

ソーシャルメディアにおける投稿およびユーザの政治的傾向予測と政治的投稿フィルタによる性能向上

佐橋優人¹ 狩野芳伸¹

¹ 静岡大学

¹{ysahashi, kano}@kanolab.net

概要

ソーシャルメディア投稿の政治的傾向を予測することを目的とし、政治的投稿であるか検出するモデルを事前に構築したうえで、データを政治的投稿かどうかでフィルタしてから、学習・評価する方法を提案する。異なるソーシャルメディアプラットフォームの投稿を用いた検証では、フィルタを用いる提案手法がフィルタなしの評価結果を大幅に上回り、国会議員投稿におけるアカウントの与野党分類では最大で F1 値が 13 ポイント向上した。

1 はじめに

ソーシャルメディアネットワーク (SNS) 上では日常的に政治的な意見が発信・共有されている [1][2]。

ユーザの政治的傾向を予測する際、政党などの政治団体や政治家に比べ、一個人の一般的なユーザは政治的ではない投稿が大半を占めるため、区別なく投稿を学習すると、特にデータ量が少ない場合に予測性能を下げる可能性がある。そこで、政治的傾向を学習する前に政治的な投稿に絞り込んだうえで、モデルのファインチューンを行う。また、投稿単位での予測と、投稿単位の予測の集合から計算したアカウント単位の予測精度を検証する。

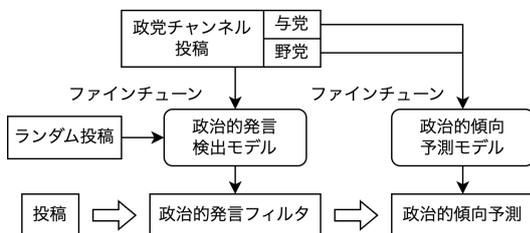


図1 提案手法の流れ

SNS の代表例として Twitter(現 X) があり、政治家に限らず一般的なユーザも活発にテキストの投稿を行っている。以前は Twitter API により容易に取得が可能であったが、現在では大規模なデータの取得が

難しい。そのため、Twitter 以外の API による収集が容易な SNS 上の投稿をもとに学習し、Twitter データで検証することで、データ数不足を解消する。

ファインチューンする事前学習モデルについて、一般的な事前学習モデルと、大規模投稿データで事前学習したモデルをそれぞれ利用し比較した。

2 関連研究

JTweetRoBERTa[3] は、RoBERTa[4] をベースモデルとし、日本語の約 6,000 万件の大規模投稿で事前学習したモデルである。一般的な書き言葉ベースのモデルとは異なり、SNS 上の表現に特化している。

Ali Rahmati ら [5] は、イランの Twitter ユーザの政治的傾向を分類するため、政府への支持者、反対者を分類する BERT や LSTM のモデルを構築、比較した。政治的発言を含む約 12 万投稿を収集、ラベル付けを行った。投稿単位の分類は F1 スコアで 75.68%、アカウント単位の分類は F1 スコアで 93.18% を達成した。

陳ら [6] は、多言語対応の BERT モデルである LaBSE をもとに、政治家の Twitter における投稿を左翼から右翼の軸上に投影し、米国の民主党と共和党の明確な分極化を示した。単語レベルのモデルより優れた性能を達成した。

3 提案手法

本研究では、SNS 上のユーザ投稿文から政治的な傾向の予測を行う。まず、一般的な投稿であるか政治的な投稿であるかの二値分類を事前学習済みモデルに対しファインチューンし、政治的投稿検出モデルを作成する。次に後段処理のモデルとして、与野党のどちら寄りか投稿であるかの二値分類をするよう事前学習済みモデルをファインチューンし、政治的傾向予測モデルを作成する。

政治的傾向の強いデータで学習した方が分類性能

が良いという仮説を検証するため、実験では、政治的投稿検出モデルによって異なる閾値でフィルタしたものをファインチューンの学習データとし、比較する。また、評価データも異なる閾値でフィルタし比較する。提案手法の流れは図 1 の通りである。

