

# 全共闘世代をテキストマイニング

近藤 伸郎・正田 備也（立教大学大学院 人工知能科学研究科）

本研究では、「全共闘世代」と呼ばれる世代とその周辺を対象としたアンケート・データをテキスト分析した。高齢化し鬼籍に入る者も少なくない全共闘世代の実態解明を目的として行われたのが本アンケートで、基本的な情報に関する設問に加えて、政治運動に参加した動機やメンタリティ、それを今どう考えているかの総括に関する設問もあり、回答者からは多数の自由記述的な言葉が寄せられた。それらを分析することで、世代の何らかの特徴を得られないか。今回は、テキストをまとめて形態素解析し、特定の二値カテゴリにしたがって分類器に学習させ、その分類器において重要度の高い特徴量を重要な単語としてあぶり出した。その上で、その重要語の特徴を考察した。

## Text Mining for “Zenkyoto Generation”

Noburo Kondo / Tomonari Masada

(Graduate School of Artificial Intelligence and Science, Rikkyo University)

In this study, we tried a text analysis of questionnaires from "Zenkyoto Generation", a generation who committed themselves to student activism in 1960s. The questionnaires include questions about their motivations and mentalities for participating in political movements and their retrospective thoughts in addition to basic information, so that we have got many open-ended answers as text data. In this study, we tried to figure out remarkable characteristics of this generation based on the text data. After performing morphological analysis, we trained classifiers according to specific binary categories and extracted important words as the words having high feature importances in the trained classifiers. We then discussed the characteristics of the extracted words.

### 1. はじめに

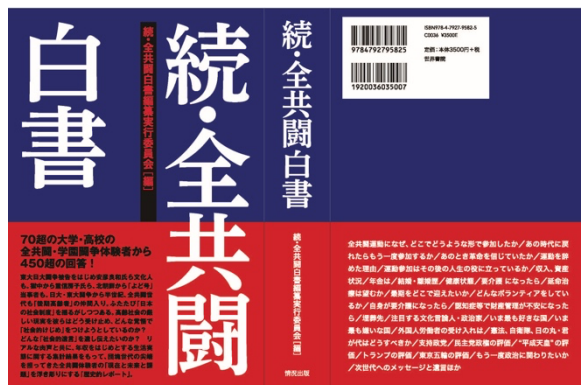


図1. 『続・全共闘白書』[1]  
Figure 1. “Zoku-Zenkyoto Hakusho”

1968年創業の雑誌社、情況出版が編んだ『続・全共闘白書』[1]という本が昨年末に刊行された。「全共闘世代」と呼ばれる世代とその周辺へのアンケート・データをそのまま収録した単行本で、700頁を超えるページ数となっている。そこには、大学や職業、年収や家族形態など基本的な質問事項に加え、「当時革命が起こると信じていたか」「もう一度あの時代に戻ったらどうするか」など、政治運動に参加した動機やメンタリティ、それを今どう考えているかの総括に関する設問もあり、回答者からは多数の自由記述的な言葉が寄せられた。今だからこそ語れる思いの丈が各人各様に書き込まれており、なかには数ページに及ぶもの

まである。これらを分析することで、全共闘世代の特徴を浮かび上がらせることができないかというのが、本研究の問題意識である。

研究方法としては、これら自由記述的なテキストを形態素解析し、特定の二値カテゴリにしたがって分類器に学習させ、その分類器において重要度の高い特徴量を重要な単語としてあぶり出す。今回は、大学入学年度が1968年以前か1969年以降かという二値カテゴリと、運動に参加した立場が一般学生としてか活動家としてかという二値カテゴリを使った。これらの二値カテゴリで自由記述を文書分類し、それぞれの場合で重要度の高い特徴量を求めることで、この世代を特徴付ける重要語を明らかにする。

### 2. データについて

本研究で分析の対象とするアンケートは、『続・全共闘白書』[1]を編むために実施されたものである。該当する世代の451人に対して、出身大学、性別、出生年、入学年など、75項目の質問項目が提示された。<sup>a</sup> 特に問73の「子供あるいは孫の世代への遺言（継承してほしいこと、ほしくないこと等）をお書きください」と、問75の「最後に、今だから話せる当時のこと、今こそぜひとも伝えたいことをご自由にお書きく

<sup>a</sup> 実際に使われたアンケート用紙は、  
<http://www.zenkyoutou.com/アンケート用紙.pdf>  
にて確認可能である。

