

卒業論文

動画の属するジャンルやチャンネルとコメントの
関係性に着目した動画再生回数予測手法の提案

川上 大凱

2024年4月11日

岐阜大学 工学部 電気電子・情報工学科 情報コース
鈴木研究室

本論文は岐阜大学工学部に
学士（工学）授与の要件として提出した卒業論文である。

川上 大凱

指導教員：

鈴木 優 准教授

動画の属するジャンルやチャンネルとコメントの 関係性に着目した動画再生回数予測手法の提案*

川上 大凱

内容梗概

本研究では、動画のメタデータから再生回数を高い精度で予測する手法を提案する。動画はジャンルやチャンネルごとに存在し、同じジャンルやチャンネルであれば視聴者数の類似傾向が見られる。我々は動画の再生回数予測精度を向上させるために、ジャンルごとの視聴者数の違いやチャンネルの特色を考慮すべきではないかと考えた。ところが、それらのデータは明示的に与えられているわけではない。そこで本研究では、動画に付随するコメントにジャンルごとの視聴者数の違いやチャンネルの特色が現れ、ジャンルやチャンネルを判別可能な情報を保有するという仮説を立てた。仮説を検証することによって、コメントがジャンルごとの視聴者数の違いやチャンネルの特色を表すデータとして使えるのかどうかを調査した。仮説を検証するため、コメントを入力データとしたジャンル及びチャンネル分類を行った。実験の結果より、コメントがジャンルやチャンネルの情報を保有していることが分かった。また、ジャンルやチャンネルの情報を特徴量にすることによって再生回数予測精度の向上が期待できるという仮説を立てた。仮説を検証することによって、コメントが再生回数予測精度に影響を与えるかどうかに加え、コメントに含まれるジャンルやチャンネルを表す情報が再生回数予測精度に影響を与えるかどうかを調査した。仮説を検証するため、自作回帰ニューラルネットワークモデルによる2種類の再生回数予測実験を行った。実験の結果より、コメントに含まれるジャンルやチャンネルの情報が再生回数予測に影響を与えることが分かった。

キーワード

*岐阜大学 工学部 電気電子・情報工学科 情報コース 卒業論文, 学籍番号: 1203033058, 2024年4月11日.

ソーシャルメディア, YouTube, メタデータ, 再生回数予測, ニューラルネットワーク, 機械学習

目次

図目次	v	
表目次	vi	
第 1 章	はじめに	1
第 2 章	基本的事項	4
2.1	BERT	4
2.2	ニューラルネットワーク	4
2.3	評価指標	5
2.3.1	Accuracy	5
2.3.2	決定係数 (R^2)	5
2.3.3	Cross Entropy Loss	6
2.3.4	RMSE Loss(Root Mean Squared Error Loss)	6
2.3.5	F 値	6
2.4	k 分割交差検証	7
2.5	対応のある 2 標本 t 検定	7
第 3 章	関連研究	9
3.1	動画の再生回数予測に関連する研究	9
3.2	視聴者の反応と動画に対する印象の関連性を調査する研究	10
第 4 章	提案手法	12
4.1	コメントデータ	12
4.2	使用データ	14
4.3	使用モデル	14
4.3.1	BERT	15
4.3.2	回帰ニューラルネットワークモデル	16
第 5 章	評価実験	18

5.1	実験1：コメントを用いたジャンル分類・チャンネル分類	18
	5.1.1 実験手順	18
	5.1.2 結果・考察	19
5.2	実験2：コメントの有意性検証実験	21
	5.2.1 実験手順	21
	5.2.2 結果・考察	22
5.3	実験3：ジャンルやチャンネルの情報の有意性検証実験	24
	5.3.1 実験手順	24
	5.3.2 結果・考察	25
第6章	おわりに	27
6.1	まとめ	27
6.2	今後の課題	28
	謝辞	30
	参考文献	31
	発表リスト	33

目次

5.1	ベースモデル	22
5.2	モデル 1(コメント有・動画タグ無)・モデル 2(コメント無・動画タグ有)	23
5.3	モデル 3(ジャンルの予測ラベル)・モデル 4(チャンネルの予測ラベル)	25

表目次

5.1	実験 1：使用データ数	19
5.2	実験 1：テスト用データに対する Accuracy と Cross Entropy Loss	19
5.3	実験 1：ジャンル分類の混同行列	20
5.4	実験 1：ジャンル分類の F 値	20
5.5	実験 1：チャンネル分類 (ゲーム実況) の混同行列	20
5.6	実験 1：チャンネル分類 (アニメコント) の混同行列	21
5.7	実験 2：テスト用データに対する精度 (10 回平均)	23
5.8	実験 2： p 値	23
5.9	実験 3：テスト用データに対する精度 (10 回平均)	26
5.10	実験 3： p 値	26

第1章 はじめに

我々は、高い精度で人気の動画を推測することによって、投稿者が動画を作成するための大きな助けとなると考えた。そこで本研究では、高い精度で人気の動画を推測するため、再生回数を高い精度で予測する手法を提案する。

高い再生回数予測精度を得るため、我々は動画の再生回数に差が表れる原因を考慮する必要があると考えた。我々は動画の再生回数に差が表れる原因として、動画の内容以外にもジャンルごとの視聴者数の違いやチャンネルの特色が存在していると考えた。動画の視聴者数はジャンルによって差が存在する。ゲーム実況のような人気ジャンルは、バイクや車などの乗り物関連のジャンルと比較して視聴者数が多い*†。また、動画の視聴者数はチャンネルによっても差が存在する。同時期に類似した動画をアップロードしていたとしても、チャンネルごとに動画の再生回数は異なる。しかし、視聴者数の違いが表れる原因は明確化することが難しい。そのため、ジャンルごとの視聴者数の違いやチャンネルの特色を表すデータは直接収集することが現実的ではない。そこで我々は、ジャンルごとの視聴者数の違いやチャンネルの特色を表すため、動画に付随するコメントに着目した。

我々は二つの仮説を立てた。一つ目の仮説は、動画に付随するコメントがジャンルやチャンネルを判別可能な情報を保有しているということである。二つ目の仮説は、ジャンルやチャンネルの情報を特徴量にすることによって再生回数予測精度の向上が期待できるということである。

動画に付随するコメントがジャンルやチャンネルを判別可能な情報を保有しているという仮説を検証することによって、コメントがジャンルごとの視聴者数の違いやチャンネルの特色を表すデータとして使えるのかどうかを調査する。仮説を検証するため、実験1としてコメントを入力としてBERT[1]を用いたジャンル分類とチャンネル分類を行った。ジャンル分類では、ジャンルを独自に7種類設定した。コメントの収集期間を設定し、収集したそれぞれのコメントにジャンルラベルを付与したものを入力データとした。チャンネル分類では、一つのジャンルに対して10個のチャンネルを用意した。コメントの収集期間は設定せず、収集したそれぞれの

*出典：LINE リサーチ <https://lineresearch-platform.blog.jp/archives/39951598.html>

†出典：株式会社G Z ブレイン eb-i Xpress <https://www.kadokawa.co.jp/topics/2508/>

コメントにチャンネルラベルを付与したものを入力データとした。

実験の結果、ジャンル分類では Accuracy が 0.87149 という結果が得られた。チャンネル分類では Accuracy が最大で 0.6785 という結果が得られた。また、ジャンルによって分類精度に大きな差 (0.1844) が存在することが確認できた。以上より、コメントがジャンルやチャンネルを判断可能な情報を保有していることが確認できた。

