

文末語を用いた社説検索プログラム

石井 真人

Editorial Search Program using Sentence Final-Words

Masato ISHII

Abstract

In the 2024 Tokyo and Hyogo gubernatorial elections, information from major media was overwhelmed by information from social media. The expansion and penetration of information from social media showed the possibility of surpassing major media, which have advantages in terms of human resources and funds.

Mobile phone users can easily access online information and can disseminate it immediately. Then, the information acquired by users spreads instantly in the cyberspace.

Therefore, false information on social media may have a greater impact on people than major media.

To avoid this situation, Internet users need to have access to authorized information as well. The information is disseminated from media that are not subject to political regulations, such as newspapers, weekly magazines, and academic journals. However, most of these media are paper media. The content is more difficult than SNS and is paid, so general Internet users tend to avoid it. The purpose of the research is to develop a computer program that can obtain authorized information using electronic media. The source of information is editorials from major newspapers. Editorials are clear opinions on major articles from each newspaper company.

Editorials are not subject to legal regulations like television broadcasts, so each newspaper can make its own claims. Reading editorials improves Internet users' information literacy, so the editorial search program is an effective vaccine against false information.

The results obtained are as follows:

(1) The search items of the editorial search program include the editorial title, keywords, publisher name, and publication date. In particular, evaluation of the editorial content is incorporated into the search items.

(2) The evaluation items are divided into three types: negative, positive, and suggestive, so multiple editorials with the same title can be compared immediately.

(3) The evaluation items are automatically determined using the frequency of use of sentence-final words and their weights. The agreement rate between the automatic determination results and the evaluators has been 69.6%.

Keywords: editorial, newspaper, sentence-final word, Quantification2, computer program, information literacy

1 はじめに

2024年行われた知事選では、それ以前の選挙結果と比較して大きな変化が報じられた。東京都知事選では、現職小池氏の最大の対抗者と見られた蓮舫氏を上回る票を得た石丸氏が注目された^[1]。また、兵庫県知事選挙では、議会の混乱や内部告発などで不利を報じられていた元知事の斉藤氏が勝利をおさめた^[2]。これらの知事選挙の共通点としては、従来の大手メディアの報道の予想を覆す得票数を獲得した点である。石丸氏は広島県の市長の経験があるとはいえ、東京では小池氏や蓮舫氏と比較すると知名度は圧倒的に低い^[3]。斉藤氏は県幹部の自殺など数々の不利な情報が流布されていたが^[4]勝利することができた。

2人の選挙運動の共通点は、まずインターネット、とくにSNSを最大に活用したこと。そして、その活用において疑惑が生じている点である。石丸氏は市民団体により、選挙費用に関して公職選挙法の疑いで告発された^[5]。また斉藤氏と選挙に関係した西宮のPR会社が、同じく公職選挙法違反の容疑で告発されている^[6]。この現状から分かることは、従来からの大手メディアが発信する情報の信頼性が低下している可能性がある^{[7][8]}、それを超す情報発信力がインターネットにはあると考えられる。発信する情報の質も大きく異なるケースが多い。

大手メディアの場合、発信する情報は少なくとも複数の人で査読、チェックされオーソライズされた内容である。したがって特定の候補者に隔たことは少ないが、その分受け取り側のインパクトは減少される。また、SNSのように個人の意見がすぐに拡散できる瞬発性は低い。特にテレビでは放送法の関係もあり量的公平性が求められるため、選挙期間の報道にも限界がみられる^{[9][10][11]}。

一方新聞は、放送法のような縛りは無くオンライン情報も充実しているので携帯性にも遜色はないが、有料であることが多い。また、活字離れもあって情報が全世代用に選別されるため、SNSの内容に比べると直感的に受け入れにくい世代もある^[12]。

しかし、選挙のような国民に直接関係する公的事業では、ネットユーザはバイアス化したネット情報への疑問を抱いて情報リテラシーを高めるべきである^[13]。それには受け手である個人がまずオーソライズされた情報を習得して予備の情報も得た状態で、SNS情報にアクセスするほうが良い。

