研究指導 中澤 真 教授

多様な Web メディアに対応可能な汎用的特徴量を用いたデマ判別システム
―拡張モダリティを利用して―

山本 晃大

1. はじめに

インターネットの一般化に伴い，フェイクニュースなどの不確かな情報に触れる機会は増えている．みずほ情報総研による調査[1]ではインターネット上でフェイクニュースを見かける頻度が「週1回以上」と回答したのは26.1%にも上る．フェイクニュースのうち，デマは誤情報をもとに危険な行為を促す事例[2]もあり，特に対策が必要である．この対策の1つにファクトチェックがあるが，デマの発信と検証結果が周知のものとなるまでにタイムラグが生じ，その間にデマが拡散してしまう恐れがある．

この問題に対して機械的にデマを自動判別する研究が行われている[3][4][5]．しかしこれらの研究の多くはSNSなどのメディアに対象を絞ることで，そのメディア固有の特徴を用いて判別することが多く，多様なメディアへの対応という点で汎用性が低い．みずほ情報総研による調査[1]にて，過去一年間に「ホームページやブログの閲覧・書き込み」をしたと回答したのは63.8%，「まとめサイト」では31.6%という結果になるなどSNSが普及した現在でもインターネットの利用者におけるSNS以外のWebサイトの利用は健在である．そのため個人ブログやまとめサイトから誤情報が波及するケースも多く存在する[6][7]．そのため，これらにも対応した汎用性の高いデマ判別システムが必要である．

それには各メディア固有の特徴量を用いず，Webメディア共通の特徴から判別する必要がある．これを実現した事例として，松田ら[4]は文章に含まれるモダリティを用いて判別を可能にした．しかしこのシステムは人手によってモダリティのラベル付けをすることが前提であり，自動判別には至っていない．

そこで本研究ではモダリティのラベル付けを自動化し，デマを含むか否かの判別を，手動でラベル付けするものと同程度の精度で実現することを目指す．

2. デマ判別システムの現状と課題

2.1 デマへの対策の現状

IPA[1]が発表する「情報セキュリティ10大脅威」の個人分野で2024年「ネット上の誹謗・中傷・デマ」が9年連続で選ばれる[8]など，インターネット上の不確かな情報が情報セキュリティ脅威として近年認識されつつある．特にデマは誤情報をもとに危険な行為を促す事例[2]もあり，拡散の防止が必要である．これらへの対策としてファクトチェックの推進，リテラシー教育，プラットフォーム運営企業による凍結措置，法整備などが実施されているが，それぞれ問題点を抱えており，抜本的な解決には至っていない．特にファクトチェックではデマの発信から検証結果の公表までに時間差が生じ，その間に誤情報が広がる可能性があるため，より迅速かつ効果的な手法の開発が求められている．

2.2 デマ判別システムの現状

デマ発信から検証結果の公表までのタイムラグが少ないデマへの対策として，アルゴリズムを用いて自動的にデマを判別して閲覧者に通知する手法が挙げられる．しかし現状のデマ判別システムの研究では，研究対象を特定のWebメディア固有の特徴量に依存する事例が多く[2][3]，多様なメディアに対応できる汎用性の高いシステムの開発は少ない[4][5]．この理由として，そのメディア固有の特徴量を活用することが判別精度の向上に直結することが挙げられる．例えば，X[2]のリポスト数やリプライ数のような指標は，そのメディア内での情報の拡散パターンやユーザの反応を具体的に反映しており，デマか非デマかを判断する上で重要な手がかりとなる．一方でこれらの特徴量を用いた判別システムは他のメディアへの応用が難しくなり，汎用性は失われてしまう．

SNSが普及した現在でもブログやまとめサイトの利用は健在である[1]ため，これらのサイトから誤情報が波及したケース[4][7]もある．そのため，多様なメディアに対応可能なデマ判別システムが必要である．

汎用性の高いデマ判別システムを開発するにはメディア固有の特徴量ではなく，共通の特徴量から判別する必要がある．

2.3 汎用性の高い特徴量

メディアの特性に依存しないデマ判別のための特徴量を得る方法として，文章自体に含まれる情報を利用する方法が挙げられる．文章は客体的な要素と主体的な要素とで構成され，主体的要素の1つにモダリティがある[9]．モダリティは「～かもしれない」や「～だろう」のような話し手の判断・発話態度を表す文法的カテゴリである[10]ため，デマ判別の特徴量として利用した研究もある．