ジャンルやチャンネルの情報を特徴量にすることによって再生回数予測精度の向上が期待できるという仮説を検証することによって、ジャンルやチャンネルの情報が再生回数予測精度に影響を与えるかどうかを調査する。また、コメントが再生回数予測精度に影響を与えるかどうかを調査する。ジャンルやチャンネルの情報を特徴量にすることによって再生回数予測精度の向上が期待できるという仮説を検証するため、実験 2 と実験 3 として自作の回帰ニューラルネットワークモデルを用いた 2 種類の再生回数予測実験を行った。ベースモデルとしてタイトル、チャンネル登録者数、動画高評価数を説明変数に持つモデルを作成した。

実験 2 では、コメントが再生回数予測精度に影響を与えるかどうかを調査した。ベースモデルに加えて、ベースモデルにコメントを加えたモデル、コメントの代わりに動画に付随する動画タグを加えたモデルを用意した。三つのモデルに対するテスト用データの予測精度を比較することによって、コメントが再生回数予測精度の向上に寄与するかどうかを調査した。10 分割交差検証を行い、有意水準を 10% として対応のある 2 標本 t 検定での片側検定を実施した。実験 3 で本実験での調査対象であるコメントを細分化した特徴を用いるため、我々は本実験での厳しい判定は必要ないと考え、有意水準を 10% と設定した。タイトルとコメント、動画タグは BERT を用いて 768 次元の文ベクトルとした。再生回数は動画単位で付与されるものであるため、コメントのベクトルは同動画内でのコメントのベクトルの平均とした。

実験の結果、コメントを説明変数に含むモデルの決定係数 (R^2) が 0.7470 であり、最も高い数値となった。次点で動画タグを説明変数に含むモデルが続き、決定係数 (R^2) は 0.7391 であった。ベースモデルの決定係数 (R^2) は 0.7264 で最も低い数値となった。 t 検定でも各モデルの評価値間に有意な差が見られた。以上の結果より、コメントが再生回数予測精度に影響を与えていることが確認できた。

実験3では、コメントに含まれるジャンルやチャンネルを表す情報が再生回数予測精度に影響を与えるかどうかを調査した。ベースモデルに加えて、説明変数に前述したジャンル分類で得られた予測ラベルを加えたモデル、そしてチャンネル分類で得られたチャンネルの予測ラベルを加えたモデルの三つを用意した。それぞれのモデルに対するテスト用データの予測精度を比較し、ジャンルやチャンネルが再生回数予測精度の向上に寄与するかどうかを調査した。10分割交差検証を行い、有意水準を5%として対応のある2標本 t 検定での片側検定を実施した。動画単位の学習を行わせるために、同動画内のコメントに対して付与された予測ラベルの中で最も付与された回数の多い予測ラベルを動画の予測ラベルとした。

実験の結果、予測ラベルを含むモデルの決定係数 (R^2) はベースモデルの決定係数 (R^2) と比較して最大で 0.0526 ほど高いという結果が得られた。また t 検定でも評価値間に有意な差が見られた。しかし実験1で得られた予測精度が低いジャンルにおいては、予測ラベルを含むモデルとベースモデルの評価値間に有意な差が見られないという結果が得られた。以上の結果より、精度向上にコメントが保有するジャンルやチャンネルの情報が影響を与えていることが確認できた。

本論文における貢献は以下の通りである。

- コメントにはジャンルとチャンネルの特徴が保有されていることを確認した。
- コメントを用いることによって再生回数予測精度が向上するということを確認した。
- ジャンルやチャンネルの情報が再生回数予測精度の向上に影響を与えていることを確認した。

本論文の構成は以下の通りである。2章では基本的事項について述べる。3章では関連研究について述べる。4章では提案手法について述べる。5章では行った実験について述べる。最後に6章では本論文のまとめと今後の課題について述べる。

第 2 章 基本的事項

2.1 BERT

BERT(Bidirectional Encoder Representations from Transformers) は Transformer の Encoder を用いた自然言語処理特化の深層学習モデルである。このモデルは双方向でのエンコード表現を実装しており、MLM(Masked Language Modeling) と NSP(Next Sentence Prediction) という二つの事前学習タスクを行うことによって、文脈を考慮した分散表現が可能となっている。また、BERT はラベルなしデータを用いた事前学習を行っている。そのため、様々なタスクに対してファインチューニングを行うことが可能であるという特徴を持つ。ファインチューニングを行う際には、BERT モデルに対して該当するラベル数分の出力層を加える。発表当時には 11 個の自然言語処理 (NLP) タスクで SoTA(State-of-the-Art) を叩き出した。主に文書分類タスクで用いられることが多い。

2.2 ニューラルネットワーク

ニューラルネットワークは人間の脳を模したネットワークを用いた機械学習手法の一つである。入力層、隠れ層 (中間層)、出力層の三階層で構成されており、結合する各層の間には重みというパラメータが存在している。目的の数値に出力結果を近づけられるよう、重みを更新していくことをニューラルネットワークの学習と呼ぶ。学習には誤差逆伝播法が用いられ、重みの更新には勾配降下法が用いられる。また非線形なデータに対応できるよう、各隠れ層では活性化関数が適用されている。隠れ層の数を増やすことによって、結果の表現力は向上する。隠れ層の数や学習率といった自由に操作できるパラメータをハイパーパラメータと呼ぶ。

2.3 評価指標

2.3.1 Accuracy

Accuracy とは分類タスクで用いられることが多い評価指標である。本稿では多クラス分類で用いられているため、以下の式で定義される。

$$Accuracy = \frac{1}{n} \sum_k TL_k$$

k はラベル、 TL_k はあるラベル k に対して予測ラベルと正解ラベルが一致しているデータの個数、 n はデータの総数である。Accuracy はモデルが予測したラベルがどの程度正解ラベルと同じだったのかを表す評価値である。しかしデータのラベルに偏りが存在する場合、うまく機能しないという問題点が存在する。本稿では分類を行う際にアンダーサンプリングを実施することによって、各ラベルのデータ数を合わせている。そのため Accuracy の問題点は解消されており、評価指標として信頼できるものとなっている。また、Accuracy は 0 から 1 の間の値を取る。

2.3.2 決定係数 (R^2)

決定係数とは回帰タスクで用いられることが多い評価指標であり、以下の式で定義される。

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

N はデータ数、 y_i は i 個目の正解値、 \hat{y}_i は i 個目の予測値、 \bar{y} は平均値である。決定係数は予測した値がどの程度正解値に当てはまるのかを表す評価値である。0 から 1 の間の値を取る。

2.3.3 Cross Entropy Loss

CrossEntropyLoss とは分類タスクで用いられることが多い評価関数であり，以下の式で定義される．

$$CrossEntropyLoss = - \sum_k y_k \log \hat{y}_k$$

k はラベル， y は正解値， \hat{y} は予測値である．

2.3.4 RMSE Loss(Root Mean Squared Error Loss)

RMSELoss とは回帰タスクで用いられることが多い評価関数であり，以下の式で定義される．

$$RMSELoss = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$$

n はデータ数 (バッチ数)， y は正解値， \hat{y} は予測値である．この評価関数では予測値と正解値の誤差に対して二乗を取る．そのため細かい誤差も大きく考慮することができるが，その分外れ値の影響を受けやすいという特徴がある．

2.3.5 F 値

F 値はモデルの予測がどの程度バランスよく行われているのかを表す評価指標である．本稿では多クラス分類で用いられたため，あるラベル A に対しての F 値について述べる．F 値は以下の式で定義される．

$$Fscore = \frac{2 \cdot Recall_A \cdot Precision_A}{Recall_A + Precision_A}$$

また，F 値は 0 から 1 の間の値を取る．

$Recall_A$ は正解ラベル A に対しての予測ラベルが，どの程度 A を再現することができたのかを表す指標であり，以下の式で定義される．

$$Recall = \frac{TL_A}{T_A}$$

TL_A は予測ラベルと正解ラベルが A で一致しているデータの個数, T_A は正解ラベルを A としているデータの総数である.