本稿では、これらのメディア間の差を埋めてネットユーザがマスメディア情報を取得する方法を検討した。対象としたメディアは新聞である。新聞は、多方面の記事を扱い専門的な内容もある。しかし重要な記事に関する新聞社の意見として社説を読むことで、社会の大きな流れは理解できることが多い。欠点としては、一定の傾向に重きを置く新聞もあるが、これは万人向けのテレビと違い長所でもある。

本研究では、主要5新聞の社説を容易に比較できる検索プログラムを検討した。社説タイトルからも検索できるが発刊日などからも検索できる。社説本文は搭載していないが、タイトルなどの検索結果から、スマートフォンなどで当該サイトから読むことはできる。

現在では新聞記事検索はスマートフォンで可能である。これらの機能との違いは、内容の判断を3段階で行っている点である。この機能で他社の社説との比較が瞬時に行えるので、ユーザは、社会の流れを概略的に把握することが可能である。なお、現時点では対応年は直近3年である。

2 実験方法

本章では、検索プログラムの内容について述べる。そして社説内容の判断基準を述べる。

2.1 検索対象とする新聞と発刊期間

本稿では、発行部数の多い朝日、読売、毎日と産経を選択した。また既報^[14]と同様の理由により、東京新聞を加えた。期間は、2022年9月1日から2024年6月30日までに発刊された分である。

2.2 社説検索に用いたツールと検索対象項目

検索プログラムは、Windows パーソナルコンピュータで動作する。Perlで記述してあるため、他

のマシンにも容易に移植可能である。検索項目は、タイトル、発刊日、新聞社名のほか表-1に示す項目が対象である。なお社説本文は、ユーザが結果をもとに好みの媒体で読めばいいので、この検索プログラムには含まれない。

〔表-1〕 検索対象項目

対象項目	内 容
キーワード	社説本文から抜き出した専門的な用語。検出されない社説もある。
判別結果	内容を否定的内容、肯定的内容あるいは提案に分類した判別結果。
判別方法	自動判別あるいは複数の評価者による判別。

2.3 判別方法

既報^[4]にて、自動判別との比較のため3人の評価者が、5新聞の判別対象の社説を表-2の3判別項目のカテゴリーに分類した。本稿でもこの結果の一部を用いて自動判別との比較を行っている。評価者による社説内容の判別項目と基準を表-3に示す。

〔表-2〕 共通文末語カテゴリーと使用頻度

共通文末語のカテゴリー	使 用 頻 度
1	1 社説中に対応する共通文末語が全く使用されていない。
2	1 社説中に対応する共通文末語が1回使用されている。
3	1 社説中に対応する共通文末語が複数回使用されている。

〔表-3〕 評価者による社説内容の判別

判別項目（カテゴリー）	判 別 基 準
否 定 (1)	主題に異なる意見を述べておりジャーナリズム色彩が強い社説。
提 案 (2)	新たな知見から主題に対し解決案としての提案内容が強い社説。
肯 定 (3)	主題をほぼ肯定し、あるいはそれを支援する内容が強い社説。

自動判別は、既報^[4]にて数量化二類による社説判別用に抽出した5紙共通に用いる46の共通文末語を用いて、判別計算結果から各共通文末語のウェイトを算出した（詳細は3.2.2）。そして、検索対象の社説に用いられた文末語の合計のウェイトを計算して、あらかじめ決定しておいた否定・提案・肯定のウェイト範囲と比較して社説内容の判別を行った。

2.4 共通文末語の抽出

既報^[4]にて用いた共通文末語のうち使用頻度の低い17語を入れ替えた。語の選択は、2023年6月から翌年2月までの社説で使用頻度の高い文末語である。抽出結果の一部を表-4に示す。

〔表-4〕 文末語抽出例

共通文末語	例 文
を求めたい	情勢を見極めた冷静かつ丁寧な情報提供を求めたい。
は明らかだ	政府の法案より説得力があるのは明らかだ。
評価できる	おおむね順調に機能したことは評価できる。

2.5 文末語のウェイト

判別対象の5社の社説について、46語の文末語使用頻度と数量化二類によるカテゴリー数量が大きく関連しているならば、その文末語にウェイトをかけることにより、総ウェイトが対象社説判別項目の指標となり得る。

