

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

菅野祐希¹ 坂野遼平¹

¹ 工学院大学大学院 工学研究科 情報学専攻

em23019@ns.kogakuin.ac.jp banno@cc.kogakuin.ac.jp

概要

近年のインターネットやスマートフォンの普及により、動画共有サービスに多くの人々がアクセスするようになった。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、視聴者とマーケティングの両方において有益となる。本研究では過去に行った動画のコメントから視聴者の感情を推定する手法の拡張として、BERT と GPT-3.5-turbo を利用した動画視聴者の感情推定を行った。結果として、単純な感情であれば BERT が向いており、複数の感情が混在したり推定の難易度が高い感情においては GPT-3.5-turbo が優勢となった。

1 はじめに

近年、スマートフォン等の普及が急速に進み、世帯保有率は 90.1% (2022 年時点) と高い数値を示している [1]。また、同時にインターネットの普及も進んでおり、スマートフォンからの個人による利用率が 71.2% となっている [1]。これらの要因から、インターネットはメジャーな情報媒体となっており、インターネット利用率がテレビ利用者率を 5 年以上連続で上回る結果となっている。このことから、インターネットは幅広い世代に浸透していることが分かる。中でも YouTube¹⁾ のアクティブユーザ数は 2022 年 1 月時点で全世界で 25 億 6200 万人となっている [1]。日本国内でもユーザ数が 7000 万人を突破し、日本でも非常に人気の動画共有サービスである [2]。YouTube 上には 2016 年の時点で 20 億本の動画が存在しているという調査結果が存在している [3]。3500 億円の経済効果や 10 万人相当の雇用を生み出したという調査結果も存在する [4]。

1) <https://www.youtube.com>

近年では動画共有サービスはビジネスやマーケティングに利用されることも多く、YouTube では広告や有名な YouTuber に対する案件としてマーケティングが行われている [5]。

動画が人の感情に与える影響は大きい。動画が感情を動かすのに適したメディアであることは Gross ら [6] によって実証されている。YouTube 上に「笑える動画」「泣ける動画」等のコンテンツも存在していることから、YouTube 上の動画が感情を動かすためのツールとして利用されている事が分かる。

我々は動画から得られる感情についての研究を行ってきた [7][8]。動画から得られる感情を取得することにより、関連動画の精度の改善等のユーザビリティの向上に用いることが出来る。しかし、動画のサムネイルやタイトルから感情を得ることは難しい。多くのユーザからの感想や意見から一般性の高い印象を取得し、利用することが出来れば客観的な印象を取得できると考え、動画のコメントを利用する。本稿ではその研究の差分として、BERT によるより良い推定の手法の検討と、大規模言語モデルである GPT-3.5 を利用した推定について提案する。

2 関連研究

堺らの研究 [9] 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²⁾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。また、Word2Vec に代表される単語の分散表現を取得する

2) <https://www.nicovideo.jp>

手法を利用することも提案している。

中澤らの研究 [10] では、BERT による感情推定の手法とその有効性について精度を確かめている。Wikipedia を利用して学習を行った BERT モデルを利用している。読売新聞のニュースを短文に分けたものを Google Forms で 7 段階の尺度でポジティブ・ネガティブを評価させ、500 個の正解ラベル付きの教師データを取得している。これらを用いてファインチューニングを行った BERT のモデルを利用して感情の値を判定している。比較として、構文解析を用いた感情値判定と形態素解析による動詞および修飾部の極性値による感情極性推定を行っている。

以上の研究ではポジティブ・ネガティブの評価がメジャーであり、複数の感情の詳細な推定は行っていない。本研究では、複数の感情についてのマルチラベル分類による推定を行っている。

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) [11] 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。感情値は 7 次元のベクトルとして算出する。提案手法の全体図を図 1 に示す。

3.1 学習用データの収集

本研究では、クラウドソーシングを利用したデータの収集を行っている。150 人を対象に、7 つの感情を想起させる動画 ID を感情ごとに 1 つずつ収集した。1 人につき 7 本の動画 ID を収集し、合計で 1050 本分の動画 ID を収集し、利用した。その後、Google 社の提供する YouTube Data API v3³⁾ を利用し、動画からコメントデータを取得している。

