

受容性を考慮した多様な意見の検索手法

伊藤 優希[†] 馬 強^{††}

[†] 京都大学工学部情報学科 〒 606-8501 京都府京都市左京区吉田本町

^{††} 京都大学大学院情報学研究科 〒 606-8501 京都府京都市左京区吉田本町 36-1

E-mail: [†]y.ito@db.soc.i.kyoto-u.ac.jp, ^{††}qiang@i.kyoto-u.ac.jp

あらまし フィルターバブル現象が問題視される昨今、ユーザに提供する情報の多様性が重要である。情報収集には多くの場合に検索システムが使用されるが、従来の検索手法の多くは情報の違いのみに着目している。情報の違いに注目した検索は、ユーザに多様な視点をもたらすという意味では有用だ。しかし同時に、攻撃的・批判的表現を含む受容できない情報も提示してしまう可能性がある。そこで本研究では、ユーザの興味に合致するツイートやレビュー文をクエリとした、受容性を考慮しつつ多様な情報を提示する手法を提案する。提案手法では、各文章に含まれるアスペクトとその極性の対の系列でモデル化し、その意見の特徴を形式的に表す。そして言及するアスペクトや極性の同異から、クエリ文に共感を示しつつ、異なる意見も述べているツイートやレビュー文を発見する。本手法は、ユーザにとって受け入れやすい多様な意見を提示することを可能とし、フィルターバブル軽減効果が期待される。

キーワード テキストマイニング、情報検索・推薦、情報信ぴょう性、意見分析、フィルターバブル

1 はじめに

インターネット上に溢れる情報量は日々増加している。今やその全ての情報を個人が閲覧し取捨選択するのは不可能に近い。そこで注目すべき情報を簡単に見つけるための方法として、個人適応型情報推薦手法が研究開発されてきた。これは個人の選好に合わせて、情報表示の優先順位を決めることを可能にするものである。しかしこうした手法を用いると、各個人の選好に過剰にフィットした推薦が行われてしまう場合がある。これにより人々は、自分の興味に合わない情報に接触する機会を喪失するという問題に直面した。この現象は、フィルターバブルとして知られている[16]。フィルターバブルの問題が指摘されてから約10年経った今でも、身近な検索システムやSNSでフィルターバブルにつながる現象が確認されている。例えばTwitterでは、気候変動懐疑論やCOVID-19などの話題で、同じ意見の者同士集まってエコーチェンバーを形成していることが確認された[7][10]。

これまで研究されてきた主なフィルターバブル軽減手法は、個人の選好に合わない情報・意見も幅広く提示することである。例えば[19]では、対話生成システムにおいてユーザの発話内容に対する文脈上自然な反対意見を提示する手法が提案された。このような手法は自分と異なる意見に出会う機会を提供する。しかし選好と異なる意見の中には、提示すると逆効果になる可能性があるものが存在する。例えばある人の価値観を攻撃するような内容を含んだ意見は、その人にとって受け入れ難い。攻撃的な意見に対して不快感を抱くことは、その意見を無視したり、この意見には信頼性がないと決めつけたり、自分の考え方と一致する意見ばかり探したりといった行動につながる[1]。つまり自分の価値観や信念を強化してしまい、フィルターバブル軽減には逆効果となる懸念がある。

従来の検索で得られる反対意見	本研究で目標とする推薦
最近電気が高騰してるの反原発派のせいだろ。ふざけんなよ。	震災で原発の危険性は十分思い知らされた。けれど技術が進んで、今の原発はあの頃より安全になってる。そろそろ再稼働もアリだと思う。
原発再稼働反対！危険すぎる！環境にやさしい太陽光や風力で対処すべき	原発は火力に比べると環境にいいし、再エネより発電量も多い。色々な意見はあると思うが、個人的には再稼働賛成。
原発の代わりに太陽光で〜とか、発電能力的に絶対無理。いまだに太陽光推してる人いるの？w	原発再稼働には私も不安を感じる。でも電力不足のまま冬が始まると、凍死者が出る可能性もある。再稼働も仕方ないのかな。
原発を再稼働しない限り、いつまで経っても日本の電力不足は解消されないだろう	

図1: 本研究で達成したい目標のイメージ図

この問題を緩和するために本研究では、文章がユーザにとって受け入れやすいか、すなわち受容性に着目した情報検索手法を提案し、ユーザの意見に共感を示しつつ、ユーザとは異なる多様な意見も提示できるようにする。

例えば図1は、原発再稼働の賛否というトピックに関して否定的意見を持っているユーザ(図の左端)に対して異なる意見、すなわち原発再稼働に肯定的な意見を提示するという状況を表している。従来通り、意見の違いだけを考慮して推薦すると、表の左列のように、ユーザにとって受け入れづらい攻撃的な内容も表示されてしまう。一方、表の右列の文章は、ユーザの意見と異なる内容を述べているが比較的受容性が高いと考えられる。よってフィルターバブル軽減という視点では、表の右列のような文章を提示する方が、左列の文章を提示するより望ましい。

提案手法ではツイートやレビュー文などの短文から、受容性が高く、かつ多様な視点をユーザに提供する文章を発見する。本手法はまず、利用者の選好に近い意見を述べた文章(クエリ文)と、クエリ文と同じトピックについて述べた文章(候補文章)を受け取り、これらを文章表現モデルに変換する。文章表現モデルとは、文章に現れる意見の特徴を抽出して形式的に表したものである。

次に各候補文章の受容性・多様性を以下の基準で評価する。

- 受容性：利用者が受け止めやすい文章か？
 - － 共感度, 客観度で判断 (いずれも 3 節で説明する).
- 多様性：利用者の意見や視野とは異なる内容の文章か？
 - － 反対度, 多様度で判断 (いずれも 3 節で説明する).

次に, 受容性と多様性の評価結果を表す数値を特徴量として, 候補文章がユーザにとって受け入れやすく, かつ多様な視点を提供しているかを 2 値分類する. 最後に分類結果や, 文章の長さなどの基準を考慮して候補文章をランキングする.

