN) [11]の適応も必要ではないかと思われる。また、分類

表7：新聞別、年別の検証結果 (確率は Predict 結果のニュース記事確率)

新聞	year	合計	0	1	2	3	4	5	6	7	8	9	10	11	12	13
朝日新聞	2010	252		217	14	2	3	4	1	6	2				2	1
	2011	276		231	24	4	2	3	2	4	3			1		2
	2012	314		261	28	6	7	7		4				1		
	2013	353	1	305	32	3	1	3		6	2					
	2014	311		238	39	19	5	5	1		2			1		1
	2015	248		209	33	4	1	1								
	Total	1754	1	1461	170	38	19	23	4	20	9				3	2
朝日新聞	比率		0.06%	83.3%	9.69%	2.17%	1.08%	1.31%	0.23%	1.14%	0.51%			0.17%	0.11%	0.23%
	平均	0.784	0.952	0.897	0.907	0.862	0.903	0.923	0.882	0.833				0.939	0.908	0.872
	std		0.066	0.075	0.069	0.072	0.081	0.062	0.081	0.064				0.083	0.091	0.084
毎日新聞	2011	241	1	218	7	2	2	2	2	6					1	
	2012	346		299	11	11		7	6	8				3		1
	2013	336		307	10	3	2	5	4	1				1	2	1
	2014	294		229	19	25	4	5	7	2			1		2	
	2015	313	1	285	19	1	3	2		2	2					
	2016	305		274	22		1	5	1	2						
	2017	291		256	18	2	2	4	1	2	2					4
	2018	249		212	15	5	3	4	1	1	4				4	
	Total	2375	2	2080	121	49	17	34	3	24	25			1	8	9
毎日新聞	比率	0.08%	87.6%	5.09%	2.06%	0.72%	1.43%	0.13%	1.01%	1.05%			0.04%	0.34%	0.38%	0.08%
	平均	0.81	0.965	0.884	0.941	0.877	0.905	0.856	0.875	0.835			0.907	0.926	0.893	0.837
	std	0.008	0.055	0.08	0.071	0.08	0.072	0.055	0.078	0.07				0.094	0.084	0.099
読売新聞	2011	413		385	13	3		6	2	4						
	2012	465		406	26	8		15	1	7	2					
	2013	458		399	40	9	1	7	2							
	2014	323	1	260	34	15	1	6		2	2	1	1			
	2015	329	1	293	25	1	4		1	2				1	1	
	2016	386		346	32	1		5		2						
	2017	369		332	22	2		5	5	1						2
	2018	319		272	38	2		6	1							
	Total	3062	2	2693	230	41		2	54	11	17	6	1	1	1	3
読売新聞	比率	0.07%	87.9%	7.51%	1.34%	0.07%	1.76%	0.36%	0.56%	0.20%	0.03%	0.03%	0.03%	0.03%	0.10%	
	平均	0.886	0.964	0.907	0.934	0.866	0.909	0.846	0.87	0.88	0.881	0.882	0.99	0.955		
	std	0.16	0.058	0.073	0.071	0.111	0.084	0.084	0.076	0.058				0.048		

コード毎の、ニュース記事確率の平均および標準偏差も示した。フィーチャー記事に分類された記事はニュース記事確率がやや低いが、標準偏差から確率の値だけで分類するのはやや難しい。これらは今後の課題である。今回の分析の結果、新聞記事中の研究結果に基づくニュース記事はプレスリリースの対応記事に比べ、5.7倍ほど見つかった。海外の研究機関の他、大学以外の研究機関の研究成果もあり直接比較は出来ないが、日経プレスリリースと対応する記事は全体のニュース記事中の一部であることが分かった。今後、新聞掲載の要因分析にも考慮が必要である。さらにフィーチャー記事の検証と適正化も今後の課題となる。ただし、この程度の記事分類でもプレスリリースとの対応を行う場合、比較記事数を大幅に削減できるので、Doc2vec等の手法で類似記事の抽出およびプレスリリースとの対応作業の自動化には有用であると思われる。

4 まとめ

大学関連の学術結果であり、原論文のDOIが特定されたプレスリリースと対応した新聞報道記事を教師データとした機械学習により、第1段階として新聞記事データからニュース記事（速報記事）に近いと思われる記事を推定した。その後、学術成果の発表に関するキーワードとの共起を調べることによって、学術成果のニュース記事を選択し（第2段階）、キーワードと共起の無かった記事をフィーチャー記事（特集記事）、それ以外をNegativeとして第3段階の3分類の訓練データを作成した。この訓練データを使い、3分類の機械学習モデルを作成し、朝日、毎日、読売新聞の記事データに対して予測（Predict）を行った。

この結果の学術成果のニュース記事（速報記事）と推定した確率が0.75以上のデータについて目視でニュース記事であるかどうか確認したところ、推定された記事の内83%-89%の割合でニュース記事であった。第1段階で訓練データに含まれていない朝日新聞の正解率がやや低かったが、実用になる水準でニュース記事を抽出できることがわかった。間違いになる記事は研究不正や受賞に関するものが多かったが、今後訓練データを作る際に排除に使ったキーワードを更新すること及び、今回試していないsequenceモデル（sepCNN）を適用することによって改善されるのではないかとと思われる。また、このモデルの学習データに関しては、正解データ（POS）の数が不正解データ（NEG）に対して非常に少なく、今後オーバーサンプリング等の手法も取り入れていきたいと思っている。しかし、現時点の分類でも目的の学術研究に関する新聞記事（ニュースおよびフィーチャー記事）の抽出により全体の3%程度まで記事を減らすことができるので、プレスリリースとの一致記事の同定が容易になり、類似記事の推定モデルの作成により自動推定も可能になるかと思われる。これも今後の課題である。

謝辞

本研究は科学研究費基盤研究C（16K00455, 20K12569）の助成を受けて行った成果である。

参考文献

- [1] 西澤 正己, 孫 媛, 「学術研究のメディア報道における定量的調査研究」, 情報知識学会誌, Vol. 22, No. 2, pp. 138-143, (2012).
- [2] 西澤 正己, 孫 媛, 「学術研究のメディア報道における定量的調査研究 -プレスリリースと新聞報

- 道の関係」, 情報知識学会誌 Vol. 23, No. 2, pp. 279-285, (2013).
- [3] 西澤 正己, 孫 媛, 「学術研究のメディア報道における定量的調査研究 -プレスリリースと2つの異なる全国紙の比較-」, 情報知識学会誌 Vol. 24, No. 2, pp. 216-223, (2014).
- [4] 西澤 正己, 孫 媛, 「学術研究のメディア報道における定量的調査研究 -プレスリリースと原論文, そして新聞報道の関連-」, 情報知識学会誌 Vol. 25, No. 2, pp. 131-2139, (2015).
- [5] 西澤 正己, 孫 媛, 「学術研究のプレスリリースと原論文に関する詳細分析」, 情報知識学会誌 Vol. 26, No. 2, pp. 116-123, (2016).
- [6] 西澤 正己, 孫 媛, 「学術研究のメディア報道における定量的調査研究 -プレスリリースと原論文, オルトメトリックスとの関連分析-」, 情報知識学会誌, Vol. 27, No. 2, pp. 213-218, (2017).
- [7] 西澤 正己, 孫 媛, 「プレスリリースと原論文, オルトメトリックス指標との関連分析 -過熱報道の分析に与える影響-」, 情報知識学会誌, Vol. 28, No. 2, pp. 121-126, (2018).
- [8] 西澤 正己, 孫 媛, 「プレスリリースと新聞報道との関連分析 -外部要因の分析-」, 情報知識学会誌, Vol. 29, No. 2, pp. 116-122, (2019).
- [9] 日経テレコン 21:
<http://t21.nikkei.co.jp/> (参照 2021年4月)
- [10] 日外アソシエーツ, 学術研究・開発研究のための言語資源コーパス:
<http://www.nichigai.co.jp/sales/corpus.html> (参照 2021年4月)
- [11] Developers Google.com Machine-Learning Guides: <https://developers.google.com/machine-learning/guides/text-classification> (参照 2021年4月)
- [12] 形態素解析エンジン MeCab (和布蕪):
<http://taku910.github.io/mecab/> (参照 2021年4月)

第29回年次大会予稿

物語会話文での自称詞と対称詞の分類

Classification of first personal pronoun and second personal pronoun in conversational sentences within novel corpus

村井源^{1*}, 石川一稀¹
Hajime MURAI^{1*}, Ishikawa KAZUKI^{1*}

¹ はこだて未来大学

Future University Hakodate

〒041-8655 北海道函館市亀田中野町116-2

E-mail: h_murai@fun.ac.jp

*連絡先著者 Corresponding Author

日本語の会話の文体は話者の属性や、話者と聴者の関係性など多種多様な要因によって複雑に変化する性質をもっている。本研究では日本語の自称詞、対称詞の表現の多様性に着目し、物語会話文においてどのような表現が用いられているかを現代日本語書言葉均衡コーパスに基づいて調査し分類を行った。また、同時に用いられやすい自称詞・対称詞の組みあわせのパターンを因子分析により抽出した。

The styles of Japanese conversation change in a complex manner depending on various factors such as the attributes of the speaker and the relationship between the speaker and the listener. In this study, we focused on the variety of expressions of Japanese first personal pronouns and second personal pronouns, and investigated and classified the appeared expressions within conversation sentences based on the Balanced Corpus of Contemporary Written Japanese. In addition, the co-occurrence patterns of first personal pronouns and second personal pronouns were extracted by factor analysis.

キーワード: 会話文, 文体, 自称詞, 対称詞, 物語テキスト

Conversational sentences, Style, First personal pronoun, Second personal pronoun, Novel text

1 はじめに

日本語の会話における文体は、人為的に作成された物語における会話においても

[1], 自然な日常会話においても[2], 話者の属性や話者と聴者の関係性などの種々の要因によって大きく異なる性質を持つ

ていることが、機能語などの分析により明らかになっている[3].

会話における文体と話者の属性の関係性の特徴を抽出できれば、言語学的な知見を得られるのみならず、自然言語処理技術の向上が見込まれ、より人間に近い文章自動生成などの応用的な用途でも有用であると考えられる。本研究ではすでに先行研究で明らかになっている機能語よる文体的特徴に加えて、日本語において特徴的であり、かつ文体的な特徴の出現しやすい対象である呼称表現[4]の差違を取り扱う。

日本語の会話には多数の自称表現、他称表現が含まれる。このような表現の代表的な語彙としては人称代名詞（「わたし」、「あなた」、「彼」「彼女」など）があるが、それ以外にも氏名を用いるもの、肩書や職業を用いるものなど多様な語彙が存在する。自称表現や他称表現のうち一人称的に用いられるものを自称詞、二人称的に会話の相手を指すものを対称詞とも呼ぶ。

自称詞、対称詞に関してはこれまで主に言語学やジェンダー論等の観点から研究が進められてきており、言葉の性差の観点からドラマでの呼称を分析したもの[5]、歌詞書き換え課題から呼称とジェンダーの意識変化を分析したもの[4]などがある。

また先行研究では呼称中で頻出の人称代名詞等を主な分析の対象とするものが多いが、緒方[6]は、より詳細な呼称表現を取り扱い、「人名・敬称」「職業名・役職名」「人称代名詞」「一般名詞類」「親族名称」などに分類している。この詳細な分類をより精緻化する目的でマンガにおける自称詞と対称詞の分類を行った研究なども存在しており[7]、自称詞、対称詞の利用がドラマ・マンガなどのメディアによっ

て異なる可能性も指摘されてきている。

これらの先行研究より自称詞、対称詞に関する種々の知見が蓄積されつつあるが、小説等の読むことを前提とした長文テキストでの自称詞、対称詞利用の傾向は明らかになっておらず、また自称詞、対称詞の利用パターンの詳細も明らかではない。そこで本研究では小説の物語テキストを対象として自称詞、対称詞の出現傾向を調査して先行研究と比較し、また出現パターンを抽出するため計量的な解析を実施した。

2 分析の対象

本研究では、一般的な物語テキストにおける会話文の特徴を分析するため、『現代日本語書き言葉均衡コーパス』[8] (The Balanced Corpus of Contemporary Written Japanese, 以下BCCWJ)をデータとして用いた。BCCWJには販売された書籍のISBNに基づくサンプリングで作られたコーパスと、図書館に所蔵された書籍のリストに基づくサンプリングで作られたコーパスが含まれているが、実際に流布している物語テキストにおける会話文の実態をとらえることを目的とするため、図書館の蔵書に基づくコーパスを用いた。また、日本語文学テキストを対象とするため十進分類で日本語の物語・小説自体にあたる913に分類された可変長のサンプルのテキスト(文章構成的なまとまりで区切られた部分を抽出したテキスト)のみを選択した。結果として条件に該当する2189サンプルが抽出されたが、そこからランダムに100サンプルを抽出し人手でタグ付けを行った。

本研究では物語中に会話として明示的に表れる部分のみに限定した分析を行う。タグ付けの対象はテキスト中の時間軸で

現在発話されている「」や『』でくくられた会話箇所とした。回想、置手紙、語り手や登場人物による想像上の会話などその場で行われていない発話的な描写は除外するが、電話の会話やファンタジーにおけるテレパシーでの会話は分析対象に含めた。結果として、「」や『』などの区切り記号を単位とすると日本語の物語文100サンプル中の5632発話が抽出された。ただし、小説では発話の途中で説明や情景描写の地の文を挟むなどの描写法が多用されるため「」の数が発話の数とは限らない。

これらの発話に対し、発話者とその性別、年齢(若年層、成年層、老年層の三種類に人手で分類)に関しては国立国語研究所で作成されたBCCWJ Speaker Information Corpus[9]を用いてタグを付与し、それに加え発話者から見た発話の相手の関係性(友人、家族、上司など)もタグとして付与した[10]。なおBCCWJ Speaker Information Corpusの話者属性付与は、会話文内の情報に合わせて地の文や、サンプル外のテキスト中の情報も含めて行われている。またBCCWJ Speaker Information Corpusには対話を構成しない発話(内言等)も含まれているが本研究では対話構造の分析を目的とするため、内言等は省略し、より単純な構成としている。

3 自称詞と対称詞の分類

抽出したBCCWJ中の物語テキスト中の会話文に対して、人手による目視で自称詞、対称詞の抽出を行った。また類似する自称詞、対称詞の語彙を分類しカテゴリーを作成した。自称詞の結果を表1、対称詞の結果を表2に示す。表中の[]は該当する複数の表現が存在することを示している(例:

[家族関係]の中には、「父」「母」「兄」「姉」などが入りうる)。

表1より自称詞の利用では種々の人称代名詞が大部分を占める結果となったが、人称代名詞(全体の91%)、氏名、家族関係等の比率に関しては先行研究の結果とほぼ一致している[7]。一方で表2の対称詞では人称代名詞が合計で約46%、氏名が37%と先行研究(人称代名詞が25%、氏名が60%、[7])とは大きく異なる数値となっている。

4 呼称の利用パターン抽出

得られた自称詞と対称詞の分類を用いて、自称詞と対称詞が用いられる文脈のパターンを抽出するために、同じ話者の会話文に出現する自称詞と対称詞のカテゴリーを一つのベクトルとして因子分析を行った。分析においては、BCCWJでの同じサンプルテキスト中の会話文において、同じ話者が同じ関係性の相手に発話する会話文をまとめて(例えば、特定の登場人物が友人と会話する場合の会話文全てなど)、分析用のデータセットを作成した。パラメータとしては、話者の属性(年齢・性別)に合わせて出現頻度の高い話者と聴者の関係性(家族、友人、親戚)を含め、自称詞と対称詞それぞれの分類に合わせてデータとした。統計的処理を行うため、自称詞と対称詞合計の出現頻度5以上の発話分のセットに限定し、かつ分析に用いる自称詞と対称詞の分類も合計の出現頻度が5以上の物に限定した。結果として、27変数、113件のデータが得られた。

平行法を用いた結果因子数が4と推定されたため、プロマックス回転を用いて因子分析を実施した。得られた因子負荷量の結果を表3に示す。

表1 自称詞の出現数とカテゴリー

カテゴリー名	語彙とパターン	数
私・私達	私,わたくし,私ども,私達,わたしや	423
俺・俺達	俺,俺さま,俺達,俺等	197
僕・僕達	僕,僕達,僕ら	79
あたし・あたしたち	あたし,あたくし,あちき,うち,あたしたち	69
我・我々	我,われ,吾,わが,我々,我ら	42
氏名	[氏名],[氏],[名],[名]+ちゃん	32
わし	儂,わし,わしら	16
肩書・職業	[肩書],[氏名]+[肩書],[氏]+[肩書],[職業],[組織],本官	15
拙者		12
自分		11
家族関係	[家族関係],[家族関係]+さん,お+[家族関係]+さん	8
あだ名		5
人数	二人	3
位置関係	こちら,お互い	3
年齢・性別	女の子	1
その他		3

表2 他称詞の出現数とカテゴリー

カテゴリー名	語彙とパターン	数
氏名	[氏],[氏]+くん,[氏]+君,[氏]+さま,[氏]+さん,[氏]+どの,[氏]+殿,ミスター+[氏],[氏名],[氏名]+さま,[氏名]+さん,[氏名]+どの,[氏名]+君,[氏名]+公,[氏名]+殿,[名],[名]+くん,[名]+クン,[名]+サア,[名]+さーん,[名]+さん達,[名]+どの,[名]+どん,[名]+君,[名]+くん,[名]+様,[名]+さん,[名]+ちゃん,お+[名]+サア,[名]+じいさん	418
あなた	あなた,あなた方,あなた様,あなたたち,あなた達,貴女,貴殿	210
君	君,きみ,キミ,きみたち,君たち	119
お前	お前,おまえ,お前さん,お前さんたち,おまえたち,お前ら	109
肩書・職業	[肩書],[肩書]+方,[肩書]+閣下,[肩書]+殿下,[肩書]+殿,[肩書]+様,氏+[肩書],名+[肩書],[職業],[職業]+サア,[職業]+さん,[職業]+の方,[職業]+屋,主上,上様,貴官,殿下	73
あんた	あんさん,あんた,あんたがた,あんたたち,あんたら,あんちゃん	39
てめえ・貴様	てめえ,てめえら,貴様,キサマ,貴様ら,こいつ,こいつら,このあま,このやろう,この野郎,野郎,やつ,奴	31
家族関係	御+[家族関係],お+[家族関係]+さん,お+[家族関係]+ちゃん,[家族関係],[家族関係]+さん,[家族関係]+者,[家族関係]+上	27
年齢・性別	爺,オヤジ,兄貴,男の子,坊や,女,あま,姐さん,姐御,奥さん,小娘	17
汝	汝,おたく,お主,おぬし,わぬし	14
自分	御自分,自分,おのれ	13
皆	皆,みなさん,皆さん,みんな	11
悪口	馬鹿者,卑劣な男,不肖の子,悪鬼風情,けだもの,ばか,ばかな男,卑怯者	8
人数	お一人,ふたり,二人,おふたり	7
誰	だれ,誰,だれか,どなた	5
あだ名		4
団体	[団体],[団体]+さん	4
位置関係	そちら,そなた	4
その他		32

表3 自称詞と対称詞の出現パターンの因子

	F1	F2	F3	F4
自称_僕・僕達	-0.7794	0.25226	-0.17665	-0.05264
対称_君	-0.7185	0.27262	-0.1794	-0.0849
自称_私・私達	0.28034	0.18497	-0.25995	-0.09131
対称_肩書・職業	0.20651	0.19309	-0.04401	-0.08798
対称_皆	0.16446	0.07571	0.07526	-0.0481
対称_人数	-0.1277	-0.02513	0.03728	-0.06261
対称_年齢・性別	0.11001	0.09383	0.06493	-0.04817
関係_友人	0.05289	-0.7061	0.03779	-0.13779
対称_氏名	0.08181	-0.4849	0.30443	-0.12315
自称_あたし・あたしち	0.11485	-0.4612	-0.3014	-0.0391
年齢	-0.14427	0.39917	-0.03281	0.12747
自称_我・我々	0.05365	0.1293	0.0819	0.01026
自称_氏名	0.05702	0.08545	0.01312	0.06494
自称_俺・俺達	0.08796	-0.13483	0.71714	0.00953
性	0.41575	-0.3541	-0.5243	-0.02865
対称_お前	0.08615	0.00311	0.4023	0.01158
対称_てめえ・貴様	0.08287	-0.00704	0.34535	-0.05033
対称_汝	0.09574	0.0497	0.24852	0.03639
対称_あなた	0.19304	0.05577	-0.19939	-0.12064
自称_自分	0.06191	0.06415	-0.07732	-0.05232
自称_拙者	-0.01261	0.00625	0.04395	-0.02923
関係_家族	-0.00311	0.11272	-0.07208	1.00921
対称_家族関係	0.04485	0.10067	-0.02323	0.40717
自称_家族関係	-0.0301	0.13055	0.06818	0.32463
自称_わし	-0.03134	0.04836	-0.01253	0.30952
対称_あんた	0.02299	-0.0027	-0.05369	0.066
関係_親戚	-0.03628	0.05374	-0.05439	-0.05601

表3より、第一因子は「僕・僕達」という自称と、「君」という対称が女性（性の正が女性，負が男性）の話者の場合には用いられにくいことを示している。よって第一因子を一般男性話者因子と名付ける。

第二因子は若年の女性が友人との会話で自称として「あたし・あたしち」を用い、かつ他称として氏名に基づく呼称を用いることを示している。よって第二因子を

友人会話女性話者因子と名付ける。

第三因子は男性が自称として「俺」、対称として氏名もしくは「お前」「てめえ」を用いることを示している。よって粗野男性話者因子と名付ける。

第四因子は家族の会話で家族の関係が呼称として用いられることを示している。よって家族話者因子と名付ける。

5 結論と今後の課題

本研究は日本語会話文での話者属性等と文体特徴の関係を調査するため、自称詞、対称詞を対象として、物語テキストでの会話文に限定して該当語彙の抽出と分類を行った。また自称詞、対称詞の出現パターンに因子分析を行い、大きく4種類の頻出パターンがあることが明らかになった。

自称詞の語彙は先行研究の結果と類似していたが、対称詞の出現傾向はマンガのテキストを対象とした先行研究と大きく異なる数値となった。よって今後会話文のメディアによる差違とその理由などのより詳細な分析が必要と考えられる。また、同じカテゴリー内でも敬称や敬意を含んだ表現と含まない表現の特徴の差違、会話文の意図や機能による文体の差違などは未解明の課題として残っており、今後の研究のさらなる発展が必要と考えられる。

謝辞

本研究は科学研究費基盤研究C「物語文と日常会話文における発話文体と話者属性の分析」、国立国語研究所「大規模日常会話コーパスに基づく話し言葉の多角的研究」、「日本語日常会話コーパスの構築」の支援を受けた。

参考文献

- [1] Hajime Murai: “Factor Analysis of Utterances in Japanese Fiction-writing Based on BCCWJ Speaker Information Corpus”, *Advances in Human-Computer Interaction*, vol. 2018, pp. 1-9, 2018.
- [2] Hajime Murai: “Factor Analysis of Japanese Daily Utterance Styles”, *LREC 2018 Joint Workshop LB-ILR2018 and MMC2018 Proceedings*, pp. 26-29, 2018.
- [3] Hajime Murai: “Japanese Daily Utterance Styles: A Factor Analysis based on Balanced Corpus”, *Proceedings of PACLIC 33, The 33rd Pacific Asia Conference on Language, Information and Computation*, pp. 539-546, 2019.
- [4] 菊地悟: 「現代大学生における言語の性差意識--男女の歌詞書き替えの課題から」, *岩大語文*, Vol. 12, pp. 109-101, 2007.
- [5] 張文博, 劉爽, 川口良: 「対称詞に見ることばの性差について—ドラマの会話を資料として—」, *文教大学大学院言語文化研究科紀要*, Vol. 4, pp. 31-53, 2018.
- [6] 緒方隆文: 「呼称のカテゴリー分析: 自称詞・対称詞・他称詞」, *筑紫女学園大学・筑紫女学園大学短期大学部紀要*, Vol. 10, pp. 1-13, 2015.
- [7] ニン・ティ・ニャン・ヴァン: 「場面によって変化する自称詞及び対称詞の研究」, *人間生活文化研究*, Vol. 2016, No. 26, pp. 499-503, 2016.
- [8] 国立国語研究所: 「現代日本語書き言葉均衡コーパス」 https://pj.ninjal.ac.jp/corpus_center/bccwj/, (参照 2021-3-8).
- [9] Makoto Yamazaki, Yumi Miyazaki and Wakako Kashino: “Annotation and Quantitative Analysis of Speaker Information in Novel Conversation Sentences in Japanese”, *Proceedings of the 11th International Conference on Language Resources and Evaluation*, pp. 1078-1081, 2018.
- [10] 村井源: 「日本語小説の会話文タグ付コーパスの開発に向けて」, 第31回人工知能学会全国大会論文集, 1D2-OS-29a-2 (PDF), 2017.

第29回年次大会予稿

モバイルゲームにおける女性キャラクターの魅力的要素の分析

Analysis of Attractive Elements of Female Characters in Mobile Games

宇田朗子^{1*}, 村井源²

Akiko UDA^{1*}, Hajime MURAI²

1 株式会社エイチ・アイ・ディ

HID Co., Ltd.

〒060-0003 札幌市中央区北3条西2丁目10番地2 札幌HSビル

2 公立はこだて未来大学

Future University HAKODATE

〒041-8655 北海道函館市亀田中野町116-2

*連絡先著者 Corresponding Author

これまでキャラクターに関する研究において、外見についての研究や男性キャラクターにのみ着目した研究は行われてきたが外見以外の魅力に着目した研究や女性キャラクターにのみ着目した研究はあまり行われていない。本研究ではキャラクターの誘因魅力と継続魅力の違いに着目し、女性キャラクターの魅力についての分析を行った。キャラクターのどのような要素に魅力を感じているかアンケートを行い、回答を収集した。アンケートの結果を用いてカイ二乗検定を行い、誘因魅力と継続魅力に魅力的要素の違いに有意な差があることが明確になった。また、アンケート結果を用いて因子分析を行い、どのような要素の組み合わせに魅力を感じるかを調査した。結果として、キャラクターを印象付ける要素が誘因魅力、あとから追加できる要素が継続魅力として有力であることが分かった。

In character research, research on appearance and research focusing only on male characters have been conducted. However, research focusing on attractiveness other than appearance and research focusing only on female characters have not been conducted much. In this research, we focus on the difference between the incentive attractiveness of a character and the continuous attractiveness of characters, and analyze the attractiveness of a female character. First, we investigated how attractive female characters are by questionnaire. Chi-square test and factor analysis were performed using the results of this questionnaire. Chi-square test revealed there was significant difference between incentive attractiveness and continuous attractiveness. In addition, factor analysis was performed to find out which combination of factors was attractive. Incentive attractiveness was an element that impressed the character, and continuous attractiveness was an element that could be added later to the character settings. As a future task, it is necessary to investigate whether similar results will be obtained with other mobile games.

キーワード: キャラクター, モバイルゲーム, アニメ, 漫画, アンケート
Character, Mobile game, Anime, Comic, Questionnaire

1 序論

モバイルゲームとは携帯電話やスマートフォン等を用いて遊ぶことができるゲームである。モバイルゲームとは携帯電話やスマートフォン等を用いて遊ぶことができるゲームである。株式会社三菱総合研究所のスマホゲームに関するアンケート結果によると、モバイルゲームに課金をする理由で最も多いものは欲しいキャラクターやアイテムが存在するためであり[1]、キャラクターはモバイルゲーム上においてユーザーに課金させる重要な要素であるといえる。また、モバイルゲームは運営されていく中で様々な要素が追加され発展していくという特徴をもつ。そのため、モバイルゲームに登場するキャラクターにも様々な追加要素が存在し、その要素が追加されるたびに、キャラクターの魅力に感じる要素も増えていくと考えた。そのため、キャラクターを好きになるきっかけの魅力的要素とキャラクターを好きでいさせる魅力的要素に違いがあると仮定した。本研究ではキャラクターを好きになるきっかけの魅力的要素を誘因魅力、キャラクターを好きでいさせる魅力的要素を継続魅力と名付け、この2つの違いに着目し、どのような要素を持つ女性キャラクターが魅力的であるかを明確にすることを目的とする。

2 関連研究

キャラクターの外見に関する研究として、漫画の登場人物の顔の分析[2]やアニメキャラクター評価のための外見の特徴抽出[3]、キャラクターの縦横比から受ける印象についての研究[4]が行われている。ま

た、キャラクターのセリフや性格を用いた性格推定[5]や少年漫画におけるキャラクターの魅力的要素の分析[6]などキャラクターに関する研究が行われている。

このようにキャラクターの魅力についての研究は行われているが、女性キャラクターのみに着目した研究やモバイルゲームのキャラクターに関する研究はあまり行われていない。

3 分析対象

分析対象として、『アイドルマスターシンデレラガールズ[7]』を選定した。『アイドルマスターシンデレラガールズ』は190人の女性キャラクターを有しており、様々な魅力的要素を抽出できると期待できる。また、『アイドルマスターシンデレラガールズ』のモバイルゲームである『アイドルマスターシンデレラガールズスターライトステージ[8]』はアプリの総ダウンロード数が2500万を超えている。加えて、『アイドルマスターシンデレラガールズ』は9年、『アイドルマスターシンデレラガールズスターライトステージ』は5年という長期間サービスが継続し続けているため人気や知名度が高いと推測される。よって、本研究の目的に沿った分析対象であると考えられる。

4 手法

4.1 アンケートの実施

作品のファンがキャラクターのどのような要素に魅力を感じるか詳細に知るため、対象作品を知っている人を対象にネット上でアンケートを実施した。表1にアンケート内容を示す。

アンケートは2020年10月6日から10月18日までの2週間実施し2631件の回答が集まった。そのうち、回答が重複しているものを除いた2603件からキャラクターのどのような要素に魅力を感じているかを分析した。

表 1 アンケートの質問項目

質問項目
アイドルマスターシンデレラガールズで好きなキャラクターは誰ですか。
キャラクターを好きになった時期がわかれば教えてください。
そのキャラクターを好きになったきっかけを教えてください。
次の外見的要素のうち、あなたがそのキャラクターを好きになったきっかけに当てはまるものはどれですか。
次の内面的要素のうち、あなたがそのキャラクターを好きになったきっかけに当てはまるものはどれですか。
キャラクターを好きになったきっかけにギャップは存在しますか。
キャラクターを好きになったきっかけにギャップが存在すると回答した人に質問します。 どの要素とどの要素のギャップが魅力的だと感じますか。○△な要素と△△な要素のように回答してください。
その他要素で好きになったきっかけについて詳しく書けることがあれば教えてください。
現在そのキャラクターを好きなどころを教えてください。
次の外見的要素のうち、現在あなたがそのキャラクターを好きな理由に当てはまるものはどれですか。
次の内面的要素のうち、現在あなたがそのキャラクターを好きな理由に当てはまるものはどれですか。
現在キャラクターを好きな理由にギャップは存在しますか。
現在キャラクターを好きな理由にギャップが存在すると回答した人に質問します どの要素とどの要素のギャップが魅力的だと感じますか。○△な要素と△△な要素のように回答してください。
その他要素で現在キャラクターが好きな理由について詳しく書けることがあれば教えてください。

4.2 アンケート結果の分析

アンケート結果よりキャラクターの魅力进行分析するため、カイ二乗検定と因子分析を行った。カイ二乗検定はキャラクターの誘因魅力と継続魅力に差異があることを明確にするために行った。因子分析は魅力を感じる要素の組み合わせを調べるため

に行った。

5 結果

5.1 アンケート結果

本アンケートの結果を項目ごとに集計した。図2はキャラクターを好きになったきっかけと現在好きな理由の差異を集計したものである。このカテゴリーを総合カテゴリーと名付ける。総合カテゴリーにおいて、男女ともに、どの項目も誘因魅力よりも継続魅力のほう数値が多いことが分かる。

図3はキャラクターの外見についての項目を集計したものである。このカテゴリーを外見カテゴリーと名付ける。キャラクターの外見に魅力を感じる要素がないという回答以外は、誘因魅力より継続魅力の数値が大きいことが分かる。

図4はキャラクターの内面についての項目を集計したものである。このカテゴリーを内面カテゴリーと名付ける。「なし」と「弱気」の項目以外は誘因魅力よりも継続魅力のほう数値が大きいことが分かる。

5.2 カイ二乗検定

アンケートの結果から誘因魅力と継続魅力に差異があることを確認するためにカイ二乗検定を行った。表2、表3、表4はそれぞれ総合カテゴリー、外見的要素、外見的要素の誘因魅力と継続魅力に対してカイ二乗検定を行った結果である。図機能のp値が0.1以下ならば+、0.05以下ならば*、0.01以下ならば**で示している。

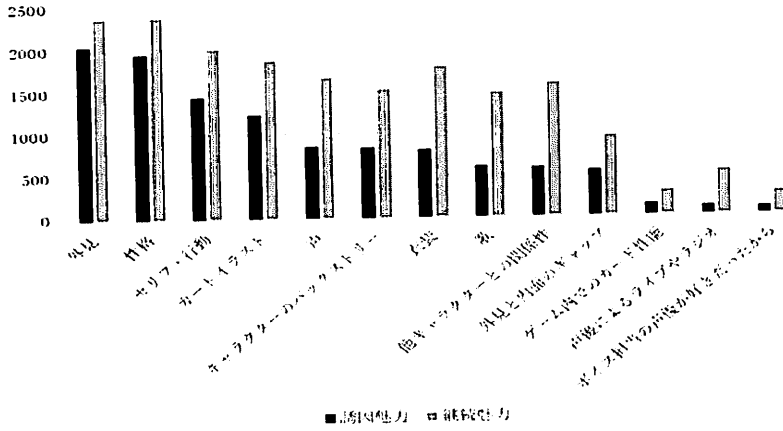


図 1 総合カテゴリーにおける誘因魅力と継続魅力の差

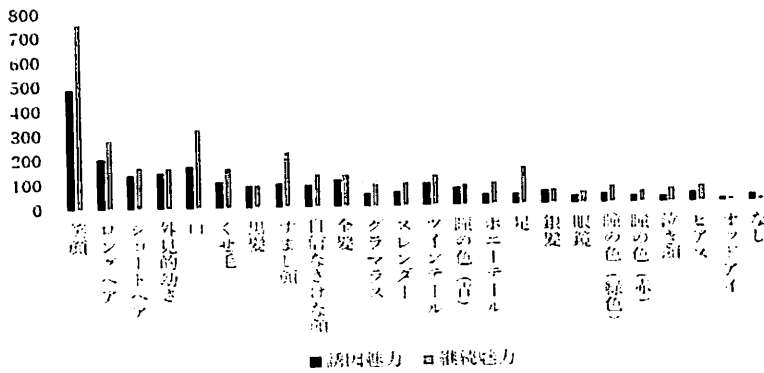


図 2 外見における誘因魅力と継続魅力の差

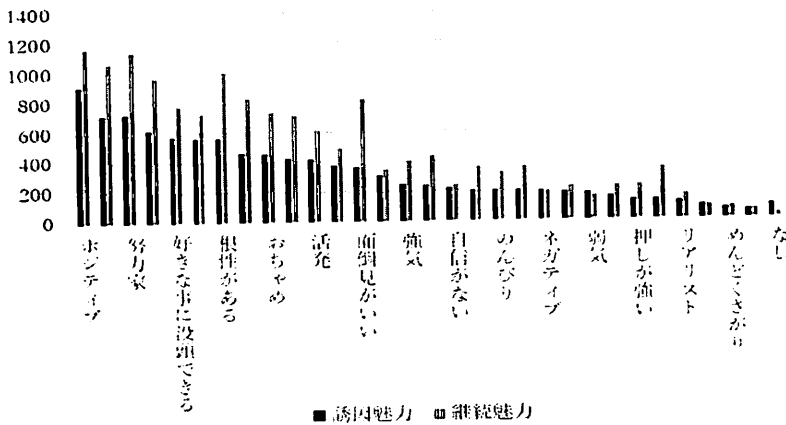


図 3 内面における誘因魅力と継続魅力の差

表 2 総合カテゴリーにおける残差分析の結果

	調整済み標準残差	p 値
外見	13.130	**
性格	11.011	**
セリフ・行動	5.609	**
カードイラスト	2.776	**
声	-3.942	**
バックストーリー	-2.092	*
衣装	-7.038	**
歌	-8.303	**
他キャラクターとの関係性	-10.513	**
ギャップ	-0.971	ns
カード性能	-2.531	*
ライブやラジオ	-11.072	**
ボイス担当の声優	-5.693	**

表 3-1 外見の要素の残差分析の結果

	調整済み標準残差	p 値
笑顔	-0.025	ns
ロングヘア	1.706	+
ショートヘア	2.957	**
外見的幼さ	3.283	**
口	-4.327	**
くせ毛	0.537	ns
黒髪	3.198	**
すまし顔	-4.369	**
自信なさげな顔	-0.968	ns
金髪	1.972	*
グラマラス	0.809	ns
スレンダー	0.356	ns
ツインテール	0.714	ns
瞳の色(青)	1.658	+
ポニーテール	-1.767	+
足	-7.921	**
銀髪	2.171	*
眼鏡	0.814	ns
瞳の色(緑色)	-0.869	ns
瞳の色(赤)	-0.343	ns
泣き顔	-3.814	**
ピアス	-1.493	ns
オッドアイ	0.035	ns

5.3 因子分析

魅力を感じる要素の組み合わせを調べるために、2603 件のアンケート結果に対して因子分析を行った。平行法を用いた結果、因子数は 41 となった。表 5, 表 6,

表 7, 表 8, 表 9, 表 10 に因子分析の結果の一部を示す。要素名で 1 とついているものは好きになったきっかけの要素、要素名の後に 2 とついているものは現在好きな理由の要素について示している。

表 4 内面的要素の残差分析の結果

	調整済み標準残差	p 値
ポジティブ	3.858	**
強い信念	0.446	ns
努力家	-0.822	ns
まじめ	-0.697	ns
好きな事に没頭できる	2.004	*
明るい	2.777	**
根性がある	-3.285	**
自分の弱さと向き合える	-3.027	**
おちゃめ	-1.066	ns
負けず嫌い	-1.698	+
活発	0.510	ns
自由奔放	1.929	+
面倒見がいい	-6.790	**
おっとり	3.431	**
強気	-0.905	ns
お姉さんキャラ	-2.613	**
自信がない	3.496	**
勝気	-2.055	*
のんびり	-0.833	ns
いじられ役	-2.163	*
ネガティブ	3.865	**
恥ずかしがり屋	2.547	*
弱気	4.871	**
弱みを見せない	0.449	ns
押しが強い	-1.489	ns
ツッコミ役	-5.990	**
リアリスト	0.391	ns

6 考察

6.1 誘因魅力と継続魅力の差について

6.1.1 総合

全体として、誘因魅力は「外見」・「性格」・「セリフ・行動」・「カードイラスト」が有意に多い。継続魅力は「声」・「バックストーリー」・「衣装・歌」・「他キャラクターとの関係性」・「カード性能」・「ラ

イブやラジオ」・「ボイス担当の声優」が有意に多い。キャラクターを知った時に目に入りやすいものが誘因魅力として、あとから知ることが多い要素や追加されることが多い要素が継続魅力として出現している。

表 5 因子分析の結果 1

要素名	因子 1	因子 2	因子 3
ネガティブ 2	0.69	-0.11	0.11
ネガティブ 1	0.68	-0.09	0.07
自信がない 1	0.66	0.00	-0.01
自信がない 2	0.65	0.03	-0.05
弱気 1	0.64	0.02	-0.04
弱気 2	0.61	0.02	0.03
自信なさげな顔 1	0.58	0.04	0.02
すぐへこたれる 2	0.55	-0.06	-0.05
すぐへこたれる 1	0.51	-0.09	-0.09
自信なさげな顔 2	0.49	0.19	0.04
泣き顔 1	0.33	0.04	-0.01
泣き顔 2	0.30	0.17	0.02
背景 2	0.03	0.69	0.07
衣装 2	-0.02	0.61	-0.09
イラスト 2	0.00	0.56	0.00
セリフ 2	0.05	0.52	0.00
人間関係 2	0.04	0.51	-0.06
関連作品 2	0.03	0.42	-0.10
性能 2	-0.05	0.36	-0.08
ギャップ 2	-0.03	0.36	-0.06
性格 2	-0.01	0.31	0.12
努力家 1	-0.04	-0.15	0.81
努力家 2	-0.07	0.06	0.75
まじめ 2	0.02	-0.01	0.72
まじめ 1	0.02	-0.22	0.71
強い信念 2	-0.07	0.18	0.39
根性がある 2	-0.01	0.21	0.38
根性がある 1	0.00	-0.03	0.37
強い信念 1	-0.05	-0.01	0.37
ライブ 2	-0.06	0.37	-0.08
歌 2	0.05	0.38	0.05

6.1.2 外見

誘因魅力は「ショートヘア」・「外見的幼さ」・「黒髪」・「金髪」・「銀髪」が有意に多い。継続魅力は「口」・「すまし顔」・「足」・「泣き顔」が有意に多い。男女ともに外見的な幼

さ・髪色・髪型が誘因魅力として出ており、ユーザーは幼い外見や髪型に魅力を感じていることが分かる。それに対して継続魅力は口・足などの性的魅力を感じやすい項目とすまし顔・泣き顔等の表情に関するものが上がっており、キャラクターをよく見てから気づく魅力や最初に見せていたものと別のものを見せてくるものがあげられている。

表 6 因子分析の結果 2

要素名	因子 4	因子 5	因子 6
勝気 1	0.91	0.08	0.06
勝気 2	0.89	0.09	0.01
強気 2	0.72	0.01	0.00
強気 1	0.71	0.00	0.00
負けず嫌い 1	0.35	-0.14	0.00
負けず嫌い 2	0.31	-0.14	0.00
押しが強い 2	0.15	0.08	-0.05
明るい 1	0.01	0.85	0.05
明るい 2	-0.02	0.81	0.02
活発 2	0.10	0.60	-0.01
活発 1	0.09	0.59	0.01
ポジティブ 1	0.00	0.55	0.04
笑顔 1	-0.07	0.35	-0.04
笑顔 2	-0.08	0.23	0.02
内面と内面 2	-0.03	0.05	0.01
女性	0.05	0.04	1.05
男性	-0.06	-0.05	-1.05
ポジティブ 2	0.50	0.00	0.12

表 7 因子分析の結果 3

要素名	因子 7	因子 8	因子 9
背景 1	0.50	-0.02	-0.12
人間関係 1	0.47	0.01	-0.03
性格 1	0.45	0.01	-0.04
セリフ 1	0.40	0.02	0.01
ギャップ 1	0.29	0.11	-0.02
関連作品 1	0.13	0.00	0.02
外面と真の顔 1	0.05	0.00	0.00
銀髪 2	0.09	1.03	-0.02
銀髪 1	0.10	0.91	-0.03
口 2	-0.17	-0.01	0.94
口 1	-0.02	-0.02	0.73
足 2	-0.09	0.00	0.35
衣装	0.31	0.01	0.06

表 8 因子分析の結果 4

要素名	因子 10	因子 11	因子 12
リアリスト 1	1.03	-0.07	-0.13
リアリスト 2	0.85	-0.04	-0.10
お姉さんキャラ 2	-0.07	0.83	0.00
お姉さんキャラ 1	-0.03	0.78	-0.02
面倒見がいい 1	0.00	0.52	-0.01
面倒見がいい 2	-0.01	0.51	0.02
スレンダー 2	-0.11	-0.01	0.93
スレンダー 1	-0.09	-0.01	0.87

表 9 因子分析の結果 5

要素名	因子 13	因子 14	因子 15
声優 2	0.51	0.01	0.02
ライブ 1	0.47	0.00	-0.03
声優 1	0.47	0.00	-0.02
ライブ 2	0.42	0.03	0.01
性能 1	0.15	0.02	-0.03
ショート 2	0.03	0.98	0.00
ショート 1	0.04	0.90	0.02
趣味に没頭 2	-0.10	-0.01	0.96
趣味に没頭 1	-0.03	0.03	0.85

表 10 因子分析の結果 6

要素名	因子 16	因子 17	因子 18
眼鏡 2	1.04	-0.02	-0.02
眼鏡 1	0.83	-0.02	-0.04
金髪 2	-0.01	1.04	-0.02
金髪 1	-0.03	0.82	-0.01
外見的幼さ 2	-0.02	-0.01	0.94
外見的幼さ 1	-0.04	-0.02	0.92

表 11 因子分析の結果 7

要素名	因子 19	因子 20	因子 21
グラマラス 2	0.99	0.02	-0.02
グラマラス 1	0.86	0.01	-0.04
おっとり 2	0.01	1.08	0.02
おっとり 1	0.02	0.72	0.00
オッドアイ 2	-0.04	0.01	0.97
オッドアイ 1	-0.03	0.02	0.85

6.1.3 内面

誘因魅力は「ポジティブ」・「好きな事に没頭できる」・「明るい」・「自信がない」・「お

っとり」・「ネガティブ」・「恥ずかしがり屋」・「弱気」・「ツンデレ」が有意に多い。継続魅力は「根性がある」・「自分の弱さと向き合える」・「面倒見がいい」・「お姉さんキャラ」・「強気」・「いじられ役」・「ツッコミ役」が有意に多い。誘因魅力として、庇護欲を掻き立てる要素か元気で明るい要素が重要であると推測できる。また、キャラクターをより魅力的に感じさせるためには他のキャラクターとの関係を強調したり増やしたりする必要があると考えられる。

表 12 因子分析の結果 8

要素名	因子 22	因子 23	因子 24
自由奔放 2	0.87	0.02	0.04
自由奔放 1	0.87	0.03	0.02
押しが強い 1	0.12	-0.01	0.02
青目 2	0.06	0.95	-0.06
青目 1	0.02	0.91	-0.03
赤目 2	0.04	-0.05	0.93
赤目 1	0.05	-0.03	0.90

6.2 因子分析の考察

因子分析の結果、外見・内面に関しては誘因魅力と継続魅力で同じ要素を魅力的だと感じているとわかる。一度魅力と感じた要素は魅力を感じる事がなく、誘因魅力がそのまま継続魅力となると考えられる。

第 2 因子から、あとからキャラクターに追加できる要素、キャラクターを知ることによって気づく要素が継続魅力として有力であると考えられる。それに対して誘因魅力は一度キャラクターの要素として決めてしまうと変えられない要素があげられている。継続的にキャラクターに魅力を感じさせるためには最初のキャラクター設定が重要であると推測できる。

また庇護欲を掻き立てるキャラクター、

芯の強いキャラクター，ムードメーカーなキャラクターが魅力的だと感じるユーザーが多く，モバイルゲームにおいて誘因魅力・継続魅力共に一貫したキャラクター性が大事であるといえる。

7 結論

7.1 結論

本研究では、『アイドルマスターシンデレラガールズ』のファンにアンケート調査を行い，その結果を用いてカイ二乗検定と因子分析によってキャラクターの誘因魅力と継続魅力の差と魅力的に感じる要素の組み合わせを分析した。その結果，キャラクターを印象付けることができる要素が誘因魅力，後から追加することが可能な要素が継続魅力であるということが分かった。

モバイルゲームは運営されていく中で様々な要素が追加され発展していく特徴がある。そのため，誘因魅力は変えることができない印象的な要素，継続魅力はあとから追加できる要素が魅力的であることは今後のモバイルゲームの発展に貢献できるのではないかと考えられる。

7.2 今後の課題

本研究では『アイドルマスターシンデレラガールズ』の一作品のみの分析を行っているため，他のモバイルゲームで同じ傾向が出るか定かではない。キャラクターが少ない作品や，アイドル以外のキャラクターが出る作品，もしくは男性アイドルが出てくる作品等と比較し，その他モバイルゲームと今回の結果にどのような差異があるのか調べる必要がある。

参考文献

- [1] 株式会社三菱総合研究所：「スマホゲームに関するアンケート結果」
https://www.caa.go.jp/policies/policy/consumer_policy/caution/internet/pdf/adjustments_index_1_160915_0004.pdf（閲覧日：2020年10月28日）。
- [2] 三浦直是，山本広志：「漫画の登場人物の顔の分析」，山形大学紀要（教育科学），vol.15, No.2, pp.213-223, 2010.
- [3] 河谷大和，柏崎礼生，高井昌彰，高井那美：「アニメキャラクターの特徴抽出に基づくアニメ度の評価」，情報処理学会研究報告グラフィクスとCDA（CG），vol.80, pp35-38, 2018.
- [4] 牟田淳：「キャラクターから感じる印象の研究」，東京工芸大学芸術学部紀要，vol.21, pp.27-40, 2015.
- [5] 田中翔一，山本博史：「フィクション上のキャラクターに対する性格推定」，情報処理学会関西支部大会論文集，2016.
- [6] 平田郁織：「少年漫画におけるキャラクターの魅力的な要素の分析」公立ほこだて未来大学卒業論文，2018.
- [7] 『アイドルマスターシンデレラガールズスターライトステージ』
<https://cinderella.idolmaster.jp/sl-stage/>（閲覧日：2020年10月30日）。
- [8] 『アイドルマスターシンデレラガールズ』
<https://cinderella.idolmaster.jp/>（閲覧日：2020年10月30日）。

第29回年次大会予稿

共起性に着目した中小企業プロジェクトの失敗リスク要因の 関係性分析

Relationship Analysis of Failure Risk Factors in Small and Medium-sized Company Projects Focusing on Co-occurrence

保田洋^{1,2*}, 川向肇², 西村治彦²

Hiroshi YASUDA^{1,2*}, Hajime KAWAMUKAI², Haruhiko NISHIMURA²

1 甲子園短期大学

Koshien Junior College

〒663-8107 兵庫県西宮市瓦林町4番25号

E-mail: h-yasuda@koshien.ac.jp

2 兵庫県立大学大学院応用情報科学研究科

Graduate School of Applied Informatics, University of Hyogo

〒650-0047 兵庫県神戸市中央区港島南町7-1-28

*連絡先著者 Corresponding Author

プロジェクトマネジメントに触れる機会が少ない中小企業で、知識や経験が乏しい人材が体系化された方法でプロジェクトを管理するのは困難である。そこで著者らはこれまで、中小企業に対して、失敗したプロジェクトの事例について内容等の収集を実施し、失敗回避に向けた分析を行ってきた。本研究では、これまでの研究から得られた失敗リスク要因に対して共起ネットワーク等を適用し、リスク対応を検討するうえでリスク要因の関係性の分析を行った。

キーワード: プロジェクト管理, QCD, リスク要因, テキスト分析, 共起ネットワーク

Project Management, QCD, Risk Factors, Text Analysis, Co-occurrence Network

1 はじめに

昨今、複雑な業務を多数の関係者が関与し分担しながら仕事を進めることが一般的な業務形態となってきており、業務自体がプロジェクト化してきているといえる。プロジェクトが複雑になるにつれて、さまざまな職種の人々が係わること

になるため、受け持った工程の進捗度合だけではなく、前後工程の進捗度合との調整も必要となる。品質(Q:Quality)やコスト(C:Cost)を意識しながら最終納期(D:Delivery)に間に合うようにプロジェクト全体をマネジメントすることが重要となり[1][2]、大企業においては知識体系化されたProject Management Body of

Knowledge (PMBOK) [3][4]等によるプロジェクトマネジメント法が普及しつつある。

しかしながら、限られた社員数の下、プロジェクトマネジメントに備える機会が少ない中小企業で、知識や経験が乏しい人材が体系化された方法でプロジェクトを管理するのは困難である[5][6]。失敗回避ためには、まず第一に、中小企業が失敗したプロジェクトを分析し、失敗要因を体系的な情報として蓄積し、失敗回避に活用していくことが必要となる。

そこで著者らは、まず、プロジェクトを遂行していくうえで抱えている問題を把握するために、中小企業に対して、失敗したと判断されたプロジェクトの事例について内容等の収集を実施し、失敗要因の分析とカテゴリー化を行った[7]。次に、失敗回避に向け、洗い出したカテゴリー対しリスク分析を行い、失敗リスク要因の対応順位付けを行った[8][9]。

本研究では、リスク要因の対応を検討するうえで、これまでの研究から得られた失敗リスク要因に対して共起ネットワーク等を適用し、リスク要因の関係性の分析を行った。

2 これまでの取り組み

2.1 中小企業プロジェクトの失敗要因の分析とカテゴリー化

プロジェクトを遂行していくうえで抱えている問題を把握するために、業務改善のため支援活動を行ってきた中小企業に対して、失敗プロジェクトの事例について内容等の収集を実施し、47社の中小企業から収集した54件の事例から失敗要

因の分析を行った[7]。

まず、失敗したプロジェクト事例の記述より失敗につながるキーワードの抽出を行った。その結果、63個のキーワードの抽出ができ、さらに、キーワード間の類似性に基づきカテゴリー化することで、18項目のサブカテゴリー、7項目のカテゴリーに分類できた(表1)。

次に、得られた7項目のカテゴリーをPMBOKに示されている10の知識エリアに対して対応付けを行った。その結果、「人的資源」、「コミュニケーション」に関する認識は高い反面、「統合」、「スコープ」、「タイム」、「コスト」に関する認識は比較的低く、さらに、「リスク」、「調達」、「ステークホルダー」の知識エリアに対応するカテゴリーは存在しないことが判明した。このことは、中小企業でのプロジェクト現場においては、大局的な視点に立てない状況下でプロジェクトが遂行されがちであることが示唆された。

表1 カテゴリー・サブカテゴリーの分類結果

カテゴリー	サブカテゴリー
スキル・知識不足	技術不足
	知識・理解不足
	作業不備
チームワーク不備	コミュニケーション不足
	情報共有不足
	無責任
やる気の低さ	やる気不足
	意識の低さ
マネジメント不備	調整能力不足
	マネジメント能力不足
	見積もりミス
設計不足	設計能力不足
	設計不備
組織体制不備	調査不足
	体制不備
	戦略ミス
要員不足	人員不足
	担当者不足

2.2 カテゴリーに基づく失敗回避に向けたリスク分析

プロジェクトの失敗回避に向け、2.1節で述べたサブカテゴリーをリスク発生要因とし、起り易さ、影響の受け易さを求めリスクマネジメントで用いられているリスクアセスメント[10-12]に基づいてリスク分析を行った[8][9]。まず、個人的な側面の強いリスク要因と組織的な側面が強いリスク要因に分類したうえでマトリクスを構成した。次に、構成したマトリクスに失敗リスク要因の対応順位付けを行った。その結果、プロジェクトの失敗要因が個人のスキル不足によるリスクが高いことが読み取れた。しかしながら、スキル不足は組織的な側面から考えると、人材育成に失敗したこととなり、「組織体制不備」、「要員不足」にも結び付くことが考えられる。そのため、リスク対応を検討するうえでリスク要因の関係性を考慮する必要があることが浮かび上がった。

3 研究方法

リスク対応を検討するうえでリスク要因の関係性を考慮するため、これまでの研究から得られた失敗リスク要因に対して共起ネットワーク等を適用し、以下のように分析を行う。

①. 71件の失敗したプロジェクト事例の失敗内容と失敗理由の記述から文脈を考慮し抽出された65個の失敗に繋がるキーワードに対して、意味上の類似性を基にリスク要因のラベル化を行う。

②. 失敗したプロジェクト事例毎に含まれるリスク要因のセット（アイテム集

合）を1レコードとする。

③. 全71件のレコードにクラスター分析と共起ネットワークを適用し、リスク要因の関係性を分析する。分析にはKH Coder (ver. 3. Beta. 02) [13][14]を用いた。

4 研究結果と考察

65個の失敗に繋がるキーワードに対して、意味上の類似性を基にラベル化することで30個のリスク要因に集約することができた。次に、プロジェクト事例毎に発生したリスク要因のセットをレコード化し、全プロジェクト事例の71件のレコードに対して、共起関係の強さの尺度としてJaccard係数を適用し、クラスター分析と共起ネットワークの分析を行った。Jaccard係数とは、レコード内でのアイテム間の積集合を和集合で除したものである。図1は、クラスター化の方法にWard法を適用し、クラスター分析を行った結果であり、クラスター数として11が析出された。図2は、比較的強く互いに結びついているアイテム同志をグループ化するmodularity[15]のサブグラフ検出を適用し、Jaccard係数が0.2以上の共起ネットワークを示したものである（円の大きさは各リスク要因の発生件数を反映）。また、それぞれの図において、各リスク要因が属するカテゴリーも記載している。図1、図2より以下のことが判明した。

①. 図1のクラスター分析によるリスク要因の分類傾向は、2.1節で述べた各カテゴリー内に留まるものではない。

②. 図2より2.1節で述べたカテゴリー内で共起が強い(Jaccard係数0.2以上)リスク要因は、「情報共有不足」と「無責

任」の共起のみであり、他の強い共起の関係は、異なったカテゴリー間で発生している。

①, ②より、リスク対応を検討するうえではカテゴリー、サブカテゴリー単位ではなく、リスク要因の共起が強い関係性のパターンを考慮に入れることが肝要であると確認できた。そのため、単独の失敗リスク要因としては対応順位付けが低いリスクであっても、共起の関係上、リスクの高い要因とともにセットで対応する必要があると考えられる。

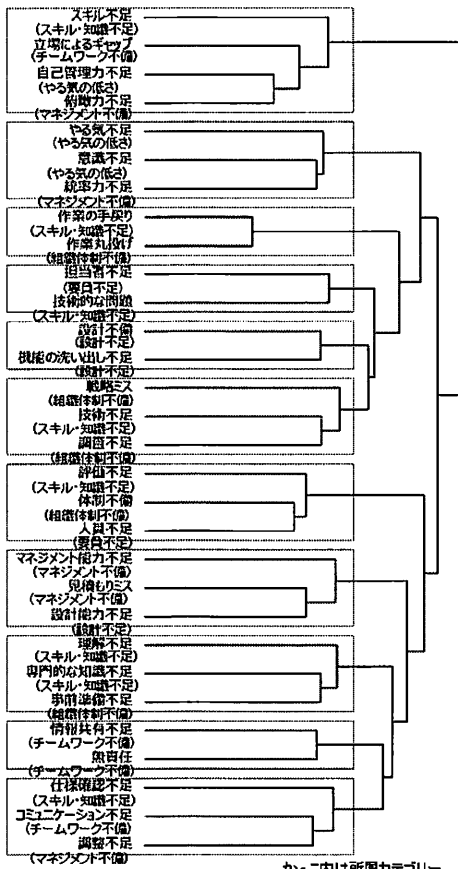


図1 クラスタ分析の結果

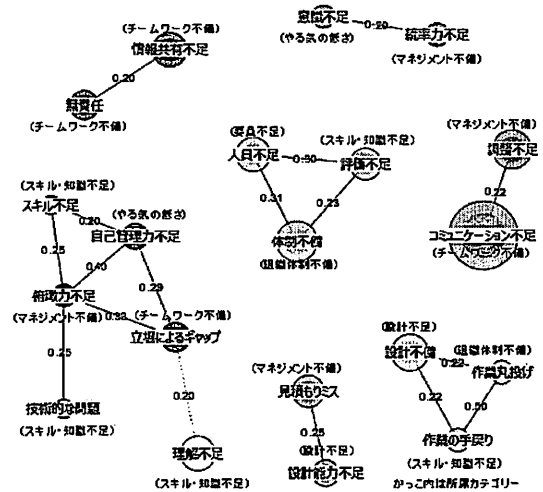


図2 共起ネットワーク分析の結果

5 おわりに

本研究では、中小企業におけるプロジェクトの失敗リスク要因の関係性が明確となった。

今後は、各リスク要因の関係性を基にリスク対応を行う必要がある。具体的には、リスク要因の関連性も考慮したうえで対応順位を設定し、リスク対応策も明記した作業チェックシートを開発を行う。次に、作業チェックシートを支援先の中小企業のプロジェクト内で試し、プロジェクトの達成度を検証するとともに、作業チェックシートの改善を図っていききたい。

参考文献

[1]佐藤知一:「プログラム&プロジェクト・マネジメント理論の全体概念(〈特集〉プロジェクト・マネジメントにおけるシステム・情報技術)」, システム制御情報学会誌, 58(6), pp. 220-226, 2014.

- [2] 広兼修: 「プロジェクトマネジメント標準 PMBOK 入門: PMBOK 第6版対応版」, オーム社, 2018.
- [3] PMI: “A Guide to the Project Management Body of Knowledge”, PMI, 1996.
- [4] ISO: “ISO21500:2012 Guidance on Project Management”, ISO, 2012.
- [5] (社)日本経済団体連合会: 「中小企業を支える人材の確保・定着・育成に関する報告書」, 2010.
<http://www.keidanren.or.jp/japanese/policy/2010/064/> (2021年4月4日参照)
- [6] 三宅由美子; 上村隆幸; 内平直志: 「中小企業の人材育成を考慮したPBL活動とその効果: 訪問看護ステーション支援プログラムの強化」, 国際P2M学会誌, Vol.12, No.1, pp.36-54, 2017.
- [7] 保田洋; 川向肇; 西村治彦: 「中小企業プロジェクトの円滑なマネジメントに向けた失敗要因の分析」, 情報知識学会誌, 30(3), pp.299-311, 2020.
- [8] 保田洋; 川向肇; 西村治彦: 「中小企業プロジェクトの失敗回避に向けたリスク分析」, 情報知識学会誌, 30(2), pp.196-199, 2020.
- [9] 保田洋; 川向肇; 西村治彦: 「中小企業プロジェクトの失敗要因の分析とリスク分析」, プロジェクトマネジメント学会誌, 22(4), pp.3-8, 2020.
- [10] ISO: “ISO 31000: 2018 Risk management — Guidelines”, ISO, 2018.
- [11] リスクマネジメント規格活用検討会: 「ISO 31000:2018(JIS Q 31000:2019) リスクマネジメント 解説と適用ガイド (Management System ISO SERIES)」, 日本規格協会, 2019.
- [12] 経済産業省: 「事業リスク評価・管理人材育成事業報告書」, 2004.
https://warp.da.ndl.go.jp/info:ndl.jp/pid/1003665/www.meti.go.jp/report/data/jinzai_ikusei2004_06.html (2021年4月4日参照)
- [13] 樋口耕一: 「KH Coder 3 リファレンス・マニュアル」, 2020.
- [14] 樋口耕一: 「社会調査のための計量テキスト分析 —内容分析の継承と発展を目指して— 第2版」, 日本規格協会, 2019.
- [15] Clauset, A.; M. E. J. Newman; C. Moore: “Finding Community Structure in Very Large Networks”, *Physical Review E*, 70(6): 066111, 2004.

第29回年次大会予稿

患者紹介時の診療情報提供書における記載内容のテキスト分析

Text Analysis of the Contents of Patient Referral Documents

from Primary Doctors

石崎 潤^{1,3*}, 吉岡 正昭², 西村 治彦³

Jun ISHIZAKI^{1,3*}, Masaaki YOSHIOKA², Haruhiko NISHIMURA³

1 西宮協立リハビリテーション病院

Nishinomiya Kyoritsu Rehabilitation Hospital

〒662-0002 兵庫県西宮市鷺林寺南町2番13号

2 大阪府済生会吹田病院

Saiseikai Suita Hospital, Osaka Saiseikai Imperial Gift Foundation

〒564-0013 大阪府吹田市川園町1番2号

3 兵庫県立大学大学院応用情報科学研究科

Graduate School of Applied Informatics, University of Hyogo

〒650-0047 兵庫県神戸市中央区港島南町7丁目1番28

*連絡先著者 Corresponding Author

高齢化に対応するため、厚生労働省は地域包括ケアシステムの下での施設間連携を推進している。その充実には、地域の施設間での診療情報提供書の情報共有と二次利用が望まれるが、紙面での自由記述によるやり取りのため、利用は患者紹介時の一過性のものとどまっている。この現状の克服のために我々は、診療情報提供システムの構築に向けて、実際の記載内容の所定様式 11 への準拠性の評価と内容構成のカテゴリ化を進めてきた。本研究では、準拠性の評価と紹介目的のカテゴリ化の結果に基づき記載内容のテキスト分析と共起ネットワークの構成を行った。その結果、準拠性や紹介目的に応じて、記載内容の傾向の違いが出現単語の種類、語数、共起関係を通して捉えられることが確認された。

キーワード: 診療情報提供書, テキスト分析, 情報共有, 地域連携

patient referral document, text analysis, information sharing, regional cooperation

1 はじめに

わが国では、2025年には65歳以上の人口

が3,657万人（高齢化率30.3%）に達するとされており[1]、今後さらなる増加が見

込まれる。これに対応するため、厚生労働省は地域の特性に合わせた、住まい・医療・介護・予防・生活支援が一体的に提供される地域包括ケアシステムの構築[2]を推進している。これには、かかりつけ医を中心に、急性期病院、地域包括支援センター、介護老人福祉施設などの中での連携が必要であり、その充実には連携機関間での情報共有が重要となってくる。

医療機関の間での現行の主要な情報提供手段は診療情報提供書である。これは、患者を紹介する際に、担当医が他の医療機関へ情報を提供するものであり、厚生労働省が保医発0305第3号[3]にて基本項目（紹介目的、傷病名、既往歴及び家族歴、症状経過及び検査結果、治療経過、現在の処方、備考）とその大枠の所定様式11を定めている。紹介元医療機関の医師は様式11をベースにこれを自由記載し、紙面により紹介先医療機関に提供しているため、利用は一過性のものにとどまっているのが現状である。今後、診療情報提供書の情報共有と二次利用に対応していくためには、提供書のデジタル化とデータ構造化が必要となる。

そのため我々は、診療情報提供システムの構築に向けて、実際の記載内容の所定様式11への準拠性の評価と内容構成のカテゴリ化を進めてきた[4-7]。本研究では、準拠性の評価と紹介目的のカテゴリ化の結果に基づき、テキスト分析と共起ネットワークの構成による記載内容の特徴抽出について検討した。

2 方法

2.1 対象データ

地域医療支援病院のS病院（標榜診療科26診療科、許可病床数500床）が2014年4月

に近隣医療機関から受けた1,623件の診療情報提供書から1,000件を対象とした。S病院では、提供された診療情報提供書は画像ファイルとして保管されている。本研究での分析を行うために、記載内容を全て画像ファイルからExcelに転記（テキスト化）を行った。

2.2 倫理的配慮

本研究はS病院倫理審査委員会の承認（2016.3.14）を受け実施した。分析に際し、医師および患者に対して連結不可能匿名化し、さらにテキスト化された記載内容中の個人を特定できる箇所についても匿名化を行った。

2.3 様式11への準拠性に基づく記載内容のテキスト分析

2.3.1 様式11への準拠性の評価

様式11への準拠性の評価にあたり、基本項目の「紹介目的」、「傷病名」、「既往歴及び家族歴」、「症状経過及び検査結果」、「治療経過」の5項目に対して、各項目に沿った内容が書中に明確に記載されているか（以下、内容充足とする）、所定欄に記載されているか（以下、所定欄記載とする）の2点について以下のように評価し、レベル付けを行った[7]。

レベル4：内容充足し、所定欄記載であるもの、レベル3：内容充足し、所定欄記載でないもの、レベル2：内容充足せず、所定欄記載であるもの、レベル1：内容充足せず、所定欄記載でないもの、レベル0：提供書中に当該項目についての記載がないもの。

各項目において、内容充足に値する判断基準としては以下のように設定した。「紹介

目的」は、依頼内容が明確であること。「傷病名」は、紹介目的に沿った傷病名（症状）の記載があること。「既往歴及び家族歴」は、病名と付随する情報、または“特記事項なし”と明記されていること。「症状経過及び検査結果」と「治療経過」は、具体的な情報の記載があること。

2.3.2 準拠性の違いのテキスト分析による比較

診療情報提供書の記載内容に対して、準拠性による違いに着目しテキスト分析を行う。具体的には、前述で付与した基本項目ごとのレベルについて共起ネットワーク分析を行い、それぞれの抽出語の傾向を分析した。

テキスト分析にはKHCoderを用い、MeCabにて形態素解析を行った[8]。記載内容は自然言語で書かれた文章であり、重複や誤記、表記の揺れなどがみられたため、それらの削除や修正によりデータの質を高めた。また、KHCoderが単語を検出する際の標準辞書のみでは、実データの解析は不十分なことが多い。そのため、自動抽出用システム（TermExtract）を用いて専門用語を自動抽出し、そこから研究者が文脈を考慮して必要な単語を強制抽出した。

2.4 紹介目的のカテゴリ化に基づく記載内容のテキスト分析

2.4.1 紹介目的のカテゴリ化

KJ法[9]の手順より診療情報提供書ごとに紹介目的の内容が類似するものを分類し、各グループに対して包括的な名称を付与すると共に、木構造の階層化の導入によりカテゴリ化を行った[4-7]。

2.4.2 紹介目的の違いのテキスト分析による比較

2.3.2で述べたのと同様の要領で形態素解析を行い、紹介目的による記載内容の違いについて分析する。具体的には、前述のカテゴリ化の結果をもとに共起ネットワーク分析を行い、それぞれの抽出語の傾向を分析した。

3 結果

3.1 様式11への準拠性に基づく記載内容のテキスト分析

3.1.1 様式11への準拠性の評価

評価に際し1,000件の内、様式11でなく所定の検診用紙による市民検診など（102件）は除外した。各項目のレベル別件数は表1の通りであり、各項目の内容充足（レベル3または4）の割合は、紹介目的：78.7%、傷病名：99.2%、既往歴及び家族歴：19.3%、症状経過及び検査結果：95.8%、治療経過：31.1%であり、紹介目的、傷病名、症状経過及び検査結果の内容充足の割合が高く、医師が重要視していることが窺える。

そこで、[紹介目的、傷病名、症状経過及び検査結果]のセットに対して、そのレベルパターンの分析を行った。その結果、3項目全てが内容充足しているもの（[333]～[444]）は683件（76.1%）であり、それ以

表1 様式11の基本5項目における各レベルの出現件数

	紹介目的	傷病名	既往歴及び家族歴	症状経過及び検査結果	治療経過
レベル0	8	3	679	2	462
レベル1	137	0	11	0	157
レベル2	46	4	35	36	0
レベル3	231	214	39	3	279
レベル4	476	677	134	857	0

参考文献

- [1] 総務省：「平成 28 年度版情報通信白書 (p.2)」, <https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h28/pdf/28honpen.pdf> (2021 年 4 月 3 日参照)
- [2] 厚生労働省：「地域包括ケアシステム」, https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/hukushi_kaigo/kaigo_koureisha/chiiki-houkatsu/ (2021 年 4 月 3 日参照)
- [3] 厚生労働省：「平成 26 年 3 月 5 日保医発 0305 第 3 号「診療報酬の算定方法の一部改正に伴う実施上の留意事項について(通知) 別紙様式 11」」, <https://www.mhlw.go.jp/file/06-Seisakujouhou-12400000-Hokenkyoku/0000041236.pdf> (2021 年 4 月 3 日参照)
- [4] 石崎潤；宮部剛実；西村治彦：「標準化を目指した診療情報提供書の内容分析とカテゴリ化」, 第 68 回日本病院学会抄録集, p. 187, 2018
- [5] 石崎潤；吉岡正昭；西村治彦：「地域連携システムの標準化を目指した診療情報提供書の分析」, 医療情報学, 39 (Suppl.), pp. 237-242, 2019
- [6] Jun, Ishizaki; Masaaki, Yoshioka; Haruhiko, Nishimura: "Analysis and Categorization for the Standardization of Medical Referral Document", 9th International Congress on Advanced Applied Informatics, and 11th International Conference on E-Service and Knowledge Management (ESKM 2020), pp. 812-813, 2020
- [7] 石崎潤；吉岡正昭；西村治彦：「地域連携での情報共有に向けた診療情報提供書の記載状況把握と内容構成の構造化」, 日本医療・病院管理学会誌 (投稿中)
- [8] 樋口耕一：「社会調査のための計量テキスト分析」, 第 2 版, ナカニシヤ出版, 2020.
- [9] 山浦晴男：「科学的な質的研究のための質的統合法 (KJ 法) と考察法の理論と技術」, 看護研究, 41 (1), pp. 11-32, 2008

第29回年次大会予稿

マンガの内容に基づく探索のためのNDC-LDとのリンク

Linking Manga to NDC-LD for Content-based Search

後藤佑斗^{1*}, 三原鉄也², 永森光晴²

Yuto GOTO^{1*}, Tetsuya MIHARA², Mitsuharu NAGAMORI²

1 筑波大学 情報学群情報メディア創成学類

College of Media Arts, Science and Technology, School of Informatics, University of Tsukuba

〒305-8550 茨城県つくば市春日1-2

2 筑波大学 図書館情報メディア系

Faculty of Library, Information and Media Science, University of Tsukuba

〒305-8550 茨城県つくば市春日1-2

*連絡先著者 Corresponding Author

Web上で提供されるマンガには内容に基づく探索のためにタグやキーワードなどの主題語が付与されているが、意味的な探索のための構造化が不十分で、その活用は限られている。本研究では日本十進分類法(NDC)の持つ主題の意味関係に即した探索のために、Web上でマンガと共にその内容を示すタグおよびマンガのあらすじの2種の情報を利用して、マンガと主題に基づくLinked Data版日本十進分類法(NDC-LD)を機械的に結びつける手法を提案する。タグを利用した手法ではNDC-LDの相関索引語とタグの文字列の類似度に基づいて同一性を判定しリンクを行った。あらすじを利用した手法ではあらすじ文中のNDCの見出し語と相関索引語についてTF-IDFを用い、特徴語として抽出されたものをリンク対象とした。更に、マンガ146作品に対してこれらの提案手法を適用し、人手による分類及び機械学習によるNDCの推定手法と比較した結果、提案手法が人手での分類の再現に有効であった。(393文字)

Manga are often provided on Web with tags or the keywords expressing subjects for the search based on contents. However, they are less structured and not useful for semantic search. This paper proposes a method for automatic linking manga and NDC-LD entities based on the subject by using two kinds of information, their subject tags and outline on the Web explaining manga and its content. The method using subject tags is to link by the similarity of tag words and correlative catchword of NDC-LD. On the other hand, The method using the outline is to link entities extracted as characteristic words of identified heading words of NDC in the outline. This paper also applied the proposed method to for 146 manga works, and compares the result with the classification by hands and the existing automatic estimation method of NDC by the machine learning. This comparison shows that the proposed method is more efficient for the reproduction of the manual classification. (160words)

キーワード: マンガ, 漫画, コミック, Linked Open Data, エンティティリンキング

Manga, Comic, Linked Open Data, Entity Linking

1 はじめに

現在,膨大な数のマンガがWeb上で出版されている。Web上でマンガを販売するサイトでは,検索のためにマンガの内容を表す単語(主題語)を「タグ」や「キーワード」として提供している。主題語には,互いに上位・下位概念の関係にあるものや類似した意味を持つものがあり,主題が複数の主題語に分かれている場合がある。

現在,こうしたマンガの検索のために提供されている主題語に意味概念に基づく構造化が十分にされているものは見られない。そのため,主題語の上位・下位概念や意味の類似などを利用した知的な探索を実現するためには,主題語同士の意味的な関係を記述した語彙が必要である。

本研究では,マンガの主題の意味関係を記述するための語彙に図書の標準的な分類体系である日本十進分類法(Nippon Decimal Classification, NDC) [1]を用いることを目的として,マンガに付与されている内容を表す記述としてタグ及びあらすじの2種の情報源とNDCを機械的にマンガと結びつける手法を提案する。

2 NDCとマンガのリンキング

NDCとマンガを結びつける試みとしては,日本財団・マンガナイトが主催するこれも学習マンガだ! [2]がある。このプロジェクトは教育や社会啓発に役立つマンガを選書し公開するものであるが,選出された200作品をNDCで分類するワークショップが行われた [3]。マンガには一般に「726.1 漫画」の分類が付与されるが,ここではNDC

が一般的な書籍に対して用いられるように対象の主題を示す分類を検討している。

「基本的には第3次区分まで,その場で第4次区分以降までの分類が可能なものは可能な限り加える」ことを基本方針として図書館司書8名を2グループに分け,100作品ずつ作品に適した分類記号の検討を行っている。このワークショップでは作業者によって付与の方針に差があったことや,議論しながらマンガに対して人手で分類を付与するには時間を要することが明らかになっている。

多数出版されるマンガについて,こうした人手によってNDCを付与していくことは困難である。そこで本研究では,こうした付与の効率化のためにマンガとNDCを機械的に結びつける手法を提案する。NDCのデータはLinked Data版日本十進分類法(NDC-LD)を利用する。NDC-LDは,ウェブ上でのNDCの利活用を目的として日本図書館協会(JLA)と国立国会図書館(NDL)の共同研究[4]により開発された,他のデータとリンクできる形で公開された機械可読なデータである。

3 NDC-LDとマンガのリンキング手法

法

3.1 タグを利用したリンキング手法

本手法では,電子書籍ストアでマンガに付与された「タグ」や「キーワード」といった主題語とNDCの相関索引語のマッチングによりマンガとNDCを結びつける。マンガに付与されている主題語とNDCの分類項

目の相関索引語のLevenshtein距離を用いて文字列の類似度を算出する。本手法では類似度が0.6より大きい場合にリンクする。例えば、「783.3 ハンドボール. ドッジボール」に対して「ハンドボール」や「ドッジボール」といった主題語が結びつく。

3.2 あらすじを利用したリンク手法

3.2.1 マンガペディアのあらすじの利用

本研究でリンクに利用するあらすじはマンガペディア[5]に掲載されているものを用いる。マンガペディアはマンガとその内容に関する百科事典であり、Web上で提供されている。マンガペディアの編纂は百科事典と同様の体制で行われており、編集者が項目を作成する際はすべてのマンガに目を通したうえで執筆しているため、記載内容に対する信用度が高い。また、マンガペディアは編纂の過程において内容や語彙がある程度統制されているため、そうした統制が見られない一般的なオンライン書店などのあらすじに比べ、主題語とのリンクに適している。

3.2.2 分類候補の抽出

取得したあらすじには人物名などNDCとのリンクには不要な固有名詞が含まれているため、まずこれらを除去する。

次に、リンクに必要な単語の候補として、あらすじ中からNDCの見出し語と相関索引語およびWeb NDL Authoritiesを用いる。マンガとNDCを結び付けやすくするにはNDCに関連付けられる単語がより多いほうが望ましい。そこで、NDCの見出し語や相関索引語に加えてWeb NDL

Authorities[6]を利用してNDCの関連単語を拡張する。Web NDL Authoritiesは典拠データをキーワード検索やNDCの分類記号検

索などにより検索することができ、各典拠データには対応するNDCの分類が定義されている。本研究ではWeb NDL Authoritiesの分類記号検索を行い、得られた普通件名の標目とその同義語、上位語、下位語、関連語を取得する。

NDCの全分類記号についてこの作業を行い得られた単語についてマンガペディアの全作品のあらすじを対象にして取得した単語と一致するものを取得し、NDCと結びつく候補単語の集合とする。

3.2.3 NDC分類とのリンク

マンガの内容に即したNDCの分類を付与するには、取得した単語の中からマンガの内容を表すものを選択する必要がある。そこで取得した候補単語の集合に対してTF-IDFを適用し、集合の中で相対的に特徴的な単語を取得する。各作品の候補単語集合の中でTF-IDFの結果が最も大きい単語が最もそのマンガの特徴を表す単語と見なし、その単語が関連付けられているNDC分類をリンクする。

しかし単純にTF-IDFの値の最も高いものを分類に採用する場合には、内容と全く関係のない単語や文章の一部などが誤って抽出された単語が、偶然特定の作品のあらすじ内に多数出現して最上位であった場合などに内容に即したものにならない。一方、仮に最上位でなくても、TF-IDFの結果が上位となった単語の多くに共通する概念があれば、その概念はマンガの内容を示すものであると推測できる。そこで、NDCの分類記号を用いて上位10位以内となった単語集合の共通の上位概念を見つける。上位10単語と結びついている分類記号の中から上位2桁および上位3桁が多くの単語に関連しているものを取得し、得られた

分類記号と関連づけられている単語の優先度を上げる。例えば、2位が726.1, 4位が726, 7位が720, 8位が726.1と結びついている場合、このマンガの内容は分類記号の上位2桁が72である概念と近いとみなしこれらの単語の優先度を上げる。

4 評価

4.1 人手により付与された分類との比較

本研究では精度や有効性を検討するため、二つの提案手法による分類付与結果と人手による分類付与結果との比較を行う。これも学習マンガだ！に選出された200作品のうちタグを取得できた152作品とあらすじを取得できた146作品の中で、それぞれ比較対象として適当であると判断したものに対して人手での分類付与結果との比較を行った。あらすじを利用した手法についてはマンガペディアの延べ13,179作品に対してあらすじの取得を試みて13,058作品からあらすじを取得することができた。その文章中からリンキングに必要な単語の候補を取得し、全単語に対してTF-IDFを適用した。結果を表1に示す。

表1 提案手法と人手で付与した分類の比較結果

手作業による分類と一致する区分	タグ		あらすじ	
	該当作品数	割合(%)	該当作品数	割合(%)
完全一致・より詳細	20	13.2	36	24.7
第四次区分	0	0	2	1.4
第三次区分	10	6.6	4	2.7
第二次区分	16	10.5	20	13.7
第一次区分	24	15.8	25	17.1
不一致	26	17.1	59	40.4
付与不可	56	36.8	0	0
合計作品数	152	100	146	100

付与された分類のうち、完全一致およびより詳細な分類が付与できた割合はタグ利用の場合が13.2%、あらすじ利用の場合

が24.7%であり、あらすじを利用した場合のほうが高い。また、付与した分類が人手で付与した分類と一致しなかった作品と分類が付与できなかった作品の合計はタグ利用の場合が53.9%(不一致17.1%と付与できなかったもの36.8%の合計)、あらすじ利用の場合が40.4%である。

4.2 NDC Predictorとの比較

さらなる評価のため、既存のNDC推測サービスであるNDC Predictor[7]とのNDCの分類付与結果の比較を行う。NDC Predictorは機械学習によるNDCの推測アプリであり、テキストエリアに貼り付けられた書誌情報からNDCの分類を第三次区分まで推測する。学習にはタイトル、出版社、著者の情報といった書誌情報を用いている。結果は確信度の高いものから順に第一候補から第三候補までを表示する。したがって、提案手法との比較のために固有名詞を除去したあらすじを用いてNDCを推測し、その結果をあらすじを用いた提案手法による結果の上位3つと比較した。タグを用いた提案手法に関しては分類の候補が一つであるため、第一候補との比較を行った。結果を表2に示す。

表2 提案手法とNDC Predictorによる分類付与結果の比較

	タグ 作品数 (完全一致作品数)	あらすじ 作品数 (完全一致作品数)	NDC Predictor 作品数 (完全一致作品数)
第一候補が一致	30(20)	42(36)	26(10)
第二候補が一致	-	10(9)	14(7)
第三候補が一致	-	5(5)	9(6)
不一致	122	89	97

あらすじを利用した提案手法によって付与された分類の第一候補の第三次区分までが人手で付与された分類と一致した作

品数は42作品, そのうち第四次区分以降も含めて分類が完全に一致した作品数は36作品であった. NDC Predictorによってあらすじから推測された分類の第一候補と人手で付与された分類の第三次区分までが一致した作品数は26作品, そのうち完全に一致した作品は10作品であった. また, 提案手法の第三候補までに人手で付与された分類と一致するものがなかった作品は89作品, NDC Predictorの第三候補までに人手で付与された分類と一致するものがなかった作品は97作品であった.