また, $Precision_A$ は予測ラベル A に対しての正解ラベルが, 実際にどの程度 A であったのかを表す指標であり, 以下の式で定義される.

$$Precision_A = \frac{TL_A}{P_A}$$

TL_A は予測ラベルと正解ラベルが A で一致しているデータの個数, P_A は予測ラベルを A としているデータの総数である.

2.4 k 分割交差検証

交差検証とは, モデルの汎化性能を評価するために用いられる手法の一つである. 交差検証を実施する際, データセットを k 個に分割するものを k 分割交差検証と呼ぶ. 本稿で実施した k 分割交差検証では $k = 10$ としている.

本稿では分割した十個のデータセットの内, 一個をテスト用データ, 一個を検証用データ, そして残りの八個を訓練用データとした. データセットを割り振った後, テスト用データを交差するよう入れ替えていき, 全 10 回分の試行に対して得られた評価値の平均を取ることによってモデルの汎化性能を評価した.

2.5 対応のある 2 標本 t 検定

対応のある 2 標本 t 検定とは, 対応関係にある二つの母集団から抽出した標本に対して, 母平均の差が存在するかどうかを調査するための検定方法である. 帰無仮説は「二つの母集団の平均に差がない」である. 検定統計量は以下の式で定義される.

$$t = \frac{\sqrt{M}(\bar{X} - \mu)}{s}$$

M はデータ数, \bar{X} は二つの標本データに対する差の平均, μ は二つの母集団の平均に対する差, s は二つの標本データに対する差の標準偏差である. μ は帰無仮説が「二つの母集団の平均に差がない」であるため, 0 が代入される. t 検定では上記の検定統計量から導かれる p 値に対して, 有意水準を満たすかどうかを調査する.

また、 t 検定に用いるデータは正規分布に従っている必要がある。本稿の実験では、同一データに対して特徴量を加える前後で検定を行うため、対応のある 2 標本 t 検定を用いた。本稿では、各評価値に対して片側検定を実施した。

第3章 関連研究

3.1 動画の再生回数予測に関連する研究

動画の再生回数予測に関連する研究は現在までに複数行われている。Abishevaら [2] は動画作成から Twitter による動画の拡散までの速度に注目し、拡散速度と最終的な動画人気との関係を調査をした。Yu ら [3] は Twitter 由来の特徴量を用いることによって、動画再生回数の急激な増加とアップロード直後の再生回数の増加を予測する研究を行った。これらはどちらも動画に対して行われた Twitter での拡散に着目し、それを要因とした動画再生回数との関係性を調査するものであった。本稿では動画投稿サイト内で取得できる特徴のみを使用することによって、投稿者と視聴者以外が介入する可能性を排除する。投稿者と視聴者以外が介入する可能性として、本研究では動画投稿サイト内で動画を視聴せずに Twitter のタイムライン上に流れてきたものをそのまま拡散するといった場合を挙げる。その可能性を排除することで、我々は再生回数予測に対するノイズとなり得るデータが混入するリスクを削減することができる考えた。

Zhou ら [4] は、関連動画推薦や検索アルゴリズムなどの視聴ユーザの動画への到達方法と、動画に対する視聴パターンの関連性について研究を行った。これは Twitter だけでなく、動画投稿サイト内での拡散方法についても着目していた。YouTube における関連動画推薦機能は動画検索アルゴリズムや Twitter での拡散と異なり、再生回数が低いマイナーな動画に対しても行われていることを確認した。動画の持続的な視聴には関連動画推薦と検索アルゴリズムによる誘導が有効であることを確認した。柴田ら [5] は、ニコニコ動画における動画再生回数の推定を行った。音楽関連の動画は線形回帰分析の結果、高い推定精度を確認した。また、音楽関連の動画は再生回数とマイリスト登録数の相関が他のジャンルよりも高いことを確認した。動画についての動画タグを使用したクラスタリングと MLP による回帰分析を組み合わせた結果、推定精度が向上した。本稿では関連動画推薦や検索アルゴリズムでは考慮しきれない動画の特徴を考慮する方法として、動画タグではなくコメントを用いる。また関連動画推薦や検索アルゴリズムと再生回数の関係性を考慮するため、推薦や検索アルゴリズムに用いられる動画タイトルを特徴量として

用いる。また、マイリスト登録数の代わりにチャンネル登録者数を使用する。

3.2 視聴者の反応と動画に対する印象の関連性を調査する研究

視聴者の反応と動画に対する印象の関連性を調査する研究も現在までに複数行われている。山本ら [6] は楽曲動画の印象推定を視聴者のリアルタイムコメントを用いて行った。サビ区間で複数出現する形容詞や形容動詞を繰り返すコメントに着目し、楽曲の印象を推定した。歌詞や音響特徴量を用いた印象推定と比較し、視聴者のリアルタイムコメントが印象推定に有意な特徴であることを確認した。土屋ら [7] は形容詞や形容動詞以外の品詞にも注目することによって、より詳細な楽曲動画の印象推定手法を提案した。メディアや印象タイプごとに適切な品詞を用いることによって、より詳細な印象推定が可能であるということが確認された。また品詞を考慮せずコメント全文を用いることによって、メディアや印象タイプに関係なくある程度の印象推定が可能であることが確認された。村上ら [8] はリアルタイムコメントに基づく動画検索ランキングの手法を検討した。使用データセットはニコニコ動画で収集可能なメタデータセットであった。再生回数とマイリスト登録数、コメント数に相関があることを確認した。また、コメント付与者の面白さ尺度とコメント付与人数を考慮したランキング手法を検討した。面白さ尺度はあるシーンに対してのコメント数であった。堺ら [9] は YouTube 動画におけるメタデータを使用した炎上動画分類を行った。動画に対する低評価数とコメント数が多いものを炎上動画と仮定した。また、コメントに対して感情分析を行った結果を感情スコアとした。動画のベクトル化には感情スコアと再生回数、コメント数、高評価数を使用した。LightGBM を用いた分類器に全ての動画ベクトルを入力した際、最も高い分類精度を確認した。Siersdorfer ら [10] は YouTube 上のコメントに対する様々な分析を行った。コメントに対する感情分析を行い、その結果を用いることによってコメントがコミュニティに受け入れられるものかどうかを自動的に判断できることを確認した。またコメントとユーザによって反応が変わる動画の間には相関関係を確認した。実験結果より、コメントで表現された感情とコミュニティによるコメント評価、そして動画のトピックに強い依存関係を確認した。本稿ではコメントを品詞で制限を加えることなく全文で使用する。そのため我々は、動画に対する視聴者の印

象に考慮できないものが存在する可能性は低いと考える.

