芸能人への感想を表す X 上のポスト集約 およびウェブ検索・RAG によるその理由の集約

横山響[1] 土田陸斗[2] 宇津呂武仁[3]
[1]筑波大学大学院 システム情報工学研究群 知能機能システム学位プログラム
[2]筑波大学 理工学群 工学システム学類 [3]筑波大学 システム情報系 知能機能工学域
{s2320808,s2110466}@u.tsukuba.ac.jp utsuro@_iit.tsukuba.ac.jp

概要

本研究は，芸能人のファンが，芸能人に関連する事柄についての情報探索を行いやすくすることを目的とする．X から特定の芸能人に関するポストを収集し，大規模言語モデル (LLM) である ChatGPT を活用して，芸能人の評価対象とそれに関連する感想のペアを抽出する．これらをキーワードとして使用し，上位 30 件のウェブページから感想の背景にある理由を探る．この理由の収集・集約では，ChatGPT に検索拡張生成 (RAG) の枠組みを適用し，収集したウェブページの内容を参照情報として活用する．人手で作成した評価対象と感想のペア，および理由のデータを参照として評価を行った結果，提案手法が高い精度を達成することが明らかになった．

1 はじめに

本論文は芸能人のファンが，芸能人に関連する事柄についての批評や関連情報の探索を行いやすくすることを目的とする．この目的を達成するため，本研究では以下のような手法を提案する．まず，X(旧 Twitter) において特定の芸能人に関連する事柄について述べているポストを収集し，それらの芸能人に関する主要な感想を収集する．評価対象と感想についての ChatGPT を用いた収集・集約方法の詳細は，3 節で説明する．その後，集められた評価対象と感想のペアをキーワードとして使用し，ウェブページを検索して上位 30 位までの内容を，それらの感想の背景にある理由を探るための情報源として収集する．ウェブページの内容収集が完了したら，各ページの内容をもとに芸能人の評価対象に対する感想の詳細な理由を収集する．この部分では，大規模言語モデル (LLM) として ChatGPT を利用する．本研究の重要な側面は，この理由収集プロセスにおける RAG [6] の枠組みの適用である．収集したウェブページの内容を参照情報として活用することにより，LLM の事前知識のみに依存する場合と比較して，より信頼性の高い結果を得ることが可能となる．最後に，各キーワードに対して得られた複数の感想の理由を，その頻度を考慮して集約しランク付けする．これにより，ユーザーは芸能人の評価対象に対する感想の理由を集約されたランク形式で容易に理解することができ，芸能人関連トピックの批評や関連情報の探索が可能となる．この部分においても ChatGPT を使用する．本論文の貢献は以下の通りである：

1. 大規模言語モデルである ChatGPT を用いて，X のポストから芸能人の評価対象に関する感想を効果的に収集・集約する新しいアプローチを提案した．
2. RAG フレームワークにおいて，ChatGPT がウェブページから芸能人に対する感想の理由を収集・集約する上で高い効果を示すことを実証した．

2 関連研究

芸能人に関する情報探索を支援する先行研究には，Twitter と Wikidata を組み合わせて大規模な芸能人プロフィールデータセットを構築する研究 [15] や，ソーシャルメディアにおける芸能人の言語使用の説得戦略を分析してその影響力を予測する研究 [2] がある．芸能人のファンを支援することに関しては，マイクロブログ・ポストにおける芸能人と感想の関係性を判定する研究 [9] や，マイクロブログ・ポストから芸能人の評価対象に関する感想を収集する研究 [13] がある．本論文はこれらの先行研究と異なり，マイクロブログ・ポストから抽出した芸能人の評価対象に関する感想の背景にある理由を

表 1　芸能人の評価対象に対する感想の収集・集約の評価結果

芸能人名	使用ポスト数	芸能人に対する感想収集		芸能人に対する感想集約	
		再現率 (# 正解 / # 参照)	適合率 (# 正解 / # 収集)	再現率 (# 正解 / # 参照)	適合率 (# 正解 / # 集約)
山田涼介	1000	0.63 (=112/179)	0.67 (=112/166)	0.43 (=29/68)	0.73 (=29/40)
二宮和也	500	0.73 (=54/74)	0.84 (=54/64)	0.41 (=13/32)	0.81 (=13/16)
菊池風磨	500	0.68 (=36/53)	0.71 (=36/51)	0.54 (=20/37)	0.67 (=20/30)
小栗旬	500	0.58 (=38/65)	0.62 (=38/61)	0.27 (=8/30)	0.40 (=8/20)
綾野剛	500	0.56 (=90/161)	0.82 (=90/110)	0.24 (=5/21)	0.63 (=5/8)
合計/ミクロ平均	3000	0.62 (=330/532)	0.73 (=330/452)	0.40 (=75/188)	0.66 (=75/114)