5 考察

タグを利用した手法とあらすじを利用した手法によって付与された分類の付与の結果を比較すると, 人手での分類と完全一致およびより詳細な分類が付与できた割合はタグが13. 2%, あらすじが24. 7%であり, 一致しなかった作品と分類が付与できなかった作品の合計はタグが53. 9%, あらすじが40. 4%であった. 特にあらすじを利用した手法では対象とした全作品に何らかの分類を付与することに成功しており, あらすじを利用したことによる情報量の増加や, 3. 2. 1で述べたような情報の統制が有効であったと考えられる. マンガペディアのあらすじの内容や語彙が統制されていることにより, 平等な条件の下で十分な量の単語を得ることができた. そのため, マンガに付与されているタグの不足やNDCの相関索引と結びつかないタグが存在していたことなど, マンガとNDCが結びつかない場合が発生するタグを利用した手法の問題点を改善できたと考えられる.

また, 提案手法によるNDCの分類の付与結果とNDC Predictorを用いたあらすじから

の分類推測の結果を比較したところ, あらすじを用いた提案手法のほうが人手で付与した分類に近い分類ができたという結果が得られた. 特に, NDC Predictorの結果の表示は第三次区分までであるため, 人手で第四次区分以降も分類が付与されている作品に関しては完全に一致する分類を付与することができない. 一方, 提案手法では第四次区分以降も付与することが可能であるため, NDC Predictorと比較してより詳細で内容に即した分類を付与することが可能となっている.

6 関連研究

NDCを利用して図書以外の実体に分類を付与する研究に, 安形ら[8]による研究がある. 安形らは分類体系としてNDCとYahoo!のカテゴリを用い, ウェブページを自動で分類する手法を提案した. この研究では単語の抽出方法として形態素解析を利用しているが, マンガのあらすじにはマンガ固有の単語が含まれており, 一般的な辞書では正しく形態素解析ができない可能性があるため本研究には適していない.

また, ウェブ上の文書に記述されている人物に対してNDCの分類記号を付与する村上ら[9]の研究がある. この研究では人物に関するウェブ上の文書からNDCの相関索引語を抽出する手法を提案し, 従来の特徴語抽出によって得られたNDC相関索引語との比較を行なっている. 文書からNDCに関する単語を抽出している点は本研究と共通しているが, 本研究ではマンガのあらすじに対して適用している点や, 抽出した単語集合に共通の上位概念を利用している点などが異なる.

7 おわりに

本論文では、主題の意味に基づいたマンガの探索のためにタグおよびあらすじを利用してマンガとNDCを結びつける2種の手法を提案した。更に、実際のマンガ146作品についてのNDC分類の付与の比較から、提案手法の内あらすじを利用した提案手法は既存手法を用いたあらすじからのNDC分類推測よりも有効であることを示した。

本研究を通していくつか課題が残った。1つ目は、あらすじから抽出する単語に文章や単語の一部など不要な単語が混じってしまう点である。解決策として、形態素解析などを利用して不要な単語の混入の防止を検討することが考えられる。2つ目は、あらすじから取得した単語が複数のNDC分類記号と関連付けられる際、最適な分類を機械的に選択することである。本研究では目視により選択したが、今後はTF-IDFの結果上位となった単語同士の関連性や共通の上位概念などを利用して機械的に選択を行える手法を検討する必要がある。3つ目は、あらすじから取得した単語の集合に共通した概念の利用である。本研究では分類記号の上位区分を利用して単語集合に共通する概念の取得を行ったが、さらなる精度向上のためには分類記号のほかに単語自体の上位概念の利用が望まれる。これにより抽象的な主題を持つマンガへの分類付与の精度が向上する可能性がある。

謝辞

本研究はJSPS科研費20K20129の助成を受けたものです。

参考文献

- [1] もりきよし原編. 『日本十進分類法』新訂10版, 日本図書館協会, 2015.
- [2] 日本財団. “これも学習マンガだ!”, <http://gakushumanga.jp>, (参照 2021-04-12).
- [3] 日本財団. “これも学習マンガだ! 図書館分類ワークショップレポート【後編】200選をNDC分類で分けてみた”, <http://gakushumanga.jp/news/これも学習マンガだ! 図書館分類ワークショップ-2/>, (参照 2021-04-12).
- [4] 中井万知子; 藤倉恵一; 橋詰秋子; 福山樹里; 神崎正英: 「日本十進分類法のLinked Data化: セマンティックWebへの対応を目指して」, 情報管理, Vol. 59, No. 4 p. 209-217, 2016.
- [5] 百科総合リサーチ・センター. “マンガペディア”. <https://mangapedia.com/>, (参照 2021-04-12).
- [6] 国立国会図書館. “Web NDL Authorities”. <https://id.ndl.go.jp/auth/ndla/>, (参照 2021-04-12).
- [7] NDL Lab. “NDC Predictor”. <https://lab.ndl.go.jp/ndc/>, (参照 2021-04-12).
- [8] 安形輝; 石田栄美; 久野高志; 野末道子; 上田修一: 「NDCの分類体系とYahooのカテゴリを使った分類」, 情報処理学会研究報告, Vol. 99, No. 39, pp. 113-120, 1999.
- [9] 村上晴美; 浦芳伸; 片岡祐輔: 「Web上の人物への図書館の分類記号の付与と人物ディレクトリの開発」, システム制御情報学会論文誌, Vol. 29, No2, pp. 51-64, 2016.

第29回年次大会予稿

新型コロナウイルス感染症による日本の研究活動への影響

The Impact of COVID-19 Infections on Research Activities in Japan

西川 開^{1*}
Kai NISHIKAWA^{1*}

1 科学技術・学術政策研究所 (NISTEP)

National Institute of Science and Technology Policy

〒100-0013 東京都千代田区霞が関3-2-2 中央合同庁舎第7号館東館16階

E-mail: k.nishikawa@nistep.go.jp

*連絡先著者 Corresponding Author

本研究では、新型コロナウイルス感染症が日本の研究活動に与えた影響を詳細に把握するために、1,717名の研究者を対象とするアンケート調査により2020年1月頃から同年9月にかけての研究活動の状況を尋ね、回答者の所属地域と研究分野による違いに着目した分析を行った。分析の結果、コロナ禍による研究活動への影響の程度や研究活動の進捗状況、コロナ禍への対応策であるデジタルツールの活用状況は地域によって大きく異なることが示された。また、研究分野によって、研究活動を行っていく上での現状の懸念等は異なる傾向にあることが明らかとなった。

In this study, a questionnaire survey of 1,717 researchers was conducted to understand the impact of COVID-19 infections on research activities in Japan from around January 2020 to September 2020. The results show that the degree of impact of COVID-19 infections on research activities, the progress of research activities, and the use of digital tools varied greatly by the respondents' regions. In addition, it became clear that the current concerns about research activities tended to differ by the respondents' research fields.

キーワード: 新型コロナウイルス感染症, COVID-19, 科学技術政策, デジタルトランスフォーメーション (DX)

Novel coronavirus infections, COVID-19, Science and technology policy, Digital transformation

1 はじめに

2020年より日本国内でも感染拡大が生

じた新型コロナウイルス感染症は、社会的活動の様々な側面に重大な影響を及ぼし

ている。こうした中、同感染症が日本の研究活動等の諸局面に与えた影響や、今後求められる支援・対応策を探る調査が実施されている[1][2]。

文部科学省が2020年5月に科学官・学術調査官等を対象に実施したアンケート調査[1]では、コロナ禍による研究活動への即時的影響や、研究活動の停滞による中長期的な課題等を自由記述形式で尋ねている。

重茂・蒲生が2020年6月に約 2,000 人の科学技術の専門家で構成されるネットワークを対象に実施したアンケート調査[2]では、コロナ禍による日本の科学技術全体への影響や、研究開発現場への影響、今後の科学技術政策のあり方を多肢選択式および自由記述式で尋ね、その結果を東日本大震災後のアンケート調査結果と比較している。また、同調査では今後の課題として、回答者の所属地域や研究分野による影響の差を考慮し、調査項目を細分化した詳細な分析を行う必要があることが挙げられている。

本研究は、コロナ禍による日本の研究活動への影響を横断的に明らかにし、コロナ禍を踏まえた科学技術政策のあり方を検討する際の示唆を得ることを目的とするという点で、上記の先行研究[1][2]と軌を一にする。他方で、本研究では回答者の所属地域や研究分野の違いに着目するとともに、調査項目を細分化することでコロナ禍が具体的にどの様な研究活動の局面に影響を及ぼしているかを尋ねているという点で、重茂・蒲生[2]に対して相補的な関係にあるといえる。

なお、本研究は、科学技術・学術政策研究所 (NISTEP) による「科学技術の状況に

係る総合的意識調査 (NISTEP 定点調査 2020) 」[3]の調査結果の一部を基に追加的に分析を行ったものである。

2 方法

2.1 調査概要

日本の大学・大学共同利用機関・公的研究機関に所属する自然科学系 (理学, 工学, 農学, 保健 (医学を含む)) の研究者1,717 名を対象にアンケート調査を実施した。調査期間は2020年9月11日から同年12月25日である。

2.2 調査項目

2020年1月頃から同年9月にかけての、新型コロナウイルス感染症による研究活動への影響等を調査した。調査項目を大別すると下記の通りである。

なお、調査項目4および5については、それぞれ「個人レベル」「部局・機関レベル」「国レベル」の3点について回答を求めた。

- 1 研究活動への影響 (多肢選択式)
- 2 研究活動の進捗状況 (多肢選択式)
- 3 研究活動を行う上でのデジタルツール等の活用状況 (多肢選択式)
- 4 現状の懸念等 (自由記述式)
- 5 今後求められる変化・対応等 (自由記述式)

2.3 回答率と回答者の属性

調査を依頼した1,717名中、回答者は1,596名であり、回答率は92.6%であった。回答者の所属は、大学が1,305名、大学共同利用機関が33名、公的研究機関が258名であった。このうち、大学の研究者につい

では、所属している大学の部局により専門分野（以下、部局分野）と所属地域を判定し、属性情報として付与した。なお、所属地域については文部科学省による資料[4]を参考として各都道府県を11地域に分類した。

2.4 自由記述回答の分析方法

2.2節で示した調査項目のうち、自由記述式の設問（4・5）における回答についてはKH Coder 3 (<https://khcoder.net/>)を用いた計量テキスト分析を実施した。なお、調査項目4に対応する設問の回答数は1,786件、調査項目5に対応する設問の回答数は1,528件であった。

3 結果

3.1 研究活動への影響

図1は、調査項目1について、新型コロナウイルス感染症が研究活動にどのような影響を与えたかを全回答者に尋ねた結果である。特に、「③研究者や学生の移動や異

動への影響」[5]において「大きなマイナスの影響あり」とする回答割合が大きい。他方で、「⑮学内業務の変化に伴う研究時間への影響」及び「⑯教育業務の変化に伴う研究時間への影響」についてはプラスの影響も一定数認められる。

図2は、調査項目1について、「大きなマイナスの影響あり」又は「ややマイナスの影響あり」を選択した回答を大学の研究者の所属地域別に集計した結果であり、各項目（行）についてマイナスの影響が大きい地域ほど濃い赤色で強調している。特に「⑨ 所属機関内の共用施設・設備の利用への影響」や「⑭ 研究資金の事務手続きへの影響」について地域間でマイナスの影響を受けた程度の差が大きい。また、北海道・東京圏・九州ではマイナスの影響が全体的に大きいのに対して、東北・甲信越・北陸ではマイナスの影響が全体的に小さい傾向にある。

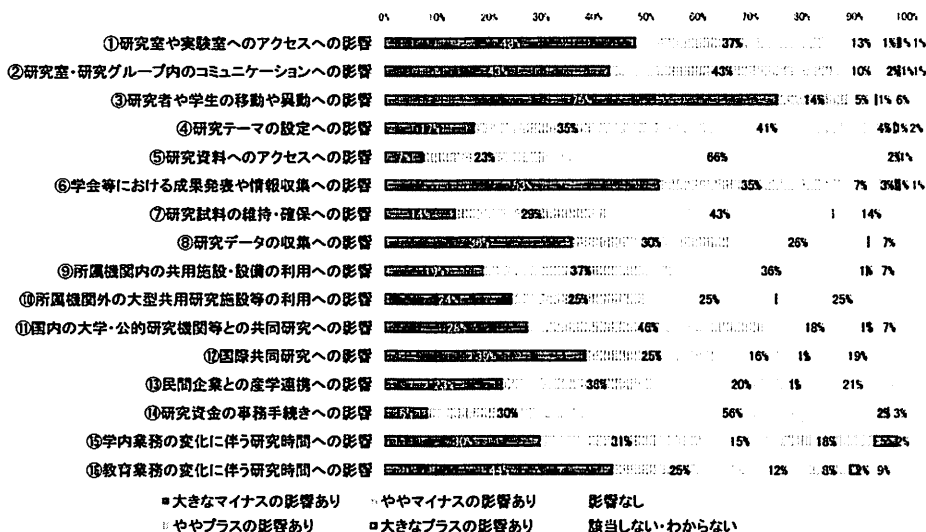


図 1 新型コロナウイルス感染症による研究活動への影響（2020年1月頃～9月の状況）

	北海道(65)	東北(99)	北関東(31)	東京圏(311)	甲信越(47)	東海(99)	北陸(34)	近畿(224)	中国(81)	四国(58)	九州(140)
① 研究室や実験室へのアクセスへの影響	89.2%	70.7%	93.5%	88.4%	83.0%	85.9%	79.4%	86.6%	79.0%	70.7%	84.3%
② 研究室・研究グループ内のコミュニケーションへの影響	90.8%	77.8%	93.5%	87.1%	80.9%	83.8%	85.3%	90.6%	84.0%	86.2%	86.4%
③ 研究者や学生の移動や異動への影響	93.8%	85.9%	83.9%	88.4%	89.4%	86.9%	91.2%	92.0%	91.4%	81.0%	90.0%
④ 研究テーマの設定への影響	55.4%	43.4%	54.8%	58.3%	52.2%	53.5%	47.1%	50.9%	45.7%	55.2%	58.8%
⑤ 研究資料へのアクセスへの影響	30.8%	21.2%	29.0%	37.5%	31.9%	30.3%	35.3%	31.7%	19.8%	32.8%	34.3%
⑥ 学会等における成果発表や情報収集への影響	89.2%	84.8%	93.5%	89.1%	72.3%	81.8%	85.3%	86.2%	91.4%	87.9%	87.9%
⑦ 研究試料の維持・確保への影響	47.7%	34.3%	45.2%	48.8%	38.3%	34.3%	32.4%	40.4%	46.9%	46.6%	49.3%
⑧ 研究データの収集への影響	69.2%	61.6%	64.5%	72.0%	57.4%	60.6%	55.9%	62.5%	59.3%	60.3%	67.1%
⑨ 所属機関内の共用施設・設備の利用への影響	58.5%	40.4%	54.5%	64.0%	46.8%	49.5%	35.3%	54.0%	56.8%	46.6%	61.4%
⑩ 所属機関外の大型共用施設等の利用への影響	49.2%	48.5%	58.1%	53.1%	46.8%	46.5%	35.3%	42.4%	55.6%	34.5%	57.9%
⑪ 国内の大学・公的研究機関等との共同研究への影響	73.8%	72.7%	67.7%	72.7%	66.0%	60.6%	64.7%	70.1%	80.2%	75.9%	79.3%
⑫ 国際共同研究への影響	56.9%	58.6%	48.4%	63.3%	56.6%	55.6%	52.9%	63.8%	66.7%	56.9%	69.3%
⑬ 民間企業との産学連携への影響	56.9%	61.6%	48.4%	58.2%	46.8%	55.6%	50.0%	50.9%	58.0%	60.3%	64.3%
⑭ 研究資金の事務手続きへの影響	43.1%	27.3%	35.5%	48.0%	27.7%	36.4%	20.6%	31.7%	32.1%	31.0%	42.1%
⑮ 学内業務の変化に伴う研究時間への影響	72.9%	61.6%	54.8%	65.3%	53.2%	60.6%	61.3%	61.2%	84.2%	60.3%	66.4%
⑯ 教育業務の変化に伴う研究時間への影響	83.1%	69.7%	77.4%	78.8%	66.0%	79.8%	79.4%	77.7%	76.5%	75.9%	81.4%

図 2 地域別新型コロナウイルス感染症の研究活動へのマイナスの影響 (2020年1月頃～9月の状況)

3.2 研究活動の進捗状況

図3は、調査項目2について、2020年1月頃から同年9月にかけての研究活動が、通常想定される進捗度合いと比べてどのような状況にあったかを全回答者に尋ねた結

果である。北関東、東京圏の順で研究活動が停滞又は停止したと回答する割合が大きいものに対して、東北・北陸は相対的に研究活動が停滞又は停止したと回答する割合が小さい。

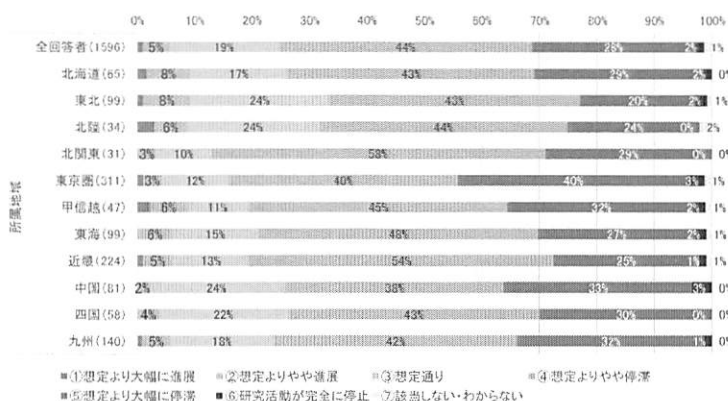


図 3 研究活動の進捗状況 (2020年1月頃～9月の状況)

3.3 研究活動を行う上でのデジタルツール等の活用状況

図4は、調査項目3について、新型コロナウイルス感染症の感染拡大が見られ始めた「2020年1月頃から本格的に活用」を選択した回答を大学の研究者の所属地域別に集計した結果である。各項目(行)につ

いて活用率が高い地域ほど濃い青色で強調している。全体的に「①テレワークシステム」と「②ウェブミーティングシステム」はコロナ禍以降に活用が進んでいることがわかる。地域別に見ると、特に東京圏、北海道、九州、近畿においてコロナ禍以降にデジタルツール等が導入されている。

	北海道 (65)	東北 (99)	北関東 (31)	東京圏 (311)	甲信越 (47)	東海 (99)	北陸 (34)	近畿 (224)	中国 (81)	四国 (58)	九州 (140)
① テレワークシステム(リモートアクセス等)	51%	49%	59%	55%	45%	46%	29%	44%	47%	41%	52%
② ウェブミーティングシステム(Zoom, Webex, Skype等)	91%	86%	77%	80%	91%	83%	79%	85%	88%	90%	88%
③ ビジネスチャット(slack, chatwork, Microsoft Teams等)	13%	21%	10%	73%	6%	28%	4%	46%	24%	17%	34%
④ ファイル共有システム(Google Drive, Dropbox等)	28%	17%	16%	20%	21%	17%	12%	18%	15%	19%	23%
⑤ プレプリントサーバへの投稿(arXiv, MedRxiv, ChemRxiv等)	0%	5%	3%	4%	2%	2%	0%	3%	2%	2%	5%
⑥ クラウド環境での論文執筆(Google Doc, Overleaf等)	11%	3%	0%	0%	9%	4%	0%	5%	10%	2%	6%
⑦ オープンデータ	0%	2%	0%	4%	9%	3%	1%	1%	2%	0%	1%
⑧ 実験機器のオンライン利用	5%	2%	0%	4%	0%	1%	0%	4%	4%	0%	2%
⑨ 実験機器の自動化	5%	1%	0%	4%	2%	0%	0%	3%	4%	2%	2%

図 4 研究活動を行う上でのデジタルツール等の活用状況 (2020年1月以降に活用開始)

3.4 現状の懸念等

図5は、研究活動を行っていく上での現状の懸念等を自由記述で尋ねた設問(調査項目4)の回答について、回答者の部局分野(赤字)と頻出語で対応分析を行った結果である。図5は、対応分析により抽出された最初の2つの成分による同時付置であり、これらの成分の累積寄与率は84.11%であった。

付置された語を見ると、工学の回答内容に特徴的な語としては、「人」「実験」「国際」「会議」「コミュニケーション」などがある。これらの語が含まれる実際の回答には、「国内外会議がオンラインのみのため、新しい人的ネットワークを築くことが極めて困難」や「実験の停滞、研究グループ内のコミュニケーション不良、実験技術の喪失」などが見られる。保健に特徴的な語としては、「確保」「遅れ」などがある。これらの語が含まれる実際の回答には、「臨床と併行しながら研究を進めているため、コロナ対応による臨床の負担増により、研究時間が十分確保しにくい」や「自分自身には特に影響はないが、実験が遅れることで研究結果の遅れ、ひいては論文の生産性に低下が生じると考える」などが見られる。農学と理学の回答は似通った傾向を持つといえるが、特に農学に特徴的な語としては、「実習」「調査」「出張」など

があり、これらの語が含まれる実際の回答には、「学生の宿泊を伴う実習(フィールドワーク等)の実施」や「とくに海外でのフィールド調査の継続が困難になり復活の見込みがない」などが見られる。特に理学に特徴的な語としては、「海外」「増加」などがあり、これらの語が含まれる実際の回答には、「海外施設を用いた研究と教育の停滞、大学院生の教育に対面の打ち合わせや実験が必要。海外施設にある実験装置を建設途中」などが見られる。

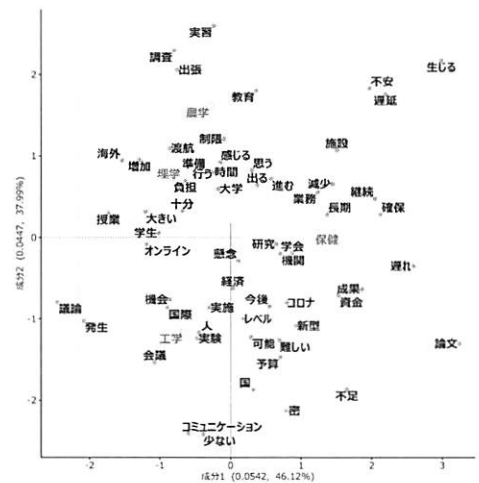


図 5 部局分野と頻出語の対応分析

3.5 今後求められる変化・対応等

調査項目5について、設問タイプ(「個人レベル」「部局・機関レベル」「国レベル」)ごとに回答内容を見ると、個人レベルおよび部局・機関レベルの回答傾向は似

通っており、具体的な回答としては「デジタル又はオンラインツールの活用や適用」に関する意見が多く見られるほか、「全てを実験だけで行う研究ではなく、データサイエンスなどの分野へ少し研究の幅を広げる努力をしている」などが見られる。国レベルでの回答としては、「検査体制（制度・インフラ・人員など）の拡充・整備」や「予算の柔軟な執行や、未実施の施策の見直しによる予算計画の再編を可能にする」などが見られる。

4 おわりに

本研究の結果から、部局分野や所属地域によって新型コロナウイルス感染症の影響の度合いや現状の懸念等が異なることがわかった。また、コロナ禍に際して今後求められる変化・対応等として挙げられる典型的な意見の一つであるデジタルツール等の活用状況についても、地域ごとに差が見られる。これらのことは、コロナ禍を踏まえた支援体制ひいては今後の科学技術政策を検討する際に、画一的ではなく、分野や地域ごとの事情に即した施策を行うことが必要となることを示唆するものである。

本研究の限界としては、調査対象者は自然科学分野の研究者に限定されており、人文・社会科学分野の研究者が含まれていないことが挙げられる。今後は後者も調査対象に含めたうえで継続的にコロナ禍の影響を注視していくことが必要となろう。

謝辞

本研究の調査項目の設計を主にご担当頂いた科学技術・学術政策研究所の村上昭義

氏、調査設計および分析についてご協力頂いた同 伊神正貫氏に感謝する。なお、本研究の内容に関する責任は著者個人に帰属する。

注記および参考文献

- [1] 文部科学省科学技術・学術審議会学術分科会（第79回）：「新型コロナウイルス感染症による学術研究への影響及び支援ニーズに関するアンケート結果（主な意見）」，参考資料3-3，2020，https://www.mext.go.jp/content/20200806-mxt_sinkou01-000009243_13.pdf（2021年4月5日参照）
- [2] 重茂浩美；蒲生秀典：「新型コロナウイルス感染症等による日本の科学技術への影響と科学者・技術者の貢献—科学技術専門家ネットワークアンケートによる東日本大震災時との比較—」，NISTEP RESEARCH MATERIAL, No. 303, 2021, 文部科学省科学技術・学術政策研究所。
- [3] 文部科学省科学技術・学術政策研究所：「科学技術の状況に係る総合的意識調査（NISTEP定点調査2020）報告書」，NISTEP REPORT, No. 189, 2021。
- [4] 文部科学省大学分科会（第135回）：「地方大学の振興等の検討のための基礎資料」，資料4，2017，https://www.mext.go.jp/b_menu/shingi/chukyo/chukyo4/gijiroku/_icsFiles/fieldfile/2017/04/13/1384455_07_1.pdf（2021年4月5日参照）
- [5] 研究者や学生の渡航・帰国，外国人研究者の着任や留学生の入国等への影響を指す。

第29回年次大会予稿

COVID-19におけるTwitterの利用傾向に関する探索的研究

Exploratory Research on Twitter Usage Trends of COVID-19

渡邊憲二^{1*}, 箕輪弘嗣¹
Kenji WATANABE ^{1*}, Hirotsugu MINOWA¹

1 岡山商科大学

Okayama Shoka University

〒700-0087 岡山県岡山市北区津島京町2丁目10-1

E-mail: k-watanabe@po.osu.ac.jp

*連絡先著者 Corresponding Author

COVID-19 が蔓延する中で、社会的な変化だけでなく、個人の生活や行動にも影響がみられている。そこで、人々の考えや社会的な反応を俯瞰することを目的として、Twitter のツイート内容から得られる質的データに着目した。Twitter のツイートには、リアルタイム性が強く、様々な情報や意見を含んだ投稿が多い。このツイート内容から得られる質的データを解析することで、COVID-19 の今後の対応や取り組みにも活用が可能となる。

本研究では、Twitter のツイート内容を対象に、テキストマイニングからツイート傾向を確認する。また、新型コロナウイルス感染の拡大によるツイート内容の変化も検討する。

The spread of COVID-19 affects not only social changes but also personal lives and behaviors. Therefore, we focused on the content of Twitter with the aim of clarifying people's thoughts and social reactions. Twitter often posts various information and opinions in real time. By analyzing the contents of Twitter, it will be possible to utilize it for future correspondence and efforts for COVID-19.

In this study, we analyzed the contents of Twitter of COVID-19 from text mining.

キーワード: COVID-19, Twitter, テキストマイニング

Keywords: COVID-19, Twitter, Text Mining

1 はじめに

2021年現在、世界中でCOVID-19（以下、「新型コロナウイルス感染症」と記載）の拡大は続いている。新型コロナウイルス感染症は、生命を脅かすことだけに留まらず、各国の政治や経済にまで影響を及ぼし、さらには人間の日常生活にまで変化をもた

らしている。

日本国内では、新型コロナウイルス感染症感染者が2020年1月に初めて確認され、新型コロナウイルス感染症の感染者数は累計23.4万人（2020年12月末）に達している。このように新型コロナウイルス感染者が増加したことで、2020年は政府による小

中高校の臨時休校要請や東京オリンピック・パラリンピックが延期となった。さらに、自治体からの自粛要請や緊急事態宣言発令等の異例な対応が続いている。個人の生活では、県外移動や不急外出の自粛だけでなく、在宅勤務やテレワーク等の新しい生活様式の導入も進んでいる。

新型コロナウイルス感染症が蔓延する中で、社会的な変化だけでなく、個人の生活や行動にも影響がみられている。そこで、人々の考えや社会的な反応を俯瞰するために、Twitterのツイート内容に注目する。Twitterのツイートには、リアルタイム性が強く、様々な情報や意見を含んだ投稿が多い。このツイート内容から得られる質的データを解析することで、新型コロナウイルス感染症への今後の対応や取り組みにも活用が可能となる。

Twitterのツイートを利用した災害時の先行研究では、村井[1]、村井[2]、河井・藤代[3]、篠田ら[4]が東日本大震災を取り上げ、ツイートの利用傾向やコミュニケーション構造の検討、さらにはツイート内容から社会状況や被災地の現状を把握することが行われている。榎本ら[5]や横田・山田[6]は、ツイートから災害時における心理状態や行動を定量的に把握している。また、峰滝[7]は、新型コロナウイルスのツイート内容から緊急事態宣言の解除がもたらした影響を明らかにしている。

本研究では、コロナ禍における人々の話題や意識を把握することを目的として、Twitterのツイート内容を対象に、テキストマイニングから全体的なツイート傾向を確認する。また、二時点間比較を用いて、新型コロナウイルス感染の拡大によるツイート内容の変化を検討する。

2 日本における新型コロナウイルス感染症の動向

ここでは、日本の新型コロナウイルス対応の動向を概観する。日本における新型コロナウイルス感染症の陽性者数を示したのが、図1である[8]。

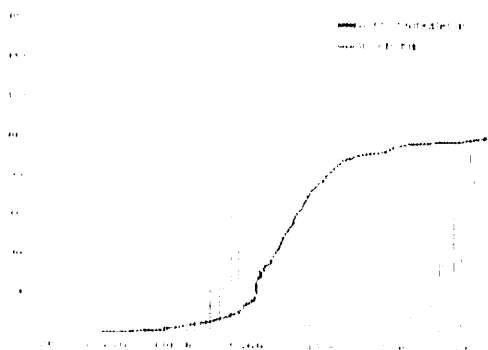


図1 日本における新型コロナウイルス感染症陽性者数の動向

2020年1月16日に、日本国内では最初の新型コロナウイルス感染者の発生が確認された。2月中旬から増え始め、3月下旬以降に新型コロナウイルス感染者が急増した。一日あたりのPCR検査陽性者数が3月27日に100名を超え、4月10日には708名にまで拡大した。その後、4月下旬から緩やかに減少しつつ、5月下旬には収束している。そして、6月下旬から再び増加し始め、6月28日には一日あたりのPCR検査陽性者数が111名となり、7月末には1,000名を上回るようになった。死亡者数をみると、国内初の新型コロナウイルスによる死亡者が2月13日に確認され、4月13日には100名に達している。さらに、4月下旬から急増し、7月下旬には1,000名を上回った。こうした新型コロナウイルス感染者数の推移

について、5月末までを第1波として、6月からが第2波としている[9].

3 データと分析手法

データ収集に関して、Twitter のツイートを収集するプログラミングコードを作成し、新型コロナウイルス感染症に関するツイートのみを収集するように対応させた。収集したツイートは、MySQL-DBへJSON-raw形式で保存した。なお、現在、低スペックのサーバで収集することができ、Excel/CSV化が実時間で実現できている。解析に用いるデータは、CSV(or Excel)形式へ変換(エクスポート)できるように、エクスポートするプログラムはPythonでコードを作成した。

本研究では、コロナ禍におけるTwitterのツイート反応から社会状況を把握するために、二時点間比較によるアプローチを行う。このアプローチを行うにあたって、分析対象の日時(以後、「分析日時」と記載)を設定する必要がある。この分析日時は、3つの基準から設定した。基準1は、一日あたりのPCR検査陽性者数の動きである。これは、収束時期から新型コロナウイルス感染者の増加によって、ツイート内容の変化をみるためである。基準2は、政府から新型コロナウイルス関連の発表や対応が行われた日時とした。これは、時事的な話題や社会的な関心事への反応をみるためである。基準3は、二時点間比較の分析日時を1カ月以内とすることである。これは、季節性、生活スタイル、社会状況などの定期的な動きや不規則な変化を極力抑えるためである。こうした3つの基準から、分析日時は6月19・20日と7月10・11日の各2日間

で、曜日は金曜日と土曜日とした。

分析日時の状況として、6月19・20日では、一日あたりのPCR検査陽性者数は54人(6/19)と65人(6/20)で、新型コロナウイルス感染者数は下げ止まった時期である。また、6月19日は、全国的に県外への移動自粛が解除されただけでなく、新型コロナウイルス接触確認アプリ(COCOA)がリリースされている。次に、3週間後の7月10・11日は、一日あたりのPCR検査陽性者数は420人(7/10)と373人(7/11)まで急増した時期である。7月10日は、観光庁からGOTOトラベルの概要が発表された。

収集したツイート数は、6月19・20日が1,361件で、7月10・11日が1,675件であった。収集したツイートは、分析の精度を高めるために、データのクレンジングを行った。クレンジングの対象として、「RT(retweet)」、「@tweet」、「URL」、「重複しているtweet広告」、「記述記号」は削除した。分析に使用するサンプル数(以下、「分析データ」と記載)は、それぞれ1,219件(6月19・20日)、1,265件(7月10・11日)となった。

分析には、テキストマイニング解析ソフトKH Coder(3.Beta.02c)を使用している。KH Coderを利用して、分析データをMeCabで形態素解析を行った。出現頻度の高い単語については、コーディング(表記揺れを含む)を定めている。出現頻度が低い(10回以下)単語は、コーディングならびに表記揺れから除外している。分析対象のコーディングは、表1に示した。出現回数が多い単語については、表2のとおりである。

表1 分析対象のコーディング

コード名	コーディングによる分類(書記振れを含む)
新型コロナ	新型コロナ, コロナ, COVID-19, COVID, SARS-CoV, SARS-COV, 新型コロナウイルス
検査	検査, PCR, PCR検査
ウイルス	ウイルス, ウイルス
日本	日本, JAPAN, japan, nihon
今日	今日, 本日, 今, 現在, now
クラスター	クラスター, 集団
COCOA	COCOA, ココア, ココアアプリ, COVID-19 Contact-Confirming Application
インストール	ダウンロード, インストール
政府	政府, 国
死亡	死亡, 亡くなる, 死者, 死ぬ
NEWS	ニュース, NEWS, News, news
都知事	都知事, 小池, 小池都知事
GOTO	GOTO, Go to, ゴートゥー
トラベル	トラベル, travel, 旅行, Travelキャンペーン
第2波	第2波, 第二波, 2波, ニ波
楽天市場	楽天, 楽天市場

表2 出現回数の多い単語リスト

出現回数の多い単語			
6月19・20日のツイート		7月10・11日のツイート	
単語	出現回数	単語	出現回数
新型コロナ	2008	新型コロナ	2000
ウイルス	456	感染	648
感染	411	ウイルス	303
マスク	316	東京	294
アプリ	207	人	189
確認	184	トラベル	165
人	176	マスク	142
接触	155	今日	136
今日	140	検査	129
対策	124	GOTO	117
日本	115	思う	109
COCOA	114	対策	108
楽天市場	93	言う	107
東京	79	増える	96
厚生労働省	78	日本	86
自粛	73	NEWS	84
6月	72	政府	78
検査	68	確認	76
インストール	67	死亡	74
死亡	66	緊急事態宣言	66
お金	61	出る	66
情報	61	自粛	64
思う	60	経済	61
予防	53	拡大	57
副業	52	都知事	56
言う	51	陽性	55
無料	51	自分	53
負ける	50	新た	52
在庫	49	7月	51

出現回数の多い単語として、「新型コロナ」、「ウイルス」、「感染」が挙げられる。6月19・20日から7月10・11日の動きをみると、「マスク」は大きく減少しているものの、「検査」、「増える」、「拡大」、「陽性」といった単語は出現回数が増えている。また、政府が関わる内容について、6月19・20日では「アプリ」、「COCOA」、「厚生労働省」、「インストール」といった単語が多く確認され、7

月10・11日では「トラベル」、「GOTO」といった単語が増えている。こうしたことから、Twitterのツイートは、新型コロナウイルス感染症の話題だけでなく、時事的な話題や社会的関心事にも一時的に強く反応していることが窺える。

4 分析結果

ここでは、MDS（多次元尺度構成法）と共起ネットワークを用いて、Twitterによるツイートの利用傾向や社会状況を把握する。MDSは、単語間の類似性や相違性を視覚化することができる分析である。特徴としては、原点付近にプロットされた場合は、ツイートで一般的に使われた単語といえる。また、原点からの距離が離れるほど、特徴的な単語であるといえる。そして、単語間の関係が強い場合には単語が近くに集まり、関係が弱い場合には離れてプロットされる。バブルのサイズは、出現回数が多いほど大きな円となる。共起ネットワークは、出現パターンの類似する単語の共起関係がみてとれる。共起ネットワークで結ばれていれば、同じツイート内で使われており、単語間の関連性が強いといえる。また、共起関係の強弱は、Jaccard係数で判断することができる。

MDSと共起ネットワークによる分析結果は、それぞれ図2と図3である。図2は、6月19・20日の結果であり、8つのクラスターが確認された。原点には、「新型コロナ」があり、その周囲には「ウイルス」や「感染」といった単語が集まっている。第一象限には、「検査」、「陽性」、「死亡」等の新型コロナウイルス感染結果に関する単語がみられる。第二象

限には、「COCOA」、「アプリ」、「政府」等の行政に関する単語が確認された。第三象限では、「仕事」、「在宅」、「自粛」等の日常生活に関する単語である。第四象限は、「楽天市場」や「販売」等のECに関する単語が集まっている。

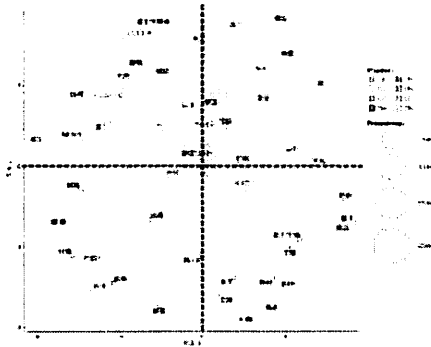


図2 MDSの分析結果 (6/19・20)

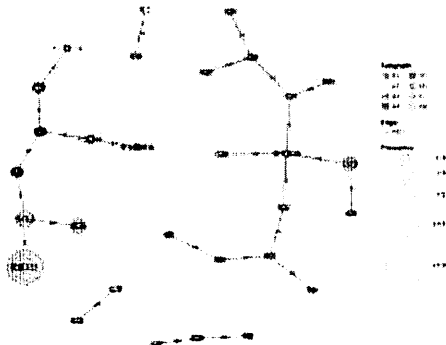


図3 共起ネットワーク (6/19・20)

共起ネットワークに関して、ツイート内容を見ると、中規模程度の話題が2つ確認されている。一つは、「新型コロナ」と新型コロナウイルス接触確認アプリ(COCOA)の関連が挙げられる。もう一つは、「楽天市場」を通じた食品やマスクの販売ならびに在庫に関する話題であった。また、「自粛解除」や「在宅副業ーお金」の日常生活に関するツイートもみてとれる。

これらのことから、ツイートの利用傾向として、時事的な内容が多くツイートされていることがみてとれる。また、行政の発表や対応だけでなく、日常生活の話題も確認された。特に、コロナ禍による新たな生活スタイルを表す多様な話題がツイートされていることが確認された。

次に、図4と図5は7月10・11日の結果である。図4では、9つのクラスターがみられた。原点には、「新型コロナ」の単語を中心に、「東京」、「感染」「ウイルス」、「増える」等の単語が近くにみられている。第一象限には、「GOTO」、「トラベル」、「政府」等の行政に関する単語が集まっている。第二象限には、「発表」、「NEWS」、「更新」といった報道に関する単語がみられた。第三象限には、「検査」、「陽性」、「研究」、「結果」、「クラスター」等の新型コロナウイルスの発生や調査に関する単語が確認できる。第四象限には、「第2波」、「来る」、「マスク」、「対策」、「仕事」、「生活」、「影響」等の生活様式や今後の動向に関する単語が現れている。

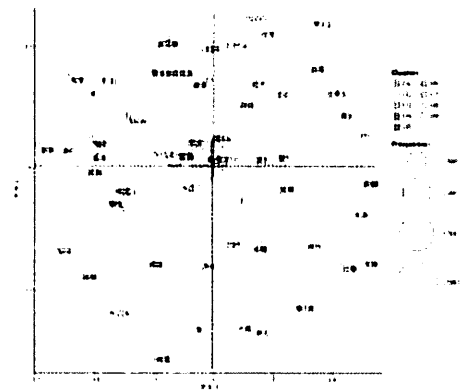


図4 MDSの分析結果 (7/10・11)

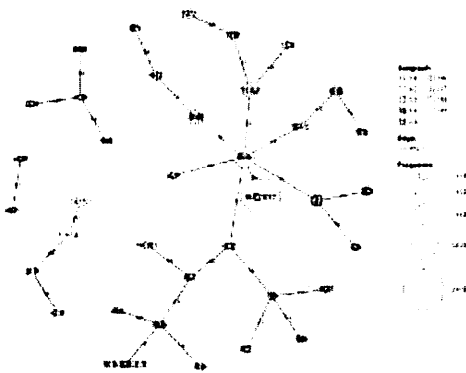


図5 共起ネットワーク (7/10・11)

共起ネットワークについて、図5では、「新型コロナー感染」を軸にした単語の繋がりが大きくなっている。特に、「感染-東京」, 「感染-ウイルス-対策」, 「感染-増える-検査-陽性」といった単語の繋がりが多くみられる。また、「GOTO-トラベル-政府-経済」といった政策に関する単語も比較的多く出現している。加えて、仕事に関する共起関係のネットワークもみられた。

これらのことから、ツイートの話題として、新型コロナウイルス感染に関する報道や情報が多くみられている。また、共起ネットワークでは、「新型コロナー感染」が話題の中心となっており、新型コロナウイルスの懸念や猛威に関する話題が多くを占めている。興味のある時事的な話題として、GOTOトラベルの概要が発表されたことや生活不安のツイートも比較的確認することができた。感染者数の増加に伴う状況変化から、今後の動向を意識するような話題が出現していることが推察される。

5 まとめ

本研究では、コロナ禍における人々の話題や意識を把握することを目的として、

Twitterのツイート内容を対象に、テキストマイニングから全体的なツイート傾向を確認した。また、二時点間比較を用いて、新型コロナウイルス感染の拡大によるツイート内容の変化を検討した。得られた知見は、以下のとおりである。

形態素解析から、新型コロナウイルス感染症が収束している時期では、政府の対応に関わる時事的な話題や単語が多く確認された。新型コロナウイルス感染者数が増加している時期では、「検査」、

「増える」、「拡大」、「陽性」といった単語の出現回数が増え、新型コロナウイルス感染症の拡大の話題が多く表れている。また、政府に関わる内容について、「トラベル」、「GOTO」といった単語が増えている。こうしたことから、Twitterのツイートは新型コロナウイルス感染症の話題だけでなく、時事的な話題や社会的関心事にも一時的に強く反応する傾向が窺えた。

分析結果から、共通の話題として、新型コロナウイルス感染に関する報道や情報が多くみられた。さらに、政府の発表や対応だけでなく、日常生活の話題も確認された。感染者数の増加に伴う状況変化として、今後の動向を意識するような話題が比較的多く出現していた。

共起ネットワークでは、新型コロナウイルス感染症が収束している時期では、時事的な内容が多くツイートされていることがみてとれた。特に、コロナ禍による新たな生活スタイルを表す多様な話題がツイートされていることが確認された。一方で、新型コロナウイルス感染者数が増加している時期では、新型コロナウイルスの懸念や猛威を表すツイートが多く

を占めているものの、時事的な内容や生活不安の話題も比較的多かった。

最後に、本研究では、コロナ禍における人々の話題や意識について、二時点間比較による探索的なアプローチを行ったに過ぎない。今後の課題として、継続的にデータ収集をしていくことで、より詳細な分析が可能である。このことを踏まえ、他稿に譲りたい。

参考文献

- [1]村井源：「東日本大震災後のTwitter利用傾向－震災関連ハッシュタグの計量的分析－」，情報知識学会誌，Vol. 22，No.2，pp.97-106，2012.
- [2]村井源：「東日本大震災でのTwitterハッシュタグの利用傾向調査－震災時のICT活用に向けて－」，情報知識学会誌，Vol. 23，No.3，pp.355-370，2013.
- [3]河井孝仁；藤代裕之：「東日本大震災の災害情報におけるTwitterの利用分析」，広報研究，第17号，pp. 118-128，2013.
- [4]篠田孝祐；榊剛史；鳥海不二夫；風間一洋；栗原聡；野田五十樹；松尾豊：「東

日本大震災時における Twitter の活用状況とコミュニケーション構造の分析」，知能と情報，Vol. 25，No.1，pp.598-608，2013.

[5]榎本甫；桑野将司；小池淳司：「災害時のソーシャルメディアと帰宅行動の関連性分析」，土木学会論文集D3（土木計画学），Vol. 70，No. 1，pp.102-112，2014.

[6]横田尚己；山田圭二郎：「熊本地震のつぶやきに見る感情極性値の時空間解析」，都市計画論文集，Vol.52，No.3，pp.1081-1087，2017.

[7] 峰滝和典：「Twitterデータを利用した「新型コロナウイルス」関連語句の分析」商経学叢，Vol. 67，No. 1，pp. 195-219，2020.

[8]厚生労働省：新型コロナウイルス感染症について

https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/0000164708_00001.html（2021年04月08日参照）

[9]鈴木基：COVID-19の致命率と重症化リスク因子について，2020

<https://www.mhlw.go.jp/content/10900000/00662183.pdf>（2021年04月08日参照）

第29回年次大会予稿

官民データ活用推進基本法の施行と自治体における

オープンデータ推進の関係

Relationship between plan for the advancement of public and private sector data utilization and open data promotion

本田正美^{1*}

Masami HONDA^{1*}

1 関東学院大学経済経営研究所客員研究員

Research Institute for Economics and Business Administration Kanto Gakuin University

〒236-8501 神奈川県横浜市金沢区六浦東1丁目50-1

E-mail: ask@honda-masami.jp

*連絡先著者 Corresponding Author

2016年末に官民データ活用推進基本法が施行された。同法施行前から、2012年には電子行政オープンデータ戦略が策定され、オープンデータの公開が国と自治体で推進されてきた。データ活用に関する法律制定が自治体におけるオープンデータ施策の推進に影響を及ぼすのか。本研究は、オープンデータ着手済の自治体の対象に、官民データ活用推進基本法に基づく官民データ活用推進計画の策定状況を確認することで、その影響の有無を確認する。

At the end of 2016, Basic Act on the Advancement of Public and Private Sector Data Utilization came into effect. Even before the law came into effect, the electronic administration open data strategy was formulated in 2012, and the disclosure of open data has been promoted by the national and local governments. Will the enactment of data utilization legislation affect the promotion of open data measures in local governments? This study confirms the influence of the plan formulation by confirming the formulation status of the “Plan for the Advancement of Public and Private Sector Data Utilization” for local governments that have already started open data.

キーワード:官民データ活用推進計画, オープンデータ, 政策形成

Keyword : Plan for the advancement of public and private sector data utilization, Open Data, Policy Making

1 研究の背景と目的

2016年に官民データ活用推進基本法が施行された。それまでに、2012年には電子行政オープンデータ戦略が策定され、公共機関におけるオープンデータの取り組みが推進されてきたところであるが、同法の施行により、さらなるデータの公開と活用が進められることとなった。

官民データ活用推進基本法は国と都道府県に対して官民データ活用推進計画の策定を義務付けている。同計画により、データの公開や活用が推進されることになり、このことから、都道府県にあってはその全てがオープンデータに着手済となっている[1]。

市区町村にあっては、官民データ活用推進基本法で官民データ活用推進計画について策定が努力義務とされており、計画策定の有無は各団体に判断が任されている。このことから、官民データ活用推進計画を策定する団体もあれば、策定しない団体もある。

前述のように、2012年段階で電子行政オープンデータ戦略が策定されており、2016年の官民データ活用推進基本法の施行よりも前から、既に自治体レベルでオープンデータの取り組みは進展している。

自治体におけるオープンデータの着手の契機には、外部からの提案、首長の意向、国の政策動向の参照、議会からの提案などがある[2]。都道府県においては、官民データ活用推進基本法による計画策定の義務化が大きな契機となっていると目されるものの、基礎自治体については計画策定が努力義務であって、計画策定がオープンデータ着手の契機になっているのか検証が必要とされる。

国は、「地方公共団体オープンデータ推進ガイドライン」や「推奨データセット」の公表など、自治体におけるオープンデータの推進を後押しする取り組みを実施しており、官民データ活用推進基本法の制定もその延長線上に位置付けられるものと考えられる。

そこで、本研究では、官民データ活用推進基本法の施行が自治体のオープンデータ政策に及ぼした影響の有無に着目する。

具体的には、自治体における官民データ活用推進計画の策定がオープンデータの着手を促すのか。あるいは、官民データ活用推進計画の策定の有無にかかわらず、自治体においてはオープンデータが推進されているのか。これが本研究のリサーチクエスションとなる。

2 本研究の対象と作業

本研究では、オープンデータに着手済の自治体に着目する。

まず、自治体でオープンデータに着手された時点を確認する。そして、それらの自治体において、どの時点で官民データ活用推進計画を策定していたのかを確認する。この作業によって、オープンデータ着手の時点と官民データ活用推進計画策定の時点を比較するが可能となり、その前後関係から、官民データ活用推進基本法施行が自治体のオープンデータの取り組みに及ぼした影響を間接的に検証出来る。

本研究では、以下の作業を行う。

- ①オープンデータ着手済自治体の着手時点の特定
- ②官民データ活用推進計画の策定時点の特定

まず①については、内閣官房情報通信技術(IT)総合戦略室が「オープンデータ取組済自治体一覧」を累次公表しているため、そのデータを利用することとする。このデータは過去分も公表されている。

データが公表されているWebページには、「なお、平成31年3月11日時点更新以降に追加・更新した自治体については、追加日・更新日を記載しています。」[3]とされている。

表1 過去の取り組み自治体一覧

ファイル名	時点
opendata_ig_mani_list_20171220	2017年12月20日
opendata_ig_mani_list_20180319	2018年3月19日
opendata_ig_mani_list_20180430	2018年4月30日
opendata_ig_mani_list_20180917	2018年9月17日
opendata_ig_mani_list_20181216	2018年12月16日
opendata_ig_mani_list_20190311	2019年3月11日
opendata_ig_mani_list_20190617	2019年6月17日
opendata_ig_mani_list_20190917	2019年9月17日
opendata_ig_mani_list_20191216	2019年12月16日
opendata_ig_mani_list_20200302	2020年3月2日

(出典：[4]より取得した)

表1にあるとおり、過去については10の時点についてオープンデータ取組済の自治体一覧を記したファイルが公開されている。その他に、2021年3月12日時点の最新版の自治体一覧のファイルが公開されている[5]。

2021年3月版のファイルでは、2019年3月11日版以降の変更について日時を追記するという対応が取られているため、この2021年3月版を土台に、表1の各ファイルの内容を反映させるかたちで、オープンデータ着手済の自治体について、いずれの時点で着手済となって記載されたのかを記載した一覧を作成することとした。

これは、あくまでも内閣官房情報通信技術(IT)総合戦略室が当該自治体のオープンデータの着手を認知し、その結果を図表1にあるいずれかの時点でファイルに反映されたことについて、それをまとめたに過ぎない。ファイルは10時点分しか作成されていないため、各自治体について正確なオープンデータの着手時点を把握することは出来ない。ただし、概ね三カ月おきにファイルは作成されており、大きな乖離はなく自治体のオープンデータの着手時点を推定することが出来るものと考えられる。

②については、①で明らかとなった最新のオープンデータ取組済の1018団体について、官民データ活用推進計画の策定の有無を確認する作業となる。

なお、2019年7月1日時点の策定済団体については、内閣官房情報通信技術(IT)総合戦略室が「地方の官民データ活用推進計画策定済団体」[6]を公表しているため、この資料も参照することとした。

1018団体については、各団体のWebサイトにアクセスし、サイト内検索を用いて「官民データ活用推進計画」と打ち込み検索を行った。加えて、Google検索と同様の用語を打ち込み検索を行うことで抜け漏れを防ぐこととした。

検索の結果、何らかの計画の存在が明らかになった場合には、その内容を確認し、いずれの時点で策定されたものなのか記録した。

3 結果

前述の作業によって得られたオープンデータ着手済の自治体について、その着手の時点と計画策定の時点は以下のとおり

である。全1018団体あるため、図表は分割している。

表2 オープンデータ取組自治体(1)

団体名	リスト掲載時点	官民データ計画
北海道札幌市	2017年12月20日	2020年3月
北海道函館市	2017年12月20日	
北海道小樽市	2017年12月20日	
北海道旭川市	2017年12月20日	
北海道室蘭市	2017年12月20日	
北海道帯広市	2019年9月17日	
北海道網走市	2020年11月4日	
北海道根室市	2021年1月21日	
北海道千歳市	2019年9月17日	
北海道滝川市	2020年10月19日	
北海道砂川市	2020年2月19日	
北海道深川市	2018年12月16日	
北海道富良野市	2018年12月16日	2021年4月
北海道登別市	2019年6月17日	
北海道新穂津村	2019年3月11日	
北海道知内町	2021年2月15日	
北海道森町	2017年12月20日	2018年5月
北海道八雲町	2017年12月20日	
北海道奥尻町	2021年3月10日	
北海道せたな町	2020年12月2日	
北海道島牧村	2021年3月2日	
北海道ニセコ町	2019年9月17日	
北海道喜茂別町	2018年12月16日	
北海道倶知安町	2019年3月11日	
北海道岩内町	2020年2月27日	
北海道積丹町	2020年8月26日	
北海道上砂川町	2020年4月30日	
北海道長沼町	2021年2月17日	
北海道沼田町	2019年3月11日	
北海道上高良野町	2019年12月2日	
北海道占冠村	2019年6月17日	
北海道増毛町	2020年3月31日	
北海道小平町	2020年5月27日	
北海道中頓別町	2020年1月17日	
北海道礼文町	2020年7月28日	
北海道利尻富士町	2020年3月31日	
北海道幌延町	2020年7月20日	
北海道平取町	2021年2月12日	
北海道新冠町	2020年7月21日	
北海道様似町	2021年1月19日	
北海道音更町	2018年12月16日	
北海道上士幌町	2020年7月21日	
北海道鹿追町	2019年3月11日	
北海道新得町	2021年3月5日	
北海道中札内村	2020年3月18日	
北海道更別村	2019年6月17日	
北海道池田町	2019年10月18日	
北海道足寄町	2019年9月7日	
北海道釧路町	2020年7月14日	2016年3月
北海道厚岸町	2020年8月3日	
北海道浜中町	2020年3月12日	
北海道別海町	2019年3月11日	
青森県青森市	2017年12月20日	
青森県弘前市	2017年12月20日	
青森県八戸市	2017年12月20日	
青森県黒石市	2019年9月17日	

表3 オープンデータ取組自治体(2)

団体名	リスト掲載時点	官民データ計画
青森県五所川原市	2020年6月16日	
青森県十和田市	2018年3月19日	
青森県三沢市	2019年9月17日	
青森県むつ市	2020年7月16日	
青森県つがる市	2020年7月21日	
青森県平川市	2021年2月8日	
青森県平内町	2019年9月17日	
青森県今別町	2019年9月17日	
青森県藤田村	2020年9月7日	
青森県外ヶ浜町	2019年6月17日	
青森県鮎ヶ沢町	2019年6月17日	
青森県深浦町	2019年9月17日	
青森県西目屋村	2020年10月21日	
青森県藤崎町	2020年11月17日	
青森県大鰐町	2020年7月27日	
青森県田舎館村	2019年6月17日	
青森県板柳町	2019年6月17日	
青森県鶴田町	2019年9月17日	
青森県中泊町	2020年12月2日	
青森県野辺地町	2019年9月17日	
青森県七戸町	2020年1月16日	
青森県六戸町	2019年9月17日	
青森県横浜町	2019年9月17日	
青森県東北町	2020年8月6日	
青森県六ヶ所村	2019年12月16日	
青森県おいらせ町	2019年6月17日	
青森県大間町	2020年7月21日	
青森県東通村	2020年7月16日	
青森県風間浦村	2020年10月15日	
青森県佐井村	2019年9月17日	
青森県三戸町	2020年8月17日	
青森県五戸町	2019年3月11日	
青森県田子町	2019年9月17日	
青森県南郷町	2019年9月17日	
青森県階上町	2020年8月4日	
青森県新郷村	2019年9月17日	
岩手県盛岡市	2018年12月16日	2019年12月
岩手県宮古市	2021年3月10日	2021年3月
岩手県一関市	2018年3月19日	
岩手県奥州市	2018年12月16日	
岩手県金ヶ崎町	2019年12月6日	
岩手県平泉町	2021年1月28日	
岩手県山田町	2021年3月8日	
岩手県岩泉町	2020年3月18日	
岩手県野田村	2020年3月31日	
宮城県仙台市	2017年12月20日	2018年12月
宮城県石巻市	2017年12月20日	
宮城県気仙沼市	2020年3月31日	
宮城県名取市	2018年9月17日	2020年4月
宮城県岩沼市	2019年6月17日	
宮城県登米市	2017年12月20日	
宮城県翠原市	2021年1月29日	
宮城県大崎市	2018年9月17日	
宮城県柴田町	2019年9月17日	2019年3月
宮城県大和町	2020年10月5日	
宮城県涌谷町	2020年12月15日	
秋田県秋田市	2018年9月17日	2021年4月
秋田県能代市	2021年2月12日	
秋田県横手市	2017年12月20日	2018年3月
秋田県大館市	2021年3月10日	
秋田県男鹿市	2019年9月17日	
秋田県湯沢市	2017年12月20日	2018年7月
秋田県鹿角市	2019年6月17日	2017年4月

表4 オープンデータ取組済自治体(3)

団体名	リスト掲載時点	官民データ計画
秋田県由利本荘市	2018年9月17日	2020年4月
秋田県にかほ市	2020年11月13日	
秋田県仙北市	2021年2月25日	
山形県山形市	2017年12月20日	
山形県米沢市	2019年6月17日	
山形県酒田市	2020年1月16日	2020年3月
山形県上市市	2019年9月17日	
山形県長井市	2020年3月31日	
山形県天童市	2019年3月11日	
山形県西川町	2021年1月27日	
山形県舟形町	2020年11月30日	2021年3月
山形県庄内町	2021年2月12日	
福島県福島市	2018年4月30日	2021年3月
福島県会津若松市	2017年12月20日	2020年3月
福島県郡山市	2018年3月19日	2019年8月
福島県いわき市	2017年12月20日	2020年2月
福島県須賀川市	2020年8月27日	2019年3月
福島県喜多方市	2017年12月20日	
福島県二本松市	2020年5月1日	
福島県田村市	2020年8月17日	
福島県伊達市	2020年3月18日	
福島県本宮市	2021年2月12日	
福島県桑折町	2020年6月11日	
福島県国見町	2020年6月16日	
福島県川俣町	2020年2月18日	
福島県大玉村	2020年6月24日	
福島県鏡石町	2020年8月5日	
福島県天栄村	2020年12月14日	
福島県下郷町	2021年1月25日	
福島県只見町	2021年1月7日	
福島県南会津町	2021年2月8日	
福島県北塩原村	2020年12月23日	
福島県磐梯町	2021年2月15日	
福島県猪苗代町	2020年9月7日	
福島県柳津町	2021年2月3日	
福島県昭和村	2021年1月14日	
福島県西郷村	2020年12月23日	
福島県泉崎村	2019年9月17日	
福島県中島村	2021年3月12日	
福島県矢祭町	2021年2月8日	
福島県塙町	2020年8月3日	
福島県鮫川村	2021年2月3日	
福島県石川町	2021年2月15日	
福島県玉川村	2021年2月26日	
福島県平田村	2021年2月15日	
福島県浅川町	2020年12月16日	
福島県三春町	2021年2月8日	
福島県小野町	2020年6月11日	
福島県広野町	2021年3月3日	
福島県楢葉町	2020年9月10日	
福島県大熊町	2021年2月3日	
福島県双葉町	2020年12月14日	
福島県葛尾村	2020年1月20日	
茨城県水戸市	2017年12月20日	
茨城県日立市	2019年6月17日	2018年2月
茨城県石岡市	2020年1月16日	
茨城県結城市	2021年3月12日	
茨城県常総市	2019年12月5日	2019年3月
茨城県常陸太田市	2020年6月11日	
茨城県高萩市	2020年6月16日	
茨城県笠間市	2017年12月20日	2020年9月
茨城県つくば市	2019年6月17日	2018年9月
茨城県ひたちなか市	2019年6月17日	

表5 オープンデータ取組済自治体(4)

団体名	リスト掲載時点	官民データ計画
茨城県筑西市	2019年3月11日	
茨城県かすみがうら市	2020年1月15日	
茨城県桜川市	2020年3月31日	
茨城県神栖市	2019年6月17日	
茨城県行方市	2021年2月12日	
茨城県つくばみらい市	2021年2月15日	
茨城県小美玉市	2020年8月3日	
茨城県大洗町	2020年10月15日	
茨城県東海村	2017年12月20日	
茨城県大子町	2019年9月17日	
茨城県美浦村	2021年3月8日	
茨城県阿見町	2020年12月25日	
茨城県五霞町	2020年6月23日	
茨城県利根町	2021年2月17日	
栃木県宇都宮市	2017年12月20日	
栃木県栃木市	2021年3月10日	2021年3月
栃木県佐野市	2018年4月30日	
栃木県鹿沼市	2019年6月17日	
栃木県小山市	2017年12月20日	
栃木県真岡市	2019年6月17日	
栃木県大田原市	2017年12月20日	2020年3月
栃木県那須塩原市	2017年12月20日	2019年3月
栃木県さくら市	2017年12月20日	
栃木県那須烏山市	2019年6月17日	
栃木県上三川町	2021年2月1日	
栃木県茂木町	2021年2月3日	
栃木県壬生町	2020年1月16日	
栃木県高根沢町	2017年12月20日	
栃木県那須町	2017年12月20日	
栃木県那珂川町	2019年3月11日	
群馬県前橋市	2017年12月20日	2021年3月
群馬県高崎市	2020年12月23日	
群馬県桐生市	2019年6月17日	
群馬県伊勢崎市	2020年3月12日	
群馬県太田市	2020年7月13日	
群馬県渋川市	2020年6月11日	
群馬県安中市	2021年3月3日	
群馬県榛東村	2020年12月25日	
群馬県中之条町	2020年8月4日	
群馬県高山村	2020年6月11日	
埼玉県さいたま市	2017年12月20日	
埼玉県川越市	2017年12月20日	
埼玉県熊谷市	2017年12月20日	2019年3月
埼玉県川口市	2017年12月20日	
埼玉県行田市	2019年9月17日	
埼玉県秩父市	2020年12月14日	
埼玉県所沢市	2019年6月17日	2019年3月
埼玉県飯能市	2018年4月30日	
埼玉県本庄市	2021年3月8日	
埼玉県東松山市	2019年6月17日	
埼玉県春日部市	2019年6月17日	
埼玉県狭山市	2020年5月29日	
埼玉県羽生市	2021年2月17日	
埼玉県鴻巣市	2017年12月20日	
埼玉県深谷市	2017年12月20日	
埼玉県上尾市	2019年6月17日	
埼玉県草加市	2018年4月30日	
埼玉県越谷市	2017年12月20日	2021年4月
埼玉県蕨市	2019年6月17日	
埼玉県戸田市	2017年12月20日	2021年3月
埼玉県入間市	2019年3月11日	
埼玉県朝霞市	2017年12月20日	
埼玉県志木市	2019年6月17日	

表6 オープンデータ取組済自治体(5)

団体名	リスト掲載時点	官民データ計画
埼玉県和光市	2017年12月20日	
埼玉県新座市	2017年12月20日	
埼玉県桶川市	2019年6月17日	
埼玉県久喜市	2017年12月20日	
埼玉県北本市	2017年12月20日	
埼玉県八潮市	2021年2月15日	
埼玉県富士見市	2018年9月17日	
埼玉県三郷市	2017年12月20日	
埼玉県蓮田市	2019年9月17日	
埼玉県坂戸市	2019年6月17日	2020年3月
埼玉県鶴ヶ島市	2019年6月17日	
埼玉県日高市	2017年12月20日	
埼玉県吉川市	2019年6月17日	
埼玉県ふじみ野市	2017年12月20日	2019年2月
埼玉県白岡市	2019年9月17日	
埼玉県伊奈町	2019年6月17日	
埼玉県三芳町	2017年12月20日	
埼玉県毛呂山町	2017年12月20日	
埼玉県越生町	2019年6月17日	
埼玉県滑川町	2019年6月17日	
埼玉県吉見町	2019年6月17日	
埼玉県鳩山町	2017年12月20日	
埼玉県東秩父村	2017年12月20日	
埼玉県神川町	2019年6月17日	
埼玉県上里町	2020年12月16日	
埼玉県寄居町	2020年8月21日	
埼玉県杉戸町	2017年12月20日	
千葉県千葉市	2017年12月20日	
千葉県市川市	2017年12月20日	
千葉県船橋市	2018年3月19日	
千葉県館山市	2019年6月17日	
千葉県木更津市	2017年12月20日	
千葉県松戸市	2017年12月20日	2018年3月
千葉県野田市	2017年12月20日	
千葉県茂原市	2019年6月17日	
千葉県成田市	2021年2月18日	
千葉県佐倉市	2019年6月17日	
千葉県旭市	2020年7月9日	
千葉県習志野市	2017年12月20日	
千葉県柏市	2018年3月19日	
千葉県勝浦市	2021年3月9日	
千葉県市原市	2018年3月19日	
千葉県流山市	2017年12月20日	2018年3月
千葉県八千代市	2018年12月16日	
千葉県我孫子市	2020年4月22日	
千葉県富津市	2019年9月17日	
千葉県浦安市	2017年12月20日	
千葉県四街道市	2019年9月17日	2021年3月
千葉県袖ヶ浦市	2019年9月17日	
千葉県八街市	2019年9月17日	
千葉県印西市	2017年12月20日	
千葉県富里市	2020年2月18日	
千葉県香取市	2021年2月8日	
千葉県大網白里市	2021年2月26日	
千葉県酒々井町	2020年10月6日	
千葉県芝山町	2021年3月3日	
東京都千代田区	2017年12月20日	
東京都中央区	2019年6月17日	
東京都港区	2017年12月20日	2018年3月
東京都新宿区	2017年12月20日	2018年3月
東京都文京区	2017年12月20日	2019年3月
東京都台東区	2017年12月20日	
東京都墨田区	2017年12月20日	2019年3月

表7 オープンデータ取組済自治体(6)

団体名	リスト掲載時点	官民データ計画
東京都江東区	2019年12月17日	2020年3月
東京都品川区	2017年12月20日	
東京都目黒区	2018年9月17日	
東京都大田区	2019年6月17日	2021年3月
東京都世田谷区	2017年12月20日	2018年3月
東京都渋谷区	2019年9月17日	
東京都中野区	2019年12月16日	
東京都杉並区	2017年12月20日	2019年7月
東京都豊島区	2017年12月20日	
東京都北区	2019年9月17日	2020年2月
東京都板橋区	2017年12月20日	2019年3月
東京都練馬区	2017年12月20日	2020年3月
東京都足立区	2017年12月20日	
東京都葛飾区	2019年6月17日	2019年3月
東京都江戸川区	2017年12月20日	
東京都八王子市	2017年12月20日	2018年7月
東京都立川市	2019年6月17日	
東京都武蔵野市	2019年6月17日	2020年2月
東京都三鷹市	2018年9月17日	2020年3月
東京都青梅市	2019年6月17日	
東京都府中市	2017年12月20日	
東京都昭島市	2021年3月2日	
東京都調布市	2017年12月20日	
東京都町田市	2017年12月20日	
東京都小金井市	2020年1月30日	
東京都小平市	2019年6月17日	
東京都日野市	2017年12月20日	
東京都東村山市	2017年12月20日	
東京都国分寺市	2019年6月17日	
東京都福生市	2017年12月20日	
東京都狛江市	2019年6月17日	2021年3月
東京都東大和市	2020年12月2日	2019年3月
東京都清瀬市	2019年3月11日	
東京都東久留米市	2018年9月17日	
東京都武蔵村山市	2017年12月20日	2021年3月
東京都多摩市	2017年12月20日	
東京都稲城市	2017年12月20日	
東京都あきる野市	2019年6月17日	
東京都西東京市	2019年6月17日	
東京都八丈町	2019年6月17日	
神奈川県横浜市	2017年12月20日	2018年5月
神奈川県川崎市	2017年12月20日	2019年3月
神奈川県相模原市	2017年12月20日	2020年3月
神奈川県横須賀市	2017年12月20日	
神奈川県平塚市	2017年12月20日	2021年4月
神奈川県鎌倉市	2017年12月20日	
神奈川県藤沢市	2017年12月20日	2020年7月
神奈川県小田原市	2019年6月17日	2020年4月
神奈川県茅ヶ崎市	2017年12月20日	2020年7月
神奈川県逗子市	2017年12月20日	
神奈川県三浦市	2017年12月20日	
神奈川県秦野市	2018年4月30日	2021年4月
神奈川県厚木市	2017年12月20日	2021年3月
神奈川県大和市	2017年12月20日	
神奈川県伊勢原市	2019年6月17日	2020年3月
神奈川県海老名市	2017年12月20日	2020年4月
神奈川県座間市	2018年9月17日	
神奈川県南足柄市	2020年3月12日	
神奈川県綾瀬市	2017年12月20日	
神奈川県葉山町	2017年12月20日	
神奈川県寒川町	2021年1月25日	
神奈川県大磯町	2020年1月16日	
神奈川県二宮町	2020年1月16日	

表8 オープンデータ取組済自治体(7)