本研究の貢献を以下にまとめる.

- 文章表現モデルを提案している (3.1 節).
- ユーザにとって受容的かつ多様な視点を提供可能な意見を発見するための評価尺度として, 共感度, 客観度, 反対度と多様度を提唱し, その推定手法を提案する (3.2 節).
- 提案手法と比較手法の分類性能・ランキング性能の評価をユーザ実験で行い, 提案手法の有効性を確認した (4 章).

2 関連研究

フィルターバブル軽減方法として, 今までにもさまざまな手法が提案されている. それらの手法は大きく 2 つに分けられる. 利用者の行動変容を促すアプローチと, 検索や推薦などの尺度に着目するアプローチである.

2.1 利用者の行動変容を促すアプローチ

利用者にフィルターバブルを軽減する行動を促すには, 提示する情報の可視化が有効とされている. Linder はソーシャルニュースサイト Reddit の各 subreddit (特定のトピックについてのコミュニティ) 間の関連をグラフ構造化し, それを都市を模した VR 空間 Blue Link City で可視化した. Blue Link City では各トピックの文脈や背景が複数の視点から見えるようになることで, フィルターバブル軽減が期待される [11]. また, ユーザがニュース記事に注釈を書き込めるようにすることで, 多様な視点に触れる機会を確保するニュース閲覧サービス Newsr も構築された [21]. ステークホルダに注目する方法も研究された. [14] はニュース記事内に出現するステークホルダの関係, 評価をグラフ化して他のニュース記事と比較させることで, ニュース記事自体の偏向を明らかにした.

情報の偏りを可視化する以外のアプローチも存在する. 例えば異なる意見に触れるまでのプロセスを容易にすることは, 利用者の行動を変化させるのに有効である. ニュース記事閲覧アプリケーションの NewsSalad は, 利用者が閲覧中のニュースの関連記事のうち, 内容が最も異なる記事へのアクセスをワンタップで行うことを可能にしている. このアプリケーションは, 利用者の意識を幅広い情報へのアクセスの重要性に向けるという点で有用だった [8]. また Google 検索の検索クエリ提案部にユーザの価値観と異なる意見につながるクエリを優先的に表示することで, ユーザに多様な視点の意見を自発的に探してもらう試みも行われた [26]. Google 検索では他に, ユーザに提示されている情報が偏っている場合に通知することでユーザ自身にフィ

ルターバブル解消を促すインターフェースも提案された [24].

また, メタ認知能力を向上させるアプローチも考えられる. メタ認知とは, 自身の認知を客観的に認知することである. この能力を向上させると, 人々は自らの認知の偏りを認識し異なる視点の意見を探すようになった [12].

2.2 推薦や検索の尺度に着目するアプローチ

利用者の行動を変化させるのではなく, 情報推薦アルゴリズムそのものを工夫するアプローチも研究されてきた. 例えば, あるセッションでユーザに提示されている情報との類似度から, 次の情報を推薦する Session-based Recommender System に, 多様性を表す重みを追加した session-based k nearest neighbor method は 従来よりも多様な情報を推薦することに成功した [3]. 元のトピックとの関連性を保ちつつも異なる視点の内容を含む情報を推薦する方法として, VAE 強調フィルタリング [2] や, sequential rule mining と session-based k-nearest neighbor method の組み合わせ手法 [4] も提案された. 多様性以外の尺度に着目した研究もある [15]. この研究で Pardos らはセレンディピティに注目し, 大学のコース推薦システムの出力を目新しいコースが従来より多く含まれるように改善した. また SNS ユーザコミュニティ間の類似性に基づいて検索結果を再ランキングしてフィルターバブルを弱める community-aware model (CAM) アプローチ [5] も提案されている.

ここまで紹介してきた全ての研究は, フィルターバブルを軽減する手法という点で本研究と共通している. しかしいずれも受容性を考慮した推薦手法ではない. また, そのような手法は著者らの知る限りは存在しない.

3 提案手法

本節では, あるトピックについて検索した際, ユーザにとって受容性が高く, かつ異なる視点をユーザに提供する文章を優先的に表示する手法を提案する. 手法の全体図を図 2 に示す.

提案手法は以下のような入出力を想定する.

- 入力：クエリ文と候補文章集合
 - － クエリ文：ある意見を述べた短文
 - － 候補文章集合：クエリ文と同じ話題に対する意見の集合
- 出力：受容性・多様性を考慮して候補文章をランキングした結果

クエリ文はユーザの選好を表す意見である. つまり図 1 の吹き出し内の文章に相当する. 提案手法ではクエリ文と各候補文章を 3.1 節で定義する文章表現モデルに変換する. 次に以下の 4 指標で候補文章集合中の各文を評価し, それがクエリ文と同じ考えの人にとって受容的かつ多様な視点をもたらす意見 (以後, 受容的かつ多様な意見と呼ぶ) になっているかどうかを分類する. そしてその結果を元に, 受容的かつ多様な意見のランキングを行う. ランキング方法は 3.3 節で詳述する.

3.1 文章表現モデル M_S

文章 S の文章表現モデル M_S を以下のように定義する.