4 実験

4.1 データの収集と前処理

本研究で使用したツイートデータは、Twitter API V2¹⁾の Academic Research ならびに X API V2 の Basic を用いて収集した。また、コメントデータは、YouTube Data API v3²⁾を用いて収集した。すべての投稿文に共通する前処理として、Unicode 正規化、メンションや URL の削除、10 文字以下の投稿の削除を行った。収集するデータの種類、条件は以下の通りである。

ランダム投稿 日本語ひらがな五十音を投稿文に含む投稿を他に制限を設けず収集するランダムサンプリングにより収集したツイート。

政党チャンネル投稿 自民党、立憲民主党、公明党、日本維新の会、国民民主党、日本共産党、れいわ新選組、社会民主党、参政党の公式 YouTube チャンネルのすべての動画のコメント。

政党フォロワー投稿 自民党、立憲民主党、公明党、日本維新の会、国民民主党、日本共産党、れいわ新選組、社会民主党、参政党の公式 Twitter アカウントをフォローしているユーザの 2022 年から 2023 年までの投稿の一部を収集したツイート。

国会議員投稿 2021 年と 2024 年の参議院選挙、2022 年の参議院選挙によって当選した国会議員のうち、Twitter アカウントが確認できた 582 名の 2019 年から 2023 年までの投稿の一部を収集したツイート。

4.1.1 政治的傾向予測用のデータセット

政党チャンネルコメントデータを自民党、公明党の動画に対するコメントには与党ラベル、他政党の動画に対するコメントには野党ラベルを付与した。野党投稿をアンダーサンプリングして件数を合わせ、それぞれ件数は 24,027 件である。

政党フォロワー投稿データについて、自民党または公明党をフォローしている場合は与党ラベル、他

政党をフォローしている場合は野党ラベルを付与した。野党投稿をアンダーサンプリングして件数を合わせ、それぞれ件数は 779 件であり、評価のみに用いる。

国会議員投稿データについて、当該議員が所属する政党が自民党または公明党の場合は与党ラベル、他政党の場合は野党ラベルを付与した。野党投稿をアンダーサンプリングして件数を合わせ、それぞれ件数は 103,463 件であり、評価のみに用いる。

4.1.2 政治的投稿検出用のデータセット

ランダム投稿を政治的な内容を含まない一般的な投稿、前述の傾向予測用データセットのうち件数の多い政党チャンネルコメントを政治的な内容をより含む投稿とみなし、与野党を結合して用いた。ランダム投稿をアンダーサンプリングして件数を合わせ、それぞれ件数は 48,054 件となった。

4.2 実験設定

どのデータセットでも、訓練・検証・評価を 8:1:1 に分割した。

事前学習済みのモデルには、SNS の投稿に特化した JTweetRoBERTa、比較用として DeBERTaV3³⁾ ベースの ku-nlp/deberta-v3-base-japanese³⁾を使用した。

政治的傾向予測モデルを作成する際の政治的投稿フィルタでは、政治的投稿検出モデル確信度が 0.99 以上、0.9 以上、フィルタ無し(すべての投稿を用いる)の 3 種類を比較した。作成した政治的傾向予測モデルをそれぞれのデータセットについて、学習データ同様に異なる閾値のフィルタを適用したデータセットに対して評価した。この評価時のフィルタは、政治的な投稿かどうかで与野党分類問題が難しくなるかを確認するために設けた。

5 実験結果

5.1 政治的投稿検出モデル

ファインチューンの結果を表 1 に示す。

表 1 政治的投稿検出モデルの評価

モデル	Accuracy	Precision	Recall	F1
JTweetRoBERTa	96.37	97.04	95.59	96.31
DeBERTa	95.56	94.78	96.35	95.56