団体名	リスト掲載時点	官民データ計画
神奈川県中井町	2019年6月17日	
神奈川県松田町	2020年1月16日	
神奈川県山北町	2020年8月25日	
神奈川県開成町	2019年9月17日	
神奈川県箱根町	2020年1月16日	
神奈川県真鶴町	2020年3月31日	
神奈川県湯河原町	2020年3月31日	
神奈川県愛川町	2018年9月17日	
神奈川県清川村	2020年7月20日	
新潟県新潟市	2017年12月20日	2018年1月
新潟県長岡市	2017年12月20日	
新潟県三条市	2017年12月20日	2019年4月
新潟県柏崎市	2020年4月22日	
新潟県新発田市	2018年9月17日	
新潟県小千谷市	2021年2月5日	
新潟県加茂市	2020年12月16日	
新潟県十日町市	2017年12月20日	
新潟県見附市	2017年12月20日	
新潟県燕市	2019年6月17日	
新潟県糸魚川市	2017年12月20日	2019年7月
新潟県妙高市	2019年6月17日	
新潟県五泉市	2020年9月3日	
新潟県上越市	2017年12月20日	
新潟県魚沼市	2020年10月26日	
新潟県南魚沼市	2021年2月10日	
新潟県弥彦村	2021年2月25日	
富山県高山市	2017年12月20日	
富山県高岡市	2017年12月20日	
富山県魚津市	2017年12月20日	
富山県水見市	2019年6月17日	
富山県滑川市	2019年6月17日	
富山県黒部市	2019年6月17日	
富山県砺波市	2017年12月20日	
富山県小矢部市	2019年9月17日	
富山県南砺市	2017年12月20日	
富山県射水市	2019年3月11日	
富山県舟橋村	2019年6月17日	
富山県上市町	2019年6月17日	
富山県立山町	2019年6月17日	
富山県入善町	2019年6月17日	
富山県朝日町	2019年6月17日	
石川県金沢市	2017年12月20日	2018年2月
石川県七尾市	2017年12月20日	
石川県小松市	2019年6月17日	
石川県輪島市	2020年3月31日	
石川県珠洲市	2017年12月20日	
石川県加賀市	2018年4月30日	2020年3月
石川県羽咋市	2019年12月5日	2019年3月
石川県かほく市	2019年6月17日	
石川県白山市	2017年12月20日	
石川県能美市	2017年12月20日	
石川県野々市市	2017年12月20日	
石川県川北町	2020年3月31日	
石川県津幡町	2019年6月17日	
石川県内灘町	2017年12月20日	
石川県志賀町	2020年3月31日	
石川県宝達志水町	2020年3月31日	
石川県中能登町	2020年1月16日	
石川県穴水町	2018年9月17日	
石川県能登町	2020年3月31日	
福井県福井市	2017年12月20日	2018年4月
福井県敦賀市	2017年12月20日	
福井県小浜市	2018年9月17日	

表9 オープンデータ取組済自治体(8)

団体名	リスト掲載時点	官民データ計画
福井県大野市	2017年12月20日	
福井県勝山市	2017年12月20日	
福井県鯖江市	2017年12月20日	
福井県あわら市	2017年12月20日	
福井県越前市	2017年12月20日	
福井県坂井市	2017年12月20日	2020年3月
福井県永平寺町	2017年12月20日	
福井県池田町	2017年12月20日	
福井県南越前町	2017年12月20日	
福井県越前町	2017年12月20日	
福井県美浜町	2017年12月20日	
福井県高浜町	2017年12月20日	
福井県おおい町	2017年12月20日	
福井県若狹町	2017年12月20日	
山梨県甲府市	2018年4月30日	2021年2月
山梨県韮崎市	2020年6月3日	
山梨県笛吹市	2021年1月21日	
山梨県市川三郷町	2020年11月10日	
山梨県身延町	2020年7月27日	
山梨県南部町	2021年1月25日	
山梨県富士川町	2021年2月22日	
山梨県昭和町	2021年2月9日	
山梨県鳴沢村	2020年8月17日	
長野県長野市	2018年3月19日	
長野県松本市	2019年6月17日	
長野県上田市	2019年3月11日	
長野県岡谷市	2018年3月19日	2019年3月
長野県飯田市	2019年9月17日	
長野県諏訪市	2020年3月31日	
長野県須坂市	2017年12月20日	
長野県小諸市	2020年3月31日	
長野県伊那市	2018年9月17日	
長野県駒ヶ根市	2017年12月20日	
長野県中野市	2017年12月20日	
長野県大町市	2020年3月31日	
長野県飯山市	2020年3月31日	
長野県茅野市	2019年6月17日	2018年4月
長野県塩尻市	2017年12月20日	
長野県佐久市	2020年3月31日	
長野県千曲市	2021年3月12日	
長野県東御市	2020年3月31日	
長野県安曇野市	2019年9月17日	2019年3月
長野県小海町	2020年3月31日	
長野県川上村	2020年3月31日	
長野県南牧村	2020年3月31日	
長野県南相木村	2021年3月10日	
長野県北相木村	2020年3月31日	
長野県佐久穂町	2020年3月31日	
長野県軽井沢町	2020年3月31日	2020年3月
長野県御代田町	2020年3月31日	
長野県立科町	2020年3月31日	
長野県青木村	2020年3月31日	
長野県長和町	2020年3月31日	
長野県下諏訪町	2020年3月31日	
長野県富士見町	2020年5月19日	
長野県原村	2020年1月15日	
長野県辰野町	2018年3月19日	
長野県箕輪町	2019年6月17日	
長野県南箕輪村	2020年3月31日	
長野県中川村	2019年6月17日	
長野県宮田村	2020年3月31日	
長野県松川町	2020年3月31日	
長野県高森町	2020年3月31日	

表10 オープンデータ取組自治体(9)

団体名	リスト掲載時点	官民データ計画
長野県阿南町	2020年3月31日	
長野県阿智村	2020年3月31日	
長野県平谷村	2020年3月31日	
長野県根羽村	2020年3月31日	
長野県下條村	2020年3月31日	
長野県赤木村	2020年12月15日	
長野県天龍村	2020年3月31日	
長野県泰阜村	2020年3月31日	
長野県喬木村	2019年6月17日	
長野県豊丘村	2020年3月18日	
長野県大鹿村	2020年3月31日	
長野県南木曾町	2020年3月31日	
長野県木祖村	2020年3月31日	
長野県王滝村	2020年3月31日	
長野県大桑村	2020年3月31日	
長野県木曾町	2020年3月31日	
長野県麻績村	2020年3月31日	
長野県朝日村	2020年3月31日	
長野県筑北村	2021年2月12日	
長野県池田町	2020年5月13日	
長野県松川村	2020年3月31日	
長野県白馬村	2020年3月31日	
長野県小谷村	2020年3月31日	
長野県坂城町	2021年2月12日	
長野県小布施町	2020年3月31日	
長野県高山村	2020年3月31日	
長野県木島平村	2020年3月31日	
長野県信濃町	2020年3月31日	
長野県飯綱町	2020年3月31日	
岐阜県岐阜市	2019年3月11日	2020年3月
岐阜県大垣市	2017年12月20日	
岐阜県高山市	2019年3月11日	
岐阜県多治見市	2019年3月11日	
岐阜県関市	2017年12月20日	
岐阜県中津川市	2019年3月11日	
岐阜県美濃市	2019年3月11日	
岐阜県瑞浪市	2019年3月11日	2020年3月
岐阜県羽島市	2017年12月20日	
岐阜県恵那市	2019年3月11日	2019年8月
岐阜県美濃加茂市	2019年6月17日	
岐阜県土岐市	2019年3月11日	2020年8月
岐阜県各務原市	2019年3月11日	
岐阜県可児市	2018年4月30日	
岐阜県山県市	2019年3月11日	
岐阜県瑞穂市	2019年3月11日	
岐阜県飛騨市	2019年3月11日	
岐阜県本巣市	2019年6月17日	
岐阜県郡上市	2019年6月17日	2020年4月
岐阜県下呂市	2019年3月11日	
岐阜県海津市	2019年3月11日	
岐阜県岐南町	2019年3月11日	
岐阜県笠松町	2019年3月11日	
岐阜県養老町	2019年3月11日	
岐阜県垂井町	2019年6月17日	
岐阜県関ヶ原町	2019年3月11日	
岐阜県神戸町	2019年3月11日	
岐阜県輪之内町	2019年3月11日	
岐阜県安八町	2019年11月15日	
岐阜県揖斐川町	2019年6月17日	
岐阜県大野町	2019年3月11日	
岐阜県池田町	2019年3月11日	
岐阜県北方町	2019年6月17日	
岐阜県坂祝町	2019年3月11日	

表11 オープンデータ取組自治体(10)

団体名	リスト掲載時点	官民データ計画
岐阜県宮加町	2019年3月11日	
岐阜県川辺町	2019年3月11日	
岐阜県七宗町	2019年6月17日	
岐阜県八百津町	2019年6月17日	
岐阜県白川町	2019年6月17日	
岐阜県東白川村	2019年11月14日	
岐阜県御嵩町	2019年3月11日	
岐阜県白川村	2019年9月17日	
静岡県静岡市	2017年12月20日	2019年3月
静岡県浜松市	2017年12月20日	2019年2月
静岡県沼津市	2017年12月20日	2019年3月
静岡県熱海市	2020年4月22日	
静岡県三島市	2017年12月20日	
静岡県富士宮市	2019年6月17日	
静岡県伊東市	2017年12月20日	
静岡県島田市	2017年12月20日	
静岡県富士市	2017年12月20日	2019年3月
静岡県磐田市	2017年12月20日	
静岡県焼津市	2017年12月20日	2018年7月
静岡県掛川市	2017年12月20日	2020年3月
静岡県藤枝市	2017年12月20日	
静岡県御殿場市	2017年12月20日	
静岡県袋井市	2017年12月20日	2019年3月
静岡県下田市	2020年10月2日	
静岡県裾野市	2017年12月20日	2018年11月
静岡県湖西市	2017年12月20日	
静岡県伊豆市	2019年6月17日	
静岡県御前崎市	2017年12月20日	
静岡県菊川市	2017年12月20日	
静岡県伊豆の国市	2017年12月20日	
静岡県牧之原市	2017年12月20日	
静岡県東伊豆町	2017年12月20日	
静岡県河津町	2019年9月17日	
静岡県南伊豆町	2017年12月20日	2018年9月
静岡県松崎町	2019年9月17日	
静岡県西伊豆町	2019年6月17日	
静岡県西南町	2017年12月20日	
静岡県清水町	2019年9月17日	
静岡県長泉町	2017年12月20日	
静岡県小山町	2017年12月20日	
静岡県吉田町	2017年12月20日	
静岡県川根本町	2017年12月20日	2018年7月
静岡県森町	2019年6月17日	
愛知県名古屋	2017年12月20日	2019年3月
愛知県豊橋市	2017年12月20日	
愛知県岡崎市	2017年12月20日	
愛知県一宮市	2017年12月20日	2021年4月
愛知県瀬戸市	2017年12月20日	2021年3月
愛知県半田市	2017年12月20日	
愛知県春日井市	2019年6月17日	
愛知県豊川市	2018年9月17日	
愛知県碧南市	2019年9月17日	
愛知県刈谷市	2017年12月20日	
愛知県豊田市	2017年12月20日	2020年7月
愛知県安城市	2018年9月17日	2021年4月
愛知県西尾市	2018年4月30日	
愛知県蒲郡市	2019年6月17日	
愛知県犬山市	2017年12月20日	
愛知県江南市	2019年4月30日	
愛知県小牧市	2017年12月20日	2021年3月
愛知県新城市	2018年12月16日	
愛知県東海市	2018年9月17日	
愛知県知多市	2017年12月20日	

表12 オープンデータ取組自治体(11)

団体名	リスト掲載時点	官民データ計画
愛知県知立市	2017年12月20日	
愛知県尾張旭市	2017年12月20日	
愛知県高浜市	2018年12月16日	
愛知県岩倉市	2017年12月20日	
愛知県豊明市	2017年12月20日	2020年4月
愛知県日進市	2017年12月20日	
愛知県田原市	2019年6月17日	
愛知県みよし市	2017年12月20日	
愛知県あま市	2020年12月1日	
愛知県長久手市	2017年12月20日	
愛知県東郷町	2017年12月20日	
愛知県豊山町	2019年9月17日	
愛知県大口町	2019年9月17日	
愛知県蟹江町	2020年3月13日	
愛知県飛島村	2019年9月17日	
愛知県阿久比町	2021年3月4日	2020年10月
愛知県東浦町	2017年12月20日	
愛知県南知多町	2019年9月17日	
愛知県美浜町	2020年8月5日	
愛知県幸田町	2019年9月17日	
愛知県設楽町	2019年9月17日	
愛知県豊根村	2019年6月17日	
三重県津市	2017年12月20日	
三重県四日市市	2017年12月20日	
三重県伊勢市	2017年12月20日	
三重県桑名市	2017年12月20日	
三重県鈴鹿市	2018年12月16日	
三重県名張市	2021年2月22日	
三重県亀山市	2019年9月17日	
三重県鳥羽市	2020年11月30日	
三重県熊野市	2021年3月5日	
三重県いなべ市	2017年12月20日	
三重県志摩市	2017年12月20日	
三重県伊賀市	2021年2月26日	
三重県木曽岬町	2019年10月19日	2019年5月
三重県東員町	2020年11月9日	
三重県菟野町	2021年2月3日	
三重県川越町	2020年11月30日	
三重県多気町	2021年1月21日	
三重県度会町	2021年2月19日	
三重県紀北町	2021年2月8日	
三重県御浜町	2020年11月19日	
滋賀県大津市	2017年12月20日	2019年3月
滋賀県長浜市	2017年12月20日	2020年3月
滋賀県草津市	2017年12月20日	2020年3月
滋賀県甲賀市	2017年12月20日	
滋賀県野洲市	2021年3月8日	
滋賀県高島市	2020年10月14日	
滋賀県多賀町	2021年3月5日	
京都府京都市	2017年12月20日	
京都府福知山市	2019年3月11日	
京都府舞鶴市	2019年3月11日	
京都府綾部市	2019年3月11日	2018年7月
京都府宇治市	2019年3月11日	
京都府宮津市	2019年3月11日	
京都府亀岡市	2019年3月11日	
京都府城陽市	2019年3月11日	
京都府向日市	2019年3月11日	
京都府長岡京市	2018年4月30日	
京都府八幡市	2019年3月11日	
京都府京田辺市	2019年3月11日	2020年7月
京都府京丹後市	2019年3月11日	
京都府南丹市	2019年3月11日	

表13 オープンデータ取組自治体(12)

団体名	リスト掲載時点	官民データ計画
京都府木津川市	2019年3月11日	
京都府大山崎町	2019年3月11日	
京都府久御山町	2019年3月11日	
京都府井手町	2019年3月11日	
京都府宇治田原町	2019年3月11日	
京都府笠置町	2019年3月11日	
京都府和束町	2019年3月11日	
京都府精華町	2019年3月11日	
京都府南山城村	2019年3月11日	
京都府京丹波町	2019年3月11日	
京都府伊根町	2019年3月11日	
京都府与謝野町	2019年3月11日	
大阪府大阪市	2017年12月20日	2018年3月
大阪府堺市	2017年12月20日	2020年8月
大阪府岸和田市	2019年6月17日	2019年8月
大阪府豊中市	2019年6月17日	2018年3月
大阪府吹田市	2019年6月17日	2019年3月
大阪府泉大津市	2017年12月20日	
大阪府高槻市	2017年12月20日	2020年4月
大阪府貝塚市	2019年3月11日	
大阪府守口市	2020年6月23日	
大阪府枚方市	2017年12月20日	2018年4月
大阪府茨木市	2017年12月20日	2018年2月
大阪府富田林市	2017年12月20日	
大阪府寝屋川市	2019年3月11日	
大阪府河内長野市	2018年4月30日	
大阪府大東市	2019年6月17日	
大阪府東大阪市	2018年3月19日	2018年4月
大阪府泉南市	2021年3月8日	
大阪府四條畷市	2020年6月29日	
大阪府岬町	2021年3月5日	
兵庫県神戸市	2017年12月20日	
兵庫県姫路市	2017年12月20日	2020年7月
兵庫県尼崎市	2017年12月20日	2019年3月
兵庫県明石市	2020年3月2日	
兵庫県西宮市	2017年12月20日	2019年3月
兵庫県芦屋市	2017年12月20日	2020年4月
兵庫県伊丹市	2021年2月5日	2019年5月
兵庫県豊岡市	2020年2月19日	
兵庫県加古川市	2017年12月20日	
兵庫県赤穂市	2018年12月16日	
兵庫県西脇市	2021年1月5日	
兵庫県宝塚市	2017年12月20日	
兵庫県三木市	2021年2月12日	
兵庫県高砂市	2018年12月16日	
兵庫県川西市	2017年12月20日	
兵庫県三田市	2017年12月20日	2020年4月
兵庫県加西市	2021年3月11日	
兵庫県養父市	2021年2月22日	
兵庫県丹波市	2020年6月2日	
兵庫県たつの市	2019年6月17日	
兵庫県多可町	2021年2月19日	
兵庫県稲美町	2021年2月5日	
兵庫県市川町	2021年2月5日	
兵庫県福崎町	2021年3月12日	
兵庫県香美町	2021年2月22日	
奈良県奈良市	2017年12月20日	2020年4月
奈良県大和高田市	2020年5月19日	
奈良県大和郡山市	2018年9月17日	
奈良県天理市	2017年12月20日	
奈良県橿原市	2020年3月13日	2018年6月
奈良県桜井市	2020年9月30日	
奈良県五條市	2020年5月18日	

表14 オープンデータ取組済自治体(13)

団体名	リスト掲載時点	官民データ計画
奈良県御所市	2020年3月31日	
奈良県生駒市	2017年12月20日	
奈良県葛城市	2017年12月20日	
奈良県宇陀市	2017年12月20日	
奈良県山添村	2020年1月15日	
奈良県三郷町	2020年3月31日	
奈良県斑鳩町	2020年4月22日	
奈良県安堵町	2019年9月17日	
奈良県三宅町	2018年4月30日	
奈良県田原本町	2019年9月17日	
奈良県曾爾村	2021年1月7日	
奈良県御杖村	2021年1月12日	
奈良県高取町	2019年12月16日	
奈良県明日香村	2020年9月25日	
奈良県上牧町	2020年3月31日	
奈良県王寺町	2020年7月17日	2019年3月
奈良県広陵町	2020年10月5日	
奈良県河合町	2020年10月26日	
奈良県吉野町	2020年6月2日	
奈良県大淀町	2021年2月5日	
奈良県下市町	2020年6月23日	
奈良県黒滝村	2020年2月27日	
奈良県天川村	2020年12月24日	
奈良県野迫川村	2020年1月15日	
奈良県下北山村	2020年6月2日	
奈良県川上村	2020年4月22日	
奈良県東吉野村	2019年12月5日	
和歌山県和歌山市	2019年6月17日	
和歌山県海南市	2017年12月20日	
和歌山県橋本市	2017年12月20日	
和歌山県有田市	2020年9月24日	
和歌山県御坊市	2020年2月28日	
和歌山県田辺市	2020年3月31日	
和歌山県紀の川市	2021年3月12日	
和歌山県有田川町	2020年7月16日	
和歌山県由良町	2020年9月17日	
和歌山県串本町	2021年1月22日	
鳥取県鳥取市	2018年3月19日	2020年3月
鳥取県米子市	2019年6月17日	
鳥取県境港市	2018年12月16日	
鳥取県北栄町	2020年11月30日	
鳥根県松江市	2017年12月20日	
鳥根県浜田市	2019年6月17日	
鳥根県出雲市	2019年6月17日	2021年3月
鳥根県益田市	2019年6月17日	
鳥根県大田市	2019年6月17日	
鳥根県安来市	2017年12月20日	2019年3月
鳥根県江津市	2019年6月17日	
鳥根県雲南市	2019年6月17日	2020年3月
鳥根県奥出雲町	2019年6月17日	
鳥根県飯南町	2019年6月17日	
鳥根県川本町	2019年6月17日	
鳥根県美郷町	2019年6月17日	
鳥根県邑南町	2019年6月17日	
鳥根県津和野町	2019年6月17日	
鳥根県吉賀町	2019年6月17日	
鳥根県海士町	2019年6月17日	
鳥根県西ノ島町	2019年6月17日	
鳥根県知夫村	2019年6月17日	
鳥根県隠岐の島町	2019年6月17日	
岡山県岡山市	2017年12月20日	2019年3月
岡山県倉敷市	2017年12月20日	
岡山県津山市	2017年12月20日	

表15 オープンデータ取組済自治体(14)

団体名	リスト掲載時点	官民データ計画
岡山県玉野市	2017年12月20日	
岡山県笠岡市	2017年12月20日	
岡山県井原市	2017年12月20日	2019年3月
岡山県総社市	2017年12月20日	
岡山県高梁市	2017年12月20日	
岡山県新見市	2017年12月20日	
岡山県瀬戸内市	2021年1月5日	
岡山県真庭市	2019年3月11日	
岡山県浅口市	2017年12月20日	
岡山県和気町	2019年6月17日	
岡山県早島町	2017年12月20日	
岡山県里庄町	2017年12月20日	
岡山県矢掛町	2017年12月20日	
岡山県鏡野町	2018年12月16日	
岡山県勝央町	2018年12月16日	
岡山県奈義町	2018年12月16日	
岡山県西粟倉村	2021年1月27日	
岡山県久米南町	2018年9月17日	
岡山県美咲町	2018年12月16日	
広島県広島市	2017年12月20日	
広島県呉市	2017年12月20日	
広島県福山市	2018年3月19日	2017年4月
広島県東広島市	2017年12月20日	
広島県安芸高田市	2020年10月9日	
広島県江田島市	2018年3月19日	
広島県府中町	2020年1月16日	
広島県熊野町	2020年10月13日	
広島県安芸太田町	2020年1月15日	
広島県北広島町	2019年6月17日	
山口県下関市	2019年12月2日	
山口県宇部市	2017年12月20日	2018年9月
山口県山口市	2017年12月20日	
山口県防府市	2017年12月20日	
山口県下松市	2020年7月31日	
山口県岩国市	2019年12月2日	
山口県光市	2018年9月17日	
山口県柳井市	2019年12月5日	
山口県美祿市	2021年2月12日	
山口県周南市	2017年12月20日	
山口県山陽小野田市	2018年4月30日	
山口県周防大島町	2021年1月14日	
山口県上関町	2019年6月17日	
山口県田布施町	2020年3月31日	
山口県平生町	2020年12月16日	
徳島県徳島市	2018年9月17日	2019年3月
徳島県鳴門市	2019年6月17日	
徳島県小松島市	2019年6月17日	
徳島県阿南市	2019年6月17日	
徳島県吉野川市	2019年9月17日	
徳島県美馬市	2019年6月17日	
徳島県三好市	2017年12月20日	
徳島県上勝町	2020年10月16日	
徳島県石井町	2020年2月18日	
徳島県那賀町	2019年9月17日	
徳島県牟岐町	2019年6月17日	
徳島県美波町	2020年1月20日	
徳島県海陽町	2020年7月27日	
徳島県松茂町	2020年3月31日	
徳島県北島町	2019年9月17日	
徳島県板野町	2020年1月16日	
徳島県上板町	2021年2月12日	
徳島県つるぎ町	2021年2月3日	
香川県高松市	2017年12月20日	2019年3月

表16 オープンデータ取組済自治体(15)

団体名	リスト掲載時点	官民データ計画
香川県丸亀市	2017年12月20日	
香川県坂出市	2020年10月9日	
香川県善通寺市	2021年3月5日	
香川県東かがわ市	2017年12月20日	
香川県三豊市	2021年3月10日	
香川県土庄町	2021年3月12日	
香川県小豆島町	2021年2月22日	
香川県宇多津町	2021年1月27日	
香川県多度津町	2021年3月4日	
愛媛県松山市	2017年12月20日	2019年3月
愛媛県今治市	2017年12月20日	
愛媛県宇和島市	2020年12月1日	2020年7月
愛媛県新居浜市	2018年3月19日	
愛媛県西条市	2021年1月21日	
愛媛県大洲市	2019年9月17日	
愛媛県伊予市	2019年3月11日	
愛媛県東温市	2019年3月11日	
愛媛県久高原町	2019年6月17日	2018年2月
愛媛県松前町	2019年3月11日	
愛媛県砥部町	2019年3月11日	
愛媛県内子町	2021年2月12日	
愛媛県伊方町	2020年11月19日	
高知県高知市	2021年3月12日	2020年3月
高知県宿毛市	2020年6月9日	
高知県いの町	2020年12月2日	
福岡県北九州市	2017年12月20日	2019年6月
福岡県福岡市	2017年12月20日	2019年6月
福岡県大牟田市	2020年2月26日	
福岡県久留米市	2017年12月20日	
福岡県直方市	2020年2月27日	
福岡県飯塚市	2019年6月17日	
福岡県田川市	2020年2月26日	
福岡県柳川市	2020年3月31日	
福岡県八女市	2020年4月22日	
福岡県筑後市	2021年2月18日	
福岡県大川市	2018年12月16日	
福岡県行橋市	2021年3月4日	
福岡県中間市	2020年3月31日	
福岡県小郡市	2017年12月20日	
福岡県筑紫野市	2019年6月17日	
福岡県春日市	2018年12月16日	
福岡県大野城市	2018年12月16日	
福岡県宗像市	2019年6月17日	
福岡県太宰府市	2018年12月16日	
福岡県古賀市	2018年12月16日	
福岡県福津市	2018年12月16日	
福岡県うきは市	2017年12月20日	
福岡県宮若市	2020年12月25日	
福岡県嘉麻市	2019年3月11日	
福岡県みやま市	2020年4月22日	
福岡県糸島市	2018年12月16日	2019年1月
福岡県那珂川市	2018年12月16日	
福岡県宇美町	2018年12月16日	
福岡県篠栗町	2018年12月16日	
福岡県志免町	2018年12月16日	
福岡県須恵町	2018年12月16日	
福岡県新宮町	2018年12月16日	
福岡県久山町	2018年12月16日	
福岡県粕屋町	2018年12月16日	
福岡県芦屋町	2020年6月4日	
福岡県水巻町	2020年10月6日	
福岡県岡垣町	2020年8月17日	
福岡県遠賀町	2020年2月28日	

表17 オープンデータ取組済自治体(18)

団体名	リスト掲載時点	官民データ計画
福岡県鞍手町	2020年2月26日	
福岡県筑前町	2021年2月3日	
福岡県大刀洗町	2017年12月20日	
福岡県大木町	2018年12月16日	
福岡県広川町	2020年3月31日	
福岡県香春町	2021年2月5日	
福岡県川崎町	2020年12月24日	
福岡県福智町	2021年3月2日	
福岡県苅田町	2020年10月14日	
福岡県みやこ町	2021年1月8日	
福岡県築上町	2021年3月9日	
佐賀県佐賀市	2019年6月17日	
佐賀県鳥栖市	2018年12月16日	
佐賀県多久市	2019年9月17日	
佐賀県伊万里市	2017年12月20日	
佐賀県小城市	2020年6月11日	
佐賀県玄海町	2021年3月3日	
佐賀県有田町	2019年9月17日	
長崎県長崎市	2020年3月18日	
長崎県佐世保市	2017年12月20日	2020年4月
長崎県大村市	2019年3月11日	2019年4月
長崎県平戸市	2020年6月23日	
長崎県松浦市	2020年7月16日	
長崎県西海市	2020年7月22日	
長崎県長与町	2020年6月29日	
長崎県時津町	2020年3月31日	
長崎県東彼杵町	2021年2月15日	
長崎県川棚町	2020年2月18日	
長崎県波佐見町	2020年8月26日	
長崎県小値賀町	2020年8月21日	
熊本県熊本市	2017年12月20日	2020年11月
熊本県荒尾市	2019年6月17日	
熊本県水俣市	2021年2月1日	
熊本県山鹿市	2019年6月17日	
熊本県菊池市	2017年12月20日	
熊本県宇土市	2021年3月3日	
熊本県宇城市	2021年2月3日	
熊本県阿蘇市	2021年3月4日	
熊本県大津町	2021年2月15日	
熊本県高森町	2021年2月3日	2020年2月
熊本県御船町	2020年1月16日	
熊本県水川町	2021年3月8日	
大分県大分市	2018年4月30日	2019年6月
大分県別府市	2019年9月17日	
大分県中津市	2019年9月17日	
大分県佐伯市	2021年2月18日	
大分県津久見市	2020年10月5日	
大分県宇佐市	2021年1月18日	2021年1月
大分県国東市	2021年3月5日	
大分県玖珠町	2021年2月22日	
宮崎県宮崎市	2020年3月31日	
宮崎県都城市	2020年1月9日	2020年4月
宮崎県延岡市	2021年1月28日	
宮崎県日南市	2017年12月20日	2021年2月
宮崎県日向市	2019年6月17日	
宮崎県串間市	2019年6月17日	
宮崎県えびの市	2020年11月18日	
宮崎県川南町	2020年7月13日	
宮崎県門川町	2021年3月5日	
宮崎県高千穂町	2019年9月17日	
鹿児島県鹿児島市	2017年12月20日	2018年3月
鹿児島県鹿屋市	2020年12月16日	2019年3月
鹿児島県垂水市	2020年2月3日	

表18 オープンデータ取組済自治体(17)

団体名	リスト掲載時点	官民データ計画
鹿児島県曾於市	2020年4月22日	
鹿児島県志布志市	2020年12月25日	
鹿児島県奄美市	2020年1月31日	2021年3月
鹿児島県始良市	2020年3月18日	
鹿児島県大崎町	2020年2月3日	
鹿児島県東串良町	2020年2月20日	
鹿児島県錦江町	2020年2月20日	
鹿児島県南大隅町	2020年2月20日	
鹿児島県肝付町	2020年2月20日	2020年3月
沖縄県那覇市	2019年6月17日	2019年4月
沖縄県浦添市	2017年12月20日	
沖縄県名護市	2020年11月9日	
沖縄県豊見城市	2017年12月20日	
沖縄県読谷村	2020年1月30日	
沖縄県北中城村	2021年3月4日	
沖縄県中城村	2021年3月4日	
沖縄県南風原町	2020年2月18日	

富良野市・秋田市・平塚市・秦野市・安城市の5市については、官民データ活用推進計画の策定作業中であることが判明したため、2021年4月として表中に記載している。

オープンデータ着手済の自治体のうち、官民データ活用推進計画が策定されていたのは157団体あった。オープンデータ着手済の自治体は1018団体あり、計画策定済の団体数は明らかに少ないと言える。

オープンデータ着手済の時点と計画策定の時点の前後関係を確認すると、オープンデータの着手が先行した事例は計画策定済157団体中126団体あった。

4 考察

以上で、オープンデータ着手済の自治体について、その着手時点と官民データ活用推進計画の策定時点を特定する作業を行った。

その結果、そもそも官民データ活用推進計画の策定が進んでいない現況が明らかとなった。データの公開と活用を促すこと

を企図して、2016年に官民データ活用推進基本法が施行され、同法に基づき基礎自治体については官民データ活用推進計画の策定が努力義務とされたものの、官民データ活用推進計画の策定は浸透していないことが明らかとなった。対して、オープンデータの取り組みは1000を超える自治体で進行している。つまるところ、法律の有無にかかわらず、オープンデータの取り組みが基礎自治体にあっては進行していたものと考えられる。

これは、オープンデータ着手済の自治体のうち官民データ活用推進計画を策定している157団体中の126団体においてオープンデータの取り組みが先行していたことから裏打ちされている。計画を策定し、その計画に基づいてオープンデータに着手するという事例もないわけではないが、それは少数であって、オープンデータの取り組みが先行していると言える。

都道府県においては、官民データ活用推進基本法で官民データ活用推進計画の策定が義務化され、それが大きな契機となってオープンデータに着手済という事態に至っている。対して、基礎自治体については、計画の策定は努力義務であり、必ずしも策定を行う必要はない。また、計画策定となれば、その検討のための作業が必要となるため、必ずしも官民データ活用推進計画の策定が浸透していないものと考えられる。そして、オープンデータ自体は当該団体の判断でいつからでも開始出来るため、必要と判断された時に、それが可能な状況にあれば着手に至っているものと考えられる。

5 結論と課題

本研究では、官民データ活用推進基本法の施行が自治体のオープンデータ政策に及ぼした影響の有無を検証した。具体的には、2021年3月段階でオープンデータ着手済の自治体について、その着手時点と官民データ活用推進基本法に基づく官民データ活用推進計画の策定時点を内閣官房情報通信技術(IT)総合戦略室公開のデータを参照しながら推定し、オープンデータ着手済の自治体にあつては、官民データ活用推進計画の策定が進んでおらず、また計画の策定がなされていても、それはオープンデータ着手後になされている傾向があることを確認した。

なお、本研究については、次のような研究上の課題が残されている。それは、本研究では、オープンデータに未着手の自治体を検証の対象に含めていないことである。オープンデータ未着手の自治体についても官民データ活用推進計画の策定状況について確認を行えば、計画策定が進む一方で、オープンデータの取り組みが進展していないという事態が明らかになる可能性がある。ただし、オープンデータ着手済の1018団体のうち、官民データ活用推進計画が策定済なのが157団体に留まっており、オープンデータ未着手の700余の自治体において計画策定が進んでいるとも考えにくい。いずれにしても、全自治体を対象に、官民データ活用推進計画の策定の有無を確認する作業が求められるところである。

また、オープンデータ着手済の自治体について、官民データ活用推進計画の策定が進んでいないと言っても、策定が努力義務である中で、150程度の団体が計画策定済

であった。この計画策定に至る事由についても別途検証を行う必要があると考える。

謝辞

本研究は公益財団法人セコム科学技術振興財団特定領域研究助成「民主制下における地方自治体の情報公開・オープンデータと情報セキュリティとの交錯に関する研究」における研究成果の一部である。

参考文献

- [1] 内閣官房情報通信技術(IT)総合戦略室：「地方公共団体におけるオープンデータの取組状況」(令和3年3月12日時点)
https://cio.go.jp/sites/default/files/uploads/documents/opendata_lg_rate_20210312.pptx (最終アクセス2021年4月11日, 以下のURLについては全て同様)
- [2] 本田正美, 梶川裕矢：「自治体におけるオープンデータ推進の政策過程」『情報文化科学研究』, vol.8, pp.1-9, 2018
- [3] 政府CIPポータル：「オープンデータ」,
<https://cio.go.jp/policy/opendata>
- [4] 内閣官房情報通信技術(IT)総合戦略室：「オープンデータ取組済自治体一覧とマップの過去データ」
https://cio.go.jp/sites/default/files/uploads/documents/opendata_lg_past.zip
- [5] 内閣官房情報通信技術(IT)総合戦略室：「オープンデータ取組済自治体一覧」, 2021
https://cio.go.jp/sites/default/files/uploads/documents/opendata_lg_manu_list.csv
- [6] 内閣官房情報通信技術(IT)総合戦略室：「計画策定済団体一覧」, 2019
https://cio.go.jp/sites/default/files/uploads/documents/sakutei_ichiran.pdf

第29回年次大会予稿

地域資料のデジタル化・オープン化の促進への一提案

-神奈川県事例から-

A proposal on the progress of digitization and openness of local

history materials - A case of Kanagawa Prefecture, Japan -

長塚隆

Takashi NAGATSUKA

鶴見大学

Tsurumi University

〒230-8501 横浜市鶴見区鶴見2-1-3

E-mail: nagatsuka-t@tsurumi-u.ac.jp

神奈川県全市町村を対象に市町村史、市町村勢要覧、広報誌等の地域資料のデジタル化とオープン化の現状について国立国会図書館デジタルコレクション、WARP、および神奈川県立図書館横断検索、各市ホームページや図書館OPACにより該当資料のメタデータおよび原資料のデジタル化とオープン化について調査し、当該地域資料のデジタル化とオープン化の現状を明らかにした。今後のデジタル化とオープン化をより一層促進するための方策を提案した。

The progress of digitization and openness of regional and municipal materials, such as history books, states and public relations magazine published by municipalities in Kanagawa prefecture has been estimated based on survey of metadata and digital materials. The survey was carried out by using search services such as the National Diet Library Digital Collections (NDL Digital), the NDL Web Archiving Project (NDL WARP), the cross search of public libraries in Kanagawa prefecture, municipal homepages and OPAC of public libraries. A proposal on the progress of digitization and openness of local history materials has been made.

キーワード: 地域資料, 市町村史, 市町村政概要, 市町村広報誌, デジタル化, オープン化

local materials, history books, state reports, public relations magazine, digitization, openness

1 はじめに

地域資料は地域レベルで過去に発生し現在まで保管されてきている資料と現在地域で

発生している資料の両方を含むとされる[1]。地域資料の範囲は広く図書、雑誌、新聞、地図、パンフレットのような印刷資料(行政刊行

物を含む)から、写真、フィルム、マイクロ資料、磁気あるいは光媒体に記録された資料、さらに古文書や記録、美術品、博物資料などの現物資料まで多様なものが含まれる。

これらの地域資料は各地の公共図書館の他に、歴史資料や現物資料は文書館・博物館・資料館、行政資料は行政情報センター・自治体史の編纂室、文学資料は文学館など多くの施設や機関に分散して収集されており、これらの組織が互いに補い合いながら地域資料の保存を担っていると言える[2]。

近年、郷土史・地域史や市政要覧・市報などを含む地域資料はデジタルで公開されることも多いが、具体的にどの程度デジタル化され、どのように公開され、どのような問題点があるのかは充分把握されていない[3-6]。

本研究では、神奈川県全市町村を対象に国立国会図書館サーチ(NDL Search)、デジタルコレクション(NDL Digital)、インターネット資料収集保存事業(NDL WARP)など全国的な検索ポータル、および神奈川県内横断検索など地域ポータル、当該地域の公共図書館OPACやデジタルアーカイブ、自治体ホームページを併用してメタデータおよびデジタル資料を調査し、デジタル化・オープン化の現状と今後の課題を提起した。

2 神奈川県の概況と地域資料

神奈川県の人口は東京に次いで912万人で全国2位で、19市6郡13町1村で構成され横浜市374万人、川崎市151万人、相模原市72万人と3市が政令指定都市である[7]。

図1は、今回調査の対象とした市町村史、市町村勢要覧、市町村広報誌について、市町村ごとに発行されている地域資料の発行開

始年を比較したものである。市町村史・市町村勢要覧の出版開始が最も早いのは横浜市中、川崎市や横須賀市などが比較的早かった。一方、市町村広報誌は戦後になり各市町村ともほぼ同時期に開始されており、市町村による差異は少ないことがわかる。

3 市町村史のデジタル化

神奈川県全市町村について、市町村史のデジタル化の現状を調査した(図2)。市町村史は各自治体の歴史や地域の現況をまとめたものであり、地域資料の中でも中心になる資料と言える。しかし、市町村史を自治体で独自にデジタル化し公開しているケースは非常に少ない。神奈川県では横浜市と湯河原町のみであった。

自治体の中でも比較的歴史が古く市町村史が早い時期に発行されている市町村では国立国会図書館によるデジタル化がなされているが、すべての市町村史がデジタル化され公開している自治体は鎌倉市、逗子市、松田町、湯河原町のみであった。

横浜市や川崎市のような大都市においても市史のデジタル化は80%、24%であった。比較的新しく市制が開始された伊勢原市、海老名市や多くの町は独自でのデジタル化は実施していないうえ、国立国会図書館によるデジタル化もされていない(図2)。

4 市町村勢要覧のデジタル化

最近では各自治体が発行する市町村勢要覧は一部の自治体では紙資料からデジタル資料やホームページへの掲載に変化しつつある。しかし、多くの自治体では依然紙資料のみのところが多い(図3)。市町村勢要覧は横

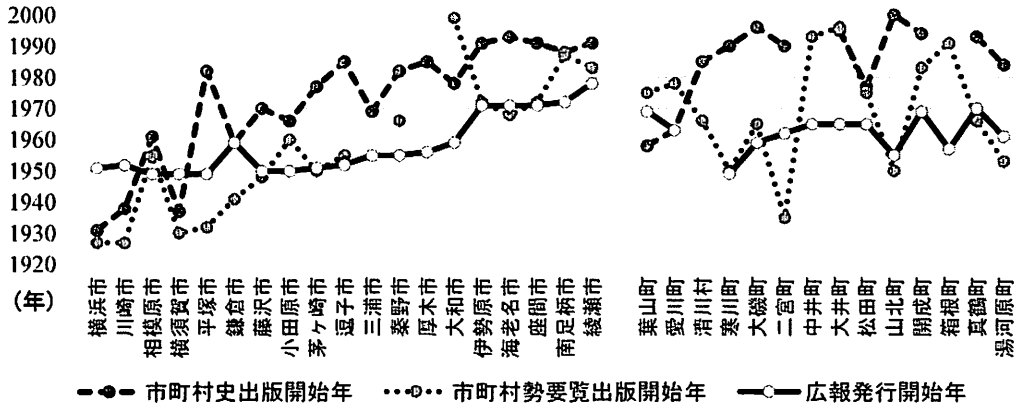


図1 市町村別の地域誌の発行開始年 (左:市, 右:町村)

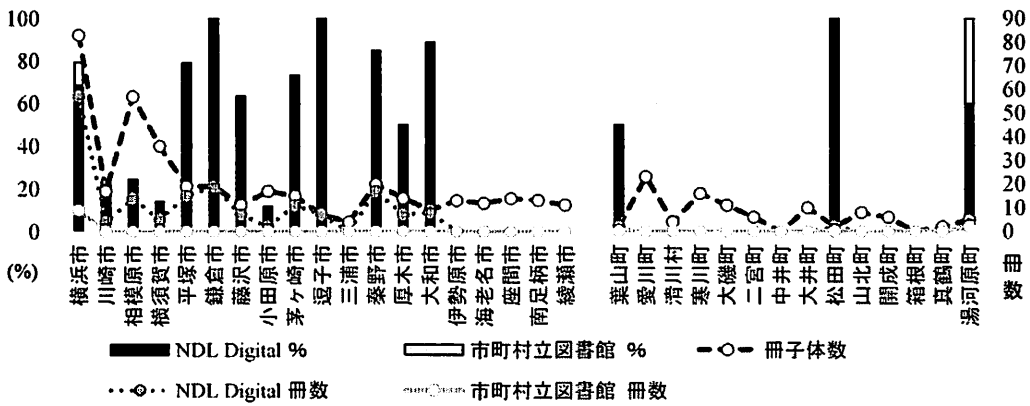


図2 神奈川県市町村史のデジタル化率 (左:市, 右:町村)

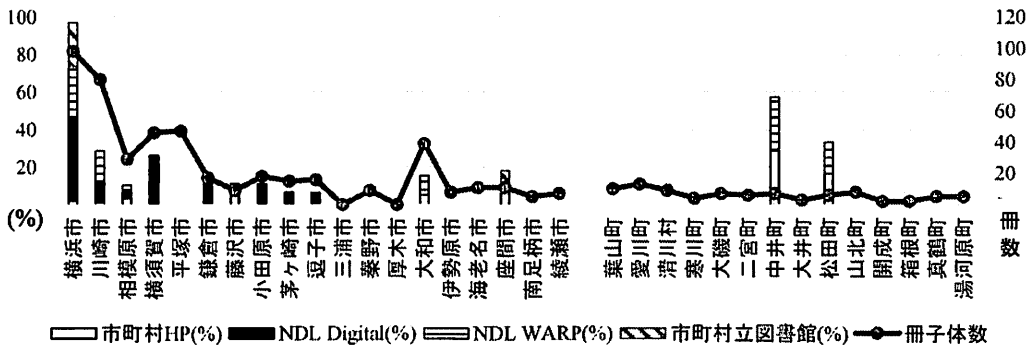


図3 神奈川県市町村勢要覧のデジタル化率 (左:市, 右:町村)

浜市や川崎市のような大都市では長期にわたり多くの刊行がされてきたが、刊行数が10に満たない町村が多く、市町村ごとに発行冊数が大きく異なっている。

最新の市町村勢要覧は自治体ホームページでPDF版が公開されることが増えているが、依然として印刷体資料のみ発行の自治体が多い。横浜市、川崎市、相模原市、藤沢市、大和市、座間市、中井町、松田町の6市2町はホームページでPDF版を公開しているが、現状ではデジタル版を公開していない自治体が多く今後の検討課題である。

5 広報誌のデジタル化

各市町村が毎月あるいは月に2回程度発行し、原則として全住民を対象に配布している広報誌は各地域での詳細な動きを経時的に知ることが出来る貴重な地域資料である。各市町村が発行する広報誌のデジタル化の現状を、図4に示した。

広報誌はほぼ70年前に発行が始まった戦後生まれの地域資料と言える。広報誌はほとんどの市町村でPDF版あるいは電子ブック版などにより、各自治体ホームページで公開されている。これらのホームページは10年以上にわたり、インターネット資料収集保存事業(NDL WARP)により收拾され公開されている。

横浜市、川崎市では広報誌デジタル版のホームページへの掲載期間がNDL WARPで閲覧可能な期間より短い(図4)。これはデジタル地域資料が自治体ホームページに掲載されるが、一定期間後にホームページから削除され利用できなくなることを示している。印刷体の広報誌は各地域の公共図書館が保存公開して

いるが、デジタル資料の保存公開は未だ担当機関が明確になっていないことを示している。

平塚市と小田原市では約70年前の創刊号からデジタル化しホームページに掲載している。しかし、このような自治体はごく少数で、ほとんどの自治体はデジタル化が20-30%と低いのが現状である。

6 地域資料オープン化の現状

図5は各市町村における広報誌のオープン化の状況である。インターネット資料収集保存事業(NDL WARP)が保存した資料は1:インターネット公開資料, 2:図書館送信限定, 3:国立国会図書館内で閲覧可の3種類に区分されて公開されている。3の資料は遠隔地の利用者にとって利用の際の大きな制約となっている。

多くの自治体はNDL WARPが収集した過去のホームページに掲載された広報誌を「1:インターネット公開資料」に指定し、広く閲覧できるようにしている。しかし、横浜市、横須賀市、平塚市、鎌倉市などは「3:国立国会図書館内で閲覧可」に指定し広報誌の閲覧を制限しており、今後の改善が必要である。

市町村史、市町村勢要覧、市町村広報誌のデジタル化とオープン化の進展度を図6に示した。市町村史と市町村勢要覧では市域でのデジタル化は40-50%であったが、町村域では10%未満と特にデジタル化が進展していないことが分かった。

これに対し、広報誌では市域と町村域でのデジタル化は、それぞれ30%程度であり大きな差異は認められなかった。

市町村史、市町村勢要覧、市町村広報誌の各地域資料により、オープン化の進

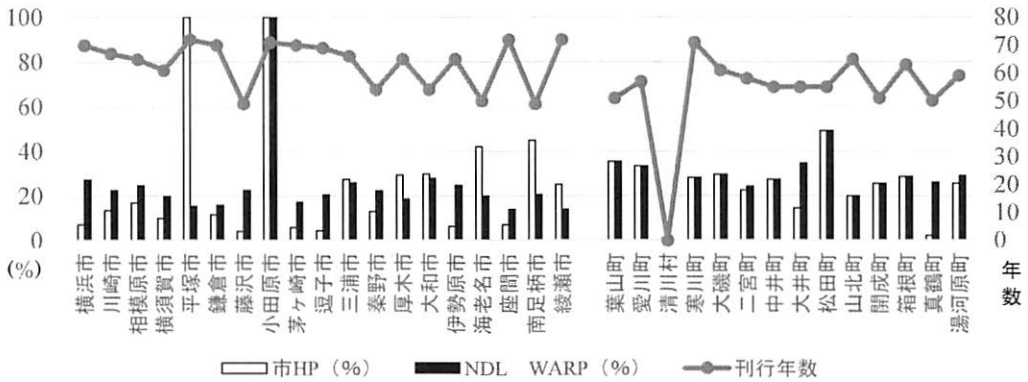


図4 神奈川県市町村広報誌のデジタル化率 (左:市, 右:町村)

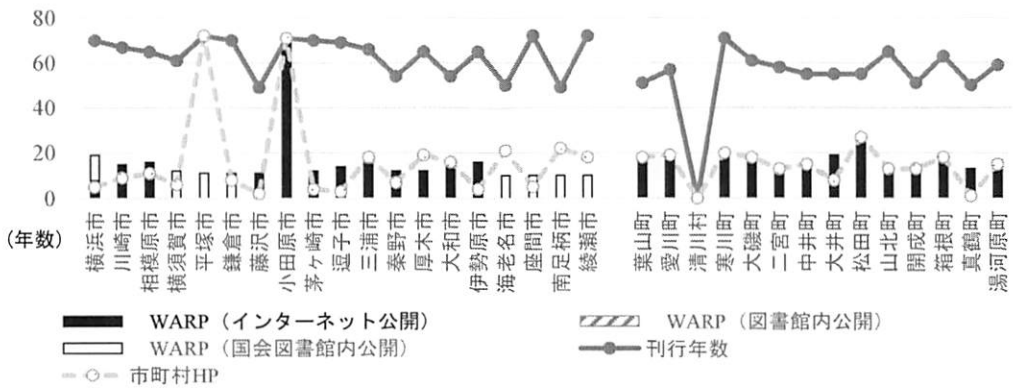


図5 神奈川県市町村広報誌のオープン化率 (左:市, 右:町村)

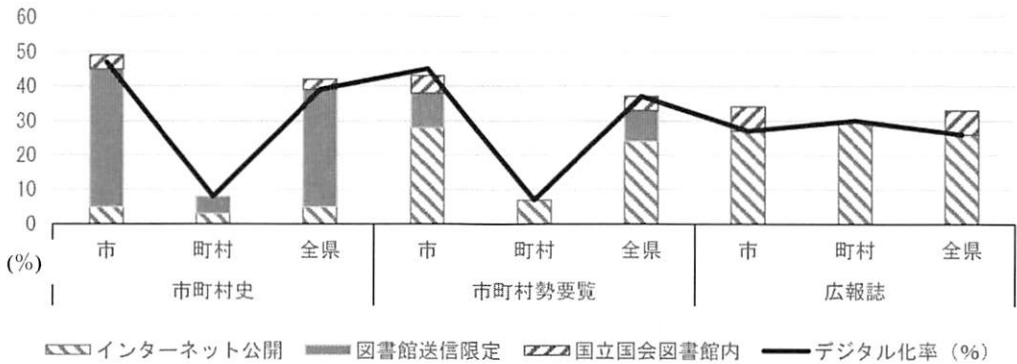


図6 神奈川県市町村地域資料のデジタル・オープン化率

注1 市史のデジタル化で横浜市立図書館と国立国会図書館で重複分は除いた。

注2 横浜市の広報誌は横浜市報を除き広報よこはまのみで集計した。

注3 広報誌のデジタル化率は各市町村のHPに掲載されたものが国立国会図書館WARPに収集されているので、重複分を除いた。ただし、地域のオープン化率の国会図書館内(7%)は除いていない。

展度は異なっており、市町村史では「2：図書館送信限定」に指定している自治体が多く見られた。これに対して、市町村勢要覧と市町村広報誌では多くの資料が「1：インターネット公開資料」に指定され、広く閲覧できるようになっていた。しかし、3種類の資料とも一部の自治体ではあるが「3：国立国会図書館内で閲覧可」に指定し広く利用して貰うべき地域資料の閲覧が制限されており今後の改善が必要となっている。

7 デジタル化・オープン化の課題

現状では政令指定都市の公共図書館でさえ地域資料のデジタル化への対応が充分ではなく、NDL DigitalやNDL WARPに多くを依存していた。また、NDL WARPが収集した資料の公開基準について自治体間で大きく相違しており、今後解決すべき課題であることが明らかになった。

今後、公共図書館がデジタル化した地域資料を保存公開できる全国規模のデジタル地域資料リポジトリが国立国会図書館等との連携により構築されることを期待したい。

参考文献

- [1] 国立国会図書館関西館図書館協力課編：「図書館調査研究レポート no. 9 - 地域資料に関する調査研究」, 国立国会図書館, 2008.
https://current.ndl.go.jp/files/report/no9/lis_rr_09_rev1.pdf
- [2] 地方史研究協議会編：「歴史資料の保存と地方史研究」, 岩田書院, 2009.
- [3] Nagatsuka, Takashi and Ishikawa, Tetsuya: "A Progress on the Digital

Collections of Local History at the Public Libraries and Other Memory Institutions in Japan", IFLA WLIC 2017 – Wrocław, Poland – Libraries. Solidarity. Society, 2017.

<http://library.ifla.org/id/eprint/1702>

[4] Nagatsuka, Takashi: "The Relationships and Linkages among the Digital Collections at Libraries in Japan. Paper presented at: IFLA WLIC 2018 – Kuala Lumpur, Malaysia – Transform Libraries, Transform Societies", 2018. <http://library.ifla.org/id/eprint/2341>

[5] Nagatsuka, Takashi: "Digital Collaboration: Linkages among Local History Collections in Japan", IFLA WLIC 2019 - Athens, Greece – Libraries, 2019.

<http://library.ifla.org/id/eprint/2692>

[6] 長塚 隆, 地域資料のデジタル化の進展をNDL SearchやJapan Searchなどの検索ポータルから推測, 情報知識学会誌, 2019, 29巻, 4号, p. 340-343.

https://doi.org/10.2964/jsik_2019_049

[7] 県勢要覧2019 (令和元年度版), 神奈川県統計センター統計管理課. 令和2年3月発行. <https://www.pref.kanagawa.jp/docs/x6z/tc10/yoran.html>

[8] 長塚 隆, 自治体史等の地域資料のデジタル化・オープン化の進展状況 -神奈川県政令指定都市の事例から-, 情報知識学会誌, 2020, 30巻, 2号, p. 155-162.

https://doi.org/10.2964/jsik_2020_012

[9] 長塚 隆, 自治体史等の地域資料のデジタル化・オープン化の進展状況 -神奈川県全市の事例から-, 情報知識学会誌, 2020, 30巻, 4号, p. 447-454.

https://doi.org/10.2964/jsik_2021_006

(以上のurlは全て2021年4月5日参照)

第29回年次大会予稿

Webシステム上のAIの学習システムを使った授業の実践

Practice of lessons using the AI learning system of the Web

山島一浩

Kazuhiro YAMASHIMA*

筑波学院大学 経営情報学部

Tsukuba Gakuin University Faculty of Business Design and Informatics

〒305-0031 茨城県つくば市吾妻3-1

E-mail: mt-is@tsukuba-g.ac.jp

*連絡先著者 Corresponding Author

AIの学習システムが画像の分類や話し言葉の理解などの有用なタスクを達成できるようになった。Webシステムで、AI計算を実行できるようになった。AIシステムに画像の認識を教えることは重要である。そこで3学年の学生にテストとトレーニングをやらせてみた。最初にトレーニング例を提供して、特定の画像がどのように見えるかを学習できるようにする。学生は、実例をやってみて、興味をもち、自分のAIシステムを組みだした。

AI learning systems are now able to accomplish useful tasks such as image classification and spoken language comprehension. You can now perform AI calculations on your web system. It is important to teach AI systems how to recognize images. I asked the third grade students to do the test and training. First, provide a training example so that you can learn what a particular image looks like. The students tried the examples, became interested, and built their own AI system.

キーワード: AI学習システム, Webシステム, インストラクショナル・デザイン

AI learning systems , Web System, Instructional Design

1 はじめに

教育の場でのインストラクショナル・デザインは、学習者の自由度を保ったままで高い学習効果が生じることを意図して、具体的な計

画を立てることである。教育活動の効果・効率・魅力を高めるための手法を集大成したモデルや研究分野、またはそれらを応用して学習支援環境を実現するプロセスである。しかし

ながら、学ぶ内容が最近のものは、それらが不足しているのが現状である。

例えばAIシステムがある。AIシステムは、画像の分類や話し言葉の理解などに有用なタスクを達成できるようになっている。Webシステムにおいても、AIを実行できるようになった。

そこに、教材の仕様を固める作業が必要である。

2 AIをテーマにした授業研究

2.1 AIの進化

AIが、推論、認識、判断など、人間と同じ知的な処理能力を持つ情報処理システムであることは、衆目の一致するところである。AIが、画像の分類や話し言葉の理解などの便利なタスクを実行できるようになったのは、2000年以降のことである。

そして、それ以降、機械学習が進歩し、学生が利用できるスマートフォンやタブレットで重要なAI計算を実行できるようになった。

2.2 AIの学習システム

ここでは、TeachableMachineを利用する。これは、機械学習モデルを簡単に作成し、Webベースでアクセスできる。表1に機能を示す。

表 1 TeachableMachine の機能

収集する	コンピュータに認識させたいクラスまたはカテゴリごとに、サンプルの画像を収集する。
トレーニングする	モデルをトレーニングした後、すぐにテストして、新しいサンプルを正しく分類できているかどうかを確認する。
エクスポートする	サイトにモデルをエクスポートする。エクスポートしたモデルはダウンロードできる。

TeachableMachineには、チュートリアルがあり、バナナの熟れぐあいを分類として、それを読むだけでこの利用学習ができる。

2.3 MITのインストラクショナル・デザイン

この授業では、MIT App Inventorのカリキュラムを参考にした。

MITのサイトでは、チュートリアルレッスンのほか、学生の探索やプロジェクト作業の提案が含まれている。そして、補助教材として、TeachableMachineを使っている。

2.4 MIT教材の構成

学生は機械学習の基本について学び、画像分類を通じてこれらの概念を実装する独自のアプリを作成する。学生はモバイルデバイスで写真を撮り、アプリはそれらの写真内のオブジェクトを識別する。各分類には信頼水準がある。これは、アプリがその分類にどれだけ自信を持っているかを示す値である。

表 2 MIT のレッスン

時間	アクティビティ
10	ユニットの概要 機械学習とは何か、どのように使用されるかについて話し合います。
25	Teachable Machineで遊ぶ 生徒はWebにアクセスし、Teachable Machineを使用して、機械学習がどのように機能するかについての基本的な理解を得ます。
10	まとめディスカッション データがどのように収集されるか、情報をどの程度使用できるか、機械学習についての考えについて話し合います。

3 インストラクショナル・デザイン

そこで大学3学年の学生に、AIを用いた授業を計画し、AIのテストとトレーニングをやらせてみた。

最初にトレーニング例を提供して、特定の画像がどのように見えるかを学習できるようにする。

表3が、90分間の授業時間計画である。

表 3 授業計画

20分	トレーニング例を提示
30分	学生が、実際に組む
10分	学生が、確認する
30分	学生が自分の着想で組んでみる

この授業は、通信で行い、学生の姿を目にすることはできない。全て Google classroom で掲示を行い、作業は、すべて学生個人で行った。

図1に、授業内容を示した Google classroom の画面を示す。



図 1 授業内容

図2にテキストの一部を示す。この内容は、Google Meet を使って説明をした。

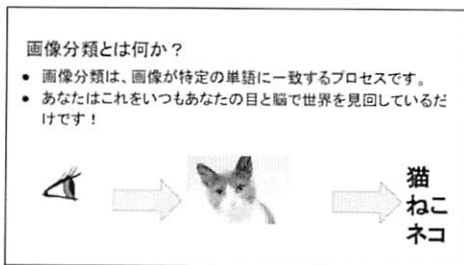


図 2 授業のテキストの一部

図3に、授業で用いたサンプルを示す。授業では、まず、私がこの写真を使って、TeachableMachine への登録を説明する(20分)。そのあと、学生が、この写真を使って

登録をする(30分)。その結果で、確認する(10分)。残りの30分で、学生が自分の着想で組んでみる。



図 3 授業のサンプル

4 授業結果

学生は、実例をやってみて、興味をもち、自分のAIシステムを組みだした。13人の受講者について、練習問題をこなし、自分の課題について取り組んだ。

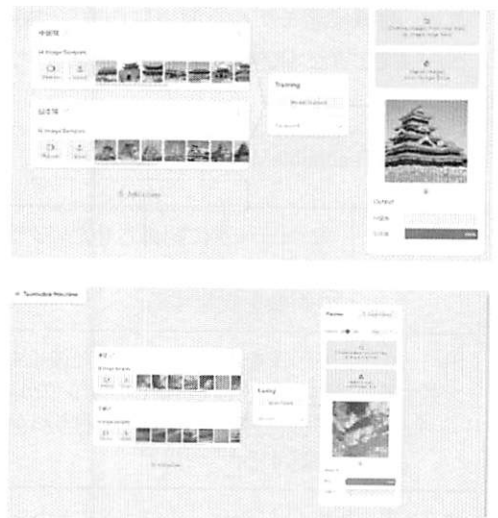


図 4 学生の出力した例

6人が、課題の応用を行った。そして、7人が、自分の関心のあるテーマに取り組んだ。その後、学生は、実例をやってみて、興味をもち、作業は終わった。

レポートでは、応用編として、枚数を多くする、テーマとして分類できるものをまとめるなどと、自分の感想を加えたものが出てきた。

この授業の結果として、卒業研究のテーマとして、取り組んだのは2名であった。

5 考察

最初にトレーニング例を提供して、特定の画像がどのように見えるかを学習できるようにしたのは、興味を引き立てるにはよいと思われる。

学生は、Google classroom の授業サイトに戻って、アプリの成功に基づいて、ミスからのモデルを更新できるはずであった。

自分のAIシステムを精度と制限に基づいて、アプリがどのように機能したかについて話し合い、モデルを改善する方法について話し合うなど、もう少し時間をとってあげるべきではなかったかと思う。

授業評価としては、実情と理想の差異を認識し、最終効果を決定することが、なんであったのかを明らかにする手順はできたと思われるが、その先の卒業論文までへの問題点の発見という到達には至らなかった。

5 おわりに

AIシステムが画像の分類や話し言葉の理解などの有用なタスクを達成できるようになったことから、Webシステムで、3学年の学生にテストとトレーニングをやらせてみた。学生からは、仕組みには、納得した。

短時間に説明を行え、実習ができ、今後の授業に入れ込むことができればと思う。

その際には、画像の枚数、学生の各自の枚数など、授業計画に加えた方がよいと反省も残った。

参考文献

- [1] 鈴木克明：e-Learning実践のためのインストラクショナル・デザイン，日本教育工学会論文誌 29 (3)，197-205，2005.
- [2] Barron Webster：Teachable Machine Tutorial，<https://medium.com/@warronwebster/teachable-machine-tutorial-banabnameter-4bffffa765866>，(2021.4現在)。
- [3] Introduction to Machine Learning：Image Classification，<https://appinventor.mit.edu/explore/resources/ai/image-classification-look-extension>，(2021.4現在)。

第29回年次大会予稿

弱視生徒を対象とした授業におけるノートのデジタル化

Digitization of Notebooks in Class for Students with Low Vision

小川雄太^{1*}, 宮本行庸²

Yuta OGAWA^{1*}, Yukinobu MIYAMOTO²

1 兵庫県立視覚特別支援学校

Special Needs Education School for the Visually Impaired in Hyogo Prefecture

〒655-0884 兵庫県神戸市垂水区城が山4-2-1

E-mail: ogaway123@gmail.com

2 神戸学院大学 経営学部

The Faculty of Business Administration, Kobe Gakuin University

〒650-8586 神戸市中央区港島1-1-3

E-mail: miyamo@ba.kobegakuin.ac.jp

*連絡先著者 Corresponding Author

弱視生徒は、アクセシビリティの設定によりICTを有効に活用している。特にiPad等のタブレット端末は、弱視生徒の学習環境を飛躍的に高めるツールの一つであると捉えられる。そこで、本研究においては、弱視生徒の日常におけるICT利用状況を踏まえ、iPadによるノートのデジタル化を図った授業実践に取り組んだ。その結果、iPadによるノートのデジタル化は、内容理解の促進、家庭学習の促進等において有効であった。

Students with low vision make effective use of ICT by setting accessibility. In particular, tablet terminals such as iPads are regarded as one of the tools that dramatically enhance the learning environment for students with low vision. Therefore, in this study, based on the daily ICT utilization situation of students with low vision, we practiced lessons by digitizing notebooks using iPad. As a result, digitization of notebooks using the iPad was effective in promoting content understanding and home learning.

キーワード: 視覚障害, 弱視, ICT教育, iPad, タブレット端末

Keywords: Visually impaired, Low Vision, ICT education, iPad, Tablet device

1 はじめに

視覚障害は、情報の障害ともいわれ、情報を保証するため、教育におけるICTの果たす役割は極めて大きい[1]。視覚障害には、視覚をほとんど利用できていない盲と、制限はあるが視覚を利用することができる弱視がある[2]。そして、弱視教育においては、視覚補助具の活用が学校教育における重要な検討課題となっており、近年はiPadをはじめとするタブレット端末の活用に注目が集まっている[2]。

タブレット端末等のICTの活用は、障害のある子どもたちの可能性を広げ、これからのインクルーシブ教育システムの推進において、合理的配慮を進めるために大きな役割を果たすことができる[3]。国レベルでも一人一台の学習用端末と、高速大容量の通信ネットワークを一体的に整備するというGIGAスクール構想が2019年に閣議決定され、ICTの学校への普及が進みつつある。また、「学校教育法等の一部を改正する法律」（2018年法律第39号）の関係法令が施行され、紙媒体の教科書を活用して学習することが困難な場合、一定の基準の下でデジタル教科書を活用できるようになった。弱視生徒は、デジタル教科書の文字拡大や音声読み上げ機能等によって、教科書へのアクセスが向上した。

元々、紙媒体の拡大教科書における拡大された文字であっても読みづらいことも多かったため、弱視生徒にとって、拡大鏡や拡大読書器等は必須の補助具であった。拡大鏡はそれぞれが所持し、持ち運ぶことも簡単ではあるが、拡大読書器はどこへでも持ち運べるものではない。また、全ての弱視生徒の自宅に設置されているとも限らず、学校以外のどこにでも設置されてい

るものではない。一方で、視覚障害者用の紙媒体の教科書は、文字や図版を拡大しているため、普通校で使用している一冊の教科書が複数冊で構成されている。そのため、持ち運びが不便であるという可搬性の問題やページや段落を探すことが困難であるという操作性の問題がある[4]。また、点字版の教科書であっても複数冊で構成されており、持ち運びや管理に関して、同様の負担が強いられている。

これら紙媒体の教科書のデメリットは、タブレット端末一台に全ての教科書を保存することが可能なデジタル教科書の登場によって解消されたといえる。紙媒体の教科書には「紙の良さ」があり、ページをめくることで、どこに何が書いてあるかという情報と記載されている内容が関連付けられるために内容が頭に入ってくるともいわれる。このような紙媒体の教科書の持つメリットとタブレット端末を活用したデジタル教科書のメリットを比較した上で、メリットの大きい方を選択することとなる。個人個人でどのように感じるかは大きく異なるものの、国レベルでタブレット端末によるデジタル教科書の普及が進められている現状を鑑みると、タブレット端末によるデジタル教科書の持つメリットは大きいといえ、弱視生徒の学習環境を飛躍的に高めるツールであると捉えられる。そして、盲学校を中心に、弱視生徒に対して、デジタル教科書が広がりつつある。しかしながら、教科書以外の教材等は、必ずしも弱視生徒にとってアクセシブルになっていない[4]。つまり、教科書以外の教材等はデジタル化がそれほど進んでいない状況にある。教科書以外の教材等の中でも、教科書とセットで学習に不可欠なもの

としてノートがある。ノートは、生徒が自分自身で作上げるものである。したがって、生徒と相談した上、教師の工夫次第で簡単にデジタル化が可能であるものの、個人の工夫によるものであるからこそ、ノートのデジタル化は大々的に進められてはいないといえる。ノートも教科別、科目別に用意すると、相当の冊数になり、持ち運びの負担や管理にも注意を要する。また、弱視生徒がノートに文字を書くには、紙媒体の教科書の文字を読むのと同様に、拡大鏡や拡大読書器等の補助具が必要である。この点に関して、デジタル教科書に倣って、iPadの活用によるデジタル化を進めることで、弱視生徒の感じている負担を減らし、学習環境を向上させることができると考えられる。ただし、学習に対するモチベーションを下げないように、視覚障害の特性に応じてICTの操作等の指導を丁寧に行うことに留意する必要がある[5]。

2 先行研究

一般的な授業におけるICT活用については数多くの先行研究の蓄積がある。ここでは、本研究の目的に照らし、弱視生徒に対するタブレット端末を活用した授業をテーマにしている研究について概観する。

小林ら(2020)[6]は、最新のICTの活用により、視覚障害ゆえに埋もれていた能力を引き出す必要性を指摘している。最新のICTの一例としてタブレット端末が挙げられる。タブレット端末の中でもiPadには、視覚障害者のアクセシビリティを補助するための手段として、音声読み上げ機能(VoiceOver)、フォントサイズの拡大、画面の拡大・縮小、背景色と文字色の反転等の機能が標準で装備されている[7]。

実際の視覚障害者のICT利用状況について、渡辺(2017)[8]は、パソコン利用率は全盲96.7%、弱視88.9%、タブレット端末利用率は全盲14.4%、弱視38.9%、スマートフォン利用率は全盲52.1%、弱視55.6%であることを報告している。年代別の利用率は明らかとはなっていないものの、10代から80代までを対象者としているため、このように、タブレット端末の利用率が低い結果になっているものと考えられる。また、ICTの利用に関する講習会の開催が都市部に集中していると渡辺が指摘していることも、特にタブレット端末の利用が進んでいない背景にあると認識できる。しかしながら、盲学校在学中の生徒は、学校でiPad等のタブレット端末をはじめとするICTの活用について、教員から直接指導を受ける機会も多くあり、渡辺が報告している利用率以上にiPadの活用が進んでいると感ぜられる。

現に、デジタル教科書の普及はさることながら、文部科学省による「学習上の支援機器等教材活用促進事業」の中で、iPad向けの教科書閲覧アプリ「UDブラウザ」が開発されている[9]。UDブラウザはiPad専用のアプリとして開発が進められているところをみても、視覚障害者にとってiPadは学習における最適なツールであると認識できる。教科書等のデジタル化を進めた授業に関する研究としては、中野ら(2014)、小倉(2014)等がある。

中野ら(2014)[10]は、iPadにインストールしたPDFファイルのデジタル教科書と紙媒体の教科書をどちらも使用できる状態の授業を半年間試み、生徒および指導する教員に対する調査を行った。その結果、iPadのデジタル教科書を今後も使いたい

という生徒が多かったことや一人ひとりのニーズに応じて拡大率を変えたり、迅速に白黒反転等ができたりするため、紙媒体の拡大教科書と比べ、学習意欲が向上していると感じている教員が多かったこと、調べ学習や家庭学習等に積極的に取り組む場面が増えたという教育効果のあったことを報告している。また、小倉(2014)[11]は、iPadにインストールしたデジタル教科書を用いた授業実践から、iPad活用の利点として、見え方に応じた表示が可能である点、直感的に操作でき利用しやすい点、複数の教科書・プリントを集約できる点等を指摘している。

以上のとおり、先行研究において、教科書や配布資料のデジタル化が弱視教育において非常に有効であることが明らかにされている。これらの先行研究における知見を踏まえると、ノートのデジタル化も有効な手段となり、弱視生徒の学習環境の向上に資することができると考えられる。そこで、本研究においては、授業実践を通して、弱視生徒に対するiPadを活用したノートのデジタル化の有効性について検討を行うこととする。

3 iPadによるノートのデジタル化

従来の授業においては、紙媒体の教科書とともに紙媒体のノートが使われており、弱視生徒にとっては、持ち運びの問題や見えづらさによる管理の問題が生じていた。そのため、デジタル教科書とともにノートのデジタル化を進めることで、弱視生徒の学習環境を飛躍的に向上させることができるといえる。具体的にはiPadのApple Pencilの使用によりPDFファイルに記入す

る方策が考えられる。Apple Pencil(第1世代)はiPad(第6世代、第7世代、第8世代)、iPad Air(第3世代)、iPad mini(第5世代)等で利用できる[12]。

物理的に必要な準備物としては、教員用と生徒用のiPad、生徒用のApple Pencilである。それに加えて、Word等で作成された空欄穴埋め方式の授業プリントを交換したPDFファイルが必要である。教員は空欄穴埋め方式の授業プリントをPCで作成し、PC上でPDFファイル化することもできる一方、iPadのWordアプリで授業プリントを作成しPDFファイル化することも可能である。いずれにしても、生徒に配布するため、事前に教員用のiPadにPDFファイルを保存しておかなければならない。そのため、PCで作成した場合はクラウド経由で送る方法、PCと接続してファイル転送する方法、メールに添付して送信する方法等によって、あらかじめiPad内に保存しておく必要がある。このような操作に慣れているのであれば、iPadへの保存は問題にならないと思われるが、これらの手順に教員が苦手意識を持っている場合には、はじめからiPadで作成の方がシームレスに授業準備を進めることができる。また、本実践においては、ノートをデジタル化するため紙媒体を配布する必要はなく、テストも同様にデジタル化して実施するため、基本的にはプリンターに接続して紙媒体として印刷する必要はない。したがって、iPad内で作業を完結させても特段の支障は生じないともいえる。iPadに保存しているPDFファイルは、特別な環境設定をせずとも、iPad同士でファイルのやり取りが可能な手段であるAirDrop[13]によって行う。生徒は受け取ったPDFファイルをアプリで開

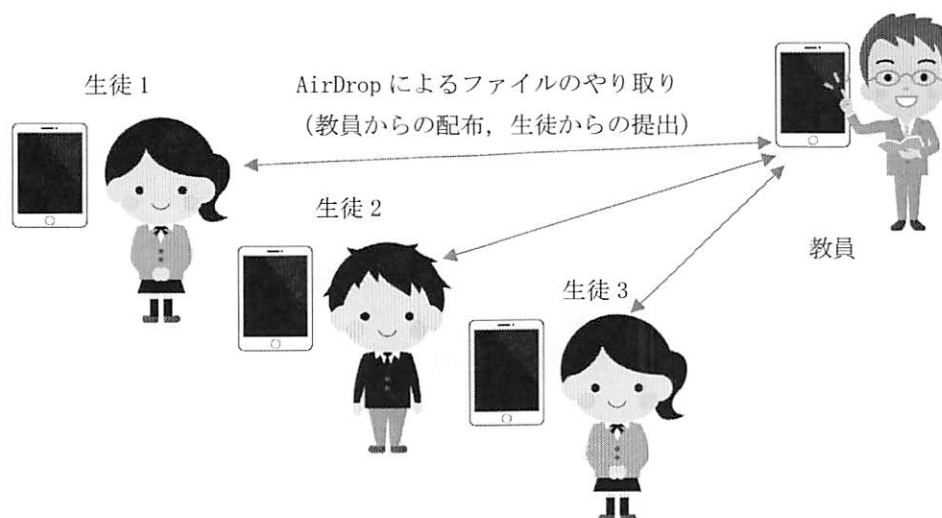


図1 授業におけるICT活用のイメージ

き、教員の説明を聞き取りながら、紙媒体のノートと同じように Apple Pencil を使って記入していくこととなる。

4 実践

4.1 対象

盲学校（視覚特別支援学校）高等部普通科において、2学年3名（弱視）を対象に、iPadによるノートのデジタル化を図った授業を2020年6月から2021年3月まで実施した。

4.2 概要

4.2.1 授業

図1に示した授業におけるICT活用のイメージのとおり、地理歴史科の地理Aにおいて授業実践を行った。教員および生徒はそれぞれApple Pencilに対応したiPadを所有している。生徒は3名とも各自の見え方に応じたアクセシビリティの設定を行った上、日常生活においてiPadを自分自身の

力で活用できている。このような生徒の実情を踏まえ、第3章に述べた方法により授業実践を行った。

4.2.2 定期テスト

定期テストにおいても授業と同様に、教員からPDFファイルを配布し、生徒がiPadを使ってApple Pencilで記入する方式で実施した。また、生徒が記入した答案用紙についてもAirDropによる提出を求めた。

4.3 ICTの活用に関する意識の把握および実践の評価

ICTの活用状況や授業におけるiPad活用の意識を把握するため、12月に質問紙調査および面接調査を実施した。生徒が3名という少人数であるため、お互いの意見交流の場として、質問項目に関する意見を自由に出してもらおう中で、教員が口頭で追加の質問を行った。また、12月の調査を踏まえて、翌1月にiPadのアクセシビリティの設定に関する面接調査を実施した。

表 1 ICT の活用レベル

	PC活用レベル		タブレット活用レベル			スマホ活用レベル		
	Windows	Mac	Windows	iPad	Android	Windows	iPhone	Android
平均	2.33	2.00	1.00	3.67	1.00	1.00	2.33	2.00
SD	0.47	1.41	0.00	0.47	0.00	0.00	1.25	1.41

n=3 (4件法)

表 2 iPad によるノートのデジタル化の効果

	内容理解 の促進	家庭学習 の促進	合理的配慮 の提供	テスト解答時 の有用性
平均	4.00	3.00	4.00	3.67
SD	0.00	0.82	0.00	0.47

n=3 (4件法)