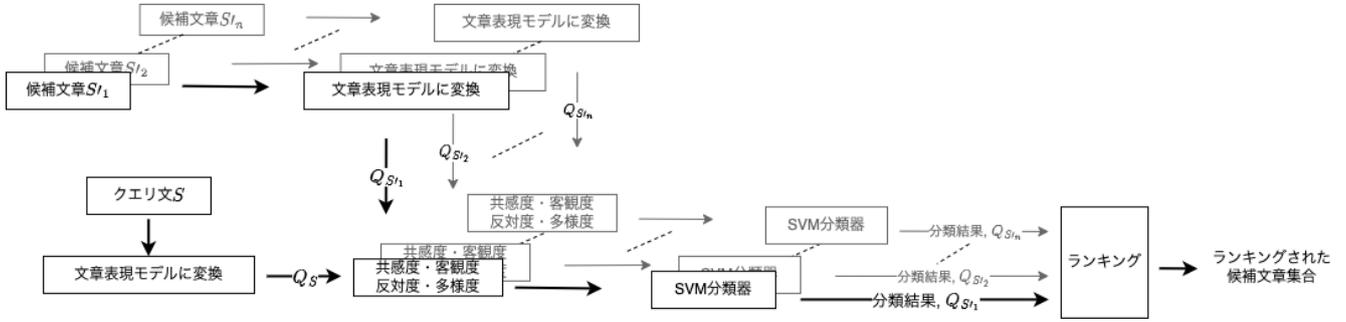


図 2: 提案手法の全体図

$$M_S = (P_S, wps), O_S \quad (1)$$

$$((a_1, p_1, aps_1), (a_2, p_2, aps_2), \dots, (a_m, p_m, aps_m))$$

モデル中の P_S は文章の極性を表す. 文章の極性 (polarity) とは, 文章内容が肯定的か否定的かといった感情を表すラベルである. 例えば「このカメラは画質が悪い」という文章の極性は negative(否定的)である. 極性と組になっている wps は, 極性予測 P_S がどの程度の確率で正しいかを示す数値である.

O_S は, S がどの程度客観的に述べられているかを数値で表したものである.

a_i は文中に出現するアスペクトを表す. アスペクトとはあるトピック内でよく議論される観点のことである. 例えば「このカメラのデザインは良いが画質が悪い」という文章は, カメラというトピックについて意見を述べている. この文中で述べられているアスペクトは, デザインと画質である. カメラに関するアスペクトには他にも, 重量, 値段などがある. さらに各アスペクトに対する極性も定義する. これは文章の書き手が, その文中に出現するアスペクトに対して表明している意見のことである. カメラの例だと「デザイン」の極性は positive(肯定的), 「画質」の極性は negative(否定的)となる. (a_i, p_i, aps_i) は, 文中に出現するアスペクト a_i とその極性 p_i , そして極性予測 p_i がどの程度の確率で正しいかを表す数値 aps_i の組である. 文章表現モデル M_S にはこれらの組の系列が含まれる.

3.1.1 文章全体極性 (P_S, wps)

文章全体極性 (P_S, wps) は, ファインチューニング済みの感情分析手法 [22] に S を入力することで求める. 本研究では感情分析手法の出力を, negative, positive に対応する 2 値ベクトルと設定した. 2 次元ベクトルのうち, negative に対応する数値が positive に対応する数値より大きい場合, P_S は negative とする. 逆の場合は positive とする. さらに P_S に対応する値を wps (= whole polarity score) とする.

3.1.2 アスペクト極性系列

アスペクトと極性の系列 $((a_1, p_1, aps_1), \dots, (a_m, p_m, aps_m))$ は, 文中に出現するアスペクトを抽出し, そのアスペクトの出現位置周辺の文章極性をアスペクトの極性とみなすことで求める. 具体的には以下の手順である.

(1) S を文末と否定の接続詞・接続助詞の位置で区切り, 部分文章の系列 (S_1, S_2, \dots, S_m) を生成する.

(2) 各部分文章 S_i について以下の手順を行う.

表 1: 手法で用いた否定の接続詞・接続助詞一覧

否定の接続詞	しかし, ところが, でも, けれど, だが, が, なのに, だけれど, だけれども, だけど, けれども, けど, それでも, それなのに, しかしながら, ですが
否定の接続助詞	でも, でも, ところで, が, けれど, けれども, けど, のに, ながら, つつ, もの

(2-1) S_i 中の名詞を全て抜き出し, 名詞の系列 $(a_{i,1}, a_{i,2}, \dots, a_{i,n_i})$ を作る.

(2-2) S_i の極性 p_{S_i} とそのスコア aps_{S_i} を求める.

(2-3) 抜き出した名詞をアスペクトとし, 部分文章の極性とスコアをそれらのアスペクトの極性とみなす. つまり系列 $((a_{i,1}, p_{S_i}, aps_{S_i}), \dots, (a_{i,n_i}, p_{S_i}, aps_{S_i}))$ を求める.

(3) 各 S_i について求めた (アスペクト, 極性, スコア) の系列を結合し, $(a_1, p_1, aps_1), \dots, (a_m, p_m, aps_m)$ を得る.

手順 (1) では, 極性の切り替わる可能性の高い位置で文章を分割する. 具体的には, 句読点と否定の接続詞・接続助詞 (表 1) を手掛かり表現として極性の境目を探し, その位置で分割する. 否定の接続詞・接続助詞は形態素解析ツール MeCab [20] を用いて特定する.

3.2 候補文章の評価

本節では式 (1) の文章表現モデルを使って, 候補文章 S' がクエリ文 S と同じ考えのユーザにとって受け入れやすいか (受容性) と, 新しい視点をもたらすか内容か (多様性) を分析する. 受容性と多様性を測る尺度は以下のように定義する.

- 共感度: クエリ文に共感する表現が入っているか
- 客観度: 客観的な口調であるか
- 反対度: クエリ文と異なる意見であるか
- 多様度: 多様な視点に言及しているか

このうち共感度と客観度は受容性を測り, 反対度と多様度は多様性を測る. これらの尺度の詳しい定義, 計算方法を以下に示す. ただし, 尺度の計算時に行う単語分割や品詞の判定には, 形態素解析ツール MeCab [20] を利用する. また文章や単語の埋め込み表現には, 感情分析手法 [22] 中の BERT モデルに文章や単語を入力して得られる最終層出力のうち, 単語全体の特徴に対応するベクトルを用いる. また S, S' の文章表現モデル $M_S, M_{S'}$ は以下のようにしているとする.