JTweetRoBERTa と DeBERTa の両モデルともに、すべての指標で 95%以上を達成し、高い性能を示し

1) <https://developer.x.com/ja/docs/x-api>

2) <https://developers.google.com/youtube/v3/getting-started>

3) <https://huggingface.co/ku-nlp/deberta-v3-base-japanese>

た。投稿に特化した JTweetRoBERTa が Recall を除き高い結果となったのは、本研究の対象テキストはすべて SNS 投稿であるためと考えられる。

5.2 政治的傾向予測モデル

ファインチューンしたモデルを政党チャンネル投稿で評価した結果を表 2 に示す。フィルタの効果を検証するベースラインであるフィルタ「無し」では、両モデルともに F1 で 80%程度である。JTweetRoBERTa の場合、学習時のフィルタ閾値を上げることで、評価時のフィルタ閾値にかかわらず F1 が 14 ポイント向上した。DeBERTa の場合も同様に、学習時のフィルタの予測確率を上げることで、大幅な性能向上が見られた。

表 2 政治的な傾向予測モデルの評価 (JTweet: JTweetRoBERTa, Acc: Accuracy, P: Precision, R: Recall)

学習条件	評価条件	Acc	P	R	F1
フィルタ	フィルタ (件数)				
JTweet					
無し	無し	81.47	79.31	85.68	82.37
0.9 以上	無し	96.68	95.83	97.67	96.74
0.99 以上	(4,798)	95.76	94.84	96.91	95.86
0.999995 以上		96.65	95.63	98.02	96.81
無し	0.9	82.01	82.86	81.19	82.02
0.9 以上	以上	96.55	96.82	96.33	96.57
0.99 以上	(4,757)	95.95	96.10	95.90	96.00
0.999995 以上		96.75	96.84	96.90	96.87
無し	0.99	82.60	85.23	79.29	82.15
0.9 以上	以上	96.87	97.43	96.33	96.88
0.99 以上	(4,718)	95.80	96.76	94.90	95.82
0.999995 以上		96.82	97.34	96.51	96.92
無し	0.999995	80.47	86.54	72.65	78.99
0.9 以上	以上	88.40	93.50	82.75	87.80
0.99 以上	(2,923)	88.00	93.35	82.18	87.41
0.999995 以上		96.51	98.29	94.93	96.58
DeBERTa					
無し	無し	81.37	78.44	87.05	82.52
0.8 以上	無し	95.57	94.38	97.06	95.70
0.9 以上	(4,798)	95.66	94.53	97.19	95.84
0.99 以上		98.58	98.46	98.46	98.46
無し	0.8	82.06	85.87	77.19	81.29
0.8 以上	以上	96.19	97.75	94.68	96.19
0.9 以上	(3,882)	96.40	97.21	95.75	96.48
0.99 以上		98.58	100.00	96.92	98.44
無し	0.9	81.30	86.07	75.17	80.25
0.8 以上	以上	91.83	95.00	88.60	91.69
0.9 以上	(2,833)	96.08	97.26	95.06	96.15
0.99 以上		96.45	95.45	96.92	96.18
無し	0.99	70.74	69.95	73.76	71.81
0.8 以上	以上	72.00	71.84	73.91	72.86
0.9 以上	(141)	73.39	72.39	78.05	75.12
0.99 以上		96.45	96.88	95.38	96.12

ファインチューンしたモデルを、国会議員投稿と政党フォロワー投稿でそれぞれ投稿単位とアカウント単位で評価した結果を表 3 と付録の表 6 に示す。

5.3 人手ラベル付投稿による評価

より信頼性の高い評価を行うため、国会議員投稿と政党フォロワー投稿の一部を手動で政治的な投稿かどうか、政治的な投稿である場合は与党寄りであるか野党寄りであるかラベル付けを実施した。人手ラベルの結果から、政治的投稿と政治的でない投稿を各 50 件、与党寄りと野党寄りの投稿を各 25 件選び、評価した結果を表 4,5 に示す。

6 分析と考察

6.1 政治的傾向予測モデルの評価データセット別分析

データセット内で評価データを同一にして比較可能にするため、評価条件フィルタが同じもの同士を比較する。学習条件はフィルタ無しの場合をベースラインとする。以下では F1 値を対象に議論する。

