

# ユーザの高評価時区間を用いた ランダムフォレスト回帰に基づく動画推薦

藤本菜々美<sup>†</sup> 北山 大輔<sup>†</sup>

<sup>†</sup> 工学院大学情報学部システム数理学科 〒163-8677 東京都新宿区西新宿 1-24-2

E-mail: [†j318240@ns.kogakuin.ac.jp](mailto:†j318240@ns.kogakuin.ac.jp), [†kitayama@cc.kogakuin.ac.jp](mailto:†kitayama@cc.kogakuin.ac.jp)

**あらまし** YouTubeなどの動画共有サービスでは、日々多くの動画が投稿されている。その膨大な数の動画から、ユーザが自身の嗜好に合った動画を見つけることは難しい。そこで、本稿では「ユーザが動画内で特に良いと感じた時区間」を高評価時区間と定義し、高評価時区間と動画の音の特徴量を用いた動画推薦を行う。高評価時区間と動画の音の特徴量を用いることで、再生回数や good 数など、動画自体のメタデータでは扱えない、動画の部分的な好みに基づく推薦が可能であると考え。評価実験では、提案手法を用いた動画推薦、good 率を用いた動画推薦の 2 手法を比較した。また、高評価時区間の割合と動画自体に対する評価の関係や、学習データ量や中間時区間の長さや予測精度の関係について、考察を行った。

**キーワード** 動画推薦, ランダムフォレスト回帰

## 1 はじめに

YouTube<sup>1</sup>などの動画共有サービスでは、日々多くの動画が投稿されている。その膨大な数の動画から、ユーザが自身の嗜好に合った動画を見つけることは難しい。従来の動画推薦では、ジャンルなどの作品情報とユーザの視聴回数を用いた推薦 [1] や、トピックを用いた推薦 [2] など、動画自体のメタデータに基づく推薦が行われている。

そこで、本稿では動画自体のメタデータではなく、動画内で時間的に変化するデータに着目する。まず、「ユーザが動画内で特に良いと感じた時区間」を高評価時区間と定義する。そして、高評価時区間と動画の音の特徴量を用いた動画推薦を行う。本稿ではコント動画を対象とするため、高評価時区間はユーザが笑った時区間となる。提案手法では、視聴しながら高評価ボタンを押すことで高評価時区間を取得するが、視聴中のユーザの笑顔を検出するなどによって、より現実的な評価を行うことができる。と考える。

本稿では、以下の 2 つの研究課題を設定する。

- RQ1 動画の音の特徴量を用いた高評価時区間の予測は可能か  
RQ2 高評価時区間の予測に基づく動画推薦は効果的か  
高評価時区間や動画の音の特徴量については、3 章で述べる。

高評価時区間と動画の音の特徴量を用いることで、再生回数や good 数など、動画自体のメタデータでは扱えない、動画の部分的な好みに基づく推薦が可能であると考え。再生回数や good 数などは動画全体に対する評価であるため、その動画にはユーザが面白いと感じた部分とそうでない部分が混在していると考えられる。この場合、ユーザが面白いと感じた部分だけでなく、面白いと感じなかった部分も推薦に影響を及ぼしてしまうため、満足度の低下につながると考える。一方、高評価時

区間を用いることで、ユーザが面白いと感じる部分に対する予測が可能となり、コント動画であれば、ユーザが好むツッコミによる笑いが多い動画などを推薦できるようになる。

本稿の構成を以下に示す。2 章では関連研究、3 章では提案手法について、4 章では予備実験の方法および結果と考察、5 章では評価実験の方法および評価、6 章では考察、7 章ではまとめを述べる。

## 2 関連研究

### 2.1 動画自体のメタデータに基づく推薦

太田ら [1] は、ユーザの潜在的嗜好を反映したアニメ作品や関連コンテンツを横断的に推薦することができるアニメ作品推薦システム “AniReco” を提案した。ジャンルや制作会社などの作品情報と、ユーザの視聴回数を用いて、推薦を行った。評価実験を行った結果、太田らの提案システムが、ユーザの嗜好に合ったアニメ作品を推薦可能であることを確認した。