ウェブページから収集・集約する．また本論文は，先行研究が共起頻度統計などの手法に依存していたのに対し，大規模言語モデルである ChatGPT を用いてマイクロブログ・ポストから芸能人の評価対象に関する感想を抽出する点でも異なる．さらに，本論文の主要な特徴の 1 つは，外部情報の参照を可能にすることで LLM が生成する出力の信頼性を向上させる RAG [6] の使用である．RAG を用いることで，LLM は外部データベースから取得した情報を参照することができるようになり，生成されるコンテンツの正確性と信頼性が向上する．本論文では，RAG を活用してウェブページの内容に基づいて感想の理由を収集することで，LLM の知識のみに依存する従来の手法と比較して，収集結果の信頼性を大幅に向上させることを目指している．最近の研究では，ツール検索と計画生成のための文脈調整 [1]，後期相互作用モデルと共同訓練による非制限ドメインのテーブル質問応答の改善 [7]，教師あり学習を必要としない Few-shot の多言語画像キャプション生成 [12]，より強力な質問応答システムのための追加コンポーネントの組み込み [14] など，様々な RAG の応用と改善が探求されている．また，高リソース言語からのプロンプトを使用した低リソース言語における Zero-shot 性能の向上 [8]，2 段階フレームワークによる非知識集約型タスクにおける検索の活用 [4]，知識集約型タスクにおける生成品質向上のための豊富な回答エンコーディングの組み込み [5] に焦点を当てた研究もある．ChatGPT に関連する研究には，エンティティリンキング [10]，対話分析 [3]，テキスト要約 [11, 16, 17] などもある．

3　ChatGPT を用いた X ポストからの感想収集

本節では，X から芸能人名を含むポストを収集し，芸能人の特定の評価対象に関する感想に言及しているポストを特定し，ChatGPT を用いてそれらを評価対象と感想のペアに集約する手順について説明する．

3.1　X ポストの収集

本論文では，X で頻繁に議論される 5 名の芸能人を選択し，2022 年 9 月 7 日から 2023 年 4 月 9 日までの期間，その芸能人名を検索クエリとしてポストを収集した．ポストの収集には Twitter Search API[1] を使用した．本論文ではリポスト以外のポストのみを使用する．

3.2　ポストからの感想の収集・集約

次に，前節で収集した特定の芸能人名を含む X ポストに対して，2 つの主要なタスクを実行する．第一に，その芸能人の特定の評価対象に関する感想に言及しているポストを収集する．第二に，収集した情報を評価対象と感想のペアに集約する．これらのタスクの枠組みとして，ChatGPT[2] モデルの `gpt-4o-2024-08-06` を使用する．ここでは，芸能人「山田涼介」を対象としたプロンプトの例を示す．プロンプトではまず，入力されたポストの中から，山田涼介の評価対象 (感想の対象) に対する感想を述べているポストを収集するよう指示する．次に，収集したポストを評価対象と感想のペアごとに集約するよう指示する．その後，出力形式を指定し，各評価対象と感想のペア，およびそれらに関連する具体的なポストを出力するよう指示する．この処理は 100 件のポストを 1 単位として実施する．プロンプトでは，評価対象に芸能人名「山田涼介」を含めないよう指示し，評価対象とそれに対応する感想が重複しないよう注意を促している．最後に，出力前に，収集も集約もされていないポストが残っていな

1) https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets
2) https://platform.openai.com/docs/models/

表 2　感想の理由の収集・集約の全体評価結果

感想理由の収集		感想理由の集約		
再現率	適合率	再現率	適合率	冗長度
0.66 (=73/111)	0.78 (=73/93)	0.66 (=19/29)	1.00 (=19/19)	1.11 (=21/19)

