YouTube 動画コメントを用いた視聴者感情の推定と感情処理能力の比較

菅野祐希 [1]　坂野遼平 [2]
[1] 工学院大学大学院 工学研究科 情報学専攻
[2] 一橋大学大学院 ソーシャル・データサイエンス研究科
em23019@ns@kogakuin.ac.jp　banno@computer.org

概要

YouTube や TikTok に代表される動画共有サービスは，現代において様々な人の行動や選択に大きな影響力を持つようになっている．それにより動画共有サービスはビジネスやマーケティングの場としても活用されるようになった．視聴者が動画を閲覧することでどのような感情を得るかという情報は，視聴者とマーケターの両方において有益となる．本稿では，オンライン動画共有サービス上にアップロードされた動画に付けられたコメントから，動画の視聴者に引き起こされる感情の推定を行う手法を提案する．BERT 及び数種類の大規模言語モデル (LLM) を用い，動画コメントを利用した各モデルの感情推定に関する能力の違いを明らかにする．提案手法では，7 種類の感情の強さを成分とした 7 次元ベクトルにより感情を表現し，推定を行う．実験の結果，100 件のコメントを使用して精細な感情の強度を推定する場合には LLM が優位であることが分かった．また，10 件の少ないコメント件数から最も強い感情を推定する場合，BERT が高いスコアを示した．

1　はじめに

近年，動画共有サービスの普及が進んでいる．調査によって，インターネットの利用時間の内訳で最も長時間利用されているものが，YouTube[1)] や TikTok[2)] に代表される動画共有サービスであるという結果が出ている [1]．

動画共有サービスはユーザの多さやテレビ視聴者よりも若い年齢層への訴求力の高さから，マーケティングの場としても扱われるようになっている [2]．

動画媒体は人間の感情へ大きな影響を与えることが研究により判明している [3]．動画サイトにおける広告やテレビ CM に代表される，動画を用いたマーケティングにおいて，視聴者の感情をコントロールすることによって商品に対する見方を変えられる事が実証されている [4]．このような事例から，動画視聴者がどのような感情を動画から受け取ったかという情報は，有用であると言える．

我々は動画から得られる感情について研究を行ってきた [5][6][7]．動画から得られる感情を取得することで，関連動画の精度改善などのユーザビリティの向上やマーケティングへの応用などに活用できる．しかし，動画から感情を取得する手法について，現在主な手法である生体情報を用いた測定法 [8] はコンテンツの数に対するスケーラビリティに問題がある．YouTube 上には 2016 年時点で 20 億本以上の動画がアップロードされている [9]．動画共有サービスの利用者が増加傾向にあるため，動画本数が増加し続けることが考えられる．先述の手法では，多くの動画に対して感情の判別をすることがコストや時間の面から難しいと考えられる．

本研究では，動画共有サービスで最もよく知られた YouTube を対象とし，多くのユーザの感想や意見を集約し一般性の高い客観的な感情の抽出を目的とし，動画のコメントを利用した感情推定手法について提案する．近年では大規模言語モデル (LLM) の開発競争が加速し，複数の LLM が市場に存在する．しかし，本稿で提案している視聴者コメントに基づく動画視聴者の感情推定については，精度や傾向が明らかではない．そこで，BERT 及び 5 種類の LLM を用い，当該タスクにおける推定能力の差異を明らかにする．具体的には，深層学習モデルとして BERT[10] を，大規模言語モデルとして GPT (3.5-turbo，4，4-turbo，4o)[11]，[12]，Claude (3 Sonnet，3 Opus，3.5 Sonnet)[13]，[14]，

1） https://www.youtube.com
2） https://www.tiktok.com

Gemini (1.5 Flash，1.5 Pro)[15]，[16]，Deepseek[17]，Llama-3-ELYZA-JP-8B[18]，[19] を使用している．

2 関連研究

LLM の感情処理能力に関連する研究として，田中らの研究である，LLM がコーパスから獲得している社会集団に対する偏見や感情について評価を行う研究 [20] や，Huang らの研究である LLM が特定状況において想起する感情について，人間との比較を行っている研究 [21] などが存在する．それらの研究において，LLM が状況に応じて感情の強度やポジティブ・ネガティブを変化させること，LLM が持つ特定の社会集団に対する感情がすでに社会調査によって得られている人間の感情傾向と一致すること等が示されている．ここから，LLM は人間に近い感性や考えを，コーパスから獲得している可能性が示唆されている．

堺らの研究 [22] では，YouTube 上の炎上動画を判別する手段として動画のコメントを利用することを提案している．ニコニコ動画[3]での炎上動画の自動検出の技術を YouTube を対象とする形で応用したものとなっている．炎上している動画にはネガティブな言動が多く含まれているという点に着目した考え方である．動画に付加されているコメントを，感情辞書を用いて-1 から+1 までの範囲で感情数値を付与する．そして，その数値からコメント全体のポジティブ・ネガティブの値を判別している．