松田ら(2022)[11]はデマを含む文章とそれ以外の文章において，モダリティの現れ方に差があるかどうかを検証した．その結果デマを含む文章にはモダリティ分類別の出現頻度，出現順序に差があることを明らかにし，判別基準として応用できる可能性を示した．

松田(2023)[4]は，さらに判別基準として応用できる可能性のある特徴を示した上で，これらの特徴を用いて

[1] 独立行政法人 情報処理推進機構

[2] https://twitter.com

機械学習による判別実験をした．その結果，出現順序の学習により，性能向上が図れることを示した．ただしこのシステムではモダリティのラベル付けを手動で行っており，完全な自動化はなされていない．本研究は先行研究の判別精度を維持しつつ，自動的なモダリティのラベル付けを実現し，そのラベルを特徴量としたデマ判別システムの開発を目指す．

3. デマ判別システムの構成

3.1 使用するツール

本研究のシステムは，Python3を用いて開発した．HTMLから本文を抽出する外部ライブラリとしてExtractContent3[3]，spaCy[4]，emoji[5]の3つのモジュールを用いた．ブログなどのWebコンテンツからテキストを抽出する場合，HTMLタグを除いただけではおすすめの記事などの関連情報も出力されてしまうため，本文のみを抽出できるExtractContent3を採用した．なお，本文とみなす閾値であるスコアは50に設定している．自然言語処理のためのライブラリであるspaCyは，文境界解析に用いた．spaCyの言語モデルのうち，句点に頼らない文境界分析が可能なものには自然言語処理ライブラリGiNZA[6]のja-ginza-electaraを使用した．絵文字に関する種々の処理のためのライブラリであるemojiは絵文字削除のために用いた．またモダリティの解析には拡張モダリティ解析器Zunda[7]を使用した．

3.2 拡張モダリティ解析器 Zunda

Zundaの動作としては，本研究の設定ではその内部でMeCab[8]，CaboCha[9]を呼び出し，形態素解析と係り受け解析をした上で処理をする．この結果をもとに主体的要素が現れる部分であるイベントを見つけ出し，拡張モダリティの情報(表 2)を付与する．これらのZundaの付与する拡張モダリティは，先行研究[4][11]のモダリティとは異なる．

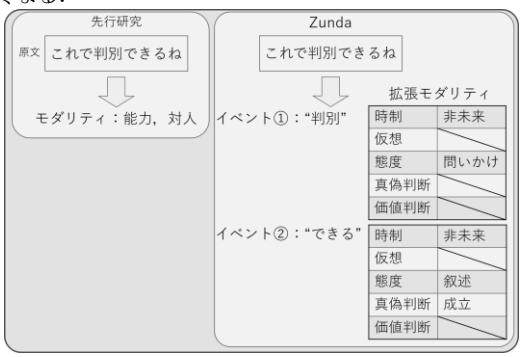

図 1 モダリティと拡張モダリティの違いの例

先行研究が定義するモダリティは「話者の知覚や感情といった心理的態度を表現する部分」[11]である．対して拡張モダリティはより広い概念であり，主体的な要素を統合した情報である．それに伴い先行研究とZundaでは情報の付与手法が異なる．

先行研究の情報付与手法は各文に対し表 1の10のラベルから1つ以上のラベルが付与されるものである．対してZundaは1つのイベントに対して，表 2左列の5つの項目に右列のタグが1つずつ付与される．例えば，「これで判別できるね」という原文をラベル付けすると図1のようになる．このように情報の付与手法が異なるため，先行研究の手法をそのまま適用することはできない．

表 1 松田ら(2022)[11]の分類([11]より筆者作成)

分類	ラベル
認知的モダリティ	断言，推測
束縛的モダリティ	強制，誘導
力動的モダリティ	意思，能力
証拠的モダリティ	経験，伝聞
対人的モダリティ	対人
モダリティなし	なし

表 2 Zunda がイベントに対し付与する情報[12]

項目	タグ
時制	未来，非未来
仮想	条件，帰結
態度	叙述，意志，欲求，働きかけ-直接，働きかけ-間接，働きかけ-勧誘，問いかけ
真偽判断	成立，不成立，不成立から成立，成立から不成立，高確率，低確率，低確率から高確率，高確率から低確率
価値判断	ポジティブ，ネガティブ