いかを再確認するよう指示している．続いて，100 ポスト単位での集約結果に対して集約処理を行う．予備実験として山田涼介に関する 1000 ポストでの分析を行い，頻度 5 回以上出現する評価対象と感想のペアが 500 ポスト時点で既に出現していることを確認したため，以降の芸能人については 500 ポスト分の集約結果を入力として，評価対象と感想のペアの重複を統合する．これにより，データセット全体での感想の一貫性を確保する．本手法の性能を評価するため，収集したポストの一部を人手でアノテーションし，参照データセットを作成した．評価は，収集タスクと集約タスクの両方について，ChatGPT が生成した出力と参照を比較して実施した．結果を表 1 にまとめる．評価の結果，芸能人に対する感想の収集タスクでは再現率 0.62 (=330/532)，適合率 0.73 (=330/452) を達成した．これは，ChatGPT が感想を含むポストを高い精度で識別できることを示している．一方，感想の集約タスクでは再現率 0.40 (=75/188)，適合率 0.66 (=75/114) という結果となった．集約タスクの再現率は比較的低いものの，適合率が 0.66 を維持していることから，ChatGPT が抽出した評価対象と感想のペアの信頼性は十分に高いと言える．

4　感想の理由の収集・集約

本節では，ChatGPT を用いてウェブページから感想の理由を収集・集約する手順について説明する．4.1 節では，3 節で収集・集約された評価対象とそれに対応する感想のペアから，ウェブページ検索のキーワードとして使用するペアを選択するプロセスについて説明する．4.2 節では，選択されたキーワードを用いてウェブページを検索し，ウェブページの内容を収集する方法について説明する．4.3 節では，収集したウェブページの内容から感想の理由を収集する手順を説明し，人手評価の結果を示す．4.4 節では，収集した感想の理由を集約する手順を説明し，人手評価の結果を示す．

4.1　評価対象と感想のペアの選択

本節では，3 節で収集・集約された評価対象とそれに対応する感想のペアから，ウェブページ検索のキーワードとして使用するペアを選択するプロセスについて説明する．4.2 節で詳しく説明するように，本研究ではウェブページの検索に Google の検索エンジン[3)]を使用する．そのため，3.2 節で集めた評価対象と感想のペアから，各芸能人について，それぞれ最も出現頻度の高いペアを選定した．この選定は，評価対象と感想のペアの中で最も代表的なものに焦点を当てることで，提案手法の基本的な性能を検証することを目的としている．以降の節では，これらの選択された評価対象と感想のペアを検索キーワードとして使用し，感想の理由の収集・集約が可能かどうかを検証する．具体的に選定された 5 件のペアについては，付録 A の表 3 の「キーワード」列に示す．以降の節の目的は，これらの選択された評価対象と感想のペアを検索キーワードとして使用して，感想の理由を収集・集約することが可能かどうかを明らかにすることである．

4.2　ウェブページの検索

まず，前節で選択したキーワードを用いて Google の検索エンジンでウェブページを検索する．次に，検索結果の上位 30 件のウェブページの内容を人手で収集する．この一連の操作をすべてのキーワードに対して実行する．例えば，「山田涼介の顔・美しい」の場合，まず「山田涼介の顔・美しい」をキーワードとしてウェブページを検索し，上位 30 件のページの内容を収集する．

4.3　理由の収集

4.3.1　手順

次に，前節で収集したウェブページの内容を用いて，各ウェブページごとに感想の理由を収集する．理由収集の枠組みとして，ChatGPT モデル `gpt-4o-2024-08-06` を使用する．ここでは，キーワード「山田涼介の顔・美しい」を対象としたプロンプトの例を示す．まず，ChatGPT に対し，ChatGPT 自身の事前知識を使用せず，コンテキストとして追加された収集したウェブページの内容を参照して，感想の理由を探すよう指示する．もし，コンテキス

3)　https://www.google.co.jp/

トに感想の理由が含まれていない場合，「×」のみを出力するように指示する．ChatGPT にウェブページの内容に基づいて感想の理由を探させることで，事実と異なるまたは存在しない情報の出力をしてしまうハルシネーションを抑制することが期待される．