判別は、否定・提案・肯定の3項目であり、自動判別ではこれらの外的変数のカテゴリー数量がこ

の順序ならば、(否定 $\leq$ 提案 $\leq$ 肯定)に設定できるから、これらの3項目の閾値を1社説の総ウェイト(W)の範囲に設定すれば良い。総ウェイトの上下の閾値を $Thw1$ 、 $Thw2$ (ただし、 $Thw1 > Thw2$)と置くと判別項目と閾値の関係は式-1で示される。

$$(\text{否定 } W \leq Thw2) \parallel (Thw2 < \text{提案 } W < Thw1) \parallel (Thw1 \leq \text{肯定 } W) \quad (\text{式-1})$$

3 実験結果と考察

本章では、2章の方法に従って、社説内容の自動判別を行った結果とそれについての考察を述べる。

3.1 出力

検索プログラムは一部の出力がコード化されており、またグラフィカルユーザインタフェースが実装されていないため、ユーザには使いにくい。これは今後改善していく予定である。

現時点では、画面指示に従いキーワードを順次入力すると、AND、OR検索で結果を出力する。新聞名や発刊日あるいは判別項目で絞りこむことも可能である。図-1に、発刊日とタイトルの一部を特定して検索した結果の画面を例示する。この例では、入力語‘世界’を含んだ社説タイトル、発刊日、新聞名、判定結果が示される。

```

MINGW32:/c:/home/paper/no105headline7/function
-----stop->y go on->n
y
Hit value no of seach(val to key OR)=8 9 10
title=T0 混迷する世界 強権政治の台頭は許されない
judge_by_val date(newspaper)=20180104(asa) opinion=falt
untj nothing
title=EN 世界経済 安定成長に影落とす保護主義
judge_by_val date(newspaper)=20180104(yom) opinion=supp
untj nothing
title=DM 岐路に立つ世界 「自国第一」からの脱却を
judge_by_val date(newspaper)=20180104(mai) opinion=sugg
untj nothing

```

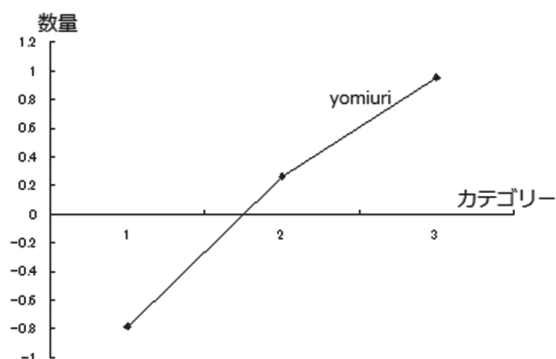
〔図-1〕出力画面

3.2 自動判別閾値の決定

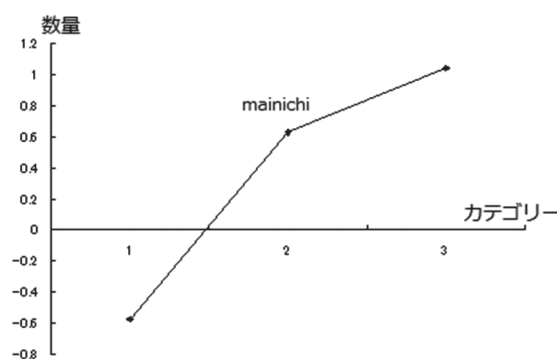
3.2.1 カテゴリーとカテゴリー数量の関係

新聞名ごとに判別計算した結果、全新聞で1軸の相関比が2軸より上回っているため1軸に関して検討した。外的基準である評価者による評価項目のカテゴリー数量は、すべての新聞がカテゴリー1から2へは増加傾向を示した(図-4~6)。なお2から1へは減少傾向を示した新聞もあったがカテゴリー3のサンプル数が他に比べて少ないことなどから、新聞全体としてカテゴリー順にカテゴリー数量は増加すると捉えた。