小川ら [2] は、ユーザが「今まで知らなかったが、推薦されて興味を持ったアイテム」の割合 (Novelty) を向上させる手法を提案した。嗜好傾向に基づくアイテムのトピック分類を行い、ユーザごとにパーソナライズされた推薦リストを作成し、トピックを用いた推薦リストの多様化を行った。その結果、小川らの提案手法が、既存の協調フィルタリング手法よりも Novelty の高い推薦を実現できることを確認した。また、平澤ら [3] は、「社会 (集団) 的に未知であるが、推薦されて興味を持つ」動画を、ソーシャルノベルティのある動画と定義した。コメント特性、動画内容特性、視聴行動特性の 3 特性を特徴量とし、ソーシャルノベルティのある動画を、Support Vector Regression (SVR) を用いて発見する手法を新たに提案した。評価実験を行った結果、多くの人が、未知かつ高い興味を持つ動画を発見できることを確認した。

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

本稿では、高評価時区間と動画の音の特徴量を用いることで、再生回数や good 数など、動画自体のメタデータでは扱えない、動画の部分的な好みに基づく推薦が可能であると考ええる。

## 2.2 動画の部分に対するアノテーションを用いた研究

中村ら [4] は、動画に対して投稿されるコメントを分析することで、コメントが動画のインデックスとして、どの程度有用なのかを明らかにした。また、印象に基づく動画検索およびランキング手法を提案し、プロトタイプシステムを実装、運用した。その結果、笑える度合いや泣ける度合いを考慮したランキングの可能性を明らかにした。また、佃ら [5] は、視聴者の動画への反応に基づいた動画の検索や動画の推薦を可能とするシステムを提案した。「ある人物が動画の後半から活躍する、泣ける動画を検索したい」といった意図を反映した検索を可能とした。また、動画に対する視聴者の反応の類似度に基づいて、動画の推薦を行った。その結果、活躍パターンや印象情報を指定した検索には、一定の需要があることがわかった。また、視聴者の反応に基づく動画の推薦では、ストーリー性のある動画ほど、精度の高い推薦が行えるという知見が得られた。

小西ら [6] は、ニコニコ動画の Time-Synchronized Comment とコメント分布の特徴を用いた、動画の分析手法を提案した。動画を、コメントしたユーザの集合とみなし、ユーザの共通の興味に基づいて、動画をクラスタリングした。実験では、ニコニコ動画に投稿された動画に対して、小西らの提案手法を適用し、既存のテキストベース手法や画像ベース手法との比較による定量的評価と、クラスタリング結果の主観的評価を行った。定量的評価の結果、小西らの提案手法は、同様のユーザベースの既存手法に比べて、テキストベース手法や画像ベース手法、カテゴリを正解データとした際の正規化相互情報量の値が高かった。また、主観的評価の結果、小西らの提案手法は、既存手法に比べて同等以上の結果を示した。

土屋ら [7] は、研究の第一段階として、ユーザが漫才を視聴した際の印象的な部分や、漫才を検索する際のクエリとして用いる言葉について、調査を行った。その結果、ユーザは、セリフを再検索する際の手がかりとしており、それらは、漫才動画内での観客の笑い声が大きいタイミングに表れる傾向があることを明らかにした。本井ら [8] は、発話のリズムに着目し、漫才対話の発話のリズムを分析した。その結果、発話におけるリズムやテンポの良さは、疑似周期的に行われているモーラのリズムユニットを、次話者が自然と受け継ぐことや、一定に刻まれているリズムが、一時的に崩されることで生まれる変調によって生まれるということがわかった。

橋原ら [9] は、利用者がどのような情報をもとにシーン探索を行うかを調査した。その結果、利用者は登場人物を主な手がかりとして、動画内のシーンを探していることがわかった。この結果をもとに、登場人物に注目したシーン探索を支援するシステムを設計、プロトタイプを作成した。坂本ら [10] は、音声・映像記録、トランスクリプト、フィールドノーツなどから、発話のリズムを動的に表現するグラフを作成し、発話状態の可視化を行うことで、“場面の手がかり”を見出した。また、質的研

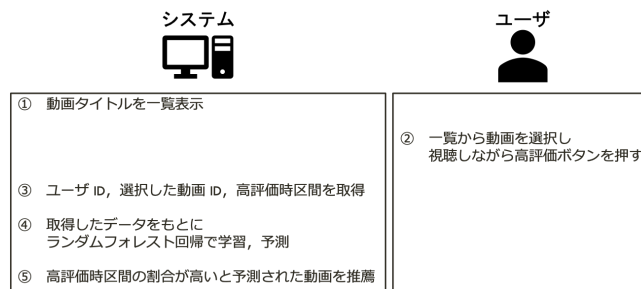


図1 提案手法の概要

究において重要な、データの解釈やコーディング作業を支援し、その妥当性を高めた。坂本らの提案手法を、小学校での授業実習を行った学部生とのインタビューの分析に適用し、その有効性を検証した。