$$M_S = (P_S, wps), O_S,$$

$$((a_1, p_1, aps_1), (a_2, p_2, aps_2), \dots (a_n, p_n, aps_n))$$

$$M_{S'} = (P_{S'}, wps'), O_{S'}$$

$$((a'_1, p'_1, aps'_1), (a'_2, p'_2, aps'_2), \dots (a'_m, p'_m, aps'_m))$$

3.2.1 共感度

共感的な表現が含まれる意見であれば、受け取り手の価値観とは異なる内容であったとしても受け入れられる可能性が高い [1], [6], [18]. 厚生労働省によると、共感とは相手の気持ちを理解しようとする姿勢である [25]. これに基づき本研究では、共感的な表現を「情報の受け取り手の考えを理解し、寄り添う態度を示す表現」と定義する。共感度は文章 S がこの定義を満たす内容を含みつつ、異なる意見も述べているかを、以下の 3 点に着目して評価する。

- 意見変化条件： S に意見の変わり目が含まれているか。例えばユーザーの意見が「このカメラは画質が悪い」である場合、「たしかに画質は悪いが、価格はお手頃だ」は共感的表現を含む反対意見である。この文章は接続助詞「が」の前で意見が分かれており、前半ではカメラに否定的な意見を、後半ではカメラに肯定的な意見を述べている。「が」のように文章の極性が切り替わる位置を、意見の変わり目と呼ぶ。文中での共感の示し方は多々あるが、上述の例のように、共感的な表現を文章の前半に置き、後半で異なる意見を述べるという構成がよく使われる。意見変化条件は、文章がこのような構造を含むかを判断する。

- 意見順条件：意見の変わり目より前は、ユーザーの意見と同意見か。

意見順条件は、項目 1 を満たす文章が本当にユーザーの意見に共感しているかを判断する指標である。例えば「たしかに価格はお手頃だが、画質は悪い」という文章は意見の変わり目を含む。しかし意見の前半は、カメラに対して肯定的な意見を述べている。よってこの文章の前半部分は、「このカメラは画質が悪い」という意見に真に共感的とは言えない。

- 話題類似度： S の共感的表現が入力と似た話題について述べているか。

話題類似度はユーザーの意見との話題の共通度を測る。「たしかに画質は悪いが、価格はお手頃だ」「たしかにデザインはダサいが、価格はお手頃だ」という文章は意見変化条件、意見順条件を共に満たす。しかしユーザーの意見「このカメラは画質が悪い」はデザインについては触れていない。よって前者の方が、後者よりもユーザーの意見に強く共感していると言える。

共感度の計算方法は以下の Algorithm 1 の通りである。ただしアルゴリズムの入力は M_S , $M_{S'}$ から取り出した以下の値である。

$$input_polarity = (P_S, wps)$$

$$input_aspects = ((a_1, p_1, aps_1), \dots (a_n, p_n, aps_n))$$

$$target_aspects = ((a'_1, p'_1, aps'_1), \dots (a'_m, p'_m, aps'_m))$$

アルゴリズム中に出現する各関数について説明する。

8 行目の SeparateAspects は、極性が切り替わる位置でア

Algorithm 1 共感度の計算

```

1: function EMPATHIC_DEGREE(input_polarity, input_aspects,
   target_aspects)
2:     ▷ アスペクトが含まれない文章が入力された場合
3:     if  $n == 0$  then
4:         return 0.5
5:     else if  $m == 0$  then
6:         return 0
7:     end if
   ▷ アスペクトが含まれている場合
8:     sep_target_asps = SeparateAspects(target_aspects)
9:     for each  $sep_i \in sep\_target\_asps$  do
10:        sep_hmeans  $\leftarrow$  emptylist
11:        for each  $(sepa_{i,j}, seppi, sepsi_{i,j}) \in sep_i$  do
12:            is_empathic_phrase =  $(seppi == P_S)$  &
    $(seppi_{i+1} = P_S)$ 
13:            asp_hmeans  $\leftarrow$  emptylist
14:            for each  $(a_l, p_l, aps_l) \in input\_aspects$  do
15:                pol_sim = PolSim( $(p_l, aps_l)$ ,  $(seppi, sepsi_{i,j})$ )
16:                word_sim = CosSim( $sepa_{i,j}$ ,  $a_l$ )
17:                asp_hmeans.add(Hmean(pol_sim, word_sim))
18:            end for
19:        end for
20:        sep_hmeans.add(max(asp_hmeans))
21:    end for
22:    return mean(sep_hmeans)
23: end function

```

スペクト系列を分割する関数である。例えば、アスペクトの系列

$((\text{画質}, \text{negative}, 0.8), (\text{デザイン}, \text{negative}, 0.9),$
 $(\text{重量}, \text{positive}, 0.7), (\text{価格}, \text{positive}, 0.8), (\text{性能},$
 $\text{negative}, 0.8))$

が与えられたとする。系列の前から二つのアスペクト (画質, デザイン) の極性は negative である。しかし 3 つ目のアスペクトは positive である。よってここを極性が切り替わる位置とみなし、系列を分割する。同様に価格と性能の間でも分割を行う。この手順を繰り返すと最終的に、以下の 3 つに分割された部分系列集合が得られる。

$((\text{画質}, \text{negative}, 0.8), (\text{デザイン}, \text{negative}, 0.9)),$
 $((\text{重量}, \text{positive}, 0.7), (\text{価格}, \text{positive}, 0.8)),$
 $((\text{性能}, \text{negative}, 0.8))$

SeparateAspects の出力は上記の部分系列集合である。これは部分系列のリスト $[sep_1, \dots, sep_{sep_num}]$ で表される。ただし、

$sep_i =$
 $((sepa_{i,1}, seppi, sepsi_{i,1}), \dots, (sepa_{i,m_i}, seppi, sepsi_{i,m_i}))$

アルゴリズム中の bool 変数 is_empathic_phrase は、共感度の 3 つの評価指標のうち意見変化条件に対応する。これは For 文で抽出した部分系列 sep_i が、 S' 中の意見の変わり目の直前の部分文章に属するものかどうかを示す。この変数は、一度入力に共