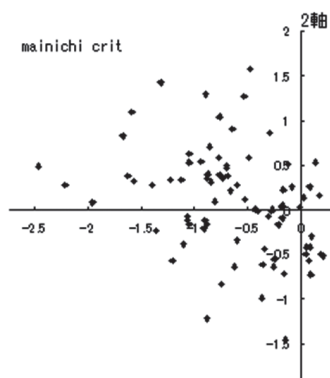
つまり、社説内容は評価が否定から肯定に向くほどカテゴリー数量が上がっていくと解釈される反面、厳しい評価ほどカテゴリー数量は小さい傾向を示す。読売と毎日新聞のカテゴリー数量とカテゴリーの関係を図-2~3に示す。また、図-4~6に毎日新聞のカテゴリーごとのサンプルカテゴリー数量を示す。



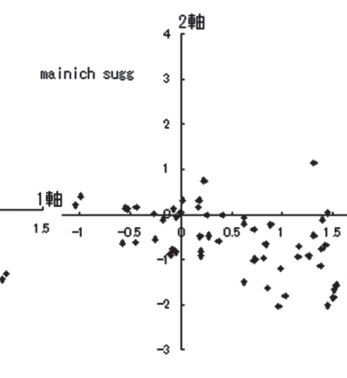
〔図-2〕 外的基準値のカテゴリー数量（読売）



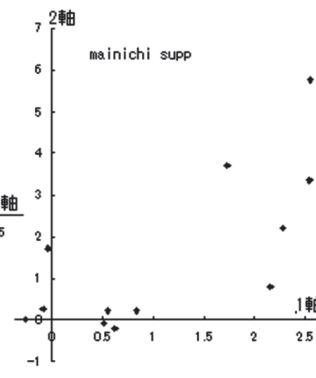
〔図-3〕 外的基準値のカテゴリー数量（毎日）



〔図-4〕 カテゴリー1の数量



〔図-5〕 カテゴリー2の数量



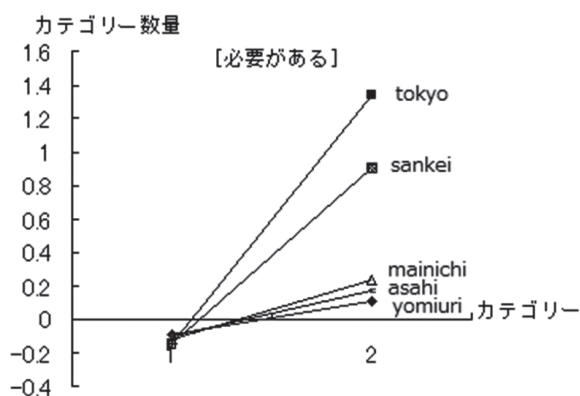
〔図-6〕 カテゴリー3の数量

3.2.2 文末語のカテゴリー数量

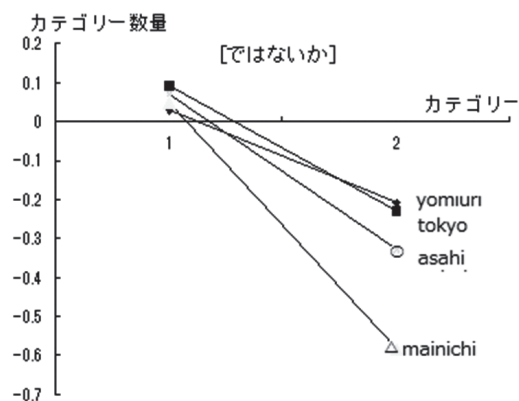
文末語の頻度別カテゴリー数量は、文末語ごとにばらつきがみられる。しかし各文末語のカテゴリー数量もカテゴリーの増加による増減の傾向がみられる。これは、文末語使用数の増減が評価者による評価項目のカテゴリー数量に関係していると捉えることができるから、文末語の総ウェイト (Tw) で評価項目を予測できることを示す。

共通文末語はカテゴリーが上がるにつれてカテゴリー数量が下がる語が一番多く次に上がる語、およびこれらの中間の語が混在する。カテゴリーが上がるとカテゴリー数量が上下する共通文末語は、その語の使用頻度増加に従い評価項目に強く関係すると解釈される。とくに、カテゴリーが上がるとカテゴリー数量が下がる共通文末語は、頻度の増加に従って否定的な評価が予想される。一方カテゴリー数量が上がる共通文末語は、頻度の増加に従って肯定的な評価が予想される。