4.4 質問紙の調査項目

4.4.1 ICTの活用レベル

ICTの活用レベルについて、以下のような項目を質問した。

- WindowsとMacそれぞれのパソコンの活用レベル (4件法) 等。
- Windowsタブレット, iPad, Androidタブレットの活用レベル (4件法) 等。
- Windowsスマホ, iPhone, Androidスマホの活用レベル (4件法) 等。

4.4.2 iPadによるノートのデジタル化の効果

iPadによるノートのデジタル化が以下のそれぞれの項目にどれくらい影響したかについて質問した。

- 内容理解の促進 (4件法) 。
どの程度、授業の内容についての理解を促進したかについて質問した。
- 家庭学習の促進 (4件法) 。
どの程度、授業以外の家庭等での学習を促進したかについて質問した。
- 合理的配慮の提供 (4件法) 。
どの程度、合理的配慮の提供といえる

と感じたかについて質問した。

- テスト解答時の有用性 (4件法) 。

どの程度、テストの解答時において有用であると感じたかについて質問した。

4.5 面接の調査項目

質問紙調査を踏まえ、さらに詳細を把握することを目的として、以下の項目について面接調査を行った。

- 日常生活におけるICT活用。
- iPadのアクセシビリティ設定。
- iPadによるノートのデジタル化の感想。
- Apple Pencilの活用の感想。

5 結果

5.1 ICTの活用レベル

ICTの活用レベルに関する調査結果を表1に示す。パソコンに関して、Windowsを持っているのは1名であり、活用レベルは平均:2.33, SD:0.47であった。Macを持っているのは1名であり、活用レベルは平均:2.00, SD:1.41であった。タブレット端末に関して、Windowsは誰も持っておらず、

活用レベルは平均:1.00, SD:0.00であった。iPadを持っているのは3名であり、活用レベルは平均:3.67, SD:0.47であった。Androidは誰も持っておらず、活用レベルは平均:1.00, SD:0.00であった。スマホに関して、Windowsは誰も持っておらず、活用レベルは平均:1.00, SD:0.00であった。iPhoneを持っているのは3名であり、活用レベルは平均:2.33, SD:1.25であった。Androidは誰も持っておらず、活用レベルは平均:2.00, SD:1.41であった。

5.2 iPadによるノートのデジタル化の効果

iPadによるノートのデジタル化の効果に関する調査結果を表2に示す。内容理解の促進は、平均:4.00, SD:0.00であった。家庭学習の促進は、平均:3.00, SD:0.82であった。合理的配慮の提供は、平均:4.00, SD:0.00であった。テスト解答時の有用性は、平均:3.67, SD:0.47であった。

5.3 日常生活におけるICT活用

日常生活におけるICTの具体的な活用について、「iPadの活用について、自分に必要なことはだいたいできる。」「タブレットやスマホは絵を描くときや調べるときに使っている。」「パソコンはZoomで使うぐらい。」「連絡手段、乗り換え検索、動画視聴、拡大鏡、ネット通販等を使っている。」との回答が寄せられた。加えて、ICTの活用に関する困難さについて、「操作している際に（知らない間に指が他のページに飛ぶ部分に触れてしまい）、ページが飛んでしまう。」「容量がどんどん増えていってしまう。」との回答が寄せられた。また、ICT全般について、「iPadを日常的に使っているため、パソコン自体は、ほとんど使っ

ていない。」、「大学進学にあたって、どのパソコン（WindowsかMacか）を購入すべきか迷っている。」との回答が寄せられた。

5.4 iPadのアクセシビリティ設定.

iPad活用のために必要なアクセシビリティの設定に関して、次のような回答が寄せられた。まず、「画面表示と明るさ」の「外観モード」を「ライト」から「ダーク」に変更していた。これは、弱視生徒にとって「ライト」では画面全体が明るすぎることから、明るさを抑えるため「ダーク」に設定するものである。次に、「アクセシビリティ」の「画面表示とテキストサイズ」の中の「さらに大きな文字」をオンにして、文字サイズを拡大するとともに、文字を太くしていた。しかし、文字サイズを大きくしても見えづらい場合もあり、また、アプリのアイコン等も見えづらいため、「アクセシビリティ」の「ズーム」をオンにし、「ズーム領域」を「フルスクリーンズーム」に設定していた。これは、画面全体がいわば拡大鏡になっているかのように、タッチ操作で動かしていくことで、目的の場所が画面全体に拡大されるものである。以上は、いずれもiPadの設定アプリから操作を行うものである。その他にもホーム画面にある「コントロールセンター」に「反転」「テキストサイズ」のショートカットキーの設定をしていた。

5.5 iPadによるノートのデジタル化の感想

iPadによるノートのデジタル化について、「（地理A以外の）他の授業でも使えたらいいなと思う。」「数学は（教科書に問題が掲載されているため）拡大読書器を使って問題文を確認した上、（拡大読書器

を使って) ノートを書いている。」「どの教科でも提出するものを (iPadのAirDropによって) 電子データで渡したい。」「紙媒体のプリントよりPDFファイルの方が管理しやすい。」「他の授業では (授業で書いた) ノートの内容を (授業以外の時間を使って) iPadに書き込んでいる。」「テスト勉強のとき, 他の教科は (授業で書いた) ノートをiPadによって写真を撮っている。(容量の都合で) テストが終わったら削除している。」との回答が寄せられた。

5.6 Apple Pencilの活用の感想

Apple Pencilの活用に関して、「ペーパーライクフィルムを使っている。」「特に問題なく文字の記入ができています。」「文字の色や太さが自由に選べる。」「紙媒体に記入するのとほとんど変わらない。」「テストもこの方式でできたので良かった。」「電池がすぐ切れる。」との回答が寄せられた。

6 考察

弱視生徒は, 文字を拡大する等のアクセシビリティの設定によってICTを日常生活において自分一人の力で活用している。視覚障害ゆえの困難さがあるものの, 晴眼者と同じようなレベルで使用できている例もある。特にiPadをはじめとするタブレット端末は画面サイズが大きく, 持ち運びも容易であるため, 日常生活において欠かすことのできない重要なツールの一つとなっている。そのため, 授業においてもiPadを積極的に活用すべきであるといえ, 現に, タブレット端末によるデジタル教科書の普及が進んでいる。

本研究においては, デジタル教科書に倣って, iPadによるノートのデジタル化を図った授業実践に取り組んだ。その結果, 内容理解の促進, 家庭学習の促進, 合理的配慮の提供, テスト解答時の有用性のいずれの調査項目においても平均値が高く, 弱視生徒に対してiPadによるノートのデジタル化は有効であることが明らかとなった。iPadを活用しない従来の学習形態であれば, 弱視生徒はノートに文字を書く場合にも, ノートに書いた文字を読む場合にも拡大鏡や拡大読書機器等を準備する必要があった。特に, 拡大読書器は高価であることに加え, サイズも大きく, 普段から容易に持ち運べるものではない。iPadによるノートのデジタル化を進めることで, 高価な拡大読書器は不要となり, 一台のiPadさえあれば, 場所を問わずに学習を進めることが可能となる。つまり, 晴眼の生徒が日常生活において勉強しているスタイルと同じように弱視生徒も勉強に取り組むことができる。例えば, 通学途上の電車内でノートを開くように, iPadを活用して勉強することが可能となり, 家庭で身体を横にして寝転がっていてもiPadを活用して勉強することが可能となる。このように晴眼の生徒が勉強するのとほとんど変わることなく, 弱視生徒も勉強することができるようになると, 視覚障害ゆえの困難さが軽減され, 学習に対する動機づけを高めることにもつながるといえる。それゆえ, iPadの活用によるノートのデジタル化は, 弱視生徒に対する合理的配慮の提供として, 学習時間の増加につながり, 内容理解にも直結したと考えられる。また, Apple Pencilの活用は「紙媒体に記入するのとほとんど変わらない。」という回答が寄せられてお

り、弱視生徒にとっては、デジタル化したノートの方が紙媒体のノートよりも使い勝手が良いといえる。テストにおいてもデジタル化したことで、普段通りにiPadとApple Pencilを活用できたことから、解答しやすかったのではないかと考えられる。しかしながら、「ページが飛んでしまう。」というような回答も寄せられたことから、授業でiPadを活用する場合、教員はこの点について特に注意を払う必要がある。

なお、本実践で活用したICTはiPadのみであり、他に特別な機器や環境整備は不要である。教員がパソコンで作成した授業プリントを紙媒体に印刷して配布するという従来の形態が、PDFファイルに変換して電子的に配布するという形態に変わっただけである。アプリに関しても標準的にインストールされているもので十分である。したがって、生徒と教員がiPadさえ所持していれば、比較的簡単に導入でき、弱視生徒に対する一つの方策として広めていくことが十分に可能である。また、GIGAスクール構想が進展することで、タブレット端末は教材として標準的な装備となり、盲学校においても通常学校（小学校、中学校、高校等）においても、タブレット端末等のICTを活用したデジタル化がますます広がっていくであろう。この点を踏まえると、少なくとも本研究におけるノートのデジタル化については、盲学校であろうと通常学校であろうと大きな違いがないものと考えられ、通常学校で行われる授業を盲学校でそのまま導入できる可能性がある。その際には、合理的配慮として視覚障害の特性に応じた対応が必要となってくるものの、盲学校と通常学校の境界を無くすことにつながり、障害のある者と障害のない者

がともに学ぶ仕組みであるインクルーシブ教育システムの構築にも関係するといえる。

7 おわりに

濱田；山本（2019）[2]は、弱視教育におけるICTの効果や課題に関する知見は現時点では十分に蓄積されていないことを指摘している。また、ICT活用の普及に関して、堀田・高橋・丸山・山西（2008）[14]は、ICTを活用した典型的な授業場面が提示され、各教員が自身の授業スタイルに組み込みやすいものを選択し、実際に自分の授業で試していくという方法が必要であると指摘している。

これらのことを踏まえると、弱視生徒に対するICT活用に関する研究や実践が今後ますます進められていかなければならない。そのためには、ICTの得手不得手に関わらず、また、特別な環境設定の有無に関わらず、全ての教員が授業におけるICT活用に取り組んでいかなければならない。その意味で、本研究においては、導入が簡単かつ弱視生徒にとって大きなメリットのある授業実践例を提示することができたといえる。

本研究においては対象者が3名と少数であったため、対象者を拡大した授業実践を行い、引き続き、検証する必要があると考えられる。また、今後の弱視教育において、iPad等のICTが活用されるような授業がスタンダードとなるよう、弱視教育に関わる教員に対して、ICTの活用が生徒の学習環境を飛躍的に高め得るツールであることを広く発信していかなければならないと考えている。

謝辞

ご協力をいただきました関係のみなさまに対して、心より感謝申し上げます。

参考文献

- [1] 独立行政法人国立特別支援教育総合研究所：「視覚障害のある児童生徒のための教科書デジタルデータの活用及びデジタル教科書の在り方に関する研究—我が国における現状と課題の整理と諸外国の状況調査を踏まえて—」, 2016.
- [2] 濱田由己, 山本利和：「弱視生徒を対象とした教育におけるタブレット端末利用の現状と課題」, 大阪教育大学紀要総合教育科学, Vol. 67, pp. 269-278, 2019.
- [3] 独立行政法人国立特別支援教育総合研究所：「特別支援教育でICTを活用しよう」, 2016.
- [4] 慶應義塾大学：「平成28年度学習上の支援機器等教材研究開発支援事業成果報告書(概要)」, 文部科学省, 2017.
- [5] 渡部良平, 大町雅志, 林祐弥, 阿部梨紗, 和田恒彦, 宮本俊和：「視覚障害教育における反転授業の可能性と課題」, 筑波大学特別支援教育研究, Vol. 10, pp. 23-30, 2016.
- [6] 小林ゆきの, 宮城愛美, 田中仁, 金堀利洋, 天野和彦, 香田泰子：「視覚障害学生の教育に関する現状と課題」, 筑波技術大学テクノレポート, Vol. 27, No. 2, pp. 3-29, 2020.
- [7] Appleホームページ
アクセシビリティ
<https://www.apple.com/jp/accessibility/vision/>
(2021年4月1日最終アクセス)
- [8] 渡辺哲也：「視覚障害者の意思疎通支援サービス, 及びICT利用状況の地域間差の分析」, 保健医療科学, Vol. 66, No. 5, pp. 523-531, 2017.
- [9] 文部科学省：「学習上の支援機器等教材活用促進事業平成28年度版」, p. 9, 2017.
- [10] 中野泰志, 富田彩, 山本一寿, 相羽大輔, 氏間和仁, 田中良広, 永井信幸, 韓星民：「ロービジョンの生徒のためのアクセシブルPDF教科書の有効性に関する研究—盲学校における実証研究—」, JSDT年次大会発表原稿集, pp. 73-74, 2014.
- [11] 小倉正幸, 山本一寿, 中野泰志, 相羽大輔, 氏間和仁：「タブレット情報端末を用いた弱視生徒の指導実践報告：拡大教科書として・学習支援機器として」, 弱視教育, Vol. 52, No. 2, pp. 1-6, 2014.
- [12] Appleホームページ
Apple Pencil
<https://www.apple.com/jp/apple-pencil/>
(2021年4月1日最終アクセス)
- [13] Appleホームページ
iPhone, iPad, iPod touchでAirDropを使う方法
<https://support.apple.com/ja-jp/HT204144>
(2021年4月1日最終アクセス)
- [14] 堀田龍也, 高橋純, 丸山紋佳, 山西潤一：「一斉授業の授業過程におけるICT活用の目的・頻度・タイミングに関する調査」, 日本教育工学会論文誌, Vol. 32, No. 3, pp. 255-291, 2008.

第29回年次大会予稿

タイピングによるプログラミング学習のための
ソースコード提示に関する一検討
**An Investigation on a Source Code Sequence for
Learning Programming by Typing**

村川猛彦^{1*}

Takehiko MURAKAWA^{1*}

1 和歌山大学

Wakayama University

〒640-8510 和歌山県和歌山市栄谷930

E-mail: takehiko@wakayama-u.ac.jp

*連絡先著者 Corresponding Author

大学1年生向けの授業において、C言語のソースコードを与えて打ち込ませる「タイピングによるプログラミング学習」を実施しており、週1回の授業を通じて文法知識などの継続的の定着が期待される。授業で使用した59個のソースコードに対して、printfや++などのトークンに分割してから集合演算を行うことで、授業で学んだ書き方を後の授業で活用しているかを定量化することを試みた。

“Learning programming by typing” is a way of learning programming where a learner types the sample source code verbatim. The author has adopted this style in his courses of the C programming language so that the university students could acquire the knowledge of the syntax continually. In this paper, we attempted to verify whether learning contents are utilized in later classes. In the evaluation, 59 source files written in C were segmentalized into tokens such as “printf” and “++”, and set operations emerged the notable tokens.

キーワード: ソースコード分析, C言語, トークン分割, プログラミング教育

source code analysis, C programming language, tokenization, programming education

1 はじめに

学校教育におけるプログラミング学習は、「何のために」「誰が」プログラムを作成するのかに関する変化と、密接な関係がある。数十年前のプログラムの用途として、弾道計算、機械制御、オペレーティン

グシステム (OS) などの開発が想起でき、それらの作成者は専門家 (技術者・研究者) や職業プログラマであった。近年では、コンピュータの低価格化や、文部科学省の「プログラミング教育」[1]の推進などにより、児童もプログラム作成に携わることと

なる。その場合の用途として、(算数・数学などにおける)問題解決、コンピュータグラフィックス(CG)、そして自己表現を挙げることができる。

より多くの人々が効果的・効率的にプログラムを作ることができるようにするため、プログラミング教育が推進されてきた。国内では、大学や専門学校における情報専門教育から、大学全体あるいは学部のすべての学生を対象とする一般情報教育へと、学習者が広がり、学習指導要領に基づき2020年度からは小学校でのプログラミング教育が導入されている。

本研究の主なスコープ(支援対象)は、大学1年生向けのプログラミング授業である。これには2つの理由がある。一つは筆者が授業を実施しており、開発・改良しているWebアプリケーションが利用できること、もう一つは、大学における週1回の授業により、文法知識などを積み重ねて学習することへの支援を想定している(それに対し、小学校などでのプログラミングの授業は例えば単元などで集中的に実施し、そこで獲得した言語の知識は必ずしも上の学年で活用されない)ためである。言語はC言語とする。

プログラミング授業を通じて、系統的かつ継続的な学習をさせるにあたり、例題となるソースコードの内容および提示順序は重要な要素となる。例えばfor文を学習させる場合、構文的には、変数や代入を使用することなく、for文を含むプログラムを書くこともできるが、「1から10まで繰り返す」といった実用的な用途を念頭に置くと、変数宣言や代入の概念を身につけた後に指導するのが適切と言える。その一方で、C言語ではインクリメント演算子(++)が、for

文の更新処理で多用される。しかしながらこの演算子をfor文より先に学習させる必要性はなく、for文の学習と同時期でもよい。

どの学習事項(文法知識など)をいつ修得させるべきかというのは、担当教員の教育観によるところもあり、客観的な検証は困難である。その一方で、既存のコンテンツ(プログラミング授業においてはソースコード)を分析するのであれば、機械的処理および人的チェックが容易となる。

このような背景のもと、本稿では、筆者が2020年度の授業で使用したC言語のソースコードに対して、forや++など、プログラミング言語における語句(トークン)に分割してから、集合演算を行うことで、授業で学んだ書き方を後の授業で活用しているかを定量化することを試みた。

2 写経型学習とタイピングによるプログラミング学習

プログラミングを理解するために採用されている学習方法の一つに「写経型学習」がある[2-3]。学習の概略は次のとおりである。はじめに「手本」となる、書籍などに書かれたプログラムの完成品をそのまま打ち込む。打ち込みを終えたら実行する。もしエラーが発生すれば、打ち込んだ内容と手本を見比べて修正する。

写経型学習のメリットとして、プログラムが実際に動くことを、学習者にとっては少ない手間でも体感できる点が挙げられる。対象とするプログラムを多数用意して(あるいは書籍で範囲を指定して)おけば、期間内にどこまでプログラムを作成したかを、学習者および指導者が把握することもできる。

一方、完成品の模写であることに起因する写経型学習のデメリットとして、創造性を育成することができない点、またプログラムを構成する行または語（トークン）に関する理解が容易ではないといった点を指摘できる。創造性の育成に関しては、別に応用問題を課すか、PBL (Project/Problem Based Learning) の活用が期待される。行またはトークン単位での意味の理解に関しては、写経対象のプログラム群をどの程度用意して学習者に取り組みさせるべきかを検討することが必要となり、これが本研究の主要な問題意識である。

写経型学習に関するワークブックや先行研究をもとに、筆者らは、タイピングアプリをもとにした学習支援のWebアプリケーションを独自に開発した[4-7]。1行のfor文をタイプする場合の流れを図1に示す。

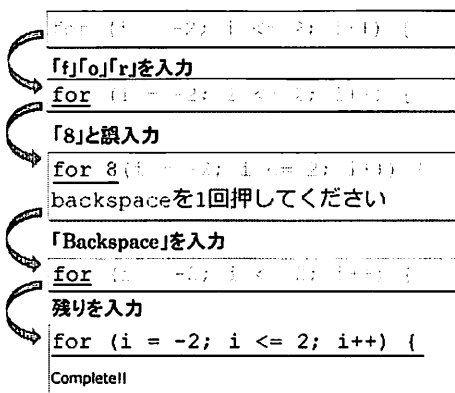


図1 タイピングの流れ[4]

初期状態では打ち込むべき対象を薄い文字色にしている。タイプすると、正しく打ち込んだ文字が黒色に変わる。空白文字は入力する必要がなく、スペース、タブ、改行のキー操作があっても無視される。誤タイプするとき、その誤タイプ文字を赤色で

表示させ、バックスペースキーで削1除しないと、次の文字入力に進めないようになっている。最後の文字（図1では「{」）まで打ち込めば、1問が終了となる。

打ち込む対象は、複数行とすることもできる。また誤タイプとその訂正、および空白文字のタイプを含め、タイプごとに経過時間を獲得してサーバに送付しており、タイピングの手が止まったり、誤タイプしたりした箇所を後で知ることもできる。文献[4]ではタイピングのみであったが、[5-6]の学習支援システムでは、問題ごとに解説ページを設け、プログラムの動きや構造、なぜその関数を使うのかなどを説明している。2021年3月に公開したシステム[7]では、タイプ後に、paiza.IO(オンライン実行環境)によるソースコードと実行結果の表示を取り入れた。

文献[5-6]では「写経型プログラミング」という表記を採用したが、本稿では「タイピングによるプログラミング学習」としている。一般に写経型学習では、動作する一つのプログラムを打ち込むのに対し、タイピングによるプログラミング学習では、1行や複数行といった、プログラムコードの一部分でもよいこととする。また1文字単位でのタイピング状況を記録し、他の学習者の状況と合わせて分析することで学習支援を図るのも、従来の写経型学習と異なる点と言える。

中田[8]はタイピングの速度・正確さと、C言語の理解度の関係について調査するため、ブラウザ上で動作するプログラミング言語学習用のタイピングソフトを構築し、授業においてデータ収集を行った。結果として、タイピング速度が実験期間内に急速に上昇する一部の学生を除いた場合、タイ

ピング速度が速い人ほどC言語の理解が高いことが示唆された。

筆者が関わった2020年度の授業での実施[5-6]では、「for」「printf」といった特定のキーワードのタイピング時間を利用してくれるたびにタイプ時間が減っていくことを想定していたが、若干の傾向は見られるものの、大きな変化は見られなかった。

3 タイピングによるプログラミング学習の授業実施状況

筆者が2020年度後期に担当した授業の概況を述べる。和歌山大学システム工学部1年次学生を対象としたもので、科目名は「情報処理ⅡA」「情報処理ⅡB」であり、本稿では以降「A科目」「B科目」と表記する。それぞれ、第3クォーター(10-11月)・第4クォーター(12-1月)に開講の必修科目であり、週1回(105分)が7回で、合格すると1単位を得る。

同じ時間帯および科目名で2つの科目を開講し、筆者はC言語、もう一つの科目はPythonを通じて、プログラミングの基本を学ぶことを目的とし、受講者(1年生全員および2年次以降の再履修者)はA科目の開講前に履修クラスを希望し、人数調整をせずに受け入れた。なお、A科目・B科目で同一の言語(担当教員)とした。300名強の対象者のうち、筆者のクラスには125名が受講した。

各授業では、1回の授業につき3つ程度の動画を視聴してから、プログラミング課題に取り組んだ。課題は「スライドに表示されたプログラムをそのまま打ち込む(写経)」「写経したあと指示に従って修正する」「仕様を満たすプログラムを自作する」に大別でき、写経と修正の課題はA科目・B

科目共通であるが、自作の課題はB科目で採用した。大学で運用している学習支援システム(Moodle)に、課題ごとにソースコードと実行結果を提出する。プログラミングと実行は、paiza.IOを推奨環境とし、他の作成・実行方法を認めた。

授業と別に、前節で述べた学習支援のWebアプリケーションを使わせた。厳密な本人確認が行えないため、利用は成績評価の対象外とした。A科目の第2回から第7回において、1つずつ、各回の授業に関連するソースコードをタイプさせた。検証(5節)においては、このプログラムを「学習支援のソースコード」と表記する。

4 トークン分割

「授業で学んだ書き方を後の授業で活用しているか」の検証にあたり、本研究では、ソースプログラムの「最小の字句的な単位」[9]となるトークンに着目した。前処理指令を含むソースコードを対象とする。

「#」から始まる行(前処理指令)について、本稿では「#include 文字列」の書式のみを対象とし、これを「#」「include」「文字列」の3つのトークンに分割する。

ソースコードのトークンに着目した技法および分析ツールとしてコードクロン検出(CCfinder)があり、山本ら[10]は、BSD系UnixのCファイルとヘッダファイルを対象として分析を試みている。その研究では異なるソフトウェアシステム間のファイルが同一であるか否かを、最低一致トークン数に基づいて判定しており、本研究と対象および手法が異なる。

A科目で最初に打ち込ませた、以下のソースコードを使用して、トークン分割の例を示す。

```
#include <stdio.h>
int main(void){
    printf("Hello, World!\n");
    return 0;
}
```

上記をトークン分割することで、「#」「include」「<stdio.h>」「int」「main」「(」「void」「)」「{」「printf」「(」「"Hello, World!\n"」「)」「;」「return」「0」「;」「}」が得られる。ヘッダファイル指定(<stdio.h>)や文字列リテラル("Hello, World!\n")は、それぞれがトークンとなる。

C言語のソースコードには、トークンの区切りを表す情報が含まれていない。多くの場合、空白文字がその区切りとなるが、「main(void){」には空白文字が含まれていないにもかかわらず、トークン分割により「main」「(」「void」「)」「{」の5つに分かれる。また文字列リテラルや文字リテラル(''など)に含まれる空白文字は、区切りではなくトークンの一部となる。

このような点に注意してトークン分割を効率よく行うため、C言語用の構文解析ソフトウェアc-lex[11]を使用した。ただしこのプログラムは、前処理指令に対応していないため、本研究で対象とするC言語のソースコードを、前処理指令行と正味のC言語の行に分け、後者に対してのみc-lexを適用してトークンを獲得し、前処理指令については処理プログラムを独自に作成してトークン分割を行った。全体的な処理プログラムをRuby言語で記述した。

複数のソースコードを対象として、トークン分割および本研究で試みる検証のための処理の流れを、集合を用いて記述する。検討対象のソースコードの数を n とし、各

ソースプログラムを C_i ($i = 1, 2, \dots, n$) とする。各 C_i は文字列である。 C_i をトークン分割した結果を T_i と表記する。この T_i は、順序付けられたトークンの並びであるが、これを多重集合とみなし、その重複度を1にして得られる、ソースコードに1回以上出現するトークンの集合を、 S_i とし、 S_i から識別子(変数名、関数名など)とリテラル(数値、文字、文字列)を除外したものを S'_i とする。ここまではソースコードごとに独立した処理で求めることができる。例えば前述のソースコードを C_1 とすると、 S'_1 の要素は「#」「include」「<stdio.h>」「int」「main」「(」「void」「)」「{」「printf」「;」「return」「}」である。

次に複数のソースコードに関わる処理を示す。 C_1 から C_i までのソースコードに1回以上出現するトークンの集合を X_i と表記する。これは $X_i = \bigcup_{k=1}^i S_k$ で求められる。 X_i から識別子とリテラルを除外したトークンの集合 X'_i は、 $X'_i = \bigcup_{k=1}^i S'_k$ である。 C_1 から C_i までをそれぞれ文書と見なしたときの、あるトークン t の文書頻度を、 $F_i(t)$ と表記する。有限集合 $\{\dots\}$ の要素数を $\#\{\dots\}$ と書くとき、 $F_i(t) = \#\{j \mid j \leq i \wedge t \in S_j\}$ と表される。

5 検証結果

「授業で学んだ書き方を後の授業で活用しているか」について、本節では次の2項目に分けて検証を行った。

一つは、「学習支援のソースコードは、対応する授業回またはそれより前に学習したトークンを使用しているか?」を検証するものであり、Q1と略記する。学習支援のソースコード番号を $q_1 (> 1)$, q_2 , ..., $q_m (\leq n)$ としたとき、 $1 \leq j \leq m$ に対して

$X'_{q_j} - X'_{q_{j-1}}$ が空集合か否かで求められる。ここでマイナス記号は集合差演算である。

もう一つは、「A科目で学習する重要なC言語のトークンが、A科目・B科目を通じてどのくらい使用されているか？」を検証する(Q2と略記する)。検証対象のトークンの集合を T とし、 $t \in T$ に対する $F_n(t)$ を求めることとなる。

評価にあたり、2020年度の授業のために用意した59個のソースファイル ($n = 59$) を使用した。このうち36個 (C_1, \dots, C_{36}) がA科目、23個 (C_{37}, \dots, C_{59}) がB科目のものである。 C_1 は前節で例示したソースコードであり、今回のソースファイルでは、 S'_1 は、 S'_2, \dots, S'_{59} それぞれの真部分集合となる。実際、「main」「void」「printf」などはどのプログラムにも出現する。学習支援のソースコード(3節)は6個 ($m = 6$) で、 $C_5, C_{12}, C_{18}, C_{25}, C_{31}, C_{36}$ である。

Q1の検証結果として、 $X'_{12} - X'_{11}$ にトークン「else」が属し、他の集合差演算の結果は空集合となった。 C_{12} はA科目の第3回授業で、学習支援システムを用いてタイプさせたもので、変数の値に対して偶数か奇数かを出力するプログラムであり、if~elseの構文が不可欠となる。第3回授業の中ではif文を解説し、elseの使用は第4回としていた。提供前の開発者間のレビューでは、筆者が「(第3回の) 授業ではelseを使用しません、試しに入れましょう」と記して許容し、プログラム解説のページで丁寧に説明するよう指示していた。

Q2の検証において、対象とするトークンおよびその出現回数を表1に示す。より詳細な、ソースファイルごとのトークンの出現状況を、表2に示す。各トークンは C_1, C_2, \dots の初出の順であり、whileは (ifと書式

が似ているので) forよりも先に取り上げたが、科目における利用頻度は高くないことが分かる。また $F_{59}(\text{for})$ と $F_{59}(++)$ が等しくなったが、詳細を見ると、forが出現し++は出現しないソースコードと、その逆が、一つずつあった。

表1 トークンごとの出現回数

トークン t	出現回数 $F_{59}(t)$
=	54
if	32
=	22
else	21
while	13
for	26
++	26
[17
char	17
double	9
<math.h>	8

6 おわりに

本稿では、C言語に関する大学1年次向けプログラミング科目においてソースコードの提供順序が適切であることを検証することを目的として、トークンに着目し分割処理を行い、ソースコード間で特定のトークンの出現状況を調査した。学習支援システムのソースコードの中で、授業で未習のトークンを使用している事例を確認できたほか、科目全体を通じて、重要なC言語のトークンの頻度を把握することができるようになった。

今後の課題として、この結果を踏まえた学習支援システムのコンテンツ増強と利用者評価、および[2]など他の教材におけるトークン出現状況の調査と本研究の対象との比較などが挙げられる。

謝辞

本稿で述べた学習支援システムのうち、ソースコードと解説文は、小高真太郎氏が作成したものであり、ここに感謝いたします。本研究はJSPS科研費19K12267の助成を受けたものです。

参考文献

- [1] 文部科学省：「小学校プログラミング教育の手引（第三版）」。
https://www.mext.go.jp/content/20200218-mxt_jogai02-100003171_002.pdf（2021年4月12日参照）。
- [2] 喜多一；岡本雅子；藤岡健史；吉川直人：「写経型学習によるC言語プログラミングワークブック」，共立出版，311p.，2015。
- [3] 岡本雅子；喜多一：「プログラミングの「写経型学習」における初学者のつまずきの類型化とその考察」，滋賀大学教育学部附属教育実践総合センター紀要，Vol.22，pp.49-53，2014。
- [4] 小高真太郎；窪田雅崇；村川猛彦：「写経型学習に基づくC言語学習支援システムの開発」，2020年電子情報通信学会総合大会，情報・システム講演論文集1，p.121，2020。
- [5] 小高真太郎：「写経型プログラミングに基づく学習支援システムの構築」，和歌山大学大学院システム工学研究科修士論文，2021。
- [6] 小高真太郎；村川猛彦：「写経型プログラミングに基づく学習支援システムの構築」，2021年電子情報通信学会総合大会，情報・システム講演論文集1，p.92，2021。
- [7] <https://github.com/takehiko/LbTyping>（2021年4月12日参照）。
- [8] 中田豊久：「プログラミング学習の理解度とソースコードタイピングに関する考察」，情報処理学会研究報告，Vol.2013-CE-121，No.7，2013。
- [9] プログラム言語C，JIS X 3010:2003。
- [10] 山本哲男；松下誠；神谷年洋；井上克郎：「クローン検出ツールを用いたソフトウェアシステムの類似度調査」，電子情報通信学会技術研究報告，SS2001-15，Vol.101，No.240，pp.25-32，2001。
- [11] <https://github.com/timpeskett/c-lex>（2021年4月12日参照）。

表2 ソースファイルごとのトークン出現状況

	=	if	==	else	while	for	++	[char	double	<math.h>
1	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0
6	1	0	0	0	0	0	0	0	0	0	0
7	1	0	0	0	0	0	0	0	0	0	0
8	1	1	0	0	0	0	0	0	0	0	0
9	1	1	1	0	0	0	0	0	0	0	0
10	1	1	1	0	0	0	0	0	0	0	0
11	1	1	1	0	0	0	0	0	0	0	0
12	1	1	1	1	0	0	0	0	0	0	0
13	1	1	1	1	0	0	0	0	0	0	0
14	1	0	0	0	1	0	0	0	0	0	0
15	1	0	0	0	1	0	0	0	0	0	0
16	1	0	0	0	0	1	1	0	0	0	0
17	1	0	0	0	0	1	1	0	0	0	0
18	1	0	0	0	1	1	1	0	0	0	0
19	1	0	0	0	0	1	1	0	0	0	0
20	1	1	1	0	0	1	1	1	0	0	0
21	1	1	1	0	0	1	1	1	0	0	0
22	1	0	0	0	0	1	1	1	0	0	0
23	1	0	0	0	0	1	1	1	0	0	0
24	1	1	1	1	0	1	1	1	0	0	0
25	1	1	0	0	0	1	1	1	0	0	0
26	1	0	0	0	0	0	0	1	1	1	0
27	1	1	1	0	0	1	1	1	1	0	0
28	1	1	1	0	0	1	1	1	1	0	0
29	1	1	1	1	0	0	0	0	0	1	1
30	1	1	1	1	0	0	0	0	0	1	1
31	1	0	0	0	0	0	0	0	0	1	0
32	1	1	0	1	0	0	0	0	0	0	0
33	1	1	1	1	1	0	0	0	0	0	0
34	1	1	1	1	1	0	0	0	0	0	0
35	1	1	1	1	1	0	0	0	0	0	0
36	1	0	0	0	0	1	1	0	0	1	0
37	1	1	1	1	0	0	0	0	0	0	0
38	1	1	1	1	0	0	0	0	0	0	0
39	1	1	0	1	0	1	0	0	1	0	0
40	1	1	0	1	1	1	1	0	1	0	0
41	1	1	1	1	0	1	1	0	0	0	0
42	1	1	1	1	0	1	1	0	0	0	0
43	1	0	0	0	0	1	1	0	0	0	0
44	1	1	0	1	0	1	1	1	1	0	0
45	1	1	0	1	0	1	1	1	1	0	0
46	1	1	0	0	0	1	1	1	1	0	0
47	1	0	0	0	0	1	1	1	1	0	0
48	1	0	0	0	0	1	1	1	1	0	0
49	1	1	0	0	0	0	1	0	0	0	0
50	1	1	0	1	1	0	0	0	0	0	0
51	1	1	1	1	1	0	0	1	1	0	0
52	1	1	1	1	1	0	0	1	1	0	0
53	1	0	0	0	1	0	0	0	0	1	1
54	1	0	0	0	1	0	0	0	0	1	1
55	1	0	0	0	0	0	0	0	1	0	1
56	1	0	0	0	0	0	0	0	1	0	1
57	1	0	0	0	0	1	1	0	1	1	1
58	1	0	0	0	0	1	1	0	1	1	1
59	1	1	1	1	1	1	1	1	1	0	0
合計	54	32	22	21	13	26	26	17	17	9	8

左端の番号は C_i の i に対応する。各セルにおいて、1は1回以上出現し、0は出現しないことを表す。

第29回年次大会予稿

専門的データベース間の相互協力は可能か
専門特化型データベースの存在意義を問う

**Is Mutual Cooperation among Specialized Databases Possible:
Questioning the Significance of Specialized Databases**

青柳明佳^{1*}
Sayaka AOYAGI^{1*}

1 国文学研究資料館

National Institute of Japanese Literature

〒190-0014 東京都立川市緑町10-3

E-mail: iwano.sayaka@nijl.ac.jp

*連絡先著者 Corresponding Author

今日、人文科学分野における専門特化型データベースへのアクセス数は減少しつつある。最大の要因としては、いわゆる教育人口の減少に伴い、主要ユーザーが減っていることがあげられる。また専門家が利用するために作られたデータベースは、その専門性の高さゆえに万人向けとはいえない。そのため、専門家の人数が減少すればおのずと利用頻度も下がることになる。人文科学分野における専門特化型データベースの認知度と利用頻度を上げ、重要性が再認識されるためには何が必要だろうか。単独では利用者数や認知度を増加させることが難しい場合は、相互協力をすれば利用頻度を上げ、重要性を認知させていくことは可能なのだろうか。以上を踏まえたうえで、公共性が高く、万人に利用されるデータベースとはどのようなものなのか、「国文学論文目録データベース」を用いて検証を試みる。

Today, the number of accesses to specialized Databases in the humanities is declining. One of the biggest reasons for this is the decrease in the number of major users due to the decline in the so-called educational population. In addition, databases designed for use by specialists are not suitable for everyone due to their high degree of specialization. Therefore, as the number of specialists decreases, the recognition of such databases will naturally decrease, and the frequency of their use will also decrease. What is needed to increase the frequency of use of specialized databases in the humanities and to reaffirm their importance? If it is difficult to increase the number of users on its own, is it possible to increase the frequency of use and raise awareness of its importance through mutual

cooperation? We will try to verify what kind of database is highly public and can be used by everyone, using the "Database of Japanese Literature Theses"

キーワード: 専門特化型データベース, 国文学論文目録データベース, テキストマイニング, 効率化, 相互協力

Specialized databases, Database of Japanese Literature Theses, text mining, improvement of efficiency, Mutual Assistance

1.はじめに

研究を進めるにあたり、どのような先行研究が行われているのかを調査することは必須の作業といえる。また現在では、先行研究の検索・参照を行うため、インターネット上に膨大な学術情報が存在している。特に、学術論文を中心とした目録データベースや、書誌情報を検索するためのリファレンスデータベースは様々な形で国内外に存在しており、国内においては、国立情報学研究所のCiNii Articlesや国立国会図書館のNDL Search、J-Stage等が有名である。同時に特定の研究分野を中心とした専門的データベースも数多く存在する。国文学研究資料館（以下、国文研）が提供する「国文学論文目録データベース」もまた、専門分野に特化した専門特化型データベースのひとつである。

このような専門性の高いデータベースは、それぞれ独自の視点から情報の収集、選択を行っている。多分野にわたる網羅的データベースも必要だが、専門分野に特化して作成されたデータベースがその特性を生かして発展し、活用されていくこともまた、必要なことではないだろうか。

本稿では、国文研の「国文学論文目録データベース（以下、本データベース）」をもとに、本データベースの特徴や、問題点

について述べ、次に特性を生かすための活用術を模索しながら、専門特化型データベースが互いに協力しつつ発展していくための方法について考察する。

2.国文学論文目録データベース

2.1 概要

本データベースは、国文学年鑑を前身とする国文学に特化した論文目録データベースである。国文研が作成・提供するもので、明治21年から現在に至る国文学関係論文の目録情報を採録対象としている。前身は冊子体の『国文学年鑑』である。現在も採録方針に大きな変わりはないが、基本的には国文研に所蔵されている資料を採録対象としているため『国文学年鑑』作成時に収録されていなかった論文は検索することができない。また本データベースの時代分類と分野は、1998年発行の『国文学年鑑』に準じたものとなっているほか、研究上の便宜を図るため、国文研OPACと連携し、書誌情報を参照できるようにするなど、いくつかの変更点が施されている。

国文学論文目録データベース Database of Research Thesis in Japanese Literature

国文学研究資料館 電子資料館 更新日: 2

▼ 一覧+詳細 < > 455/455 表示: 50

22701 - 22735 / 22735

22701	平安朝物語における皇子女の居住空間―『源氏物語』を中心に 天野ひろみ, あまのひろみ, J01120, 国文論叢, 54, 2019,
22702	『源氏物語』における『古今和歌集』の引き歌について 徳岡涼, とくおかりょう, J00740, 国語国文研究と教育, 57, 2019,

ID	01279199
題名	平安朝物語における皇子女の居住空間―『源氏物語』を中心に
執筆者名	天野ひろみ
執筆者名(よみ)	あまのひろみ
請求	J01120
誌著名	国文論叢
通巻	54
開始頁	1

特集=選詩人・夏目漱石【授業実践】漱石「目面に題す」を使った「戯塞ぎ」

図1: 国文学論文目録データベース 一覧+詳細画面

2.2 特徴

本データベースの大きな特徴は、目録情報に加えてキーワードが付与されていることである。データ入力時に、その論文中で言及されている主要な「作品名」「作者名」を標準化したデータとして付与している。そのため、たとえば「俊頼」で検索した場合にも、「源仲正の寄物型恋題歌群―『夫木和歌抄』『三勇和歌集』を資料として」や「継子・鳥・鍋の歌―口承社会と書承社会」とのように、表面上源俊頼との繋がりが見えないものでもヒットする。これは研究者、特にまだ研究を始めて日が浅いものにとってはとても重要なことである。『源氏物語』や『枕草子』などでは論文のタイトルに作品名がないものが多い。それでもタイトルに「桐壺」や「末摘花」とあれば『源氏物語』を取り上げたものだとわ

かるし、「春はあけぼの」や「すさまじきもの」とあれば『枕草子』がテーマだということがわかる。しかし、「総体としての和歌の捉え方と〈よみ人〉概念の変遷―一条期を中心として」などはキーワードとして『源氏物語』が付与されていないければ『源氏物語』で検索したとしてもヒットしない。本データベースで『源氏物語』を検索したときに、タイトルや誌著名にその文字列がなくてもヒットするのは論文キーワードとして『源氏物語』が付与されているからである。また、「蜀山人」と「大田南畝」のように同一人物だが別号を持つ作者についてもどちらでも引けるように作られている。

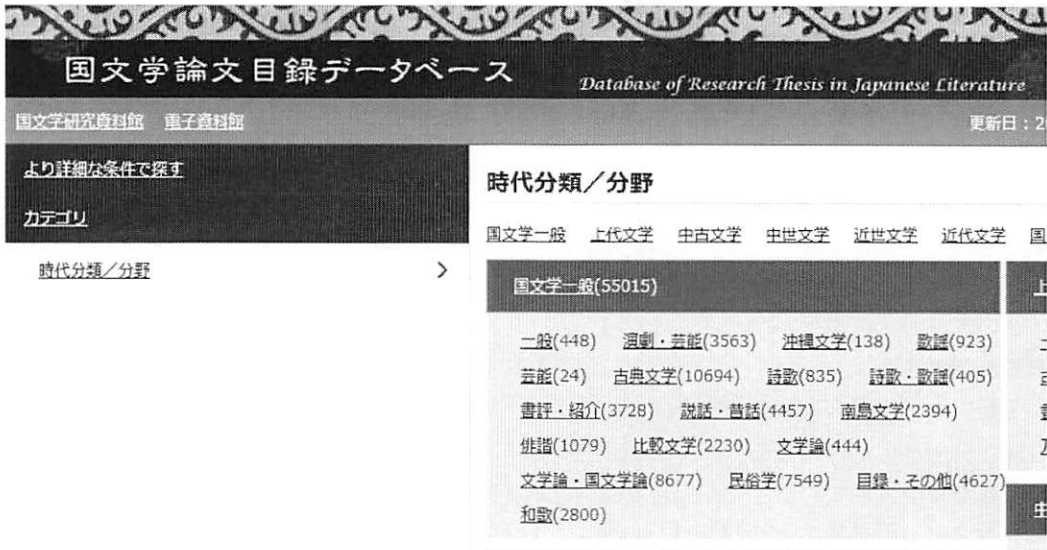


図2：時代分類/分野

2.2 検索方法

本データベースは、画面に表示される時代分類/分野から検索を行うことができる。また、詳細検索画面から時代分類や発表年、論文タイトル、執筆者などを組み合わせた複合的な検索を行うことも可能である。つまり、2000年に発表された論文にはどのような傾向があるか、という傾向分析を行うことができる。このように、国文学研究者が先行論文の調査をするために使われるのではと思われがちだが、社会の動向等を調査する目的にも役立つことができるのである。

2.3 問題点

本データベースの問題点として、論文発表からデータベースとしての公開までに1～2年ほどの遅れが生じるということが挙げられる。これは、人力で論文を読み込み、キーワードの付与や時代分類を行って

るためである。また、年4～6回ほど更新されているが、何が更新されたのか、年代等を含め公表されていない。これは隠しているわけではなく、更新の際に全てのデータを入れ替えており、作業もその回にどのようなデータが追加されたのか知ることができないためである。

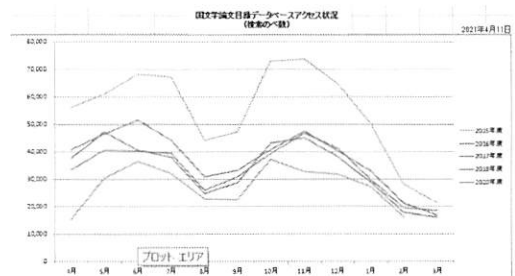


図3：国文学論文目録データベース
アクセス数推移

また、アクセス数の分布をみると3～4月、8～9月の利用者数が少なく、5～6月、11～12月のアクセス数が多い。これは、国文研

からアクセスできるデータベース全体の傾向といえる。3～4月、8～9月は長期休暇であり、大学へ来る学生も少ない。また、5～6月はデータ検索の方法論等の講義が行われるため、アクセス数が増える。11～12月は卒論やレポートなどの提出時期が近づくことから、やはりアクセス数が増えるのであろう。このことから、データベースへのアクセスは学生が主であるという推測ができる。

また、本データベースは研究者向けに作られたデータベースのため、PCでの運用を想定して設計されている。しかし近年、スマートフォンの普及によりパソコンの個人所有率は減少傾向にあり自宅にパソコンを所有していない学生も少なくない。2020年のアクセス数はそれ以前のものにくらべて5～10%程減少しているが、これはコロナ禍により大学が休校あるいはオンライン授業となっており、自宅から本データベースにアクセスすることができなくなったからではないだろうか。

図表 1-6 年齢階層別インターネット端:

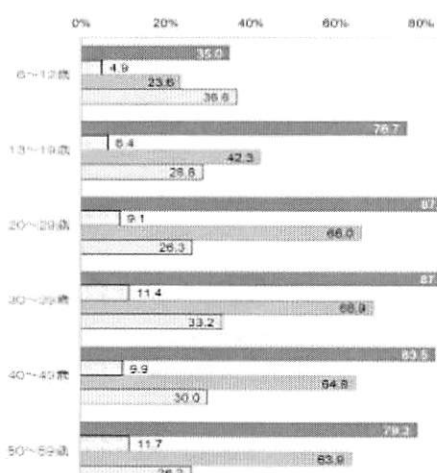


図 3: 令和元年 年齢階層別インターネット端末の利用状況

3. 人文科学分野における専門特化型データベースの活用法

最初に述べたとおり、研究において先行論文の参照は必須である。

Cinii ArticlesやNDL Sarch、J-Stageなどは、より多くの分野を対象とした「網羅的」データベースといえる。一方、本データベースは、採録対象を国内で発表された日本文学・日本語学・日本語教育とする「専門特化型」データベースである。専門特化型データベースの最大の特徴は、検索の利便性にある。網羅的データベースの場合、検索条件によっては膨大な検索結果が表示されてしまうことがある。専門特化型データベースを用いれば、少なくともその専門とする範囲でのみ検索が行われるため、効率的に探すことができるようになる。また、目録情報に付与されたメタデータを記録することで次回以降、副次的に利用することもできる。採録対象の関係で、網羅的データベースには掲載されていないデータもあることが専門特化型データベースの利点といえるだろう。

また、2節でも述べたように本データベースは発表年、時代分類等で検索し、学界の動向分析等にも利用できる。2000年には沢井清が「国文学の研究動向—国文学論文目録データベース（1926年～1996年）を利用した調査」を発表している。このように、特定分野の文献を数量的に扱い、分野ごとの比率を見たり、経年的な動きを見たりという使い方もできる。また、特定分野に特化したデータベースのため、当該分野の研究一覧として使用することもできる。

専門特化型データベースは、その専門分野の研究に必要な他のデータベースと横断

的な検索を目指すことが求められる。本データベースは、現在、国文研のOPACと連携し、資料を参照することが可能である。今後は、CiNiiのように一次データへのリンクを整備することや、付与されているキーワードの可視化、検索画面のスマートフォン対応などが課題としてあげられるであろう。一次情報へのリンクについては、著作権や出版社の許諾等様々な問題が絡むためすべてのデータをリンクさせることは難しい。しかし、キーワードの可視化や時代分類におけるユーザーインターフェースの整備、検索画面のスマートフォン対応などについては比較的容易に進めることができるのではないだろうか。登録速度を早めることも課題のひとつだが、こちらはテキストマイニングツールを用いるなど工夫することで解決が可能ではないかと考える。

専門特化型データベースは、その専門性ゆえに今はまだ万人向けとはいえない。主要ユーザーである日本語・日本文学を学ぶ学生や研究者は年々減少しつつある昨今、アクセシビリティやユーザーインターフェースの整備を行うことにより、利用者層を広げることができれば、今以上に発展性のあるデータベースとして活用できるの

ではないだろうか。こうした取り組みを通じて、専門特化型データベースとしての質を維持しつつも間口を広げた万人に使いやすいデータベース構築に貢献していきたいと考えている。

参考文献

- [1] 相田満：栗城大地：野本忠司：「国文学研究論文目録データベースの高次資源化と汎用化をめざして」, じんもんこん 2019論文集, pp. 211 - 216, 2019-12-07.
- [2] 青柳明佳：篠原泰彦：「人文科学系データベースにおけるユーザビリティの検証—Google Chrome 拡張機能を用いた取り組み—」, 人文科学とデータベース, No25, pp. 9-16, 2020-02-29.
- [3] 浅田徹：「「国文学論文目録データベース」の、あまりよく知られていないこと」, 国文研ニュース, No41, pp. 6-7, 2015.
<https://www.nijl.ac.jp/pages/news/041/041zentai.pdf>
- [4] 総務省：「令和元年通信利用動向調査の結果」, 報道資料, 4p, 2020.
https://www.soumu.go.jp/johotsusintok/ei/statistics/data/200529_1.pdf

第29回年次大会予稿

炎上の早期解決を促す情報解析システムの実現のため、発言の立場ごとに発言内容を分類する手法の研究

Research on a method to classify the content of statements according to the position of the statement in order to realize an information analysis system that promotes early resolution of flaming

横山拓也¹, 箕輪弘嗣^{1*}, 渡邊憲二¹

Takuya YOKOYAMA¹, Hirotsugu MINOWA^{1*}, Kenji WATANABE¹

¹ 岡山商科大学

Okayama Shoka University

〒700-0087 岡山県岡山市北区津島京町2丁目10-1

E-Mail: minowa@po.osu.ac.jp

*連絡先著者 Corresponding Author

近年、SNSは若い人を中心に普及し、それに付随して、企業も率先して利用しているメディアである。しかし、SNSでは炎上と呼ばれる激しい議論が生じてしまう事例がある。本研究では、炎上の早期解決を支援するためのシステムの開発を目指している。本論文では、システム実現のため、Twitterで炎上したツイートのスレッドを対象に複数の機械学習法を用いて意見の賛成派、反対派の分類を機械学習で試みた結果、1例だけだが、86.6-88.1%の精度で分類できた事について報告する。

In recent years, SNS has become widespread mainly among young people, and along with that, companies are also taking the initiative in using it. However, there are cases in which a fierce debate called flaming occurs on SNS. In this research, we aim to develop a system to support the early resolution of flames. This paper report that we tried to classify automatically between pros and cons of opinions using multiple machine learning methods for a tweet thread that burned on Twitter, and obtained results of 86.6-88.1% accuracy.

キーワード: 意思決定支援, 炎上, SNS, ANN, SVM, ロジスティック回帰

Keyword: Decision Making support, Flame, SNS, ANN, SVM, Logistic regression

1. 緒言

近年, Social Network Service以下SNSと呼

ばれるサービスが流行している。SNSは本人の発言や行動に関する事象に対する情報交換などに利用されている反面、意

見の対立や不適切な発言などで、多方面から非難や議論が生じてしまう事がある。これを炎上と呼ぶようになった。炎上には、刑事事件に発展した事例や悪質なデマ、過去には炎上によって倒産に追い込まれた事例[1]や名誉が傷つけられ裁判に至る事もある。裁判となれば、両者が何で争っているか把握するため、弁護士は、各者または各グループの発言を解析し、要約して、裁判官に理解してもらえる準備をしなくてはならず負担を要する。近年は炎上監視するサービス[2]もでており、早期に炎上を納めることが必要となっているがそのための人員を割くことは企業や個人ではとても対処しきれないので機械を用いることで対処することが最も効率が良い方法だと考えられる。

そこでコンピュータを用いた、炎上スレッドの解析システムについて研究する。

本ソフトウェア・システムでは、情報科学の自然言語処理技術を用いて、スレッドの立場毎の発言に分類し、各発言の要約や特徴を分析する事で、炎上の解決に有用となるだろう情報の解析、要約を実施する事を想定している。

本研究では、スレッド内の発言内容から、発言の立場を特定、分類する自然言語処理ベースのテキスト・マイニング手法について研究した内容を報告する。本論文では、提案システムの設計を示し、提案する手法の手順や実施結果について報告する。

2. 研究内容

2.1 全体のシステム

提案するシステムは次の通りである。

(ア) 発言を立場に応じたグループ毎へ

分類

(イ) 各グループの発言の特徴量を解析

(i) 感情の解析・可視化

感情を表す単語から感情の強さを元に算出する手法の確立を想定している。

(ii) 信ぴょう性を量る手法の確立

信ぴょう性は、デマへの批判や訂正意見の多さ、議論の内容の中身が、感情的か、知的かなどで測定できる可能性があり、これらの特徴量を元に算出する手法の確立を想定している。デマでない事を明らかにする事は難しいが、デマに対する反論が多ければ、それは情報の不確かさを示しているといえ、その不確かさが高い情報である事を判別できる事で、デマに扇動される人が減る事を期待できる。

(ウ) 各グループの発言の要約

自然言語処理分野では、文章の要約の機能が研究されてきており、それらを使う事で要約できる機構をもたせる事を想定している。

(エ) 要約や特徴量に基づく助言の提示

上記で得られた数値を元に、対応策を助言できる機構をもたせる事を想定している。デマか否かを助言し、注意を促すなどが実現できると考える。

本論文では(ア)について取り組んだ。その内容は2.2節から述べる。

2.2 AIによる立場の分類の実現

本節では、AIを用い、発言の立場(例：賛成派、反対派、中立派など)を求める手法を研究する。一般的に言われるAIは、機械学

習, 深層学習を使用したものが多く, 本研究では機械学習を用いた手法で研究した。

2.3 提案手法

本研究では, 発言の立場に応じたグループに発言内容をコンピュータで分類する手法を提案する。機械学習の分類は, 万能ではないが, 主張内容が近いのであれば, グループに分類は可能であると考えた。以降に, 分類する提案手法の手順について述べる。

発言の類似度の測定を, 本論文では, 発言中の各単語の有無で選別する。そのため, 発言中の単語の有無をベクトルで表現するone-hot-vectorという構造を利用する。One-hot-vectorは, 表 1の通り, 要素1が「私」という単語に割り当てられていたとすると, 「私」という単語が存在する文があれば, 要素1が1となる。「私」がなければ0が入る。

Bag-of-Word(以下BoW)は, 対象となる文中全ての単語を基に生成したベクトルに対し, 単語の有無でone-hot-vector形式化したものである。

表 1 one-hot-vector

	要素 1	要素 2	要素 3
私	1	0	0
は	0	1	0

2.3.1 手法の手順

Step1 データ解析用の抽出, 整形
激しい議論がなされている一連のテキストデータを解析しやすいファイル形式にスクレイピングする。Webスクレイピングとは, ウェブサイトから情報を抽出するコンピュータソフトウェア技術を指す。

解析用に保存する形式は様々ある。例えば, 表形式に保存したければTSV, CSV, Excelがある。Excelは特に複数のテーブルを保存できる特徴がある。ExcelやCSVは編集が容易な特徴がある。

また, データの構造(型)が決まっているのならRDBMSなMySQL, PostgreSQLなどがある。また, MySQLの様にサーバの用意がむずかしくなければSQLiteという単一に利用できる形式がある。

もし, データの構造が決まっていないのならNoSQLに属するMongoDBなどの手段がある。こちらはJSON, BSON, YAMLなどの形式を変換せずDBへ格納する事ができる。この構造は望ましいのは, 近年のSNSが提供しているAPIはJSON形式が多く, サービス提供側のAPIの改修によって提供されるJSONデータの構造も変化する。そのため, 事前にデータ長やデータ型が定められるRDBMSでは変更の負担が大きかった事に起因する。しかし, 現在はMySQLやPostgreSQLなど著名なRDBMSは, JSON型という可変データを格納できる型を提供してくれており, RDBMSが再び着目されるようになっていく。

Step2 前処理

前処理では, ノイズ, 悪影響になりえそうな情報の除去やデータ構造の整形を行う。データの除去では, 本手法において意味がない顔文字に使われた記号, 数値, 「を」, 「は」などの助詞などである。

データの整形は, NoSQLのまま保存できる事(=非構造のまま保存)と一見反する事項である。しかし, NoSQLは非構造のまま保存できるだけで, 処理する際は, 処理対象を選び, データ型に即して適切に解析す

る必要がある。

Step3 Bag-of-word (BoW) の生成

次に形態素解析で発言内のテキストを単語に分割する。分割して得られた単語を One-shot-vector 形式に落とし込み Bag-of-word を生成する。

Step4 クラス値の付記

生成した BoW のベクトルに、クラス列を追加する。このクラス列は 2.1 節で説明した、機械学習法が、説明変数を基に導出しようとする目的変数を意味する。つまり、目的変数=クラス列の各セルとなる。このクラス列を、発言が属するグループ(賛成派、否定派など)を One-hot-vector 化で、Group ID へ落とし込む。例えば、賛成派=0、否定派=1 と定める事を指す。

Step5 弁別モデルの生成

機械学習法を用い、弁別モデルを生成する。

以上で、発言グループを生成するモデルは完成となる。

2.4 実装と評価実験

Step1 で対象の議論を Excel に落とし込む対象を、Twitter のスレッド[3]とした。スレッドとは、つぶやきの連なりを指す。スレッド[3]では、コロナ禍においてマスク着用疑問を持つ娘さんを父が後押しし、マスク着用を奨励しないツイートをした事で、沢山の人から反論が寄せられたスレッドである。その賛否両論のつぶやきは現在 209 件となった。本スレッド[3]を Python による Web スクレイピングで、Excel に書き

出した。書き出せた項目を表 2 に示す。

表 2 スレッドより抽出した項目

列名	内容
SC_NAME	表示名
SCREEN_ID	つぶやいた者の ID
DATE	つぶやいた日付
TEXT	つぶやいた内容
REPLY_COUNT	リプライ (返信) の数
RETWEET_COUNT	リツイートの数
FAVORITE_COUNT	Favorite(お気に入り)の数

Step2 に従って前処理を行う。前処理した結果は次の通りである。

- 処理に問題があった NAN (=No ANswer) なレコードは解析から除外した。
- スクレイピング後の text 列には、ツイートの文字列が無いレコードがある。これは、画像のみを添付したツイートであった。

Step3 に従って Bag-of-word を生成する。形態素解析には janome [4] を利用した。Python で利用できる形態素解析パッケージは多数あるが、janome は Python のみで実装されており、新語や固有語や複合語に強い辞書 ipadic-NEologd も使う事ができる。janome を用いて形態素解析した結果を scikit-learn パッケージの CountVectorizer クラスで、BoW を作成した。

Step4 に従い、クラスを表す vector 値を付記する。本事例では、マスク着用の賛成派と反対派である。そこで、各つぶやきにおける発言がマスク着用の賛成派と反対派のどちらの視点で書かれているか、筆者が判断し、Web スクレイピングでスレッド[3]を抽出した Excel へ追記した。この列デー

タがStep4で述べたクラス列に該当する。本Stepでは、そのExcel上の賛成派を0, 反対派を1とone-hot-vector化して、Step3で生成したBoWの各レコードに追記する。

2.5 性能評価とその結果

性能評価の手順は次のとおりである。

Step 1. 訓練データと評価データの分割
訓練データと評価データの使い分けは次のとおりである

1. 訓練データは、議論のグループを識別するモデルを生成するために用いる。
2. 一方、評価データは、生成したモデル

を評価するために用いる。

Step 2. 弁別モデルの生成

Step 1で定めた訓練データに対して機械学習法を用い、弁別モデルを生成する。

Step 3. モデルの性能評価

Step 2で定めた評価データを用い、Step5で生成した弁別モデルの分類性能を評価する。

評価結果は表 3となった。平均スコアが求めるべき性能を表しており、Python コードより得られた数値を少数第 4 位で四捨五入し百分率(%)に変換した。表 3の結果を踏まえた考察を次節で述べる。

表 3 弁別性能スコア

	平均スコア [%]	1 回目	2 回目	3 回目	4 回目	5 回目
ANN	88.1%	92.6%	88.9%	85.2%	85.2%	88.5%
Linear SVM	86.6%	92.6%	88.9%	85.2%	81.5%	84.6%
ロジスティック回帰	88.0%	92.6%	92.6%	85.2%	85.2%	84.6%

2.6 考察

表 3の結果より、本スレッドの分類性能の測定値の平均は高い精度で実現できる事を確認した。約88%の精度は、値的には高い。しかし、実用化のためにはもっと正確性の向上が必要である。その機械学習法もほぼデフォルトのパラメータ設定であるので、パラメータや前処理の最適化を実施していく事で、分類性能を100%近く高められる事が望ましい。精度のさらなる向上には、前処理などを強化していく必要があると考える。

例えば、前処理において、2.4節で述べた通り、・の前処理(Step2)で、text欄が空白の

ツイートを、今回は煽っている内容の画像が多かったので、マスク賛成派に対する煽り画像としてクラス値を全て「賛成派」にした。しかし、削除以外の方法もありそうである。例えば、近年の深層学習を用いる事で、写っている被写体名の抽出や、情景の抽出[5]ができてきているので、画像に写っている内容がわかれば、煽っている内容かそうでないのか、がわかるだろう。煽っているか否かを明らかにできる事と、前後の文脈関係から賛成派、反対派などの立場を明らかにできるかもしれないと考えられる。

3. 結言

SNS上の炎上による訴訟, 広告料獲得のためのデマへの扇動などの問題が生じている. その解決のために両者の意見を解析するAIシステムの開発を目指している. 本論文では, 機械学習法を利用して, 2者の意見を弁別する手法を作成, 評価実験を実施した. 結果, ANNで約88.1%, Linear SVMで約86.6%, ロジスティック回帰で約88.0%の精度で弁別できる事が明らかになった.

今後は, 弁別性能の向上や, 両者の主張の要約をして, ユーザの判断・理解の促進する機能, デマか真実かの信頼度を明らかにする機能, 感情的度合いを可視化する機能などを実装して, ユーザの作業負担や不確かな情報に踊らされないようきっかけを与える意思決定支援システムへと発展させたい.

参考文献

[1] 「『バイトテロ, 一生許せない』あのそば店社長からの手紙」, *NewsPicks*, 12

月 16, 2013.

<https://newspicks.com/news/260500/> (参照 1月 21, 2021).

[2] 「投稿監視サービスの比較! メリット・依頼すべき内容・主要7サービス」, *アスピック | クラウドサービス比較・活用サイト*.

<https://www.aspicjapan.org/asu/article/969> (参照 1月 23, 2021).

[3] 「いずみchang (久保田 泉) さんはTwitterを使っています」, *Twitter*.
https://twitter.com/IZUMI_chang_/status/1306381620337754115 (参照 10月 15, 2020).

[4] T. Uchida, *Janome: Japanese morphological analysis engine*.

[5] 秀彰柳澤, 拓朗山下と裕渡辺, 「主要キャラクターの抽出を目的とした漫画キャラクター画像のクラスタリング」, *映像情報メディア学会誌*, vol. 73, no. 1, pp. 199-204, 2019, doi: 10.3169/itej.73.199.

第29回年次大会予稿

Narration Adder : 自動ナレーション追加システムの開発

Develop Auto Narration Adder System

岩田和樹¹, 箕輪弘嗣^{1,*}

Kazuki IWATA¹, Hirotsugu MINOWA^{1,*}

1 岡山商科大学

Okayama Shoka University

〒700-8601 岡山県岡山市北区津島京町2-10-1

E-mail : minowa@po.osu.ac.jp

*連絡先著者 Corresponding Author

本研究では,Text-to-Speechサービスを利用し,PowerPoint内のテキストを音声ナレーションへ変換し,印加するコードの開発を試みた.結果, Google TTSを利用した音声ナレーションを印加できるようになった.ただし,それだけだと聞き取りづらいという問題点があった.しかし,タイトル前後に無音を印加する工夫をしたことで,タイトルと本文に区切りがついて,聞き取りやすくなったと感じられる.また,新たな問題点を発見することができた.使用したライブラリでは,スライドを開いた時に,添付されたオーディオファイルを自動で再生を開始するauto-play設定をONにする事ができていない.今後は,今回実現できなかったことを研究し,より完成度の高いものを実現しようと考えている.

In this research attack, I liked the Text-to-Speech service, and since it was in PowerPoint, I returned Rona and started developing the code now. It's like splitting the space using Google TTS, there is no point in trying to get it with just this. Discover the point It is not possible to set the automatic play setting to play automatically by tapping, tapping, and tapping. Or, he said that he researched what was achieved and attached a box with a degree of appearance.

キーワード: ナレーション,自動化,オーディブル

Keywords: Narration, Automation, Audible

1. 緒言

新型コロナウイルス禍の今,全国の大学は対応に追われている.そうなれば,教員は即座にオンライン授業の準備をしなければならなくなる.

私は,PowerPointを活用した講義に注目した.PowerPointを使った場合,スライドを用いて説明すると思うが,スライドだけだと物足りなさを感じる.また,教員側がスライドに書かれていることを説明しようとするときかなり時間がかかってしまう可能性がある

のだ。そこで、説明を付記したスライド、すなわちスライドに自動でナレーションを付加できるシステムがあれば上記の問題が解決する。また、プログラムなので低コストに抑えることができる。この研究を成功させることができれば、教員の負担を軽減するだけでなく、学生の学習効率向上にも役立つのではないかと考えた。

2. ナレーションの有用性・必要性

本章ではナレーション印加の有用性と必要性について述べていく。

2.1. オーディブルのニーズ

現在、オーディブルが流行っていることから、さまざまなニーズが考えられる。しかし、情報を得ることが目的のビジネス書では、発話において抑揚があっても、特に効果はないと考えられる。つまり、臨場感を持って内容を知りたいというニーズより、読み上げられたナレーションを聞く「ながら聞き」をしたいというニーズが大きいのではないだろうか。また、資料を自分で読むだけだと視覚しか使わないが、ナレーションもあると視覚だけではなく聴覚も使うことになる。情報量が多いほうがより記憶に残りやすいと考えられる[1]。

2.2. 教員の負担

ロックダウンになると、次の日から教科書からの資料をオンライン化しなければならず、教員の負担は大きい。オンライン講義の実施が可能であればよいが、貧困が進む家庭環境では、オンライン講義に耐えられるネット回線がない[2]、罹患者の回復ペースがバラバラ[3]などの理由で、同時に講義できな

い問題等がある。それを乗り切る方法として、VODによる講義の実施が考えられる。VODの場合、事前にスライド等に内容を書き出すことができ、画面を録画しながら音声を付け足すことができる。ただし、スライドの修正のたびに、ナレーションの修正は時間がかかりすぎてしまう。これに自動読み上げを追加できれば、教員の負担を減らせられる。また、画面を録画しながら音声を付け足していたら、データのファイルサイズが大きくなりすぎてしまう問題もある。PowerPointをHTML5に変換すれば動画へ変換するよりずっと講義用データのファイルサイズは小さいのだ。

3. 研究内容

本章では、PowerPointのファイルに機械的なナレーションを追加するソフトウェアについて研究した事を説明する。

3.1. システム全体像

提案システムの概略図を図1に示す。本システムではPowerPointのテキストをクラウド型のText-to-speechサービスへ送信し、音声化されたmp3ファイルを取得する。そして、取得したmp3ファイルをPowerPointファイルの各スライドに貼り付けていくという処理を行う。

これにより、生成されたスライドを開くと、自動で音声ナレーションが流れるだろう。このPowerPointのファイルをmpeg4(.mp4)形式へ変換すると、ナレーション付き動画を生成することができる。PowerPointのversion 2007から保存ファイルはzip圧縮されたXMLファイルとなった。XML(Extensible Markup Language)とは、タグを使って文章を構造化するマークアップ言語のひとつである。XML

以外に有名なマークアップ言語として HTML があるが, XML はデータ記述用の言語であるのに対して, HTML は Web ページを記述するための言語である。また, XML はユーザーが独自のタグを定義することが可能で, データの意味や構造を視覚的に分かりやすく記述することができるという特徴もある。

XML形式は, 値だけではなく, アトリビュートという機構にサブ情報を印加できるので, より詳しい構造を示す事ができる。並列に要素を並べて同系列にデータの追加もできるので, PowerPointアプリの発展に応じてデータの追加が容易となっている。

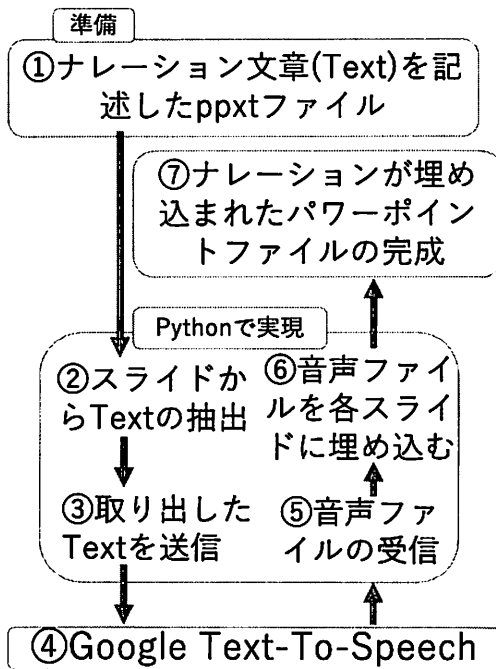


図 1 システム概略図

4. 実装

本章では, 2章で説明したシステムを, Pythonを用いて実装した経緯, 問題点, 結果を示す。

4.1. Text-to-speech のアプリの比較

本システムを開発するに当たり, テキストを音声化するText-to-Speech(以下TTS)について調べた。調査結果は表1に示す。

4.2. ナレーションの空白の実現

Google TTSサービス無料版では, 無音の追加はできない。そのため, タイトルといった音声で聞く時には, 一区切りあって欲しいところに無音の時間がない。そのため, そのまま音声ナレーションを追加すると, スライドのタイトルと本文が続けて音読されてしまい, 聴いた者はタイトルという事を認識しにくいという問題がある。

スライドの説明でタイトルは, 新しい内容であるという区切りである。視覚的にはスライドがあるとはいえ, 音声的にもタイトルである事が明確となるように区切りをつけたい。

そのため, スライド内のタイトルと本文のテキストを別々にTTSサービスへ送付して音声化し, タイトルの音声の前後に, 他の音声との連結の際, その間に約x秒の無音を追加する事にした。xは1~2がはいる事を想定している。無音を追加する事で, ナレーションを区切る事ができた。主観的には, タイトルと本文の区切りを聴いている者に示す事ができると考えている。

表 1 Text-To-Speech のアプリ/サービス比較

No	製品名(URL)	料金	Cloud での 利用	販売元 / 開発元	音質	MEMO
1	Google text-to-speech[4]	無料 / 有料	○	Google	普通	Python 用パッケージがあり、簡単に利用できる。発売元が Google といたこともあり、信用できるソフトである。
2	Balabolka[5]	無料	×	Ilya Morozov	普通	Microsoft Text Speech など音声 API を利用できるソフト。安定している。
3	SoftTalk[6]	無料	×	CNCC	低い	簡単に使用できるが、Balabolka に比べ音質が低く感じた。
4	Microsoft Azure Text speech[7]	無料	○	Microsoft	普通	Windows でしか利用できないようである。
5	Windows Speech API (SAPI)[8]	無料	×	Microsoft	普通	5までリリースされている様子である。
6	初音ミクシリーズ (バーチャル・シンガー・ソフトウェア)[9]	有料	×	Crypton (コア開発は YAMAHA)	高い	著名な初音ミクの歌声を作る事ができるパッケージ。歌らしく聞こえるには抑揚、間の間隔などが必要と思われる、細かく指定できる可能性がある。Python からの利用は不明。

4.3. 結果

ここでは開発したコードをPythonが使える環境で実行する。そうするとソースコードが表示される。ソースコードの一部を説明する。まず、音声を追加したいppxtファイルのファイル名を入れる。次に、音声を追加したときのppxtファイルのファイル名を設定している。それ以下に関しては大まかにいうと、PowerPointのスライドからText（スライドに書かれた文章）を取り出し、取り出したTextをGoogleTTSに送信。その後、音声ファイルを受け取り、各スライドに埋め込んでいるのだ。最終的に、任意でつけたファイル名で保存されて、音声付きファイルの完成である。

新しく追加されたppxtファイルを開くと、各スライドの真ん中あたりに再生ボタンが追加されている。それをクリックして再生すると、スライドに書かれているすべての文字を音読してくれるのだ。

4.4. 考察

4.3節で説明した通り、Google TTSを利用した音声ナレーション印加できるようになった。工夫として、4.4節で説明した通り、タイトル前後に無音を印加する工夫をした。そうすることで、タイトルと本文に区切りがついて、聞き取りやすくなったと考えられる。

実際に作ってみたが、まだ、ナレーションの無機質な感は否めない。インターネット上ではTTSを利用して、なめらかに話しているプロダクトもあるので、その仕組みを調べ取り入れたい。

また、問題点もある。PCの設定や環境によっては、プログラムが動かないことがある。自分のPCでは問題なく動いても、他人のPCだと予期していないところでエラーが起きる

のだ。こういった設定や環境でエラーが起きるのか、しっかりと情報を集めて対処していきたいと考えている。

5. 結言

本研究では、Text-to-Speechサービスを利用し、PowerPoint内のテキストを音声ナレーションへ変換し、印加するコードを開発した。

本論文では、ナレーション自動付加の必要性について説明した。また、設計、実装と今後の課題を説明した。今後は、この研究で実現できなかったオートプレイONの自動化についての研究や人間らしいナレーションに近づける。

参考文献

- [1] 宮川 雅充, 中司 智之と青野 正二, 「音環境の印象に及ぼす視覚情報と聴覚情報の影響」, 騒音制御, vol. 26, no. 1, pp. 53–59, 2002, doi: 10.11372/souonseigyō1977.26.53.
- [2] 株式会社インプレス, 「この国の『通信格差』。学生の通信環境問題を考える【西田宗千佳のイマトミライ】」, *Impress Watch*, 4月 06, 2020. <https://www.watch.impress.co.jp/docs/series/nishida/1245112.html> (参照 1月 24, 2021).
- [3] 「【解説】感染から回復まで、何日くらいかかる? 新型ウイルス」, *BBC ニュース*, 4月 27, 2020. <https://www.bbc.com/japanese/features-and-analysis-52438079> (参照 1月 24, 2021).
- [4] 「Text-to-Speech: 自然な音声合成 | Google Cloud」. <https://cloud.google.com/text-to-speech?hl=ja> (参照 1月 22, 2021).

- [5] 「Balabolka」. <http://www.cross-plus-a.com/jp/balabolka.htm> (参照 1 月 24, 2021).
- [6] 「SofTalk の詳細情報: Vector ソフトを探す!」. 参照: 1 月 24, 2021. [Online]. Available at: <https://www.vector.co.jp/soft/winnt/art/se412443.html>.
- [7] 「Text to Speech | Microsoft Azure」. <https://azure.microsoft.com/ja-jp/services/cognitive-services/text-to-speech/> (参照 1 月 22, 2021).
- [8] Karl-Bridge-Microsoft, 「Windows 10 の音声、音声、および会話 - Windows applications」. <https://docs.microsoft.com/ja-jp/windows/apps/speech> (参照 1 月 24, 2021).
- [9] 「クリプトン・フューチャー・メディア (株) | Sound the future.」 <https://www.crypton.co.jp/> (参照 1 月 22, 2021).

