

政治イデオロギーによる小説・映画の嗜好の違いの特徴分析

DENG JUNFU 樋口 心 掛谷 英紀
筑波大学

概要 本研究は、政治的イデオロギーの異なりによって、政治と直接的に関わらないジャンルにおける嗜好に差異性があるかどうかについて調べることを目的とする。保守派とリベラル派の高評価したビデオ作品の概要を形態素解析して機械学習を行った結果、保守派とリベラル派のどちらが好む作品の概要かが高確率で判定された。素性の出現頻度から、保守派は名作や人間ドラマを好み、リベラル派は刺激性の強い映画を好む傾向が見られた。また、両派の高評価した小説作品について調べたところ、保守派は文芸作品と文学賞受賞作家の作品を好む一方、リベラル派はライトノベルを好むことが分かった。

1 はじめに

自然言語処理の手法を使い、上院議員や最高裁判事のイデオロギーの傾向を分析する研究[1, 2]や、政党の特徴を分析する研究[3]が多く行われ、政治的イデオロギーにより言葉遣いが異なることが明らかになっている。

橋本らは政治家に対するレビューと電子書店のレビューを最大エントロピー法で学習した結果を利用した書籍推薦システムを提案している[4]。東らは web 上の言語資源に基づき自己組織化マップで国会議員の分類と可視化の研究を行っている[5]。掛谷らは国会議事録に基づき短命大臣の発言の特徴を分析している[6]。また、Dupree らは、普段差別に反対しているはずの米国リベラル派が、相手の人種によって違った言葉遣いをする傾向があることを見出している[7]。

最近、人々が持つイデオロギーの違いは、言葉遣いだけではなく政治と直接関連の無い趣味についても異なる傾向を示すことが明らかになっている。樋口らは Amazon のブックレビューから、保守派とリベラル派を分類したところ、政治以外のジャンルの本についても、読む本の傾向に違いがあることを見出している[8]。

そこで、本研究では樋口らの研究を踏まえ、イデオロギーの異なりによって政治と直接的に関係のないジャンルにおける人の嗜好の差異性を掘り下げることを目的とする。樋口らは保守派ユーザーとリベラル派ユーザーの全レビュー及びレビューした商品名を用いて、それぞれの特徴を分析したが、高評価も低評価も、レビューの集中する商品も

そうでない商品も含んでいるため、嗜好の差異を細かく調べるには分析対象および内容を絞って考察する必要があると考えられる。

研究の一貫性を保つため、本研究は樋口らの収集したデータベースを利用し、保守派とリベラル派の高評価した作品の内容に重きを置き、その特徴にどのような差異があるかを分析する。研究対象は樋口らが2018年12月17日から9日間の期間に収集したカスタマーレビューのユーザーの高評価した作品とする。

本論文の構成は以下の通りである。まず、第2章で映画概要サイトなどから抽出したビデオ作品の概要を用いた分類実験の手法、目的、結果を説明する。第3章では、保守派とリベラル派の高評価した小説について、ジャンル、作品、作家など共通点と差異性を調べる。最後に第4章でまとめを行う。

2 ビデオコンテンツの概要を用いた分類実験

2.1 目的と手法

本章では、保守派とリベラル派のアマゾンユーザーが高評価したビデオ作品(DVD, Blu-ray)の特徴の差異を検証する。保守派とリベラル派の定義は、過去15年間の年間ベストセラー上位20位の書籍のうち、保守派・リベラル派と広く認知されている著者の作品に高評価、またはその反対の作品に低評価をつけたユーザーを、それぞれ保守派・リベラル派ユーザーとするという樋口の基準を踏襲する。この基準に適合した本のリストを表1に示す。

表1 保守・リベラル派を分ける基準とする書籍

保守派		リベラル派	
タイトル	著者	タイトル	著者
儒教に支配された 中国人と韓国人の悲劇	ケント ギルバード	官僚の責任	古賀茂明
人間の分際	曾野綾子	母-オモニー	姜尚中
大放言	百田尚樹	しがみつかない生き方	香山リカ
呆韓論	室谷克実	悩む力	姜尚中
人間にとって成熟とは何か	曾野綾子	おひとりさまの老後	上野千鶴子
老いの才覚	曾野綾子		
国家の品格	藤原正彦		
美しい国	安倍晋三		