感じてからそのあとに入力と反対の意見を述べているフレーズを見つけるために使用される。例えばクエリ文「このカメラは画質が悪い」に対して、「このカメラは画質が良くないとは聞いていたが、使ってみると予想以上に画質が悪い」のような候補文章があったとする。これは文章の前半と後半が接続助詞「が」で別れているものの、前半も後半もカメラについて negative なことしか述べていない。 $sepp_{i+1}$ を考慮することによって、このような表現が共感表現として抽出されることを防ぐ。

15 行目の PolSim は意見順条件を満たす度合いを計算する関数である。候補文章とクエリ文のアスペクトとその予測スコアである、 (a'_i, p'_i, aps'_i) と (a_j, p_j, aps_j) から以下のように求める。

a' が共感的表現内の場合

$$\begin{cases} \frac{aps'_i + aps_j}{2} & (p'_i == p_j) \\ \frac{(1 - aps'_i) + aps_j}{10} & (p'_i \neq p_j) \end{cases}$$

a' が共感的表現外の場合

$$\begin{cases} \frac{aps'_i + aps_j}{2 * 2} & (p'_i == p_j) \\ \frac{(1 - aps'_i) + aps_j}{10 * 2} & (p'_i \neq p_j) \end{cases}$$

このスコアは、 a' がクエリ文中のいずれかの a に共感しているかどうかを測る指標なので、極性が一致していない場合はスコアを低くしている。また共感表現中にある a' に点数が高く入るようにしている。なお式中の定数には、チューニングを行い、最終的なランキング結果が最も良かった数値を採用した。

16 行目の CosSim は話題類似度を計算する関数である。候補文章とクエリ文のアスペクトからそれぞれ一つずつ選び、それらの単語埋め込み表現のコサイン類似度を返す。

3.2.2 客観度

客観的な事実に言及している意見は、受け手とは反対の意見であっても比較的受け入れやすい。客観度を評価することによって、そのような文章の発見を目指す。

主観的な意見は動詞、形容詞、副詞が含まれている割合が高い傾向にある [13]。よって本研究では、それらの割合が低い文章を客観的とみなす。客観度 $O_{S'}$ は以下のように定義される。

$$O_{S'} = 1 - \frac{\text{文章中の動詞} \cdot \text{形容詞} \cdot \text{副詞の個数}}{\text{文章中の全単語数}}$$

3.2.3 反対度

反対意見を述べる文章中は、ユーザが触れたことのない情報や視点を含む可能性が高いため、ユーザに多様な視点を提供するのに有用である。反対度を評価することによって、文中にユーザの意見とは反対の内容が述べられているかを測る。

反対度は (P_S, wps) と $(P_{S'}, wps')$ から求める。

$$P_S \text{ と } P_{S'} \text{ が同じ場合 反対度} = 1 - wps'$$

$$P_S \text{ と } P_{S'} \text{ が異なる場合 反対度} = wps'$$

3.2.4 多様度

多様度を評価することで、多彩な視点や情報が盛り込まれた文章を発見することを目指す。例えばトピック A について話し

Algorithm 2 ランキングの生成

```

1: function RANKING( $M_S, S_{correct}, models$ )
2:   for each  $i = 1$  to  $|S_{correct}|$  do
3:     text =  $S_{correct}[i]$ 
4:     if  $P_{text} \neq P_S$  then
5:        $\triangleright P$  は極性 ( $M_S$  や  $models$  から取得)
6:       ranking1.add(text)
7:     else
8:       ranking2.add(text)
9:   end if
10: end for
11: return Sort(ranking1) + Sort(ranking2)
12: end function

```

ている 2 つの文章「A には B という問題があるから反対である」「A には B という問題や C という問題があることが知られているが、D や E といった長所もある」について、前者は B という観点しか述べていないが、後者は B, C, D, E とより多くの観点到に触れている。後者のような文章は、様々な意見を閲覧者に提供すると考えられる。

多様度の計算には、文章表現モデル $M_{S'}$ から取り出したアスペクトと極性の系列 $\{(a'_1, p'_1, aps'_1), \dots, (a'_n, p'_n, aps'_n)\}$ を用いる。まずはアスペクトを BERT モデルに入力することで、アスペクトの埋め込み表現 emb'_i を得る。

$$emb'_i = BERT(a'_i)$$

次に n 個のアスペクトの埋め込み表現のうち、任意の 2 つのコサイン類似度を全て求める。求めた nC_2 個のコサイン類似度の平均を多様度 $div_{S'}$ とする。

$$div_{S'} = \frac{1}{n} \sum_{k=1}^n \sum_{j=1}^{k-1} \frac{emb'_k \cdot emb'_j}{|emb'_k| \times |emb'_j|}$$

3.3 ランキング

ここまでで求めた 4 種類のスコアをもとに候補文章集合を分類し、ランキングする。分類には SVM 分類器を用いる。分類器の入出力は以下である。

- 入力：反対度, 共感度, 客観度, 多様度
- 出力：入力を元に、文章を受容的かつ多様な視点をもたらずか否かで 2 値分類した結果

分類器の詳しい学習方法は 4.2.2 節で説明する。

ランキングは Algorithm2 のように求める。ただしランキングアルゴリズムに与えられる入力のうち、 M_S はクエリ文 S の文章表現モデルである。また $S_{correct}$ は候補文章のうち、分類器によって受容的かつ多様な意見と分類された文章の集合を表すリスト $[S'_1, \dots, S'_n]$ 、 $models$ は $S_{correct}$ に含まれる文章の文章表現モデルのリスト $[M_{S'_1}, \dots, M_{S'_n}]$ である。アルゴリズム中で、受容的かつ多様な意見かどうかの正例ラベルは *correct* と表記している。

アルゴリズム中に出現する関数 Sort は、引数のリストに含まれる各文章を文章表現モデル生成時 (3.1.2 節) と同様に文末・接続詞・接続助詞で分割したとき、いくつかの部分文章に分割さ