ださい」は、自由記述の質問項目であり、多くの回答者が長文を書き込んでいる。他にも、各設問での想定を超える言葉が書き込まれるケースが多くあった。<sup>a</sup> よって、本研究では、可能な限りテキストを拾うという観点から、これらの「設問外回答」のテキストも全て結合させて、いくつかの前処理を行った上で<sup>b</sup>「自由記述欄」として分析の対象にすることにした。「特になし」やそれに類する回答の意思がない旨の回答、また、あまりに短い故に形態素解析で品詞を絞った後で空白になるものも除いた結果、451人の回答中、有効回答は440であった。

### 3. 分析手法

回答者ごとの自由記述を一つの文書とみなし、まずはSudachiPyのSplitMode=Cを用いて形態素解析した。デフォルトの辞書Coreだと、例えば、活動家の固有名詞である「藤本敏夫」が「藤本

<sup>a</sup> 例えば「ア 自らの信念で」「イ 友人先輩に誘われて」「ウ 社会正義から」「エ 時代の雰囲気」という選択肢で運動を参加した理由を問う設問に対して、「見なかった様に生きることは出来る。しかしそれは『見たということだ』これから一生自分を偽って生きるのは嫌だった。」というものである。

<sup>b</sup> 明らかに「設問外回答」として入力している部分に関してのみ、テキストを拾った。例えば、「もう一度あの時代に戻ったらどうするか」の設問で、「するが、今度は正しく闘いたい。闘えると思う。それとは別に、せつかく大学を占拠したのだから、全共闘運動はもっと学術・研究とか文化・芸術とかの運動をやればよかったと思う。」というものは、「する」の部分を取り除き、それ以降のテキストを有効とした。また、「社会主義の有効性を失ったと思うか」という設問の回答で、「社会主義の定義によるが、自由、平等、博愛の精神は有効性は失われていない。」というものについては、「社会主義」という文言が設問と重複するため取り除き「定義によるが、」以降の部分を採用した。このようにテキストを拾った設問は、問2、問3、問4、問5、問6、問7、問9、問10、問11、問12、問13、問14、問15、問45、問46、問47、問48、問49、問50、問51、問52、問53、問54、問65、問68、問69、問70、問72、問73、問74、問75とした。今回は、イシューとしては憲法や安保などの全共闘運動に大きく関わりのある項目は取り入れ、「死後どのように葬られたいですか」といった政治から離れたものや、「2020東京五輪開催をどう評価されますか。」「トランプ米大統領をどう評価されますか。」といった個別すぎる論点に関しては排除した。

と「敏夫」に分割されるなど、今回の分析では不適切な分割が生じたので、固有名詞なども充実したFullの辞書のほうを用いた。品詞に関しては、助詞や補助記号など、テキストの特性を表す上で重要でないと考えられる品詞は除外し、名詞・動詞・形容詞のみをピックアップした。さらに、SlothLibのWebページ<sup>c</sup>から取得したストップワードのリストに、自明に分析に貢献しないと思われる単語<sup>d</sup>を加え、これらを除外した。

今回のテキスト分析では、形態素解析の結果をそのまま用いるのではなく、一部、自動的に分かち書きできなかったものを手作業で後処理した。例えば、「羽田 闘争」や「10 8」、「sect 6」などは、分割されては意味をなさないため、それぞれ「羽田闘争」「10・8」「SECT6」と後処理で置き換えている。

このようにして単語に分割された回答者ごとのテキストを、scikit-learnのTfidfTransformerで文書ベクトルに変換した後、質問項目のなかから選ばれた注目に値する二値カテゴリにしたがってラベリングし、これを分類器に学習させた。そして、学習後に各特徴量の重要度を計算することで、二つのカテゴリを識別するのに寄与している単語を明らかにする。

全共闘運動に関しては、運動の「最高揚期」である1968年までと、いわゆる「内ゲバ」が惨憺たるものになっていく1969年以降では雰囲気が違うと言われている。そのため、今回は大学入学年を問う質問の結果を用いて、{1968年以前入学のクラス, 1969年以後入学のクラス}という二値カテゴリを採用し、全共闘世代のなかでも細かい年長・年少によってテキスト内容に注目すべき差が生じているかを検証する。この二値カテゴリによって回答者ごとの自由記述をラベリングしたものを教師データとし、分類器に学習させた。

また、全共闘運動自体は「大衆運動」であるという言説があり、いわゆる党派活動家の運動とは区別されるべきだという意見も根強い。そこで、「全共闘運動あるいは何らかの政治社会運動に参加しましたか。」という設問への回答を用いて

<sup>c</sup> <http://svn.sourceforge.jp/svnroot/slothlib/CSharp/Version1/SlothLib/NLP/Filter/StopWord/word/Japanese.txt>