第4章 提案手法

本稿では二つの仮説を立てた。一つ目の仮説は、動画に付随するコメントがジャンルやチャンネルを判別可能な情報を保有しているということである。二つ目の仮説は、ジャンルやチャンネルの情報を特徴量にすることによって再生回数予測精度の向上が期待できるということである。

動画に付随するコメントがジャンルやチャンネルを判別可能な情報を保有しているという仮説の検証のため、動画の属するジャンルやチャンネルのラベルを付与したコメントを入力データとしたジャンル分類とチャンネル分類を行う。ジャンルの正解ラベルを l_{genre} 、チャンネルの正解ラベルを $l_{channel}$ とする。また分類によって得られたジャンルの予測ラベルを p_{genre} 、チャンネルの予測ラベルを $p_{channel}$ とする。

また、コメントが再生回数予測に影響を与えるのかを検証するため、コメントから生成した文ベクトルを説明変数に加えたモデルを用いて再生回数予測実験を行う。コメントに着目した理由については4.1節で詳しく述べる。

また、ジャンルやチャンネルの情報を特徴量にすることによって再生回数予測精度の向上が期待できるという仮説を検証するため、コメントに含まれるジャンルやチャンネルの情報を特徴量とした再生回数予測実験を行う。コメントに含まれるジャンルやチャンネルの情報を特徴量とするため、 p_{genre} と $p_{channel}$ を特徴量として用いる。

4.1 コメントデータ

我々は動画の再生回数予測精度を向上させるために、ジャンルごとの視聴者数の違いやチャンネルの特色を考慮すべきではないかと考えた。しかし、ジャンルごとの視聴者数の違いやチャンネルの特色をデータとして収集するためには二つの問題が存在する。一つ目はジャンルごとの視聴者数の違いやチャンネルの特色を表すデータが存在しないことである。そして二つ目は収集可能なメタデータには動画の内容に則していないものが存在することである。我々はこれらの問題を、コメントを用いることによって解消できると考えた。

一つ目の問題は、ジャンルごとの視聴者数の違いやチャンネルの特色を表すデータが存在しないことである。1章で述べた通り、ジャンルやチャンネルによって動画の再生回数に差が生まれることは明確である。そこで我々は動画の再生回数に差が表れる原因として、動画の内容以外にもジャンルごとの視聴者数の違いやチャンネルの特色が存在していると考えた。しかし、動画のデータを収集する際、ジャンルごとの視聴者数の違いやチャンネルの特色を直接表すデータは存在しない。そこで我々は、ジャンルごとの視聴者数の違いやチャンネルの特色を別のデータから推定することによって、その問題を解消できると考えた。

二つ目の問題は、収集可能なメタデータには動画の内容に則していないものが存在することである。メタデータとはコンテンツに付随する周辺データのことであり、動画のメタデータにはタイトルや動画タグ、概要欄、サムネイル画像、字幕の有無など多くのデータが存在している。しかし投稿者が自由に設定可能であるメタデータには、動画の内容を表していないものが存在することがある。例えば動画タグでは、他のコンテンツの視聴者を誘導するために、関係のないコンテンツの動画タグを設定している投稿者が存在する。また設定ミスにより、動画タグを一つも設定していない投稿者や他のカテゴリのタグを設定したままの投稿者も存在する。そういった形でメタデータが設定されている場合、そのメタデータは動画の属するジャンルの違いやチャンネルの特色を表現しているとは言えない。そこで我々は、動画の属するジャンルやチャンネルの特色を適切に表現するために、データが客観性を持つ必要があると考えた。

以上の問題点を解消するため、我々は不特定多数の視聴者が動画を見た上で感じたことや思ったことを書き残すものであるコメントに着目した。我々は、動画を見た上で書き残されるコメントには動画のジャンルの違いやチャンネルの特色に言及しているものが存在していると考えた。また我々は、不特定多数の視聴者によって残されるコメントには客観性が担保されていると考えた。

4.2 使用データ

使用するデータは YouTubeDataAPI* の Search 機能を用いて収集した。Search 機能はリクエストしたキーワードと関連する動画やチャンネルの情報を返す。本稿では、2023 年 1 月 1 日から同年 6 月 30 日までの動画を対象としてデータを収集した。ゲーム実況、キャンプ、アニメコント、ニュース、歌ってみた、踊ってみた、ネタ動画の 7 ジャンルを設定してデータの収集を行った。データ収集時に用いたキーワードは、それぞれ”ゲーム実況”, ”キャンプ”, ”アニメコント”, ”ニュース”, ”歌ってみた”, ”踊ってみた”, ”ネタ” である。収集したコメントには、そのコメントが属する動画の収集キーワードに対応する l_{genre} を付与した。そのため p_{genre} と l_{genre} は 0 から 6 の値を取る。

チャンネル分類における使用チャンネルは登録者数 10 万人以上であり、尚且つ収集データ内出現頻度がジャンルごとの上位 10 位以内であるチャンネルのみとした。出現頻度を考慮した理由は、登録者数のみでデータを収集した場合に該当ジャンルと関係ない動画が混入してしまうことを可能な限り防ぐためである。チャンネル分類に使用するデータのジャンルは、代表としてゲーム実況とアニメコントの二つを選定した。チャンネル分類を行うためのコメントを十分量確保するため、選定した 10 チャンネルの動画全てに対して期間を設けずコメントを収集した。収集したコメントには、該当する $l_{channel}$ を付与した。そのため $p_{channel}$ と $l_{channel}$ は 0 から 9 の値を取る。またチャンネル登録者数を s_n 、動画高評価数を f_n とする。 s_n と f_n は値を取りうる範囲が広く、重みの更新が上手くいかないことがある。そのため、 s_n と f_n は常用対数を取ることによって特徴量のスケールングを行う。

4.3 使用モデル

本研究では BERT と自作の回帰ニューラルネットワークモデルを用いて実験を行う。扱えるトークンの最大長を 512 とし、満たないものに関しては 0 での Padding 処理を実施する。バッチサイズを 32 とし、ミニバッチ学習を実施する。

*<https://developers.google.com/youtube/v3/docs?hl=ja>

検証用データの loss が 50 回連続で下がらなかった場合，earlystopping を実施する．各精度評価値は小数点第 5 位で四捨五入を行う．実験に使用するデータにはアンダーサンプリングを実施する．

4.3.1 BERT

ジャンル及びチャンネル分類と文字列データのベクトル化に自然言語処理を得意とする BERT[1] を用いる．複数文が存在する長文のコメントに対して精度の高い分類やベクトル化を行うためである．BERT の事前学習モデルは，東北大学自然言語処理研究グループの日本語学習モデル[†]を使用した．文字列のベクトル化では，BERT モデルの最終隠れ層の出力を用いる．入力文をトークン化し，トークンごとに生成された 768 次元の単語ベクトルの平均を取ったものを文ベクトルとする．本稿ではタイトルと動画タグ，コメントをベクトル化する．

タイトルを s_title ，トークン化されたタイトルの i 番目のトークンの単語ベクトルを v_i ，トークンの総数を N_{title_token} とする．また動画タグを s_tags ，トークン化された動画タグの j 番目のトークンの単語ベクトルを v_j ，トークンの総数を N_{tags_token} とする．タイトルのベクトル v_{s_title} と動画タグのベクトル v_{s_tags} は以下の式で定義される．

$$v_{s_title} = \frac{1}{N_{title_token}} \sum_{i=1}^{N_{title_token}} v_i$$
$$v_{s_tags} = \frac{1}{N_{tags_token}} \sum_{j=1}^{N_{tags_token}} v_j$$

しかし再生回数は動画単位で付与されるものであるため，コメントは動画に付随した全てのコメントの文ベクトルの平均を取った．コメントを c ，トークン化されたコメントの a 番目のトークンの単語ベクトルを v_a ，トークンの総数を N_c ， b 番目のコメントベクトルを v_b ，コメント総数を n_c とする．コメントのベクトル

[†]<https://github.com/cl-tohoku/bert-japanese>

$v_{c_pervideo}$ は以下の式で定義される.

$$v_{c_pervideo} = \frac{1}{n_c \cdot N_c} \sum_{b=1}^{n_c} \sum_{a=1}^{N_c} v_a$$

4.3.2 回帰ニューラルネットワークモデル

再生回数予測実験には回帰ニューラルネットワークモデルを用いる. モデルは入力層, 隠れ層 1 層, 出力層の 3 層構造とした. 複雑な特徴量を用いるためニューラルネットワークを使用した, 再生回数予測における特徴量の違いを確認するためのモデルであるため, 基本的な構造のモデルとした. 活性化関数には ReLU 関数を用いた. 隠れ層は 128 次元とした.