本稿では、動画に対するコメントではなく、高評価時区間と動画の音の特徴量を用いる。動画に対するコメントは、複数のユーザで構成されているため、ある程度の再生回数を有する動画に対してのみ有効であると考えたためである。高評価時区間は各ユーザが各動画に対して付与するものであり、動画の音の特徴量はユーザからの影響を受けるものではないため、提案手法で用いる特徴量はいずれもコメントのような制約はなく、新規の動画も対象にすることができる。

## 3 提案手法

### 3.1 概要

動画を視聴中、ユーザが面白いなど感じる部分には何か特徴があると考えた。そこで、各時間におけるユーザの状態と動画の音の特徴量を用いることにした。ユーザの状態は、高評価状態、中間状態、冷静状態のいずれかであると考えた。例えば、コント動画を視聴している場合、ユーザが笑っていない状態から突然笑い、再び突然笑わなくなるとは考えにくい。笑っている状態とそうでない状態の間には、どちらもいえない状態が存在すると考えたため、中間状態も含めた3つの状態をユーザの状態とした。本研究では、以前の時区間におけるこれらの状態と、動画の音の特徴量を用いて、次の時区間の状態を予測するモデルを考える。

提案手法の概要を図1に示す。はじめに、YouTube Data API v3<sup>2</sup>から、動画 URL、動画タイトル、投稿日時、再生回数、good 数を取得する。取得した動画 URL をもとに、youtube.dl<sup>3</sup>を用いて動画の音をダウンロードする。ユーザは、表示された動画タイトル一覧から動画を選択し、視聴しながら高評価ボタンを押す。ユーザが視聴を終了すると、ユーザ ID、選択した動画の動画 ID、高評価時区間をシステムが取得する。取得したユーザ ID、選択した動画の動画 ID、高評価時区間をもとに、ランダムフォレスト回帰で学習、予測を行う。その後、高評価時区間の割合が高いと予測された動画を推薦する。

2 : <https://developers.google.com/youtube/v3>

3 : <http://ytdl-org.github.io/youtube-dl/>

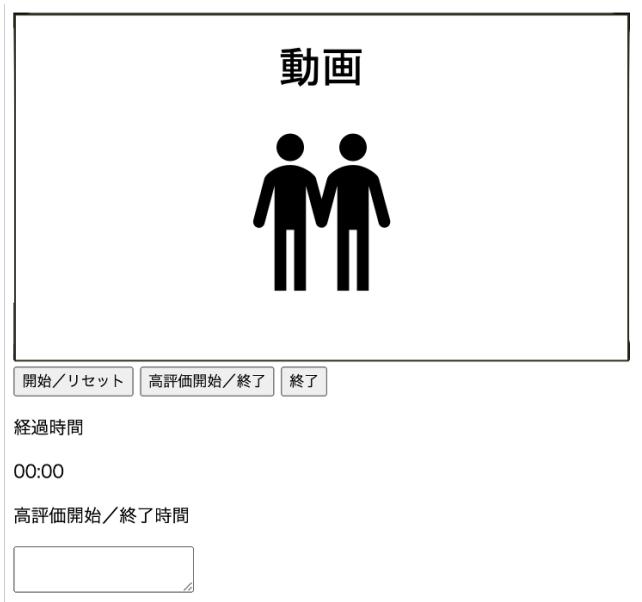


図2 高評価時区間入力 UI

### 3.2 高評価時区間入力 UI

高評価時区間の取得方法を、図2に示す高評価時区間入力UIに沿って説明する。まず、視聴開始と同時に「開始/リセット」を一度押す。これによってタイマーが計測を開始し、「経過時間」に反映される。「開始/リセット」をもう一度押すと、「経過時間」と「高評価開始/終了時間」がリセットされる。視聴中、特に良いと感じた場合は「高評価開始/終了」を一度押す。そのまま動画を視聴し、特に良いと感じなくなった場合は、再度「高評価開始/終了」を押す。「高評価開始/終了」を押すと、その時点の「経過時間」が「高評価開始/終了時間」に反映されるため、「高評価開始/終了時間」の奇数番目は高評価開始時間、偶数番目は高評価終了時間となる。対応する高評価開始時間から高評価終了時間までを、高評価時区間と定義する。視聴終了後、「終了」を押すとタイマーが停止し、ユーザID、選択した動画の動画ID、高評価開始/終了時間がシステムに送信される。