<sup>d</sup> 具体的には、「いう」「いる」「ある」「する」「こと」「ない」「なる」「もの」「できる」「ほしい」「対する」「みる」「おく」「とき」「いく」「くる」「思う」「よる」「言う」「持つ」「くれる」「やる」「出す」「知る」「しまう」「来る」「かかる」「おる」を加えた。





これら 438 人の回答者を、{1968 年以前入学のクラス, 1969 年以後入学のクラス}という二値カテゴリへ分割すると、図 5 のような分布となった。1968 年以前に入学した人は 345 人で、これらの人が書いたテキストの長さを文字数でカウントすると平均は 598.85 文字、1969 年以後の 93 人については平均文字数 425.95 文字となった。

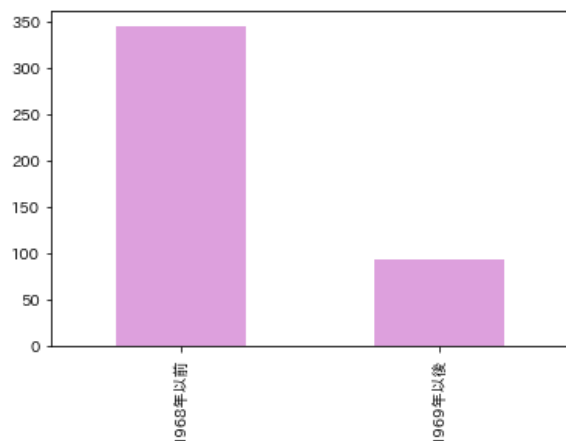


図 5. 1968 年以前入学と 1969 年以後入学の分布  
Figure 5. Distribution of “before 1968” and “after 1969” classes

入学が 1968 年以前か 1969 年以後かという二値カテゴリでテキストをラベリングし、これを訓練データとして LinearSVC に入力した。得られた線形モデルで係数の絶対値が大きい単語を、表 1 と表 2 に示す。どちらのクラスの重要語であるかは係数の符号で判定した。参考までに、各単語のコーパス全体での出現回数も示してある。

表 1. 1968 年以前入学クラスの重要語 20  
Table 1. Top 20 most important words (before 1968)

単語 Words	重要度 (係数の絶対値) Feature importances	頻度 Frequencies
権力	0.7967	111
健康	0.7720	53
人々	0.7452	71
生活	0.7286	123
多い	0.7072	65
強化	0.6943	36
考える	0.6731	331
憲法	0.6725	298
失う	0.6560	31
独立	0.6158	48
社会的	0.6147	80
批判	0.6142	50
辺野古	0.6091	15
戦争責任	0.6006	25
つく	0.5954	163

安倍	0.5947	114
民主主義	0.5905	110
変化	0.5881	51
受け入れる	0.5659	27
愛する	0.5652	11

表 2. 1969 年以後入学クラスの重要語 20  
Table 2. Top 20 most important words (after 1969)

単語 Words	重要度 (係数の絶対値) Feature importances	頻度 Frequencies
意見	1.2873	29
老後	1.2315	11
市民	1.1755	69
愚問	1.1574	4
社会活動	1.1531	11
武器	1.1373	16
安保条約	1.0146	31
可能性	0.9095	21
変革	0.8699	65
問題意識	0.8528	8
反原発	0.8518	26
人脈	0.8434	4
市議会議員	0.8384	2
社会民主主義	0.8354	16
決める	0.8233	19
苦労	0.8180	15
立場	0.8112	54
属州	0.7938	3
解決	0.7681	33
現在	0.7624	144

分類手法のハイパーパラメータを、10-fold クロスバリデーションによって分類精度で評価し、Optuna を用いてチューニングしたところ、min\_df が 1、max\_df が 0.4799、C が 2.5496、intercept\_scaling が 51.45 のときに、最適な分類精度 0.7991 となった。この値は、1968 年以前入学の人の割合 78.77%とは異なっており、全てのインスタンスを多数派のクラスへと単純に属させる分類とは異なる分類結果が得られている。

一方、運動に参加した立場に基づく {一般学生として参加のクラス, 活動家として参加のクラス} という二値カテゴリに関しては、有効回答数は 419 で、図 6 のような分布になった。一般学生として参加のクラスに属する 159 人が書いたテキストの長さの平均は 406.83 字、活動家として参加のクラスに属する 260 人については平均 602.07 字となった。