再生回数は動画単位で付与されるものであるため, 動画のジャンル予測ラベル p_{v_genre} は同動画内のコメント c に対して実験 1 で付与された p_{genre} の内, 付与回数が最も多い予測ラベルとした. p_{genre} が付与された c を c_p_{genre} , $p_{channel}$ が付与された c を $c_p_{channel}$ とする. p_{genre} が同動画内のコメントに付与された回数を $C(p_{genre})$ とすると, 動画のジャンル予測ラベル p_{v_genre} は以下の式で定義される.

$$p_{v_genre} = \max(C(p_{genre}))$$

同様に動画のチャンネル予測ラベル $p_{v_channel}$ は以下の式で定義される.

$$p_{v_channel} = \max(C(p_{channel}))$$

$\max()$ は最大値を求める関数である.

再生回数予測精度の比較を行うため, 5 種類のモデルを構築する. モデルは全て動画再生回数 V を出力とする. それぞれのモデルの説明変数を I とおくと, ベースモデルは, $I = [v_{s_title1}, v_{s_title2}, \dots, v_{s_title768}, s_n, f_n]$ とした. モデル 1 は, $I = [v_{s_title1}, v_{s_title2}, \dots, v_{s_title768}, s_n, f_n, v_{c_pervideo1}, v_{c_pervideo2}, \dots, v_{c_pervideo768}]$ とした. モデル 2 は, $I = [v_{s_title1}, v_{s_title2}, \dots, v_{s_title768}, s_n, f_n, v_{s_tags1}, v_{s_tags2}, \dots, v_{s_tags768}]$ とした. モデル 3 は, $I = v_{s_title1}$

, $v_{s_title2}, \dots, v_{s_title768}, s_n, f_n, p_{v_genre}$]とした。モデル4は, $I = [v_{s_title1}, v_{s_title2}, \dots, v_{s_title768}, s_n, f_n, p_{v_channel}]$ とした。

第 5 章 評価実験

本稿での実験目的は、二つの仮説を検証することである。一つ目の仮説は、動画に付随するコメントがジャンルやチャンネルを判別可能な情報を保有しているということである。二つ目の仮説は、ジャンルやチャンネルの情報を特徴量にすることによって再生回数予測精度の向上が期待できるということである。一つ目の仮説に対する実験を実験 1 とする。二つ目の仮説に対する実験を実験 2 と実験 3 とする。

5.1 実験 1：コメントを用いたジャンル分類・チャンネル分類

本実験では、動画に付随するコメントがジャンルやチャンネルを判別可能な情報を保有しているという仮説について検証する。コメントを入力とした BERT 分類タスクによって、コメントがジャンルやチャンネルを判別可能な情報を保有しているかどうかを調査した。

5.1.1 実験手順

BERT を用いたジャンル・チャンネル分類の実験手順を示す。

1. YouTubeAPI を用いてコメントを収集する。
2. 各コメントに対して正解のラベルを付与する。
3. BERT にコメントを入力し、ラベルを予測する。
4. テスト用データでの精度を確認する。

ジャンル分類の場合、2. と 3. はジャンルのラベルに対して行う。チャンネル分類の場合、2. と 3. はチャンネルのラベルに対して行う。データ数の詳細は表 5.1 で示す。データセットの割合は訓練用：検証用：テスト用の順に 8:1:1 とした。精度評価にはテスト用データにおける Accuracy と Cross Entropy Loss を用いた。

5.1.2 結果・考察

テスト用データに対する Accuracy と Cross Entropy Loss を表 5.2 に示す。なお、Cross Entropy Loss はバッチごとの Loss の平均を取ったものとした。

ジャンル分類における Accuracy はチャンスレート (0.143) を大きく超えている。加えて Cross Entropy Loss は 1 を切っている。また表 5.3 を確認すると、どのジャンルにおいても分類精度が高いことが分かる。よってコメントはジャンルを表現する情報を保有していると言える。また、ジャンルによって分類精度が変化するのかわを確認するため、ジャンルごとの F 値を用いて精度差を確認する。ジャンル分類の F 値を表 5.4 に示す。表 5.4 を確認すると、ほとんどのジャンルで F 値が 0.8 付近を取った。しかしアニメコトは他のジャンルと比較して大きく F 値が下がる。

チャンネル分類における Accuracy は、両者ともチャンスレート (0.1) を大きく超えている。また、表 5.5 と表 5.6 を確認すると、両者とも分類がしっかり行われていることが分かる。よってコメントはチャンネルを表現する情報を保有していると言える。しかし、ゲーム実況とアニメコトの間では Accuracy に大きな差 (0.1844) が存在する。ゲーム実況は Accuracy が 0.65 を超えているが、アニメコトの Accuracy は 0.5 を切っている。Cross Entropy Loss も Accuracy と同様に、

表 5.1 実験 1：使用データ数

チャンネル分類 (ラベル数 10)		ジャンル分類 (ラベル数 7)
ゲーム実況	アニメコト	
300,000	100,000	87,500

表 5.2 実験 1：テスト用データに対する Accuracy と Cross Entropy Loss

	チャンネル分類		ジャンル分類
	ゲーム実況	アニメコト	
Accuracy	0.6785	0.4941	0.8149
Cross Entropy Loss	0.9422	1.5215	0.5468

表 5.3 実験 1：ジャンル分類の混同行列

	pre 0	pre 1	pre 2	pre 3	pre 4	pre 5	pre 6
label 0(ゲーム実況)	548	40	35	6	57	12	2
label 1(キャンプ)	12	597	45	8	12	16	10
label 2(アニメコント)	12	41	555	35	13	25	19
label 3(ニュース)	4	25	51	596	5	9	10
label 4(歌ってみた)	35	30	28	7	581	13	6
label 5(踊ってみた)	20	31	46	8	33	550	12
label 6(ネタ動画)	6	31	61	6	16	14	566

表 5.4 実験 1：ジャンル分類の F 値

	ゲーム実況	キャンプ	アニメコント	ニュース	歌ってみた	踊ってみた	ネタ動画
F 値	0.8197	0.7987	0.7298	0.8726	0.8200	0.8215	0.8543

表 5.5 実験 1：チャンネル分類 (ゲーム実況) の混同行列

	pre 0	pre 1	pre 2	pre 3	pre 4	pre 5	pre 6	pre 7	pre 8	pre 9
label 0	2,189	40	80	138	175	40	73	119	82	64
label 1	22	2,310	122	98	52	193	24	49	81	49
label 2	46	131	1,889	128	56	468	41	48	131	62
label 3	94	94	127	2,043	108	129	36	56	198	115
label 4	170	60	75	150	2,163	44	30	188	55	65
label 5	22	143	333	85	37	2,206	24	23	92	35
label 6	70	51	131	81	36	129	2,344	72	62	24
label 7	173	77	89	152	237	73	60	1,993	82	64
label 8	81	122	194	220	87	196	32	54	1,735	279
label 9	129	141	129	290	199	116	29	84	400	1,483

ゲーム実況とアニメコントで大きな差 (0.5793) が生まれる結果となっている。混同行列でもアニメコントはゲーム実況と比較して、間違っただけの分類が行われていることが多い。以上の結果より、我々はアニメコントにおけるコメントに動画の内容の違いを表現するデータが少なかったのではないかと推測する。