また項目のうち，時制と態度はイベントに対し必ず付与されるが，図 1イベント①のように仮想，真偽判断，価値判断は付与されない場合もある．

3.3 システムの構成概要

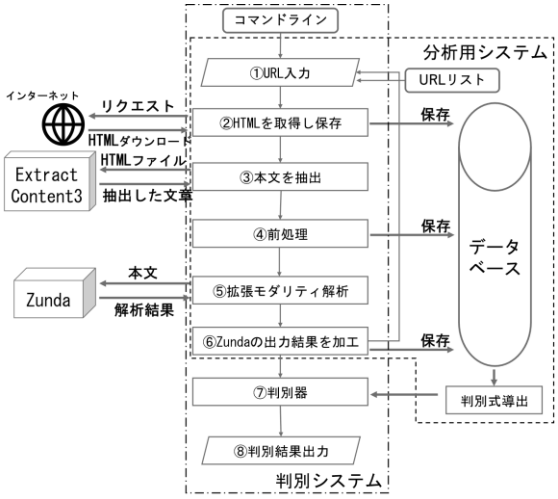

図 2 システム構成図

システムの構成を図 2に示す．URLの入力をうけ，そのWebページを解析し，判別分析でデマか否かを出力する．

まず①ではURLを取得する．分析用システムではURLを記述したリストから入力する．このURLに該当す

[3] https://github.com/kanjirz50/python-extractcontent3
[4] https://spacy.io/
[5] https://github.com/najeira/emoji-python
[6] https://megagonlabs.github.io/ginza/
[7] https://jmizuno.github.io/zunda/
[8] http://taku910.github.io/mecab/
[9] https://taku910.github.io/cabocha/

るHTMLをインターネットから取得してデータベースに保存する(②). ダウンロードに10秒以上かかる場合は判別システムではタイムアウトしエラーを出力, 分析システムではそのURLはスキップする. 次にExtractContent3を呼び出し, HTMLを解析することでタグと関連情報を取り除いて本文を抽出する(③).

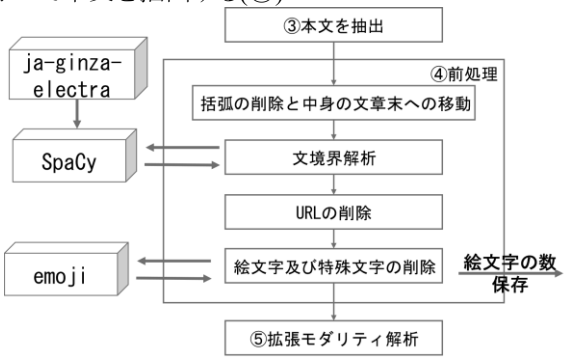

図 3 前処理(④)の概要

④では前処理として, 図 3の手順にしたがって本文を加工する. Zundaは1文ごとに改行された状態での入力を前提としているため, 取得したデータに文単位で改行コードを追加する処理が必要になる. 一般的に文の境界には句点があるため, 改行位置が決定できる. しかしブログやSNSの記事では句点のない文章も多く存在するため, spaCyを利用して句点に頼らない文境界の解析が必要になる. 続いてZundaの誤認識を引き起こすURL, 丸括弧, 特殊文字を削除する. ただし, 丸括弧の内に含まれる文は処理対象となるため, これらは文章末に移動させる. また同じく処理の対象とならない絵文字を, emojiを用いて削除する. ただし, 絵文字はモダリティと同じく話者の態度を示す部分であり, その量はWebページの著者のそのWebページに対する態度・丁寧さなどに影響を受けると考えたため, 特徴量としてその個数をカウントし, データベースに保存する.

次に前処理された本文をZundaで解析する(⑤). Zundaは1つのイベントに対して表 2に示した各項目のタグのどれを付与すべきか判断し最大5つのタグを付与する. 各項目で付与すべきタグがない場合は, その項目については付与されない. このタグの種類別の出現頻度を説明変数として用いる.

⑥ではWebページごとの説明変数としてイベント数あたりのタグの構成頻度をタグごとに以下の式のように求める. まずZundaの出力結果から21種類のタグが, それぞれいくつのイベントに付与されたかカウントし, その文章のイベント数あたりの付与されたタグ数を1種類ずつ求める. 説明変数を, 出現頻度をイベント数で除したものとするのはWebページの本文が長いほど, タグを付与するイベントが増えるため, 出現頻度が大きくなるからである. これを標準化するためイベント数で割る.