第 29 回年次大会予稿

画像認識 AI を利用した危険診断システムの研究

Research on danger diagnosis system using

image recognition AI

田中蓮¹, 箕輪弘嗣^{1*}

Ren TANAKA¹, Hirotsugu MINOWA^{1*}

1 岡山商科大学

Okayama Shoka University

〒700-0087 岡山県岡山市北区津島京町 2 丁目 10-1

E-mail: minowa@po.osu.ac.jp

*連絡先著者 Corresponding Author

現代の日本は高齢化社会と言われ介護は誰しものが他人事ではなく、親の面倒を見ないといけない可能性がある。そのため、誰しものが職業上の訓練や準備など無しに介護に従事する必要性が生じる可能性がある。そこで本研究はAIによる見守りの目を増やし、事故を未然に防止するため、危険察知システムを開発する。介護者の負担を軽減させることを目的に深層学習を用い介護の中でも最も危険な事案である転倒に焦点を当て、どんな危険があるかを明らかにし、そのシステムの設計をした。本論文では、介護における危険を調査した結果と、危険察知システムの設計内容について報告する。

Today's Japan is said to be an aging society, and there is a possibility that long-term care will have to be taken care of parents rather than by others. Therefore, it may be necessary for everyone to engage in long-term care without vocational training or preparation. Therefore, in this research, we will develop a danger detection system in order to increase the number of eyes watched by AI and prevent accidents. With the aim of reducing the burden on caregivers, we used deep learning to focus on falls, which is the most dangerous case in long-term care, clarified what kind of danger there is, and designed the system. In this paper, we report the results of investigating the dangers in long-term care and the design contents of the danger detection system.

キーワード: 介護, 危険, 転倒, 深層学習, 社会システム / 安全システム

Keyword: Long-term care, Danger, Fall, Deep-learning, Social System / Safety System

1. 介護事故に関する調査

介護における事故や怪我を調べてみた所、表 1 に示す通り、様々な要因がある[1].

表 1 介護に関する事故/怪我の種類

事故/怪我の要因	症状
急激な温度差による血圧の変化	心筋梗塞, 脳梗塞
飲食物の嚥下能力の低下	窒息, 肺炎
歩行中, 立ち上がり時の転倒	骨折
ベッドからの転落	骨折
段差などの踏み外し	骨折
人為的な事故	脱臼, 骨折, やけど

表 1 の全ての危険を検知するシステムを作る事は大変困難である。そのため、最も解決すべき優先順位を探るため、介護事故の割合を調べた、その結果を図 1 に示す。図 1 の通り、厚生労働省の報告では事故原因は「転倒・転落・滑落」が最も割合が多い事がわかった。転倒する事の危険性は、歩くといった活動力を減少させてしまい、要介護レベルが進むことである。寝たきりとなり、介護が必要となりやすくなることは大いにありそうである。更に「死亡に至った転倒・転落・墜落」の内訳は、図 2 に示す資料によると「スリップ、つまずき、よろめき」が86.7%も原因の大半を占める事がわかった。これらの資料結果が確かであれば「スリップ、つまずき、よろめき」を防止する事で介護事故の多くを防止する事ができる事がわかる。そこで、本研究では、「転倒・転落・滑落」を引き起こす「スリップ、つまずき、よろめき」を検知できるソフトウェアシステムの開

発をめざしている。

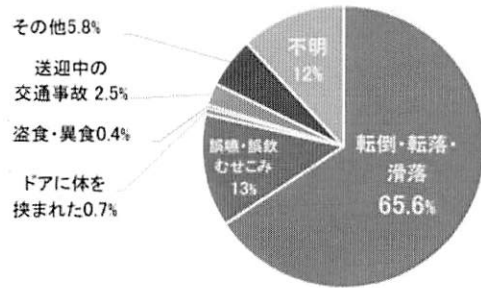


図 1 事故状況分類[3]

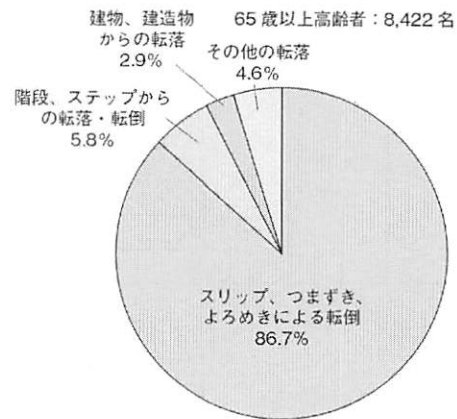


図 2 死亡へ至った転倒/転落/墜落の内訳 [4]

2. 画像認識向け深層学習技術

ディープラーニング (Deep learning, 深層学習) なる技術は、画像認識の AI と呼称され期待されている。しかし、一般的に想像されるような AI には、まだ程遠く技術的に課題が残されている。その課題について本研究における解決案や工夫を混ぜながら示す。

2.1. 画像収集

前述の部分も合わせて、収集画像について

ては表 2の通り定めた. 表 2における分類の後の数字は, 分類レベルを示している.

分類1は厚生労働省報告276事例事故状況分類[3]の「転倒, つまずき」である.

表 2 収集画像の内訳

分類 1	分類 2	分類 3	分類 4	補足
転倒	つまずき	段差	敷居	敷居の高さが重要
			(孫などのおもちゃ階段)	種類, 大きさ, 固さなどが重要
			ぬいぐるみ	大きさ, 固さなどが重要
		非段差	電気コード	電気コードの位置, 張っているかいないかが重要
			敷物 (カーペットなど)	厚みやめくれているかが重要
			布団	
	スリッパ	-	小物	大きさ, 固さなどが重要
			紙	
			ビニール	
			水	量が重要
			油	量が重要
			フローリング	種類が重要
			布製品	

3. システム設計

本章では, 目的のシステムを実現するためのシステム設計について研究した内容を報告する.

3.1. 深層学習を用い画像から判別するシステムの概要

目的をなすため, 深層学習を用いて診断モデルを生成する基本的な図 3 に示す.

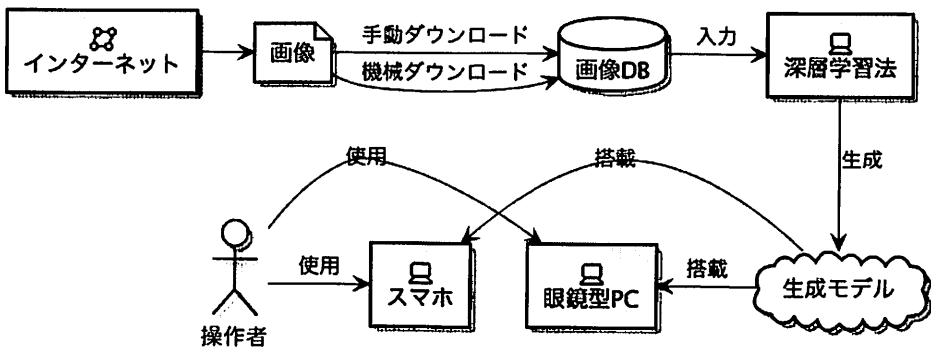


図 3 システムの流れ

2章で述べた通り、深層学習を用いる事で、訓練に用いた画像データから、画像に写っている対象物を認識することができる。そのため、危険が写った画像を沢山集めて深層学習を用い、学習したモデルを作れば、そのモデルは、危険を検知できるはずである。問題は、その危険を機械が判断するためには、多量の学習の画像が必要であり、その画像を人が集めるのは大変である。そのため、個人レベルで収集する以外に、図 3 に示すシステムの通り、ソフトウェアによる巡回で集める仕組みを併せ持つことが望ましい。このインターネットや不特定他者から、危険な画像を収集できる事は望ましい。この仕組みを踏まえて設計した結果を図 4 に示す。図 4 のシステムでは、画像の収集の手段はクローラーによる定期的な画像の収集、運営者による手動での画像収集、クラウドソーシングによる画像収集を想定している。しかし、これらの方法だけでは、本システムで検知できなかった介護者を怪我させてしまう要因の情報が上がってこない。そこで、提案として、図 4 のとおり、匿名ユーザから介護者に危険を及ぼした要因の画像を収集できる窓口を用意する事を提案する。この提案にはリスクを伴う。というのは、ユーザが善意ある者ばかりとは限らず、画像以外のフォーマットや、サイズが大きな画像が送られるなど意図しないデータが届く可能性がある。そのため、画像フォーマットか否か、画像ファイルとして大きすぎないかをチェックや、画像が本募集に合わない内容の画像であるか否か、深層学習を用いて、予め過去の spam 画像と類似しているか、訓練用画像とかけ離れているかをチェックする仕組みを設ける事が望ましいだろう。収集された意図したデータは Pending フォルダへ、危険診断に有用ではない画像は Spam

フォルダへ移る。Pending は、運営者が閲覧して、表 1 のいずれか適切な画像グループへ手動で分類される。訓練データ画像として、分類された画像データを用いて、定期的に診断モデルを更新していく事を想定している。また、匿名ユーザが画像をアップロードできないようにスマホアプリ内に uploader ソフトを内蔵する事で容易にアップロードできる仕組みを目指す。

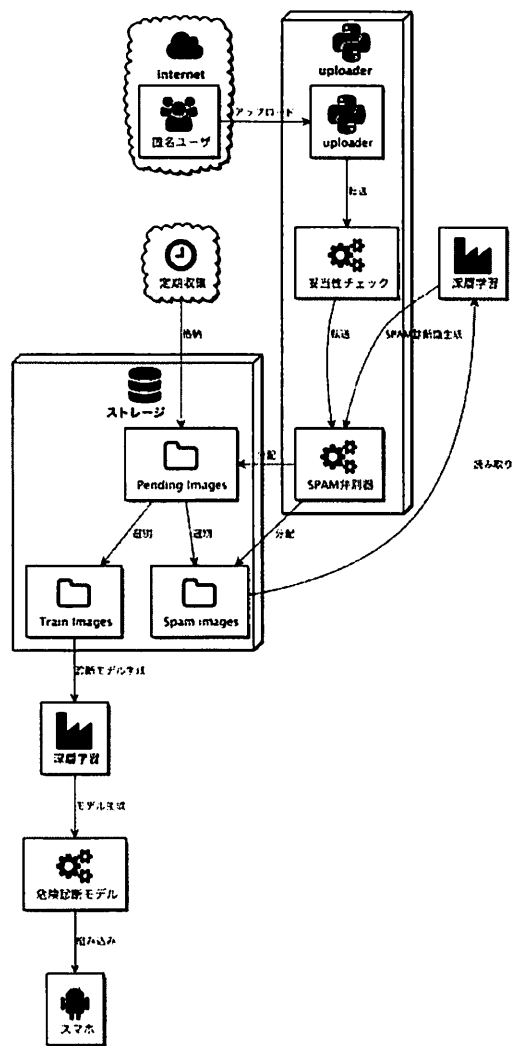


図 4 収集プロセス

3.2. 画像の収集方法

利用できそうな画像収集方法を調べ、次の通りまとめた。

3.2.1. 画像収集候補

「段差」「スリップ」などの事故を引き起こす要因の画像を取得しなければならない。その画像の取得先としては、下記がある事をリストアップした。

(1) 検索エンジン

(ア)GOOGLE 画像検索

(イ)Bing の画像検索

(2)Flickr

(ア)画像置き場として知名度があり、詳細を今後調査する。

(3)ImageNet

(ア)機械学習、深層学習用画像の公開サイト。介護事故の事例があるかは今後明らかにする。

(4)Facebook

(ア)投稿に事故の情報がある可能性がある。

スマホアプリからの画像収集図 4 の生成したモデルをスマホのアプリとして搭載予定である。搭載したアプリを使ってもらい検出できればよいのだが、場合によっては危険なのに危険として抽出できない要因もあるだろう。その要因の画像を多数アップロードしてもらえば、本システムの危険の検出精度が向上する事を期待できる。

3.2.2. 画像の収集手段

(1) 使用者からの不足画像の報告

図 5 に示す様に、使用者に危険の診断が漏れてしまった対象の画像をアップロードしてもらおう。危険の診断システムゆえ、診断漏れは合ってはならない問題であるが、完全

なシステムの完成がずっと先になってしまいかねない。

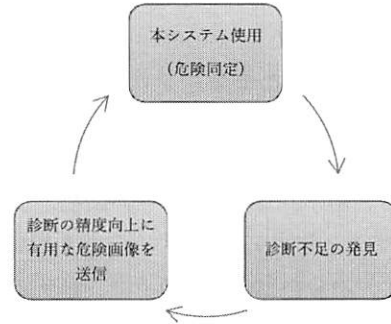


図 5 ユーザからの危険画像提出のサイクル

(2) クラウドソーシングを利用した収集

本研究で利用するクラウドソーシングサービスは、クラウド(Cloud)サービスとは違い、Crowd(群衆)サービスである。サービス例には、Yahoo!クラウドソーシングがある。このクラウドソーシングを利用して、介護中の危険な画像に心当たりがある会員をあたってもらい、その会員から画像を upload してもらおう。会員はインターネットで世界に広がっており、様々な画像を収集できることができる。

(3) クローラーによる画像の機械収集

自動で収集する手段としてクローラーがある。クローラーは、ソフトウェア・ロボットによる分散してネットを巡回して画像を収集するシステムである。自動化でき、便利だが、収集すべき画像の正確性は最も劣る。

4. 試作・調査実装

本システム実現のため、画像の機械収集の例を 4.1.1 節に示す。

4.1.1. マルチラベルの印加

Python の keras (TensorFlow) で多数の認識候補を表示する CNN を使った表示を行った結果を図 6 に示す。図 6 は、介護対象の方が喉につまらせてしまう事故がある餅を認識した例である。rice-cake (餅) である事を約 94% の高確率で検出できた事を示す。介護士や要介護者と餅が写り込んでいる場合は、スマホやスマートグラスを通して、現実空間上に注意喚起メッセージを表示できるだろう。



図 6 餅の識別例

5. 結言

本研究では、介護事故を防止するために深層学習を利用した危険検出システムを提案した。その際、介護事故に関する調査を実施し、転倒に関する事故が多い事を突き止め、転倒、転落、つまずきが起こる要因を調査した。今後は、このシステムを実装していきたい。

参考文献

- [1] 有料老人ホーム・介護施設検索 LIFULL 介護. 「【入居者 100 人に聞いた】多くのケースは怪我や病気—老人ホーム入居のきっかけ | LIFULL 介護」. 参照 2020 年 12 月 31 日. <https://kaigo.homes.co.jp/manual/voice/kentou/disease/>.
- [2] 有料老人ホーム検索探しつくす. 「命に関わる, 介護事故. 起こらないためにどうする? もし起こってしまったら家族は? | 介護のコラム | 介護 |」. 参照 2021 年 4 月 10 日. <https://www.sagasix.jp/column/care/accident/>.
- [3] 「『介護サービスの利用に係る事故を防ぐ注意ポイントを紹介』報告書」. 参照 2020 年 12 月 17 日. http://www.kaigo-center.or.jp/report/pdf/h30_kaigojiko_houkou_20180402.pdf.
- [4] 「統計局ホームページ/人口推計/人口推計 (平成 29 年 10 月 1 日現在) - 全国: 年齢 (各歳), 男女別人口・都道府県: 年齢 (5 歳階級), 男女別人口 -」. 参照 2020 年 12 月 31 日. <https://www.stat.go.jp/data/jinsui/2017np/index.html>.
- [5] 「転倒・骨折予防の取り組み | 健康長寿ネット」. 参照 2021 年 1 月 25 日. <https://www.tyojyu.or.jp/net/topics/tokushu/koreisha-undoki-kenko/tento-kossetsuyobotorikumi.html>.

第 29 回年次大会予稿

因子分析を用いた恋愛小説における文体的特徴の抽出 Extraction of Stylistic Features in Romance Novels Using Factor Analysis

白鳥 孝幸^{1*}, 村井 源¹

Takayuki SHIRATORI^{1*}, Hajime MURAI¹

¹ 公立はこだて未来大学

Future University Hakodate

〒 041-8655 北海道函館市亀田中野町 116-2

* 連絡先著者 Corresponding Author

質の高い物語を自動生成させるためには、各物語類型の物語構造や文体的特徴を明らかにする必要がある。本研究では、長編恋愛小説を対象に、因子分析によって恋愛ジャンルにおける文体的特徴を明らかにすることを試みた。その結果、「好」・「悲」の二感情において、女性キャラクターは序盤、男性キャラクターは終盤の描写が恋愛ジャンルを特徴付けているなど、登場人物の性別によって描写の特徴が異なることが確認できた。

In order to automatically generate high-quality stories, it is necessary to clarify the narrative structure and stylistic characteristics of each story type. In this study, we attempted to clarify the stylistic characteristics of the romance genre through factor analysis of full-length romance novels. As a result, it was confirmed that the characteristics of the descriptions of the two emotions of "like" and "sad" differed depending on the gender of the characters, such as the descriptions in the beginning of the story for female characters and the descriptions in the end of the story for male characters, which characterized the romance genre.

キーワード: 計量文体学, 自然言語処理, 物語論

Stylometry, Natural Language Processing, Narratology

1 はじめに

近年、人工智能に物語を自動生成させる試みがなされている。松原ら^[1]は、星新一のショートショート全編を分析し星新一のショートショートのような短編小説の自動生成を試み、公募文学賞である「星新一賞」において一次選考を通過することに成功している。しかし、自動生成された作品自体の質はいまだ高いとはいえない。その理由の一つとして、物語という媒体の特徴や性質が十分に明らかになっていないことが挙げられる。

しかし、すべての物語をまとめて分析し、その特徴を出すことは難しい。そこで、物語ジャンルごとに対象を絞り、それぞれのジャンルにおける特徴を分析する研究がなされている。例えばこれまでミステリー^[2]、ホラー^[3]、神話^[4]、RPG^[5]などの特徴分析がなされている。しかしこれまで、比較的メ

ジャーなジャンルである「恋愛」の物語について計量的な分析がなされた研究は少ない。

また、感情という観点から物語を分析する研究もなされている。Reagonら^[6]は、単語ごとに感情値を設定し、時間経過による感情値の変化について調べ、物語を6つの類型に分類することに成功している。また、白鳥ら^[7]は感情語極性辞書を用いることによって、短編小説におけるハッピーエンドとバッドエンドの特徴の抽出に成功している。

そこで、本研究では感情語極性辞書（以下、極性辞書と呼ぶ）を作成し、恋愛小説の文体特徴の抽出を行った。極性辞書とは、ポジティブまたはネガティブな単語（以下、極性語と呼ぶ）が収録された辞書のことである。極性辞書を使用することにより、登場人物の感情状態を時系列ごとに抽出し、物語を計量的に分析することが可能となる。

また、極性辞書を使用するにあたっての分析対象としては、登場人物の感情の揺れ動きが大きく描かれているジャンルが好ましいが、恋愛小説はこれに合致する。そこで本研究では極性辞書で収集したデータを因子分析することによって、恋愛小説における物語の特徴を明らかにすることを目標とする。

2 対象作品

本研究では、対象作品として長編恋愛小説 35 作品を選定した。本研究で使用した作品一覧を表 1 に示す。

対象作品の選定は、2020 年 5 月 12 日時点においてデータが公開されている四つの恋愛小説のランキングサイト [8][9][10][11] において、2 回以上登場している長編小説を対象作品とした。ただし、同一著者の作品が複数存在する場合、読書メーター [12] における登録者数を指標に、最も登録者数の多い 1 作品のみを分析対象とした。これは、特定の著者の文体が結果に影響を及ぼすことを防ぐためである。

3 手法

3.1 極性辞書の作成

本研究では、極性辞書を用いて特徴の分析を行った。極性辞書を作成するにあたり、長岡技術科学大学の自然言語処理研究室が作成した「SNOW D18: 日本語感情表現辞書」[13] (以下、SNOW と呼ぶ) を参考にした。SNOW は手作業でノイズの除去と感情カテゴリの付与を行っており、辞書ベースでの感情解析において現行で公開されている極性辞書の中で最も適していると考えられる。また、SNOW では 48 の感情カテゴリが用意されているが、本研究では Ekman の提唱した基本 6 感情 [14] に加え独自に 4 つを追加した計 10 感情を用いてカテゴリの再付与を行った。分析で使用した極性辞書の総単語数は 1563 語であった。以下の表 2 と表 3 に極性辞書の一部と付与した感情カテゴリの定義表を示す。

表 1 対象作品一覧

作品名	著者	出版年
伊豆の踊子	川端康成	1926
ジョゼと虎と魚たち	田辺聖子	1985
ノルウェイの森	村上春樹	1987
黄色い目の魚	佐藤多佳子	1993
天使の卵	村山由佳	1994
恋	小池真理子	1995
いちご同盟	三田誠広	1998
恋愛中母	山本文緒	1998
A2Z	山田詠美	2000
GO	金城一紀	2000
世界の中心で、愛をさけぶ	片山恭一	2001
アジアタムプルー	大崎善生	2002
ツ、イ、ラ、ク	姫野カオルコ	2003
愛がなんだ	角田光代	2003
ある愛の詩	新堂冬樹	2004
そのときは彼によろしく	市川拓司	2004
夜のピクニック	恩田陸	2004
夜の果てまで	盛田隆二	2004
肩越しの恋人	唯川恵	2004
100 回泣くこと	中村航	2005
ナラタージュ	島本理生	2005
クローズド・ノート	梶井縉介	2006
君と会えたから …	喜多川泉	2006
夜は短し歩けよ乙女	森見登美彦	2006
流れ星が消えないうちに	橋本紡	2006
英丘	石田衣良	2006
九月の恋と出会うまで	松尾由美	2007
陽だまりの彼女	越谷オサム	2008
ふがいない僕は空を見た	窪美澄	2010
勝手にふるえてろ	錦矢りさ	2012
三日間の幸福	三秋縹	2013
ぼくは明日、昨日のきみとデートする	七月隆文	2014
君の隣をたべたい	住野よる	2014
桜のような僕の恋人	宇山佳佑	2014
君の名は。	新海誠	2016

表 2 作成した極性辞書 (一部抜粋)

極性語	極性値	感情カテゴリ
咽び泣き	-1	悲
頑張る	1	喜
感傷	-1	恐
気遣う	1	好
恐恐	-1	恐
虐げる	-1	嫌

表 3 感情カテゴリの定義表

感情	定義	単語数
喜	嬉しさを喜びを感じ、楽しそうにしている様子	218
悲	悲しみや哀しみを感じ、落ち込んでいる様子	374
嫌	嫌悪感や敵意を抱いている様子	143
怒	怒りを示したり、感情を昂らせている様子	171
恐	不安や緊張、恐れを抱いている様子	172
驚	驚きを示している様子	220
好	人物に対し友愛や性愛・好感を抱いている様子	180
恥	照れや恥ずかしさ・羞恥心を抱いている様子	26
安	安らぎを感じ、感情状態が落ち着いている様子	39
苦	精神的または肉体的苦痛を受け、苦しんでいる様子	110

3.2 時系列ごとの極性語の抽出

極性辞書を用いて、分析用のデータの作成を行った。まず、形態素解析を行った対象のテキストから極性語を抽出した。また物語の性質として、三幕構成や起承転結など、時系列で物語の展開に起伏を与え、そこから面白さが生まれているということが一般的に知られている。そこで本研究では、物語を時系列で5分割し、それぞれの部分における極性語のデータを収集した。また、時系列はセンテンス数を基準とし、分割されたそれぞれの部分のセンテンス数がほぼ等しくなるように分割を行った。

3.3 性別ごとの極性語の抽出

恋愛小説において、男性の登場人物と女性の登場人物に特徴的な感情パターンの差異が存在するという仮定のもと、5分割したそれぞれの部分における男性メインキャラクターと女性メインキャラクターの極性語を抽出した。なお、キャラクターを示す名詞に直接係り受けしていない極性語であっても、当該シーンでのキャラクターの感情状態を表すような

間接的な感情の描写法は文学作品においては広く用いられている。そこで本研究では「センテンスの中で男性メインキャラクターを表す単語と共起している極性語は、そのキャラクターに関する極性語である」と仮定し、女性の場合も同様とした。それをもとに、作品内において男女のメインキャラクターを表している単語を調べ、その単語を含むセンテンスにおける極性語を抽出した。

3.4 使用する特徴量の選定

分析で使用する特徴量の選定を行った。まず、各感情の出現頻度を特定するため、5分割時における各感情の平均出現回数を算出したグラフを図1に示す。

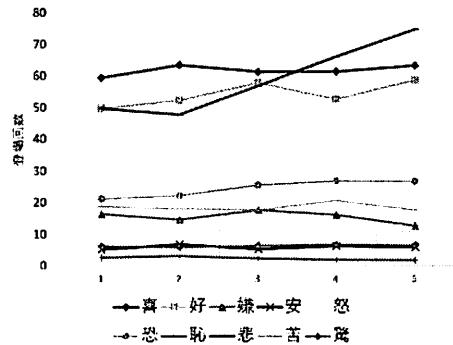


図 1 時系列ごとの各感情の平均出現回数

図1より、「喜」・「好」・「悲」が他の感情に比べて突出して多く出現していることが分かる。登場回数の少ない感情の特徴量が分析におけるノイズになるのを防ぐため、本研究では「喜」・「好」・「悲」に関する特徴量のみを分析用データとした。

3.5 因子分析

5分割したそれぞれにおける男女メインキャラクターにおける「喜」・「好」・「悲」の3感情の登場回数の特徴ベクトルとして与え、因子分析を行った。因子数はガットマン基準・MAP基準・スクリー基準・平行法で算出した因子数の平均である4とし、回転法はプロマックス回転を用いた。

4 結果と考察

因子分析の結果を表4に示す。ここでは説明変数の表記として、時系列で5分割したうちの一目を「1/5」のように表し、続いて感情、性別の順に表記している。

表4 因子分析の結果

	因子1	因子2	因子3	因子4
1/5_喜_男性	0.74	-0.09	0.25	0.16
2/5_喜_男性	0.76	-0.27	0.57	0.03
3/5_喜_男性	0.63	0.12	0.15	0.37
4/5_喜_男性	0.54	-0.08	0.30	0.43
5/5_喜_男性	0.71	-0.21	0.09	0.31
1/5_好_男性	0.72	0.45	-0.22	0.04
2/5_好_男性	0.80	0.08	-0.05	-0.03
3/5_好_男性	0.83	0.17	-0.14	-0.11
4/5_好_男性	0.89	0.07	-0.16	-0.07
5/5_好_男性	0.80	0.12	-0.03	-0.10
1/5_悲_男性	0.79	0.24	-0.28	0.06
2/5_悲_男性	0.75	-0.15	-0.10	-0.02
3/5_悲_男性	0.91	-0.11	0.08	0.05
4/5_悲_男性	0.92	0.02	-0.08	-0.13
5/5_悲_男性	0.87	-0.17	0.05	0.06
1/5_喜_女性	0.06	0.05	0.79	0.23
2/5_喜_女性	-0.07	0.17	0.86	-0.04
3/5_喜_女性	-0.12	0.16	0.84	0.18
4/5_喜_女性	-0.30	0.30	0.56	0.50
5/5_喜_女性	-0.10	0.57	0.34	0.45
1/5_好_女性	-0.11	0.97	-0.22	0.05
2/5_好_女性	-0.06	0.85	0.20	-0.02
3/5_好_女性	0.00	0.76	0.20	-0.24
4/5_好_女性	-0.06	0.69	0.34	-0.17
5/5_好_女性	0.19	0.84	0.06	0.03
1/5_悲_女性	0.16	0.88	-0.12	0.19
2/5_悲_女性	0.15	0.80	0.01	0.21
3/5_悲_女性	-0.10	0.72	0.33	0.12
4/5_悲_女性	0.17	0.45	0.62	-0.32
5/5_悲_女性	-0.03	0.42	0.54	-0.03

今回抽出することのできた4つの因子についてそれぞれ考察を行う。それぞれの因子において、最も高い因子負荷量を示した変数上位5つを表5～表8に示す。

4.1 因子1・因子2に関する考察

まず、対照的な特徴を持つ因子1と因子2について考察を行う。表5と表6より、因子1は男性に関する極性語、因子2は女性に関する極性語のみが高い因子負荷量を示していることが分かる。また、因子負荷量が高い上位5つの変数に注目すると、因子1では全てが中盤から終盤の「好」または「悲」、因子2ではうち4つが序盤から中盤の「好」または「悲」であり、「喜」の極性語は因子負荷量が高くないことが分かる。

4.1.1 対照的な感情の共起

「好」と「悲」という一見相反するように見える感情が共起して出現する理由として、以下の2点のような理由があげられる。

一つ目は、物語における片思いの描写である。まず、恋愛小説における典型的なプロットの1つに、「主人公がとある登場人物を好きになり、イベントや交流を経てその人物と両思いになる」というものがある。このような、物語のなかで「交際」または「両思い」という目標掲示を行っている作品では、大半の部分が主人公の片思いに関する描写である。そのため、相手のことが好きだがその恋が実っていないという状況が発生し、「好」と「悲」が共起したと考えられる。

二つ目は、恋愛中に発生する障害である。これは、恋愛物語における「両思いにある(あった)登場人物らに降りかかる災難とその顛末」という物語要素においてしばしば描かれ、恋人の病気や事故、浮気などもこれに含まれる。例えば主人公の恋人が病気になる、主人公はその恋人が好きだからこそ深い悲しみを覚える。そのため、相手に対する好意とその状況に対する悲観が同時に出現し、「好」と「悲」が共起したと考えられる。また、このような要素は先に述べた典型的なプロットにも部分的に描かれており、物語の中に起伏を生じさせ、読者を飽きさせない効果があると考えられる。

4.1.2 恋愛小説における描写の性差

表5と表6から、男性では終盤、女性では序盤の因子負荷量が高いことが確認できる。

まず因子1が示すように、男性メインキャラクターの終盤における「好」と「悲」の因子負荷量が

高い理由として、作中で罹患するキャラクターの性差が挙げられる。分析対象 35 作品中、物語終盤でメインキャラクターが罹患する作品は 12 作品にのぼったが、いずれの作品も主人公（語り手）の性別が男性であり、またいずれの作品も罹患する人物の性別が女性であった。このことから因子 1 は、主人公である男性メインキャラクターがヒロインを病魔から救うことができないやせなさ、ヒロインが罹患したことで深まる相手への好意が表れた因子であると言える。以上のことから、因子 1 を「ヒロイン喪失因子」と命名する。

次に因子 2 が示すように、女性メインキャラクターの序盤における「好」と「悲」の因子負荷量が高い理由として、女性メインキャラクターの「真っ直ぐな片思い」または「アバンチュールの・初期衝動的な好意」が挙げられる。後者の具体例としては、妻帯者への好意または浮気・年齢差または身分差のある相手への好意などである。実際に、因子 2 の因子得点が高かった上位 10 作品のうち女性が語り手の作品は 8 作品あり、その 8 作品全てが先に述べたトピックのいずれかを含んでいた。このことから因子 2 は、主人公である女性メインキャラクターが男性メインキャラクターに対し好意を抱くもなかなか振り向いてもらえない悲しさ、または初期衝動的な禁断の恋とそれに対する後ろめたさ・葛藤が表れた因子であると言える。以上より、因子 2 を「男性への片思い・禁断の恋因子」と命名する。

また、恋愛小説における女性は男性に比べ、一般的に規律違反とされている「浮気」というトピックを物語として昇華しやすい可能性も示唆される。

表 5 因子 1 を構成する主な要素

	因子負荷量
4/5_悲_男性	0.92
3/5_悲_男性	0.91
4/5_好_男性	0.88
5/5_悲_男性	0.83
3/5_好_男性	0.80

表 6 因子 2 を構成する主な要素

	因子負荷量
1/5_好_女性	0.97
1/5_悲_女性	0.88
2/5_好_女性	0.85
5/5_好_女性	0.84
2/5_悲_女性	0.80

4.2 因子 3・因子 4 に関する考察

次に因子 3 について考察を行う。表 7 から、高い因子負荷量を示しているのはほぼ女性メインキャラクターに関する要素で、「喜」に関するものが多いことが分かる。また、中盤から終盤にかけて「悲」も高い因子負荷量を示している。これは、序盤から中盤にかけて恋人役とうまくいっていた女性メインキャラクターが、終盤に差し掛かる際に災難・試練が起こるといった要素を示していると考えられる。また、終盤 5 分の 5 における「喜」と「悲」がどちらも高い因子負荷量を示していることから、ハッピーエンドとバッドエンド双方の要素を持つ因子であり、これは先に述べた災難・試練の結果に依存すると考えられる。以上より、因子 3 を「女性終盤降下因子」と命名する。

表 7 因子 3 を構成する主な要素

	因子負荷量
2/5_喜_女性	0.86
3/5_喜_女性	0.84
1/5_喜_女性	0.79
4/5_悲_女性	0.62
2/5_喜_男性	0.57

最後に因子 4 について考察を行う。表 8 から、男女それぞれのメインキャラクターにおける、終盤の「喜」が高い因子負荷量を示していることが分かる。これは、男女のメインキャラクター同士が最終的に両思いになるという恋愛小説における典型的なハッピーエンドを端的に表した要素であると考えられる。以上より、因子 4 を「ハッピーエンド因子」と命名する。

表 8 因子 4 を構成する主な要素

	因子負荷量
4/5_喜_女性	0.50
5/5_喜_女性	0.45
4/5_喜_男性	0.43
3/5_喜_男性	0.37
5/5_喜_男性	0.31

以上より、恋愛小説はこれらの要素を複合的に組み合わせることによって、物語のパリエーションや物語としての面白さを生じさせているということが推測される。

5 まとめ

本研究では、極性辞書を用いて収集したデータに対して因子分析を行い、恋愛小説における特徴的な要素を複数抽出した。

基礎統計として、本研究で採用した 10 感情の平均登場回数を調べたところ、「喜」・「悲」・「好」が他に比べ大きく上回っていることを示した。したがって、本研究で採用した感情のうち、この 3 感情が恋愛小説における基本感情であると考えられる。

因子分析では、男女のメインキャラクターに関する極性語のデータから四つの因子を抽出し、考察を行った。因子 1 では物語の終盤で恋人の危機に悲しむ男性、因子 2 では序盤で情熱的な恋愛をする女性など、性別によって描写が異なる傾向があることが明らかになった。

6 今後の展望

本研究に関する具体的な課題として、以下のようなのが挙げられる。

6.1 対象作品の拡充

本研究では客観的な基準を用いて恋愛長編小説 35 作品を分析対象としたが、分析の正当性をさらに上げるためには、基準を見直した上での作品数の拡大や人気のある短編形式の恋愛小説の追加、他ジャンルの作品の分析・比較を行うことが望ましい。

6.2 分析手法の再検討と精度向上

本研究では時系列で物語を 5 分割し分析を行ったが、他の分割数での分析は行っていない。分割数を増やす、あるいは減らすことによって、より特徴的な要素の発見ができる可能性がある。また、本研究では感情語のみを辞書に収録していたが、白鳥^[7]の用いた極性辞書のように、カテゴリを粒度によって 2 つに分け、感情語だけではなく、ポジティブまたはネガティブな気持ちを誘発させるようなイベントに関する単語を追加することによって、新たな知見が得られる可能性がある。

また、本研究で作成・使用した極性辞書は筆者のみで作成を行っているため、複数の分析者に感情の定義表を用いた分析を行ってもらい、一致度の検定を行うことで客観性を担保する予定である。

6.3 恋愛小説における物語構造の抽出

本研究では恋愛小説を構成する特徴的な要素をいくつか明らかにしたが、物語類型としての恋愛小説における典型的な物語構造を明らかにすることはできていない。物語構造を明らかにすることによって、本研究で得られた特徴的な要素の信頼性がさらに高まると推測される。また恋愛小説においては、展開ユニットから展開パターンを抽出するだけでなく、登場人物どうしの人間関係の変化を記述することで、さらに詳細なパターンが判明することが期待される。

参考文献

- [1] 松原仁, 佐藤理史, 赤石美奈, 角薫, 迎山和司, 中島秀之, 瀬名秀明, 村井源, 大塚裕子:「コンピュータに星新一のようなショートショートを創作させる試み」, The 27th Annual Conference of Japanese Society for Artificial Intelligence, 2D1-1, 2013.
- [2] 山田康貴, 村井源:「ネットワーク分析を用いたミステリー小説における場面の依存関係の分析」, 情報知識学会, Vol. 30, No. 2, pp. 206-213, 2020.
- [3] 鈴木諒輔, et al.:「物語と情景描写を自動生成する統合的システムの検討と開発」, 研究報告

- エンタテインメントコンピューティング, pp. 1-8, 2018.
- [4] 吉田拓海, 村井源:「物語自動生成に向けて物語要素間の関係に着目した神話物語の構造分析」, じんもんこん 2020 論文集 2020, pp. 143-148, 2020.
- [5] 中村祥吾, 村井源:「クエスト構造に注目したロールプレイングゲームの物語構造分析手法の提案」 じんもんこん 2020 論文集 2020, pp. 149-156, 2020.
- [6] Reagan, Andrew J., et al.: "The emotional arcs of stories are dominated by six basic shapes." EPJ Data Science, Vol. 5, No. 1, pp. 1-12, 2016.
- [7] 白鳥孝幸, 村井源:「計量文体学を用いた喜劇性と悲劇性の抽出」, 情報知識学会, Vol. 30, No. 2, pp. 276-282, 2020.
- [8] Amazon:「恋愛小説特集の売れ筋ランキング」, <https://www.amazon.co.jp/gp/bestsellers/books/2515232051> (2020年5月13日参照).
- [9] BOOK-OFF:「恋がしたくなる! 恋愛小説100」, <https://www.bookoffonline.co.jp/book/files/novel-love-renai/> (2020年5月13日参照).
- [10] BookLive:「【2020年に読みたい!】読者が選ぶ、おすすめ恋愛小説ベスト50ランキング!」, <https://booklive.jp/feature/index/id/reinainovel> (2020年5月13日参照).
- [11] rank1:「恋愛小説おすすめランキングBEST80【2020最新版】」, <https://rank1-media.com/I0001035> (2020年5月13日参照).
- [12] 株式会社トリスト:「読者メーター」, <https://bookmeter.com> (2020年5月13日参照).
- [13] 山本和英:「日本語感情表現辞書」, <http://www.jnlp.org/SNOW/D18> (2020年6月10日参照).
- [14] Ekman, Paul: "An argument for basic emotions", Cognition & emotion, Vol. 6, No. 3-4, pp. 169-200, 1992.

第 29 回年次大会予稿

機能語に基づく物語会話文の文体分類と推定 Stylistic Classification and Estimation of Story Conversation Sentences Based on Function Words

逢坂 駿也^{1*}, 村井 源¹

Shunya OSAKA^{1*}, Hajime MURAI¹

¹ 公立はこだて未来大学

Future University Hakodate

〒 041-0804 北海道函館市亀田中野町 116-2

* 連絡先著者 Corresponding Author

物語テキストでの発話者と会話文を紐付けたデータはさまざまな応用が可能であるが、従来発話者の特定が困難であった。そこで本研究では、会話文の話者推定に有用な特徴として文体に着目し、日本語物語テキストの会話文を文体に基づき人手で分類した。また機能語を用いて文体の推定を行うアルゴリズムを開発した。先行研究の手法では分類の正答率は 5 割ほどであったが、提案手法では 8 割を超える結果となった。

The data in which the speaker and the conversational sentence are linked in the story text can be applied in various ways, but it has been difficult to identify the speaker in the past. Therefore, in this study, we focused on the writing style as a useful feature for speaker estimation of conversational sentences, and manually classified the conversational sentences of Japanese story texts based on the writing style. We also used an algorithm to estimate the writing style using functional words developed. In the method of the previous research, the correct answer rate of the classification was about 50

キーワード: 登場人物, 話者推定, 文体, 機能語

Characters, Speaker Estimation, Stylistic, Function words

1 背景

複数の登場人物が現れる小説を読む際、文章内の発話文を通じて登場人物らの人間関係や立場、話の流れなどを理解している。具体的には、登場人物らの恋愛関係や家族関係、そしてライバル関係などの描写の多くは、文章内の発話文によって表現されている。また、物語を読む際、読者は会話文の前後の文脈や、示された登場人物の属性などの情報を考え合わせることで話者を解釈して読み進めることが前提となっている。

物語における発話者と発話文を紐付けたデータは重要度が高く、さまざまな応用方法がある。応用方法の一つとして、人工知能による物語の自動生成 [1] が挙げられる。しかし、手作業で多くの物語を分析し、発話者と発話文を紐付けたデータの作成を行うには時間やコストがかかる。そこで、発話者と発話文を自動的に推定することで、時間やコストの削減ができ、データを膨大に用意できるだろう。話者推

定のアルゴリズムを研究することで、単に話者が推定できるようになるだけでなく、発話と話者の関係性とそこに影響する言語学的、文化的、社会的な背景が明らかになる。そして、物語の自動生成における登場人物の会話文自動生成に活用することができ、チャットボットなどの対話システムへの応用も可能である。また、技術的な応用だけではなく、言語学自体の発展や、国語教育における読解の仕方の指導方法の改善など人文科学的な方面への貢献も可能である。

物語における登場人物の話者推定の研究は、機械学習などのアプローチで主に英語で執筆された作品を対象に行われている。しかし、いまだ十分に高い精度が得られていない [2] という問題がある。また、日本語で執筆された物語では、研究自体がさほど行われていないという問題もある。

そこで本研究では、会話文の話者推定に有用な特徴として文体に着目し、機能語を用いて文体の推定を行うアルゴリズムを開発することを目的とした。

2 関連研究

文章の話者推定は主に英語の文章に対していくつかの方法で提唱されている。まず、話者推定タスクの初期の研究例として、Pouliquen [3] らは、ニュースの文章に対して、発話を示す動詞を利用し、その前後の名詞と発話の紐付けを行うことで推定を行った。しかし、話者が明示的ではない場合、紐付けが不可能であった。

英語の物語作品を扱った研究例として、Elson [4] らは、各物語に対して登場人物のリストを構築し、学習モデルにリストから1人を選択させることで推定を行った。また、He [2] らは、発話文の推定精度を78割まで引き上げることに成功したが、話者が比喩的表現や代名詞などで言い換えられている文や同段落内に名前などの人物情報が含まれていない文に対しての精度は5割程度であった。

日本語の物語作品を扱った研究例として、神代 [5] らは、物語テキストから登場人物同士の関係を推定し、関係図の自動構築を行う手法を提案した。その過程で、物語中に含まれる各発話文についての話者を同定し、その会話の中身から話し手同士の関係を推定した。また、Du Yulong [6] らは、物語小説における対話文のコーパスを自動構築するため、物語小説における発話文の話者を特定し、連続する台詞を話者の情報とともに対話として抽出する手法を提案した。しかし、正解率は0.72であり高いとはいえない。また、宮島 [7] によると日本語の物語では省略表現が多用されているが、英語では日本語に比べ表現を省略することが少ない。よって、日本語の話者推定の場合は英語と異なり明示的な話者情報と話者交替だけでは十分な精度を得ることができないといえる。

3 対象データ

本研究では分析対象およびデータソースとして現代日本語書き言葉均衡コーパス (BCCWJ) [8] を用いた。

BCCWJは国立国語研究所によって作成された、現代日本語の書き言葉の全体像を把握するために構築されたコーパスであり、書き言葉に関する代表的な均衡コーパスである。均衡コーパスとは、ある言語の特徴や性質を知るために、その言語の多様性をできるだけ忠実に反映するようにバランスよくサン

ブルを収集して構築される言語資源である。様々なジャンルにまたがって1億430万語のデータが格納されている。本研究ではBCCWJの図書館サブコーパス内の「文学」ジャンルのデータを用いた。このデータには1986年から2005年までの日本文学のテキストデータが計3487作品含まれている。

また、本研究では同じく国立国語研究所によって作成された、BCCWJ Speaker Information Corpus [9] も用いた。これは、BCCWJの中の日本語で書かれている物語テキストの会話文のみが抽出され、話者の性別や年齢などの情報が付与されているコーパスである。

4 方法論

4.1 ケーススタディ的分析

本研究を開始するにあたり、第2章で述べた、日本語の物語では省略表現が多用されているという考えが適切であるかの確認のため、ケーススタディ的分析を行った。その方法として、BCCWJの物語書籍のテキストからランダムサンプリングし、20の文章の発話文に対して筆者がどのような情報を元に発話者を推定しているのかを検証した。その結果、ケーススタディ分析に基づいて大きく3種類の情報を元に推定しているのではないかという仮説を立てた。以下に簡易的な説明を記す。

話者名

発話文の前後の文に発話者の名前が明示されている場合である。以下の例文では発話文の直後に「ジョン・パーリコーン」と言う話者名が明示されていること、そして「ささやく」と言う発言を意味する動詞が話者の直後に記されていることから、話者を高い確率で特定することが可能である。

例文

「まだ…歌うなよ」と、ジョン・パーリコーンがささやいた。

交替

発話者が2名以上で、発話を交替して行っている場合である。He [2] らやDu Yulong [6] らの研究では、複数の発話文が連続して出現するとき、その話者は交替することが多いと言う性質を利用して話者の特定を行った。

例文

「いいのかって?」「新橋設計事務所とのコネを、つけたがっていたんじゃないのか」「ええそれは…」「どうするんだい」

文体

発話者が2名以上で、口調(文体)が大きく異なる場合である。話者同上の関係性の推測が可能であり、以下の例文では丁寧口調の人物とそうでない人物が交替で発話していることから、二名の間に上下関係があることが推測可能である。

例文

「こっちはもう、課長にバトンタッチしちゃったよ」「え、また喋われたんですか」「大声で言うなよ」

「話者名」に着目して話者推定を行う研究に関しては、多数の先行研究 [2], [3], [6] が存在している。特に Du Yulong [6] らは以下の2パターンの手法で小説内の話者名の推定を行った。一つ目は、固有表現抽出によって「人名」として検出された単語もしくは単語列を登場人物として抽出する方法である。固有表現抽出とはテキストに出現する人名や地名などの固有名詞を抽出する技術である。二つ目は、ソーラスで人物に相当する意味クラスを持つ名詞を登場人物として抽出する方法である。また、He [2] らは発言を意味する動詞をまとめてリストを作成することで話者推定に利用した。

また、「交替」に関しては、Du Yulong [6] らが、物語小説における対話文のコーパスを自動構築する過程で、話者の交替パターンにおける発話文の話者推定を行っている。連続する台詞の話者が交互に交替すると仮定し、既に特定された明示的話者・暗黙的話者から、その前後に出現する台詞の話者を推定した。ただし、登場人物が3名以上で順不同で発話する場合や、同じ話者が間を置いて二連続で発話する場合もあり、得られる情報は推定に有用ではあるが、交替の情報のみで話者推定が十分に遂行可能なのではない。

話者が2名の場合には交互に発話するという仮定で比較的精度よく推測可能である。

一方で、日本語では話者の文体に特徴を持たせることで話者を区別する役割語 [10] が存在するが、これを用いた話者推定の研究はまだ行われていない。役割語とは、年齢や性別、性格といった話者の人物

像を想起させる特定の言葉遣いを指す言葉である。文体の差異を利用することで、発話者が3名以上で話者交替を行っている場合においても、話者を唯一に特定することが可能なことがあると考えられる。よって、本研究では「文体」に着目し、発話文文体の差異から話者推定するというアプローチを試る。

4.2 文体特徴について

物語テキストの発話文の文体特徴についての研究として村井 [11] は、物語内の発話文で頻出する助詞・助動詞の出現パターンに対して因子分析を用いて文体特徴のグループ分けを行った。因子分析とは多変量解析手法の一つであり、複数の変数を持つデータの元にある少数の要因を特定する際に用いられる。その結果主要な因子が8あることが判明した。以下に簡易的な説明を記す。(村井 [11] より一部引用)

第1因子：中立

特定の属性のない一般的な機能語による文体。

例文

お願いする

第2因子：気さく

くだけた会話的な表現による文体である。主に仲の良く気が知れた相手に用いられる。

例文

お願いするよ

第3因子：方言

「～や」「～ねん」のような現実では大阪を含む関西圏で用いられる方言である。また、漫画やフィクションなどの世界においても関西圏のキャラクターに用いられる。

例文

お願いや

第4因子：丁寧

丁寧な表現による文体である。主に目上の人間相手や関わりの浅い人間相手に用いられる。また、話者の上品さを強調する場合や、公的な場であることを示す場合にも用いられる。

例文

お願いします

第5因子：女性

近年の日常生活では主に年配の女性に用いられる「～わね」「～わよ」のような文体である。一方でフィクション内では年配に限らず若い女性にも用いられることがある。

例文

お願いするわ

第6因子：粗野

「～ぜ」「～ぞ」のような粗暴な人物や敵対関係を表現する文体である。

例文

お願いするぜ

第7因子：老人

「～じゃ」のような高齢者的な表現の言葉である。命水 [10] によると、18世紀後半以降役割割語として芝居で使われ、明示時代以降に漫画などによって世に定着した文体である。

例文

お願いじゃ

第8因子：疑問

発話中での疑問形式に対応した文体。

4.3 先行研究に基づく手法

本研究では、先行研究 [11] の8の因子から「中立」「疑問」を除く6の因子を用いて、物語の発話文から文体を自動で判別するプログラムを作成し、精度の検証を行った。「中立」「疑問」は、他の6の因子と比較して特徴的な属性がなく、物語上の登場人物の関係性の推測を行う際には不必要であると考えた。

以下の表1に検証の対象となる主要の6つの文体名と判別に用いる機能語をまとめる。

物語テキストの会話文がまとめられている BC-CWJ Speaker Information Corpus [9] をランダムサンプリングし、200程度の会話文に6つの文体のタグ付けを手動で行った。そして作成したプログラムで文体を判別し、その正解率の確認を行った。文体の判別は、文章の中に文体の特徴となる機能語が含まれているかを形態素解析エンジン MeCab [12] を用いて計測した。その結果を以下の表2に示す。

表1: 文体名と機能語

文体名	機能語
丁寧	「です」「ます」「ござる」
女性	「わ」「ね」「の」「かしら」
方言	「や」「へん」「ねん」
気さく	「じゃん」「よ」「もの」
粗野	「やがる」「ぞ」「ぜ」
老人	「じゃ」「のう」

表2: 先行研究に基づく手法による正解率

文体名	正解	不正解	データ数	正解率
丁寧	17	13	30	0.567
女性	3	8	11	0.237
方言	0	2	2	0
気さく	3	1	4	0.750
粗野	0	3	3	0
老人	1	0	1	1
合計	24	27	51	0.471

4.4 提案手法

表2より、先行研究に基づく手法によって文体判別を行った結果、正解率は5割弱となった。そこで各文体に対して、判別対象となる機能語の調整を行った。

以下の表3に検証の対象となる主要の5つの文体名と判別に用いる機能語をまとめる。

「丁寧」文体に関しては、元の3種類の機能語「です」「ます」「ござる」に「ください」を追加し、表記ゆれに対応するため「下さい」も同様に追加した。また、判別対象を終止形に限らず「でした」「ました」といった活用形も判別対象に追加した。そして、「～ですか」といった疑問の発話文も判別対象とした。

「女性」文体に関しては、元の4種類の機能語「わ」「ね」「の」「かしら」から「ね」を除外した。その理由として判別対象が「ね」のみだと、「～だよね」のようにくだけた会話的な表現が混入してしまうためである。そこで、「わね」「のね」のように複合的な文末表現に限定した。また、「わよ」「のよ」も判別対象とした。

「気さく」文体に関しては、元の3種類の機能語「じゃん」「よ」「もの」から「よ」を除外した。その理由として判別対象が「よ」のみだと、「～わよ」の

表 3: 提案手法による文体と機能語リスト

文体名	機能語
丁寧	「です」「ます」「ござる」「くだ(下)さい」「ません」「ました」「でした」「でしょう」
女性	「わ」「の」「かしら」「のよ」「わよ」「のね」「わね」
気さく	「じゃん」「もの」「～よ(のよ, わよを除く)」「～ね(のね, わねを除く)」
粗野	「やがる」「ぞ」「ぜ」
老人	「じゃ」「のう」

ように女性的な表現が混入してしまうためである。そこで「だよ」のように対象を狭めることに決定した。また、「～だね」のように語尾が「ね」の発話文はくだけた会話的な表現であるため判別対象に加えた。しかし「わね」「のね」は女性的な表現であるため、判別対象外にした。

「老人」「粗野」文体に関しては、変更を加えなかった。また「方言」文体に関しては、判別対象である機能語が「や」「へん」「ねん」のように関西弁の方言であるが、データソースである BCCWJ Speaker Information Corpus [9] への登場頻度が極めて低いことや、関西弁以外の方言への対処が困難であることから本研究では判別対象の文体から除外した。

判別対象となる機能語の調節を行った後、4.3 節と同様にして作成したプログラムによる判別を行い、精度の検証を行った。

5 検証結果と考察

4.4 節での検証結果を以下の表 4 に示す。

表 4: 提案手法による正解率

文体名	正解	不正解	データ数	正解率
丁寧	104	8	112	0.929
女性	32	8	40	0.800
気さく	26	9	35	0.743
粗野	21	5	26	0.808
老人	7	5	12	0.583
合計	190	35	225	0.844

表 2 と表 4 を比較すると、「丁寧」「女性」「粗野」文体に関しては正解率が高くなったことがわかる。また、正解率がいずれも 8 割以上であるため、高精度での判別が可能になったといえるだろう。

次に「気さく」文体に関しては正解率の変化が極めて少なかった。正解率はどちらも 7 割 5 分ほどであり、安定した文体判別ができていえるだろう。

また、「老人」文体に関しては正解率が低くなったことがわかる。しかし、先行研究に基づく手法の際は「老人」文体のデータ数が 1 であったことや、判別対象の機能語を変更していないことから、正解率が減少したのではなく今回初めて適正な正解率を得られたといえるだろう。

そして、全体的には正解率は 8 割を上回った。先行研究に基づく手法によって文体判別を行った際の正解率が 5 割弱であったことから、高精度での文体の判別が可能になったといえるだろう。

6 今後の課題

本研究では、会話文の話者推定に有用な特徴として文体に着目し、日本語物語テキストの会話文を文体に基づき人手による分類を行った。そして機能語を用いて文体の推定を行うアルゴリズムを開発した。先行研究の手法では分類の正答率は 5 割ほどであったが、提案手法では 8 割を超える結果となった。

今後の課題として、まず 4.4 節で作成した各文体の機能語リストの客観性を担保する必要がある。本研究では、機能語の割り振りを分析者一人で行ったため、現状では作成した機能語リストは主観的である。したがって評価実験により複数人による一致度の検証を行う予定である。

また、本研究では、各文体の判別を村井 [11] の研究を参考に、助詞・助動詞などの機能語を用いて推定を行った。しかし、機能語だけではなく、呼称表現からも文体判別が可能であると考えられる。例えば物語内の発話文に「バカヤロー」のような粗暴な表現が含まれている場合、その発話者は「粗野」な

人物な可能性がある。また、発話者の一人称が「ワシ」の人物は「老人」といえるだろう。これらのように特徴的な表現をまとめ辞彙を構築することで、文体判別をより高精度で行うことが期待できる。

そして、本研究で開発したアルゴリズムを応用し、発話者の文体の差異から物語の話者を推定するアプローチの検討を行う予定である。

参考文献

- [1] 松原 仁, 佐藤理史, 赤石美奈, 角薫, 迎山和司, 中島秀之, 瀬名秀明, 村井源, 大塚裕子:「コンピュータに星新一のようなショートショートを創作させる試み」, The 27th Annual Conference of Japanese Society for Artificial Intelligence, 2D1-1, 2013.
- [2] He, H. et al.: "Identification of speakers in novels", In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pp. 1312 - 1320, 2013.
- [3] Pouliquen, B. et al.: "Automatic detection of quotations in multilingual news", In RANLP, 2007.
- [4] Elson, D, K. et al.: "Automatic attribution of quoted speech in literary narrative", In AAAI, 2010.
- [5] 神代大輔, 高村大也, 奥村学:「物語テキストにおけるキャラクタ関係図自動構築」, 言語処理学会第14回年次大会発表論文集, Vol. 14, pp. 380-383, 2008.
- [6] Yulong Du, 白井清昭:「小説からの対話コーパスの自動構築」, 言語処理学会第25回年次大会発表論文集, Vol. 25, pp. 623-626, 2019.
- [7] 宮島敦子:「文章表現と会話における日本語の主語の省略」, 東京外国語大学留学生日本語教育センター論集, No. 44, pp. 133-146, 2018.
- [8] 国立国語研究所コーパス開発センター:『現代日本語書き言葉均衡コーパス』DVD版 (Version1.1), 国立国語研究所, 2015.
- [9] Makoto Yamazaki, Yumi Miyazaki and Wakako Kashino: "Annotation and Quantitative Analysis of Speaker Information in Novel Conversation Sentences in Japanese", Proceedings of the 11th International Conference on Language Resources and Evaluation, pp. 1078-1081, 2018.
- [10] 金水敏:「ヴァーチャル日本語 役割語の謎」, 岩波書店, 2003.
- [11] Hajime Murai: "Factor Analysis of Utterances in Japanese Fiction-writing Based on BCCWJ Speaker Information Corpus", Advances in Human-Computer Interaction, Vol. 2018, Article ID 5056268, 9 pages, 2018.
- [12] Kudo, Taku: "MeCab: Yet Another Part-of-Speech and Morphological Analyzer". <https://taku910.github.io/mecab/> (2021年4月2日参照)

第29回年次大会予稿

多言語分析のための感情辞書の検討

Examining Emotional Dictionary for Multilingual Analysis

大川 慎^{1*}
Shin OHKAWA^{1*}

1 名古屋市立大学大学院経済学研究科

Graduate School of Economics, Nagoya City University

〒467-0802 愛知県名古屋瑞穂区瑞穂町山の畑1

E-mail: c173651@ed.nagoya-cu.ac.jp

*連絡先著者 Corresponding Author

近年では、インターネット上において様々な言語のコンテンツが投稿されており、複数言語のテキストを対象とした分析も多く行われている。ここで、感情分析において、言語横断的な分析に利用できる感情辞書はあまりない。単一言語での分析を想定された辞書も利用できれば有益であると思われる。本研究では、辞書に含まれる感情極性値の他言語への適用について、複数言語に翻訳された小説のテキストを用いて検証する。検証した結果、一定の有効性を確認することができた。

In recent years, content in various languages has been posted, and many analyzes have been tried on texts in multiple languages. Here, there are not many sentiment dictionaries that can be used for cross-linguistic analysis in sentiment analysis. It would be useful if dictionaries intended for analysis in a single language could also be used. In this study, we examined whether the emotional polarity values contained in dictionaries were valid for other languages using novel texts translated into multiple languages. As a result, we were able to confirm a certain degree of effectiveness.

キーワード: テキストマイニング, 感情辞書, 多言語テキスト, 小説, WordNet

Text Mining, Emotional dictionary, Multilingual Text, Novel, WordNet

1 はじめに

近年では、ソーシャルメディアの普及により、インターネット上には、ロコミ等のユーザーによるコンテンツが多く投稿されている。これらは、ユーザー生成コンテ

ンツ (User Generated Content: UGC) と呼ばれ、様々な言語でデータが日々蓄積されている。推計によると、インターネット上では英語と中国語の利用者が多いが、日本語を含めたその他の言語も広く利用されて

いる。(図1) また、ビジネスにおいては、マーケティング等への利用を念頭に、これらのコンテンツの感情を分析しようとする試みが多く存在する。テキストの感情を抽出するためには、分析過程で感情極性辞書を利用することがある。本研究では、感情極性辞書の複数言語間での利用について検証を行う。

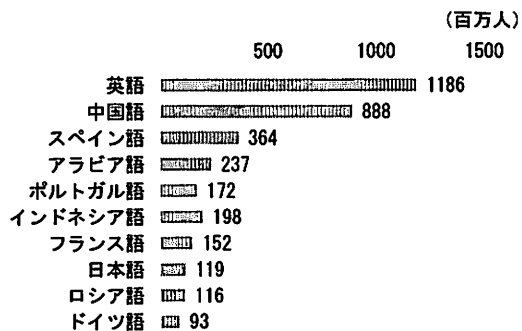


図1 言語別利用者数

(Internet World Statsより筆者作成)

2 感情分析と感情辞書

感情分析とは、「特定のエンティティに対する書き手の意見を明らかにするタスク」と定義されるテキストに含まれる感情を判断する方法論の総称である。[5] 分析手法には様々なアプローチが存在するが、テキストの感情を判断するためには、感情極性を算出する必要がある。

ここで、感情極性辞書に含まれる単語の感情極性値を利用した手法がある。例えば山元ら(2018)は、感情極性を算出する際に、日本語の感情極性辞書に含まれる感情極性値を利用している。その上で、辞書に含まれていない未知の単語に対する解決方法を提案している。[1] また、藤平(2020)は、感情極性辞書を利用した多言語テキス

トの分析手法を提案している。[2] ここでは、英語とドイツ語のテキストに対して日本語に変換し感情極性値を利用している。

日本語以外の様々な言語においても、感情極性辞書を利用した研究が行われているが、多くは英語を対象としている。したがって、他の言語では、利用できる資源が相対的に少なく制約がある状況となっている。[6] 豊富な英語の資源を他の言語にも利用できれば有益と考えられる。

ここで、英語を基にした感情極性辞書として、SentiWordNet3.0が構築されている。[7] SentiWordNetでは、WordNetのsynonym set (synset) と呼ばれる語彙概念ごとに感情極性値が付加されている。

なお、WordNetとは、Princeton大学にて心理学の研究を行うために開発された英語の概念辞書である。そして、WordNetを活用した英語以外の言語での成果も統合した「Open Multilingual Wordnet」がBondらにより構築されている。[8]

本研究では、WordNetとそれを元に構築されたSentiWordNet3.0を活用する。単語に対応するsynsetを利用し、複数言語に対して感情極性辞書の適用を試み、感情極性値の有効性を検証していく。

3 分析

3.1 分析データ

検証するデータとして、本研究では表1に示す英語を原著とする2つの小説と翻訳者による他言語のテキストを用いる。著名な作品は、多くの言語に翻訳され広く読まれているが、言語間においてテキストの持つ感情は類似しているであろう。もちろん、文単位においては、言語間で差異が存在す

るが、全体的には似た感情になると考えられる。以上の理由から、本研究において、感情極性値の他言語での有効性について検証を行う分析対象とする。

なお、短い文ではなく文書全体のようなテキストを対象とした場合、様々な感情が含まれているため、そのテキストに対する感情極性の算出が有効でないこともあり得る。この問題に対応するために、文書からアスペクト情報を抽出し、感情極性を得るアスペクトベースの感情分析も行われている。しかしながら、本研究では、複数言語における感情極性辞書の有効性を検討することを目的としている。したがって、テキストを短い段落に分割することで、各段落について単一の感情のみ含まれると見做し分析を行う方針とした。

表1 分析データ

小説		
不思議の国のアリス(ルイス・キャロル著) Alice's Adventures in Wonderland by Lewis Carroll		
バスカヴィル家の犬(アーサー・コナン・ドイル著) The Hound of the Baskervilles by Sir Arthur Conan Doyle		
言語		
英語(原著)	日本語	中国語
スペイン語	フランス語	イタリア語

3.2 分析手法

始めに、日本語の翻訳テキストに対して、500文字を目安に、小説の各章を10~20個程度の段落に分けた。そして、これを基準にし、対応する他言語のテキストについても分割した。次に、分割した小説のテキストに対して、それぞれKH coderを利用し単

語を抽出した。なお、本研究では、名詞を抽出している。ここで、単語と紐付けられたsynsetに対して、SentiWordNet3.0の感情極性値を割り当てた。感情極性値については、Positive値とNegative値に分かれているため、0から1の範囲に正規化した値の差分を最終的な段落のスコアとしている。ここで、段落における感情極性値の算出方法については、横田らの方法を参考に(1)式を用いて算出した。[3]

$$\frac{1}{tf} \sum_{i=1}^n sf_{i} \cdot score_{i} \quad (1)$$

tf : 対象テキスト内における総synset数
 sf_{i} : i 番目のsynset数
 $score_{i}$: i 番目のsynsetの感情極性値

また、辞書間の比較を行うために、日本語については、高村らの「単語感情極性対応表」を利用した分析も行った。[4] この場合においては、synsetの代わりに単語数と対応する単語の感情極性値を利用し、各段落の感情極性値を算出している。

4 結果と考察

4.1 結果

分析結果は表2及び図2の通りである。表2では、SentiWordNet3.0における言語間での相関係数を計算し関連性を見ることができる。そして、小説全体の傾向を見るために、平均値も算出している。また、図2については、一部の言語について感情極性値の推移を示したものである。日本語及び英語におけるSentiWordNet3.0の結果に加えて、単語感情極性対応表による分析結果も併せて示している。なお、図中の値は0から1の範囲に正規化している。

表2 言語間の相関行列

不思議の国のアリス							平均:	0.220
	英語	日本語	中国語	スペイン語	フランス語	イタリア語		
英語	1.000							
日本語	0.129	1.000						
中国語	-0.003	0.308	1.000					
スペイン語	0.264	0.095	0.042	1.000				
フランス語	0.406	0.241	0.150	0.240	1.000			
イタリア語	0.382	0.148	0.113	0.385	0.405	1.000		

バスカヴィル家の犬							平均:	0.380
	英語	日本語	中国語	スペイン語	フランス語	イタリア語		
英語	1.000							
日本語	0.297	1.000						
中国語	0.394	0.282	1.000					
スペイン語	0.487	0.314	0.372	1.000				
フランス語	0.505	0.171	0.262	0.324	1.000			
イタリア語	0.544	0.330	0.424	0.495	0.493	1.000		

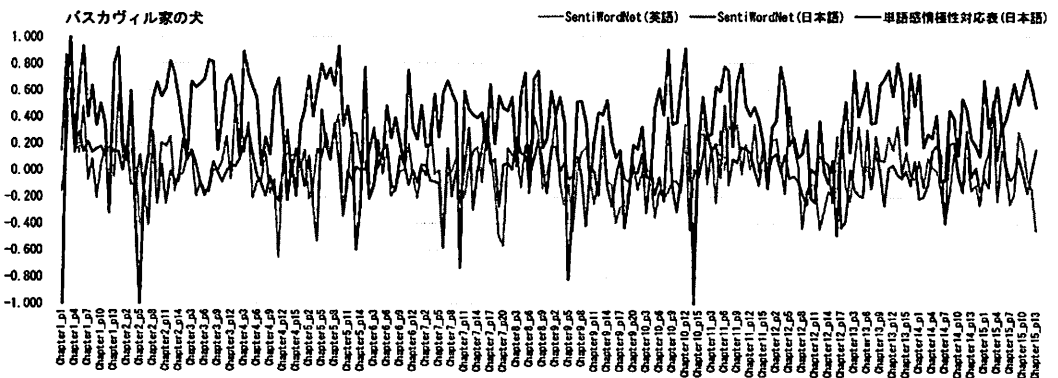
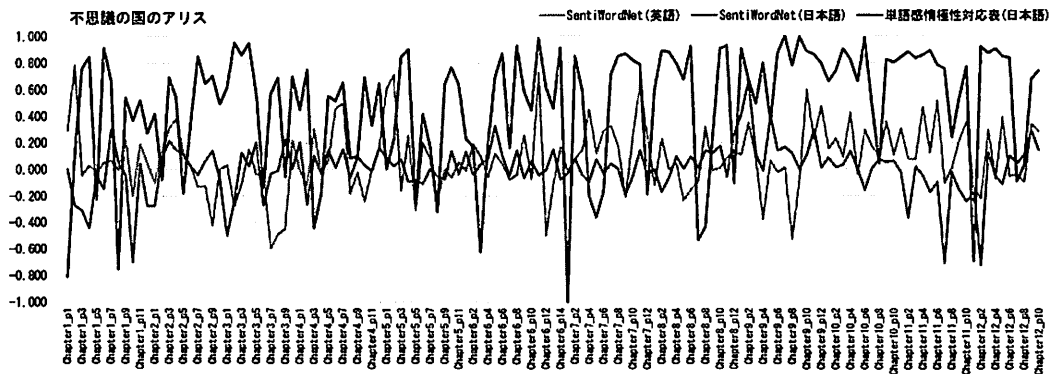


図2 感情極性値の推移

4.2 考察

まず、言語間の相関係数を見ると、小説により異なる結果を得た。「不思議の国のアリス」では、相関係数が相対的に低い傾向が見られる。一方で、「バスカヴィル家の犬」では、一定程度の相関関係が見られた。論理性が重要な推理小説においては、言語間で一定程度の関連性を見いだすことができた。これに対して、論理性を無視したナンセンス文学においては、感情の適用が困難であることが示唆される。

なお、同じテキストを複数言語での感情分析に対応しているAmazon Comprehend 及びGoogle Cloud Platformの感情分析パッケージでも分析した。Amazon Comprehend については、Positive値とNegative値の差分をスコアとして利用した。ここでも、「バスカヴィル家の犬」の言語間の相関係数が高くなっている。(表3)

表3 感情分析における相関係数

	Amazon Comprehend	Google Cloud Platform
不思議の国のアリス	0.502	0.408
バスカヴィル家の犬	0.514	0.463

また、全体的には一定程度の相関は見られたものの、英語や日本語のような構造が大きく異なる言語間では相関が低く、スペイン語とイタリア語のように構造が似ている言語間では高くなっている。

日本語については、「単語感情極性対応表」を利用した分析も行ったが、相関があるとは言えなかった。(表4) 辞書間で感情極性値の生成方法やスコアの種類は様々である。したがって、複数辞書を評価し利

用するには課題を残した。

表4 辞書間の分析結果

不思議の国のアリス	-0.030
バスカヴィル家の犬	0.108

5 おわりに

本研究では、英語の感情極性辞書であるSentiWordNet3.0を他言語に適用し、複数言語に対して感情極性値の有効性を検証した。言語間の相関係数から、言語により大きく異なっているものの、複数言語に対する一定の有効性を確認することができた。また、WordNet以外の辞書においては、他言語で利用するために翻訳等が必要となり、単語の対訳が異なると、感情極性値にも影響する可能性が考えられる。この点については今後の検討課題である。

ところで、テキストの感情をより詳細に分析するためには、辞書の感情極性値以外の情報を組み合わせるなど、別の枠組みも必要である。インターネット上の多言語のロコミ等UGCにおいては、例えばユーザー自身による評価情報等も利用することが可能である。

今後は、他の情報も考慮しながら、感情極性値の活用しつつ、UGCにおける多言語テキストの感情分析を試みる予定である。

参考文献

[1] 山元悠太；古崎晃司；駒谷和範：

「LinkedOpenDataを利用した単語極性予想によるレビュー感情分析」，言語処理学会，第24回年次大会 発表論文集，pp. 340-343, 2018.

[2] 藤平啓汰：「感情辞書を用いたWebテキストの多言語感情抽出」，情報処理学会第82回全国大会講演論文集，2020.

[3] 横田尚己；山田圭二郎：「熊本地震のつぶやきに見る感情極性値の時空間解析」，日本都市計画学会，都市計画論文集，Vol. 52, No. 3, pp. 1081-1087, 2017.

[4] 高村大也；乾孝司；奥村学：「スピンモデルによる単語の感情極性抽出」，情報処理学会論文誌，Vol. 47, No. 2, pp. 627-637, 2006.

[5] ACM: “Communications of the ACM”, Vol.56, No.4, pp.82-89, 2013.

[6] K Dashtipour; S Poria; A Hussain; E Cambria; A Y. A. Hawalah; A Gelbukh; Q Zhou: ‘Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques’, Cogn Comput 8, pp757-771, 2016.

[7] Stefano Baccianella; Andrea Esuli; Fabrizio Sebastiani: ‘SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining’, Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10), pp2200-2204, 2010.

[8] Francis Bond; Kyonghee Paik: ‘A survey of wordnets and their licenses’, In Proceedings of the 6th Global WordNet Conference (GWC 2012). Matsue, pp64-71, 2012.

第29回年次大会予稿

女性就業者数促進のための計量分析アプローチの研究

Study about Quantitative Analysis Approach for the Promotion of the Number of Female Workers

植杉京佳^{1*}, 大槻明²

Kyoka UESUGI^{1*}, Akira OTSUKI²

1 日本大学経済学部

Nihon University, College of Economics

〒101-8360 東京都千代田区三崎町一丁目3番2号

E-mail: ecky18264@g.nihon-u.ac.jp

2 日本大学経済学部

Nihon University, College of Economics

〒101-8360 東京都千代田区三崎町一丁目3番2号

E-mail: otsuki.akira@nihon-u.ac.jp

*連絡先著者 Corresponding Author

内閣府男女共同参画局の資料でも示されているように、現在日本では、女性活躍に資する働き方改革の推進が国を主体となって進められている。そこで本研究では、女性就業者数が多い地域と少ない地域の特徴的な違いや共通性を明らかにするための計量アプローチについて提案する。実証実験では、e-statで公開されている、「労働」、「福祉・社会保障」データを用いて試行的に本提案アプローチを検証した。

In Japan, it is ongoing that promote work style reform now for successful women by Cabinet Office. So, this study proposed the quantitative analysis approach for reveal some differences and common points each region. We verified the effects of this approach in evaluation experiment using e-stats data (Labor Statistics Data, Social Security Statistics Data).