### 3.3 ランダムフォレスト回帰で用いる特徴量

ランダムフォレスト [11] とは、複数の機械学習モデルを組み合わせることで、より強力なモデルを構築するアンサンブル学習の一つであり、複数の決定木を組み合わせたモデルである。ランダムフォレストは調整すべきパラメータが少なく、変数選択の必要もないため、機械学習モデルの中でも比較的単純で扱いやすいという利点がある。

ランダムフォレスト回帰で用いる各特徴量と次元数を表1に示す。いずれも1秒ごとのデータが格納されている。高評価は、高評価時区間に1、高評価時区間の前後1秒（以下、中間時区間）に0.5、その他の時区間（以下、冷静時区間）に0を格納している。テンポは、音楽分析用のPythonパッケージである librosa [12] で取得した動画のテンポを格納している。ただし、librosa で得られるテンポは約0.02秒ごとであるため、1

表1 ランダムフォレスト回帰で用いる各特徴量と次元数

特徴量	次元数
高評価	1
テンポ	1
音の大きさ	1025
$n$ 秒前までの高評価	$n$
テンポの指数平滑移動平均	1
音の大きさの指数平滑移動平均	1025

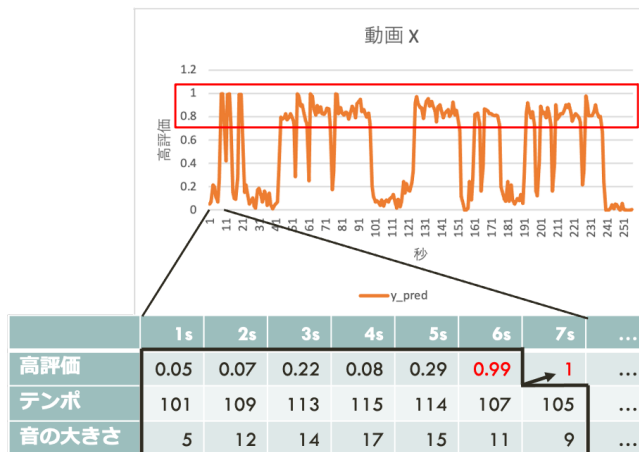


図3 予測例

秒あたりの平均テンポを算出し、格納している。音の大きさも、librosa で取得した動画の音の大きさを格納しており、次元数は1025である。テンポと同様に、1秒あたりの平均の音の大きさを算出し、格納している。単位はデシベルである。 $n$ 秒前までの高評価は、高評価の1秒前から $n$ 秒前までの高評価を格納しており、次元数は $n$ である。動画開始0から $n-1$ 秒は、1から $n$ 秒前の高評価データが存在しないため、0を格納している。テンポの指数平滑移動平均は、テンポ配列の指数平滑移動平均を格納している。また、音の大きさの指数平滑移動平均は、音の大きさ配列の指数平滑移動平均を格納しており、次元数は1025である。指数平滑移動平均 (EMA) は、以下の式(1)(2)(3)で表すことができる。 $\alpha$ は平滑化定数である。

$$\alpha = \frac{2}{m+1} \quad (1)$$

$1 \leq n < m$  のとき、

$$EMA_n = \frac{1}{n} \sum_{i=1}^n x_i \quad (2)$$

$m \leq n$  のとき、

$$EMA_n = EMA_{n-1} + \alpha \times (x_n - EMA_{n-1}) \quad (3)$$

秒数  $n$  および指数平滑移動平均の区間数  $m$  は、予備実験で決定する。予備実験については4章で述べる。

### 3.4 ランダムフォレスト回帰による推薦

3.3節で説明した特徴量をもとに、対象動画のうち、ユーザが視聴した動画をランダムフォレスト回帰で学習し、残りの動画の高評価を予測する。学習データには、各動画の高評価時区間、

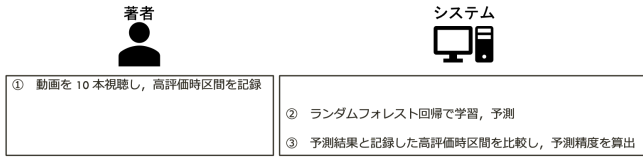


図4 予備実験の概要

中間時区間、冷静時区間のうち、最も秒数が長い時区間を2番目に長い時区間に合わせた、アンダーサンプリングデータを用いる。 $n$ 秒前までの高評価の秒数 $n=6$ 、指数平滑移動平均の区間数 $m=7$ の場合の予測例を図3に示す。予測は、テストデータの各動画に対して、図3下部の表のように、黒枠部のデータを入力としたときの予測値を挿入することを繰り返す。予測値0.75以上(図3赤枠部)を予測高評価時区間とし、各動画における予測高評価時区間の割合を算出する。その後、予測高評価時区間の割合が高い動画を推薦する。