4.3.2 評価

ここでは，ChatGPT が収集した感想の理由を，人手で収集した参照用の感想の理由と比較して評価する．評価は 5 個のキーワードで実施する．4.2 節で各キーワードについて得られたウェブページの内容に基づき，第一著者が人手で感想の理由を収集した．与えられたウェブページ d に対する ChatGPT の出力した理由の多重集合 S(d) と，ウェブページ d に対して人手で準備された参照理由の多重集合 R(d) に基づいて，再現率と適合率を以下のように定義する：

$$\text{再現率} = \sum_d |R(d) \cap S(d)| / \sum_d |R(d)|,$$

$$\text{適合率} = \sum_d |R(d) \cap S(d)| / \sum_d |S(d)|$$

評価は収集した各ウェブページに対して実施され，各キーワードの評価結果としてミクロ平均を使用する．全体の評価結果は，評価用の全 5 キーワードの評価結果のミクロ平均として測定される．理由収集の全体の評価結果を表 2 に示す．詳細な評価結果については，付録 A の表 3 に示す．結果として，理由収集において，再現率は 0.66，適合率は 0.78 と高い性能が達成された．

4.4 理由の集約

4.4.1 手順

次に，前節で収集した感想の理由を集約する．ここでも，ChatGPT モデル `gpt-4o-2024-08-06` を使用する．プロンプトでは，各ウェブページから収集した感想の理由で言及されている内容に基づいて理由をカテゴリ分けすることで理由の集約を実行するよう指示する．

4.4.2 評価手順

ここでは，ChatGPT を用いて集約された感想の理由を，人手で集約された参照用の理由と比較して評価する．評価は 5 個のキーワードで実施する．人手で集約された参照用の理由は，第一著者が理由収集ステップで使用した参照データから類似の理由をグループ化することで作成された．評価のため，まず基本となる 2 つの集合を定義する．1 つ目は，ChatGPT が集約した理由の多重集合 S' であり，2 つ目は人手で準備された参照理由の集合 R である．次に，S' を R との対応関係に基づいて 2 つの多重集合に分割する．S' の要素のうち，R に対応する理由を持つものの多重集合を S'_r，対応する理由を持たないものの多重集合を $S'_{\neg} r$ と定義する．さらに，これらの多重集合から重複を除いた集合を定義する．S'_r については，R の同一の理由に対応する複数の理由を 1 つに集約することで S_r を得る．同様に，$S'_{\neg} r$ についても類似した理由を 1 つに集約することで $S_{\neg} r$ を得る．最後に，S_r と $S_{\neg} r$ の和集合として S を定義する．これらの集合に基づいて，再現率，適合率，冗長度を以下のように定義する．

$$\text{再現率} = |S_r|/|R|,\quad \text{適合率} = |S_r|/|S|,$$

$$\text{冗長度} = |S'_r|/|S_r|$$

全体の評価結果は，全 5 キーワードの評価結果のミクロ平均として計算される．

4.4.3 評価結果

理由集約の評価結果を表 2 に示す．詳細な評価結果は，付録 A の表 3 に示す．結果として，例を含む理由集約では，再現率が 0.66，適合率が 1.00，冗長性が 1.11 となった．これは，ChatGPT が感想の理由を高い性能で収集・集約できることを示している．特に，適合率が 1.00 という性能を達成していることは，ChatGPT が出力した理由が正確であることを意味する．再現率 0.66 は，参照データの理由の 6 割以上を適切に抽出できていることを示している．また，冗長度が 1.11 という理想に近い値となったことは，ChatGPT が無駄な重複なく理由を集約できていることを示唆している．

5 おわりに

本論文では，芸能人のファンが芸能人に関する情報を批評・探索する能力を向上させる手法を提案した．提案手法によって得られた結果を，人手で収集・集約した感想，およびその理由と比較して評価を行い，ChatGPT がより高い性能を示すことを確認した．今後は，提案手法の性能をより包括的に評価するため，Mistral Large[4] や LLaMA 3 70B[5] など，同程度のパラメータ数を持つ大規模なモデルと ChatGPT を比較することを計画している．

4) https://mistral.ai/news/mistral-large/
5) https://ai.meta.com/blog/meta-llama-3/

謝辞

本論文は，一部，科研費 21H00901， 電気通信普及財団 2023 年度 研究調査助成， 弥生株式会社共同研究 の支援を受けたものである．

参考文献

[1] R. Anantha and D. Vodianik. Context tuning for retrieval augmented generation. In **Proc. UncertaiNLP**, pp. 15–22, 2024.

[2] Y. Chang, P. A. Wang, H. Hung, K. Khóo, and S. Hsieh. Examine persuasion strategies in Chinese on social media. In **Proc. 35th PACLIC**, pp. 108–118, 2021.

[3] S. E. Finch, E. S. Paek, and J. D. Choi. Leveraging large language models for automated dialogue analysis. In **Proc. 24th SIGDIAL**, pp. 202–215, 2023.