れるか求め、その降順に文章を並び替える。このとき部分文章数が同じ候補文章が複数あれば、より長い文章を上位とする。

4 実験

4.1 実験の概要

実験では、提案手法が候補文章を受容的かつ多様な意見か否かに分類する精度を求める。また、分類結果から生成されるランキングの性能も求める。そして比較手法との精度比較を行い、提案手法を評価する。比較手法には、ニューラルネットを用いて受容的かつ多様な意見を発見する手法(4.2.3節)を用いる。

4.2 学習

評価前に、まずは文章表現モデルを生成するための感情分析手法、そして分類に用いる SVM、ニューラルネットの学習を行う。

4.2.1 感情分析手法の学習

感情分析手法 [22] の学習には Twitter 日本語評判分析データセット [23] を用いた。このデータセットは Twitter 上の投稿をトピックごとに収集し、

- (1) 肯定的な意見を述べている
- (2) 否定的な意見を述べている
- (3) 肯定的な意見も否定的な意見も述べている
- (4) 肯定的な意見も否定的な意見も述べていない

といった基準で評価したものである。評価は各文章が基準に該当するか否かの 2 択で行われている。本研究ではこのデータセットを、再ラベリングして利用した。まず (1) に該当する文章を positive, (2) に該当する文章を negative とした。また (3) に該当する文章には、その文中で書き手が最も主張したい意見と一致するラベルを付与した。(4) に該当する文章は使用しなかった。なお学習時間の都合から、このデータセットのうち 1000 余の文章を negative, positive の比率が 1:1 になるよう選んで使用した。

4.2.2 提案手法の分類器の学習

SVM 分類器の学習にも Twitter 日本語評判分析データセット [23] を利用した。データ作成の際にはまず、クエリ文と候補文章をデータセット内からランダムに選び、提案手法で候補文章の共感性、客観度、反対度、多様度を求めた。そしてこれらの 4 つのスコアと、候補文章がクエリ文に対して受容的かつ多様な意見になっているかどうかを表す教師ラベルをデータセットとして保存した。ただし、Twitter 日本語評判分析データセット [23] で (1) と (2) 両方に該当する、または (3) に該当するとラベル付けされていた文章を正例とした。負例と正例の比率は 1:1 とした。

SVM の実装には scikit-learn [17] ライブラリの LinearSVC クラスを使用した。

4.2.3 ニューラルネットの学習

比較手法として、ニューラルネットを使って提案手法の処理の流れを再現した手法を用意する。本手法ではまずクエリ文と候補文章を、提案手法の文章表現モデルをもとにした形式表現に変換する。モデル形式を変更したのは、提案手法の文章表現

表 2: 5 つの候補文章集合の詳細。レビュー文は The Multilingual Amazon Reviews Corpus [9] から取得。

トピック	情報源
安倍元総理の国葬実施についての賛否	Twitter(2022/7/10)
給食での黙食強制に関する賛否	Twitter(2022/12/3)
上野千津子氏の祝辞(東大入学式)の賛否	Twitter(2019/4/15)
原発再稼働の賛否	Twitter(2021/12/1)
スマホ画面保護フィルムのレビュー	[9] のコーパス

モデルをニューラルネットに入力可能な形式にするためである。次にそれらの形式表現を反対度、共感性、客観度、多様度に対応するネットワークに入力する。そうして得られたベクトルを元に、候補文章が受容的かつ多様な視点を提供している意見かどうかを 2 値分類する。最後に、各候補文章のランキングスコアを提案手法のランキング手法(3.3節)に則って決定する。

ニューラルネットの学習にも Twitter 日本語評判分析データセット [23] を再ラベリングしたデータを用いた。データ作成の際にはまず、クエリ文と候補文章をデータセット内からランダムに選び、それらをニューラルネットに入力可能な文章表現に変換した。そしてこれらの 2 つの文章表現と、候補文章がクエリ文に対して受容的かつ多様な意見になっているかどうかを表す教師ラベルをデータセットとして保存した。正例とみなす基準、データセット内のデータの比率は 4.2.2 節と同様である。

4.3 データセット

次に、実験時に使用する候補文章集合とクエリ文を準備した。候補文章集合は Twitter や Amazon レビューコーパスから 5 つのトピックについて収集した。それぞれの候補文章集合に含まれる文章数は 50 文である。詳細は表 2 に示している。また 5 つの候補文章集合それぞれに対して、2 種類ずつクエリ文を準備した。2 つのクエリ文は各候補文章集合のトピックに対して肯定的意見を示す文章を 1 つ、否定的意見を示す文章を 1 つ選んで決めた。このようにして、候補文章集合とクエリ文の組を計 10 種類用意した。

さらに各候補文章について、分類の正解ラベル、ランキングの正解適合度を求めるために、ユーザ実験を行った。ユーザ実験では収集した各候補文章が、クエリ文と同じ意見を持つ人にとってどのように受け取られるかを評価した。この過程で、各候補文章は最低 3 人に評価された。評価項目は以下である。

- 共感性: クエリ文の考えに寄り添っているか
- 客観度: 客観的な口調で述べられているか
- 反対度: クエリ文と異なる意見を述べているか
- 多様性: 幅広い観点から根拠や意見を述べているか
- 受容性: クエリ文と同じ考えの人にとって、受け入れやすい意見か
- 多様性: クエリ文と同じ考えの人に、今まで見えていなかった意見・視点を提供しているか

これらを 5 段階で評価した結果をもとに、各文章がクエリ文に対して真に受容的かつ多様な意見かどうかを表す分類ラベルの正解を決定した。その際、受容性と多様性の平均が 3 より大き

表 3: 精度指標を求める式に出現する用語の定義 (混合行列)