国会議員投稿による評価 JTweetRoBERTa の場合、評価時フィルタ「無し」の投稿単位評価を除いた全組み合わせにおいて、学習時フィルタ「0.9 以上」で性能向上した。アカウント単位の場合の性能向上が顕著であり、すべての組み合わせにおいて 8 ポイント以上の向上、最大で 13 ポイント向上した。

DeBERTa の場合、投稿単位評価では性能向上は見られず、評価時のフィルタ「無し」もしくは「0.9 以上」の場合のアカウント単位の評価のみ、7 ポイント程の向上が見られた。JTweetRoBERTa に比べ、DeBERTa の政治的投稿検出モデルが確信度を低く出す傾向があり、フィルタにより必要以上に投稿が除かれたことが原因の一つと考えられる。

政党フォロワー投稿による評価 JTweetRoBERTa では評価時フィルタ「0.95 以上」で性能向上が見られた。国会議員投稿とは違い、投稿単位・アカウント単位ともに性能向上は 2 ポイント前後であった。

DeBERTa の場合、投稿単位、アカウント単位ともに性能向上は殆どなかった。フィルタ「0.99 以上」の評価件数が極端に少ないためと考えられ、フィルタの閾値には改善の余地がある。

6.2 人手ラベル付投稿の評価の分析

政治的投稿検出モデルの評価では、JTweetRoBERTa の場合、すべての指標で 70%以上を達成し

表 3 国会議員投稿による評価 (JTweet: JTweetRoBERTa, Acc: Accuracy, P: Precision, R: Recall)

学習条件		評価条件										
モデル	フィルタ	フィルタ	投稿単位				アカウント単位					
			件数	Acc	P	R	F1	件数	Acc	P	R	F1
JTweet	無し	無し	185,842	55.56	52.83	78.55	63.17	580	56.90	54.76	99.01	70.52
	0.9 以上			61.08	58.56	67.73	62.81		78.79	73.99	91.39	81.78
	0.99 以上			57.63	55.04	69.16	61.30		67.93	62.55	95.70	75.65
	無し	0.99 以上	100,329	58.46	56.74	81.46	66.89	573	58.64	55.98	98.33	71.34
	0.9 以上			64.06	62.93	73.54	67.82		81.85	75.79	96.00	84.71
	0.99 以上	60.19	59.00	74.45	65.83	69.81	64.08	96.33	76.96			
無し	0.999995 以上	22,378	59.97	58.17	85.32	69.17	541	60.63	57.70	97.57	72.52	
0.9 以上			64.37	63.27	77.05	69.48		75.42	70.23	93.40	80.18	
0.99 以上	61.12	59.91	79.04	68.16	67.47	62.61	96.53	75.96				
DeBERTa	無し	無し	185,842	59.70	55.97	79.43	65.67	580	63.79	59.09	99.01	74.01
	0.9 以上			60.03	58.14	62.92	60.44		78.28	75.43	86.42	80.56
	0.99 以上			52.06	50.40	74.64	60.17		55.86	54.28	96.69	69.52
	無し	0.9 以上	114,333	59.40	55.28	79.28	65.14	575	64.17	59.56	98.34	74.19
	0.9 以上			60.01	57.49	62.95	60.10		78.78	75.79	87.38	81.17
	0.99 以上	52.14	49.99	75.00	59.99	55.30	54.07	97.01	69.44			
無し	0.99 以上	2,177	56.82	49.52	72.92	58.99	414	59.90	58.28	83.94	68.80	
0.9 以上			59.21	51.92	56.85	54.27		63.04	63.83	68.81	66.23	
0.99 以上	50.11	43.67	59.12	50.23	53.86	54.55	74.31	62.91				

表 4 人手ラベル付した投稿による政治的な投稿検出モデルの評価 (JTweet: JTweetRoBERTa, Acc: Accuracy, P: Precision, R: Recall)