キーワード：女性の職業生活における活躍の推進，計量分析

Keywords: Promotion of Women's Participation and Advancement in the Workplace, Quantitative Analysis

1. はじめに

近年、コロナ禍によるテレワークの普及や資格取得のための支援や各種制度など、様々な取り組みが各社、自治体でなされており、働き方改革による働き方の見直しが注目されている。

その中でも特に、内閣府男女共同参画局[1]の資料で示されているように、女性活躍に資する働き方改革の推進が、国が主体となって進められている。このことから本研究では、女性就業者数が多い地域と少ない地域の特徴的な違いや

共通性を明らかにするための計量アプローチについて研究した。実証実験では、e-statで公開されている、「労働」、「福祉・社会保障」データを用いて試行的に本提案アプローチを試行した。

2. 先行・関連研究

高城ら[1]は、女性就業者の就業意欲と職業観の変化について計量アプローチから分析を行うために、資料収集を行った。具体的には、就業以前に持っていた職業観と現在の就業状況や職業観について、社会人女性7名を対象にインタビュー調査を行い、得られたデータから「離職・転職理由」と「学生時代の就業意欲と現在の満足度・就業意欲の関係」に着目して検討を行った。

泉[2]は、山梨県内の非合併山村自治体を対象として、全国平均及び県平均の女性就業率と比較検証することにより、対象自治体において、第2次産業の女性就業者が減少していることと、第3次産業の女性就業者が増加していることを明らかにした。

高城らの研究から、女性就業者数の促進というテーマ自体の重要性がうかがえるが、この研究では計量分析の前段階の作業にとどまっている。泉は計量分析までを行っているが、単純な女性就業率の比較にとどまっている。

対して、松本ら[3]は、e-sat等から都道府県単位で労働環境（男女比、有給消化率、非正規雇用割合、所定内労働時間、通勤時間）のデータを取得し、機械学習のアプローチからこれらのデータを分析することで、女性就業率の低い都道府県と高い都道府県を明らかにしている。しかし、松本らの分析は都道府県レベルでの分析にとどまっており、さらに女性就業率の低い都道府県と高い都道府県それぞれの理由や要因までは明らかにされていない。

ゆえに本研究では、松本の知見をさらに自治

体レベルに掘り下げる形で分析を行うとともに、女性就業者数が多い地域と少ない地域の特徴的な違いや共通性を明らかにするための計量アプローチについて研究する。

3. 提案アプローチ

3.1. 分析対象データ

分析対象データは e-stat (<https://www.e-stat.go.jp/>) から取得する。具体的には、表1に示す通り、「人口」、「労働」、「福祉・社会保障」、「教育」関係データをそれぞれ取得した。取得したデータのイメージを表2に示す。なお、本稿では誌面の都合上、「労働」、「福祉・社会保障」データを用いた分析を試行したので、この結果について述べる。

3.2. 分析アプローチ

3.2.1. クロス集計

松本が述べる女性就業率の高い都道府県（青森、山形、徳島、高知、長崎、熊本、宮崎、鹿児島）と低い都道府県（愛知、茨城、三重、滋賀）に分けたうえで行サンプルに設定し、表1の各データを列変数に設定することで分析対象データ（モデル）を準備する。このイメージを表3に示す。なお、表3のデータは、例えば表2列目のように、女性の就業数を各データ項目の人数で割ることにより、各データ項目の女性比率に変換することで準備した。

そして、表3を元に都道府県レベルでクロス集計を用いて分析することにより特徴的なデータセットを抽出し、これらのデータを自治体レベルにまで掘り下げて抽出したうえで次節のクラスタリング（機械学習）を行う。ここでいう特徴的なデータセットとは、例えば、表3の8列目（第1次産業に従事する就業者数割合）は、女性就業率の高い都道府県（青森、山形、徳島、高知、長崎、熊本、宮崎、鹿児島）は、女性就業率の低い都道府県（愛知、茨城、三重、滋賀）よりも相対的に高い。このような特徴的なデータセットをいくつか抽出したうえで次節のクラスタリングを行う。

表4から、女性就業割合の高い都道府県の平均値と低い都道府県の平均値の開きについて第3四分位点を参考閾値として考える。例えば表4であれば、全列項目の「ave2 - ave」行の第3四分位点は「約3.1%」であるため、この値を参考に列項目を抽出（ここでは3.1%よりも大きい列項目を抽出）すると、抽出される項目は4列目、8～10列目、13～14列目となる。これらの特徴をまとめると次のようになる。

- ・ (8～10列目) 女性就業割合の高い都道府県は、低い都道府県よりも第一次、第三次産業に就いている割合が相対的に高い。対して女性就業割合の低い都道府県は、高い都道府県よりも第二次産業に就いている割合が相対的に高い。
- ・ (13～14列目) 全女性就業者数に占める自市区町村で就業している女性就業者数の割合は、女性就業割合の高い都道府県の方が低い都道府県に比べて相対的に高い。逆に、他県で従事している女性就業者数の割合は、女性就業割合の低い都道府県の方が

高い都道府県に比べて相対的に高い。

- ・ (4列目) 就業者数に占める家事を仕事としている人数の割合は、女性就業割合の低い都道府県は、高い都道府県よりも相対的に高い。

4.1.2. 福祉・社会保障データ

福祉・社会保障データのクロス集計の結果を表5に示す。4.1.1節と同様に第3四分位点を算出し、この値を参考にし得られた特徴を次のとおりまとめる。

- ・ 女性就業者数に占める「有料老人ホーム数(8列目)」の割合は、女性就業割合の低い都道府県に比べて高い都道府県の方が相対的に高い。
- ・ 女性就業者数に占める「児童福祉施設等数(9列目)」及び「保育所等数(10, 11列目)」の割合は、女性就業割合の低い都道府県に比べて高い都道府県の方が相対的に高い。

表4. クロス集計の結果(労働データ)

	労働人口		就業者数・家事 の割合(仕事 (女)/就業者 数(%)		専業主婦割合 (女)/就業者 数(%)		自営労働人口 (女)/労働人口 (%)		非労働人口・ 学生(女)/労働 人口(%)		第1次産業就業者数(女)/就業者数(%)			第2次産業就業者数(女)/就業者数(%)			第3次産業就業者数(女)/就業者数(%)			雇用者数(正規 パート)/全労働 人口(%)		雇用者数(パート /正の割合)		自営労働で就業している就業者数(女)/就業者数(%)		他県で就業している就業者数(女)/就業者数(%)	
	(女)/労働人口(%)	就業者数(女)	(女)/就業者数(%)	専業主婦割合(女)	自営労働人口(女)	労働人口(%)	第1次産業就業者数(女)	第2次産業就業者数(女)	第3次産業就業者数(女)	雇用者数(正規)	雇用者数(パート)	パート/正の割合	自営労働で就業している就業者数(女)	他県で就業している就業者数(女)													
全国	41.9%	42.1%	15.7%	34.0%	64.8%	37.9%	3.0%	14.6%	58.3%	64.8%	31.4%	57.6%	1.3%														
宮城	41.8%	42.3%	13.8%	31.7%	64.0%	35.5%	6.7%	30.2%	59.3%	65.0%	31.5%	59.5%	7.0%														
三重	44.2%	44.6%	14.7%	33.1%	64.4%	35.2%	4.7%	29.1%	63.1%	63.0%	33.4%	77.0%	2.9%														
岩手	42.6%	42.9%	15.0%	33.3%	64.9%	34.1%	3.0%	34.0%	59.7%	63.5%	32.2%	62.9%	5.7%														
ave	42.6%	43.0%	14.7%	33.2%	64.5%	35.7%	4.4%	32.0%	60.3%	64.1%	32.1%	64.4%	4.23%														
青森	45.4%	45.6%	10.8%	35.7%	65.2%	32.3%	12.9%	20.1%	63.8%	68.0%	29.3%	79.8%	0.2%														
山形	44.7%	45.1%	10.6%	34.5%	64.5%	30.2%	10.2%	30.2%	57.8%	71.1%	25.9%	74.3%	0.5%														
福島	44.9%	45.4%	12.0%	31.8%	63.4%	31.0%	8.3%	23.0%	63.1%	70.0%	27.7%	63.7%	1.1%														
茨城	46.1%	47.0%	10.7%	33.5%	61.2%	28.3%	15.1%	16.8%	65.2%	66.3%	32.1%	75.8%	0.4%														
新潟	44.8%	45.3%	11.4%	33.3%	65.2%	30.7%	14.3%	19.4%	64.4%	67.6%	32.0%	87.3%	1.0%														
栃木	46.3%	45.9%	11.2%	33.5%	64.1%	29.7%	11.5%	22.5%	64.9%	66.7%	30.9%	73.0%	2.7%														
群馬	46.4%	47.0%	11.2%	34.0%	64.0%	29.3%	15.1%	21.5%	62.0%	67.1%	30.9%	87.5%	1.3%														
埼玉県	49.1%	49.6%	12.8%	34.1%	64.8%	46.3%	13.1%	21.3%	64.7%	65.4%	53.3%	83.9%	1.9%														
千葉県	44.2%	44.9%	9.8%	34.2%	62.8%	51.3%	5.4%	14.1%	72.4%	61.7%	35.4%	60.5%	0.1%														
ave2	45.4%	45.9%	11.2%	33.5%	64.1%	34.4%	12.0%	21.2%	64.3%	67.1%	30.4%	78.9%	0.976%														
ave2 - ave	2.8%	3.1%	-3.5%	3.3%	-0.4%	-2.3%	7.6%	-10.4%	4.1%	3.0%	-1.3%	12.5%	-3.27%														

表5. クロス集計の結果(福祉・社会保障データ)

	児童福祉施設等数(正規)		児童福祉施設等数(パート)		児童福祉施設等数(パート)		児童福祉施設等数(パート)		児童福祉施設等数(パート)		児童福祉施設等数(パート)		児童福祉施設等数(パート)		児童福祉施設等数(パート)		児童福祉施設等数(パート)		児童福祉施設等数(パート)	
	児童福祉施設等数(正規)	児童福祉施設等数(パート)	児童福祉施設等数(パート)	児童福祉施設等数(パート)	児童福祉施設等数(パート)	児童福祉施設等数(パート)	児童福祉施設等数(パート)	児童福祉施設等数(パート)	児童福祉施設等数(パート)	児童福祉施設等数(パート)	児童福祉施設等数(パート)	児童福祉施設等数(パート)	児童福祉施設等数(パート)	児童福祉施設等数(パート)	児童福祉施設等数(パート)	児童福祉施設等数(パート)	児童福祉施設等数(パート)	児童福祉施設等数(パート)	児童福祉施設等数(パート)	児童福祉施設等数(パート)
全国	0.000%	0.018%	0.063%	0.055%	0.006%	0.009%	0.026%	0.127%	0.087%	0.050%	0.013%									
宮城	0.001%	0.020%	0.023%	0.039%	0.008%	0.008%	0.014%	0.106%	0.090%	0.025%	0.023%									
三重	0.000%	0.023%	0.067%	0.044%	0.009%	0.005%	0.033%	0.147%	0.128%	0.067%	0.015%									
岩手	0.002%	0.018%	0.062%	0.025%	0.007%	0.010%	0.005%	0.121%	0.095%	0.044%	0.014%									
ave	0.001%	0.023%	0.064%	0.031%	0.006%	0.006%	0.020%	0.125%	0.100%	0.049%	0.016%									
青森	0.001%	0.027%	0.065%	0.031%	0.007%	0.015%	0.092%	0.212%	0.166%	0.005%	0.034%									
山形	0.002%	0.021%	0.069%	0.035%	0.003%	0.006%	0.053%	0.141%	0.094%	0.023%	0.015%									
福島	0.003%	0.058%	0.039%	0.045%	0.027%	0.022%	0.024%	0.184%	0.136%	0.091%	0.033%									
茨城	0.009%	0.027%	0.059%	0.044%	0.306%	0.005%	0.080%	0.241%	0.202%	0.254%	0.034%									
新潟	0.000%	0.037%	0.017%	0.043%	0.011%	0.016%	0.044%	0.231%	0.199%	0.030%	0.033%									
栃木	0.001%	0.038%	0.019%	0.038%	0.010%	0.017%	0.034%	0.200%	0.173%	0.032%	0.034%									
群馬	0.009%	0.039%	0.018%	0.044%	0.010%	0.016%	0.034%	0.200%	0.183%	0.029%	0.030%									
埼玉県	0.000%	0.041%	0.017%	0.046%	0.009%	0.016%	0.052%	0.195%	0.163%	0.011%	0.020%									
千葉県	0.000%	0.017%	0.024%	0.019%	0.009%	0.008%	0.097%	0.193%	0.150%	0.027%	0.031%									
ave2	0.002%	0.035%	0.023%	0.040%	0.012%	0.009%	0.075%	0.200%	0.163%	0.261%	0.034%									
ave2 - ave	0.001%	0.012%	0.018%	0.045%	0.011%	0.003%	0.055%	0.075%	0.063%	0.213%	0.017%									

4.2. クラスタリングの結果

クラスタリングの結果, 図1に示すように11クラスタに分割された.

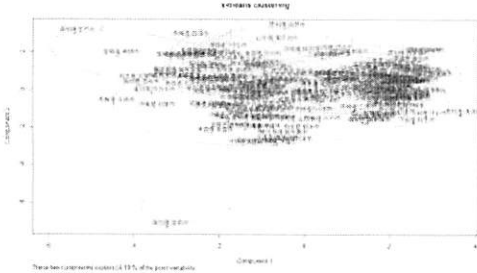


図1. クラスタリングの結果

11クラスタの特徴を表6に示す. まず, クラスタid4 (以下クラスタ4) とクラスタ6は, 「女性割合低 = 0」のため, 「高 / 低」列には便宜的に100.00を設定した. そして, 全クラスタのメンバ数平均は18.6であるため, この閾値未満のクラスタは分析対象外とした. ここで, 女性就業者数割合の高い自治体が最も集まっているクラスタはクラスタ2であり, 女性就業者数割合の低い自治体が最も集まっているクラスタはクラスタ9であったため, これら2クラスタを取り上げて特徴の考察を行う.

表6. 11クラスタの特徴

id	メンバ数	女性割合低	女性割合高	高/低
1	47	7	40	5.71
2	21	1	20	20.00
3	23	17	6	0.35

表7. クラスタ2の特徴

	家事	第1次産業	第2次産業	第3次産業	自市区町村	他県で就業	有老人ホーム数	児童福祉施設等数	保育所等数	公設保育所等数
三重県 鳥羽市	0.1468	0.14	0.17	0.6576	0.79	0	0.0006	0.0019	0.0019	0.0019
青森県 弘前市	0.1085	0.14	0.16	0.6328	0.88	0	0.0011	0.0024	0.0015	0
青森県 五所川原市	0.1231	0.14	0.2	0.6301	0.77	0	0.0012	0.0017	0.0017	0.0001
高知県 室戸市	0.1319	0.2	0.17	0.6215	0.88	0.01	0	0.0048	0.0044	0.002
高知県 須崎市	0.0953	0.19	0.18	0.591	0.81	0	0	0.0022	0.002	0.0004
高知県 高毛市	0.104	0.15	0.18	0.6657	0.81	0.02	0.0013	0.0035	0.0026	0.0022
長崎県 島原市	0.1	0.15	0.19	0.641	0.85	0	0.0005	0.0024	0.0021	0.0001
長崎県 平戸市	0.126	0.2	0.18	0.6098	0.89	0	0.0001	0.0028	0.0026	0.0004
長崎県 対馬市	0.1153	0.2	0.13	0.6693	0.99	0	0	0.0018	0.0017	0.0013
長崎県 壱岐市	0.121	0.2	0.15	0.6449	0.99	0	0.0019	0.0012	0.0012	0.001
長崎県 南島原市	0.1179	0.24	0.2	0.5613	0.84	0	0.0004	0.0023	0.0022	0.0001
熊本県 阿蘇市	0.0932	0.18	0.22	0.5978	0.89	0	0.0002	0.0019	0.0019	0.0006
宮崎県 小林市	0.1167	0.21	0.19	0.5822	0.87	0	0.0016	0.0021	0.0017	0.0005
宮崎県 串間市	0.1033	0.27	0.15	0.5742	0.87	0.06	0.001	0.0031	0.0031	0.0005
宮崎県 えびの市	0.1317	0.23	0.2	0.5559	0.83	0.02	0.0005	0.0014	0.0014	0
鹿児島県 阿久根市	0.1251	0.14	0.26	0.5922	0.81	0	0.0002	0.0019	0.0017	0.0002
鹿児島県 指宿市	0.1436	0.23	0.13	0.6389	0.94	0	0.0007	0.0017	0.0015	0.0001
鹿児島県 西之表市	0.1312	0.26	0.12	0.6157	0.97	0	0.001	0.0023	0.0018	0
鹿児島県 垂水市	0.131	0.19	0.23	0.5828	0.84	0	0.0003	0.0021	0.0021	0
鹿児島県 伊佐市	0.1302	0.18	0.24	0.5749	0.91	0.01	0.0003	0.0026	0.0023	0
沖縄県 宮古島市	0.1146	0.18	0.13	0.6286	0.96	0	0.0008	0.0028	0.002	0.0009
ave	0.1195	0.1914	0.1800	0.6128	0.8752	0.0057	0.0006	0.0024	0.0021	0.0006
ave / total_ave	0.9456	1.7589	0.7457	0.9924	1.3141	0.1744	0.9722	1.3950	1.4899	0.2036

4	8	0	8	100.00
5	5	3	2	0.67
6	6	0	6	100.00
7	55	43	12	0.28
8	8	1	7	7.00
9	19	18	1	0.06
10	6	5	1	0.20
11	7	2	5	2.50
ave	18.6			

表7にクラスタ2の特徴を, 表8にクラスタ9の特徴をそれぞれ示す. ここでいう特徴とは, 全クラスタ (1~11) における表4, 5で示した元データ (205自治体のデータ) の平均値に占める, クラスタ2, 9の同データの平均値の割合を算出したものである. 「ave」がクラスタ2, 9それぞれの平均値, 「ave / total_ave」がクラスタ全体の平均値でクラスタ2, 9の平均値をそれぞれ割ったスコアを表す. 「ave / total_ave >= 1.0」を閾値として特徴を見た場合, クラスタ2は次の項目が全国 (205自治体) 平均よりも高いという特徴が明らかとなった.

- ・ 第一次産業に就いている割合
- ・ 自市区町村に就業している女性割合
- ・ 女性就業者数に占める「児童福祉施設等数」及び「保育所等数」の割合

同様にクラスタ9で全国 (205自治体) 平均よりも高い項目は次の通りであった.

- ・ 家事に従事している割合
- ・ 第二次, 第三次産業に就いている割合
- ・ 他県で就業している割合

表8. クラスタ9の特徴

	家事	第1次産業	第2次産業	第3次産業	自市区町村	他県で就業	有料老人ホーム数	児童福祉施設等数	保育所等数	公共保育所等数
愛知県津島市	0.1761	0.02	0.29	0.6453	0.5	0.01	0.0002	0.001	0.0008	0.0001
愛知県江南市	0.1671	0.01	0.32	0.6272	0.5	0.03	0.0003	0.001	0.0009	0.0008
愛知県知多市	0.1643	0.02	0.34	0.6025	0.46	0	0.0001	0.0009	0.0008	0.0007
愛知県尾張旭市	0.1674	0.01	0.26	0.7027	0.4	0	0.0008	0.0017	0.0009	0.0005
愛知県岩倉市	0.1516	0.01	0.29	0.6368	0.37	0.01	0.0001	0.0019	0.0011	0.0007
愛知県日進市	0.1618	0.01	0.28	0.6708	0.4	0	0.0006	0.0012	0.0007	0.0005
愛知県常陸市	0.1803	0.08	0.29	0.6017	0.44	0.03	0.0003	0.0015	0.0009	0.0003
愛知県清海市	0.1587	0.01	0.29	0.6406	0.41	0.01	0	0	0	0
愛知県北名古屋	0.1622	0.01	0.31	0.6421	0.48	0.01	0.0001	0.0014	0.0008	0.0008
愛知県弥富市	0.1633	0.04	0.28	0.6264	0.5	0.09	0.0005	0.0017	0.001	0.0009
愛知県あま市	0.1722	0.02	0.32	0.6029	0.43	0.01	0.0001	0.001	0.0007	0.0005
愛知県鶴巻市	0.1158	0.04	0.3	0.6325	0.49	0.08	0	0.0017	0.0015	0.0002
愛知県新城市	0.1461	0.06	0.24	0.6451	0.43	0	0.0002	0.0009	0.0006	0.0001
愛知県かすみがうら市	0.1321	0.11	0.31	0.5656	0.46	0.02	0.0002	0.0014	0.0011	0.0004
愛知県つくばみらい市	0.1418	0.04	0.29	0.6239	0.36	0.14	0.0001	0.0018	0.0014	0.0006
愛知県津市	0.1396	0.01	0.3	0.6356	0.56	0.11	0	0.001	0.0007	0.0002
愛知県守山市	0.1547	0.02	0.32	0.6315	0.53	0.07	0.0001	0.0011	0.0008	0.0003
愛知県東海市	0.1495	0.02	0.32	0.6259	0.44	0.07	0.0001	0.0017	0.001	0.0006
熊本県合志市	0.1241	0.05	0.26	0.6728	0.41	0	0.0003	0.0022	0.0015	0
ave	0.1541	0.0311	0.2951	0.6233	0.4511	0.0361	0.0002	0.0013	0.0009	0.0004
ave / total ave	1.2193	0.2853	1.2232	1.0256	0.6772	1.1087	0.3733	0.7808	0.6526	0.1500

5. 分析結果の考察

前章の分析結果を受けて考察を行う。労働データの観点からは、女性就業割合の高さと第一次産業との親和性が高く、自市区町村で就業している女性就業者数の割合が高いことが明らかとなった。農林水産省の「女性の農業における活躍推進に向けた検討会[4]」の報告書の中でも、女性の参画が農業経営にプラスになるという提言がなされており、また、三原[5]も女性起業という観点から考察していて、特に農村女性起業の伸びが顕著である（1997年から6年間で2倍に伸びている）ことを報告している。ゆえに、女性就業者数促進のための一方策として、農業の活性化が考えられる。しかし、本研究の主なコンセプトは分析（計量）アプローチを提案することであり、分析の結果得られた知見の有効性検証については別の研究や考察が必要である。一例を示すと、そもそも農業に必要な土地をどう準備するのか、といった問題や、ただ農業を活性化させるだけではなく、仁平[6]も指摘しているように、生産された農作物を販売するためのネットワークの構築支援やマーケティング支援も重要であるため、実際に農業を活性化させるためには別の研究や考察が必要であると考えられる。

一方、福祉・社会保障データの観点からは、女性就業割合の高い都道府県では、女性が安心して働くために、子育ての負担を軽減するための保育関係の環境整備が重要である旨示唆された。

6. おわりに

本研究では、女性就業者数が多い地域と少ない地域の特徴的な違いや共通性を明らかにするための計量アプローチについて提案した。そして、実証実験では、e-statで公開されている、「労働」、「福祉・社会保障」データを用いて試行的に本提案アプローチの検証を行った。今後は、様々なデータで本提案アプローチを実証することで、女性活躍の促進に資する知見導出を目指していきたい。

参考文献

- [1] 高城佳那, 山田一之, 太田さつき, 久保田貴之, 漁田武雄, 日隈美代子: 女性就業者の就業意欲と職業観の変化について—予備的検討—, 環境と経営, 第25巻, 第1号, pp. 145-151, 2019年.
- [2] 泉桂子: 山梨県内の非合併山村自治体における高齢化・町村財政・女性就業率の推移—水源林地元村に着目して—, 林業経済 69 (3), pp. 1-16, 2016.
- [3] 松本大地, 大槻明: 機械学習を用いた労働環境分析から考察する地域間格差とその解決策に関する研究, 2018年度社会情報学会関東支部研究発表会.
- [4] JAcom: 「女性の力農業経営にプラス—女性活躍で検討会が報告書」, <https://www.jacom.or.jp/nousei/news/2020/12/201225-48554.php> (2021/3/8閲覧済み).
- [5] 三原育子: 農村における女性起業の経営的性格と課題, 国立女性教育会館研究紀要 vol. 9, pp. 73-83, 2005.
- [6] 仁平章子, 伊庭治彦: 女性農業者の六次産業への取り組みに関する一考察—ネットワークの機能を視点として—, 農林業問題研究, 第 196号, pp.15-20, 2014.

第 29 回年次大会予稿

PAC 分析を利用したオタクの「布教」に対する 意識や態度についての研究

A study of attitudes and awareness of Otaku's "Fukyo" using PAC analysis

細川 怜椰^{1*}, 松村 敦², 宇陀 則彦², 堤 智昭³

Reiya HOSOKAWA^{1*}, Atsushi MATSUMURA², Norihiko UDA², Tomoaki TSUTSUMI³

1 筑波大学 大学院 人間総合科学学術院 人間総合科学研究群 情報学学位プログラム

Graduate Program in Informatics; Graduate School of Comprehensive Human Sciences; University of Tsukuba

〒 305-8550 茨城県つくば市春日 1-2

E-Mail: s2021688@s.tsukuba.ac.jp

2 筑波大学 図書館情報メディア系

Faculty of Library, Information and Media Science, University of Tsukuba

E-Mail: matsumur@slis.tsukuba.jp, uda@slis.tsukuba.ac.jp

3 筑波大学 人文社会系

Faculty of Humanities and Social Sciences, University of Tsukuba

E-Mail: tsutsumi.tomoaki.gn@u.tsukuba.ac.jp

* 連絡先著者 Corresponding Author

本研究ではオタクの定義を「特定の趣味に深くのめりこむ人々」とし、特に「アニメ・漫画にのめり込む人々」に焦点を当てる。「オタク文化」の一つに、「布教」と呼ばれる行為がある。「布教」を推薦行為と捉えた研究を行いたい、「布教」そのものがまだ明らかになっておらず、基礎研究が必要である。本研究では、「布教」の実体を明らかにする足がかりとして、オタクが「布教」を行う際の意識や態度を明らかにすることを目的とした。手法には PAC 分析を用いて、2020 年 9 月から 2021 年 3 月までに 6 名の協力者に調査を行った。本稿ではそのうち 3 名についての分析結果を報告する。分析結果より、「布教」において布教相手の存在を意識していることがわかった。意識の方向性は「相手の好み」のように共通している部分もあったが、三者三様であった。

In this study, the definition of otaku is "people who are deeply absorbed in a particular hobby," with a particular focus on people who are absorbed in anime and manga. One of the aspects of otaku culture is the act of "Fukyo". However, "Fukyo" itself has not yet been clarified, and basic research is needed. The purpose of this study is to clarify the attitudes and awareness of otaku in the process of "Fukyo" as a stepping stone to clarifying the reality of "Fukyo". PAC analysis was used as the method, and a survey was conducted on six participants from September 2020 to March 2021. In this paper, we report the results of the analysis of three of them. The results of the analysis show that they were aware of the existence of their Fukyo partner in their "Fukyo". The direction of their awareness is different among the three, although there are some common aspects such as "preference of the partner".

キーワード: オタク文化, 布教, 個人別態度構造分析, 情報推薦

Keywords: otaku culture, Fukyo, Personal Attitude Construct-Analysis, recommendation

1 はじめに

現在、日本ではオタクと呼ばれる人々が広く認知されている。オタクは1980年代ごろから存在しているが、その表記がひらがなの「おたく」からカタカナの「オタク」、そして近年では「ヲタク」に変化しつつあるように、意味する対象も時代とともに変化しつつある。例えば東によると、オタクは「コミック、アニメ、ゲーム、パーソナルコンピュータ、SF、フィギュアそのほか、互いに深く結びついた一群のサブカルチャーに耽溺する人々の総称」^[1]と定義されている。しかし、相田^[2]が指摘するように、オタクを定義づける論点は時代や論者によって全く異なる。加えて昨今では、“美容オタク”や“筋肉オタク”など、必ずしもサブカルチャーに属さないような事柄に関してもオタクという言葉が用いられており、オタクの定義は曖昧さを増している。したがって、本研究ではオタクを「特定の趣味に深くのめりこむ人々」と定義し、中でも特に「アニメ・漫画にのめり込む人々」に着目する。

オタクたちが独自に形成した文化は「オタク文化」と呼ばれる。例えば、作品の舞台となった場所を訪れる「聖地巡礼」やコミックマーケットのような「同人誌即売会」は典型的なオタク文化と言える。このオタク文化は永田^[3]が指摘するように、オタクでない人にはその価値がわからず、オタクの中だけで高文脈化している。したがって、その独自性や特異性を明らかにすることは、オタク文化の真の価値を示すとともに、オタクでない人々に新しい価値を提供するという意味をもつ。事実、オタク文化は様々な分野から注目を集め、研究がなされてきた^{[4][5]}。

本研究で着目する「布教」も、このオタク文化の一つである。オタク文化における「布教」は、オタクが自らの好きなものを他者に教え広める行為であると考えられている。しかし、オタク文化として高文脈化された「布教」が一体どのような行為であるのか、その実体を明らかにし新しい価値を提示した研究はまだない。例えばこの「布教」を一種の推薦行為であると捉えることができれば、情報推薦の視点から「布教」の価値を世に示せるのではないかと。ところが、一般的な製品推薦、Eコマースにおける類似製品の推薦やスマートフォンのアプリ推薦に関する研究のように、布教を分析する基礎となるような、布教そのものに着目した研究は報告されていない。まずは「布教」の実体を明らかにしなければ、

他の推薦行為との比較研究もままならない。

そこで本研究では、オタクが「布教」を行う際の意識や態度を明らかにすることを目的とする。「布教」という未開拓の視点において初めから「布教」そのものに着目するのではなく、まずは「布教」を行う人々の視点を切り口とすることで、オタクの意識や態度から「布教」の手がかりを得られると考えられる。したがって本稿では、予備的な調査により手法として妥当だと判断したPAC分析を用いて、「布教」を行うオタクの意識や態度に迫る。

本稿ではこれ以降、特に強調する場合以外、布教にカギカッコはつけない。

2 関連研究

オタク文化に着目した研究はこれまでに多数報告されている。例えば、オタク文化を生み出す日本のマンガ・アニメ・特撮を対象にその文化記号論的解釈を試みた出原^[6]の研究や、オタク文化の経済的価値を考察した山口ら^[7]の報告などが挙げられる。他にも、和田^[4]の研究では、東京・秋葉原や大阪・日本橋のような「オタクの街」における、オタク文化の集積過程とまちづくりのプロセスを明らかにしている。また、入江^[5]は、聖地巡礼するオタクに聞き取り調査を行うことで、彼らの「ゲスト」としての実態を明らかにしている。これらの研究はいずれもオタク文化に着目しているが、「布教」を対象としていない。オタクおよびオタク文化を宗教社会学と結びつけ「聖地巡礼」や「同人誌展示即売会」などを論じた今井^[8]の一連の研究においても、「布教」は取り上げられていない。

ただし、布教が「自らの好きなものを他者に教え広める行為」であるとする、そこには布教するオタクと布教される他者が存在することになる。その点を考慮すると、永井ら^[9]が行ったオタク同士によるコミュニティ形成とその維持についての報告から、オタクが布教によってコミュニティを形成している可能性が考えられる。しかし、この報告の中でも布教についての言及はされていない。

ここで、笹倉^{[10][11]}の研究や枝井ら^[12]の研究に着目する。笹倉は、クライアントがアニメや漫画について他者に語る行為について臨床心理学の視点からプロセス化を試み、その結果、語るに至る背景にある行動やその心情を考察するに至っている。布教と“語る行為”は同列ではないものの、どちらもア

アニメや漫画に影響を受けて他者へと発信している行為である。また枝井は、中学生や高校生がマンガ・アニメ・ゲームに没頭することが思春期の彼らに与える影響について分析している。対象がアニメ・マンガ・ゲームに触れている点は本研究と共通している。このように心理学分野においては、漫画やアニメについて語る行為を質的調査でプロセス化する試みがなされている。

これらの研究より、本研究においてオタクの「布教」に着目することは、オタク文化の研究に新たな視点をもたらすという点において意義があると考えられる。また、質的調査によって行為そのものに焦点を当てる試みが存在することから、今回は PAC 分析と呼ばれる個人の態度に着目する手法を採用した。

3 調査手法

3.1 PAC 分析

本研究では、「布教」におけるオタクの意識や態度を明らかにするため、PAC 分析を用いて調査を行った。PAC とは Personal Attitude Construct の略で、個人別態度構造のことである。内藤哲雄によって開発され、主に心理学や教育学の分野で用いられている。この手法では、いくつかの手順を踏むことで「個人ごとに態度やイメージの構造を分析」^[13] できる。

PAC 分析は分析の手順があらかじめ決められている。大まかには、1. テーマに対する自由連想、2. 連想項目間の類似度評定、3. 非類似度距離行列を用いたクラスター分析、4. 協力者への半構造化インタビュー、5. 総合的解釈、の5つの手順を踏む。内藤（2002）^[13] に従ってそれぞれの手順を説明する。

1. テーマに対する自由連想

調査者はあらかじめ、調査したいテーマを連想刺激文に落とし込んでおく。連想刺激文は、協力者が自由連想する上でのお題のようなものである。調査時には、作成した連想刺激文を協力者に提示し、協力者が思いついた言葉やイメージを自由に挙げてもらう。このとき挙げられた言葉やイメージを連想項目と呼ぶ。連想項目には、協力者によって重要度の順位づけとイメージ（+、-、0）付与を行ってもらう。

2. 連想項目間の類似度評定

2つの連想項目間の直感的な類似度を、調査者が設定した任意の尺度で協力者に評定してもらう。これを全ての連想項目間について行う。

3. 非類似度距離行列を用いたクラスター分析

類似度評定から得られた非類似度行列を用いてクラスター分析を行う。分析結果はデンドログラム（樹形図）に出力する。調査者はデンドログラムを見て、協力者の態度を解釈するのに妥当だと思われるクラスター分けを行い、デンドログラムに書き込む。

4. 協力者への半構造化インタビュー

作成したデンドログラムを用いて半構造化インタビューを行う。具体的には、デンドログラムを協力者に提示し、連想項目や各クラスターについて、喚起されるイメージやクラスターとしてまとまった理由の解釈などを発話として得る。その際、協力者の発話内容だけでなく、表情や仕草なども観察する。

5. 総合的解釈

1~4の手順により得られた様々なデータを用いて、協力者の態度構造を解釈していく。得られたデータには連想項目、非類似度行列、デンドログラム、発話をはじめ、連想項目を連想した順位や重要度の順位、イメージ、発話中の仕草などが挙げられる。はじめにクラスターごとの解釈を行い、それぞれのクラスターに命名を行う。その後、全ての要素を総合して最終的な解釈を行い、その結果を PAC 分析の結果とする。

3.2 対象と期間

調査は2020年9月から2021年3月まで、コロナ禍の情勢を鑑みて全てオンライン上で行った。対象は、オタクとして布教を行った経験がある大学生・大学院生とした。対象の世代を絞った理由は、世代間の意識差が布教の態度に反映される可能性を考慮したからである。募集はスノーボール・サンプリングによって行い、6名（女性4名、男性2名）集まった。今回報告するのは、そのうちの3名（女性2名、男性1名）である。

3.3 調査の流れ

調査は Zoom によるオンラインミーティングにて行った。

まず google フォームを用いて、協力者に事前アンケートを記入してもらった。事前アンケートを行った理由は 2 つある。1 つは、協力者の布教に関する基本的な情報を入手することである。もう 1 つは協力者の負担軽減である。初めから PAC 分析で布教について想起することは協力者の負担になると考えたため、布教について考えるアイスブレイクの役割を想定している。

その後、3.1 の PAC 分析の手順に従って、解釈・イメージの報告までの分析を行った。連想刺激文は「あなたは漫画やアニメを布教するとき、意識することや注意すること、思いつくことはありますか？自分の経験や考えたことを自由に言葉にしてください。単語でも文章でも構いません。」を使用した。自由連想から類似度評定までの作業には、土屋が開発した PAC 分析支援ツール「PAC-Assist2+」^[14]を使用した。

オンライン環境下において、協力者に画面を操作してもらう際には、google リモートデスクトップのリモートサポート機能を用いた。クラスター分析には統計処理ソフト R を使用した。インタビューでは、画面共有機能を用いて協力者と共にデンドログラムを見ながら、各クラスタの内容の解釈、及びクラスタ同士の内容比較、全体構造についてのイメージを問うた。このとき、インタビュー音声は録音した。

4 結果と考察

本稿では表 1 に示した 3 名の協力者について、結果と考察を報告する。なお、表 1 の情報は全て調査当時のものである。

表 1: 調査協力者の詳細

協力者	学年	年齢	性別
F	学部 3 年	21	女性
G	学部 4 年	22	女性
I	学部 3 年	22	男性

考察を報告するに伴い、表記を以下の通りとする。クラスタは“クラスタ”もしくは“CL”と表記する。

連想項目は「」で示す。クラスタ名は<>とする。

4.1 協力者 F の総合的解釈

F の連想項目とその重要度・イメージ、デンドログラムは図 1 の通りであった。全ての連想項目のうち、プラスイメージ、“0”イメージ、マイナスイメージはそれぞれ 3 つずつであった。

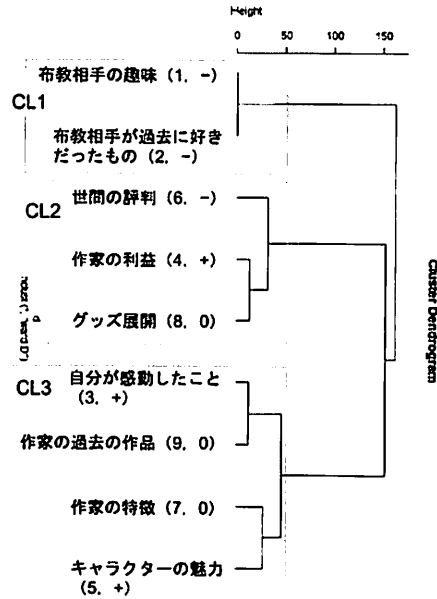


図 1: F のデンドログラム

4.1.1 CL1: 共感を得るために考慮すべき相手のこと

「布教相手の趣味」を考慮していることから、相手に合わせた布教を行っていることがわかる。「布教相手が過去に好きだったもの」から現在の好みを推測していることから、相手の嗜好を重視していることが伺える。ここで F 自身、布教に限らず相手に好きになってもらうためには共感を得る必要がある、と述べていることから、F は普段から好意を得るために相手の共感を引き出せるよう働きかけていることが読み取れる。

また、親しい友人に対しては、自身が好きでなくとも友人の嗜好に近いものを情報提供の一環として

軽率に布教してしまうと述べている。相手との関係性と理解度によっては、F自身の嗜好に限らず相手にすすめることも布教と認識していることがわかる。Fにとって、友人に対する布教は気軽に行えるものであることが推察される。

したがって、CL1は<共感を得るために考慮すべき相手のこと>と名付けられる。

4.1.2 CL2：製作者への正当な評価を広める

自分や布教相手ではなく、作家の懐に直結するグループであると述べている。布教した結果、相手が「グッズ展開」などに興味を持ってコンテンツにお金を落とし、「作家の利益」となればいいと語っている。作家が裕福になってほしいという感情も報告している。これらより、Fは作品を通してコンテンツ製作者を意識して布教していることがわかる。

また、布教するにあたって「世間の評判」を売り文句にすると語っている。F自身は評判を気にすることは無いが、布教するには相手にとって良い情報を与えた方がいいとも発話している。第三者の意見を交えることでコンテンツの面白さに説得力を持たせようとしていることが伺える。相手に与える情報を戦略的に取捨選択しているとも捉えられる。

このようにFは、コンテンツ製作者の存在を意識しつつ、相手への情報を選別して布教に臨んでいる。ここで、Fのコンテンツ製作者には報われてほしいから感想やお金を報酬として還元したい、自分以外の人にもそれ相応の評価をしてほしい、という発話に注目する。この発話から、Fは布教によって、自分の好きなコンテンツの製作者に対して、より多くの人に正当な評価をしてもらうために布教を行っていることが示唆される。コンテンツの製作者への正当な評価を増やすという目的が含まれていることが伺える。したがってCL2は<製作者への正当な評価を広める>クラスタであると言える。

4.1.3 CL3：共感を得たいが表現が難しい

Fは、「作家の過去の作品」を見たときに「自分が感動したこと」を相手に伝えようとしている。また、作品によって自身が好きだと感じる部分は違うことに触れ、キャラクターが好きな場合は「キャラクターの魅力」を伝え、作家が好きな場合は「作家

の特徴」を伝えると語っている。F自身の感情を前面に押し出しているクラスタであると言える。

Fは、これらF自身の感情を他人にわかってもらいたいと述べており、Fは相手に共感してほしいと考えていることが示唆される。しかし、感情を実際に言葉にすることは難しく伝えにくいとも語っていることから、自分に共感してほしいがうまく伝えられずにもどかしい思いをしていることが伺える。

したがってCL3は<共感を得たいが表現が難しい>と名付けられる。

4.1.4 全体考察

Fの布教は、<製作者への正当な評価を広める>つつF自身への<共感を得たいが表現が難しい>ため、<共感を得るために考慮すべき相手のこと>を意識していると捉えることができる。

しかしFは、布教相手がリアルだった場合とインターネット上だった場合とで、相手への対応が全く異なることを述べている。リアルな相手の場合は、コミュニケーションの一環であり、相手につまらない時間を過ごさせないようにしようという意識があると語っている。一方、インターネット上で繋がっている相手の場合は、相手の趣味重視ではなく自分の趣味重視で投げつけるようなイメージで布教していると報告している。Fの報告によると、リアルな相手の場合は<共感を得るために考慮すべき相手のこと>を最重要視し、<製作者への正当な評価を広める>ことはあまりせず、<共感を得たいが表現が難しい>ため表に出さないようにしている。それに対してインターネット上の相手の場合は、<共感を得るために考慮すべき相手のこと>はほとんど意識しておらず、<製作者への正当な評価を広める>ために奔走し<共感を得たいが表現が難しい>部分も積極的に発信する。このようにリアルとインターネット上とでは、Fの布教の態度が正反対となる。

ここから、Fにとって布教とは相手によって態度が大きく異なる行動であることがわかる。リアルな場合は相手の態度から存在を強く意識するため自分を抑えているが、インターネットを介する場合は相手の態度が見えないため存在を気にせず自分を解放していると捉えることができる。

4.2 協力者 G の総合的解釈

G の連想項目とその重要度・イメージ、デンドログラムは図2の通りであった。連想項目のイメージは、プラスが4つ、“0”が4つだった。マイナスはなかった。

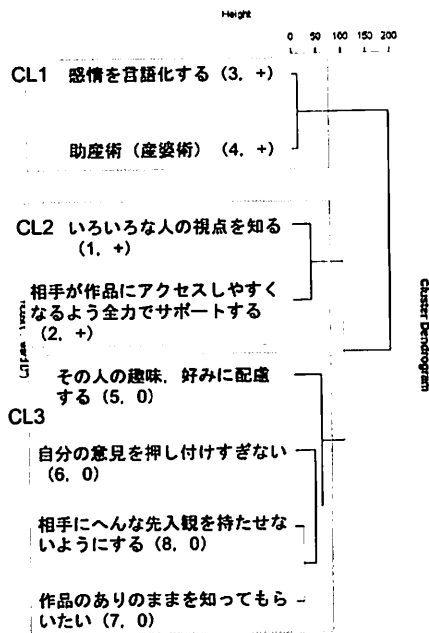


図 2: G のデンドログラム

4.2.1 CL1：発信による自身の感情の同定

G は、布教したいことがらについて言葉にすることではっきりと気持ちが強くなると述べており、相手に話すことで自分の中の「感情を言語化」していることがわかる。G 自身がコンテンツを見ることで得た曖昧な感情が、布教相手との対話により明確となり、言語化によって強化されていると捉えられる。すなわち G にとっては、布教が感情の同定の役割を果たしていると考えられる。

また、G は布教によって、自分にはない相手の考えから新しい語彙をもらっていると述べている。この相手の言葉を借りるような状態を「助産術（産婆術）」と表している。G のみでは至らなかった思考や言葉を、他者の介入によって発見していると言える。

つまり、G は相手にコンテンツの情報を発信することで、逆に新しい表現や言葉を獲得していると考

えられる。したがって、CL1 は「発信による自身の感情の同定」と名付けられる。

4.2.2 CL2：自分のために相手に尽くす

布教において G は、「相手が作品にアクセスしやすくなるよう全力でサポート」している。相手を支援することで、布教の成功度をあげようとしていることが伺える。

しかし、サポートをするのは根底に「いろいろな人の視点を知りたい」という気持ちがあるからだとも述べている。ここで G は、自身と異なる専門知識を持った人物に作品の新しい見方を教わり世界観が広がったという体験を報告している。つまり、相手のために手を尽くすだけではなく、自分のために支援を行っているという意識を持っていると言える。したがってこのクラスタは「自分のために相手に尽くす」クラスタであると言える。

4.2.3 CL3：相手への配慮と戦略性

「作品のありのままを知ってもら」うために「相手に変な先入観を持たせないように」していることがわかる。相手にコンテンツを素直に受け取ってほしいと考えていることが示唆される。「自分の意見を押し付けすぎない」ようにしていることから、なるべく相手の思考を邪魔しないように配慮していることがわかる。また、「その人の趣味・好みに配慮」する姿勢も見られる。このクラスタは誰かに布教するときに気をつけていることが集まっているという発言からも、このクラスタが G の相手に対する気配りのクラスタであることが見て取れる。

一方で G は、相手を変な先入観を持ってしまった場合、その人特有の視点を探れなくなることが悲しいと述べている。また、趣味や好みに配慮する裏では、この人ならコンテンツを気に入ってもらえたり自分に共感してもらえたりするのではないかという打算が働いているとの発言もある。このように CL3 では、相手に対する配慮の裏側に、G 自身の戦略性が潜んでいることが伺える。したがって、CL3 は「相手への配慮と戦略性」と名付けられる。

4.2.4 全体考察

連想項目はプラスイメージと“0”イメージのみである。よって、布教を行うことに前向きな姿勢であ

ると考えられる。この前向きさは、G自身が<発信による自身の感情の同定>を行うため、<相手への配慮と戦略性>によって<自分のために相手に尽くす>という態度から来ていると思われる。G自身、客観的には面倒臭そうな作業であっても、布教として行うならば苦痛はなく負担もないと語っており、根幹が相手ではなく自分にあることで布教を楽しんでいることが伺える。

つまりGは布教について、相手のために行っているという意識がありつつも、根幹は自分のためという認識を持っていると言える。相手とのやりとりが巡り巡って自分に利益をもたらす状況を楽しんでいる可能性も考えられる。

4.3 協力者Iの総合的解釈

Iの連想項目とその重要度・イメージ、デンドログラムは図3の通りであった。連想項目のイメージは、プラスが2つ、マイナスが5つであった。“0”は見られなかった。

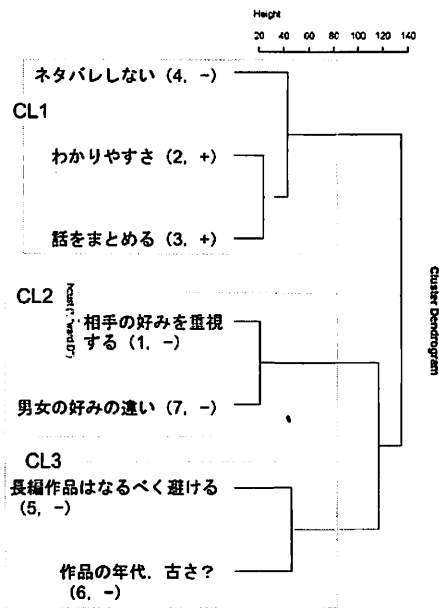


図3: Iのデンドログラム

4.3.1 CL1:自発的に工夫できる部分とできない部分がある葛藤

CL1の連想項目から、Iは布教する相手にどれだけ「ネタバレしない」で「わかりやす」く「話をまとめ」られるかを意識していることがわかる。興味を持ってもらえるよう、話の内容から相手に配慮していると考えられる。また、このクラスは相手によらず常に大事にすべきとも述べていることから、Iはどんな相手に対しても自発的に工夫していることが伺える。

ここで、イメージやデンドログラムの形からもわかるように、「わかりやすさ」「話をまとめる」と「ネタバレをしない」にはIの意識の違いが見られる。前者は得意で苦労しないと語る一方で、後者は大事だと思ふもののネタバレなしに話を進めることは難しいと語っている。イメージがプラスとマイナスに分かれていることから、ネタバレに関して、Iが重要視したい理想とうまくできない自分の現実のギャップに葛藤を抱いていることが伺える。ネタバレについてはCL1内で唯一、相手の存在によって重要度が変化すると発言もある。Iにとっては全て十分に気をつけたい事柄だが、相手次第ではネタバレすることがより良い選択となる可能性も理解していることが示唆される。したがって、CL1は<自発的に工夫できる部分とできない部分がある葛藤>と命名される。

4.3.2 CL2:自分の好みより相手の好みを優先

何かをすすめる上では、自分の好みを殺して「相手の好みを重視する」必要があると述べている。好みを殺すという強い言葉を使っている点や、重要度が1位である点からも、Iが布教において相手の好みを最も考慮していることがわかる。「男女の好みの違い」は、これまで男性にしか布教したことがないため考慮したことがなく重要度も低いが、必要があれば重要度は上がっていた可能性を語っている。布教において、“好み”という要素が重要な地位を占めていると考えている可能性がある。

この好みという点について、Iは自らが面白いと思っていた作品でも相手から色よい返事が来なかった体験を語っている。相手の好みを無視するとコンテンツそのものに嫌悪感を抱かれ、そもそも興味を持ってもらえなくなり、布教の根底を揺るがしかねない問題となるとも述べている。ここからIは、布

教をするには興味を持ってもらえる土壌が必要であると考えていることが伺える。嫌悪感により興味を持ってもらえなくなるくらいならば、あえて刺激せずフラットな感情でいてほしいと思っている可能性が示唆される。

このように相手の好みを重要視する一方で、CL2の連想項目のイメージはどちらもマイナスである。この点についてIは感情的なやりづらさを語っており、相手に合わせることで自分の好みを前面に出せないという状況が、相手の好みと自分の好みの間で板挟みになり葛藤に繋がっていることが考えられる。まとめると、CL2は<自分の好みより相手の好みを優先>と名付けられる。

4.3.3 CL3：好みの不和によるストレスを回避

あまりに長いと紹介してもみんな見たがらず、めんどくさいと思われがちなため「長編作品はなるべく避ける」ようにしていると語っている。コンテンツが布教に適するかどうかを考慮していることがわかる。また、「作品の年代、古さ」によってはほとんど市場に出回っていないものもあるためおすすめの重要度が下がると述べており、流通の状況も考慮していることが読み取れる。Iは布教において、相手のストレスとなるような要素をなるべく排除しようとしていると考えられる。

一方でIは、昔の長編作品を小学校の頃から見ていた経験があり、思い入れが深いと述べている。好きだからこそ紹介を避けたくはないが、軽率に紹介しないほうが良いとわかっているとも語っている。ここから、自身の好きなものを否定されたくないという思いから、あえて布教する際に考慮に加えているとも捉えられる。

必ずしも視聴を拒否されるわけではないためあまり意識する必要はないと語り、重要度も高くはないが、明らかにIの布教への態度に葛藤をもたらしているクラスである。したがって、CL3を<好みの不和によるストレスを回避>と名付ける。

4.3.4 全体考察

連想項目はマイナスイメージが多い。ここまでの考察から、Iは理想と現実によって葛藤が生じている項目にマイナスをつけることがわかっている。し

たがって、布教に対して全体的に葛藤を抱えていることが推察される。

最終的には作品の話と一緒にできるようになることが、布教の面白さであり一番の利点であると語っている。そのため<自分の好みより相手の好みを優先>したり、<好みの不和によるストレスを回避>していると考えられる。Iにとって布教とは、好きなものをすすめるだけではなく、好きなものを通してコミュニケーションをとる手段であると捉えることができる。そのために、人を惹きつけるような内容を作らないといけないと述べており、伝える内容を<自発的に工夫できる部分とできない部分がある葛藤>がある態度が見える。

つまりIは布教を行う際、簡潔さとわかりやすさを両立させつつ相手の興味を引けるよう配慮していると言える。一方で、相手に配慮することで自分自身の考えや好みをあまり外に出せず、布教という行為に葛藤が生まれていることも示唆される。

5 おわりに

本研究では、オタクが「布教」を行う際の意識や態度を明らかにすることを目的として、6名の大学生・大学院生にPAC分析を用いた調査を行った。本稿ではそのうち3名の結果を報告した。

協力者Fは、布教を行うことで相手からの共感を得つつ、相手に作品やその製作者への評価を何らかの行動によって示してほしいと考えていることがわかった。また、リアルな相手とインターネット上の相手と、布教の意識が異なることが示された。協力者Gは、布教を相手のために行っているという意識を持ちつつも、根幹には自分のためにやっているという認識があることがわかった。布教を行うことに楽しさを見出していることも示唆された。協力者Iは、布教によって相手の作品への興味を引けるよう、簡潔さとわかりやすさを意識しつつ相手の好みに配慮していることがわかった。しかし、相手と自分自身の考えや好みが違うことを理解しているために、布教という行為に葛藤が生まれていることも明らかとなった。

以上の結果から、布教を行うオタクは何らかの形で布教相手の存在を意識していることが示された。特に相手の好みについては3名とも意識していたが、意識するに至る背景は三者三様であった。したがって、オタクの「布教」では行為者であるオタクとそ

の相手の存在が重要な意味を持っていることが示唆された。

今後は、本稿で報告した3人の考察を、事前アンケートの結果を含めてさらに深めていく。また、今回報告できなかった3名の分析を進めていくとともに、男女比を考慮して調査対象者を増やしていく。

参考文献

- [1] 東浩紀: 「動物化するポストモダン オタクから見た日本社会」. 動物化するポストモダン. 講談社, 2001.
- [2] 相田美穂: 「おたくをめぐる言説の構成: 1983年~2005年サブカルチャー史」, 広島修大論集. 人文編, Vol. 46, No. 1, pp. 17-58, 2005.
- [3] 永田大輔: 「コンテンツ消費における「オタク文化の独自性」の形成過程」, ソシオロジ, Vol. 59, No. 3, pp. 21-37, 2015.
- [4] 和田崇: 「オタク文化の集積とオタクの参画を得たまちづくり: 大阪・日本橋の事例」, 経済地理学年報, Vol. 60, No. 1, pp. 23-36, 2014.
- [5] 入江由規: 「「ゲスト」へと変貌したオタクたち: アニメ聖地巡礼者の交流から」, フォーラム現代社会学, Vol. 13, pp. 58-70, 2014.
- [6] 出原健一: 「相同性: 「オタク文化」の場合」, 彦根論叢, No. 388, pp. 18-31, 2011.
- [7] 山口麻希; 西崎雅仁: 「オタク文化の経済価値に関する経営学的考察」, 経営情報学会 全国研究発表大会要旨集, Vol. 2010f, pp. 38-38, 2010.
- [8] 今井信治: 「オタク文化と宗教の臨界: 情報・消費・場所をめぐる宗教社会学的研究」. 晃洋書房, 2018.
- [9] 永井睦美; 福田豊: 「オタクのコミュニティ形成・維持機能について」, 日本社会情報学会全国大会研究発表論文集, Vol. 25, pp. 96-99, 2010.
- [10] 笹倉尚子: 「漫画やアニメについて他者に語るプロセス」, 心理臨床学研究, Vol. 28, No. 1, pp. 16-27, 2010.
- [11] 笹倉尚子: 「漫画やアニメについて他者に語るプロセス: 他者に語る行為の背景について」, 京都大学大学院教育学研究科紀要, Vol. 56, pp. 195-207, 2010.
- [12] 枝井栄利子; 守屋英子: 「思春期における自立の支えとなるもの: マンガ・アニメ・ゲーム等にハマるといふ観点から」, 茨城大学教育実践研究, No. 33, pp. 185-199, 2014.
- [13] 内藤哲雄: 「PAC分析実施法入門: 「個」を科学する新技法への招待」. ナカニシヤ出版, 改訂版, 2002.
- [14] 土屋義郎: 「PAC-Assist2」. <http://www.kanazawa-it.ac.jp/~tsuchida/lecture/pac-assist.htm>.

第29回年次大会予稿

二次創作作品を対象とした物語特徴分析

Characteristic analysis of stories in Fan Fiction

石川一稀^{1*}, 村井源¹

Kazuki ISHIKAWA^{1*}, Hajime MURAI¹

1 公立はこだて未来大学

Future University HAKODATE

〒041-8655 北海道函館市亀田中野町116-2

E-mail: g2121004 @fun.ac.jp

*連絡先著者 Corresponding Author

昨今、意欲的に進められている物語の分析は一次創作作品を対象としたものであり、二次創作作品を対象としたものは少ない。そこで本研究では二次創作作品研究の第一歩として、各作品の特徴を表すと考えられるジャンルや登場人物などの属性を探索的に調査し、複数属性に基づく作品のデータ化を行った。得られたデータに対して計量的な分析を行った結果、二次創作作品の典型的な3種類の物語パターンなどが明らかになった。

In recent years, most of the analyses of story's have been conducted on primary works, and only a few have been conducted on Fan Fiction. In this study, as a first step in the study of fan fiction, we exploratively surveyed the attributes such as genre and characters that are considered to represent the characteristics of each work, and created data of the works based on multiple attributes. As a result of quantitative analysis of the obtained data, three typical types of story patterns of Fan Fiction were revealed.

キーワード: 物語分析, 特徴分析, 二次創作

Story analysis, Feature analysis, Fan Fiction

1 はじめに

近年、小説・漫画などの媒体から物語を分析、物語構造を抽出し物語自動生成へ繋げていく形で研究は今まで盛んに行われている。自動生成されたショートショート

で文学賞の一つである「星新一賞」の一次選考を突破し、話題になったことは記憶に新しい[1]。星新一の作品を対象とした研究では、物語の流れや話のオチについてパターン化なども研究が成されてきた[2,3]。こ

の他にもミステリやループ物[4,5,6], RPGやホラー[7,8]などのジャンル別の物語の特徴, 特有の展開などが研究され, 一次創作作品に対しては様々な視点・手法によって物語研究が進んできた。

国内で行われている二次創作に関する研究は, 二次創作作品が売買される市場構造や二次創作をする作家などに焦点を当てた分析・考察がこれまで行われてきた[9]. しかし二次創作作品の物語自体を分析対象にした国内の研究は確認できていない。

国外の研究を見てもその例は少ないが, Fan FictionまたはFan Maidという名前で二次創作が浸透・研究されている。海外ではFan Fictionの研究がハリー・ポッターなどを題材に行われており, 原作とFan Fictionでの登場人物の注目度に関する統計的な分析[10]や, 作品の要素間の関係性をオントロジー化する試みなどが行われてきた[11]. これらの先行研究では二次創作作品そのものを対象とするが, 物語の種類や展開をパターン化する研究とはいえない。

これまで二次創作作品の物語分析が行われなかった背景には, いくつかの理由があげられる。その理由の中で最大の要因となるのが, 上記でも触れたように著作権に関する問題である。昨今SNSやwebサイトなどで散見されるイラストや小説などといった二次創作作品は, 作品の権利者に許諾をとって掲載されている作品は数少ない。しかし, 近年では作品の権利者側が二次創作をする上でのガイドラインを作成し, その存在を公に認めているケースも増えてきている[12]. このことから現代では公の場で二次創作作品が発表できる環境が整いつつある。加えて, 日本では世界有数の同人即売会である『コミックマーケ

ット』が一年に二回の頻度で開催されるほど二次創作の文化が大変盛んである。1970年代から現代まで続くコミックマーケットは多くの二次創作作品を生み出し, 世に送り出してきた。2019年12月末に開催された『コミックマーケット97』では総入場者数は75万人, 出点サークル数は3万2千にものぼった[13]. これらのことから, 研究そのものの新規性や文化的な観点からも二次創作作品の分析を試みる価値は大いにありと考えられる。

本研究は二次創作作品の物語に焦点を当て, その特徴の分析・パターンを抽出することを目的とした新たな試みである。

2 分析対象の選好

2.1 分析対象となる一次創作作品

分析対象データとなる作品は, その原作の権利者が二次創作作品の存在を公的に認めているものが好ましい。それに加えて, 計量的な分析に耐えうるほどの作品数が必要である。その条件を満たす作品として, 本研究では『東方Project』[14]という作品を分析対象の原作と定めた。1995年から現在に至るまで25年間もの作品が続き, ブロック崩しゲームから始まった本作は6作目以降からシューティングゲーム(以下STG)へと移行し, そこからアクションゲームや小説, 漫画など様々なメディアで展開されてきた人気タイトルである。

加えて人気タイトルであるというだけでなく, 『東方 Project』を分析対象とする最大の利点は, その二次創作ガイドライン[15]の寛容さと二次創作作品の量である。

二次創作ガイドラインには物販や流通を規制するものと作品内容に関するものなどの大きく分けて2種類が存在する。『東方Project』の二次創作ガイドラインは, 二つの

観点から見ても非常に自由度が高い。比較対象として2012年10月3日に公開されたフリーゲーム『魔女の家』[16]の二次創作ガイドライン[17]をあげる。これには原作と比較し、同一性が損なわれる、他作品のキャラを混在させる、舞台や時代などに関する世界観を失う、エンディングを改変する、といった二次的創作を禁止している。しかし『東方Project』ではそういった禁止事項はない。これにより二次創作作品の自由度は他の作品よりも高くなり、多くの二次創作作品が許容されている。

二次創作作品数に関しては有志の集計のため参考数字になるがwebサイトに投稿されているイラストの数は250万件以上である[18]。この件数には同人誌などで頒布された紙媒体の作品は含まれてはいない。さらに年に2回開催されている『東方Project』の二次創作同人誌のみを扱った同人即売会『博麗神社例大祭』では2018年のイベント来場者数は5万人以上にも上り、2019年に国内で行われた『東方Project』関連の同人即売会は150回以上にも昇る[18]。これだけ大量の二次創作作品が生み出され、かつそれらの存在が公的に認められている作品は他に類を見ない。よって『東方Project』は本研究の分析対象として適している一次創作作品といえる。

2.2 分析する二次創作作品の選出方法

分析対象となる作品はクオリティの高いものが好ましいが、二次創作作品の客観的クオリティを担保するのは難しい。二次創作作品は商業作品と違い、売上金額や発行部数などの数字を表に出すことはない。そこで本研究では、SNS等を利用したアンケートを用いて、『東方Project』のファンからクオリティの高い二次創作作品を募った。

アンケートは全2回行った。1回目のアンケートではクオリティの高い二次創作作品を募集し、479作品が集まった。そこから得票数が高かったものを厳選し、2回目のアンケートでは厳選した作品を上げ、これらがクオリティの高いものだと同意できるかを調査しつつ、さらに上記にない作品を推薦する項目を設けた。

2回のアンケートの結果、得票数の多かった作品の内、入手困難なものを除いた上位50作品を本研究の分析対象とした。本研究では物語の特徴分析を目的としている為、物語性が薄い「ギャグ」を取り扱った作品や成人向けの作品を除外した。

3 分析手法

二次創作作品を分析するにあたって、どのような要素を特徴とし抽出すべきか。本研究ではそれを明らかにするために探索的な手法を検討した。作成したリストの作品について、二次創作作品の特筆すべき特徴と考えられる要素を抽出し記述した。主に次のような手順を用いた。

- A. 二次創作作品の重要な要素を仮説に基づいて抽出する項目を考案する。
- B. 作品を読み、考案した項目を記述する。

この他にも抽出した項目や作品自体に特筆すべきものがあつた際には自然言語で記述し、後の考察や項目の検討に利用する。分析作品の全50作品を分析し二次創作作品の物語構造において特徴がよく現れていると考えられる項目を表1にまとめた。この項目は、一つの物語に対し複数該当することもある。表1の各項目について解説を行う。

「物語ジャンル」は分析対象となる作品が、原作のストーリーラインと比べて、どのようなストーリーラインを採用しているかを記述するも

のである。作品に対する物語のジャンルについて明確な定義は存在しないため、表2のような定義も設けた。

「アクション」は物語進行上の障害や問題を武力による決闘で解決する物語ジャンルである。今回の定義では決闘上のルールの有無については考慮しないものとした。

「冒険譚」は特定の物体や概念的な思想などの獲得、発見の過程、結果を物語として描いた物語ジャンルである。主人公が特定の物体を探す為に旅に出るというストーリーライン

などは、これに該当する。

「ホラー」は怪異といった科学で説明困難な得体の知れないものにより引き起こされる恐怖体験や、人間同士で行われる凄惨な攻撃的行為の様を描いた作品が該当する。

表1 分析対象作品から抽出した項目

物語ジャンル
オリジナルキャラクターの有無
キャラクターの生死
時系列
作品の舞台

表2 物語ジャンル定義表

物語ジャンル	概要
アクション	作中で起こる問題・障害を武力で持って解決しようと行動する過程・結果を描く
冒険譚	特定の物体や概念的な思想などを発見・獲得する過程・結果を描く
ホラー	オカルト・怪異などの科学では説明できない類いのモノによって恐怖を描く
成長譚	物語を通して精神的・肉体的な成長を描く
ミステリ	作中で謎や問題が提示または発覚し、それを推理・推論などを用いて解決する
日常	日常生活における小規模な問題解決を描く
青春	登場人物同士の関係性の発展を描く

「成長譚」は登場人物の精神的・身体的成長の過程、結果が描かれる物語ジャンルである。具体例を挙げるなら、逆上がりができない子供を主人公とし、その練習過程やチャレンジすることで鍛えられる精神の成長を描くような物語が「成長譚」に該当する。

「ミステリ」は同じ物語ジャンルのアクションの決闘部分を推理などに置き換えた物語ジャンルである。決闘(暴力的行為)を行わず、発生した問題・障害に対して知識や推理を用いて解決することを主軸とした物語が該当する。主な該当作品として、推理物と呼ばれる作品群があげられる。

「日常」はその名の通り、登場人物の日常風景を描くことを主とした物語ジャンルである。その特徴として、作中での主人公の目的が他ジャンルと比べて非常に小さく、日常

生活に収まるくらいに小さいものである。

「青春」は、登場人物間の人間関係発展や深まりを描くことを主とした物語ジャンルである。本研究では深まる関係性を特定せず、恋愛・友愛・家族愛など幅広く集計した。

「オリジナルキャラクターの有無」とは、原作となる物語に登場しないキャラクターが分析作品中に登場したかどうかを記述したものである。登場人物を「オリジナルキャラクター」とする際は、名前が作中で明記されていることとそのキャラクターが発話していることの二つの条件を満たしたものとした。しかし演出の関係上、明らかに名前が伏せられている場合や、作中で死亡しているために回想以外の発話が不可能であるなどの場合は、条件に当てはまらずとも「オリジナルキャラクター」として認めた。

「キャラクターの生死」は、作中で登場人物が死亡した場合に記述する項目である。

「時系列」は原作の物語に対してどのような時間軸の作品かを記述する項目である。本研究では、作品ごとの時系列を表3のように定めた。「前日譚」は原作で起こる出来事の直前の時系列であり、物語の行く末が原作の出来事に繋がることを判断基準とした。

「並行譚」は原作の物語進行と抱き合わせで進行し、その進行を別の視点で観測している物語作品が該当する。

「後日談」は原作で起こった出来事の直近の時系列であり、物語の発端が出来事に起因している場合に該当する。

そして「後日談」の中で既存キャラクターの死亡、世界の滅亡など様々な要因で原作の時系列の中には含むことができない遙か遠い時系列のため原作の時系列に収まらない作品を新たに「後日談(非連続)」としてまとめた。

「作品の舞台」では、物語の舞台がどこで展開しているかを記述した。作品の主要メディア媒体での舞台を「内」、その他メディア等での舞台を「外」とした。これらとは別に二次創作作品を大まかなパターンごとに分類するべく、3つのパターンを立案した。それが表4である。本研究では、対象作品を分析し、該当するパターンを記述した。複数個に該当する場合は、その全てを記述した。

表3 時系列定義表

時系列	概要
並行譚	原作同じ時系列で進行する。物語を観測する視点は原作と異なっても構わない。
前日譚	描かれた物語の時系列が原作での出来事の前に該当し、結末が出来事へ繋がっている。
後日談	描かれた物語の時系列が原作での出来事の後に該当し、結末を踏まえて原作の時系列を壊さない。
後日談(非連続)	描かれた物語の時系列が原作での出来事の後に該当し、結末を踏まえると原作の時系列が壊れる。

表4 考案した二次創作パターン

パターン	概要
救済	原作での不幸な境遇やコンプレックスなどを救済点とし、それを解決する。
設定考察	原作に既にある設定・要素の内、成り立ちや本質が不明瞭であるモノを考察で補う。
設定流用	原作の要素の一部を改編または抹消し、本来も世界観では起こりえない物語。

「救済」は物語の中でキャラクターが持つ不幸な境遇や報われない展開、各々が持つコンプレックスなどを救済点として上げ、それを解消・解決する物語パターンである。

「設定考察」は原作に既に存在する要素・設定に関して、その成り立ちや本質が不明瞭である際に、その不明瞭部分を考察し、その考察を物語にするパターンである。

「設定流用」は原作の要素や設定を一部改編・抹消などして本来の世界観では起こ

りえない物語を展開するパターンである。この改編・抹消は主に原作で起こった出来事をなかったことにするや世界に存在するルールの消失、改編などがあげられる。その他にも他作品の要素と掛け合わせるクロスオーバー作品もこのパターンに該当する。

4 分析

作品リスト上位50作品の分析対象作品を作成した抽出項目、定義表に基づいて分析した。

作成した50作品分のデータを対象に本研究では、作品にまつわる16個の抽出要素を変数として因子分析を行った。因子間に相関関係があると想定し、プロマックス回転を採用した。因子数は平行法より3と決定した。

5 結果と考察

因子分析の結果、以下の表5のような結果が得られた。

表5 因子分析の結果

	因子1	因子2	因子3
アクション	0.14484	-0.01562	-0.03165
冒険譚	0.09488	-0.01904	0.01784
ホラー	0.34234	0.07571	-0.14435
ミステリ	-0.0722	0.24078	-0.17185
成長譚	-0.12084	0.15478	0.28657
日常	-0.20568	-0.01338	-0.06205
青春	-0.07521	0.02924	0.46519
オリキャラの有無	0.05617	0.04709	0.42911
死亡	0.41266	0.28848	0.27204
並行譚	-0.22085	0.32217	-0.44713
前日譚	-0.59934	0.33565	0.50138
後日談	-0.52989	-1.0005	0.02413
後日談(非連続)	1.03544	0.28581	-0.01884
救済	-0.01732	-0.11917	0.30391
設定考察	-0.06946	0.34714	0.17501
設定流用	-0.00507	-0.01315	-0.35717

因子1では正の値として「ホラー」が0.34、「死亡」が0.41、「後日談(非連続)」が1.03である。また負の値として「前日譚」が-0.59、「後日談」が-0.52、となっている。

因子2では正の値として「並行譚」が0.32、「前日譚」が0.33、「設定考察」が0.34である。負の値として「後日談」が-1.00と出ている。

因子3では正の値として「青春」が0.46、「オリジナルキャラクターの有無」が0.42、「前日譚」が0.50、「救済」が0.30となっている。負の値と

して「並行譚」が-0.44、「設定流用」が-0.35となっている。

本研究では、閾値を因子負荷量の絶対値が0.3を超えるものと定め、各因子で該当するものについて考察を述べる。

5.1 因子1についての考察

因子1について、物語ジャンルの「ホラー」と「キャラクターの死亡」についての親和性の高さは想像に難くなく、その結果原作の時系列の中には組み込めない独自の物語となるパターンが多いのだと読み取れる。ホラーな物語を展開し、既存の登場人物が死亡してしまうと前日譚として物語を原作の出来事に繋げることが難しいことが読み取れる。本研究で取り扱った原作『東方Project』の世界観がオカルト要素を内包していることから、ホラーな物語を展開しやすいということが因子1に大きな影響を与えていると考えられる。加えて原作に登場するキャラクターを死亡させる行為や世界の終末を描く行為は、原作に則したジャンルの物語でなければ行うことは難しく、原作ではできないような物語を描きたいという作り手の意思を強く反映した因子とも読み取ることができる。因子負荷量が大きく、物語ジャンルが一つしか出現していないことから、この因子1を「非連続式ホラー型死亡」因子と命名する。

5.2 因子2についての考察

因子2について、原作で起こった出来事によって露出した要素や設定を考察し、それによって前日譚や並行譚の時系列で物語にすることで、要素・設定についての考察を作り手が読者に対して打ち出す物語の魅せ方が読み取れる。この傾向は後日談で語るよりは、原作の出来事へ繋げる形で描ける前日譚や、原作と同時進行で進む物語を別視点から見る並行

譚の方が考察に説得力を出すことができるからではないかと考えられる。「設定考察」と「後日談」の因子負荷量の絶対値が大きいことや二つの時系列で一つの二次創作パターンを表現することから、この因子2を「非後日談式設定考察」因子と命名する。

5.3 因子3についての考察

因子3について、キャラクター同士の関係性が発展していく中で、キャラクターが抱えている救済点を解決・解消していくストーリーラインが読み取れる。キャラクター同士というのは、既存キャラクター同士、既存キャラクターとオリジナルキャラクターの二つのパターンが存在する。加えてキャラクター同士の関係性を深めていく際、そのキャラクターの血縁者や、生息区域の施設に務める人物や職場の人間などが登場する流れが多く、それら二つの要因が「オリジナルキャラクターの有無」の因子負荷量を大きくしているものと考えられる。時系列として前日譚が用いられるのは、原作で新たな設定が露出した際、実際のところはわからないものの、実は交友関係があるのではないかと、共通する要素があるのなら交友関係があってもおかしくないのではないかと、という想像が働き、そこに特定のキャラクター同士の関係性を深めたいという作り手の願望が合わさったことで、物語として出力されたのだと読み取ることができる。因子負荷量が大きく、時系列・物語ジャンル・二次創作パターンが一つずつ現れていることから、この因子3を「前日譚式青春型救済」因子と命名する。

6 結論と今後の展望

6.1 まとめ

本研究では、『東方Project』を原作とするクオリティの高い二次創作作品をアンケートによ

り収集・選定した。分析対象作品リストに基づき、二次創作作品の特徴を抽出するための属性を探索的に抽出し、これらの属性に基づいて二次創作作品のデータ化を行った。得られた各作品の属性に対して計量的な分析を行い、二次創作作品における物語やパターンなどの特徴の抽出を試みた。その結果、因子分析では「非連続式ホラー型死亡」、「非後日談式設定考察」、「前日譚式青春型救済」の三つの因子が抽出され、典型的な抽出項目の組み合わせの傾向が明らかになった。

6.2 今後の展望

本研究で因子分析を行うことで、二次創作作品の特徴的な要素・要素ごとの組み合わせのパターンを抽出に成功した。これにより、今後二次創作作品の自動生成を行う際、趣向に合わせてジャンルや登場人物、その特性などの物語に関する要素を選定することができるようになる。しかし本研究には物語全体の流れが含まれていない。物語の起承転結の中で、誰が・何処で・何をした、といった物語の流れの抽出には至っていない。今後の方針として、物語自動生成を目指すために、二次創作作品におけるストーリーラインの抽出、または既に研究されている一次創作作品のストーリーラインの転用が可能であるか研究することが想定される。また、本研究では『東方Project』を原作とした二次創作作品の分析を行ったが、これらの結果が現存する二次創作作品の普遍的な特徴・パターンとは言い切れない。原作が異なる場合、それに伴って特徴・パターンが変化する可能性は大いに考えられる。そのため別ジャンルまたは同ジャンルの作品を原作とした二次創作作品の分析をすることが求められる。

参考文献

- [1] 松原仁, 佐藤理史, 赤石美奈, 角薫, 迎山和司, 中島秀之, 瀬名秀明, 村井源, 大塚裕子:「コンピュータに星新一のようなショートショートを創作させる試み」, 第27回人工知能学会全国大会予稿集, 2D1-1, 2013.
- [2] 豊澤修平, 村井源:「物語自動生成のための文脈依存性を考慮した文章表現抽象化」じんもんこん2020論文集, pp. 135-142, 2020.
- [3] 佐藤知恵, 村井源, 往住彰文:「星新一ショートショート文学の物語パターン抽出」情報知識学会誌, Vol. 20, No. 2, pp. 123-128, 2010.
- [4] 山田康貴, 村井源:「ネットワーク分析を用いたミステリー小説における場面の依存関係の分析」, 情報知識学会誌, Vol. 30, No. 2, pp. 203-213, 2020.
- [5] 村井源:「複雑的構造を持つ物語の自動生成に向けて」, 情報知識学会誌, Vol. 30, No. 2, pp. 214-219, 2020.
- [6] 栗原将風:「多変量解析を用いた2000年以降の日本のループ物語作品の傾向分析」, 平成30年度 公立はこだて未来大学卒業論文, 2019.
- [7] 石川一稀, 中村祥吾, 宇田朗子, 斉藤勇璃, 根本さくら, 長野恭介, 山内拓真, 白石智誠, 太田和宏, 稲垣武, 村井源, 平田圭二, 迎山和司, 田柳恵美子:「ロールプレイングゲームシナリオにおけるクエスト構造に着目した物語分析」, 情報知識学会誌, Vol. 30, No. 2, pp. 256-262, 2020.
- [8] 鈴木諒輔, 佐々木奨之, 袴田翔, 田中瑞穂, 三浦隆太郎, 城田晃希, 高橋翔太, 南部太雅, 山田康貴, 吉田拓海, 松浦史佳, 松原千里, 寺島啓悟, 津沢慎吾, 渡邊広基, 村井源, 迎山和司, 田柳恵美子, 平田圭二, 角薫, 松原仁, "物語と情景描写を自動生成する統合的システムの検討と開発", 情報処理学会研究報告, Vol. 2018-EC-50, No. 28, 1-8, 2018.
- [9] 立花晃:「コミックマーケットをめぐる二次創作市場の構造と二次創作者の動機に関する研究」, 第42回日本計画行政学会全国大会研究報告要旨集, pp. 35-38, 2019.
- [10] David Bamman: "Beyond Canonical Texts: A Computational Analysis of Fanfiction", Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 2048-2053, 2016.
- [11] Federico Ponzola: "Linked-Potter: an example of ontology for the study of the evolution of literature and reading communities", 2020 JADH2020.
- [12] ニトロプラス:「ニトロプラス キラル作品の著作物利用について」, <https://www.nitroplus.co.jp/license/> (最終閲覧日2021-1-17) .
- [13] 有限会社コミケット:「コミックマーケット97 アフターレポート」, <https://www.comiket.co.jp/info-a/C97/C97AfterReport.html> (最終閲覧日2021-1-17).
- [14] 東方Project, 上海アリス弦楽団, 1996.
- [15] 上海アリス弦楽団:「東方Projectの二次創作ガイドライン」, <https://touhou-project.news/guideline/> (最終閲覧日2021-1-17).
- [16] 魔女の家, ふみー, 2012.
- [17] ふみー:「『魔女の家』の利用規約」, <https://majonoie.karou.jp/guideline.html> (最終閲覧日2021-1-17).
- [18] 上海アリス弦楽団:「数字でみる「東方Project」」, <https://touhou-x.jp/> (最終閲覧日2021-1-17).

第29回年次大会予稿

神話物語と神話を原型にした現代物語の構造比較

Comparison of the structure of mythological stories and modern stories based on mythological prototypes

吉田拓海^{1*}, 村井源¹

Takumi YOSHIDA^{1*}, Hajime MURAI¹

¹ はこだて未来大学

Future University Hakodate

〒041-8655 北海道函館市亀田中野町116-2

E-mail: g2120050@fun.ac.jp

*連絡先著者 Corresponding Author

神話は物語の原型であると考えられており、物語自動生成において神話物語は大きな価値があるといえる。しかし、シンプルかつ基本的な物語構造をもつ神話とは異なり、現代物語の構造は複雑であると考えられている。そこで、神話物語の構造と神話を原型にした現代物語の構造を比較し、原型から物語生成を行う方法を考察する。分析の結果、神話物語の主人公の行動と現代物語の主人公の行動に異なる構造が存在することが明らかになった。

Mythology is considered to be the original form of storytelling, and mythological stories are of great value in automatic story generation. However, unlike myths, which have a simple and basic narrative structure, modern narratives are considered to have a complex structure. In this paper, we compare the structure of mythological stories with that of modern stories based on myths, and discuss how to generate stories from the original. As a result of the analysis, it is clear that there is a different structure between the behavior of the protagonist of the mythical story and that of the protagonist of the modern story.