	予測値が正例	予測値が負例
真のラベルが正例	真陽性	偽陰性
真のラベルが負例	偽陽性	真陰性

い文章を正例とみなした。また受容性と多様性の平均を、各候補文章の適合度とした。

4.4 精度評価と考察

4.4.1 評価手法

本研究では、提案手法と比較手法のアブレーション研究を行った。各手法は共感性、客観度、反対度、多様度を求めるための個別の要素から構成されている。よって提案手法では

- prop-emp: 共感性だけを利用する手法
- prop-obj: 客観度だけを利用する手法
- prop-opp: 反対度だけを利用する手法
- prop-div: 多様度だけを利用する手法
- prop-all: 4 指標全てを利用する手法

の 5 種類、比較手法では

- comp-emp: 共感性ネットワークだけを利用する手法
- comp-empopp: 共感性・反対意見ネットワークを利用する手法

- comp-empoppobj: 共感性・反対意見・客観性ネットワークを利用する方法

- comp-all: 4 指標のネットワーク全てを利用する手法
- の 4 種類の手法を評価する。それぞれの手法の評価は、以下の手順で行った。

(1) クエリ文と候補文章を評価対象手法に入力して候補文章の 2 値分類結果 (受容的かつ多様な意見か否か) を得て、分類の正解率、適合率、再現率、F1 スコアを求める。受容的かつ多様な意見であると分類された候補文章の集合を $S_{correct}$ とする。

(2) $S_{correct}$ 中の文章の順位と適合度 (後述) を求め、nDCG を計算する。

(3) 1 から 2 の手順を 10 クエリ分繰り返し、正解率、適合率、再現率、F1 スコア、nDCG の平均を求める。

評価に用いられた精度指標は以下のように計算した。

$$\text{正解率} = \frac{\text{真陽性} + \text{真陰性}}{\text{候補文章総数 (50 文)}}$$

$$\text{適合率} = \frac{\text{真陽性}}{\text{真陽性} + \text{偽陽性}}$$

$$\text{再現率} = \frac{\text{真陽性}}{\text{真陽性} + \text{真陰性}}$$

$$\text{F1 スコア} = \frac{2}{\frac{1}{\text{適合率}} + \frac{1}{\text{再現率}}}$$

$$\text{nDCG} = \frac{\sum_{k=1}^{|S_{correct}|} \frac{\text{評価対象手法で求めた適合度}_k - 1}{\log_2(k+1)}}{\sum_{k=1}^{|S_{correct}|} \frac{\text{ユーザ実験で求めた正解の適合度}_k - 1}{\log_2(k+1)}}$$

式中の用語の定義を表 3 に示す。

提案手法や比較手法では、各ランキング対象文章の適合度を、推定された極性ラベル (*negative*, *positive*) と部分文章数 (3.3

表 4: 分類・ランキング性能の評価結果表

評価対象	正解率	適合率	再現率	F1	nDCG
prop-emp	0.466	0.78	0.166	0.267	0.956
prop-obj	0.57	0.619	0.824	0.695	0.923
prop-opp	0.46	0.571	0.451	0.494	0.917
prop-div	0.618	0.784	0.534	0.618	0.939
prop-all	0.644	0.715	0.698	0.695	0.927
comp-emp	0.43	0.829	0.009	0.157	0.977
comp-empopp	0.42	0.632	0.101	0.17	0.845
comp-empoppobj	0.436	0.618	0.116	0.193	0.961
comp-all	0.532	0.737	0.361	0.478	0.930

節) を基準とした 5 段階評価で推定する。極性がクエリ文と異なるかつ、部分文章数が 5 以上の場合は適合度 5、部分文章数が 4 以下の場合は適合度 4、極性がクエリ文と同じなら適合度 3 とする。適合度が 3 以上 5 以下の整数で表されるのは、4.3 節で正解の分類ラベルを計算した際、受容性と多様性の評価値平均が 3 より大きいものを正例とみなしたことに対応している。

提案手法、比較手法それぞれの分類精度、ランキング精度の平均はそれぞれ表 4 のようになった。

4.4.2 結果と考察

実験結果から以下のことが分かった。

- 提案手法は比較手法と比べて、ランキング性能を損なわずに高い分類性能を達成している。

- 4 つの指標のうちいくつかを選んで使う場合より、4 指標全て用いた場合の方が分類やランキングの妥当性が高い。

- 提案手法で発見を目指した受容的かつ多様な意見は、実際にユーザにとって受容性が高く新たな視点を提供可能である。

結論の 1 つ目は prop-all と comp-all を比較することで確認できる。分類性能の評価値である正解率・適合率・再現率・F1 スコアは適合率以外、prop-all の方が優れている。このことから、提案手法は受容的かつ多様な意見をより正確に判別していると言える。また nDCG の値からランキング性能にも大きな差がないことがわかった。

各手法の nDCG を比べると、comp-emp の値が最も高くなっている。これは comp-emp が 50 件中 1 から 3 件程度しか正例と判定しておらず、他の手法と比べるとそもそもランキング対象になる文章が極端に少ないことが原因である。これは再現率や F1 スコアが低いことから確認できる。この他にも提案・比較手法の両方において、部分的に指標を利用した手法の方が、4 指標を利用する手法よりも nDCG 値が高いことがある。しかしいずれの場合も、F1 スコアが 4 指標を利用した手法より低い。これらのことから、指標を部分的に利用する場合、nDCG は高いがそもそもランキング対象として抽出された文章が少ない、という現象が起こりやすいと言える。よって 4 つの指標全てを利用する方が、より質の高い分類・ランキングを行うことができる。以上が結論の 2 つ目である。

提案手法内で prop-all と他の手法を比べると、適合率は prop-div、再現率は prop-obj の手法に劣るものの、prop-all はどの評価指標に対してもバランスよく高い性能を達成していることが

読み取れる。しかし prop-all は、特にランキング性能において、prop-emp などの単独指標の手法に劣る傾向がある。これには先ほど述べた原因の他に、提案手法内のハイパーパラメータやスコア算出アルゴリズムに原因がある可能性も考えられる。評価に使用した候補文章集合は各 50 文しか含んでおらず、データが少なすぎる可能性もある。パラメータチューニングや追加のデータ収集などを行なって、これらを改善することを今後の課題としたい。