## 4 予備実験

### 4.1 実験方法

以下の2つの値を決定するため、予備実験を行った。

- $n$ 秒前までの高評価の秒数 $n$
- テンポおよび音の大きさの指数平滑移動平均の区間数 $m$

予備実験の概要を図4に示す。まず、著者の1人が動画を10本視聴し、高評価時区間を記録した。この記録した高評価時区間が正解データとなる。記録した高評価時区間をもとに、ランダムフォレスト回帰を用いて、10分割交差検証で評価を行った。ランダムフォレスト回帰は、scikit-learnを用いて行った。パラメータは $n\_estimators=100$ を用い、他はデフォルト値を用いた。ランダムフォレスト回帰で用いる特徴量のうち、 $n$ 秒前までの高評価の秒数 $n$ を $1 \leq n \leq 20$ 、テンポおよび音の大きさの指数平滑移動平均の区間数 $m$ を $1 \leq m \leq 20$ とし、すべての組み合わせで行った。予測結果と記録した高評価時区間の誤差が0.25未満である割合を算出し、予測精度とした。1つの組み合わせにおいて、10本それぞれをテストデータとした場合の予測精度が得られるため、平均予測精度を算出し、最も平均予測精度が高かった秒数 $n$ と区間数 $m$ の組み合わせを採用する。

### 4.2 実験結果と考察

秒数 $n$ ごとの平均予測精度を図5に示す。秒数 $n$ ごとの平均予測精度を比較したところ、 $n=8$ における平均予測精度が約0.532で最も高かった。次に、区間数 $m$ ごとの平均予測精度を図6に示す。区間数 $m$ ごとの平均予測精度を比較したところ、 $m=1$ における平均予測精度が約0.553で最も高かった。最後に、各秒数 $n$ と各区間数 $m$ における平均予測精度を図7に示す。図7は、値が小さくなるにつれて緑に近づき、値が大きくなるにつれて赤に近づくものである。 $n=1$ の場合は、 $m$ の値に関係なく平均予測精度が低いことがわかる。また、 $m$ の値が小さい方が平均予測精度が高い傾向にある。結果として、 $n=9$ 、 $m=2$ における平均予測精度が約0.583で最も高かった。以上