上記基準に基づいて定義した保守派とリベラル派のユーザーについて、高評価レビューが多い作品を調べる。具体的には、以下の基準に基づいて作品を抽出する。

- (1) 保守派とリベラル派のそれぞれについて、平均の評価が星4以上で評価者数2人以上のものを抽出する。
- (2) 上記基準で両カテゴリに重複するものは除去する。
- (3) ストーリー性のない作品は除去する。
- (4) 上記基準では保守派の作品が偏って多くなりすぎるので、保守派については評価者数2人の作品は、評価が星4.3以上の作品以外を除去する。

以上の基準に基づき、樋口らの収集したデータベースから「保守派の高評価したビデオ作品」および「リベラル派の高評価したビデオ作品」を抽出した結果を表2に示す。保守派には人間ドラマ作品、リベラル派には猟奇性やエロティシズムを感じさせる作品が並んでいる。

抽出した作品について、「映画.com」と「ウィキペディア」などから作品の概要を収集する（映画.comにない作品のみウィキペディアなどから収集する）。いずれも、2019年11月に収集したもので、収集した作品概要の平均文字数（形態素解析後の素性数）は359(67)、標準偏差はそれぞれ141(26)と279(52)である。形態素解析については、2019年8月に更新された新語辞書 neologd の適用した MeCab を用い[8]、名詞と動詞の基本型を素性としている。

作品中に頻出する人名、時間を表す言葉、作品名など内容に関わらない名詞はストップワードとして除去し、Bag of Words と fasttext[9] で作成した分散表現2種類のベクトルを作成する。分散表現の場合は Wikipedia2019 日本語版のダンプデータを使い、cbow方式、epoch数15、デフォルトの学習率で300次元の分散表現辞書を作成し、辞書か

ら対応形態素の分散表現ベクトルを取得する。単語の重要度および稀少性の影響も調べるため tf-idf 法で処理したパターンとそうでないパターンの2パターンでそれぞれ4種類のベクトルを作成する。

機械学習の手法として、サポートベクターマシン(SVM)、ランダムフォレスト(ExtraTreesClassifier)、ナイーブベイズの3種類を用いて、上述した4種類の埋め込みベクトルと組み合わせ分類実験を行う。以上述べた手法の概要を図1に示す。その上、両カテゴリの素性から出現頻度が大きく異なる名詞と形容詞を抽出し、両カテゴリの特徴を分析する。

表2 保守派・リベラル派が好むビデオ作品

保守派	リベラル派
いま、会いにゆきます	22年目の告白 私が殺人犯です
おくりびと	エロマンガ先生3
ALWAYS 続・三丁目の夕日	ベルリン・天使の詩
俺は、君のためにこそ死ににいく	クライマーズ・ハイ
セッション	L.A. コンフィデンシャル
モアナと伝説の海	好色 元禄(秘)物語
県庁おもてなし課	ちはやふる 上の句
硫黄島からの手紙	エロマンガ先生4
紅の豚	白い肌の異常な夜
結婚できない男	サヨナラCOLOR
Mr. & Mrs. スミス	歩いても 歩いても
ヒトラー最期の12日間	太陽を盗んだ男
アウトレージ 最終章	エロマンガ先生6
アイアンクロス	相川結 結歩道
ヒトラー親衛隊《SS》装甲師団	~ 美ら風 memories ~
シン・ゴジラ	エロマンガ先生5
サウンド・オブ・ミュージック	危険な情事
かもめ食堂	
おやすみなさいを言いたくて	
下妻物語	
A2	
15時17分、パリ行き	
007 ロシアより愛をこめて	
007 リビング・デイル	