3 提案手法

本提案手法では，感情を日本語感情表現辞書 (JIWC-Dictionary) [23] に基づき，「悲しい」「不安」「怒り」「嫌悪感」「信頼感」「驚き」「楽しい」の 7 つに分類する．これら 7 種類の感情を，各成分 0 以上 1 以下の 7 次元の単位ベクトルとして扱う (以下，“感情ベクトル”とする)．以降に述べる手法では，入力をコメント，出力を感情ベクトルとする．

3.1 BERT による感情推定手法

本研究では，事前学習済みの BERT として，東北大学乾・鈴木研究室が公開している日本語版の Wikipedia コーパスによって学習が行われている日本語の BERT トークナイザ[4]を使用している．

感情ごとのコメントの傾向の学習に，Fine-tuning を用いる．学習用のコメントをクラウドソーシング[5]を用いて収集した．150 名の実験協力者を対象に，7 つの感情を得られる動画を 1 本ずつ提示させた．各感情 150 本ずつ，合計 1050 本の動画 ID を得た．

各動画から，YouTube Data API を用いてコメントを人気順で 100 件収集し，絵文字や英字，数字等を削除したものをファインチューニングした．

3.2 大規模言語モデルによる感情推定手法

LLM では追加の学習を行わず，コメントからどのような感情を推定できるかをプロンプトを用いて判別させる．

大規模言語モデルに推定を行いたい動画のコメントを入力する．1 件ずつ入力を行い，プロンプトによって指定されたフォーマットで感情値の返答を行う．最終的にそれらの平均の値を取り，単位ベクトルに変換することで感情ベクトルとして出力する．プロンプトを以下に示す．

プロンプト 1 感情推定プロンプト

```
"あなたには入力した動画のコメントから,動画が視聴者に与える感情の推定を行っていただきます．"
"条件として,感情は必ず次に示す 7つの選択肢から選択してください．"
"[悲しい],[不安],[怒り],[嫌悪],[信頼],[驚き],[嬉しい]"
"感情の強さに応じて，0,1,2,3,4の 5段階で評価してください．"
"感情が強く含まれる場合は大きい数字,感情が含まれない場合は小さい数字で評価してください．"
"今日は新発売のケーキを買った.最高の一日だ．"
"[悲しい:0][不安:0][怒り:0][嫌悪:0][信頼:0][驚き:1][嬉しい:4]"
"愛猫が亡くなってしまった.泣きながら一日眠ってしまった．"
"[悲しい:4][不安:2][怒り:0][嫌悪:0][信頼:0][驚き:2][嬉しい:0]"
```

GPT や Llama，Claude は，入力の例と理想の返答について設定することが可能である．Gemini に関しては入力例と返答例を送信することが不可能であるため，プロンプト 1 の下 4 行の例示を行っていない．例示として使用した文章は作成した物であり，

3） https://www.nicovideo.jp
4） https://github.com/cl-tohoku/bert-japanese?tab=readme-ov-file
5） https://crowdworks.jp

文章の感情値は実験協力者 5 名にアンケートを取った平均である．

4　評価・結果

二種類の手法によって評価を行った．一つ目は 100 件のコメントから 7 次元感情ベクトルを推定する精度の評価であり，二つ目は少数のコメント (10 件) から最大感情を推定する精度の評価である．

評価の際に用いる正解データとして，クラウドソーシングを利用した．7 次元感情ベクトルの推定精度評価では，100 名に対して動画 3 本を視聴した際に得られた感情の値を収集した．最大感情の推定精度評価では，100 名に対して 30 本の動画からランダムに 3 本を割り当て，各動画 10 名分の感情の値を収集した．

クラウドソーシングによる感情の評価は，1 から 5 の 5 段階で評価を行った．値が大きいほど感情が強い．正解データとして用いる際に，感情が得られない場合の数値を 0 とするため，0 から 4 の 5 段階評価として扱った．

BERT は Fine-tuning 時に学習に用いたコメントの動画数と入力 token 数によって 4 種類のモデルを用意している．入力した動画数は 350 本と 1050 本，入力 token 数は 256token と 512token に分かれている．それぞれを組み合わせた 4 通りのモデルを使用している．表においては (動画数/token 数) で示されており，動画 350 本，入力 token が 256token で学習されたモデルは (350/256) と示されている．

4.1　7 次元感情ベクトルの推定精度評価

7 次元感情ベクトルの推定精度評価では，推定を行いたい動画のコメント 100 件を入力，感情ベクトルを出力とする．正解データとの類似度を算出し，評価を行う．評価には以下の 3 本の動画を用いた．正解データの感情ベクトルを付録 A に示す．