タグiの候補となるのは21個のタグである. 本環境では「働きかけ-勧誘」「低確率から高確率」「高確率から低確率」の3つのタグは付与が確認できなかったためこれらを除く. 時制のタグ「未来」「非未来」はどちらか一方が必ず付与されるため独立でなく, 多重共線性が生じるのを避けるため「未来」のみを用いる. したがって「働きかけ-勧誘」「低確率から高確率」「高確率から低確率」「非未来」の4つのタグを除き, 拡張モダリティによる説明変数はX_1~X_{17}の17個とする. また④で保存した絵文字の個数も同様にイベント数で割り, 18番目の説明変数X_{18}とする. よって説明変数18個で構成された18次元ベクトルをもとに判別する.

$$(Webサイトjの持つ特徴量) = (X_1, X_2 \cdots X_i, \cdots X_{18})$$

$$説明変数X_i = \frac{(タグiの出現頻度)}{(Webサイトに含まれるイベント数)}$$

判別システムでは⑥の結果を当該URLに対する文章の説明変数の観測値としてデータベースに保存する. URLリストに未処理のものがあれば, ①から繰り返し, すべて処理済みであれば判別式を導出する. 判別システムでは⑦に進み, ⑥の結果を判別器に入力しその結果が正であればデマ, そうでなければ非デマと判別する.

4. 判別精度の検証

4.1 デマの定義

大辞林[16] はデマに「1.根拠・確証のないうわさ話」, 「2.政治的効果をねらって, 意図的に流される虚偽の情報」という2つの語釈をつけている. 他の辞書[17][18][19]でも共通して同様の語釈である.

大辞林は現代語としての一般的な語釈が1つ目となるため, 本研究では1つ目を採用し, 根拠が不十分なうわさ話をデマとして扱う.

4.2 サンプル収集

システムの精度検証のためデマをサンプルとして収集した. 収集したサンプルは主にブログやニュースサイトである. ブログはジャンルが幅広く, 投稿数が十分に多いAmebaブログ[10]を利用した.

表 3 サイト分類別サンプルの数

サイト分類	デマ/非デマ	記事数
Amebaブログ(Ⅰ)	デマ	71
Amebaブログ(Ⅱ)	非デマ	80
Ⅰの参考文献	デマ	5
Yahoo!ニュース	非デマ	3
ファクトチェック済みサイト	デマ	16

デマが指摘[14][15]されていた「ワクチン」「環境問題」というキーワードでAmebaブログの検索機能を用いて対象記事を収集し, それぞれの記事に対して手動でデマと非デマとのラベル付けをした. さらに収集したブログ記事内でリンク先のデマを含むWebページについても収集し, 同じトピックについて扱ったYahoo!ニュースの記事を, 内容を確認したうえで非デマとして収集した. 加えて対応可能なWebメディアを増やすためファクトチェック専門メディア「リトマス」[11]に掲載されたもののうち, 「誤り」, 「不正確」, 「ミスリード」, 「根拠不十分」とされた, 原文が確認できたWebページをデマとして収集した. 表 3

[10] https://ameblo.jp/

[11] https://litmus-factcheck.jp/

の記事数は収集したWebページのうち，本システムで処理できたものである．

これらを分析用システムで分析し判別式を導出する．

4.3 デマ判別の結果

収集したサンプルすべてで判別式を作成し，判別精度を検証した．これをメディアごとに検証したものを表 4 に示す．Webページに対してデマと判別したもののうち実際にデマだったページの割合が適合率であり，実際にデマ内容が記載されたWebページの集合に対して，それをデマと正しく判別できた割合が再現率である．

表 4 全体と各サイト分類における判別精度

全体	正解率	0.674
	適合率	0.667
	再現率	0.761
Amebaブログ（Ⅰ，Ⅱ）	正解率	0.660
	適合率	0.619
	再現率	0.732
Ⅰの参考文献 と Yahoo!ニュース記事	正解率	0.667
	適合率	0.625
	再現率	1.000
ファクトチェック済みサイト	再現率	0.813

4.4 考察

ここでは先行研究[4]との比較を行う．先行研究では時系列学習を用いた判別と構成頻度を用いた判別の2つを行っており，その正解率は前者が0.71，後者は0.52~0.46となっている．なお本研究と先行研究では検証に用いたサンプルに違いがあり，注意が必要である．

本研究の全体での判別精度は0.67となっており，先行研究の出現頻度のみを学習させたモデルより精度が高くなった．これは拡張モダリティが一般的なモダリティより広い概念であり，含む情報が多かったためであると考えられる．一方で出現順序を学習させたモデルの精度にはわずかに劣っている．本研究では出現頻度のみを学習させたが，拡張モダリティの出現順序も情報として利用すれば判別精度が向上する可能性がある．