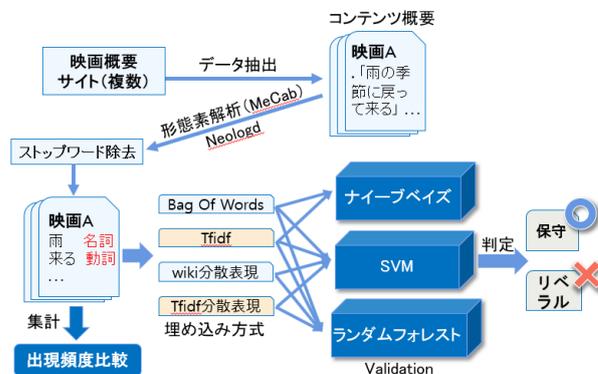


図1 システム概要

2.3 分類実験の結果と考察

図1の流れに従い、leave-one-out法を用いて合計12種類の学習モデルにおける分類実験を行った。結果を図2に示す。

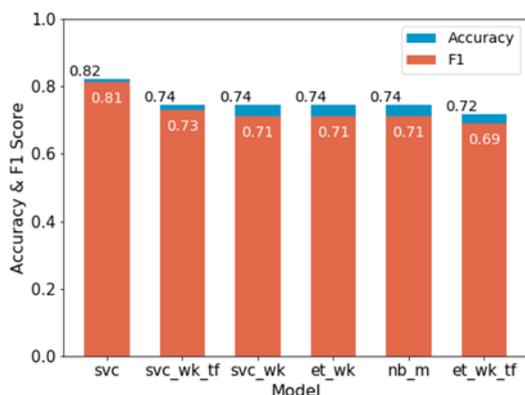


図2 実験結果の比較(上位6種のみ)。svcはSVM(Bag Of Words)、svc_wk_tfはSVM(tfidf付きwiki分散表現)、svc_wkはSVM(wiki分散表現)、et_wkはランダムフォレスト(wiki分散表現)、nb_mはナイーブベイズの多項分布モデル(Bag Of Words)、et_wk_tfはランダムフォレスト(tfidf付きwiki分散表現)

結果から Bag of Words 方式でベクトル化して SVM(kernel="linear")での分類は最もよい精度および f1 値を得た。一方、分散表現を使ったランダムフォレストの学習は 74.4%(f1 値 71.2%)の結果を出したにも関わらず、Bag of Words を使った場合は全く学習できなかった。

名詞、動詞によって保守派かリベラル派かを言い当てる正解率が最高 82%に達した結果(表3)から、イデオロギーの異なりによって、ビデオ作品における人の嗜好の差異性が、その作品概要からも判断できることが分かる。

表3 SVM(bag of words)の判定結果

	再現率	適合率	保守派判定数	リベラル派判定数
保守派	87%	83%	20	3
リベラル派	75%	80%	4	12

次に、各カテゴリにおいて名詞と動詞の出現頻度に大き

く異なる素性を抽出する。特定の映画のみに頻出する単語を特徴として誤認されないように、ビデオコンテンツ2本以上に出現する素性のみを選択する(表4)。

表4 両カテゴリの特徴的な名詞と動詞

保守派	回	リベラル派	回	保守派	回	リベラル派	回
ある日	8	犯人	8	ある	9	くる	8
受賞	6	事件	8	知る	9	せる	5
計画	6	刑事	6	描く	9	殺す	5
回	5	部	6	続ける	6	ある	5
脚本	5	担当	4	られる	6	持つ	4
ほか	5	生徒	4	出会う	5	抱く	4
公開	5	様子	3	言う	5	描く	3
彼ら	4	私	3	務める	4	見る	3
ドラマ	4	それぞれ	3	感じる	4	誘う	3
アカデミー賞	4	映画化	3	やってくる	4	つける	3
目	4	警察	3	生まれる	3	会う	3

表4から、保守派の好まれるコンテンツに「受賞」「アカデミー賞」の名詞や、「ある日」「ドラマ」などが多く現れ、元の概要テキストから確認すると保守派はアカデミー賞などの受賞作品やヒューマンドラマを好む傾向が見られる。一方、リベラル派には「犯人」、「事件」、「刑事」の名詞や、「殺す」の動詞がよく現れ、リベラル派は凶悪事件や、犯罪に関わる作品を好む傾向が窺える。

3 小説作品の差異分析

本章では、保守派とリベラル派の高評価した小説作品を中心に、ジャンル(ここのジャンルはアマゾンの書籍ジャンル)、作家、作品などの差異を調べることを目的とする。具体的には、以下の基準に基づいて作品を抽出する。

- (1) 保守派とリベラル派のそれぞれについて、平均の評価が星4以上で評価者数2人以上のものを抽出する。
- (2) 対立派の高評価数が1/3未満である場合、保守派、リベラル派の好む作品とする。
- (3) (2)に当てはまらない場合、両派が好む作品とする。