[4] Z. Guo, S. Cheng, Y. Wang, P. Li, and Y. Liu. Prompt-Guided Retrieval Augmentation for Non-Knowledge-Intensive Tasks. In **Findings ACL**, pp. 10896–10912, 2023.

[5] W. Huang, M. Lapata, P. Vougiouklis, N. Papasarantopoulos, and J. Pan. Retrieval Augmented Generation with Rich Answer Encoding. In **Proc. 13th IJCNLP and 3rd AACL**, pp. 1012–1025, 2023.

[6] P. Lewis, E. Perez, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. In **Proc. 34th NeurIPS**, pp. 483–498, 2020.

[7] W. Lin, R. Blloshmi, B. Byrne, A. de Gispert, and G. Iglesias. LI-RAGE: Late Interaction Retrieval Augmented Generation with Explicit Signals for Open-Domain Table Question Answering. In **Proc. 61st ACL**, pp. 1557–1566, 2023.

[8] E. Nie, S. Liang, H. Schmid, and H. Schütze. Cross-Lingual Retrieval Augmented Prompt for Low-Resource Languages. In **Findings ACL**, pp. 8320–8340, 2023.

[9] Y. Nozaki, K. Sugawara, Y. Zenimoto, and T. Utsuro. Tweet review mining focusing on celebrities by MRC based on BERT. In **Proc. 36th PACLIC**, pp. 757–766, 2022.

[10] R. Peeters and C. Bizer. Using ChatGPT for entity matching. **arXiv preprint arXiv:2305.03423**, 2023.

[11] D. Pu and V. Demberg. ChatGPT vs Human-authored Text: Insights into Controllable Text Summarization and Sentence Style Transfer. In **Proc. 61st ACL-SRW**, pp. 1–18, 2023.

[12] R. Ramos, B. Martins, and D. Elliott. LMCap: Few-shot multilingual image captioning by retrieval augmented language model prompting. In **Findings of ACL**, pp. 1635–1651, 2023.

[13] K. Sugawara and T. Utsuro. Developing a dataset for mining reviews in tweets focusing on celebrities' aspects. In **Proc. 7th ABCSS**, pp. 466–472, 2022.

[14] W. Tan, Y. Li, et al. Reimagining retrieval augmented language models for answering queries. In **Findings of ACL**, pp. 6131–6146, 2023.

[15] M. Wiegmann, B. Stein, and M. Potthast. Celebrity Profiling. In **Proc. 57th ACL**, pp. 2611–2618, 2019.

[16] H. Zhang, X. Liu, and J. Zhang. Extractive Summarization via ChatGPT for Faithful Summary Generation. In **Findings of EMNLP**, pp. 3270–3278, 2023.

[17] H. Zhang, X. Liu, and J. Zhang. SummIt: Iterative text summarization via ChatGPT. In **Findings of EMNLP**, pp. 10644–10657, 2023.

表 3　感想の理由の収集・集約の評価結果

キーワード	感想理由の収集		感想理由の集約		
	再現率	適合率	再現率	適合率	冗長度
山田涼介・顔・美しい	0.20 (=1/5)	1.00 (=1/1)	0.20 (=1/5)	1.00 (=1/1)	1,00 (=1/1)
二宮和也・映画「ラーゲリより愛を込めて」・感動的	0.77 (=33/43)	0.77 (=33/43)	0.83 (=5/6)	1.00 (=5/5)	1.00 (=5/5)
菊池風磨・魅力・好き	0.55 (=27/49)	0.75 (=27/36)	0.67 (=10/15)	1.00 (=10/10)	1.00 (=10/10)
小栗旬・鎌倉殿の 13 人・小栗旬の演技がすごい	0.92 (=11/12)	0.92 (=11/12)	1.00 (=2/2)	1.00 (=2/2)	2.00 (=4/2)
綾野剛・結婚・驚き	0.50 (=1/2)	1.00 (=1/1)	1.00 (=1/1)	1.00 (=1/1)	1.00 (=1/1)
合計/ミクロ平均	0.66 (=73/111)	0.78 (=73/93)	0.66 (=19/29)	1.00 (=19/19)	1.11 (=21/19)

A　感想の理由の収集・集約の詳細な評価結果

表 3 は，4 節に記載した感想の理由の収集・集約について，詳細な評価結果を示したものである．