BERT にコメントを入力する際、入力可能なトークン数に上限が存在する。そのため、コメントデータ取得の際、関連度順で上位 100 件のコメントを取得するように設定を行っている。取得したコメントデータは前処理を行い、BERT に入力できる形式に整える。

3.2 BERT による推定

本研究では、事前学習済み BERT モデルとして、東北大学乾研究室の提供している事前学習済みモデル [12] を利用している。感情名のリストを作成し、フォルダ名及びテキストファイル名に感情名を付加

する。各感情の傾向をファインチューニングするために、各感情名の付加されたテキストファイルからファインチューニングを行うように設定している。

ファインチューニングを行った BERT に推定を行いたい動画のコメントを入力する。各感情を想起させる動画のコメントと入力したコメントがどれだけ近いかを出力する。

3.3 GPT-3.5-turbo による推定

BERT と比較を行う対象として、大規模言語モデルである GPT-3.5-turbo [13] を使用する。OpenAI の API を利用し、GPT-3.5-turbo を用いて推定を行いたい動画のコメントを直接入力する。動画のコメントは 1 文ずつ入力し、プロンプトを用いて各感情ごとに 0 から 4 の 5 段階で評価を行わせる。コメント内に含まれると予測される感情が強いほど数値が大きくなる。1 動画につき 100 コメントを入力し、全てのコメントに付加された感情の値の平均を感情値とする。入力には以下のプロンプトを用いる。

ソースコード 1 感情推定プロンプト

- ```
1 {"role": "system", "content": "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けます。"},
2 {"role": "system", "content": "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。"},
3 {"role": "system", "content": "[悲しい],[不安],[怒り],[嫌悪],[信頼],[驚き],[嬉しい]"},
4 {"role": "system", "content": "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。"},
5 {"role": "system", "content": "感情が強くな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。"},
6 {"role": "user", "content": "今日は新発売のケーキを買った。最高の一日だ。"},
7 {"role": "assistant", "content": "[悲しい:0][不安:0][怒り:0][嫌悪:0][信頼:0][驚き:1][嬉しい:4]"},
8 {"role": "user", "content": "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。"},
9 {"role": "assistant", "content": "[悲しい:4][不安:2][怒り:1][嫌悪:0][信頼:0][驚き:0][嬉しい:0]"},
10 {"role": "user", "content": "text"}
```

3) <https://developers.google.com/youtube/v3>

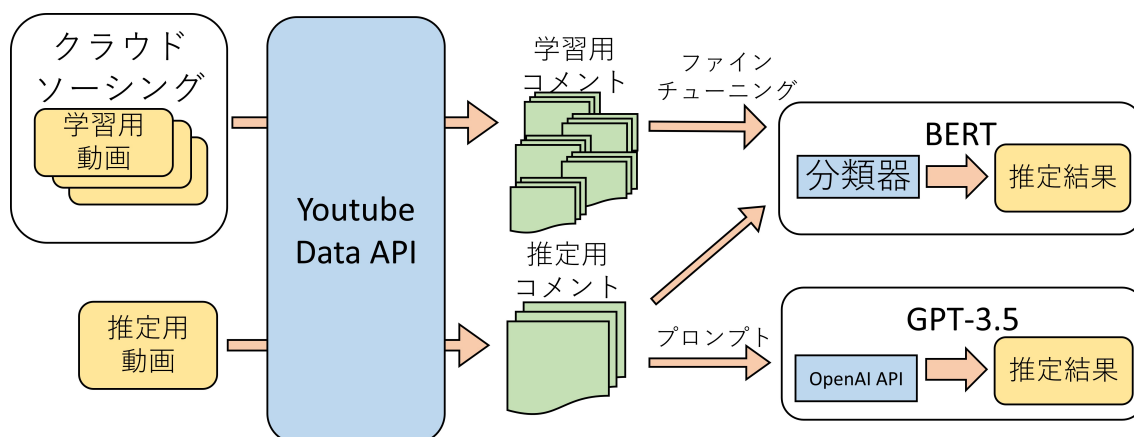


図1 提案手法概要図

最後の text には推定用のコメントを1文ずつ入力する。

## 4 評価手法

### 4.1 評価用データ

評価に用いる動画として、以下の3本の動画を選定した。3つの動画は異なる感情を与えると予想したものを選定した。

感情推定の精度を検証する際に、以下の3つの動画についてクラウドソーシングを利用し100名にアンケートを行った。各動画を視聴した際に各感情がどの程度想起されたかを0から4の5段階評価で評価を行わせ、アンケートの結果を平均したものを感情値とした。