これらの基準に従うと、保守派の高評価した作品は76、リベラル派の高評価した作品は111、両派とも高評価した作品は35冊抽出された。各分類における本の各ジャンルが占める割合を図3に示す。

抽出結果から、保守派のユーザーは文芸作品と文学賞受

賞作家の作品を好み、リベラル派はライトノベルを好むことがわかる。また、両派ともに好まれているのは文学賞受賞作品が多い。さらに、各カテゴリに高評価された作家を作品数順にリストして表5に示す。

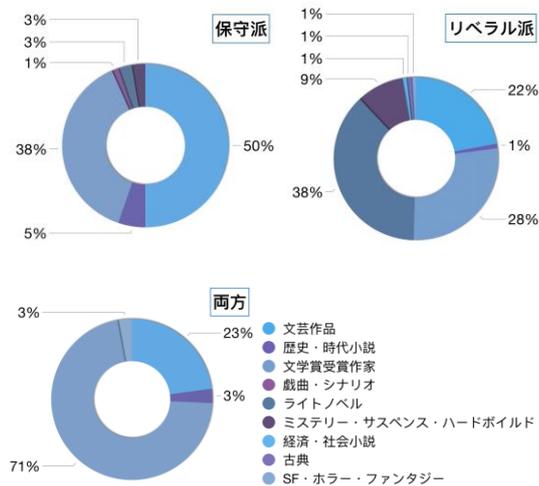


図3 各カテゴリが高評価した本のジャンル

表5 各カテゴリの高評価した作家

保守派	作品	リベラル派	作品	両派	作品
池井戸潤	12	東野圭吾	10	池井戸潤	8
百田尚樹	10	池井戸潤	8	東野圭吾	6
東野圭吾	9	村上春樹	5	中島京子	2
篠田節子	3	吉田修一	3	重松清	2
西村賢太	3	宮部みゆき	3	司馬遼太郎	2
村上春樹	3				

池井戸潤と東野圭吾の作品は両派とも多く好評されている。その内訳をみると、「下町ロケット」と「容疑者 X の献身」は、保守派、リベラル派とも高評価をした人数が多いことが分かった。

4 まとめ

本研究は、政治的イデオロギーの異なりによって、政治と直接的に関わらないジャンルにおける嗜好に差異性があるかどうかについて、ビデオ作品および小説を対象に検証を行った。

まず、保守派とリベラル派の高評価したビデオ作品の概要を形態素解析して、名詞と動詞のみ使い、4種類の埋め込みベクトルと3種類の分類器で合計12種類の学習モデル

で分類実験を行った。結果から、最高82%(81%)の正解率(f1値)で分類でき、ビデオ作品の内容を表す名詞と動詞のみ、保守派とリベラル派のどちらかに好かれるかをある程度の確度で判定することが確認された。また、素性の出現頻度から、保守派は名作や人間ドラマを好み、リベラル派は刺激性の強い映画を好むことが分かった。

次に、両派の高評価した小説作品について調べた。保守派は文芸作品と文学賞受賞作家の作品を好むが、リベラル派はライトノベルを好むことが分かった。また、両派ともに好かれる作家は、池井戸潤と東野圭吾ということがわかった。現在、世界中で保守派とリベラル派の分断が起きているが、両者に共通して支持される作品を分析することで、社会の分断を抑制するためのヒントが得られる可能性もある。

今後は、池井戸潤や東野圭吾のように、保守派とリベラル派の双方に支持される作品について、保守派とリベラル派でレビューの内容に違いがあるかを分析する予定である。また、海外においても、保守派とリベラル派で好む映画や小説に同様の違いがあるかを調べることも今後の課題である。

参考文献

[1] Daniel Diermeier et al., British Journal of Political Science, 2011.
 [2] Benjamin E. Lauderdale et al., American Journal of Political Science Volume, 2014.
 [3] 畑中允宏 他, 言語処理学会第15回年次大会, 2009.
 [4] 橋本悠 他, 言語処理学会第16回年次大会, 2010.
 [5] 東宏一 他, 言語処理学会第17回年次大会, 2011.
 [6] 掛谷英紀 他, 知能と情報 Vol.31(2), 2019.
 [7] 樋口心 他, 日本選挙学会, 2019.
 [8] MeCab, taku910.github.io/mecab/
 [9] Piotr Bojanowski et al., Trans. of the Association for Computational Linguistics, vol.5, 2017.