肯定的、否定的な内容が予想される代表的な共通文末語とカテゴリーの関係を図-7～8に示す。



〔図-7〕 文末語とカテゴリー数量－必要がある



〔図-8〕 文末語とカテゴリー数量－ではないか

3.2.3 共通文末語ウェイト

共通文末語の使用頻度で評価項目の予測を行うため、全5紙共通の共通文末語のウェイト設定を行った。共通文末語の1軸のカテゴリー数量は、各紙ごとに異なるが全紙の平均値も各紙への影響と同じく全紙的な影響を持っている。全紙のカテゴリー数量が小さければ、その予想される評価項目は否定的な内容の可能性が高く、大きければ肯定的となる。それらの中間の範囲ならば提案的な内容が予測される。

そこで5紙の1軸のカテゴリー数量平均値を算出して影響の低い語を明確化するため正規化した。その結果0に近い共通文末語を切り捨てて27語に絞り込んだ。

カテゴリー数量平均値は、個別のウェイト平均値 (hw) の正負は判断できない。そこで5紙うち3紙以上がカテゴリー増加に対してマイナスであれば hw にマイナスをつけてこの値を共通文末語のウェイトとした。

一社説の総ウェイト (W) は式-2で表される。

$$W = \sum_{i=0}^n hw(i) \quad (\text{式-2})$$

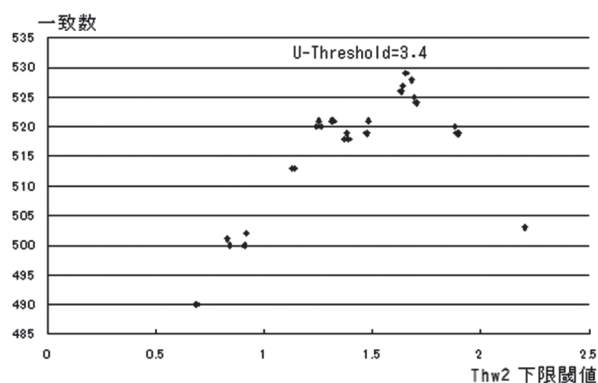
ここで n は、対象社説に使用された共通文末語数-1であり、 $hw(i)$ は、検出された文末語のウェイトである。

3.2.4 判別項目の閾値の設定

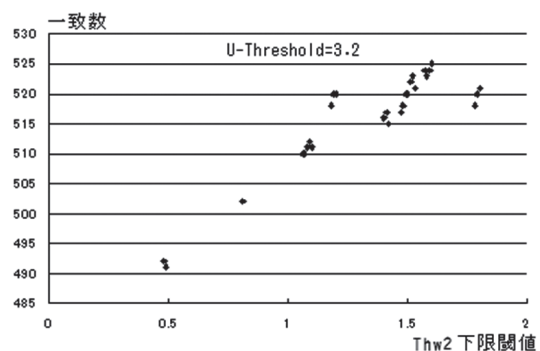
計算された総ウェイト (W) はそのままでは否定的か肯定的か判断することはできない。評価者の判定項目と一致させるためには W が式-1にしたがって3項目のひとつに合致させる必要がある。

そこで評価者が判定した項目に W を合わせるため、 $Thw1$ と $Thw2$ の値を変えて評価者の判定結果と最大的一致数を出した値を3項目の閾値とした。使用したサンプルは2022年6月～2024年5月までの5紙合計760部である。

$Thw1$ の値は0.1きざみで0.51～3.5の範囲で、また $Thw2$ は-0.41から1.65の範囲で変化させた。そして評価者が判定した結果との一致数を計算して一致数最大の $Thw1$ と $Thw2$ を求めた。閾値 $Thw1$ 閾値 $Thw2$ と一致数の関係を図9～10に示す。



〔図-9〕 $Thw1$ を3.4に固定した W の一致数



〔図-10〕 $Thw1$ を3.2に固定した W の一致数

図-9より、最大的一致条件は $Thw1$ と $Thw2$ の値はおおよそ3.4と1.65であった。そして、そのときの評価者の判定結果との一致率は69.6%であった。