- 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)<sup>4)</sup>
- 動画2: back number - 手紙 (full)<sup>5)</sup>
- 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】<sup>6)</sup>

### 4.2 BERT による推定の評価

事前学習済み BERT にファインチューニングを行い、分類器として使用した。ファインチューニングごとにわずかに分類性能が変化するため、ファインチューニングを行い動画のコメントを入力し、最終的な感情の値を算出する手順を10セット行った。最終的にそれらの平均を BERT による感情推定の値として扱う。

本研究では、ファインチューニングのパラメータを変化させてそれらの精度の変化を比較した。トークン数とファインチューニングに用いるテキストファイルの個数を変化させている。トークン数は256トークンと512トークン、テキストファイルの個数は各感情50本(合計350本)と各感情150本(合計1050本)で変化させている。以下では「(トークン数) - (1感情ごとのテキストファイル数)」と表記する。256トークンで、1感情ごとに50ファイルを学習させている場合は「256-50」と表記する。

### 4.3 GPT-3.5-turbo による推定の評価

評価用動画のコメントを100文取得し、1列100行のCSVファイルとして出力する。提案手法に示したプロンプトに1文ずつ入力し、出力された感情値を平均したものを GPT による感情推定の値として扱う。

### 4.4 類似度の比較

出力された感情ベクトルは形式が統一されていないため、以下の数式を利用して単位ベクトル化を行う。 $\vec{E}mo$  は動画の感情ベクトル、 $\vec{E}mo_n$  は動画の感情ベクトルの特定の成分を示している。

$$\text{単位ベクトル成分} = \frac{\vec{E}mo_n}{|\vec{E}mo|} \quad (n = 1, 2, 3, \dots, 7)$$

$$|\vec{E}mo| = \sqrt{\sum_{n=1}^7 \vec{E}mo_n^2}$$

求めた単位ベクトルを利用し、コサイン類似度を用いてベクトルの類似度を比較した。

4) <https://youtu.be/uJBHO2NXvxQ>

5) <https://youtu.be/woRV5VxJDkU>

6) <https://youtu.be/OtJvFyqqeo0>

表 1 各手法とアンケート結果のコサイン類似度

| 動画   | BERT   |        |               |         | GPT-3.5       |
|------|--------|--------|---------------|---------|---------------|
|      | 256-50 | 512-50 | 256-150       | 512-150 |               |
| 動画 1 | 0.8340 | 0.8061 | 0.7621        | 0.7376  | <b>0.9561</b> |
| 動画 2 | 0.8609 | 0.8539 | <b>0.9035</b> | 0.8885  | 0.8586        |
| 動画 3 | 0.8896 | 0.8359 | <b>0.9189</b> | 0.8901  | 0.8665        |

## 5 結果・考察

### 5.1 結果

コサイン類似度によるアンケート結果との比較を表 1 に示す。動画ごとに最も類似度の高い手法を太字で示している。また、各動画、各手法ごとの単位ベクトルの成分を補足の表 2, 3, 4 に示す。

動画 1 は GPT-3.5-turbo が最も高く、動画 2, 動画 3 においてはトークン数 256, テキストファイル数が各感情 150 個 (合計 1050) 個でファインチューニングを行った BERT が最も高いという結果になった。

### 5.2 BERT による推定

動画 1 においてはテキストファイル数が 350 本の BERT が成績が良く、動画 2 と 3 においてはテキストファイル数が 1050 本の BERT の方が成績が良い。補足の表 2 のアンケート結果に示される通り、動画 1 は突出して高い感情が無い。「信頼」「驚き」「喜び」の 3 つの感情を人に強く与えることが分かる。