最後に結論の 3 つ目について、prop-all などの nDCG が 0.9 を超えていることから、本手法で発見することを目指した文章は、受容性と多様性を兼ね備えていると結論づけられる。

5 おわりに

本研究では意見やレビューの文章表現モデルと、ユーザにとって受け入れやすく、尚且つ多様な視点をユーザにもたらす意見の発見手法を提案している。提案手法はまず、入力された文章から文章全体の極性、アスペクトごとの極性の系列などを取り出すことで、意見の特徴を表した文章表現モデルを生成する。そして作成したモデルをもとに、受容性を評価する「共感度」「客観度」と、多様性を評価する「反対度」「多様度」を用いて候補文章をランキングする。

また比較手法も構築し、提案手法との性能比較を行った。実験結果から、両手法とも受容的かつ多様な意見の発見を容易にすること、提案手法が比較手法に主に分類性能で勝ることが確認された。フィルターバブル軽減を狙って選好と異なる意見を表示する際に提案手法を応用すれば、ユーザにとって受け入れやすい文章を優先的に表示でき、異なる意見を無意識に排斥してしまう現象が起りにくくなると期待される。今後の課題としては、手法のパラメータ調整や評価済み候補文章の追加収集などが挙げられる。

謝 辞

本研究の一部は科研費（19H04116）による。

文 献

- [1] Ecker, et al. The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, Vol. 1, No. 1, pp. 13–29, 2022.
- [2] Gao, et al. Mitigating the filter bubble while maintaining relevance: Targeted diversification with vae-based recommender systems. pp. 2524–2531, 2022.
- [3] Gharahighehi, et al. Making session-based news recommenders diversity-aware. In *Proceedings of the Workshop on Online Misinformation-and Harm-Aware Recommender Systems*, pp. 60–66. CEUR Workshop Proceedings, 2020.
- [4] Gharahighehi, et al. Personalizing diversity versus accuracy in session-based recommender systems. *SN Computer Science*, Vol. 2, No. 1, p. 39, 2021.
- [5] Grossetti, et al. Reducing the filter bubble effect on twitter by considering communities for recommendations. *International Journal of Web Information Systems*, Vol. 17, No. 6, pp. 728–752, 2021.
- [6] Hyland-Wood, et al. Toward effective government commu-

- nication strategies in the era of covid-19. *Humanities and Social Sciences Communications*, Vol. 8, No. 1, p. 30, 2021.
- [7] Julie Jiang, et al. Social media polarization and echo chambers in the context of COVID-19: Case study. *JMIRx Med*, Vol. 2, No. 3, p. e29570, aug 2021.
- [8] Kiritoshi Keisuke, et al. A diversity-seeking mobile news app based on difference analysis of news articles. In *Database and Expert Systems Applications: 26th International Conference, DEXA 2015, Valencia, Spain, September 1-4, 2015, Proceedings, Part II 26*, pp. 73–81. Springer, 2015.
- [9] Phillip Keung, et al. The multilingual amazon reviews corpus. *arXiv preprint arXiv:2010.02573*, 2020.
- [10] Kolic, et al. Quantifying the structure of controversial discussions with unsupervised methods: a look into the twitter climate change conversation. 2022.
- [11] Linder, et al. Pop the feed filter bubble: Making reddit social media a vr cityscape. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 619–620, 2018.
- [12] Grace Liu. Moving up the ladder of source assessment: Expanding the craap test with critical thinking and metacognition. *College & Research Libraries News*, Vol. 82, No. 2, p. 75, 2021.
- [13] Yin Mingyong, et al. A subjective expressions extracting method for social opinion mining. *Discrete Dynamics in Nature and Society*, Vol. 2020, p. 2784826, 2020.
- [14] Tatsuya Ogawa, et al. News bias analysis based on stakeholder mining. Vol. 94-D, pp. 578–586, 2011.
- [15] Pardos, et al. Combating the filter bubble: Designing for serendipity in a university course recommendation system. 2019.
- [16] Eli Pariser. *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin, 2011.
- [17] Fabian Pedregosa, et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, Vol. 12, pp. 2825–2830, 2011.
- [18] Luiza A. Santos, et al. Belief in the utility of cross-partisan empathy reduces partisan animosity and facilitates political persuasion. *Psychological Science*, Vol. 33, No. 9, pp. 1557–1573, 2022. PMID: 36041234.
- [19] Yoshida Satoshi, et al. Generating dialogue sentences to promote critical thinking. In *Database and Expert Systems Applications: 31st International Conference, DEXA 2020, Bratislava, Slovakia, September 14–17, 2020, Proceedings, Part I 31*, pp. 354–368. Springer, 2020.
- [20] Kudo Taku, et al. Applying conditional random fields to Japanese morphological analysis. pp. 230–237, July 2004.
- [21] Wood, et al. Rethinking engagement with online news through social and visual co-annotation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–12, New York, NY, USA, 2018. Association for Computing Machinery.
- [22] Yang, et al. Pyabsa: Open framework for aspect-based sentiment analysis. 2022.
- [23] Suzuki Yu. Filtering method for twitter streaming data using human-in-the-loop machine learning. *Journal of Information Processing*, Vol. 27, pp. 404–410, 2019.
- [24] 片岡雅裕ほか. フィルターバブルを気づかせるシステムの提案. 人工知能学会全国大会論文集, Vol. JSAI2015, pp. 1H21–1H21, 2015.
- [25] 厚生労働省. 共感:用語解説|こころの耳:働く人のメンタルヘルス・ポータルサイト. <https://kokoro.mhlw.go.jp/glossaries/word-1541>. Accessed: 2022-12-28.
- [26] 今田隆人ほか. フィルターバブルを回避するナッジの考案. 人工知能学会全国大会論文集, Vol. JSAI2021, pp. 3H4GS11a01–3H4GS11a01, 2021.