アニメコントとゲーム実況の間に精度差はあったものの、動画に付随するコメントからジャンルやチャンネルを判別可能であることが確認できた。よって、動画に付随するコメントがジャンルやチャンネルを判別することができるだけの情報を保有しているということが確認できた。

表 5.6 実験 1：チャンネル分類 (アニメコメント) の混同行列

	pre 0	pre 1	pre 2	pre 3	pre 4	pre 5	pre 6	pre 7	pre 8	pre 9
label 0	554	18	39	32	59	51	85	35	57	70
label 1	88	390	38	39	103	45	172	26	31	68
label 2	66	17	384	24	109	63	89	68	78	102
label 3	81	18	65	434	66	46	57	52	54	127
label 4	61	40	46	26	556	110	70	24	22	45
label 5	69	34	38	22	110	515	68	41	50	53
label 6	23	30	18	22	42	55	717	7	34	52
label 7	54	21	101	39	96	39	88	382	51	129
label 8	41	14	42	32	42	37	99	34	544	115
label 9	84	15	51	33	65	54	85	57	91	465

5.2 実験 2：コメントの有意性検証実験

本実験では、ジャンルやチャンネルの情報を特徴量にすることによって再生回数予測精度の向上が期待できるという仮説について、一つ目の検証を行う。3種類のモデルを構築し、テスト用データの予測精度を比較することによって、コメントが再生回数予測精度の向上に影響を与えるかどうかを調査した。

5.2.1 実験手順

回帰ニューラルネットワークモデルを用いた再生回数予測実験の手順を示す。

- コメントの有意性検証実験
 1. チャンネル登録者数・動画高評価数をスケーリングする。
 2. BERT を用いてタイトルとコメント、動画タグからそれぞれ 768 次元の文ベクトルを生成する。
 3. 1. と 2. で用意した特徴量を説明変数としたベースモデル、モデル 1、モデル 2 を構築する。
 4. それぞれのモデルのテスト用データに対する精度を比較する。

それぞれのモデルの説明変数を I とおくと、ベースモデルは図 5.1 の通り、 $I = [v_{s_title1}, v_{s_title2}, \dots, v_{s_title768}, s_n, f_n]$ とした。モデル 1 は図 5.2 の通り、 $I = [v_{s_title1}, v_{s_title2}, \dots, v_{s_title768}, s_n, f_n, v_{c_pervideo1}, v_{c_pervideo2}, \dots, v_{c_pervideo768}]$ とした。モデル 2 は図 5.2 の通り、 $I = [v_{s_title1}, v_{s_title2}, \dots,$

, $v_{s_title768}$, s_n , f_n , v_{s_tags1} , v_{s_tags2} , ..., $v_{s_tags768}$]とした。

本実験は計 6,133 個の動画 (629,516 コメント) に対して行った。またデータセットは再生回数の偏りが無いよう、順番のシャッフルを実施した。データセットの割合は訓練用：検証用：テスト用の順に 8:1:1 とした。精度評価にはテスト用データにおける決定係数 (R^2) と RMSE Loss を用いた。特徴量の有意性を検証するため、10 分割交差検証を実施した。交差検証で得られた評価値は正規分布に従うものとし、有意水準を 10% として対応のある 2 標本 t 検定を実施した。実験 3 で本実験での調査対象であるコメントを細分化した特徴を用いるため、我々は本実験での厳しい判定は必要ないと考え、有意水準を 10% と設定した。帰無仮説を「2 群の評価値の母平均に差がない」とし、片側検定を行った。

5.2.2 結果・考察

実験結果を表 5.7 に示す。また、 t 検定についての結果を表 5.8 に示す。表 5.7 では、モデル 1 が最も高い精度であった。次点でモデル 2 が高い精度であった。ベースモデルは最も精度が低いという結果となった。

ベースモデルの精度が低い理由を、我々は説明変数の情報量が少ないためであるとする。表 5.8 を確認すると、ベースモデルと他の二つのモデル間での p 値は全

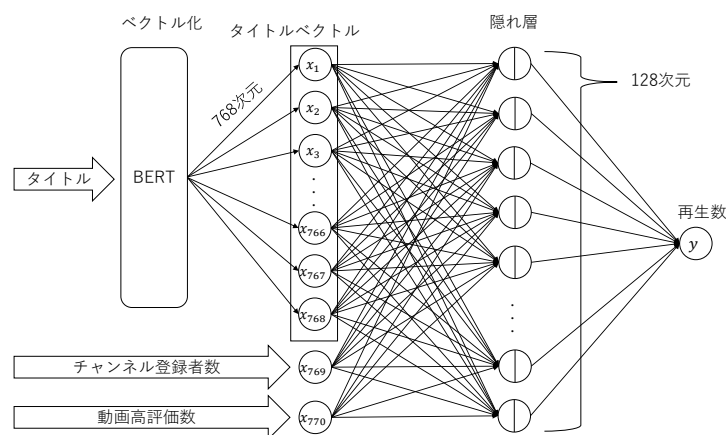


図 5.1 ベースモデル

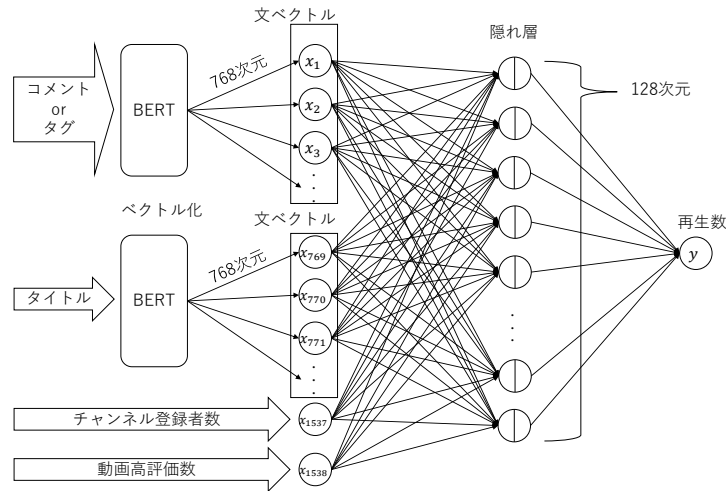


図 5.2 モデル 1(コメント有・動画タグ無)・モデル 2(コメント無・動画タグ有)

表 5.7 実験 2：テスト用データに対する精度 (10 回平均)

	ベースモデル	モデル 1	モデル 2
決定係数 (R^2)	0.7264	0.7470	0.7391
RMSE Loss	0.6927	0.6660	0.6769

表 5.8 実験 2： p 値

	ベースモデル		モデル 1 とモデル 2
	モデル 1	モデル 2	
決定係数	7.69978e-05	0.0078	0.0592
RMSE Loss	9.0836e-05	0.0076	0.0548

て 0.1 以下となっていることが分かる。よって帰無仮説は棄却され、ベースモデルと他の二つのモデルの評価値における母平均には有意な差があると言える。ベースモデルと他の二つのモデルの違いに着目すると、これらのモデルの違いは説明変数の情報量であることが分かる。ベースモデルは v_{s_title} と s_n , f_n の 3 つを説明変数としている。タイトルは 4.3.1 節で述べた通り 768 次元の文ベクトルとなっているため、ベースモデルにおける説明変数の次元は 770 次元となっている。一方で、他の二つのモデルは $v_{c_pervideo}$, または v_{s_tags} が追加されていることにより