この二値分類についても同様に LinearSVC にクラス境界を学習させ、得られた線形モデルの係数の絶対値が大きかった単語を表 3、表 4 に示す。

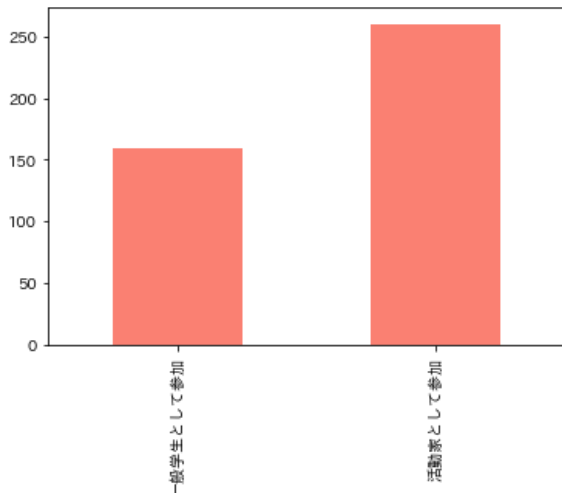


図 6. 一般学生として参加と活動家として参加の分布  
Figure 6. Distribution of “students” and “activists” classes

与える	0.9095	39
わかる	0.8976	45
沖縄	0.8790	69
救援	0.8570	14
考える	0.8506	331
結果	0.8259	41
文化	0.7817	67
日本国	0.7488	41
反戦	0.7400	76
仕事	0.7383	67
平等	0.7286	65
一人	0.7168	53
全て	0.7030	45
過去	0.6934	28
減らす	0.6856	10

表 3. 一般学生クラスの重要語 20

Table 3. Top 20 most important words (students)

単語 Words	重要度 (係数の絶対値) Feature importances	頻度 Frequencies
生きる	0.9768	227
追求	0.9010	16
安保	0.8537	152
肯定	0.8499	10
戦争責任	0.8168	25
重視	0.8156	10
地球	0.8068	19
子ども	0.8006	22
国民	0.7964	178
軍隊	0.7921	68
趣味	0.7883	9
行使	0.7510	23
明日	0.7385	8
改正	0.7374	66
中退	0.7247	11
武器	0.7241	16
堅持	0.7124	21
関心	0.7049	41
平和	0.6980	291
回答	0.6845	33

表 4. 活動家クラスの重要語 20

Table 4. Top 20 most important words (activists)

単語 Words	重要度 (係数の絶対値) Feature importances	頻度 Frequencies
外交	1.0283	37
課題	0.9938	55
強化	0.9549	36
現在	0.9473	144
就職	0.9371	68

この二値分類についてもハイパーパラメータを Optuna で同様にチューニングした結果、min\_df が 5、max\_df が 0.5177、C が 0.7829、intercept\_scaling が 38.27 のとき、ベストな分類精度 0.6469 となった。活動家として参加した人の割合 62.05%とは異なる結果となっており、やはり、全インスタンスを多数派のクラスに単純に属させる分類とは異なる結果が得られている。

## 5. 考察

どちらの分類でも、二クラスのサイズに隔たりがあるが、全てを多数派のカテゴリに分類した際の精度を超える精度にはなっていない。文書数が少ないため精度はあまり良くないが、今回は未知文書のクラスを予測することが目的ではない。また、ハイパーパラメータの値を Optuna の与えた最適値から多少動かしても上位の重要語はほとんど変化しなかった。よって、重要な特徴量を抽出するという目的は果たせていると考える。

まず、大学入学年度による分類であるが、1968年以前の「年長世代」に対応する重要語には、後期高齢者が多いこともあるのか、「健康」という言葉が入っている。「死の直前まで健康でいたい」など「団塊世代の最重要課題」の問いに対して、健康を挙げている人が多かった。また、「生活」に関しても、同じ問いで、1968年入学のある男性は最重要課題に「公的年金のみでは生活が成り立たず、いつまで仕事を続けられるかという事」を挙げている。最重要語に入っている「権力」に関しては、「国家権力」「反権力闘争」という文脈で使われて、「最後にこれだけは言いたい」の項目で、1965年入学のある男性は「権力と対峙したあの恐怖に耐え細やかながらも仲間と連帯して抵抗の意思表示を行っていた経験が、『世の中に怖いものなんて、それほどはない』という確信が

持てた。」と述べている。政治に関する重要語では、「憲法」「辺野古」「戦争責任」「民主主義」などがあるが、これらは団塊の世代である全共闘世代のなかでも特に年長組が、戦争に連続した世代であることの現れだろう。