テキストファイル数を増加させることによって、強いと考えられる感情をより強く、弱いと考えられる感情をより弱く判定する傾向が存在する。テキストファイルを 350 本分学習した結果と比較しても、1050 本分学習した結果の方が両極端な結果となっていることが分かる。また、トークン数を増加させると精度が下がる。

動画 2 や動画 3 のような、強く感じられる感情が単一であり、「悲しみ」や「喜び」などの特にメジャーでわかりやすい感情を強く与えられる動画に関しては性能が高いと考えられる。

### 5.3 GPT-3.5-turbo による推定

動画 1 において非常に高い精度を出したが、それ以外の動画では比較的低い精度を出している。動画 1 では表 2 にも示されている通り、「信頼」「驚き」「喜び」の複雑な 3 つの感情全てが高いという事を推定している。

対して表 3 に示される通り、動画 2 では「悲しみ」が非常に低い。実際のコメントを確認したところ、以下のようなコメントが散見された。

- 好きな人や恋人ではない大事な人のことも歌にしてくれる、本当に大好きです
- 自分の中でこの曲が一番泣ける歌で、この曲聴いて親のありがたみを再認識する。
- 自分の親がどれだけ大切な存在か改めて知れた素敵な曲

動画 2 の楽曲が家族愛を歌ったものであり、コメントにも家族愛に関する内容が多く散見された。GPT はこれらの詳細な感情を読み取り、「信頼」や「喜び」を強く感じていることが分かる。

GPT は単純な感情を読み取るよりも、より複雑な感情が含まれる動画の感情推定に向いている可能性が高いと考えられる。

## 6 おわりに

本研究では、オンライン動画共有サービスの動画コメントから視聴者がどのような感情を得たかを推定する手法を提案した。BERT を 1050 本の動画でファインチューニングし、分類器とする手法と、GPT-3.5-turbo を利用しコメントから直接感情を推定する手法について説明した。

結果として、2 つの感情や、「喜び」「悲しみ」のような与える感情が分かりやすい動画は BERT が優れた精度を示した。対照的に、3 つ以上の感情などが複雑に現れる動画は GPT-3.5-turbo が優れた精度を示した。

今後は他の大規模言語モデルや自然言語処理モデルに関しても同様に感情推定の精度を測定することを考えている。また、動画に含まれる詳細な感情ベクトルの推定とは別に、他の評価指標を設定することによって感情推定の精度を測定することを考えている。

## 参考文献

- [1] 総務省. 令和 5 年版情報通信白書, (2023 年 7 月).
- [2] Think with Google. Youtube が人気な理由とは, (2023 年 12 月 20 日閲覧). <https://www.thinkwithgoogle.com/intl/ja-jp/marketing-strategies/video/youtube-recap2022-1/>.
- [3] Tiernan Ray BARRON's. Youtube' s 2 billion videos, 197m hours make it an 'immense' force, says bernstein, (2023 年 12 月 22 日閲覧). <https://www.barrons.com/articles/youtubes-2-billion-videos-197m-hours-make-it-an-immense-force-says-bernstein-1462978280>.
- [4] YouTube Oxford Economics. Youtube impact report 2021 年日本における youtube の経済的・社会的・文化的影響, (2023 年 12 月 22 日閲覧). <https://services.google.com/fh/files/blogs/jpyoutubeimpactreport2021.pdf>.
- [5] YouTube Advertising. オンライン動画広告キャンペーン - youtube 広告, (2023 年 11 月 28 日閲覧). [https://www.youtube.com/intl/ALL\\_jp/ads/](https://www.youtube.com/intl/ALL_jp/ads/).
- [6] Robert W.Levenson and J.Gross . Emotion elicitation using films. **IN COGNITION AND EMOTION**, pp. 87–108, 1995.
- [7] 菅野祐希, 坂野遼平. Youtube コメントを用いた動画の感情推定. In **DEIM Forum 2023**, 2023.
- [8] 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**, 2023.
- [9] 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.
- [10] 中澤政孝, 亀井且有, 前田陽一郎, クーパーエリック. Bert を用いた短文の感情極性推定手法の提案とその有効性. 第 36 回ファジィシステムシンポジウム講演論文集 (FSS2020 オンライン), pp. 177–180, 2020.
- [11] SOCIOCOM Social Computing Laboratory since 2015. 日本語感情表現辞書 jiw-c-dictionary, (2023 年 12 月 22 日閲覧). <https://sociocom.naist.jp/jiw-c-dictionary/>.
- [12] 東北大学乾研究室. Pretrained japanese bert models released / 日本語 bert モデル公開, (2023 年 12 月 22 日閲覧). <https://www.nlp.ecei.tohoku.ac.jp/news-release/3284/>.
- [13] OpenAI. Models - openai api, (2023 年 12 月 19 日閲覧). <https://platform.openai.com/docs/models/overview>.