- 動画 1: “【衝撃】火山にゴミを捨てて処理する場合に起こること” (VIENCE バイエンス)[6)]
- 動画 2: “back number - 手紙 (full)”[7)]
- 動画 3: “貫禄ありすぎて、父親と間違われる引きこもり生徒【ジェラードン】”[8)]

表 1 に結果を示す．

3 つの動画において最も高いスコアを示したモデルは，動画 1 が GPT-4 の 0.9950，動画 2 が Gemini-1.5-Pro の 0.9458，動画 3 が GPT-4-turbo の 0.9877 となっている．LLM については，ほとんどが 0.9 以上のスコアを示しており，0.95 を越えるモデルも多い．BERT は LLM と比較してスコアが低く，その中では 350file 250token で学習されたものは性能が比較的高く，1050file 256token で学習されたものは動画 2 と動画 3 において高い性能を示している．

4.2　最大感情の推定精度評価

最大感情の推定精度評価では，推定を行いたい動画のコメント 10 件を入力とし，感情ベクトルを出力とする．クラウドソーシングを用いて 10 名から 7 つの感情を得られる動画を収集し，その中から各感情の動画が 4 から 5 本になるようランダムに選定した 30 本の動画について感情推定を行っている．感情ラベルが付与された動画の本数を付録に示す．

各モデルから出力された感情ベクトルの中で最も大きいベクトル成分と，正解データの中で最も大きいベクトル成分を比較する．一致した動画の数を評価の基準としている．

表 2 に結果を示す．正解数は 30 本の動画中で最大感情が正解データと一致した数，正解率はその割合，コサイン類似度平均は 30 本の動画ごとのコサイン類似度の平均を表している．

最もスコアの高いモデルは，BERT の 1050file 256token で Fine-tuning を行ったモデルとなり，正解数は 22 で正解率は 73 %となった．次に性能が高いモデルとして，BERT の 1050file 512token モデル，Gemini-1.5-Pro と Deepseek となった．ランダムに感情を選択した場合，正解率は 1/7 (約 14 %) となるため，ランダムに感情を選択した場合よりも性能が高く，推定する能力は全てのモデルにおいて存在していることが分かる．

BERT は Fine-tuning の際に，入力可能な Token に上限がある．そのため，学習用のコメントを 1 動画につき 1 つのテキストファイルとして入力している関係上，入力されているコメントは人気順で上位のコメントに限られる．そのため，上位 10 件のコメントのみを使用した最大ベクトルの一致率評価において高い性能を示した可能性がある．それに対し，LLM は多数のコメントを平均することによって高い性能を持っていたが，少ないコメントではコメント一つ一つの持つ感情のバラつきを精細に感じ取ることで平均した際に想定していない感情が最も高く

6） https://youtu.be/uJBHO2NXvxQ
7） https://youtu.be/woRV5VxJDkU
8） https://youtu.be/OtJvFyqqeo0

表 1　言語モデルごとの 7 次元感情ベクトルと正解データのコサイン類似度

Model	BERT				GPT			
	350/256	350/512	1050/256	1050/256	3.5-turbo	4	4-turbo	4o
動画 1	0.8340	0.8061	0.7621	0.7376	0.9796	0.9950	0.9592	0.9925
動画 2	0.8609	0.8539	0.9035	0.8885	0.8997	0.9427	0.9349	0.9231
動画 3	0.8896	0.8359	0.9189	0.8901	0.9522	0.9397	0.9877	0.9509

Model	Gemini		Claude			Deepseek	ELYZA-llama
	1.5-flash	1.5-Pro	3 sonnet	3 opus	3.5 sonnet		
動画 1	0.7720	0.9486	0.9582	0.9856	0.9775	0.9867	0.9397
動画 2	0.8196	0.9458	0.9456	0.9421	0.9354	0.9154	0.9202
動画 3	0.8487	0.9789	0.9804	0.9611	0.9672	0.9408	0.9774

表 2　モデルごとの最大感情の推定結果

	BERT				GPT			
	350/256	350/512	1050/256	1050/512	3.5-turbo	4	4-turbo	4o
正解数	17	19	22	21	14	18	17	17
正解率	57%	63%	73%	70%	47%	60%	57%	57%
コサイン類似度平均	0.8652	0.8619	0.8831	0.8786	0.7843	0.8354	0.8321	0.8155

	Gemini		Claude		Deepseek	ELYZA-llama
	1.5-flash	1.5-Pro	3 sonnet	3.5 sonnet		
正解数	17	21	20	16	21	17
正解率	57%	70%	67%	53%	70%	57%
コサイン類似度平均	0.6699	0.8015	0.8480	0.8258	0.8245	0.7758

なる可能性を持つ.