キーワード: 計量文献学, 神話, 物語分析, 物語自動生成

Stylometry, Myth, Story analysis, Automatic story generation

1 はじめに

近年、人工知能の発展は目覚ましいもの

であり様々な分野で活躍している。物語自動生成については、松原ら[1]は星新一のショートショートを分析することで物語の

自動生成を目指した。分析した結果を基に自動生成された物語は星新一賞の一次選考を通過することに成功した。このことから、物語の構造分析の結果から、物語の自動生成が可能であることが示された。また現段階においては、自動生成の基となるデータから作成することが高水準の物語の自動生成につながると考えられる。分析対象は、対象となるデータ数が多く、物語の構造が明確であり、物語の特徴に関する研究が存在することが望ましい。

これらの条件を満たす題材として、英雄神話が挙げられる。英雄神話は世界各地に点在しており、多くの物語が存在している。また、Lévi-Strauss[2]、Campbell[3]などの研究から、英雄神話に特定の構造があると考えられている。神話学において多くの先行研究があるため、神話の特徴に関する知見が得やすい。神話学研究における分析は優れた分析者による主観要素が大きい結果であると考えられるが、吉田らは分析における評価基準を設け分析方法に客観性を持たせ、計量的に複数神話の比較分析を行った[4]。その結果、物語の特徴を抽出に成功し、従来の神話学研究成果と矛盾しなかった。

しかし、シンプルかつ基本的な物語構造をもつ神話とは異なり、現代物語の構造は複雑であると考えられている。そこで本研究では、神話物語の構造と神話を原型にした現代物語の構造を比較し、原型から物語生成を行う方法を考察する。

2 分析対象

本研究では、物語の特徴を計量文献学の手法を用いて客観的な分析を行うことを

目的としているため、計量的な分析に耐えられる作品数が必要となる。

神話に関しては、英雄神話は地域、文化によって異なる物語が豊富に存在する。そのため英雄神話は対象となる作品数は条件を満たす。また、多くの物語が神話をモチーフにしていること[5]や、神話学において多くの先行研究が存在していることから、神話物語の特徴に関する知見が得やすいため本研究で得られた特徴について評価を行うことができる。以上の理由から本研究では、神話を分析対象にした。

先行研究の豊富さから多くの知見が利用できることため、ギリシア神話を分析対象とした。また、文化的差異が現れやすく、比較研究が容易であることから近隣地域の神話である北欧神話、ケルト神話を分析対象に加えた。

各神話の分析対象作品の物語数は表1のとおりである。物語数は名前の付いた登場人物の物語上の退場で計算した。分析には、ギリシア神話は岩波書店の『ギリシア神話』[6]、北欧神話は新潮社の『エッダー古代北欧歌謡集』[7]、ケルト神話は筑摩書房の『ケルトの神話—女神と英雄と妖精と』[8]を用いた。

表1 神話物語の分析対象作品の物語数

研究対象作品	物語数
ギリシア神話	53
北欧神話	26
ケルト神話	21
合計	100

現代物語については、神話をモチーフにした作品をAmazon[9]、まんが王国[10]、ebook japan[11]、コミックシーモア[12]でキーワード「神話」で検索し、売上ランキン

グ上位の中から完結済みの作品を選出した。各作品の分析した巻数と物語数は表2の通りである。物語数は神話物語と同様に名前の付いた登場人物の物語上の退場で計算した。

表2 現代物語の分析対象作品の物語数

研究対象作品	巻数	物語数
聖闘士星矢[13]	7	34
南Q阿伝[14]	6	21
バビロニアの獅子[15]	4	5
巨竜戦記[16]	3	5

3 分析手法

本研究では、物語とは課題解決のプロセスであると考え、課題を物語を構成する要素の単位とし課題間の関係から物語の構造を分析する。課題とは物語内での解決すべき問題や障害を指し、主人公の行動の目標となる。解決すべき問題や障害は、例えば物や人が傷つけられた、または盗まれたなどの、主に主人公、もしくは主人公が敵対していない集団が受けた被害が当てはまる。ギリシア神話の『ヘーラクレースの獅子退治』は、獅子（敵）が牛（主人公と同じ集団に所属する存在）を殺したため、獅子退治が課題となる。北欧神話の『スリュムの歌』は、トール（主人公）のミョルニル（持ち物）が盗まれた（欠けている）ため、ミョルニル探しが課題となる。ケルト神話の『エマーとの結婚』は、ク・ホリン（主人公）に嫁がない（欠けている）ため、ク・ホリンの結婚が課題となる。

上記の課題間の関係から物語の構造を構築するために本研究は以下の手順で分析を行った。

1. 物語分割
2. 課題の抽出
3. 課題の構成要素のカテゴリ分類

まず、物語分割を行う。分割基準は表3に示す。これらの基準は、演劇や映像作品などで考えられている物語の単位であるシーン[17]と、登場人物に着目したProppを始めとする構造主義における物語の構造分析の単位である機能[18]を組み合わせた分割基準である。

表3 物語分割の基準

シーン	場所/舞台の移動
	物語内時間が不連続となる時間経過
機能	主要な登場人物の増減
	物語内で語られる主体の交代

次に、分割した物語から課題を抽出する。課題は課題の「発生」、「目的」、「告知」、「行動」、「結果」で構成される。基準は表4に示す。課題の抽出について、ケルト神話の『クーリーの牛争い』を例に示す。『クーリーの牛争い』はコノートの女王メイヴが夫と財産自慢をすることから始まる。財産自慢に負けそうになったメイヴは価値のあるクーリーの牛を手に入れよう命令を出す。この課題を抽出すると表5のようになる。このようにして物語から課題を抽出していく。

次に課題の各構成要素の抽出した物語内容を、カテゴリに分類する。発生は5種類、目的は4種類、告知は3種類、行動は7種類、結果は6種類となる。各カテゴリの定義を表6に示す。

表 4 課題の構成要素とその定義

課題の構成要素	定義
発生	課題の原因が登場する
目的	課題の内容, 解決方法, 行動の理由
告知	課題を認識し, 解決が始まる
行動	課題を解決するための行動
結果	行動の結果

表 5 課題の抽出例

	物語内容
発生	財産自慢に負けそうだから
目的	財産自慢に勝つために
告知	王の命令
行動	クーリーの牛を手に入れようとする
結果	—

表 6 課題のカテゴリ分類とその定義

		定義
発生	被害にあう	課題の原因によって, 傷つけられる・妨害されるなどの被害にある場合.
	欠如している	課題の原因が, 人や物などなくなる・足りていないなどの欠けている状態である場合.
	害を与える	課題の原因が, 禁を破る・傷つけるなど自身の加害行動である場合.
	予言	課題が発生する前に, 課題の解決方法が伝えられる場合.
	契約	約束, 決まり事などにより課題が発生する場合.
目的	原因の排除	課題の原因を取り除くことが解決する場合.
	欠如の補充	欠けているものを取り戻す, または別のもので置き換えることで解決する場合.
	予言に従う	特定の行動で解決する場合.
	契約を守る	契約を守ることで解決する場合.
告知	依頼	他者から伝えられることで, 解決のための行動を始める場合.
	自発	自分から, 解決のための行動を始める場合.
	緊急	突然襲われる・災害に遭うなど解決せざる負えない状況の場合.
行動	戦闘	戦闘行為をする場合.
	探索	人や物を見つけ出すための行動の場合.
	頼む	解決のために, 他者に行動を求める場合. 情報や許可を求める場合も含む.
	移動する	解決のために, 場所を移動する場合. 自身が移動する場合だけでなく, 他社や物を移動させる場合も含む.
	受諾する	解決のために, 特定の行動や条件を受け入れるの場合.
	拒否する	解決のための行動をとらない場合.
	神の助け	超自然的存在や主人公以外のキャラクターによって解決される場合.
	結果	原因の排除
結果	欠如の補充	欠けているものを取り戻す, または別のもので置き換えた場合.
	到着する	特定の場所へ移動した場合.
	発生を防ぐ	行動により, 新たな課題が発生しなくなる場合.
	被害にあう	解決に失敗し, 新たに被害を受ける場合. または, 原因を取り除けない場合.
	欠如する	解決に失敗し, 新たに欠如する場合. または, 欠如したままの場合.

4 結果

分析した作品ごとの各カテゴリの出現

頻度を表7にまとめる. また, 神話物語と現代物語について各カテゴリの出現頻度の有意差を調べるために, 項目ごとに χ^2 乗

検定の残差分析を行った。残差分析の結果を表8にまとめる。表8中の記号▲は5%有意水準で多い、▲▲は1%有意水準で多い、▽は5%有意水準で少ない、▽▽1%有意水準

で少ないことを示している。表9は各作品について、物語数と課題数、物語当たりの平均課題数をまとめた表である。

表7 カテゴリ出現頻度

		ギリシア 神話	北欧神話	ケルト 神話	聖闘士 星矢	南の阿伝	ア バビロ ンの 獅子	巨 竜 戦 記
発 生	被害にあう	81	22	32	68	43	12	10
	欠如している	83	38	26	29	32	22	9
	害を与える	45	15	13	3	0	2	0
	予言	2	9	5	2	1	1	2
目 的	契約	28	15	12	27	2	4	1
	原因の排除	125	53	35	64	43	11	10
	欠如の補充	112	50	37	45	32	25	9
	予言に従う	6	14	5	1	2	0	2
告 知	契約を守る	25	9	23	17	1	5	1
	依頼	93	51	40	39	24	13	6
	自発	129	64	43	45	33	11	10
行 動	緊急	66	12	16	46	21	17	6
	戦闘	119	21	28	47	26	9	4
	探索	36	20	12	20	15	3	5
	頼む	30	25	14	1	6	2	2
	移動する	19	10	8	14	0	4	2
	受諾する	34	29	21	9	13	7	3
	拒否する	15	9	9	7	1	3	2
	神の助け	21	4	5	22	15	11	2
結 果	原因の排除	100	31	27	48	31	9	6
	欠如の補充	102	51	23	29	24	13	7
	到着する	25	12	7	14	6	4	0
	発生を防ぐ	12	10	10	0	3	1	0
	被害にあう	30	21	23	13	3	3	4
欠如する	10	10	7	6	2	3	1	

5 考察

本研究の結果から、現代物語の複雑性は複数の登場人物の視点変更によって起こると考えられる。

まず、現代物語は主人公が課題の解決に関与しないという特徴があると考えられる。表8から神話物語と比較し、現代物語では「行動-神の助け」が多いことがわかる。

これは、課題が主人公以外の登場人物の行動によって解決される特徴を示している。同様に表8から現代物語は「行動-頼む」が少ないことから、主人公と協力関係にない登場人物による課題の解決が多いことが示されている。以上のことから、現代物語では主人公が課題の解決に関与していない場合が多いことが特徴であると考えられる。

表8 神話物語と現代物語についての
残差分析

		神話物語	現代物語
発生	被害にあう	▽▽135	▲▲133
	欠如している	147	92
	害を与える	▲▲73	▽▽5
	予言	16	6
	契約	55	34
目的	原因の排除	213	128
	欠如の補充	199	111
	予言に従う	▲25	▽5
	契約を守る	57	24
告知	依頼	184	82
	自発	▲236	▽99
	緊急	▽▽94	▲▲90
行動	戦闘	168	86
	探索	68	43
	頼む	▲▲69	▽▽11
	移動する	37	20
	受諾する	84	32
	拒否する	33	13
	神の助け	▽▽30	▲▲50
	結果	原因の排除	158
欠如の補充	176	73	
到着する	44	24	
発生を防ぐ	▲▲32	▽▽4	
被害にあう	▲74	▽23	
欠如する	27	12	

表9 神話物語と現代物語の課題数について

	課題	物語	平均課題
ギリシア神話	290	53	5.5
北欧神話	139	26	5.3
ケルト神話	100	21	4.8
神話物語	529	100	5.3
聖闘士星矢	130	34	3.8
南Q阿伝	78	21	3.7
バビロニアの獅子	41	5	8.2
巨竜戦記	22	5	4.4
現代物語	271	65	4.2

次に、現代物語は神話物語と比較して、主人公が課題の原因ではないことが特徴

であると考えられる。表8より「発生-被害にあう」が多く「発生-害を与える」が少ないことから、現代物語では主人公が被害者であることが特徴的である。また、表8より現代物語の「告知-緊急」が多いことは、主人公が巻き込まれる課題が多いことを示している。以上のことから、現代物語では主人公が原因の課題が少ないと考えられる。これは、現代物語では読者が作中人物に共感することが重要であると考えられているため、主人公に非のない行動をとらせるのは確からしいといえる。

現代物語の主人公が課題の解決に関与しない特徴と主人公が課題の原因ではない特徴から、現代物語では主人公以外の登場人物の行動によって課題の発生・解決が起こると言え、主人公以外の登場人物の行動でも物語が進むと考えられる。つまり、神話物語における主人公の行動が現代物語では主人公以外に割り当てられているとも言える。また、表9から神話物語と現代物語の物語当たり含まれる平均課題数に大きな差はないといえる。よって、現代物語が煩雑な課題数により複雑化しているとは考えられない。以上のことから、現代物語の複雑性は、神話物語と同様の課題に対して、複数の視点で描いていることが原因であると考えられる。このことから、神話を原型として物語を生成する際、神話物語の主人公の行動を分割し複数人に割り当てることで現代物語らしい物語を生成することが可能になると考えられる。

6 結論と今後の課題

本研究では、神話物語の構造と神話を原型にした現代物語の構造を比較し、原型から物語生成を行う方法の考察を行った。神

話物語と比較し、現代物語は主人公以外の行動によって物語が進む特徴があり、神話物語における主人公の行動を現代物語では複数人に割り当てていると考察した。

今後の課題として、本研究の考察である現代物語は神話物語の主人公の行動を複数に割り当てることで構成されていることを検証する必要がある。また、本研究の客観性をより高めるために、複数分析者による計量的な一致度の検証を行う必要がある。

参考文献

- [1] 松原仁；佐藤理史；赤石美奈；角薫；迎山和司；中島秀之；瀬名秀明；村井源；大塚裕子：「コンピュータに星新一のようなショートショートを創作させる試み」, The 27th Annual Conference of Japanese Society for Artificial Intelligence, 2D1-1, 2013.
- [2] Claude, Lévi-Strauss (荒川幾男(訳), 生松敬三(訳), 川田順造(訳), 佐々木明(訳), 田島節夫(訳))：「構造人類学」, みすず書房, 1972.
- [3] Joseph, Campbell (倉田真木(訳), 斎藤静代(訳), 関根光宏(訳))：「千の顔を持つ英雄」, 早川書房, 2015.
- [4] 吉田拓海；村井源：「物語自動生成に向けて物語要素間の関係に着目した神話物語の構造分析」, じんもんこん 2020 論文集 2020, pp. 143-148, 2020.
- [5] Christopher, Vogler (岡田勲(訳))：「神話の法則」, ストーリーアーツ&サイエンス研究所株式会社, 2002.
- [6] Apollodoros (高津春繁(訳))：「ギリシア神話」, 岩波書店, 1953.
- [7] 谷口幸男(訳)：「エッダー古代北欧歌謡集」, 新潮社, 1973.
- [8] 井村君江：「ケルトの神話—女神と英雄と妖精と」, 筑摩書房, 1990.
- [9] Amazon：https://www.amazon.co.jp/ (2020年4月23日参照)。
- [10] まんが王国：https://comic.k-manga.jp/ (2020年4月23日参照)。
- [11] ebook japan：https://ebookjapan.yahoo.co.jp/ (2020年4月23日参照)。
- [12] コミックシーモア：https://www.cmoa.jp/ (2020年4月23日参照)。
- [13] 車田正美：「聖闘士星矢」, 集英社, 1995.
- [14] 光永康則：「南Q阿伝」, 講談社, 2012.
- [15] 氷栗優：「バビロニアの獅子」, 祥伝社, 2011.
- [16] 本田真吾：「巨竜戦記」, 講談社, 2019.
- [17] Robert, Stam; Sandy, Flitterman-Lewis; Robert, Burgoyne (丸山修(訳), 深谷公宣(訳), エグリントンみか(訳), 森野聡子(訳))：「映画記号論入門」, 松柏社, 2006.
- [18] Vladimir, Iakovlevich, Propp (北岡誠司(訳), 福田美佐代(訳))：「昔話の形態学」, 水声社, 1987.

第 29 回年次大会予稿

クエスト構造に注目したロールプレイングゲームの

物語構造と物語内容分析手法の提案

Proposal of method for analyzing story structure and content of role-playing games focusing on quests structure

中村 祥吾^{1*}, 村井 源¹

Shougo NAKAMURA^{1*}, Hajime MURAI¹

¹ 公立はこだて未来大学

Future University Hakodate

〒041-8655 北海道函館市亀田中野町116-2

*連絡先著者 Corresponding Author

近年、ロールプレイングゲームの物語の構造抽出と自動生成が可能であることが明らかになった。しかし、分析に取り入れられていたのは物語における連続構造だけであり、より複雑かつ自由な物語展開の自動生成には入れ子構造、並列構造と言った物語構造を取り入れる必要がある。また、内容についても分析に取り入れる必要がある。そこで、本研究においてはロールプレイングゲームにおけるクエスト間の連続構造、入れ子構造、並列構造といった基本的な物語構造と物語内容の記述・分析手法を提案した。提案手法を用いて合計 331 個のクエストにおける基本的な物語構造の記述と抽出を行い、それと関連させて物語内容も記述と抽出を行った。そして、抽出結果と実際の物語の比較を行うことで、提案手法の有効性を検証した。今後の課題として、提案手法の客観性の担保を行う検証の必要がある。

Recently, it was clarified that structure extraction and automatic generation of the story of the role-playing game were possible. However, only the continuous structure in the story was adopted in the analysis. The story structure such as nesting structure and parallel structure should be adopted for the automatic generation of the more complicated and free story development. The story contents also need to be included in the analysis. In this paper, I propose a method to describe and analyze basic narrative structure, such as continuous structure, nested structure, and parallel structure between quests in role-playing games. Using the proposed method, we describe and extract the basic narrative structure in a total of 331 quests, and the narrative contents are also described and extracted in relation to them. The effectiveness of the proposed method is verified by comparing the extracted results with actual

stories. As a future problem, it is necessary to verify the objectivity of the proposed method.

キーワード: 物語, 構造分析, ロールプレイングゲーム(RPG)

Stories, Structural analysis, Role-playing games (RPG)

1 まえがき

近年, 種々の媒体から物語の構造を分析・抽出を行い, その結果を元に物語の自動生成をすることは可能であると示されてきた。例えば, 星新一のような掌編小説を創作させる松原らによる試み[1]が行われている。

「きまぐれ人工知能プロジェクト 作家ですのよ」と呼ばれるこの試みは, 実際に星新一賞の一次選考を通過することに成功した。このように物語の自動生成を行うにあたり, 物語論の古典的な手法[2]を用いつつ計量的に分析・抽出を行う方法が村井らにより提案されてきた[3, 4]。物語を計量的に分析する際, 特定のジャンルに着目して物語構造を数値として抽出し, その構造に基づいて物語を自動生成するといった手法が試みられている。実際に物語構造の抽出, 物語の自動生成が可能であることを示されたジャンルとして, 探偵小説[5, 6], 怪談や戦闘物[7]などがあげられる。

英雄譚や神話などに強く影響を受けており物語の構造がよく似通っている Role-Playing Game (RPG)の物語の構造抽出と自動生成も, 同様の手法を用いて齊藤ら[8]により成功している。しかし, 現状では入れ子構造, 並列構造などを考慮できていない物語構造が存在しており, 十分に分析できたとは言えない。

本研究では, 分析手法として基本的な物語構造である連続構造, 入れ子構造, 並列構造に注目した物語構造の計量的な分析手法を提案する。

2 分析対象

分析対象として, 物語構造の分析のために物語の質が良い作品(以下, 良作)を選ぶ必要があった。品質の良さを客観的に保証するため, 2005~2018年における年間売上数[9]を参考に対象作品を選定した。また, 分析可能な作品数の担保, 今後の比較検討のしやすさ, 分析結果の安定化, 以上の理由から, シリーズとして続いている作品を対象とすることとした。よって, 年間売上上位 10 位に登場する作品の売上数をシリーズ毎に計上し, シリーズ総合売上が多い上位 3 シリーズを選定した。次に, 3 シリーズそれぞれにおいて累計売上数上位から 1 作品ずつを分析対象作品とした。これらより, 『ポケットモンスター』[10]シリーズから『ポケットモンスター 赤/緑』(以下, ポケモン赤), 『ドラゴンクエスト』[11]シリーズから『ドラゴンクエストIX 星空の守り人』(以下, DQ9), 『FINAL FANTASY』[12]シリーズから『FINAL FANTASY VIII』(以下, FF8), 以上 3 作品を最終的な分析対象作品として選定した。

3 分析手法

3.1 クエスト単位での物語分析

本研究では, RPG の物語はクエストの連続で構成されているとみなし, クエスト間の関係を抽出することを目的とした物語構造の分析手法[8, 13, 14]を用いた。そのため, 物語をクエスト単位で区切り, クエスト ID を割り振ることで物語の分割を行った。ここで, 本研究において「クエスト」とは, 「主人公達

が冒険する先々で起こる、報酬や成長を伴う障害や試練の始まりから終わりまでの一連の流れ」と定義した。さらにクエストを「発生」「経過」「結末」の3シーンに区切った。

「発生」はクエストに取り組むきっかけを描写している場面とした。「経過」は発生と結末の間の、主人公がクエスト達成に向けて取り組んでいる場面とした。「結末」はクエストを達成して結果的にどうなるかが描写されている場面とした。

3.2 物語構造分析手法

まず、本研究におけるそれぞれの構造の定義と記述方法[14]の概要を説明する。

連続構造とは、あるクエストの前にはこのクエストが、後にはこのクエストが続いているといった構造である。入れ子構造とは、あるクエスト(以下、親)のあるシーンの途中で別のクエスト(以下、子)が挟まっている構造である。並列構造とは、あるクエストが進行している最中に、同時に別のクエストも進行している構造である。

この様な、あるクエストにおける基本的物語構造の記述と抽出を行った。

3.3 物語内容分析手法

最初に、実際にキャラクターがとった行動とその結果を「内容」として自然言語で記述した。そして、頻出する「内容」に抽象的なタグをつけて分類していった。例えば、怪物と戦った、兵器と戦った、などと自然言語で記述する。これらで描写されている内容は戦闘行為である。よって、これらは「戦闘」という抽象的なタグに分類していく。このようにしてタグ付け・分類を行なった結果、RPGの物語内容は74種に分類された。そして、似通ったタグを16種のカテゴリーにまとめた。この16種のカテゴリーを「大カテゴリー」、74種のタグを「小カテゴリー」として

表2-1, 2-2に示す。そして、シーン毎に内容を記述していった。また1シーンにつき複数のタグを記述することも可能とした。

3.4 カイ二乗検定

カイ二乗検定を行うことで、ある事象が起こる頻度に有意な偏りがあるかを調べることができる。本研究では、抽出した物語内容の3作品間での傾向を分析するために使用した。また、入れ子構造と並列構造に注目した場合の傾向も分析することにした。

入れ子構造においては、ある内容のクエストの親または子としてどのような内容のクエストが現れる傾向がRPG作品全般にあるかを分析した。また、ある入れ子構造の組み合わせがどの作品で有意に偏りが存在するかを調べた。並列構造においても同様に、RPG全体での傾向と作品間での比較を行なった。

入れ子構造において3作品間の比較を行なった際、ポケモン赤におけるクエスト数が少なかったため、いくつかのカテゴリーを削除して検定を行った。これはコクランルールに則るためである。そこで、他の2作品における削除した部分の傾向を調べるために、補助的な検定としてDQ9, FF8間での比較も行なった。

行なった検定は6種であり、3作品間のクエスト内容の比較、入れ子構造における3作品間の比較、DQ9とFF8間の補助検定、RPG全体での傾向の検定、並列構造における3作品間の比較、RPG全体での傾向の検定である。以下に、検定の際に行なったデータ加工について説明する。

- ・ 3作品間比較：コクランルールに則り、データの20%以上で期待数が5未満にならないようにいくつかのカテゴリーを除外した。

- 入れ子構造においては
- 3 作品間比較：各行各列を見た時にデータ数の少ない組み合わせのものを削除した。また、コ克蘭ルールに則ってデータの 20%以上で期待数が 5 未満にならないようにいくつかのカテゴリーを除外した。
 - DQ, FF 比較：3 作品間比較を行なった際にポケモン赤の影響で除外されたカテゴリーの中で、「ハプニング」を加えた場合の検定を行なった。
 - 全体：各行各列を見た時にデータ数の少ない組み合わせのものを削除し、さらに残った行列にコ克蘭ルールに則って削除した。
- 並列構造においては、3 作品間比較と全体のそれぞれで入れ子構造と同様の処理を行なった。

表 2-1 クエスト内容タグ一覧

大カテゴリー	小カテゴリー
移動・探索	探索, 捜索, 発見, 移動, 撤退逃走, 追跡
移動経路	移動経路の開通, 移動経路の封鎖, 移動経路の変更, 移動経路の限定, 目的地への到達
アイテム	アイテム入手, 移動手段入手, アイテム譲渡, アイテム使用, 運搬
情報提示	謎伏線の提示, 謎の存在の提示, 謎の施設の提示, 情報入手分析, 謎の回収, 伏線の回収, 世界設定の開示, 謎の存在の開示, 目的開示, 衝撃の事実の開示, 黒幕の開示

表 2-2 クエスト内容タグ一覧

会話	会話
戦闘	戦闘
ハプニング	敵の奇襲, 敵の復活, 敵の増援, 会敵, 危機, 誘拐, 拘束監禁, 暴走, 凶兆, 理不尽な要求, 迷子, 仲間が怖気づく, 能力の喪失
救助	救助
生死不明	生死不明
死亡	死亡
生還	生還
陣営変化	陣営変化, 仲間との合流, 分断行動
成長	成長
関係構築	関係構築
発生動機	依頼任務, 情報入手分析(誘導), 決意
結末	成功, 部分的成功, 勝利, 達成, 立場の向上, 防衛, 妨害, 失敗, 敗北, 立場の悪化, 中断
その他	移動(隠密), 待機, 拘束監禁の解放, 能力回復, 騙す, 災害, 困窮, 殺害, 激励

4 分析・考察結果

ポケモン赤と DQ9 と FF8 の物語内容の分析をカイ二乗検定によって行った。3 作品間比較を表 4 に示す。入れ子構造を絡めた 3 作品間比較の一部抜粋したものを表 5、RPG 全体での傾向の検定を表 6 に示す。並列構造を絡めた、3 作品間比較

の一部抜粋したものを表 7、RPG 全体での傾向の検定を表 8 に示す。各表においては、右に「▲」と書かれているカテゴリーが有意に多く、右に「▽」と書かれているカテゴリーが有意に少ないことを意味している(5%有意水準)。

物語内容を抽出した結果を物語構造と絡めて順番に考察していく。

表 4 ポケモン赤, DQ9, FF8 比較

	ポケモン赤		DQ9		FF8	
移動探索	111	▲	130		163	▽
移動経路	57	▲	62		41	▽
アイテム	41	▲	48	▲	16	▽
情報提示	18	▽	89	▲	69	▽
情報開示	7	▽	51	▲	42	
会話	46		105		153	
戦闘	40	▲	30	▽	57	
ハプニング	3	▽	53		120	▲
救助	7		18		40	▲
陣営変化	3	▽	32		54	▲
関係構築	5		16		35	▲
発生動機	8	▽	92	▲	91	
結末	35		84		138	▲
その他	1	▽	10		32	▲

表 5 3 作品比較(入れ子構造)一部抜粋

親子	ポケモン赤		DQ9		FF8	
情報提示_移動探索	53		30	▽	133	▲
会話_移動探索	67	▲	26	▽	154	▲
結末_移動探索	107	▲	48	▽	70	▽
移動探索_情報提示	10	▽	47		92	▲
移動探索_会話	27	▽	72		93	
情報提示_会話	29		55		109	▲
会話_会話	33	▽	46	▽	154	▲
結末_会話	44	▲	49		57	▽
移動探索_結末	17	▽	56		82	
会話_結末	25		31	▽	112	▲

表 4 より、同時期に売り出されている人気のゲーム同士であっても物語内容に違いがあることが計量的に明らかとなった。ポケモン赤などの低年齢層・ライトユーザー向けのゲームでは移動探索、移動経路、アイテム、戦闘などが優位に多

い。これは移動中に障害物が現れ、それをアイテムや戦闘をすることで取り除き、また移動をするといった単純な物語にしていると考えられる。FF8 などの高めの年齢層・ヘビィユーザー向けのゲームでは別のカテゴリーが多い。主人公の行動

の動機付けのためにハプニング、NPC との深い人間関係を描くために救助、陣営変化、関係構築、劇的かつ独特な物語展開のため結末、その他が優位に多い。そして、両者の中間である DQ9 のようなゲームではアイテム、情報提示、情報開示、

発生動機が多い。単純にアイテムを使用するだけではなく、アイテムなどの背景・設定を説明することでより物語を複雑にしている。しかし、「発生動機」によって行動指針を明確に示すことで物語の進行を容易にしていると考えられる。

表 6 全体比較(入れ子構造)

	移動探索		移動経路		情報提示		会話		ハプニング		結末
移動探索	147	▽	126	▲	149	▲	192		112		155
情報提示	216		68	▽	121		193		118		137
会話	247		98		120		233		141		168
ハプニング	176	▲	52	▽	71		129		123	▲	84
陣営変化	78	▽	27	▽	42		90		78	▲	95
発生動機	111		52		72		120		70		88
結末	225	▲	112	▲	97		150	▽	82	▽	147

表 7 3 作品比較(並列構造)一部抜粋

	ポケモン赤		DQ9		FF8
移動探索_移動探索	517	▲	22	▽	76
移動探索_移動経路	76	▲	11		2
移動探索_情報提示	46	▽	26	▲	51
移動探索_アイテム	284	▲	30		15
移動探索_会話	163		20	▽	83
移動探索_結末	113	▲	4	▽	35
情報提示_会話	16	▽	26	▲	50
アイテム_アイテム	150	▲	42	▲	2
アイテム_会話	91		32	▲	18
会話_会話	50	▽	22		106

表 8 全体比較(並列構造)

	移動探索		情報提示		アイテム		会話		ハプニング		結末
移動探索	615	▲	123	▽	329	▲	266	▽	64	▽	152
情報提示	123	▽	57	▲	70		92		56	▲	40
アイテム	329	▲	70		194	▲	141		21	▽	71
会話	266	▽	92		141		178		85	▲	85
ハプニング	64	▽	56	▲	21	▽	85	▲	82	▲	34
結末	152		40		71		85		34		50

入れ子構造を絡めた RPG 全体としての特徴としては、「移動探索」をしている途中で道が塞がっているまたは情報収集の必要が出てきたり、「ハプニング」や「陣営変化(お

そらく、仲間やライバルが敵になる展開)が行われたりすると、「移動探索(おそらく「撤退逃走」)」「ハプニング(敵の奇襲など)」が起きるなど一般的な物語の展開が抽出されて

いる。

並列構造を絡めた RPG 全体の特徴としては、同カテゴリーの組み合わせ(「移動探索」_「移動探索」など)が多いこと、「移動探索」と「ハプニング」は何と組み合わせるかがはっきりとしていることが挙げられる。同カテゴリーの組み合わせが多い理由は、RPG の物語では並列的に進行させるクエストや、攻略順が自由にしてあるクエストは、内容や物語上の意味を同等にするまたは対比させるためだと考えられる。「移動探索」の組み合わせにおいては、新たな情報や出来事が移動中に並列的に発生するようなことでプレイヤーを混乱させないために、「情報提示」や「会話」や「ハプニング」関連の出来事を起こさないよう配慮していると考えられる。「ハプニング」の組み合わせにおいては「移動探索」の組み合わせとは逆で、プレイヤーをより混乱させ興奮させるために「情報提示」や「会話」や「ハプニング」が並列的に発生するようになっていると考えられる。

このように、本研究の手法によって抽出された結果からは各作品と RPG 全体における様々な傾向を計量的に分析できたと言える。特に、表 6, 8 などの RPG 全体の傾向は、一般的な RPG の物語自動生成に有用であると考えられる。更に対象ユーザーに応じて、表 5, 7 などの作品毎の傾向を用いて差別化を図ることが有効と考えられる。

5. まとめ・展望・課題

本研究では、基本的な物語構造に注目した RPG の物語構造と物語内容の分析手法を提案し、物語構造と合わせて物語内容を分析することを試みた。結果として、作品毎の物語内容の比較、RPG 全体の物語内容の傾向の分析、入れ子・並列構造を絡めた物語内容の分

析を行うことができた。得られた傾向や比較結果は作品や RPG 全体のそれぞれの要素に合致していると考察できた。

また今後の課題として、クエストの区切り方、物語内容の記述・考察などにおける客観性の担保が挙げられる。これについては、本研究におけるクエストの区切り方や基本的物語構造、タグの内容を明確に定義した上で、第二分析者と記述・解釈の一致度の検定を行う予定である。

今後の展望として、本研究の手法で抽出された物語構造と絡めた物語内容を用いることで、RPG におけるクエストと物語プロットの自動生成に向けての一助になることが期待される。抽出された基本的な物語構造がどれほどの確率でどのような構造としてつながり、また内容はどのようなクエストとなるかを整理することで、クエストと物語プロットの自動生成は可能と考えられる。また、これらの質や精度を向上させるためにも分析作品を増やしていくことを予定している。

参考文献

- [1] 松原仁, 佐藤理史, 赤石美奈ほか: 「コンピュータに星新一のようなショートショートを創作させる試み」, *The 27th Annual Conference of Japanese Society for Artificial Intelligence*, 2D1-1, 2013.
- [2] ウラジーミル・プロップ (北岡誠司, 福田美智代訳). 昔話の形態学, 水声社, 1987.
- [3] 村井源, 松本斉子, 佐藤知恵, 往住彰文: 「物語構造の計量分析に向けて-星新一のショートショートの特徴-」, *情報知識学会誌*, Vol. 21, No. 1, pp. 6-17, 2011.
- [4] Hajime Murai: "Automatic Extraction of Reversal-Type Punch Lines in Shinichi Hoshi's Flash Fictions", *Journal of the Japanese*

Association for Digital Humanities, Vol. 2, pp. 31-47, 2017.

[5] 村井源: 「推理小説の自動生成のためのトリックと推理想行動の構造化」, 人工知能学会第二種研究会ことば工学研究会資料, 2020, SIG-LSE-B903, p. 27-32.

[6] 豊澤修平, 工藤はるか, 石田晃大, 遠藤史央里, 川瀬稜人, 菊池亮太, 工藤健太郎, 栗原将風, 櫻井健太郎, 佐藤好高, 玉置秀基, 根本裕基, 原科充快, 久野露羽, 平田郁織, 村井源, 椿本弥生, 角薫, 松原仁: 「推理小説プロットを自動生成し映像化する統合的インタラクティブシステムの開発と評価」, 情報処理学会研究報告人文科学とコンピュータ, Vol. 2018-CH-116, No. 13, pp. 1-5, 2018.

[7] 鈴木諒輔, 佐々木奨之, 袴田翔, 田中瑞穂, 三浦隆太郎, 城田晃希, 高橋翔太, 南部太雅, 山田康貴, 吉田拓海, 松浦史佳, 松原千里, 寺島啓悟, 津沢慎吾, 渡邊広基, 村井源, 迎山和司, 田柳恵美子, 平田圭二, 角薫, 松原仁: 「物語と情景描写を自動生成する統合的システムの検討と開発」, 情報処理学会研究報告, Vol. 2018-EC-50, No. 28, pp. 1-8, 2018.

[8] 斉藤勇璃, 白石智誠, 太田和宏, 根本さくら, 石川一稀, 宇田朗子, 小川卓也, 友広純々野, 中村祥吾, 山内拓真, 西川和真, 宍

戸建元, 長野恭介, 蓬畑旺周, 稲垣武, 村井源, 迎山和司, 田柳恵美子, 平田圭二, 角薫, 松原仁: 「シナリオ・視覚要素・音響効果を統合的に自動生成するゲームシステムの構築」, The 32th Annual Conference of the Japanese Society for Artificial Intelligence, 4C2-GS-13-03(PDF), 2020.

[9] ゲームソフト年間売上 - Game Compass, <http://gcompass.sp.land.to/rank/>, (参照 2020-09-4).

[10] ポケットモンスター, 任天堂株式会社, 1996.

[11] ドラゴンクエスト, 株式会社スクウェア・エニックス, 1986.

[12] FINAL FANTASY, 株式会社スクウェア・エニックス, 1987.

[13] 中村祥吾, 石川一稀, 稲垣武, 宇田朗子, 太田和宏, 斉藤勇璃, 白石智誠, 長野恭介, 根本さくら, 山内拓真, 村井源, 平田圭二, 迎山和司, 田柳恵美子: 「クエストの連続構造を用いたRPG向け物語の自動生成」, 人工知能学会第二種研究会ことば工学研究会資料, SIG-LSE-B903, pp. 15-20, 2020.

[14] 中村祥吾, 村井源: 「クエスト構造に注目したロールプレイングゲームの物語構造分析手法の提案」, じんもんこん 2020 論文集, pp. 149-156, 2020.

第 29 回年次大会予稿

情報収集行動の振り返り支援を目的とする閲覧履歴提示手法 A Visualization Method of Browsing History to Support Reflection on Information Gathering Behavior

佐藤 千尋^{†*}, 高久 雅生^{††}

Chihiro SATO^{†*}, Masao TAKAKU^{††}

[†] 筑波大学 大学院 人間総合科学学術院 人間総合科学研究群 情報学学位プログラム

Graduate Program in Informatics; Graduate School of Comprehensive Human Sciences; University of Tsukuba

〒305-8550 茨城県つくば市春日1丁目2

E-Mail: chihiro@klis.tsukuba.ac.jp

^{††} 筑波大学 図書館情報メディア系

Faculty of Library, Information and Media Science; University of Tsukuba

〒305-8550 茨城県つくば市春日1丁目2

E-Mail: masao@slis.tsukuba.ac.jp

* 連絡先著者 Corresponding Author

閲覧履歴は生活の様々なシーンにおけるインターネットを通じた情報収集行動を記録したデータで、個人経験を記録・保存するライフログの一部である。本研究は日常生活の記録としての閲覧履歴活用を目指す。用途として過去の行動を分析し自己理解や行動改善を獲得する振り返りに着目する。提案システムは閲覧履歴内の主題分析によって Web ページの内容想起、主題と閲覧時系列に基づく閲覧履歴の可視化によって閲覧の背景想起を支援する。本システムの振り返り支援効果を検証するため、実際の閲覧履歴を入力に用いて評価を行った。結果、内容想起について、Web ページの特徴語抽出が成功した場合は十分な支援が得られた。背景想起について、主題と時系列の組み合わせにより可視化した閲覧履歴によって当時の思考や生活イベント等が想起された。提案システムを用いた振り返りは当時の情報収集行動や生活に関する有益なリフレクション獲得を促すと示唆された。

Browsing history is the data that records the information gathering behavior through the Internet in many situations in our lives, and is a part of the lifelog that records and stores personal experiences. This study aims to utilize the browsing history as a record of daily life, focusing on self-reflection, which is the process of analyzing past actions to gain self-understanding and behavior improvement. The proposed system supports the recall of contents of browsed web pages by analyzing the topics in the browsing history, and the recall of the background of browsing by visualizing the browsing history based on the topics and the browsing time series. In order to investigate the effectiveness of the proposed system, we conducted an evaluation using actual browsing histories as input. The results showed that the system provided sufficient support for content recall when the feature words of the web pages were successfully extracted. As for the background recall, the visualization of the browsing history using the combination of the topics and the time series helped the subject recall the thoughts and events of the time. Also the proposed system provided useful reflections on the information gathering behavior and life at that time.

キーワード: Web ページ閲覧履歴, 情報収集行動, 振り返り, 主題抽出, 可視化

Keywords: Web browsing history, information gathering behavior, reflection, topic extraction, visualization

1 はじめに

インターネットを通じた情報収集は今や人々の生活に溶け込んでいる。2019年の総務省の調査^[1]では、個人のインターネット利用率は約89.8%である。またインターネットを毎日少なくとも1回は利用すると回答した者は全利用者の71.1%である。加えて情報収集を目的にインターネットを利用すると回答した者は64.2%である。

現在はスマートフォンのように Web ブラウザを

内蔵したポータブルデバイスが普及し、生活の様々なシーンで情報収集が可能である。Web ブラウザが生成する情報収集記録が Web ページ閲覧履歴である。閲覧履歴には生活の様々なシーンにおける情報収集が記録されていることから、日常も特別なイベントも包括した生活の記録となりうる。なお降は Web ページ閲覧履歴を閲覧履歴と表記する。

閲覧履歴はライフログの一部である^[2]。ライフログは多様なセンサで捉えた個人経験のデジタル記録で、個人的アーカイブとして保存される^[3]。個人的アーカイブ活用を目的としてライフログを分



図 2: 閲覧履歴に対する振り返りの流れ

析・提示する手法は、ライフログデータベースの MyLifeBits^[4] 以降盛んに研究されている。

Web ブラウザの世界シェア第 1 位は Google Chrome である^[5]。Chrome では閲覧履歴が図 1 のように Web ページの閲覧時刻・アイコン・タイトル・URL を時系列順に並べたリストとして提示される。



図 1: Chrome による閲覧履歴提示

リスト型提示が提供する情報は閲覧した Web ページのメタデータに留まっている。Web ページの内容を知りたい場合、タイトル等に基づく推測、もしくは該当ページの再訪問が必要である。また全 Web ページが一列に表示され、異なる背景のもと閲覧した Web ページの区別が容易でない。以上の理由からリスト型提示は生活の記録として使用しづらい。

記録の用途には記録が作成された当時の振り返りがある。振り返りとは過去の行動や出来事等を想起し、その分析や考察を通じて自己理解や行動プロセス改善につなげる行為である。振り返りの成果として得られる自己理解や行動プロセス改善はリフレク

ションと呼ばれる。Choe ら^[6] は個人的なデータの適切な可視化は文脈の想起を支援し、過去の行動に関する考察や新たな気づきを得る充実した振り返りを促すと示唆している。

本研究では、閲覧履歴に対する振り返りの流れを図 2 のように定義する。初めに閲覧した Web ページの内容を想起する。次に想起した内容に基づき、閲覧の背景にある興味・思考や行動のプロセス・生活イベント等の背景を想起する。続いて想起した背景を通じて当時の自分を振り返り、リフレクションを獲得する。内容と背景の想起がうまくいかない場合、振り返りもうまくいかず、有益なリフレクションが得られない。

本研究は、当時の情報収集行動や生活について有益なリフレクションを獲得可能な振り返りを支援する閲覧履歴提示システムの構築を目的とする。本システムが情報収集行動や生活の振り返りを支援すると確認できたなら、日々多くの情報に触れるユーザを記憶や情報管理の面で補助する実用システムとして提案可能となる。

著者らによる 2020 年の発表^[7] では、Web ページの主題と閲覧時系列とを組み合わせた閲覧履歴提示システムとその評価実験から得られた改善点を報告した。本稿では主題分析と可視化の手法を中心に改善した提案システムの構築経緯と機能の説明、システム評価とその結果を報告する。

2 関連研究

2.1 振り返り支援

リフレクションはあるプロセスについて反省・考察する振り返りを通じて得られる学習効果である。榎本ら^[8] はリフレクションがメタ認知と相互に影響し合い、自己理解や過去に習得した知識の応用を深めるとまとめている。情報学分野でも情報にまつわる行動の振り返りやリフレクション獲得を支援する研究がある。齋藤ら^[9] は Web 情報検索プロセス提示による検索行動のリフレクション支援を研究している。Deng ら^[10] は Web 閲覧履歴の振り返り促進を目的とし、時系列や位置情報等を利用する再発見型検索アルゴリズムを提案している。

閲覧履歴にはユーザの生活における多様なニーズから生じた情報収集行動が記録されており日常生活の振り返りに役立つと考えられる。しかし閲覧履歴と日常生活を結びつけて振り返りを促す研究は広く知られておらず、本研究でその支援手法を提案する。

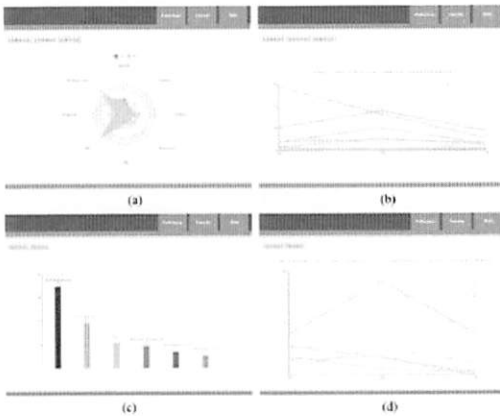


図 4: Gu らによる主題分類の統計的可視化 [17]

2.2 ライフログ活用

ライフログは人々の生活を記録したデータである。デバイスの記憶容量拡大やセンサーの高機能化などにより、ライフログの蓄積を支える環境が拡充されている。膨大なデータは適切な組織化や提供によって活用の幅が広がり、ライフログも例外ではない。最近ではライフログに対する対話的検索アプローチを比較する Lifelog Search Challenge が開催されており、様々なプロジェクトが参加している [11][12]。

検索を行うには初めに目的を定める必要があり、記憶にない行動や生活イベント等を検索アプローチを通じて想起するのは手間がかかる。本研究ではライフログに記録された行動や生活イベント等の想起を促す取り組みとして、ライフログを日記のように閲覧するシステムの構築を目指す。

Prelipcean ら [13] は写真や移動履歴等のライフログから旅行中の行動記録を自動生成している。また村上ら [14][15] は閲覧履歴含む複数のライフログからユーザの思考や興味等を再現し、その想起を支援するシステムを開発している。

しかし日常的な情報収集行動を時系列を保持したプロセスとして提示する研究は広く知られていない。本研究は閲覧履歴に着目し、当時の行動プロセスや思考、生活イベント等の振り返りを支援するシステムを構築する。

2.3 閲覧履歴可視化

閲覧履歴は様々な目的に基づき可視化されている。Xu ら [16] は情報検索プロセスの意味的関連に対する振り返り支援を目的とし、図 3 のように時系列・知識グラフ・検索結果を組み合わせて提示している。

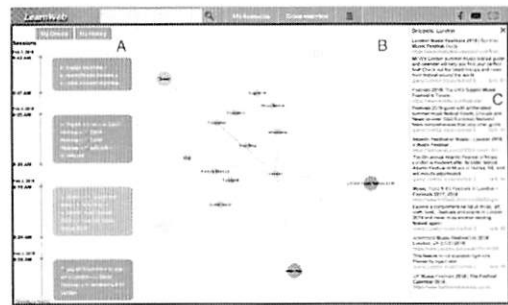


図 3: Xu らによる時系列・知識グラフ・検索結果画面を組み合わせた可視化 [16]

Gu [17] らはユーザの関心分析を目的とし、閲覧履歴に対する主題分類の結果を図 4 のように統計的に提示する手法を開発している。

以上の閲覧履歴可視化手法からは閲覧した内容と時系列両方の読み取りは支援されておらず、情報収集行動やその背景にある生活の振り返りを目的としていないことが伺える。本研究では情報収集行動や生活の振り返りを目的とした閲覧履歴可視化を行う。次章にて具体的な提案システムについて説明する。

3 提案手法

本研究の目的は Web ページ閲覧履歴を用いて情報収集行動の振り返りを支援するシステムの提案である。振り返り支援として主題提示、閲覧過程提示の手法を用いる。主題提示では Web ページの特徴量抽出とクラスタリングを用いて主題を抽出し、可視化する。閲覧過程提示では Web ページの主題が時系列に沿って変遷する様子を可視化する。本システムは Mac 上で開発され、開発言語には Python3.7.5 を使用している。入力には Google Chrome が生成する閲覧履歴データベースで、出力は主題提示と閲覧過程提示の手法による閲覧履歴提示である。システムの構成を図 5 に示す。

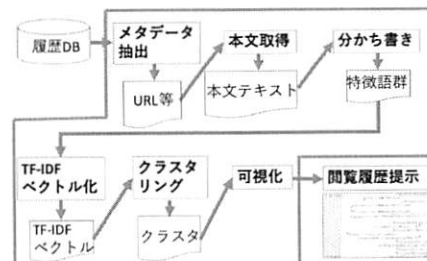


図 5: 提案手法を実現するシステムの全体像



図 6: 振り返りにおける内容想起の位置付け

メタデータ抽出から TF-IDF ベクトル化までの説明は佐藤らの報告 [7] を参照されたい。なお利用したライブラリは以下の通りである。

- メタデータ抽出: sqlite3^[18]
- 本文取得: Requests^[19], BeautifulSoup^[20]
- 分かち書き: MeCab^[21]
- TF-IDF ベクトル化: scikit-learn^[22]

以下、クラスタリングと可視化について説明する。

3.1 クラスタリング

閲覧履歴内 Web ページ群の主題を分析するため、各 Web ページの特徴量を表す TF-IDF ベクトルに対してクラスタリングを行う。クラスタリング手法は g-means 法を用いた。hamerly ら^[23]は、ガウス分布に基づいてクラスタ数を決定する g-means 法の方が BIC に基づいてクラスタ数を決定する x-means 法よりも正解に近い数のクラスタを得られることを検証している。g-means 法クラスタリングは次の手順で行われる。

1. クラスタ数の初期値を得る
2. 各クラスタの中心点をランダムに割り振る
3. 全ての点 (対象データ) と各クラスタの中心点との 2 乗ユークリッド距離を計算する
4. 全ての点について中心点が最も近いクラスタに割り当て直す
5. クラスタの割り当てが変化しなくなるまで、3. 4. を繰り返す
6. 各クラスタについて、クラスタ内データがガウス分布に従うか統計的検定で確かめる
 - (a) ガウス分布に従う場合はクラスタを保持
 - (b) ガウス分布に従わない場合はクラスタを 2 分割し、6. に戻る
7. クラスタの割り当てが変化しなくなるまで 6. を繰り返す

なお利用したライブラリは pyclustering^[24] である。

3.2 可視化

閲覧履歴の背景にある当時の興味・思考・行動・生活イベント等を再現するため、閲覧履歴を閲覧した Web ページの主題と時系列に基づき可視化する。可視化では主題分析結果と閲覧履歴の 2 つを出力する。利用したライブラリは matplotlib^[25] である。

3.2.1 主題分析結果の可視化

主題分析結果可視化の目的は主に閲覧した内容想起の支援である (図 6 参照)。

ここでは 3.1 で得られたクラスタ群とそれぞれの特徴語を一覧表示する。クラスタの特徴語として、クラスタ内で TF-IDF 値が上位 5 件の特徴語とその値を表示する。また循環カラーマップの hsv^[26] を用いて各クラスタに色ラベルを付与する。

図 7 に可視化の一部を示す。左から順に python を用いた配列の結合、日時データの処理、matplotlib を用いたグラフの描画を主題とする Web ページ群であることが伺える。

連結(0.598)	datetime(0.498)	matplotlib(0.621)
python(0.508)	python(0.491)	legend(0.448)
結合(0.478)	日付(0.487)	python(0.361)
配列(0.435)	変換(0.444)	横線(0.36)
リスト(0.416)	文字列(0.409)	グラフ(0.331)

図 7: 主題分析結果の可視化

3.2.2 閲覧履歴の可視化

閲覧履歴可視化の目的は主に閲覧した背景想起の支援である (図 8 参照)。



図 8: 振り返りにおける背景想起の位置付け

閲覧当時の興味・思考や行動のプロセス・生活イベントといった背景は、閲覧した Web ページの主題が時系列に沿って変遷する様子によって表現可能であると考えられる。

可視化では閲覧日時を縦軸に、クラスタの番号を横軸にとった空間に閲覧履歴内の Web ページを配置する。画面の上に行くほど閲覧日時は新しくなる。これは既存の Web ブラウザにおけるリスト型提示に倣っている。1 つの Web ページをクラスタの色ラベル・閲覧時刻・タイトルの 3 要素で構成する。

また、主題と時系列のいずれも切り替わる位置を閲覧履歴におけるセッションの区切りとみなす。Web ページ間の閲覧日時の差が 600 秒以上、かつクラスタ番号が異なっている場合にセッションの区切りと判定する。可視化ではセッションが異なる Web ページ間に実線を引いて表現する。

図 9 は可視化の例である。下から順に、日英翻訳、メール、学年暦、python での日時データ処理や配列処理、つくば市内の新型コロナウイルス感染症等に関する Web ページの閲覧が伺える。なおセッションを区切る実線上に配置された Web ページは、線の下にあるセッションに属す。

表 2: 主題分析結果提示の評価結果 (DS:データセット)

DS	背景の数	背景の具体例 (想起を促したクラスタ番号)
0	4	新年の挨拶がいつまで使えるか調べた (6,16)
1	9	Teams と Google カレンダーの連携を試みた (1,2,9,11~21)
2	9	CSR などベクトル圧縮手法を実装した (11,16)
3	24	スクレイピングにおける timeout 処理を実装した (13,14,15)
4	9	エスプレント号やザメンホフ氏について調べた (25,41,42)
5	14	IT・ソフトウェア業界や企業を研究した (4)
6	8	matplotlib での図のサイズ設定を調べ実装した (1,20,22,23)
7	23	VSCoDe での ipynb ファイル編集手法を調べた (44)
8	17	Slack と連携できるタスク管理サービスを探した (53,54)
9	16	matplotlib で散布図の凡例の実装を試みた (36,75)
10	25	unicode decode error の対処法を調べ、解決した (15,21)

表 3: 閲覧履歴提示の評価結果 (DS:データセット)

DS	背景の数	背景の具体例
0	7	授業資料を Google ドライブにアップロードした
1	9	19 時頃に友人同上で予定を共有する話になり、Google カレンダーと Teams の連携を試みた
2	4	matplotlib で可視化を実装しつつ、メール通知や聞いている音楽に気を取られた
3	12	ベクトル圧縮手法調査中に先輩とおしゃべりが始まり、話題になったタイの景色等を調べた
4	2	講演で聞こえた「ザメンホフ」「エスプレント号」等を検索した
5	3	k-means 法の課題であるクラスタ数推定を調べ、g-means 法を実装した
6	1	淡色を示すカラーコードを調べようとしたら淡い色の洋服のページを開いてしまった
7	15	企業説明会に参加した後に採用マイページを作成し、採用日程をカレンダーに登録した
8	10	就活相談会で話題に出たテストセンターを調べたり就職課の情報を確認したりした
9	11	自宅の TV をモニタとして使い始めたので PC 画面を TV に映す設定を調べ、企業説明会を視聴した
10	10	採用試験対策を長時間行い、煮詰まって友達の研究室へ行き、勧められた企業や漫画を調べた

4.2 データセット

筆頭著者の閲覧履歴を提案システムにて分析・可視化した。全データの概要を次に示す。

- 総ページ数: 5,365 ページ
- 期間: 2021/1/2 16:35~2021/3/31 19:27

評価用データセットとして閲覧履歴を 1 週間 (168 時間) 毎に区切り、計 11 セットを作成した。各評価用データセットのページ数、クラスタ数、セッション数、期間を表 1 に示す。

表 1: 評価用データセット概要

番号	ページ	クラスタ	セッション	期間 (2021)
0	230	28	23	1/2~1/7
1	921	76	38	1/11~1/19
2	180	35	36	1/20~1/28
3	1206	114	66	1/28~2/4
4	248	55	40	2/5~2/12
5	311	57	49	2/14~2/21
6	127	30	28	2/22~3/1
7	557	97	54	3/2~3/9
8	454	94	51	3/10~3/18
9	588	88	65	3/18~3/26
10	543	139	67	3/26~3/31

5 結果

5.1 主題分析結果提示の評価

各評価用データセットにおいて想起された背景の数と具体例を表 2 に示す。データセット全体として、4 から 25 の背景が想起され、各背景には平均 5.15 個のクラスタが属する。

想起を促したクラスタの例として、表 2 のデータセット 1 におけるクラスタの提示を図 10 に示す。図 10 のうち実線で囲まれているのは「Teams と Google カレンダーの連携を試みた」背景の想起を促したクラスタ番号 12・20 である。

それぞれの特徴語は「teams(0.58)、同期(0.552)、outlook(0.538)、会議(0.333)、カレンダー(0.332)」、「カレンダー(0.707)、共有(0.623)、google カレンダー(0.592)、予定(0.417)、設定(0.337)」である。

5.2 閲覧履歴提示の評価

各評価用データセットの計測結果と具体例を表 3 に示す。データセット全体として、1 から 15 の背景が想起され、各背景には平均 9.68 個のクラスタが属する。

具体例として、表 3 のデータセット 1 における「帰宅前 19 時頃に友人同上で予定を共有する話にな

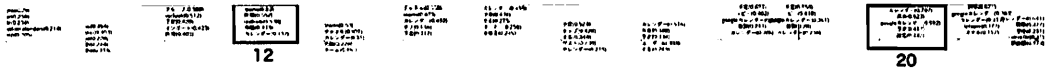


図 10: 「Teams と Google カレンダーの連携を試みた」背景の想起を促した主題分析結果提示 (データセット 1)

表 4: 獲得したリフレクション

データセット	リフレクションの例
0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10	1日の始まりにメールやカレンダー、新卒採用マイページを見ていることが多い
3, 5, 7	隔週のゼミ中に他の学生のテーマに関して論文や記事を探している
3, 5, 9	システム構築を本文抽出、ベクトル作成、クラスタリング、可視化と進めているが時々別の工程を手直ししようとしている
6, 7, 8, 9, 10	2月下旬から予稿執筆と就職活動を並行しており、就職活動の優先度が高かった
1, 2, 3	作業中に通知や音楽で気が散っており、対策すると効率上がりそう
2, 9	グラフの凡例表示やベクトル圧縮など実装を諦めた機能があった
0, 1, 5, 9, 10	pythonの辞書やリストの扱いをしよっちゅう調べている

表 5: 各評価用データセットのページ数・クラスタ数・背景数

データセット	ページ数	クラスタ数	背景数
0	230	28	4
1	921	76	9
2	180	35	9
3	1206	114	24
4	248	55	9
5	311	57	14
6	127	30	8
7	557	97	23
8	454	94	17
9	588	88	16
10	543	139	25

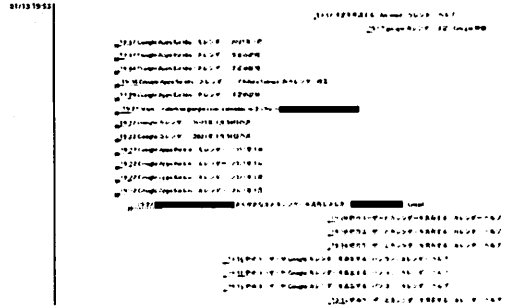


図 11: 「帰宅前 19 時頃に友人同上で予定を共有する話になり、Google カレンダーと Teams の連携を試みた」背景の想起を促した閲覧履歴提示 (データセット 1)

り、Google カレンダーと Teams の連携を試みた」背景の想起を促した閲覧履歴提示の一部を図 11 に示す。個人情報が含まれるため、一部を黒塗りした。

5.3 リフレクション獲得の評価

5.1 節、5.2 節から得られた自己理解、行動プロセス改善等のリフレクションは 12 件であった。表 4 にリフレクションの元となったデータセットと具体例を示す。

6 考察

6.1 主題分析結果提示の効果

提案システムの主題分析結果提示により閲覧履歴内の特徴語を読み取ることができた。「Teams と Google カレンダーの連携を試みた」や「CSR などベクトル圧縮手法を実装した」のように独立した短時間の背景の想起に効果があった。

主題分析結果提示では、複数のクラスタから 1 つの背景が想起される場合やクラスタから背景が想起されない場合があった。表 5 に各データセットのページ数とクラスタ数、主題分析結果から想起された背景の数の対応を示す。

主題分析結果では特徴語抽出精度が想起支援を左右している。図 12 のように特徴語抽出の精度が良ければ閲覧した内容や背景が想起される反面、図 13 のように精度が良くない場合ほとんど何も想起できなかった。

いつ(0.596)
あけましておめでとうございます(0.401)
あけましておめでとう(0.24)
新年(0.214)
使える(0.162)

図 12: 特徴語抽出の成功例 (データセット 0 におけるクラスタ番号 6)

15(0.727)
webā(0.382)
mynavicorporationä(0.258)
okā(0.258)
offā(0.258)

図 13: 特徴語抽出の失敗例 (データセット 0 におけるクラスタ番号 16)

特徴語抽出精度には限界があるため、今後さらなる想起支援として Web ページタイトルの表示を検討する。

また図 7 のように近隣同士のクラスタが類似した主題を持ち、背景の想起を支援する場合は観測された。要因と考えられるのは、クラスタ群を統計的検定に基づき適宜分割する g-means 法クラスタリングである。

ただ、表 2 に示したデータセット 9 の例では「matplotlib で散布図の凡例の実装を試みた」背景の想起を支援したのは 36,75 と離れたクラスタであった。クラスタ内の Web ページを確認したところ、36 には Google 検索結果、75 には質問投稿サイトがそれぞれ含まれていた。同様の主題でもプラットフォームが異なると離れたクラスタに割り当てられる場合があり、改善策の検討を要する。

6.2 閲覧履歴提示の効果

提案システムの閲覧履歴提示により時系列に沿った主題の変遷を読み取ることができた。「19 時頃に友人同上で予定を共有する話になり、Google カレンダーと Teams の連携を試みた」や「ベクトル圧縮手法調査中に先輩とおしゃべりが始まり、話題になったタイの景色等を調べた」のように、連続した長時間の背景の想起に効果があった。

本章では 3 つの機能を中心に効果を考察する。

6.2.1 閲覧日時とクラスタ番号を用いた Web ページ配置

横軸上の位置が主題の異同を表し、内容の想起が促された。また短時間で複数の主題を閲覧している場合には共通する背景があると示唆され、生活イベントや行動プロセス等の想起が促された。

反対に、特徴語を確認したい場合は主題分析結果提示を見る必要があり、想起の妨げとなった。今後は閲覧履歴提示上での特徴語表示を検討する。

6.2.2 セッション区切り線

閲覧履歴のセッションは当時どんな背景があったか考えるきっかけとして働いた。また幅広の提示画面を閲覧する際にガイドラインとして役立った。

6.2.3 クラスタの色ラベル

各クラスタに付与された色ラベルは 6.2.1 項に述べた主題の異同を伝える視覚効果を補強した。

提案システムではクラスタ数に応じ動的に色ラベルを付与する目的で、循環型カラーマップからクラスタ番号に基づき色ラベルを生成している。すなわち近隣のクラスタは近い色のラベルが与えられる。

隣り合うクラスタの主題分析結果が類似している場合、含まれる Web ページの内容が同様であると視覚的に読み取りやすくなる。しかし隣り合うクラスタの主題が異なる場合、視覚的に混乱を生む可能性がある。

改良のため、クラスタ番号ではなく主題の類似度に基づき色ラベルを付与する手法を検討する。

6.2.4 その他

ブラウザ以外のアプリケーションを中心に使っている場合、当時の行動に関して閲覧履歴から得られる情報が少ない。他アプリからブラウザへの遷移履歴も合わせて提示できれば当時の背景想起に役立つ可能性がある。

また閲覧履歴の主題分析結果が多い場合、提示がかなり幅広になり可視性が下がる。評価では主題分析結果が 55(データセット 4, ページ数 248) 以上になると見づらさを感じた。今回用いたデータセットは全て 1 週間分だが、それでも現在の可視化手法の限界が示唆された。

主題分析の結果の数はページ数に比例する傾向がある。長期間の閲覧履歴を振り返る場合、現在の可視化手法には改善が必要である。

6.3 リフレクション獲得効果

提案システムは閲覧履歴を通じて有益な自己理解や行動プロセス改善等のリフレクション獲得を目指している。

評価を通じて得られたリフレクションは「1日の始まりにメールや新卒採用マイページを見ていることが多い」のような習慣、「隔週のゼミ中に学生のテーマに関して論文や記事を探している」のような生活イベントやそれに応じた情報収集行動、「本文抽出、ベクトル作成、クラスタリング、可視化、と進めているが、ときどき別の工程を手直ししようとしている」のような自身の行動理解、「作業中に通知や音楽で気が散っており、対策すると効率が上がろう」のような行動改善案等が得られた。

今回の評価では1週間ごとのデータセットを用いたが、リフレクションは複数のデータセットを総合して得られた。閲覧履歴全体にはリフレクションを促す要素が遍在していることが推察される。

またリフレクションは主題分析結果提示と閲覧履歴提示の双方から想起された背景を総合的に振り返った結果獲得された。主題分析結果によって閲覧した内容を思い出し、続く閲覧履歴可視化によってその内容のWebページを閲覧した背景が時系列を伴って想起された。振り返り支援においては、閲覧履歴に対する主題分析も主題と時系列を組み合わせた可視化も効果を持つと考えられる。

筆頭著者の主観では、記憶になかった当時の生活や行動に関するリフレクションが十分に得られた。しかし振り返りやリフレクション獲得は主観によるところが大きく、筆頭著者のみを対象とする評価では効果があるとは言い切れない。

7 おわりに

本研究は、当時の情報収集行動や生活について有益なリフレクションを獲得可能な振り返りを支援する閲覧履歴提示システムの構築を目的とする。図2のうち内容想起支援として閲覧履歴内Webページ主題分析、背景想起支援として主題と閲覧時系列に基づく閲覧履歴可視化を中核とするシステムを構築した。

さらに、筆頭著者の約3ヶ月に渡る閲覧履歴データセットを用いて提案システムの効果を評価した。内容想起について、主題分析においてWebページの特徴語抽出が成功すれば十分に支援可能であった。背景想起について、主題分析結果のみの提示は独立した短時間の背景の想起を促した。また、分析された主題と時系列を組み合わせた閲覧履歴提示は連続した長時間の背景の想起を促した。

加えてリフレクション獲得について、筆頭著者は提案システムを通じた振り返りを行い、当時の情報収集行動や生活に関する自己理解や行動プロセス改善等を獲得した。

以上の評価から、提案システムは当時の情報収集行動や生活に関する有益なリフレクション獲得を支援すると示唆された。

本研究の限界として、目的とするリフレクション獲得が主観的效果である点が挙げられる。そのため今回の筆頭著者1名のみに対する評価では効果が確認されたとはいえない。今後は協力者を募って同様の評価実験を行い、提案システムの効果をさらに検証する。

また、可視化手法にもいくつかの改善点が示唆された。より効果的な振り返り支援を目指すにあたり、可視化手法の研究・改善は今後の重要な課題である。

参考文献

- [1] 総務省: 「令和2年版情報通信白書インターネットの利用状況」, 2021/4/8 参照. <https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r02/html/nd252120.html>.
- [2] 総務省: 「「利用者視点を踏まえたICTサービスに係る諸問題に関する研究会」第二次提言の公表」, 2021/4/8 参照. https://www.soumu.go.jp/menu_news/s-news/02kiban08_02000041.html.
- [3] Dodge, Martin; Kitchin, Rob: “Outlines of a world coming into existence: pervasive computing and the ethics of forgetting”, *Environment and planning B: planning and design*, Vol. 34, No. 3, pp. 431–445, 2007.
- [4] Gemmell, Jim; Bell, Gordon; Lueder, Roger; Drucker, Steven; Wong, Curtis: “MyLifeBits: Fulfilling the Memex Vision”, *Proceedings of the Tenth ACM International Conference on Multimedia, MULTIMEDIA '02*, p. 235–238, New York, NY, USA, 2002. Association for Computing Machinery.
- [5] statcounter: “browser Market Share Worldwide”, 2021/4/8 参照. <https://gs.statcounter.com/browser-market-share>.
- [6] Choe, Eun Kyoung; Lee, Bongshin; Zhu, Haining; Riche, Nathalie Henry; Baur, Dominikus: “Understanding Self-Reflection: How People Reflect on Personal Data through Visual Data Exploration”, *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare, PervasiveHealth '17*, p. 173–182, New York, NY, USA, 2017. Association for Computing Machinery.

- [7] 佐藤千尋; 高久雅生: 「Web ページ閲覧履歴を用いた情報収集行動の振り返り支援」, 情報知識学会誌, Vol. 30, No. 2, pp. 220–229, 2020.
- [8] 榎本晶; 緒方紀子; 水戸美津子: 「看護におけるリフレクションとメタ認知の考察」, 聖徳大学研究紀要, Vol. 28, pp. 151–156, 2017.
- [9] 齋藤ひとみ; 三輪和久: 「Web 情報検索におけるリフレクションの支援」, 人工知能学会論文誌, Vol. 19, No. 4, pp. 214–224, 2004.
- [10] Deng, Tangjian; Zhao, Liang; Feng, Ling; Xue, Wenwei: “Information re-finding by context: A brain memory inspired approach”, *Proceedings of the 20th ACM international conference on Information and knowledge management*, pp. 1553–1558, 2011.
- [11] Gurrin, Cathal; Joho, Hideo; Hopfgartner, Frank; Zhou, Liting; Ninh, V-T; Le, T-K; Albatal, Rami; Dang-Nguyen, D-T; Healy, Graham: “Overview of the NTCIR-14 lifelog-3 task”, *Proceedings of the 14th NTCIR conference*, pp. 14–26. NII, 2019.
- [12] Gurrin, Cathal; Le, Tu-Khiem; Ninh, Van-Tu; Dang-Nguyen, Duc-Tien; Jonsson, Bjornor; Lokoc, Jakub; Hurst, Wolfgang; Tran, Minh-Triet; Schoffmann, Klaus: “Introduction to the Third Annual Lifelog Search Challenge (LSC’20)”, *Proceedings of the 2020 International Conference on Multimedia Retrieval, ICMR ’20*, p. 584–585, New York, NY, USA, 2020. Association for Computing Machinery.
- [13] Prelipcean, Adrian C; Gidófalvi, Győző; Susilo, Yusak O: “MEILI: A travel diary collection, annotation and automation system”, *Computers, Environment and Urban Systems*, Vol. 70, pp. 24–34, 2018.
- [14] 村上晴美; 平田高志: 「WWW からの情報獲得・整理支援・思考・興味空間ブラウザ」, 情報処理学会研究報告自然言語処理 (NL), Vol. 2001, No. 20 (2000-NL-142), pp. 167–174, 2001.
- [15] Murakami, Harumi; Mitsuhashi, Kenta: 「A System for Creating User’s Knowledge Space from Various Information Usages to Support Human Recollection」, *International Journal of Advancements in Computing Technology*, Vol. 4, No. 22, pp. 496–508, 2012.
- [16] Xu, Luyan; Fernando, Zeon Trevor; Zhou, Xuan; Nejdil, Wolfgang: “LogCanvas: visualizing search history using knowledge graphs”, *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pp. 1289–1292, 2018.
- [17] Gu, Yeong Hyeon; Yoo, Seong Joon; Piao, Zhegao; Lin, Yinhe; Yan, Jiangzhi; Park, Jung Hwan: “User preference analysis and visualization through the browser history of smart devices”, *Proceedings of the 2015 International Conference on Big Data Applications and Services*, pp. 264–267, 2015.
- [18] Foundation, Python Software: 「sqlite3 — SQLite データベースに対する DB-API 2.0 インタフェース — Python 3.9.4 ドキュメント」. <https://docs.python.org/ja/3/library/sqlite3.html>, 2021/4/8 参照.
- [19] Reitz, Kenneth: 「Requests: HTTP for Humans™ — Requests 2.25.1 documentation」. <https://docs.python-requests.org/en/master/>, 2021/4/8 参照.
- [20] Richardson, Leonard: “Beautiful Soup: We called him Tortoise because he taught us.”. <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>, 2021/4/8 参照.
- [21] 工藤拓: “MeCab: Yet Another Part-of-Speech and Morphological Analyzer.”. <https://taku910.github.io/mecab/>, 2021/4/8 参照.
- [22] developersscikit-learn : 「sklearn.feature_extraction.text.TfidfVectorizer — scikit-learn 0.24.1 documentation」. https://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html, 2021/4/8 参照.
- [23] Hamerly, Greg; Elkan, Charles: “Learning the k in k-means”, *Advances in neural information processing systems*, Vol. 16, pp. 281–288, 2003.
- [24] pyclustering: “pyclustering.cluster.gmeans.gmeans Class Reference”. https://pyclustering.github.io/docs/0.9.2/html/d8/d3c/classpyclustering_1_1cluster_1_1gmeans_1_1gmeans.html, 2021/4/8 参照.
- [25] team”, ”The Matplotlibdevelopment : 「Matplotlib: Python plotting — Matplotlib 3.4.1 documentation」. <https://matplotlib.org/>, 2021/4/8 参照.
- [26] team, The Matplotlibdevelopment : 「Matplotlib でのカラーマップの選択 — Matplotlib3.4.1 ドキュメント」. <https://matplotlib.org/stable/tutorials/colors/colormaps.html>, 2021/4/8 参照.

第 29 回年次大会予稿

時系列を考慮したクエリ満足度の推定

The Query Satisfaction Prediction Considering Time Sequence

楊 之卓^{1*}, 高久 雅生²

Zhizhuo YANG^{1*}, Masao TAKAKU²

- 1 筑波大学 大学院 人間総合科学学術院 人間総合科学研究群 情報学学位プログラム
Graduate Program in Informatics; Graduate School of Comprehensive Human Sciences; University of Tsukuba
〒305-8550 茨城県つくば市春日 1-2
E-mail: s2021714@s.tsukuba.ac.jp
- 2 筑波大学 図書館情報メディア系
Faculty of Library, Information and Media Science; University of Tsukuba
〒305-8550 茨城県つくば市春日 1-2
E-mail: masao@slis.tsukuba.ac.jp
*連絡先著者 Corresponding Author

情報検索プロセスでは、ユーザが複数のクエリを入力し検索を繰り返すことで目的を達成することが多い。従って、ユーザの検索意図をより正確に発見するため、検索行動からユーザのクエリに対する満足度の推定は重要な課題の 1 つとなる。本研究ではクエリを時系列に並べ、ユーザのクエリレベルの満足度を推定する手法を提案する。提案手法は、クエリ満足度の予測によく使われているクリックデータ、間隔時間とクエリ修正を特徴量とする。加えて、修正クエリと先行クエリの検索結果との関連はクエリ満足度に影響を与えるかを考慮する。これら 4 つの特徴量を組み合わせてクエリ満足度の推定を行う。評価実験から、クエリとドキュメントの関連度により、推定効果がクエリ修正のみと比べて約 6 ポイント向上させたことが分かったが、4 つの特徴量を使う際、71.27%の正解率となり、最良の結果が得られなかった。本論文で、それに対して失敗分析を行い、改善策について検討し報告する。

In information seeking process, a user often submits multiple queries and repeats the search to achieve the purpose. Therefore, in order to discover the user's search intention more accurately, predicting the query level satisfaction from user's search behavior is one of the important issues currently. In this research, we arrange the queries which submitted in a session by time sequence, and propose a method for predicting user's query level satisfaction. The proposed method uses click data, interval time and query reformulation as indicators, which are often used to predict query satisfaction. In addition, we examine the effect of similarity between reformulated query and the search results of the previous query on query satisfaction. We predict query satisfaction by combining these four features. From the evaluation experiment, it shows that comparing with query reformulation only, the estimation effect has been improved 6 points approximately by using the similarity between the query and the document, but when using all of four features, the accuracy is 71.27% which was not the best result. In this paper, we discuss results with a failure analysis, and further improvement measures.