モデル	Acc	P	R	F1
国会議員投稿				
JTweet	73.00	70.90	78.00	74.28
DeBERTa	58.00	63.33	38.00	47.50
政党フォロワー投稿				
JTweet	72.00	71.15	74.00	72.54
DeBERTa	56.00	57.89	44.00	50.00

た。DeBERTa の場合、F1 で 50%前後と政党チャンネル投稿での評価結果とは大きく異なり、性能が低かった。DeBERTa が誤って政治的であると判定したものは、比較的文章長が短かった。誤って政治的ではないと判定したものは、政党名や国内における政策など直接的な表現が少ないものが多かった。

政治的傾向予測モデルの評価では、学習時フィルタが「0.9 以上」「0.99 以上」で性能向上が見られ、JTweetRoBERTa の場合、国会議員投稿による評価では F1 で 15 ポイント向上した。一方、政党フォロワーによる評価の場合、両モデルともに性能向上が軽微であり、絶対性能も低かった。フォローしている政党がそのまま与野党支持と一致せず分類タスクとして難しい可能性を示唆している。DeBERTa が誤って与党と判定したものは、与党を比喻を用いて批判するなど文脈の理解が必要なものであった。一方、DeBERTa が誤って野党と判定したのは、天皇制など前提知識が必要なものであった。学習件数が少

表 5 人手ラベル付した投稿による政治的な傾向予測モデルの評価 (JTweet: JTweetRoBERTa, Acc: Accuracy, P: Precision, R: Recall)

モデル	フィルタ	Acc	P	R	F1
国会議員投稿					
JTweet	無し	60.00	56.76	84.00	67.74
	0.9 以上	80.00	72.73	96.00	82.76
	0.99 以上	72.00	66.67	88.00	75.86
DeBERTa	無し	74.00	67.65	92.00	77.97
	0.9 以上	78.00	75.00	84.00	79.25
	0.99 以上	58.00	54.55	96.00	69.57
政党フォロワー投稿					
JTweet	無し	44.00	44.83	52.00	48.15
	0.9 以上	58.00	59.09	52.00	55.32
	0.99 以上	52.00	52.17	48.00	50.00
DeBERTa	無し	44.00	44.83	52.00	48.15
	0.9 以上	50.00	50.00	40.00	44.44
	0.99 以上	52.00	51.85	56.00	53.85

なく前提知識まで学習できなかった可能性がある。

7 おわりに

SNS 投稿の政治的傾向を予測するにあたり、データを事前に政治に関連するものに絞り込むことにより、学習時、評価時ともに性能が向上することを示した。また、異なる SNS プラットフォーム上の投稿文でも政治的投稿・政治的傾向ともに十分な性能で分類を行えることを示した。学習データを事前にタスクのトピックに特化して絞り込む方法は、政治に限らず一般に適用可能であり今後試みていきたい。

謝辞

本研究はJSPS 科研費 (JP22H00804) , JST さきがけ (JPMJPR2461) , JST AIP 加速課題 (JPMJCR22U4) , およびセコム科学技術財団特定領域研究助成の支援をうけた。

参考文献

- [1] 大谷卓史. Sns は世論を製造するか? 情報管理, Vol. 57, No. 6, pp. 420–422, 2014.
- [2] 金丸恭文, 柳川範之. Sns 時代の政策決定メカニズム. NIRA オピニオンペーパー, Vol. 68, pp. 1–6, 2023.
- [3] 高須遼, 狩野芳伸. JTweetRoBERTa: 大規模 sns 投稿テキストによる事前学習と各種タスクによる性能検証. 言語処理学会第 30 回年次大会発表論文集, 2024.
- [4] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. **ArXiv**, Vol. abs/1907.11692, , 2019.
- [5] Ali Rahmati, Ehsan Tavan, and Mohammad Ali Keyvanrad. Predicting content-based political inclinations of iranian twitter users using bert and deep learning. **AUT Journal of Mathematics and Computing**, Vol. 4, No. 2, pp. 145–154, 2023.
- [6] 陳景慧, 水野貴之, 土井翔平. 多言語埋め込みとグラフニューラルネットワークに基づく twitter テキストからの政治的立場の類似性の推定. 人工知能学会全国大会論文集, Vol. JSAI2024, pp. 4C1GS1103–4C1GS1103, 2024.
- [7] Pengcheng He, Jianfeng Gao, and Weizhu Chen. Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing. **ArXiv**, Vol. abs/2111.09543, , 2021.