1969年入学以降の「年少組」はどうであろうか。最重要語が「意見」なのは、リベラルな価値観が反映されてのことだと考えられる。例えば、パルシステム生活協同組合の理事長で有名な1969年入学の若森資朗さんは、「段階の世代の最重要課題」として「分断と対立をなくしたい。反対意見を尊重する寛容な社会へ」と述べている。

「武器」に関しては、自衛隊、憲法関連が多く、意外にも、連合赤軍に関してやゲバ棒などの当時の運動の暴力の文脈で触れているものはなかった。(1つだけ傘を武器にしたという記述があるのみだった。)1969年以後に関しては、データが少ないこともあるのか、年少世代を特徴付ける言葉として確信が持ちづらいものも多い。「老後」は、1968年以前にも多く見られたし、1970年入学の男性の「今のところ「老後」とか「終活」という概念はない」などという使われ方もあった。

「属州」「愚問」に関しては、一人の方が4つ以上記述欄に「愚問」と書いているなど、データの偏りがそのまま反映されている。(「属州」も同様。)

次に、運動に参加した立場で見てみる。「一般学生として参加」したグループの重要語には特徴的なものが多いように思われる。「生きる」は、必ずしも「一般学生」に限ったものではなく、活動家の方たちのテキストにもたくさん見られたが、「毎年毎年を楽しく生きること」「前向きに生きる」のようなポジティブな言説が多かった。「地球」「子ども」「国民」「趣味」「平和」など、多くの人に共感しやすいテーマも目立つように思われる。ある女性は、「何らかの政治活動、社会運動に参加の意思をお持ちですか」の問いに対して、「地域でできることをしています。子ども、災害、など。」と答えている。また、「団塊世代の最重要課題」として、「子どもや孫たちに、思いを伝えていく」「生きてきた歴史を子ども達に伝える」など「子ども」という言葉が入った回答が複数人から寄せられている。温暖化や原発など地球環境を気にしたものも多い。「趣味」に関しては、「67歳まで必死になって仕事をしてきたので、今は好きな事や趣味を精一杯楽しみたい。」というような回答もあった。

「活動家」に関しては、「沖縄」「救援」「反戦」など具体的な活動内容を示す言葉が目立つ。最重要語になっている「外交」は、「日米安保」の関係で述べている人が多かった。興味深いのは、「就職」「仕事」という言葉が入っていることである。

特に「就職」に関しては、「運動から離れた理由」や「運動参加による損害」で顕著に見られた。活動家の方が損失を負っている、ということだろう。

「一般学生」との差が現れた結果だと考えられる。「日本国」は、「日本国憲法」絡みが多かったが、「日本国」という単位をナショナリズムとして肯定している表現と否定している表現がどちらも見られた。いずれにせよ、国際・国内政治を議論する活動家ならではの特徴と考えられるかもしれない。「平等」に関しては、社会主義の理念そのものであり、例えば、「日本社会はどうあるべき」という問いに対する「生まれながらの不平等のない社会、みんなが安心して暮らせる社会の実現。搾取することが当然という価値観からの断絶。」という回答など、社会主義の有効性は失われていないと考える回答が多かった。この文脈のほか、ジェンダー関係の文脈でも「平等」の語は多く見られた。

## 6. 今後の展望

今後の展望としては、今回はシンプルに線形モデルの係数の絶対値を重要度として分析を行なったが、機械学習の解釈性[2][3]の観点から、より立ち入った分析をすることも考えられる。

また、今回は二値カテゴリを目的変数としたが、テキストの方を目的変数として学習させた分類器に依拠して分析を行うことも考えられる[4]。例えば、今回特定したいいずれかの重要語を含むか含まないかを二値の目的変数、カテゴリカルな設問項目を説明変数にし、回答者の様々な属性と言葉の選び方の相関を考察することも、今後の課題として考えられる。

## 参考文献

- [1] 続・全共闘白書編纂実行委員会 編著 (2019). 『続・全共闘白書』, 情況出版.
- [2] Qingyuan Zhao and Trevor Hastie (2019): "Causal Interpretations of Black-Box Models," *Journal of Business & Economic Statistics*, DOI: [10.1080/07350015.2019.1624293](https://doi.org/10.1080/07350015.2019.1624293)
- [3] Christoph Molnar (2020). *Interpretable Machine Learning A Guide for Making Black Box Models Explainable*. <https://christophm.github.io/interpretable-ml-book/> (accessed 2020-11-8)
- [4] Naoki Egami, Christian J. Fong, Justin Grimmer, Margaret E. Roberts, and Brandon M. Stewart (2017): "How to Make Causal Inferences Using Texts," <https://scholar.princeton.edu/bstewart/publications/how-make-causal-inferences-using-texts> (accessed 2020-11-8)