キーワード: 検索行動, クエリ修正, 満足度推定, クエリ類似度

Keyword: search behavior, query reformulation, satisfaction prediction, query similarity

1 はじめに

1.1 研究背景

検索プロセスでは、ユーザが検索結果から情報を継続的に取得し、それらの情報に基づいてクエリを修正し、反復的に検索を繰り返すことで自身の求める情報ニーズを達成する^[1]。ユーザの各段階の検索意図はクエリとして表現されているが、検索対象に対するユーザの認識が欠如していたり、情報ニーズが不明確な場合では、数語からなるクエリで検索意図を完璧に表すことは容易でなく、試行錯誤をしながら複数のクエリを提出することがよく見られる。その場合、検索プロセスの進行に伴い、検索対象に対する認識が深まり、情報ニーズが不明確な状態から徐々に明確になり、クエリが情報要求に応じて、最後のニーズを満たすまで変わる。また、旅行計画を立てるための目的地情報検索や、論文執筆のための資料検索などのような複雑なタスクを検索する際に、入手したい情報が複数の要素から構成されていることで、それぞれの関連するクエリをそれぞれ提出し検索する必要がある^[1]。

このようにユーザの検索意図の遷移はクエリに反映されている。クエリ修正の分析を通じて、ユーザがクエリに満足しているか不満であるかを推定することができ、満足したクエリは、ユーザの検索意図を適切に表現した成功クエリと言える。

成功クエリには、大きく2種類がある：(1) 結果の成功 (task-level success)：検索目的を達成すること；(2) 過程の成功 (query-level success)：ユーザは現在のクエリからある一定の有用な情報を入手しており、検索を推進したこと (例：有用な検索キーワードを見つけた)。最終的に情報ニーズ全体を満たせる

検索タスクは成功であるとみなすことが一般的である。しかし、近年多くの研究がタスクレベルの成功以外に、クエリレベルの成功に着目している^[2]。本研究では、クエリレベルの成功を中心として検討する。

クエリ修正はクエリ満足度を推定する有用な特徴量であることが検証されてきた^[2]。従来の研究において、クエリ修正は常にクエリ不満足を示した予測因子として使われている。しかし、クエリ修正がタスクレベルの情報ニーズへの不満足を示すことはわかるが、クエリレベルの検索意図に満足したか否かについては検討の余地がある。

1.2 研究目的

実際の検索プロセスにおいて、検索結果から何らかの有用な情報が入手できれば、最終的な検索目的を達成されないものの、そのクエリによって検索プロセスそのものは推進されたと考えられる。従って、そのクエリが成功したと推定できると考えられる。その観点から、本研究では、修正されたクエリと元クエリの検索結果の関連度なども特徴量とみなし、クエリ満足度を推定する手法を提案する。またその有効性を検証する。

1.3 研究意義

クエリ満足度を推定できれば、ユーザが満足しているクエリ群を発見し、それらのクエリを類似した検索意図を持つユーザに推薦するような応用も期待できる。このように、ユーザが検索意図を表現するクエリを見つげるためのプロセス、コストを削減し、検索効率を向上させることができる。

2 関連研究

2.1 ユーザ行動とクエリ満足度

ユーザの検索行動分析に関する研究は盛んに行われている。その中で、検索ログデータを用いてユーザの検索満足度を推定することが重要な課題の1つである。

Fox ら^[1]はブラウザのアドオンを使用して、ユーザの関心に対する潜在的な指標 (implicit measures) とユーザ満足度間の関連を発見した。彼らはクリックスルー (Clickthrough, CTR)、滞留時間 (dwell time)、セッション終了行動を組み合わせて、検索結果ページの満足度を適切に予測できることを指摘した。Huffman と Hochster^[4]は幾つかの単純な関連性メトリクスを使用してユーザの最終的な検索セッションレベルの満足度の推定に与える影響を検証した。その関連性メトリクスとは、クエリを提出した後検索エンジンによって取得されたドキュメントの関連度で算出する。

Hassan ら^[2]はクリックデータに加えクエリの変更がクエリ不満足を表す重要な評価指標の1つであるということを示した。それら2つの指標をとともに考慮すると、ユーザのクエリに対する満足度の推定効果を向上させることが示された。

本研究では、既存研究に基づいて、クエリレベルの満足度を推定することを目指し、クエリ修正は単純な不満を指示する指標とするだけでなく、修正されたクエリと先行クエリの検索結果の関連度も考慮する。

2.2 クエリ修正

クエリ修正における既存の多くの研究はクエリ修正タイプの予測や分類器を作成することに注目している。関口ら^[5]は検索クエリの修正における分類が文字編集内容を基準とした分類と、ユーザの修正意図を基準と

した分類の2種類に分けた。前者の代表例では、Huang ら^[6]は単語の置き換えや、追加、削除などの文字編集によるクエリ修正のタイプ分類を定義し、各修正タイプにマッチングルールを制定した。後者のタイプの研究として、Jansen ら^[7]は検索ユーザの意図を用いて、Assistance、Content Change、Generalization、New、Reformulation、Specialization という6つのクエリ修正タイプを提案した。その中で、汎化 (Generalization)、絞込 (Specialization) という2つのクエリ修正タイプは本研究も用いる。

Huang ら^[6]は語彙に基づくクエリ修正分類器を作成した。それはセマンティックなクエリ修正は対象外となり、クエリ文字列における文字の置き換え、追加、削除、語句修正といった作業に注目した分類方法である。しかし、セマンティックなクエリ修正も含めてクエリ修正を予測することが重要である。

Hassan ら^[2]は WordNet Wu and Palmer measure (wup)^[8]をもとに、Jaccard 係数でセマンティック類似度の計算手法を提案した。

本研究は、クエリ修正タイプ及びクエリ類似度を用い、クエリ修正予測を行い、クエリの語彙、キーワード、セマンティックという3つの観点からクエリ類似度を計算する。一方で、本研究で Wu and Palmer の手法^[8]に基づいて異なるセマンティック類似度の計算手法を提案する。

3 クエリ修正予測

クエリ修正とは、ユーザが同様な意図に基づいて、より良い結果を得るため、元クエリを修正して再検索するという行動である。ここで、Hassan ら^[2]によるクエリ修正の定義を採用することで、検索意図が変更され、新た

な情報を目指して後続クエリは先行クエリの修正ではないとする。クエリ修正はクエリ満足度の推定に用いる重要な特徴量として、この章で、クエリ修正予測の手法を紹介する。

修正されたクエリは、同じ意図に基づいていることにより、元クエリと構文的に類似している可能性がある。そのため、クエリ間の類似度は、クエリ修正の予測に役立つと考えられる。先行研究により、クエリ類似度の手法については、語彙を基準とした手法 (lexical similarity) と、意味に基づいた手法 (semantic similarity) の2種類に大きく分類することができる。

3.1 語彙の共通度

クエリ類似度を測るため、Jaccard 類似度と呼ばれている2つのクエリのうち、共通単語が占める割合を計算するのが語彙を基準とした手法の代表的な1つである。Jaccard 類似度はベクトル間距離が求められ、以下の式で定義される。クエリの単語が、集合の要素と見なされ、式を用いて計算を行い、その結果が語彙の類似度として扱われる。

$$Jaccard(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (1)$$

語彙の共通度はクエリ間の類似度の表現に一定の効果があるが、大きな制限があることが過去の研究で示されてきた^[2]。まず、語彙の共通度が余計な単語の影響を受けやすい。余計な単語とは、主な検索意図が表明されていない単語であると考えられる。例えば、Q1: 「hotel in Tsukuba Ibaraki」と、Q2: 「restaurant in Tsukuba Ibaraki」という2つのクエリが連続的に提出された場合、Jaccard 類似度により、Q1とQ2の語彙共通度は60%と得られることが分かる。しかし、実際にQ1、Q2では、ホテルとレストランという2つの

違う検索対象がある。60%の類似度を得るのが「in Tsukuba Ibaraki」という地名がクエリの長さの大部分を占めるため、地名が重複している場合、ユーザが異なるオブジェクトを探しても、語彙共通度が高い可能性がある。

3.2 キーワードレベルの類似度

語彙の共通度の欠点を多少解消するため、キーワードという概念が導入された。キーワードレベルの類似度では、クエリはキーワードに分割され、キーワードレベルでどの程度共通しているかを計算する。そのため、クエリセグメンテーションが必要である。

クエリセグメンテーションはクエリを個々のフレーズもしくはセマンティックユニットに分割するプロセスである^[9]。近年、クエリセグメンテーション手法が多く、学者から提案された。既存手法において、教師あり学習がよく使われている^[9]。しかし、教師あり学習手法は優れた効果を実現するため大量なレベル付き訓練データが必要となることにより、応用するのが困難であり、コストが高くなる問題がある。そして、一部の研究では、訓練データを用いずに、教師なし学習を用いてクエリセグメンテーションを行う手法を提案した^{[10][11]}。その中で、代表的な一つは自己相互情報量 (point-wise mutual information, PMI) である^[6]。PMIとは、2つの事象の間の関連度合いを図る尺度であり、以下の式で算出できる。

$$PMI(x_i, x_{i+1}) = \frac{P(x_i, x_{i+1})}{P(x_i)P(x_{i+1})} \quad (2)$$

n 個単語で構成されるクエリ q を $q = (x_1, x_2, \dots, x_n)$ とする場合、 (x_i, x_{i+1}) はクエリ q に連続的に出現したバイグラムであり、 $P(x_i, x_{i+1})$ はバイグラム (x_i, x_{i+1}) が同時に出現する確率、 $P(x_i)$ 、 $P(x_{i+1})$ は単語 x_i 、 x_{i+1}

個々の出現確率である。

本研究で、セグメンテーションは、二つの連続的な単語間 PMI が特定の閾値 γ を下回るたびに導入される。ここで、閾値 γ を 1.0 に設定する。単語および ngram の出現頻度を求めるため、公開されている自然言語コーパス Google Books Ngram Dataset (Version2) [12][14] を用いる。

3.3 セマンティック類似度

クエリセグメンテーションにより、個々の単語の代わりにセマンティックな単位での類似度に取り組むことが可能となる^[10]。2つのキーワードは直接に語彙上の共通がなくても、意味的に類似することがよく見られる。Wu and Palmer (1994) は、深さと 2 単語の最小共通上位概念 (Least Common Subsumer, LCS) を用いて、下記の式で類似度を算出する^[8]。

$$wup(w_1, w_2) = \frac{2 \times \text{depth}(LCS)}{\text{depth}(w_1) + \text{depth}(w_2)} \quad (3)$$

ここで、深さ (depth) は WordNet の階層構造において最上位の概念からの最短経路長によって定義される。LCS は、2つの概念に共通する上位語のうち最も近い上位語を指す。

2つのクエリのセマンティック類似度を測るため、シンプレックス法 (simplex method) を用い、最も良いキーワードマッチングを探す。クエリ q_1 と q_2 のセマンティック類似度を計算する本手法の手順は、以下の通りである。

(1) 2つのクエリ $q_1 = (x_1, x_2, \dots, x_m)$, $q_2 = (y_1, y_2, \dots, y_n)$ を単語ペア (x_i, y_j) ($0 < i < m, 0 < j < n$) ごとに LCS 値を計算する。その中に、最大の LCS 値を単語ペア (x_i, y_j) の LCS 値として扱う。

(2) クエリセグメンテーションにより、

$q_1 = (k_1, k_2, \dots, k_k)$, $q_2 = (k'_1, k'_2, \dots, k'_v)$ が得られる。そして(1)で計算された単語ペア (x_i, y_j) の LCS 値を使い、 $k \times v$ 個キーワードレベルの LCS 行列 $M_{s \times t}$ (s, t はキーワードの長さ) を生成する。

(3) 2つのクエリのキーワード内の意味的に類似する単語の位置やキーワードの長さが異なることに加え、単語は一回しかマッチングできないため、LCS 行列に最適な単語マッチングを探すため、行列 $M_{s \times t}$ にハンガリアン法を施し、平均 LCS 値が最大のマッチングを求め、その結果をキーワード間のセマンティック類似度とする。

(4) (3) より、キーワードレベルの LCS 行列 $M_{k \times v}$ を生成する。同様にハンガリアン法を施し、クエリのセマンティック類似度を計算する。

3.4 その他の特徴量

Jones ら^[11]は単語の文字列の編集距離特徴は類似クエリの検出に役立つことを指摘した。レーベンシュタイン距離とは、2つの文字列がどの程度異なっているかを示す距離の一種であり、クエリの表記揺れ、誤字脱字、入力変換ミスなどの語句修正の検出に有用である。ここで、レーベンシュタイン距離が 2 に下回る場合、後続クエリは先行クエリから修正して得るものと見なす。

また、Jansen ら^[7]は、絞込 (specification) と汎化 (generalization) のクエリ修正タイプを定義した。それは直前のクエリと同じ話題のクエリで、検索ユーザがより詳細な、もしくは幅広い情報を得るための修正である^[9]。そして、本研究で、クエリの絞込および汎化に判定する場合、クエリ修正とする。

ここで使っているクエリの絞込・汎化は、

Jansen ら¹⁷⁾により以下のように検出される。

- ・ 絞込：先行クエリに文字を追加したものの
- ・ 汎化：先行クエリから文字を削除したものの

4 クエリ満足度推定

本研究では、あるクエリ Q1 に対し、クエリログデータからユーザが Q1 に満足するかどうかを推定する。ここで、クエリ満足度を推定するため、使われる特徴量を紹介する。

4.1 クリックデータ

先行研究により、クリックデータはクエリ満足度を推定するために非常に有用なパラメーターと指摘されてきた。クリックデータはユーザがクエリに対して満足しているとみなせる。だが、実際の検索プロセスでは、誤ってクリックしたり、クリックした後 Web サイトのコンテンツが自分の情報ニーズと一致しないことを発見したりすることが時々起こる。その場合、Web ページ滞留時間の導入が必要となる。Fox ら (2005) は滞留時間が 30 秒を超えるのが満足度と高い相関を持つ可能性を示した¹⁸⁾。

本研究では、データセットの制限から、クリック時間情報が取得できないこともあり、クリック滞留時間の代わりにクエリ提出の間隔時間によって予測を行う。その場合、30 秒以上滞留したクエリが満足されているとみなす。

4.2 クエリ修正

クリックデータに加え、クエリ修正はクエリ満足度の推定に役立つ特徴量であることが多くの研究者から確認された¹²⁾。通常、検

索プロセスは、ユーザが検索エンジンにクエリを提出しその意図が満足でき、検索行動を終了したり、別の関連がない検索に移動したりすることである。一方、ユーザが意図を満たさない場合、検索に戻り、クエリを修正して再提出をすることが多い。また、意図を満たさないまま、検索を中止することもある。それに従って、多くの研究にはクエリ修正をクエリの不満の指標と見做している。

4.2.1 クエリ修正予測モデル

この節で、クエリ修正予測に実施した実験を紹介する。本研究で、第3章で紹介された4つの異なる特徴量を用い、モデルを立て、クエリ修正を予測し、比較実験を行う。現段階で、クエリ修正モデルはヒューリスティック法で、クエリ類似度の閾値を用いる。既存研究により、閾値を 0.35¹²⁾にする。

モデル 1. 語彙 (Lexical) : ここで、3.1 節で紹介した語彙の共通度を用いる。語彙の共通度が閾値に上回る場合、クエリ修正と判定される。

モデル 2. キーワード (Keyword) : このモデルで使う特徴量は、3.2 節で説明したキーワード類似度である。モデル 1 と同様に、キーワード類似度が閾値より高ければ、クエリ修正と判定される。

モデル 3. セマンティック (Semantic) : 3 つ目のモデルには 3.3 節で説明したセマンティック類似度を使用する。クエリ修正の判定基準はモデル 1、2 と同じである。

モデル 4. セマンティックとその他の特徴量 (Semantic with others) : このモデルでは、モデル 3 に基づいて、3.4 節で説明したその他の特徴量を用いて判定した上で、セマンティック類似度で予測を行う。予測手順は以下の通りである。

- (1) 2 つの連続的なクエリの Levenshtein

編集距離が2以下の場合、もしくはクエリの検索意図を絞込・汎化をし、後続のクエリを提案する場合、クエリ修正と判定される。

(2) それ以外のクエリは、セマンティック類似度が閾値より高い場合、クエリ修正と判定される。

4.3 クエリとドキュメントの関連性

検索タスクレベルの成功はインクリメンタルプロセス^[2]なので、タスクレベルの満足はクエリレベルの満足によって累積されて得られる。そして、本研究では、ユーザがクエリを修正して再検索を行う際に、先行クエリの検索結果から何らかの有用な情報を取得し、その情報に基づいてクエリを修正するとき、先行クエリに対してある程度満足されていると考えられる。

修正されたクエリ、特に修正された単語はユーザの検索意図をよく表している。クエリとドキュメントの関連性を計算するため、クエリの修正部分に着目し、それと先行クエリのあとに確認された Web ドキュメント間の関連性を検討すべきである。

4.3.1 本文抽出

URL から Web ページの本文部分を取得する。本研究でまず URL により Web ページの HTML ファイルを取り出す。そして Python で BeautifulSoup ライブラリを用い、余分のタグを外し、本文のテキスト部分を抽出する。

4.3.2 関連度の計算

前述の通り、クエリとドキュメントの関連度は主に修正された単語とドキュメントの関連度で表示することができる。TF-IDF とは文書 d における単語 t の重要度を計算する手法である。ここで、修正された単語に対して下記の式を用いた先行クエリに確認したドキュメントとの関連度を計算する。

$$tf(t_i, d_j) = \frac{n_{i,j}}{\sum_{k \in j} n_{k,j}},$$

$$idf(t_i) = \log\left(\frac{|D|}{|d: d \ni t_i|}\right) \quad (4)$$

$tf(t_i, d_j)$ はある単語 t_i のある文書 d_j における出現頻度であり、 $idf(t_i)$ は文書集合にある単語 t_i が含まれる文書の割合の逆数である。

$$tfidf(t_i, d_j) = tf(t_i, d_j) \cdot idf(t_i) \quad (5)$$

$tfidf(t_i, d_j)$ は文書 d_j におけるある単語 t_i の TF-IDF 値である。複数の語からなる場合は TFIDF の平均値を用いる。

4.4 クエリ満足度推定

本研究でクエリ満足度推定のため、前述した幾つかの特徴量の様々な組み合わせによる6つのモデルを用いて推定を行う。

(1) 間隔時間のみ (Interval time only): クエリ提出の間隔時間は唯一の特徴量としてクエリ満足度を推定する。ここで、4.1 節で述べたように、30 秒を閾値にして、これより低い間隔時間は、クエリが不満足であることを示すとみなされる。

(2) クリック数のみ (Click num only): ここで、クリックが満足クエリの指標として扱われる。一つのクエリに対し、クリックがあれば、すなわちクリック数が1以上であれば、ユーザはクエリに満足するとする。

(3) クエリ修正のみ (Reformulation only): 先行研究により、クエリ修正はユーザがクエリに不満足していることを示す。そして、第3章で説明した手法を用いて、クエリ修正と判定されれば不満足とする。

(4) クエリと間隔時間 (Reformulation, Interval time): ここで、クエリ修正はクエリ不満足を示す特徴量である。だが、クエリ修正ではない場合、間隔時間を用い、30 秒以上である間隔時間を持つクエリしか満足と判

定されない。

(5) クエリ修正、クエリとドキュメント関連度(Reformulation, TFIDF): 後続クエリの修正は先行クエリの検索結果と一定な関連がある場合、先行クエリにある程度満足していると見なすことができるということを 4.3 節で論述した。そして、クエリ修正ではない場合、クエリに満足していると判定されるが、クエリ修正であっても、クエリとドキュメント関連度がある (0 より大きい) 場合、クエリ満足と判定される。

(6) 全ての特徴量 (All features): ここで、間隔時間、クリック数、クエリ修正、クエリとドキュメント関連度を組み合わせて使い、クエリ満足度の推定を行う。クエリ修正である場合、(5) と同様に判定されているが、クエリ修正ではない際、間隔時間とクリック数を用いる。

しかし、注意すべきなことは、セッションの終了時間が不明であるため、セッション内に最後のクエリの間隔時間が計算できないことになる。この場合、基本的にクリック数を用いてクエリ満足度を推定することにする。それに対して、クエリ間隔時間がある場合、間隔時間を用い直接に判定を行う。

5 データと評価実験

5.1 データ

クエリ満足度の推定に向けて、ユーザの検索ログデータが必要である。本研究で公開されている Webis Search Mission Corpus 2012 (Webis-SMC-12)^[13]を用いる。Webis-SMC-12 は 2006 に公開された AOL データセットから抽出し得られ、127 名のユーザの 8840 件の検索エンジンインタラクションが含まれている。2 人の注釈者が人手でこれらのイン

タラクションを 1378 のミッションに分割した。データセットはユーザ ID、クエリ、クエリ提出時間、クリックランク、クリックされた URL、ミッション ID、コメントという 7 つのデータ項目からなる。

5.2 データ前処理

5.2.1 セッション分割

本研究で、検索セッションとは、同一ユーザ ID に対し、アイドル時間 (idle time) が 30 分以上になるたびに、セッションの中断を導入し、30 分以内であるインタラクションを 1 つのセッションとして定義する。

5.2.2 対象データ抽出

本研究では、クエリ修正がクエリ満足度推定の特徴量の 1 つとして使用されているため、1 つのクエリしかないセッションを削除し、クエリ数が 2 以上であるセッションを対象データとする。

また、データセットに同一のクエリを重複に提出することがよく見られる。提出時間が異なることで、全く同じクエリが複数のクエリとして存在することがある。これは、対象データの抽出や、クエリ類似度、特にクエリとドキュメントの関連度の計算に影響を与える。そのため、本研究で同一のクエリを連続的に提出される際に、複数のクエリを 1 つに併合する。

そして、上記の操作を行い、最終的に 5678 件の検索エンジンインタラクション、3521 クエリ、1049 検索セッションを取得した。それは対象データとして本研究で使用されている。

5.3 評価用データセット

提案手法でクエリ満足度の推定を評価するため、評価用データセットを作成する必要

がある。処理対象データからランダムに 10 名ユーザの 543 件クエリ、159 セッションを抽出し、人手で満足度付け (SAT: 満足, DSAT: 不満足) を行った。現段階で、満足度付けは筆頭著者 1 名の作業で行った。

5. 4 評価実験

提案手法の有効性を明らかにするため、クエリ修正予測およびクエリ満足度推定についてそれぞれ比較評価実験を行った。

5. 4. 1 クエリ修正予測

クエリ修正予測の評価用データセットは、対象データにより、Webis-SMC-12 の作成者が検索意図に基づいて付けたミッション ID^[13]を用いた。本研究で、同じセッション内に、先行クエリと同じ検索意図、すなわち同じミッション ID を持つクエリは先行クエリの修正であることにする。

上記の定義に従って、評価用データセットを作成し、評価実験を行った。表 1 に、正解率 (Acc)、True Positive (クエリ修正) の F1 値 (Ref. F1)、および True Negative (クエリ修正ではない) の F1 値 (No-Ref. F1) を示す。

結果として、セマンティック類似度とその他の特徴を組み合わせて用いる場合、最も良い正解率を得ることが分かる。

表 1 クエリ修正予測 (Lexical vs. Keyword vs. Semantic vs. Semantic with others)

	Acc	Ref. F1	No-Ref. F1
モデル 1	48.66%	53.67%	43.93%
モデル 2	40.25%	40.32%	40.17%
モデル 3	65.25%	75.80%	38.42%
モデル 4	72.29%	81.61%	43.81%

5. 4. 2 クエリ満足度推定

本研究で、クエリ満足度の推定は二値分類問題であると考えられる。クエリが満足 (SAT) と不満足 (DSAT) という 2 つのクラスに分類させるため、SAT と DSAT のクラス別の適合率 (precision)、再現率 (recall)、F1 値 (F1) という 3 つの指標を用い評価を行った。また、クエリ満足度推定の正解率も計算し、その結果は表 2 に示す。

表 2 により、クリック数のみ (Click number only) の場合ではクエリ修正、クエリとドキュメント関連度 (Reformulation, TFIDF) よりわずかに優れた正解率を得て、全ての特徴量を用いる場合、最高の SAT 適合率及び DSAT 再現率に達成した。

表 2 クエリ満足度推定のパフォーマンス

	Accuracy	F1 SAT	F1 DSAT	Precision SAT	Recall SAT	Precision DSAT	Recall DSAT
Interval Time Only	50.64%	86.20%	42.11%	78.42%	95.70%	72.13%	29.73%
Click number Only	75.87%	81.93%	63.71%	90.00%	75.19%	53.99%	77.70%
Reformulation Only	69.61%	76.05%	58.44%	89.12%	66.33%	46.59%	78.38%
Reformulation, interval time	69.43%	75.52%	59.31%	90.46%	64.81%	46.54%	81.76%
Reformulation, TFIDF	75.32%	81.59%	62.57%	89.19%	75.19%	53.33%	75.68%
All Features	71.27%	77.26%	61.00%	91.07%	67.09%	48.41%	82.43%

6 考察

クエリ満足度推定の評価実験により、クエリ数のみを使う場合に、最も良い結果を得た。従って、本実験ではクエリ修正、及びクエリとドキュメントの関連度の有効性は確かめられなかった。しかしながら、クエリ修正、クエリとドキュメントの関連度という2つの特徴量を用いる際、二番目に良いパフォーマンスを示した。また、全ての特徴量を使用し、クエリ満足 (SAT) を推定する際に最良の適合率を得て、またクエリ不満足に最高の再現率に達した。それが一定程度に提案した手法の有効性を示したと思われるが、またさらに論証する必要がある。

評価実験における失敗パターンに対し失敗分析を行った。失敗を引き起こす可能な原因には以下の4点を挙げられる。

(1) クエリ修正の間違い。クエリ満足度を推定する際に使われている主要な特徴量の1つとして、クエリ修正の精度はクエリ満足度推定の効果に密接に関わっている。LCS を使い、クエリのセマンティック類似度を計算するとき、wordnet の同義語のパスの長さを利用する。従って、LCS 値の計算は wordnet に対する依存性が高いため、wordnet に含まれていない URL や、ブランド名などのパターンがクエリとして提出する際に、同じの検索意図を持ってても類似度が低いもしくは0であることが多い。類似したクエリを識別するため、クエリ修正タイプや編集距離という概念を導入し、LCS 値が0のクエリパターンは200件近く減少したが、まだ多く見られる。

(2) データセットの古さ。実際に、検索ログデータはプライバシーの問題があるので、現時点で公開ログデータセットがほと

んどなく、適切なログデータセットを探すことが非常に困難である。本実験で使っているデータセットは2006年に公開されたが、その時点からWeb環境が大幅に変化してきた。これにより、特にクエリとドキュメント関連度を計算する際に、無効になってしまったURL、またはWebコンテンツが変わったということがよく見られ、関連度を計算するのが困難となる。

(3) 評価用データセットの不適合。クエリ満足度推定用の評価データセットは現時点で筆頭著者の1人の注釈者ラベル付きをした。しかし、データの信頼性を保証するため、少なくとも2人の注釈者が必要である。

(4) 推定モデルの不完全。現在の推定モデルには、各特徴量の組合せはヒューリスティックなルールに従って、閾値を用いて構築したので、そのルールの適切性や正確性に欠ける恐れがある。そして、各特徴量を入力値として機械学習モデルを用い、クエリ満足度の推定を行うことが今後取り組むべき重要な課題である。

7 おわりに

本研究はセッション内のクエリを時系列に並べ、クエリ修正を用いることにより、クエリ満足度推定の効果を向上させる手法を提案することを目指し、クエリとドキュメント関連度という新しい特徴量の有効性を検証する。評価実験により、クエリとドキュメント関連度とクエリ修正を用いる場合、75.32%の正解率を達成し、クエリ修正のみ、または間隔時間のみ使って推定する際と比較して、それぞれ約6ポイント、25ポイントの推定正解率を向上させることができるが、その手法の有効性は未だ十分な検証がで

きなかったことが分かる。

そして、今後の課題として、クエリ満足度の推定に関して特徴量やモデルの改善が必要である。クエリ修正については、多様なセマンティック類似度を測る手法を比較して検討する。

推定モデルに対し、現時点で閾値を用い、ヒューリスティック法のモデルしか構築しなかったため、各特徴量の使用順を考慮せず、入力値として機械学習モデルを構築し、クエリ満足度の推定を試すのが今後の目標の1つである。また、複数のモデルを実装し、比較実験を行い、より優れたモデルを探る。

より信頼度の高い結果を得るため、評価用データセットの改善も必要である。2人の注釈者が満足度ラベル付きを行い、意見が合わない場合、3人目の注釈者から判断をもらう。最高な得点を取る満足度は、クエリの最終的な満足度にする。このように、最大限に評価用データセットの客観性を保証することを求めたい。

参考文献

- [1] 梅本和俊, 山本岳洋, 中村聡史, 田中克己: 「検索意図の遷移検出に基づく動的なクエリ推薦に向けた行動ログデータの分析」, 研究報告 ヒューマンコンピュータインタラクション(HCI), Vol. 2012, No.24, pp. 1-8, 2012
- [2] Ahmed Hassan, Xiaolin Shi, Nick Craswell, Bill Ramsey: “Beyond clicks: query reformulation as a predictor of search satisfaction”, In Proc. CIKM’13, pp. 2019-2028, 2013.
- [3] Steve Fox, Kuldeep Karnawat, Mark Mydland, Susan Dumais, Thomas White: “Evaluating implicit measures to improve web search”. ACM Trans. Inf. Sys., Vol. 23, No.2, pp.147-168, 2005.
- [4] Scott B. Huffman, Michael Hochster: “How well does result relevance predict session satisfaction?”, In Proc. ACM SIGIR’07, pp. 567-574, 2007.
- [5] 関口裕一郎, 杉崎正之, 内山匡, 藤村滋, 望月崇由, 鈴木智也: 「検索クエリログを用いたクエリ変更意図の自動推定」, In DEIM Forum2011, B2-3, 2011.
- [6] Jeff Huang, Efthimis N. Efthimiadis: “Analyzing and evaluating query reformulation strategies in web search logs”, In Proc. CIKM, ’09 pp. 77-86, 2009.
- [7] Bernard Jansen, Amanda Spink, Chris Blakely, Sherry Koshman: “Defining a session on Web search engines”, Journal of the American Society for Information Science and Technology, Vol. 58, No. 6, pp. 862-871, 2007.
- [8] Zhibiao Wu, Martha Palmer: “Verb semantics and lexical selection”, In Proc. ACL’94, pp.133-138, 1994.
- [9] Bin Tan, Fuchun Peng: “Unsupervised query segmentation using generative language models and wikipedia”, In Proc. WWW’08, pp. 347-356, 2008.
- [10] Matthias Hagen, Martin Potthast, Benno Stein, Christof Braeutigam: “The power of naive query segmentation.” In Proc. ACM SIGIR, pp. 797-798, 2010.
- [11] Rosie Jones, Benjamin Rey, Omid Madani, Wiley Greiner: “Generating query substitutions”, In Proc. WWW’06, pp.387-396, 2006.
- [12] Jean-Baptiste Michel, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, William Brockman, The Google Books Team,

Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez Lieberman Aiden: “Quantitative Analysis of Culture Using Millions of Digitized Books”, *Science*, Vol. 331, No. 6014, 176-182, 2011.

[13] Matthias Hagen, Jakob Gomoll, Anna Beyer,

Benno Stein: “From Search Session Detection to Search Mission Detection”, In Proc. OAIR’13, pp. 85-92, 2013.

[14] Martin Trenkmann: “{PhraseFinder} -- Search millions of books for language use”, <https://phrasefinder.io> (参照 2021 年 4 月 12 日).

第29回年次大会予稿

N-gramを用いた日記文章の文章構造の分析

Analysis of the Sentence Structure of Diary Sentences

Using N-gram

松田雛乃^{1*}, 村井源¹

Hinano MATSUDA^{1*}, Hajime MURAI¹

1 公立はこだて未来大学

Future University Hakodate

〒041-8655 北海道函館市亀田中野町116-2

*連絡先著者 Corresponding Author

本研究では日記の重要性に着目し、日記文章の自動生成を目指す上で有用となる文章構造の分析を行った。執筆者の属性による影響を想定し、老年男性・老年女性・若年男性・若年女性の4カテゴリにおいてブログから日記文章を収集した。収集したブログ記事中の各文の機能を人手で分類して記号化し、n-gramを用いて文の時系列的構造を分析した。その結果、日記文章の基本的な文章構造は執筆者の属性に影響されないと判明した。

This study focused on the importance of a diary. In order to realize automatic generation of diary sentences, we analyzed the sentence structure. Assuming the influence of the author's attributes, we collected diary sentences from blogs in four categories: old men, old women, young men, and young women. The functions of each sentence in the collected blog articles were manually classified and symbolized. The time series structure of sentences was analyzed using n-gram. As a result, it was found that the basic sentence structure of the diary sentence is not affected by the attributes of the authors.

キーワード: 日記, 文章構造, n-gram

Diary, Sentence Structure, N-gram

1 背景

日記は、史料としての価値や文学作品としてのエンタテインメント性[1]や、行動や感情を把握する記録媒体[2][3]としての機能性など様々な要素を持ち合わせており、読み物・書き物の両面において価値ある存在である。

日記の活用のためには、ある程度の期間

で連続して日記を記述することが望ましい。しかし、書き手の気質やライフスタイルによっては、日記の習慣づけが難しい人も存在すると考えられる。よって、日記の自動生成の成功は、多忙な生活の中でも日記を書きたいと望む人々の支援となり、人々の生活に寄与するものであると考えられる。

日記は決まった形式がなく、口語調に近い文章である。そのため、年齢や性別などの書き手の属性の違いが文章構造に影響を及ぼす可能性が考えられる。よって、人が読んで自然に感じられる日記文章の自動生成を行うためには、書き手の属性による文章構造の差異の分析が必要であると考えられる。

よって、本研究では、日記文章の対象データとして人気の高いブログの記事を用い、属性ごとの文章構造の分析を行い、書き手の属性に対応した日記文章の自動生成を目指すにあたって、どのように文章を展開させるかという指針を立てるためのデータとする。

2 関連研究

文章構造の分析を試みた研究の一つとして、藤原浩史の『枕草子』「ありがたきもの」の国語学的解釈[4]があげられる。この研究では『枕草子』において理解の難しい章段を取り上げ、国語学的手法による文章構成の考察と章段の主題の解釈を行っている。

また、科学論文に対し文章構造の分析を行った幾田ら[5]の研究も存在する。この研究では、論文のアブストラクトに対し、Doc2Vecを用いてアブストラクトの良し悪しの判定を学習させている。そして、アブストラクトの良し悪しの要因について考察するためにムーブ分析を用い、文章構造の分析を行っている。

『枕草子』の文章構造の考察へのアプローチは国語学的であり、科学的ではない。また、科学論文のアブストラクトや本文には決まった形式が存在し、日記のように書き手の属性の違いが文章構造に及ぼす影響は少ないと考えられる。

3 対象データ

本研究では、書き手の年齢や性別などの属性で文体特徴が変化する可能性を考慮した上で日記文章を分析する。そのため、ブログは書き手の情報を元に、男・女の性別と、若年・老年という年齢層の二つの属性を組み合わせる4カテゴリに分類し、カテゴリに沿うランキング上20サイトまでを選定した。表1は、対象データの各カテゴリにおいて選定した書き手の人数と、データサイズとブログエントリ数の内訳と合計である。

表1 各カテゴリの書き手の人数とデータサイズとファイル数

カテゴリ	人数	データサイズ	ファイル数
老年男性	20	52.1MB	43120
老年女性	20	64.6MB	27506
若年男性	20	10.7MB	4534
若年女性	20	9.5MB	4642
合計	80	136.9MB	79802

表1のデータソースは、「にほんブログ村」[6]というブログ記事のランキングが掲載されたサイトのランキングとカテゴリ機能を用い、特定の基準に従って日記に近いブログを選定した上で収集を行った。

基準とは、ブログ記事が、書き手がその日にしたことや思ったことを書いているかどうかである。本研究は、人が自身の記憶や思考を綴る日記を対象としているため、いわば日記らしい記事を選定する必要がある、広告としての記事を多く含むブログは除外した。

収集した日記文章から、本研究において分析対象となるファイルのランダムサンプリングを行った。各カテゴリ20人、合計80人の書き手から5ファイルをランダムに選定し、総計400ファイルを用いて文章構造の分析を行った。

4 方法

本研究では、n-gramを用いて文章構造の分析を行った。n-gramは、神話の物語構造の分析[7]や、Twitterにおける顔文字の感情分析[8]などに役立てられている。本研究は3-gramを用いて対象データの分析を行った。以降にその詳細を述べる。

4.1 文の機能の記号化

はじめに、4カテゴリからランダムサンプリングされた対象データの日記文章を一文ごとに区切った。基準は以下の通りである。

- ・「。」「?」「!」など、句点の機能を持つ記号で区切る
- ・句点の機能を持つと判断できる場合に顔文字や絵文字で区切る
- ・括弧で区切る
- ・句点に成り得るものがない場合、文として不自然な途切れがない限り改行で区切る

なお、「?」「!」や括弧については、一文中の効果・セリフ・補足などに使われていると判断できる場合は区切らないこととした。

続いて、日記文章の通読を行った。区切った一文ごとに、「トピック提示」・「トピック掘下」・「意見感想」・「名乗り／挨拶」・「まとめ」・「和歌」・「その他」という文の機能を示すタグを自動で付与し、文章内における各文の記号化を行った。タグ名とその定義を表2に示す。

「トピック提示」は、まず、文章の初めの一文が意味のわかる内容である場合に付与した。文章が続くなかで新たな話題に切り替わった場合には新たな話題を示す一文に付与し、また、トピックについて掘り下げるなかで出現した要素に対しても掘り下げが行われた場合、掘り下げが行われた要素を示す一文に付与した。また、新たな要素もなく、先に出現した

「トピック提示」の文と全く同一の内容の文に対しても「トピック提示」の付与を行った。また、トピックとして意味は読み取れるものの、提示されるトピックの断片であると判断できる文にも「トピック提示」の付与を行った。

表2 文の機能のタグの定義表

タグ名	定義
トピック提示	トピックを提示する文
トピック掘下	トピックを説明・補足などで掘り下げる文
意見感想	トピックについて、書き手が考え・気持ちを述べる文
名乗り／挨拶	書き手の名乗りまたは挨拶の機能を持つ文
まとめ	複数トピックを総括するトピックの始点の文に付与
和歌	短歌・俳句・川柳など、特定の形式を持つ詩
その他	上のどれにも当てはまらない文

「トピック掘下」と「意見感想」は、両者とも「トピック提示」において示されたトピックについて述べる文へのタグであるが、「トピック掘下」は書き手の主観を含まない事実である文に付与し、「意見感想」は書き手の考え・感情に基づく文に付与した。「トピック掘下」と「意見感想」の判断が曖昧である文が出現した場合、本研究の対象データがネットから収集した文章であり、読み取れる書き手の意図は想像上のものであることを加味し、以下の判断基準を設けた。読み手側からの事実検証が不可能である文においては、荒唐無稽な内容ではなく、過剰な表現や悪意ある装飾がなされていない文には「トピック掘下」をタグとして付与し、それ以外の文を「意見感想」を付与した。

「まとめ」については、あるトピックがそれ以前に登場した複数のトピックの内容を総括し、書き手が考えや感想を述べていると判断できる場合に、「トピック提示」と同様にトピックの始点として付与した。

「名乗り／挨拶」については、書き手の名乗り・自己紹介のほか、「おはよう」「こんにちは」「こんばんは」などの出会いの挨拶やそれに類似した言葉や、「さよなら」「おやすみ」「いただきます」などの別れの挨拶を含む文に対して付与した。また、ブログ特有の要素である「ようこそ」「おかえり」などの意味合いの挨拶を含む文も「名乗り／挨拶」タグを付与した。

4.2 トピックIDの付与

次に、トピックの提示順と区切りを示すトピックIDの付与を行った。

基本的に「トピック提示」「まとめ」というタグを付与した文を基準にIDを付与した。「トピック提示」「まとめ」の出現順に、その「トピック提示」「まとめ」に関連した意見や感想、説明であると判断できる各文に1番からトピックを付与し、新たなトピックが提示された場合には2番,3番...と出現順にIDを付与した。

「トピック提示」タグを基準にしない例外は、ある一つのトピックが「名乗り／挨拶」「その他」「和歌」で始まっている場合と、一つのトピック内で先に出現した「トピック提示」の文と全く同一の内容と判断して「トピック提示」を付与した文が含まれる場合と、トピックの断片であるとして「トピック提示」の付与を行った文が含まれる場合である。「名乗り／挨拶」「その他」は、直後で「トピック提示」を付与された文と同じIDを付与し、「和歌」は文中の詩に関連性のあるトピックを含む「トピック提示」と同じIDを付与した。同一内容あるいは断片の文に対しての「トピック提示」については同一内容・断片の大元の「トピック提示」と同じIDを付与した。

なお、分析対象から除外する文に対してはトピックIDに0を付与した。分析から除外する基準は以降の通りである。

- ・ ブログについての応援や読者登録などを

読者に願う文

- ・ 記号のみの羅列
- ・ 意味が読み取れない文

なお、ブログを書くことそのものについての意見やアドバイスなどを主題としたトピックは除外しないこととした。

4.1の文の機能の記号化と4.2のトピックIDの付与を踏まえて作業を行ったデータの一部を表3に示す。

表3 若年男性の日記文章に記号化とID付与を行った結果(一部抜粋)

本文	機能	ID
勉強勉強日記	トピック提示	1
本当はね、駿台高2全国模試の結果を乗せようと思ったんですが、何か見れないので春休みの日常の話を(新古今和歌集の省略)	トピック提示	1
たぶんね、web返却+紙返却ではなくて、紙返却だけを選んでしまったんだと思います。	意見感想	1
痛恨のミス。	意見感想	1
さて、本題の今日やったこと。	トピック提示	2
今日は掃除した。	トピック掘下	2
以上。	その他	2
一応毎日、最低限これはやろうと欲していることはあって、[1]数学の問題集2ページ[2]英語の長文1つ[3]入門英文問題精講(竹岡広信)[4]嵐本国語1問[5]単語王2ユニット[6]英語のなんかを聞くですね。	トピック提示	3
明日からは頑張りますたぶん	意見感想	3
最後に一押しお願いします()にほんブログ村大学受験ランキング	その他	0

4.3 n-gram

4.1と4.2を経て完成したデータに対し、3-gramを用いて分析を行った。トピックIDごとに、文章パターンを図1のように分割した。

分割した文章パターンのうち、連続して出現する機能タグはひとまとまりとして捉えて変換を行った。例えば、図1のトピック1(「トピック提示」→「トピック提示」→「意見感想」→「意見感想」)は、「トピック提示」→「意見感想」へと変換される。

機能	トピックID
トピック提示	1
トピック掘下	1
意見感想	1
意見感想	1
トピック提示	2
トピック掘下	2
その他	2
トピック提示	3
意見感想	3
その他	0

トピック1	トピック2	トピック3
トピック提示	トピック提示	トピック提示
トピック提示	トピック掘下	意見感想
意見感想	その他	
意見感想		

図1 表3に対する文章パターンの分割

変換を行った各文章パターンを用いて、各ファイルの各トピックごとの3-gramを作成した。老年男性・老年女性・若年男性・若年女性の4カテゴリにおいて、作成した3-gramの出現頻度をカウントした。

5 結果

各カテゴリにおける3-gramの出現頻度上位5を以降の表4-表7に示す。なお、「トピック提示」は「提示」、「トピック掘下」は「掘下」、「意見感想」は「意／感」、「名乗り／挨拶」は「名／挨拶」、「その他」は「他」として表示している。

表4 老年男性の3-gramの出現頻度

3-gram			出現頻度
提示	掘下	意／感	127
掘下	意／感	掘下	55
提示	意／感	掘下	50
意／感	掘下	意／感	39
提示	掘下	他	23

表5 老年女性の3-gramの出現頻度

3-gram			出現頻度
提示	掘下	意／感	176
掘下	意／感	掘下	82
意／感	掘下	意／感	80
提示	意／感	掘下	49
掘下	意／感	名／挨拶	17

表6 若年男性の3-gramの出現頻度

3-gram			出現頻度
提示	掘下	意／感	138
意／感	掘下	意／感	98
掘下	意／感	掘下	82
提示	意／感	掘下	74
掘下	意／感	名／挨拶	19

表7 老年男性の3-gramの出現頻度

3-gram			出現頻度
提示	掘下	意／感	171
意／感	掘下	意／感	117
提示	意／感	掘下	101
掘下	意／感	掘下	94
掘下	意／感	名／挨拶	20

表4-表7の結果から、各カテゴリで多少の前後はあれど、以下の文章パターンが出現頻度上位4位であると読み取れる。

- ・「トピック提示」→「トピック掘下」→「意見感想」
- ・「トピック提示」→「意見感想」→「トピック掘下」
- ・「意見感想」→「トピック掘下」→「意見感想」
- ・「トピック掘下」→「意見感想」→「トピック掘下」

6 考察

結果から、一般的な日記文章の基本構造は、「トピックを提示したのちトピックの掘り下げを行い、感想を述べる」構造と「トピックを提示したのちそれへの意見感想を述べ、補足や説明を行う」構造の2種と、「意見感想を述べてからトピックを掘り下げ、再び意見感想を述べる」構造と「トピックを掘り下げた後から感想を述べ、再びトピックを掘り下げる」構造の2種であり、合計4種で成り立つと考えられる。前半2種はそれ単体で一つのトピックを成り立たせられるため、図2のような基本構造が考えられる。

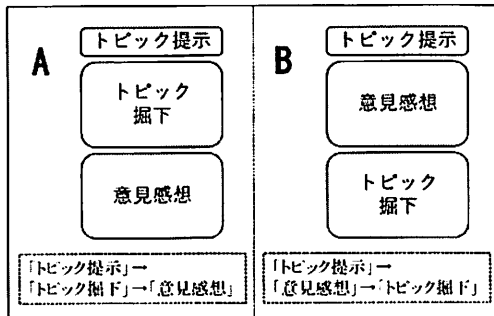


図2 日記文章の基本構造

便宜上、「トピック掘下」が上にある構造をA、「意見感想」が上にある構造をBとすると、後半2種はA・Bのそれぞれに連なる形で成立する構造であると考えられる。後半2種は互い違いに連続する可能性もあり、トピックにおける連続構造が長い場合も考えられる。

7 結論

本研究では、日記文章の書き手の属性の違いに着目し、年齢性別の異なる4カテゴリにおいて文章構造の分析を行った。日記文章において一文が果たす機能について記号化を行い、n-gramを用いて分析した。

結果、書き手の属性の違いは基本的な文章構造に影響を及ぼさないと判明した。また、日記文章の基本構造は、「トピックを提示したのちトピックの掘り下げを行い、感想を述べる」構造と「トピックを提示したのちそれへの意見感想を述べ、補足や説明を行う」構造の2種と、「意見感想を述べてからトピックを掘り下げ、再び意見感想を述べる」構造と「トピックを掘り下げから感想を述べ、再びトピックを掘り下げる」構造の2種であり、4種の基本構造で成り立つことが判明した。

今後の展望として、本研究で行った記号化について、客観性の担保のために定義表を用いた複数分析者での一致度検証を行う必

要がある。

また、本研究の結果と、日記文章のトピック抽出と文体特徴の分析[9]の結果を併せて日記文章の自動生成に役立てたいと考える。

参考文献

- [1] Webster, J: Daddy-Long-Legs, 1915. (岩本正恵訳:「あしながおじさん」, 岩波書店, 1970.)
- [2] 北村智, 河井大介:「日記式調査法による情報行動の時間的連続性・断続性の検討—携帯インターネット利用とPCインターネット利用の比較分析—」, 情報通信政策レビュー, Vol. 5, pp. 114-143, 2014.
- [3] 橋内久美, 金丸隆太:「日記による「もつれた怒り」の低減」, 茨城大学教育学部紀要, 教育科学, Vol. 68, pp. 519-532, 2019.
- [4] 藤原浩史:「『枕草子』「ありがたきもの」の国語学的解釈」, 中央大学国文, Vol. 59, pp. 1-12, 2016.
- [5] 尾上洋介, 馬場一貴, 小山田耕二:「人工知能を用いた科学論文の文章構造可視化」, 可視化情報学会誌, Vol. 38, No. 151, pp. 23-27, 2018.
- [6] ムラウチドットコム:「にほんブログ村」, <https://blogmura.com/> (参照 2021-4-2).
- [7] 吉田拓海, 村井源:「英雄神話物語の構造化に向けての物語特徴の抽出」, 情報知識学会誌, Vol. 30, No. 2, pp. 270-275, 2020.
- [8] 風間一洋, 水木栄, 榎剛史:「Twitterにおける顔文字を用いた感情分析の検討」, 人工知能学会第30回全国大会, 3H3OS17a4, 2016.
- [9] 松田雛乃, 村井源:「日記文章の文体的特徴の抽出」, 情報知識学会誌, Vol. 30, No. 2, pp. 289-296, 2020.

第 29 回年次大会予稿

エッセイに使用されているレトリックの分類と計量分析

Classification and quantitative analysis of the rhetoric used in the essay

根本さくら^{1*}, 村井源¹
Sakura NEMOTO^{1*}, Hajime MURAI¹

1 公立はこだて未来大学

Future University HAKODATE

〒041-8655 北海道函館市亀田中野町116-2

E-mail: b1017015@fun.ac.jp

*連絡先著者 Corresponding Author

これまで日本語エッセイでは、用いられているメタファー以外のレトリックや、文体的な特徴などの計量的分析はほとんど試みられていない。そこで本研究では、エッセイに平均的に用いられるレトリック等を明らかにするために筆者の異なる2作品を対象とし、使用されているレトリックのデータ化を行った。また得られたデータに基づき、 χ^2 二乗検定、因子分析などを用いてエッセイで使用されているレトリックの特徴を計量的に抽出した。

There have been few attempts to quantitatively analyze the rhetoric and stylistic characteristics of Japanese essays other than the metaphors used. In this study, we collected data on the rhetoric used in two essays written by different authors in order to clarify the rhetoric used in essays on average. Based on the data obtained, the characteristics of the rhetoric used in the essays were quantitatively extracted using the χ^2 -square test and factor analysis.

キーワード: エッセイ, レトリック, 表現技法

Essay, Rhetoric, Expression techniques

1 はじめに

エッセイは、自由な形式で自分の意見や経験を述べた散文である。社会的に高い評価を受ける優れたエッセイは、筆者の思い

や考え、筆者が過ごすありふれた日常を、表現や語り口の技巧によって読者に面白おかしく伝えることで作品に昇華させ、エンターテインメント性をもたせるという性質

をもっている。また、エッセイにおけるレトリックは、エッセイにおいて重要な「日常を面白おかしく伝えること」の一端を担っており、レトリックを効果的に用いることで、よりドラマチックなエッセイを書くことが可能になる。これまで、エッセイに関するさまざまな研究が行われてきた。しかし、日本語エッセイに使用されているレトリックについての分析は試みられていない。そこで本研究では、日常について描かれている日本語エッセイに用いられているレトリックの分類や、それらの計量的分析を行う。そして最終的には、人間が読んで面白いと思えるようなエッセイの自動生成システムの実現に向け、基礎データを構築することを目指す。

2 関連研究

エッセイに関する研究はこれまで多数行われている。例えば向田邦子のエッセイについては、ユーモア表現についての分析[1]や、感情表現についての分析[2]が行われている。また、エッセイにおけるモノの描写に着目し、特にモノ描写が顕著にみられる女性作家、幸田文と向田邦子が描くモノに対する関わりを分析した研究[3]や、幸田文の食に関する描写について分析した研究[4]も存在する。また、土屋賢二のエッセイにおけるずれた表現の面白さについて分析した研究[5]も行われている。しかし、これらはエッセイにおけるレトリックを分析したものではない。また、分析する作家を限定しているため、エッセイにおける平均的な表現や描写等を抽出できるものではない。

一方で文章表現に関する科学的な分析も行われてきており、特にメタファーに関しては、現状さまざまな計算論的アプローチ

が行われている[6]。しかし、メタファー以外にも含めたレトリックの出現傾向や、レトリックの文章内での使用傾向を計量的に分析している研究はほとんど行われていない。

2 分析対象

本研究では、『ものかんづめ』、『大河の一滴』の2作品を分析対象とした。『ものかんづめ』は主に筆者自身に起こった日常的な出来事を面白おかしく描写している作品である。また、『大河の一滴』は、筆者自身の世間に関する考えを述べている作品である。

これらを分析対象とした理由は、7週間連続でBOOKOFFの「日本のエッセイ・ノンフィクション(文庫)」週間ランキングの上位にランクインしており、高い評価を得ているエッセイ作品であると推測されるためである。また、この2作品は作者の性別、エッセイで取り扱う内容などいくつかの点で顕著な違いがあり、これらを分析することで、エッセイに平均的に使用されているレトリックの明確化や、話題の違いによるレトリックの違いの明確化が可能になると考えられる。

3 対象レトリック

本研究では、分析対象のエッセイに使用されているレトリックを、33個のレトリックに分類した。分類する33種類のレトリックは、「日本語のレトリック」[7]内で扱われているレトリック30種類に、「レトリック感覚」[8]、「レトリック事典」[9]、「吹き出しのレトリック」[10]の内容も加味して3種類のレトリックを追加したものである。また、33種類に当てはまらないものは「その他」に

分類した。表1に分類に使用したレトリックを
示す。また、表2に独自に追加したレトリックを示す。

表1 レトリック一覧

レトリック	説明	レトリック	説明
隠喩	類似性にもとづく比喩。	挿入法	カッコやダッシュなどの使用によって、文章の主流とは異なる言葉を挿入する表現法。
直喩	「～のよう」などによって類似性を直接示す比喩。	省略法	文脈から復元できる要素を省略する表現法。
擬人法	人間以外のものを人間に見立てて表現する比喩。	黙説法	途中で急に話を途絶することによって、内心のためらい等を表す表現法。
共感覚法	触覚、味覚、嗅覚、視覚、聴覚の五感の間で表現を借用する表現法。	倒置法	通常の語順を逆転させる表現法。
くびき法	ひとつの表現を二つの意味で使う表現法。	対句法	同じ構文形式のなかに対象的な意味の言葉を入れる表現法。
換喩	ものどもの隣接関係にもとづいて指示を横すべりさせる表現法。	声喩	音を文字列で表現する。
提喩	類と種の間関係にもとづいて意味範囲を伸縮させる表現法。	漸層法	しだいに盛り上げてピークを形成する表現法。
誇張法	事実以上に大げさな言いまわしをする表現法。	逆説法	常識を否定し、常識と相反することを書く表現法。
緩叙法	表現の程度をひかえることによって、かえって強い意味を示す表現法。	諷喩	一貫したメタファーの連続からなる文章。
曲言法	伝えたい意味の反対の表現を否定することによって、伝えたい意味をかえって強く表現する表現法。	反語法	相手のことばを引用してそれとなく批判を加える表現法。
同語反復法	まったく同じ表現を結びつけることによって意味をなす表現法。	引喩	有名な一節を暗に引用しながら表現する。
撞着法	正反対の意味を組み合わせて、かつ矛盾に陥らずに意味をなす表現法。	パロディ	有名な文章や定型文を茶化し場柄引用する。
婉曲法	直接言にくいことばを婉曲的に表現する表現法。	文体模写法	特定の作家、作者の文体を真似る。
逆言法	言わないと書いて実際には言う表現法。	含意法	伝えたい意味を直接言うのではなく、ある表現から推論される意味によって間接的に伝える表現法。
修辭的疑問法	形は疑問文で意味は平叙文という表現法。	挿入法	カッコやダッシュを用いて、文章の本流とは関係のないことを挿入する。

表2 独自に追加したレトリック3種

レトリック	説明
疑惑	「～かもしれない」、「～だろう」というように、断定しない言い方を する表現法。
中断	「お前は……可愛い。」のように、 途中で言いよどみつつも最後まで言 い切るもの。
頓絶法	文頭または文末に「……」や「—」 が使用されている文章。

4 結果

4.1 χ 二乗検定

レトリックの出現回数の下限を10回として χ 二乗検定を行った。 χ 二乗検定の残差分析結果を表3に示す。 p値が0.05以下ならば*、0.01以下ならば**で示している。また、+は、調整済み標準偏差の値が正ならば+で、負のならば-で示している。

表3 χ 二乗検定の結果

レトリック	ものかんづめ			大河の一滴		
	出現頻度	p値	+ -	出現頻度	p値	+ -
反復法	149	**	-	307	**	+
隠喩	125	**	-	264	**	+
声喩	132	**	+	95	**	-
直喩	79	**	+	79	**	-
修辭的疑問法	24	**	-	118	**	+
疑惑	8	**	-	69	**	+
誇張法	62	**	+	1	**	-
擬人法	39	**	+	16	**	-
曲言法	4	**	-	51	**	+
頓絶法	44	**	+	4	**	-
中断	8	*	-	28	*	+

表3より『ものかんづめ』では、『大河の一滴』に比べ、「声喩」、「直喩」、

「誇張法」、「擬人法」、「頓絶法」が有意に多いということがわかる。値が有意に多くなったレトリックについて考察していく。

まず「直喩」は、筆者の身の回りに起こった現象を「～のようだ」や「～のごとく」と明示したうえで、別のものに見立てて表現するために用いていると考えられる。「直喩法」の例として、筆者が台所で笑い転げている姉のことを“コオロギのように笑い始めた”と表現している部分や、筆者が変質者に驚き逃げたときのことを“オランウータンのごとく走り出した”と表現している部分が挙げられる。次に「擬人法」は、エッセイのテーマであるものや、筆者の身の回りにあるものを、人間に見立てて表現するために用いていると考えられる。「擬人法」の例として、筆者の家で飼っていたスズムシに対し“さて、このスズムシ界では無口でニヒルな男などまるっきり相手にされないのだ”と表現している部分が挙げられる。そして「声喩」は、擬音語・擬態語を駆使しながら状況描写をする際に用いていると考えられる。“私は待合室でジュースをゴクゴク飲みながら”や“男は何かボソッとつぶやいた”の部分が「声喩」に該当する。加えて「誇張法」は、出来事を事実よりあえて少し大げさに表現するために用いていると考えられる。例えば、筆者は足に水虫を患ったことで生じた痒みを、今まで生きてきた中で感じたすべての痒みを覚えているわけではないにも関わらず“それまでの人生で経験した事のない痒み”と表現している。また、筆者が働いていたアルバイト先でミツパチフェアが実施され、そのフェアが失敗した際も、筆者は死んでいないにも関わらず“私

達は経営者共々ハチノムサシの如くフェアに挑んで、死んだのである」と表現している。最後に「頓絶法」は、筆者の困惑や呆れ、悲しみを、文頭や文末における沈黙で表現するために用いていると考えられる。例えば、恋愛に関する詩を綴っていたノートが、筆者が上京して数年後に両親に見つかってしまったことに対し“……今頃になってあれが……”と文頭と文末に沈黙を用いながら呟いている。また、会社のパソコンを一台壊したことで障害が伝播し、会社のパソコンをすべて壊してしまった際の“こんなにもパソコン同士は仲が良かったのか……”という文の文末にも沈黙が用いられている(引用テキストは[11]より)。これらの5つの表現を文章に取り入れることで、文章にさまざまな効果をもたせることが可能になると考えられる。まず、「直喩」や「擬人法」を用いることで、日常的なものに面白みをもたせ、文章に表現的な目新しさを付加することが可能になると考えられる。加えて「声喩」は、状況描写に用いることで、出来事をより躍動的に、より詳細に表現することが可能になると考えられる。特に[飲む]や[つぶやく]、[よろめく]等の一般的によく使用される動詞に用いることで、読んでいて飽きない文章づくりの一端を担っていると考えられる。さらに「誇張法」は、感情表現や出来事を大げさに表現し、滑稽さを生み出す効果や、読者の感情的な反応を引き出す効果を発揮していると考えられる。結果として、『ものかんづめ』は、これらの効果によって読者の目をひきつけ、滑稽さを強調し、笑いを誘発していると考えられる。

表4より『大河の一滴』では、『ものかんづめ』に比べ、「反復法」、「隠

喩」、「修辭的疑問法」、「疑惑」、「曲言法」が有意に多いということがわかる。値が有意に多くなったレトリックについて考察していく。まず「反復法」は、文章をリズムカルに表現する際や、並列な関係のものを並べて記述する際に用いていると考えられる。「反復法」の例として、“ときに跳びはね、ときに歌い、ときに黙々と海へ動いていくのである”という部分や、“政治や、経済や、医療や”という部分が挙げられる。次に「隠喩」は、一言では表しにくく概念的なものを、想像しやすいものに例えて強調するために用いていると考えられる。「隠喩」の例として、“愚かしくも滑稽な劇の演じられるこの世間という円形の舞台”という文や、“「生老病死」の重い枷をはめられた人間」という文が挙げられる。そして「修辭的疑問法」は、読者・聞き手に、自分の考えを訴えかける際に用いていると考えられる。例えば、“「泣きながら生まれてきた」人間が、「笑いながら死んでいく」ことは、はたしてできないものなのだろうか”や“しかし、一方で現代の人間の存在そのものを悪と見て、そこから出発する生きかたもあるのではないか”の部分「修辭的疑問」に該当する。加えて「疑惑」は、自分の考えをあえて断言せずに読者に伝えるために用いていると考えられる。例えば、“ネガティブで悲観的な人生観と笑う人もいるかもしれない”という文の[かもしれない]という言い回しや、“それが人生の知恵というものだろう”という文の[だろう]という言い回しに現れている。最後に「曲言法」は、実際に伝えたいこととは反対のことを否定することで、伝えたいことを強調するために用いていると考えられる。例えば、

“なんとなくさびしい気がしないでもない”という文の“さびしい気がしないでもない”という言い回しや、“たかだか三、四百年の時が経過したくらいで、人生のありようが変わるはずがないではないが”という文の[人生のありようが変わるはずがない]という言い回しに現れている(引用テキストは[12]より)。『大河の一滴』も『もものかんづめ』と同様、内容に見合ったレトリックが使用され、それぞれのレトリックがさまざまな効果をもたらしていると考えられる。まず、「隠喩」は、他作品でもよく用いられるレトリックである。しかし本作品における「隠喩」は、特にアレゴリー的な理解の促進という要素が強く、教訓や風刺の意味合いも込めて使われていると考えられる。加えて「曲言」や「疑惑」、「修辭的疑問」を用いることで、断言しない言い回しや伝えたいことの反対のことを否定する表現、平文をあえて疑問文の言い回しをすることが可能になる。これにより、筆者の考えを強く強調する効果や、議論の構図を明確化する効果が生まれ、相手を説得するような効果が期待できるのではないかと考えられる。特に「疑惑」は、断定した言い回しをしないことで、筆者の考えに対する読者からの批判を防ぐ効果も持っていると考えられる。結果として『大河の一滴』は、これらの効果によって教訓等をまじえつつ読者を説得し、筆者の考えを読者に納得させることができるような作品となっていると考えられる。

4. 2 因子分析

各シーンにどのレトリックが何度出現しているかというデータに対して、因子分析を行った。

因子分析に使用したシーンベクトルの作成手順を説明する。

まず、エッセイをシーンごとに分割した。分割基準は「話題が変化している」、「時系列が変化している」の二つである。「話題が変化している」については、作者が作品内で設定した各エッセイタイトルが記述されている部分、また、エッセイ内の話題変更を示す空行が挿入されている部分を「話題が変化している」とみなした。また、「時系列が変化している」については、一日で起こった出来事について記述されているエッセイの場合は「午後五時」や「夕飯のとき」のように時間帯が明示的に示されている部分、数日に渡る出来事について記述されているエッセイの場合は、「次の日」や「一週間後」のように日の変化が明示されている部分を「時系列が変化している」とみなしシーンの区切りとした。

次に、行をシーン数、列をレトリック名、要素を各シーンにおけるレトリックの出現頻度としたシーンベクトルを作成した。ベクトルを作成する段階で合計頻度2以下のレトリック、合計頻度10以下のシーンを削除した。

最後に、R言語を用いてプロマックス回転による因子分析を行った。因子分析の結果の表は、因子負荷量が0.3以上である値は太字と太線の枠で示した。平行法を使った結果、因子数は2となった。因子分析の結果を表4に示す。

因子1について考察する。因子1において、因子負荷量0.3以上に相当する要素は、「反復法」、「修辭的疑問法」、「中断法」、「隠喩」、「声喩」、「直喩」、「緩叙法」、「曲言法」であった。これらのレトリックは、両方の作品において作品内での使用頻度が比較的高いものが多く含

まれている。よって因子1を「平均レトリック因子」と名付けた。

次に因子2について考察する。因子2において、因子負荷量0.3以上に相当する要素は、正の値が「修辭的疑問法」, 「疑惑」, 「曲言法」, 負の値が「声喩」, 「誇張法」, 「頓絶法」であった。正の値で出現しているレトリックは『もものかんづめ』に比べ、『大河の一滴』に多く現れているレトリックであった。反対に負の値で出現しているレトリックは、『大河の一滴』に比べ、『もものかんづめ』に多く現れているレトリックであった。よって因子2を「特徴レトリック因子」と名付けた。

表4 因子分析の結果

修辭技法	因子1	因子2
反復法	0.88	0.01
修辭的疑問法	0.64	0.36
中断法	0.58	-0.18
隱喩	0.58	0.15
声喩	0.51	-0.34
直喩	0.39	0.05
緩叙法	0.31	0.14
挿入法	0.13	-0.08
疑惑	0.23	0.63
曲言法	0.30	0.53
誇張法	-0.03	-0.40
頓絶法	0.03	-0.36
擬人法	0.07	-0.29
その他	-0.03	0.16

「修辭的疑問法」, 「疑惑」, 「曲言法」がセットで登場することで、強い言い方を避けつつ筆者の考えを読者に強く伝える効果が期待されるのではないかと考え

る。また、「声喩」, 「誇張法」, 「頓絶法」がセットで登場することで、ある出来事に対する筆者の呆れや困惑といった感情を、より大げさに、より躍動的に描く効果が期待されるのではないかと考える。

本研究で行った因子分析における因子数は2であった。これは、他にも作品を分析することで因子数が増える可能性がある。また、レトリック平均因子において、因子負荷量0.3以上のレトリックが要素の半分以上になっている。しかし、他にも作品を分析することで、レトリック平均因子における因子負荷量0.3以上の要素が少なくなり、エッセイで平均的に使用されるレトリックがさらに明確になる可能性がある。

5 まとめと今後の展望

日本語エッセイに使用されているレトリックの分類と、それらの計量的分析をすることを目標に、各作品のレトリック抽出を行った後、出現したレトリックの個数に対する χ^2 乗検定と、エッセイの各シーンにおいて登場しているレトリックのデータをもとに因子分析を行った。結果、本研究の研究手法を用いることで、エッセイに平均的に使用されているレトリックや、エッセイのテーマに合わせて使用されているレトリックを抽出できることがわかった。それぞれのエッセイに関しては、日常的な出来事がテーマとなるエッセイの場合は、文章に滑稽さをもたせる目的で、目新しい例え表現が可能なレトリックを使用する場合や、出来事を大げさに描くことが可能なレトリックを使用する場合があるということがわかった。また、筆者の思想や考えを述べるエッセイの場合は、筆者の考えを際立たせ読者を説得するという目的で、議論の

構図や目的を明確化することができる、いわゆる弁論や演説でも使用されるようなレトリックを用いていることがわかった。

本研究では2作品のみについての分析を行ったため、分析結果は必ずしも一般的なレトリックの傾向を示すとは言えない。したがって、他のエッセイの分析を行う必要がある。また、本研究ではレトリックのみを分析し、文中に含まれている慣用句的な表現の分析は行わなかった。今後の分析においては、慣用句等の表現も視野に入れ分析を行っていききたい。また、複数の分析者にレトリック表を用いた分析を行ってもらい、一致度の検定を行うことで客観性を担保する予定である。

参考文献

- [1] 葛西文夫：「向田邦子の個性的ユーモア表現」，笑い学研究，Vol. 12，pp. 82-87，2005.
- [2] 水藤新子：「向田邦子の感情表現：『思い出トランプ』を対象に」，中央学院大学人間・自然論叢，Vol. 44，pp. 59-76，2017.
- [3] ファルトウシナヤ・エカテリーナ：「日本の女性作家のエッセーにおけるヒトとモノとの関係:幸田文と向田邦子を中心に」，北海道大学大学院文学研究科研究論集，Vol. 10，pp. 49-67，2010.
- [4] ファルトウシナヤ・エカテリーナ：「幸田文のエッセーにおける「食」」，北海道大学大学院文学研究科研究論集，Vol. 11，pp. 133-147，2011.
- [5] 木村寛子：「共感が生まれるしくみ-土屋賢二のエッセイにおける「ずれ」を中心に-」，山梨英和大学紀要，Vol. 10，pp. A83-A94，2011.
- [6] 内海彰：「比喩理解への計算的アプローチ—言語認知研究における計算モデルの役割—」，認知科学，Vol. 20，No. 2，pp. 249-266，2013.
- [7] 瀬戸賢一：『日本語のレトリック』，岩波書店，2002.
- [8] 佐藤信夫：『レトリック感覚』，講談社，1992.
- [9] 佐藤信夫，佐々木健一，松尾大：『レトリック事典』，大修館書店，2006.
- [10] 『ふき出しのレトリック』，<http://balloon-rhetoric.atwebpages.com/> (最終閲覧日 2021-1-18).
- [11] さくらももこ：『もものかんづめ』，集英社，1991.
- [12] 五木寛之：『大河の一滴』，幻冬舎，1999.

事務局からのお知らせ

〔1〕 個人会員の皆様へ、2021年度の会費納入のお願い

1年分の年会費は正会員8千円、学生会員・ユース会員・シニア会員は4千円です。過去数年分未納のかたは合計額を納入くださるようお願いいたします。

(振込手数料はご本人負担ですが、赤色の払込取扱票(振込手数料受取人負担)をご使用になれば無料です。)

請求書が必要な方はその旨、情報知識学会事務局にメールでお知らせください。

振込先

- 1) 郵便振替口座 00150-8-706543 情報知識学会
- 2) ゆうちょ銀行 〇一九店(セイロイキョウ店) 当座 0706543 情報知識学会

(1) 会費の納入年月の確認方法

お手元に届いた学会誌の封筒の宛名ラベルには、ご自分の年会費の納入日が年度毎に西暦下2桁、月(2桁)、日(2桁)の6桁の数字で印字されています。会費未納年度には〔未納〕と表示されております。

お振り込みの後、事務局に通知が届き、宛名ラベルに印字、発送するまで10日ほどかかりますので、ご了承ください。

〔2〕 学会誌送付先、会員種別、メールアドレスの変更について

会員種別、学会誌送付先、メールアドレスが変わられ、変更手続きがお済でないかたはご連絡ください。年会費を納入していただいているのに学会誌やメールマガジンが届かないのでは申し訳ありませんので、変更の情報を事務局:office@jsik.jpまで、メールでお知らせください。新・旧の情報を並べてお書きいただくと確認できるので助かります。

〔3〕 新規入会申込方法

入会ご希望の方は情報知識学会ホームページ <http://www.jsik.jp>から「本会について」→「入会案内」→「入会申込フォーム」に必要な事項を入力・送信してください。

あるいは申込用紙をpdf形式、doc形式でダウンロードし、ご記入のうえ下記の事務局へ電子メール・FAX送信または郵送などをお願いいたします。

情報知識学会事務局

〒164-0003 東京都中野区東中野4-27-37 (株)アドスリー内

FAX:050-3730-8956 E-Mail:office@jsik.jp URL:<http://www.jsik.jp/>

情報知識学会誌 編集委員会

編集委員長	芦野 俊宏	東洋大学		
副編集委員長	常川 真央	中央大学		
編集委員				
	相田 満	国文学研究資料館	天野 晃	国立情報学研究所
	石塚 英弘	筑波大学名誉教授	宇陀 則彦	筑波大学
	江草 由佳	国立教育政策研究所	大槻 明	日本大学
	岡 伸人	近畿大学	岡部 晋典	愛知淑徳大学
	岡本 由起子	欧州情報協会	小川 恵司	凸版印刷(株)
	梶川 裕矢	東京工業大学	五島 敏芳	京都大学
	阪口 哲男	筑波大学	佐藤 翔	同志社大学
	孫 媛	国立情報学研究所	高久 雅生	筑波大学
	高田 良宏	金沢大学	田良島 哲	国立近現代建築資料館
	時実 象一	東京大学	中川 修	大日本印刷(株)
	長田 孝治	ロゴヴィスタ(株)	長塚 隆	鶴見大学名誉教授
	中山 堯	神奈川大学	西澤 正己	国立情報学研究所
	西脇 二一	奈良大学	根岸 正光	国立情報学研究所名誉教授
	原 正一郎	京都大学	原田 隆史	同志社大学
	藤田 桂英	東京農工大学	細野 公男	慶應義塾大学名誉教授
	村井 源	公立はこだて未来大学	村川 猛彦	和歌山大学
	村田 健史	情報通信研究機構	森 純一郎	東京大学
	山下 雄一郎	産業技術総合研究所	山地 一禎	国立情報学研究所
	山本 昭	愛知大学		(五十音順)

情報知識学会第29回(2021年度)年次大会実行委員会

実行委員長	村川 猛彦	和歌山大学		
委員	河瀬 彰宏	同志社大学	本田 正美	関東学院大学
	村井 源	公立はこだて未来大学		(五十音順)

■複写をされる方に

本誌に掲載された著作物を複写したい方は、(社)日本複写権センターと包括複写許諾契約を締結されている企業の従業員以外、著作権者から複写権等の行使の委託を受けている次の団体から許諾を受けて下さい。

著作物の転載、翻訳のような複写以外の許諾は、直接本会へご連絡ください。

〒107-0052 東京都港区赤坂 9-6-41 乃木坂ビル 学術著作権協会

TEL: 03-3475-5618 FAX: 03-3475-5619 E-mail: naka-atsu@muji.biglobe.ne.jp

アメリカ合衆国における複写については、次に連絡してください。

Copyright Clearance Center, Inc. 222 Rosewood Drive, Danvers, MA. 01923, USA

TEL: 978-750-8400 FAX: 978-750-4744 URL: http://www.copyright.com/

情報知識学会誌 Vol. 31, No.2 2021年5月22日発行 編集・発行 情報知識学会

頒布価格 3000円

日本学術会議協力学術研究団体

情報知識学会 (JSIK: Japan Society of Information and Knowledge)

会長 原田 隆史

事務局 〒164-0003 東京都中野区東中野 4-27-37 (株) アドスリー内

FAX : 050-3730-8956

E-mail : office@jsik.jp URL : http://www.jsik.jp/



Journal of Japan Society of Information and Knowledge

~~~~~目 次<表紙から続く>~~~~~

|                                                           |                      |     |
|-----------------------------------------------------------|----------------------|-----|
| タイピングによるプログラミング学習のためのソースコード提示に<br>関する一検討 .....            | 村川猛彦                 | 244 |
| 専門的データベース間の相互協力は可能か 専門特化型データベースの<br>存在意義を問う .....         | 青柳明佳                 | 252 |
| 炎上の早期解決を促す情報解析システムの実現のため、発言の立場<br>ごとに発言内容を分類する手法の研究 ..... | 横山拓也, 箕輪弘嗣, 渡邊憲二     | 258 |
| Narration Adder : 自動ナレーション追加システムの開発<br>.....              | 岩田和樹, 箕輪弘嗣           | 264 |
| 画像認識 AI を利用した危険診断システムの研究 .....                            | 田中蓮, 箕輪弘嗣            | 270 |
| 因子分析を用いた恋愛小説における文体的特徴の抽出 .....                            | 白鳥孝幸, 村井源            | 276 |
| 機能語に基づく物語会話文の文体分類と推定 .....                                | 逢坂駿也, 村井源            | 283 |
| 多言語分析のための感情辞書の検討 .....                                    | 大川慎                  | 289 |
| 女性就業者数促進のための計量分析アプローチの研究 .....                            | 植杉京佳, 大槻明            | 295 |
| PAC 分析を利用したオタクの「布教」に対する意識や態度についての研究<br>.....              | 細川怜椰, 松村敦, 宇陀則彦, 堤智昭 | 301 |
| 二次創作作品を対象とした物語特徴分析 .....                                  | 石川一稀, 村井源            | 310 |
| 神話物語と神話を原型にした現代物語の構造比較 .....                              | 吉田拓海, 村井源            | 318 |
| クエスト構造に注目したロールプレイングゲームの物語構造と<br>物語内容分析手法の提案 .....         | 中村祥吾, 村井源            | 325 |
| 情報収集行動の振り返り支援を目的とする閲覧履歴提示手法<br>.....                      | 佐藤千尋, 高久雅生           | 333 |
| 時系列を考慮したクエリ満足度の推定 .....                                   | 楊之卓, 高久雅生            | 343 |
| N-gram を用いた日記文章の文章構造の分析 .....                             | 松田雛乃, 村井源            | 355 |
| エッセイに使用されているレトリックの分類と計量分析<br>.....                        | 根本さくら, 村井源           | 361 |

## お知らせ

|             |     |
|-------------|-----|
| 事務局より ..... | 369 |
|-------------|-----|

情報知識学会誌 第31巻2号 2021年5月22日発行

編集兼発行人 情報知識学会 〒164-0003 東京都中野区東中野 4-27-37 (株)アドスリー内

E-mail : office@jsik.jp

URL : <http://www.jsik.jp/>

(振替 : 00150-8-706543)

学術刊行物 ISSN 0917-1436