A 付録

以下の投稿例は実例をもとに我々が作例したものである。

A.1 人手ラベル付投稿による政治的投稿検出モデルの評価分析の具体例

非政治的ラベルであるが、DeBERTa で誤答、JTweetRoBERTa で正解した例
- 海洋生物を大切にしよう - 毎日色々あるのが心配ですが、良い年末をお過ごしください。
政治的ラベルであるが、DeBERTa で誤答、JTweetRoBERTa で正解した例
- ただいまショッピングモールにて、街頭演説しています！次は夕方から市役所にて！ - 紛争地域における即時停戦・平和の実現を求める声明を採択した自治体を共有しました。

A.2 人手ラベル付投稿による政治的傾向予測モデルの評価分析の具体例

与党ラベルであるが、DeBERTa で誤答、JTweetRoBERTa で正解した例
- X 政権自体が粉碎機のような。... は粉碎機にかけるように隠滅してきた。 - X 党の批判者が増えているように感じます。... 一度立ち止まって課題を整理し、批判を受け入れて方向転換して欲しかったですね。
野党ラベルであるが、DeBERTa で誤答、JTweetRoBERTa で正解した例
- X 高校の模擬選挙にて各政党が高校生に政策をアピールして、生徒が支持する政党を投票する取組。残念だったのは選挙会場に自民党のポスターだけ貼られていなかったこと - 天皇制を「仕組み」と表現する人々は共産主義者。その理由は仕組みなら取り除くことが可能だから。

A.3 政治的予測モデルの評価

表 6 政党フォロワー投稿による評価

学習条件		評価条件										
モデル	フィルタ	フィルタ	投稿単位					アカウント単位				
			件数	Acc	P	R	F1	件数	Acc	P	R	F1
JTweet	無し 0.9 以上 0.99 以上	無し	1,223	53.56	52.46	73.44	61.20	168	69.64	81.62	81.02	81.32
				55.60	54.45	67.21	60.16		68.45	83.87	75.91	79.69
	無し 0.9 以上 0.99 以上	0.99 以上	690	54.37	53.65	62.62	57.79	63.69	83.33	69.34	75.70	
				54.64	53.21	74.71	62.15	67.15	78.18	80.37	79.26	
	無し 0.9 以上 0.99 以上	0.999995 以上	177	50.28	47.41	79.01	59.26	70	65.71	69.49	87.23	77.36
				48.59	46.09	72.84	56.46	57.14	65.45	76.60	70.59	
DeBERTa	無し 0.9 以上 0.99 以上	無し	1,223	54.78	53.33	74.92	62.30	168	69.05	80.14	82.48	81.29
				53.15	52.80	57.21	54.92		58.93	82.08	63.50	71.60
	無し 0.9 以上 0.99 以上	0.9 以上	780	49.88	49.82	68.03	57.52	67.26	80.60	78.83	79.70	
				53.85	51.91	75.07	61.37	157	68.79	80.31	80.95	80.63
	無し 0.9 以上 0.99 以上	0.99 以上	14	54.36	52.85	60.89	56.59	61.15	82.83	65.08	72.89	
				50.13	49.28	71.92	58.48	64.33	79.66	74.60	77.05	
無し 0.9 以上 0.99 以上	0.99 以上	14	42.86	54.55	66.67	60.00	11	54.55	66.67	75.00	70.59	
			42.86	57.14	44.44	50.00		45.45	66.67	50.00	57.14	
				42.86	55.56	55.56	55.56	54.55	71.43	62.50	66.67	