## A 結果詳細

各手法による動画ごとの感情ベクトルを以下に示す。感情が強いほど数値が高く、感情が弱いほど数値が低い。

表2 感情値の単位ベクトル成分 (動画1)

| 手法           | 悲しみ    | 不安     | 怒り     | 嫌悪     | 信頼     | 驚き     | 喜び     |
|--------------|--------|--------|--------|--------|--------|--------|--------|
| アンケート        | 0.1487 | 0.2588 | 0.1197 | 0.1352 | 0.5465 | 0.6083 | 0.4577 |
| BERT256-50   | 0.2613 | 0.4541 | 0.3655 | 0.3971 | 0.5164 | 0.3850 | 0.1392 |
| BERT512-50   | 0.2550 | 0.4980 | 0.3722 | 0.4037 | 0.5029 | 0.3417 | 0.1255 |
| BERT256-150  | 0.1784 | 0.5816 | 0.2734 | 0.5094 | 0.3940 | 0.3585 | 0.1074 |
| BBERT512-150 | 0.2247 | 0.5701 | 0.3147 | 0.5155 | 0.3856 | 0.3185 | 0.0978 |
| GPT-3.5turbo | 0.0227 | 0.1527 | 0.1885 | 0.2210 | 0.4615 | 0.5186 | 0.6402 |

表3 感情値の単位ベクトル成分 (動画2)

| 手法           | 悲しみ    | 不安     | 怒り     | 嫌悪     | 信頼     | 驚き     | 喜び     |
|--------------|--------|--------|--------|--------|--------|--------|--------|
| アンケート        | 0.7051 | 0.0673 | 0.0972 | 0.0747 | 0.4709 | 0.2791 | 0.4285 |
| BERT256-50   | 0.6312 | 0.3186 | 0.3508 | 0.3593 | 0.2430 | 0.2957 | 0.3184 |
| BERT512-50   | 0.6549 | 0.3463 | 0.3674 | 0.3115 | 0.2055 | 0.2762 | 0.3173 |
| BERT256-150  | 0.6036 | 0.3672 | 0.1953 | 0.2078 | 0.2329 | 0.3285 | 0.5073 |
| BBERT512-150 | 0.7296 | 0.3765 | 0.2398 | 0.2498 | 0.2253 | 0.2449 | 0.3087 |
| GPT-3.5turbo | 0.3158 | 0.0325 | 0.0000 | 0.0000 | 0.5873 | 0.1653 | 0.7260 |

表4 感情値の単位ベクトル成分 (動画3)

| 手法           | 悲しみ    | 不安     | 怒り     | 嫌悪     | 信頼     | 驚き     | 喜び     |
|--------------|--------|--------|--------|--------|--------|--------|--------|
| アンケート        | 0.0621 | 0.1002 | 0.1623 | 0.1599 | 0.2458 | 0.5489 | 0.7566 |
| BERT256-50   | 0.2308 | 0.2597 | 0.4037 | 0.3822 | 0.2633 | 0.3532 | 0.6133 |
| BERT512-50   | 0.1820 | 0.2317 | 0.5353 | 0.4141 | 0.2030 | 0.3234 | 0.5562 |
| BERT256-150  | 0.1673 | 0.1878 | 0.4332 | 0.3096 | 0.2819 | 0.3513 | 0.6711 |
| BBERT512-150 | 0.1656 | 0.1691 | 0.4792 | 0.3587 | 0.2773 | 0.3299 | 0.6325 |
| GPT-3.5turbo | 0.0030 | 0.0574 | 0.0416 | 0.0772 | 0.6119 | 0.2228 | 0.7515 |