1538 次元の説明変数を持つ。以上より、我々は特徴量の次元数が増えることで情報量が増加し、予測精度が向上すると考える。

次にモデル 1 とモデル 2 を比較する。表 5.8 を確認すると、二つのモデル間での p 値は全て 0.1 以下となっていることが分かる。よって帰無仮説は棄却され、モデル 1 とモデル 2 における評価値の母平均には有意な差があると言える。以上の結果より、コメントが再生回数予測に影響を与えていることが確認できた。

5.3 実験 3：ジャンルやチャンネルの情報の有意性検証実験

本実験では、ジャンルやチャンネルの情報を特徴量にすることによって再生回数予測精度の向上が期待できるという仮説について、二つ目の検証を行う。3 種類のモデルを構築し、テスト用データの予測精度を比較することによって、コメントに含まれるジャンルやチャンネルを表す情報が再生回数予測精度に影響を与えるかどうかを調査した。

5.3.1 実験手順

回帰ニューラルネットワークモデルを用いた再生回数予測実験の手順を示す。

- ジャンルやチャンネルの情報の有意性検証実験
 1. 5.1 節で得られたジャンルの予測ラベル・5.1 節で得られたチャンネルの予測ラベルを用意する。
 2. チャンネル登録者数・動画高評価数をスケーリングする。
 3. BERT を用いてタイトルとコメントからそれぞれ 768 次元の文ベクトルを生成する。
 4. 1. と 2. と 3. で用意した特徴量を説明変数としたモデル 3 とモデル 4, そしてそれぞれにデータを合わせたベースモデルを構築する。
 5. モデル 3 とモデル 4 のテスト用データに対する精度を、各モデルにデータを合わせたベースモデルのテスト用データに対する精度と比較する。

それぞれのモデルの説明変数を I とおくと、ベースモデルは図 5.1 の通り、 $I =$

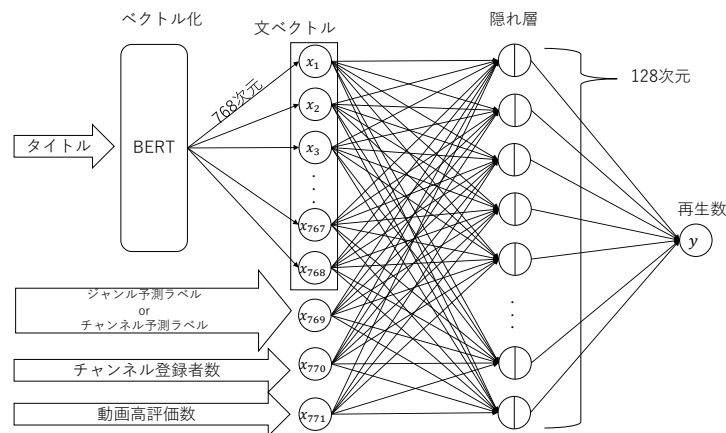


図 5.3 モデル 3(ジャンルの予測ラベル)・モデル 4(チャンネルの予測ラベル)

$[v_{s_title1}, v_{s_title2}, \dots, v_{s_title768}, s_n, f_n]$ とした。モデル 3 は図 5.3 の通り， $I = [v_{s_title1}, v_{s_title2}, \dots, v_{s_title768}, s_n, f_n, p_{v_genre}]$ とした。モデル 4 は図 5.3 の通り， $I = [v_{s_title1}, v_{s_title2}, \dots, v_{s_title768}, s_n, f_n, p_{v_channel}]$ とした。

データセットに用いた動画数は，ゲーム実況が計 11,679 個の動画 (30,000 コメント)，アニメコントが計 6,133 個の動画 (10,000 コメント)，ジャンル分類が計 3,897 個の動画 (8,750 コメント) であった。データセットは再生回数の偏りが無いよう，順番のシャッフルを実施した。データセットの割合は訓練用：検証用：テスト用の順に 8:1:1 とした。精度評価にはテスト用データにおける決定係数 (R^2) と RMSE Loss を用いた。特徴量の有意性を検証するため，10 分割交差検証を実施した。交差検証で得られた評価値は正規分布に従うものとし，有意水準を 5% として対応のある 2 標本 t 検定を実施した。帰無仮説を「2 群の評価値の母平均に差がない」とし，片側検定を行った。

5.3.2 結果・考察

実験結果を表 5.9 に示す。また， t 検定についての結果を表 5.10 に示す。表 5.9 では，全てのモデルがベースモデルよりも高い精度であった。また表 5.10 では，

アニメコントを除いた二つのモデルに対して p 値は全て 0.05 以下となっている。よって帰無仮説は棄却され、モデル 3 とモデル 4(ゲーム実況) はベースモデルと比較して評価値における母平均に有意な差があると言える。しかし、モデル 4(アニメコント) に関しては帰無仮説が棄却されず、ベースモデルと比較して評価値における母平均に有意な差があるとは言い切れない。アニメコントは 5.1.2 節で述べた通り、分類精度がゲーム実況のチャンネル分類と比べて低い。そこで我々は、チャンネルの特徴が上手く考慮できていない場合、そのデータを特徴量に追加しても再生回数予測精度は上がらないと考えた。以上の結果より、再生回数予測精度の向上にコメントが保有するジャンルやチャンネルの情報が影響を与えていることが確認できた。

表 5.9 実験 3：テスト用データに対する精度 (10 回平均)

	ジャンル		チャンネル			
	モデル 3	ベースモデル	ゲーム実況		アニメコント	
			モデル 4	ベースモデル	モデル 4	ベースモデル
決定係数 (R^2)	0.2714	0.2188	0.4307	0.3815	0.3380	0.3374
RMSE Loss	0.6432	0.6661	0.6525	0.6801	0.6113	0.6116

表 5.10 実験 3： p 値

	モデル 3	モデル 4	
		ゲーム実況	アニメコント
決定係数 (R^2)	1.4671e-05	2.0492e-06	0.3577
RMSE Loss	1.7274e-05	1.1992e-06	0.3349

第6章 おわりに

6.1 まとめ

本研究では、動画のメタデータから再生回数を高い精度で予測する手法を提案した。我々は動画の再生回数予測精度を向上させるために、ジャンルごとの視聴者数の違いやチャンネルの特色を考慮すべきではないかと考えた。

しかし、ジャンルごとの視聴者数の違いやチャンネルの特色をデータとして収集するためには二つの問題が存在する。一つ目はジャンルごとの視聴者数の違いやチャンネルの特色を表すデータが存在しないことである。二つ目は収集可能なメタデータには動画の内容に則していないものが存在することである。

一つ目の問題を解消するため、我々はジャンルごとの視聴者数の違いやチャンネルの特色を別のデータから推定することを考えた。また二つ目の問題を解消するため、我々は動画の属するジャンルやチャンネルの特色を適切に表現するためにデータが客観性を持つ必要があると考えた。そこで我々は、不特定多数の視聴者が動画を見た上で感じたことや思ったことを書き残すものであるコメントに着目した。

本稿では、二つの仮説を立てた。一つ目の仮説は、動画に付随するコメントがジャンルやチャンネルを判別可能な情報を保有しているということである。二つ目の仮説は、ジャンルやチャンネルの情報を特徴量にすることによって再生回数予測精度の向上が期待できるということである。

動画に付随するコメントがジャンルやチャンネルを判別可能な情報を保有しているという仮説を検証するため、実験1としてコメントを入力としてBERTを用いたジャンル分類とチャンネル分類を行った。実験の結果、ジャンル分類ではAccuracyが0.87149という結果が得られた。チャンネル分類ではAccuracyが最大で0.6785という結果が得られた。また、ジャンルによって分類精度に大きな差(0.1844)が存在することが確認できた。以上より、コメントがジャンルやチャンネルを判断可能な情報を保有していることが確認できた。