5　おわりに

本研究では，オンライン動画共有サービスにおける動画視聴者の感情を自然言語処理によって推定する手法を提案し，実際の視聴者の感情と提案手法による推定結果を比較することによって評価を行った．深層学習モデルとして Fine-tuning を行った BERT を，大規模言語モデル (LLM) として GPT-4 を始めとした複数の LLM を使用し，これらモデル間の感情推定能力の差異を明らかにした.

また，本研究では複数の LLM による同一の手法による感情推定を行った．これによって，各 LLM モデルの持つ感情推定能力にどのような差が存在するかについて結果を示すことが出来た.

今後は，BERT の Fine-tuning の方法やデータの前処理方法に工夫を加えることで，推定能力の向上を図ることを考えている．また提案手法による感情推定を応用し，動画の検索や類似動画の推薦を行うシステムを検討していきたい.

参考文献

[1] 総務省. 令和 5 年度情報通信メディアの利用時間と情報行動に関する調査報告書.

[2] OXFORD ECONOMICS. Youtube impact report 2022 年日本における youtube の経済的，文化的，社会的影響. Technical report, YouTube.

[3] Robert W.Levenson and J.Gross . Emotion elicitation using films., 1995.

[4] 大友章司, 竹島久美子, 広瀬幸雄. 感情状態が商品広告の情報処理方略に及ぼす影響について. 人間環境学研究, Vol. 8, No. 2, pp. 123–132, 2010.

[5] 菅野祐希, 坂野遼平. Youtube 動画コメントを用いた動画の感情推定 -ルールベース及び bert の精度比較-. In **DEIM forum 2023**, No. 4a-6-1, 2023.

[6] Yuki Kanno and Ryohei Banno. A comparative study of estimation of video viewer emotion using youtube video comments. In **TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON)**, pp. 1030–1033, 2023.

[7] 菅野祐希, 坂野遼平. オンライン動画サービスにおける bert 及び gpt-3.5 を用いた視聴者感情の推定. 言語処理学会 第 30 回年次大会, pp. 1067–1072, 2024.

[8] Mohammad Soleymani, Maja Pantic, and Thierry Pun. Multimodal emotion recognition in response to videos. **IEEE Transactions on Affective Computing**, Vol. 3, No. 2, pp. 211–223, 2012.

[9] BARRON's, Tiernan Ray. Youtube's 2 billion videos, 197m hours make it an 'immense' force, says bernstein.

[10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.
[11] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners, 2020.
[12] OpenAI. Gpt-4 technical report, 2024.
[13] Anthropic. Introducing the next generation of claude anthropic.
[14] Anthropic. Introducing claude 3.5 sonnet anthropic.
[15] Gemini Team. Gemini: A family of highly capable multimodal models, 2024.
[16] Gemini Team. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context, 2024.
[17] DeepSeek-AI. Deepseek llm: Scaling open-source language models with longtermism, 2024.
[18] Llama team. The llama 3 herd of models, 2024.
[19] ELYZA, Inc. 「gpt-4」を上回る日本語性能の llm「llama-3-elyza-jp」を開発しました | elyza, inc.
[20] 田中邦朋, 笹野遼平, 武田浩一. 大規模言語モデルに含まれる社会集団間の感情の抽出. 言語処理学会 第30回年次大会, 2024.
[21] Jen tse Huang, Man Ho Lam, Eric John Li, Shujie Ren, Wenxuan Wang, Wenxiang Jiao, Zhaopeng Tu, and Michael R. Lyu. Emotionally numb or empathetic? evaluating how llms feel using emotionbench, 2024.
[22] 堺雄之介, 竹内幹太, 伊東栄典. コメントを利用した炎上動画検出に関する検討. 情報処理学系研究報告, Vol. Vol.2021-ICS-203, No. 9, 2021.
[23] SOCIOCOM Social Computing Laboratory. 日本語感情表現辞書 jiwc-dictionary.

A　7 次元感情ベクトルの推定精度評価用動画の感情ベクトル

表 3　正解データの感情ベクトル

感情	sad	anxiety	angry	disgust	trust	surprise	joy
動画 1	0.1487	0.2588	0.1197	0.1352	0.5465	0.6083	0.4577
動画 2	0.7051	0.0673	0.0972	0.0747	0.4709	0.2791	0.4285
動画 3	0.0621	0.1002	0.1623	0.1599	0.2458	0.5489	0.7566

B　最大感情の推定精度評価用動画の感情ラベル

表 4　選定動画のラベル

感情	sad	anxiety	angry	disgust	trust	surprise	joy
動画数	4	4	5	4	4	4	5