4 結論

既存メディアの情報が、ネット情報にあっさり覆される事態がこれほど顕著に示された事例が2024年の東京と兵庫の知事選であった。このニュースはSNSの情報拡散力と浸透力のパワーが、人材と

資金で圧倒する既存メディアを凌ぐ怖さも見せつけた。現代は情報化社会と言われるが情報化社会の課題が明確化された現象であった。

ネット情報は容易に手に入り、理解するハードルも低い。しかもネットユーザは先有傾向に沿った情報にアクセスする傾向があるならば、それに否定的な情報は受け入れがたい^[15]。それゆえ、一旦ユーザに取り込まれたネット情報はバイアスがかかって拡散しやすい。またフェイク情報も、スタートでは明らかな詐称的表現がみられなくても、拡散しながら影響力を増大させていく怖さがある。

この状況の対応には、ネットユーザが精査された情報も取り込む環境が必要である。それはさまざまな制約を比較的受けにくいメディアの情報であり、たとえば新聞、週刊誌や学術誌などである。しかしそれらは紙媒体が主で、表現や内容がネットよりも難解である。しかも有料であるためネットユーザに「コスパ」が合わない^[12]。

本稿では、オーソライズされた情報を電子媒体で比較的容易に取得できるプログラム開発を試みた。対象とするコンテンツは大手新聞社の社説である。社説は各新聞社の意見や主張を明確に記述しており、結果への責任体制もある。そして法的制約がテレビより低いため、主要記事について各社各様の多彩な主張が記述されている。これは、ネット情報に依存しがちなユーザには、フェイク情報への効果的なワクチンとなる。得られた結果を以下に示す。

- ①社説プログラムの検索機能は、タイトル、刊行社名、刊行年月日、内容の評価項目など多様である。
- ②評価項目は、否定的・肯定的・提案型の3種に分類されているので、意見の異なる複数の新聞を比較することができる。
- ③自動評価計算は文末語の使用頻度で決定したが、評価者との一致率は69.6%であった。

提案した検索プログラムはまだプロトタイプであり、ユーザインタフェースも不十分である。また、評価項目の評価者との一致率は低く実用的には改善が必要である。

今後は、これらの不具合を是正して使いやすい検索プログラムに改良していく予定である。

参考文献

- [1] “石丸伸二氏、165万票で2位 蓮舫氏3位 東京都知事選の結果確定”、毎日新聞、2024/7/8
- [2] “兵庫県知事選制した斉藤さん、バンザイなし”、読売新聞オンライン、2024/11/18
- [3] “低知名度の石丸伸二氏躍進 「国政当然考える。例えば衆院広島1区」”、毎日新聞、2024/7/7
- [4] “知事を告発した職員を「死をもって抗議」に追い込んだ兵庫県の「懲戒」 公益通報者を守れぬ保護法の実態”、東京新聞オンライン、2024/4/14
- [5] “公選法違反で刑事告発 石丸伸二 都知事選文書改ざんを暴く”、週刊文春、2025/2/13
- [6] “なぜ斉藤知事とPR会社が告発されたのかSNS運用と知事側の説明”、朝日新聞、2024/2/7
- [7] 中島岳志、“兵庫県知事とネット社会 極端に走る群集心理加速”、東京新聞、2024/11/26
- [8] 水島宏明、“劣化目立つテレビの選挙報道 周辺ばかり 中身薄く”、東京新聞、2022/7/28
- [9] “政治的公平の現在地 ネット「原則なし」極端化 米国「公平廃止」社会が分断”、東京新聞、2023/3/8
- [10] “政治的中立どう捉える”、朝日新聞、2023/6/24
- [11] “放送法は番組守る盾”、東京新聞、2024/7/17
- [12] “マスメディア 持続可能性を探る”、東京新聞、2023/8/13
- [13] “ネット世論の歪み 一部の声が多数派に リスク認識を”、朝日新聞、2023/9/15
- [14] 石井真人、“文末語を用いた社説内容の判別”、メディア情報研究 10、p77-84、2024
- [15] 稲増一憲、“マスメディアとは何か「影響力の正体」”、中公新書、p51-53