ジャンルやチャンネルの情報を特徴量にすることによって再生回数予測精度の向上が期待できるという仮説を検証するため、実験2と実験3として自作の回帰ニューラルネットワークモデルを用いた2種類の再生回数予測実験を行った。

実験2では、再生回数予測においてコメントが与える影響を調査した。実験の結果、コメントを説明変数に含むモデルの決定係数 (R^2) が 0.7470 であり、最も高い数値となった。次点で動画タグを説明変数に含むモデルが続き、決定係数 (R^2) は 0.7391 であった。ベースモデルの決定係数 (R^2) は 0.7264 で最も低い数値となった。t 検定でも各モデルの評価値間に有意な差が見られた。以上の結果より、コメントが再生回数予測精度に影響を与えていることが確認できた。

実験3では、再生回数予測においてコメント内のジャンルやチャンネルの情報が与える影響を調査した。実験の結果、予測ラベルを含むモデルの決定係数 (R^2) はベースモデルの決定係数 (R^2) と比較して最大で 0.0526 ほど高いという結果が得られた。また t 検定でも評価値間に有意な差が見られた。しかし実験1で得られた予測精度が低いジャンルにおいては、予測ラベルを含むモデルとベースモデルの評価値間に有意な差が見られないという結果が得られた。以上の結果より、精度向上にコメントが保有するジャンルやチャンネルの情報が影響を与えていることが確認できた。

6.2 今後の課題

本稿で行った実験にはいくつかの懸念事項が存在している。一つ目は、手法自体の汎化性能が低い可能性があることである。使用ジャンルや再生回数予測実験に用いる特徴量、そしてデータを収集する期間は我々が独自に設定している。そのため、偏った結果が出力されている可能性が存在する。我々は、ジャンルの設定方法に YouTube 側の機能として用意されている動画カテゴリ設定を用いることによって、その問題を解消できると考える。しかし YouTube 側で用意されている動画カテゴリ設定は、動画タグのように投稿者の設定によっては動画に全く関係のないものが設定されている場合がある。そのためコメントを用いたジャンル分類だけではなく、コメント内のどの単語がジャンル予測に大きな影響を与えているのかを調査し、動画カテゴリを適するものに修正する手法を考える必要がある。また、データの収集期間を延ばすことによって手法の更なる汎化が可能になると考える。

二つ目に、少数ではあるが本実験で収集したデータ内にも設定したジャンル以外の動画が含まれてしまっていることである。本稿では、データの収集方法として

YouTubeDataAPI の Search 機能を使用している。Search 機能ではリクエストしたキーワードと関連する動画やチャンネルの情報を収集する。しかし 4.1 節で述べたように、動画タグには関係のないコンテンツの動画タグが設定されている場合や、動画タグが一つも設定されていない場合が存在する。これにより動画が別のジャンルのキーワードと関連づけられてしまうことがある。そのためデータ収集の際、YouTubeDataAPI の Search 機能以外の方法を考える必要がある。

三つ目に、アップロードされる前の動画にはコメントが付与されないことである。本稿では、コメントを用いることによって再生回数予測の精度が向上するという結果を確認した。しかし、アップロードされる前の動画にはコメントを付与することが不可能である。そのため我々は、投稿前から存在しているメタデータを用いて将来的に付与されるであろうコメントを生成する手法について考案する必要があると考える。

謝辞

ご指導いただいた鈴木先生。時に厳しく、時に優しく、時にすごく厳しい愛のある指導で私は間違いなく成長することができました。先生との問答では私が問題点を理解できるまで長時間付き合ってくださいました。論文執筆の際にはお忙しいのにも拘わらず時間を取っていただきました。全てとても嬉しかったです。感謝の念に堪えません。本当にありがとうございました。

事務補佐員の井尾さんには、色々な手続きをする際に助けていただきました。ご迷惑をおかけした時もありましたが、井尾さんのお陰で学会にも出ることができました。本当にありがとうございました。

心優しい先輩方にはたくさんご迷惑をおかけしました。それでも、親身に研究についての相談に乗っていただいたこと、とても嬉しかったです。私が自分の研究に胸を張れるようになったのは先輩方のお陰です。また、先輩方の研究を見て聞いて学ぶことも多かったように感じています。私も先輩方を目指して努力を続けていこうと思いました。本当にありがとうございました。

新しく研究室に来た後輩たちは、今や私の研究のモチベーションの一つとなっています。後輩たちにきちんとした先輩の背中を見せられるようにと、論文の執筆作業や実験を頑張ることができました。本当にありがとうございました。

親愛なる同期たち。付き合いも長いし今更多くは語らないけれど、ありがとう。お前ら、サイコーだぜ。

最後に、私の論文作成に関わっていただいた全ての人に改めて感謝を申し上げます。皆様のお陰で本論文を書き終えることができました。本当にありがとうございました。

参考文献

- [1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [2] Adiya Abisheva, Venkata Rama Kiran Garimella, David Garcia, and Ingmar Weber. Who watches (and shares) what on youtube? and when?: using twitter to understand youtube viewership. In *Proceedings of the 7th ACM international conference on Web search and data mining, WSDM 2014*. ACM, February 2014.
- [3] Honglin Yu, Lexing Xie, and Scott Sanner. Twitter-driven youtube views: Beyond individual influencers. In *Proceedings of the 22nd ACM international conference on Multimedia*, pp. 869–872, 2014.
- [4] Renjie Zhou, Samamon Khemmarat, Lixin Gao, Jian Wan, and Jilin Zhang. How youtube videos are discovered and its impact on video views. *Multimedia Tools and Applications*, Vol. 75, , 05 2016.
- [5] 柴田知親, 伊東栄典. 回帰分析による cgm 動画再生回数推定. 第 10 回データ工学と情報マネジメントに関するフォーラム (DEIM2018), pp. 1–8. 電子情報通信学会 データ工学研究専門委員会, 2018.
- [6] 山本岳洋, 中村聡史ほか. 視聴者の時刻同期コメントを用いた楽曲動画の印象分類. *情報処理学会論文誌データベース (TOD)*, Vol. 6, No. 3, pp. 61–72, 2013.
- [7] 土屋駿貴, 大野直紀, 中村聡史, 山本岳洋. ソーシャルコメントからの音楽動画印象推定手法の提案. 第 8 回データ工学と情報マネジメントに関するフォーラム (DEIM2016), pp. 1–7. 一般社団法人電気情報通信学会, 2016.
- [8] 村上直至, 伊東栄典. 視聴者コメントに基づく動画検索手法. *電子情報通信学会 2010 年総合大会講演論文集*, Vol. 500, pp. 24–24, 2010.
- [9] 堺雄之介, 伊東栄典. 動画サイトにおける視聴者コメントの特徴抽出. *人工知能学会研究会資料 知識ベースシステム研究会 124 回 (2021/11)*, pp. 17–22. 一般社団法人 人工知能学会, 2021.

- [10] Stefan Siersdorfer, Sergiu Chelaru, Wolfgang Nejdl, and Jose San Pedro. How useful are your comments? analyzing and predicting youtube comments and comment ratings. In *Proceedings of the 19th international conference on World wide web*, pp. 891–900, 2010.

発表リスト

- [1] 川上大凱, 鈴木優 『視聴者のコメントから推測される YouTube 人気コンテンツ間の組合せ相性に対する分析』 東海関西データベースワークショップ 2023, 2023.
- [2] 川上大凱, 鈴木優 『動画の属するコミュニティ情報を考慮した動画再生回数予測手法の提案』 第 16 回データ工学と情報マネジメントに関するフォーラム, 2024.