4.5 システムの課題

また収集したファクトチェック済みのサイトのうち2割程度が処理できないなど本システムで処理できないWebサイトが複数みられた．これはクローラの作りこみの問題，本文抽出の失敗，Zundaがイベントはないと判断するなどが要因として挙げられる．今後システムとしての完成度を向上させるにはこの部分の改善が必要である．

5. むすび

本研究では拡張モダリティを特徴量としたデマ判別システムを開発した．その結果，判別精度は67%で先行研究の構成頻度を特徴量とした判別方式と比べ高い結果となったため，拡張モダリティがデマ判別の特徴量となりうることを示した．今後この判別手法を改善することで，デマと指摘されていない未知のデマについても自動的に判別することができるようになる可能性がある．

一方で先行研究の時系列情報を学習させたモデルよりも判別精度が低い結果となった．今後，この情報の学習によって精度向上が図ることができる可能性がある．これを実現することが今後の課題である．

参考文献

[1] みずほ情報総研株式会社，“日本におけるフェイクニュースの実態等に関する調査研究”，総務省，https://www.soumu.go.jp/main_content/000715293.pdf，2020-03(2023-02-03).

[2] NEWS AGENCIES，“Iran: Over 700 dead after drinking alcohol to cure coronavirus”， ALJAZEERA，https://www.aljazeera.com/news/2020/4/27/iran-over-700-dead-after-drinking-alcohol-to-cure-coronavirus，2020-04-27（2023-06-25）.

[3] 渡邉研斗ほか，“Twitter 上での誤情報と訂正情報の自動分類”，言語処理学会第 19 回年次大会発表論文集 (2013 年 3 月)，pp.178-181,2013.

[4] 松田美慧，”言語学的モダリティに基づくデマの検知に関する研究”， 2022 年度 ISS Square シンポジウム (2023/3/3)，No227，2023.

[5] 柳裕太ほか，“画像付きフェイクニュースとジョークニュースの検出・分類に向けた機械学習モデルの検討”，研究報告知能システム（ICS），pp1-8，2019.

[6] 読売新聞，“まとめサイトに「NATO軍、日本に駐屯を検討」の偽情報…ネット掲示板から転載”，https://www.yomiuri.co.jp/national/20231209-OYT1T50027/，2023-12-09（2023-12-12）.

[7] 篠原修司，“「トヨタ社長豊田章男氏がワクチンは人口削減のための遅効性の毒と発言した」との個人ブログのデマが広がる“，https://news.yahoo.co.jp/expert/articles/96b53f891e74a7467133366e5b23970418897f7d，2022-8-2（2023-12-12）.

[8] 独立行政法人情報処理推進機構，“プレス発表「情報セキュリティ 10 大脅威 2024」を決定”，https://www.ipa.go.jp/pressrelease/2023/press20240124.html，2024-01-24(2024-01-25).

[9] 鈴木重幸，日本語文法・形態論，むぎ書房，1972.

[10] 青木博史，高山善行，日本語文法史キーワード事典，ひつじ書房，2020.

[11] 松田美慧ほか， ”モダリティと感情影響に注目したデマの構文解析”，コンピュータセキュリティシンポジウム 2022 論文集，pp.753-758，2022.

[12] 本多正幸，“判別分析における多重共線性問題に対する分析安定化手法の総合的研究”，https://kaken.nii.ac.jp/grant/KAKENHI-PROJECT-05680247/，1993(2024-02-07).

[13] 江口萌ほか，“モダリティ、真偽情報、価値情報を統合した拡張モダリティ解析”，言語処理学会第 16 回年次大会論文集，pp.852-855，2010.

[14] 中村塁，片岡裕，“新型コロナワクチンに関する誤情報・デマへの取り組みまとめ”，Yahoo!Japan，https://about.yahoo.co.jp/topics/20210804.html，2021-08-04(2023-12-25).

[15] 読売新聞，“ツイッターでデマ急増、マスク氏買収後「野放し」”，https://www.yomiuri.co.jp/economy/20230202-OYT1T50258/，2023-02-03(2023-12-25).

[16] 松村明，大辞林，三省堂，1988.

[17] 日本大辞典刊行会，日本国語大辞典第十四巻，小学館，1975.

[18] 新村出，広辞苑，岩波書店，2018.

[19] 北原保雄，明鏡国語辞典，大修館書店，2020.