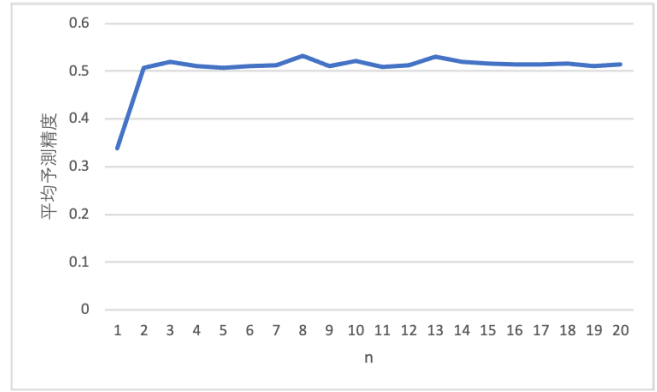


図5 秒数 $n$ ごとの平均予測精度

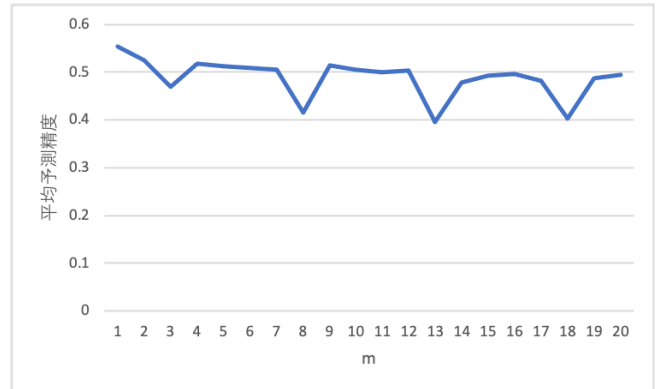


図6 区間数 $m$ ごとの平均予測精度

の結果から、最も平均予測精度が高かった $n=9$ 、 $m=2$ の組み合わせを採用した。

## 5 評価実験

### 5.1 実験方法

提案手法を用いた動画推薦、good率を用いた動画推薦の2手法について、動画自体に対する評価、被験者が実際につける正解の高評価時区間の割合、予測精度を比較する。対象動画は、ジャルジャルタワー JARUJARU TOWER<sup>4</sup>の2021年12月14日時点での最新動画50本のうち、3分以上8分未満の動画39本である。極端に短い動画は、十分なデータが得られない可能性が高いと考えたため、3分以上とした。また、極端に長い動画は被験者の負担になると考えたため、8分未満とした。再生回数やgood数は、投稿されてからの経過日数に影響されるため、good率を採用した。good率は、YouTube Data API v3から取得したgood数を再生回数で除算することで求めた。対象動画と各動画のgood率を表2に示す。なお、表2はgood率降順に並び替えたものである。被験者は大学生6名である。

評価実験の概要を図8に示す。まず、被験者はユーザIDを入力する。次に、表示された動画タイトル一覧から動画を5本選択し、視聴しながら高評価ボタンを押す。被験者が視聴を終了すると、ユーザID、選択した動画の動画ID、高評価時区間をシステムが取得する。取得したユーザID、選択した動画の

4: <https://www.youtube.com/user/comtekaigi>

m

n, m	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0.379	0.320	0.375	0.424	0.314	0.325	0.339	0.314	0.352	0.279	0.307	0.318	0.306	0.329	0.385	0.396	0.326	0.331	0.346	0.302
2	0.580	0.511	0.484	0.502	0.508	0.508	0.511	0.510	0.530	0.516	0.539	0.512	0.460	0.469	0.512	0.514	0.480	0.443	0.473	0.563
3	0.558	0.559	0.567	0.492	0.522	0.547	0.504	0.534	0.502	0.497	0.443	0.518	0.491	0.470	0.467	0.497	0.507	0.536	0.527	0.481
4	0.542	0.520	0.540	0.482	0.503	0.496	0.533	0.552	0.549	0.508	0.526	0.525	0.473	0.499	0.438	0.483	0.493	0.520	0.510	0.526
5	0.560	0.518	0.545	0.497	0.539	0.522	0.523	0.526	0.504	0.548	0.497	0.463	0.464	0.483	0.507	0.522	0.493	0.514	0.467	0.464
6	0.570	0.557	0.565	0.508	0.535	0.487	0.541	0.508	0.515	0.556	0.487	0.519	0.493	0.465	0.474	0.467	0.486	0.494	0.476	0.512
7	0.570	0.523	0.553	0.538	0.507	0.535	0.556	0.512	0.516	0.505	0.485	0.508	0.473	0.468	0.507	0.495	0.476	0.505	0.529	0.500
8	0.561	0.492	0.570	0.529	0.517	0.470	0.487	0.498	0.528	0.554	0.514	0.520	0.528	0.487	0.498	0.534	0.535	0.459	0.463	0.502
9	0.571	0.583	0.563	0.506	0.505	0.562	0.514	0.506	0.531	0.488	0.520	0.510	0.441	0.513	0.473	0.487	0.494	0.461	0.479	0.500
10	0.566	0.553	0.559	0.525	0.511	0.540	0.540	0.545	0.531	0.526	0.508	0.522	0.455	0.491	0.521	0.501	0.494	0.517	0.517	0.505
11	0.554	0.560	0.545	0.486	0.496	0.535	0.465	0.510	0.506	0.517	0.519	0.506	0.514	0.462	0.520	0.523	0.494	0.466	0.474	0.518
12	0.568	0.554	0.543	0.539	0.546	0.506	0.523	0.552	0.525	0.491	0.491	0.546	0.486	0.514	0.465	0.513	0.525	0.460	0.440	0.477
13	0.571	0.518	0.570	0.503	0.533	0.555	0.535	0.549	0.520	0.492	0.512	0.518	0.502	0.495	0.498	0.516	0.491	0.470	0.538	0.490
14	0.556	0.541	0.572	0.553	0.537	0.515	0.475	0.539	0.535	0.539	0.531	0.560	0.453	0.500	0.488	0.537	0.461	0.489	0.501	0.517
15	0.548	0.534	0.547	0.516	0.561	0.509	0.555	0.516	0.548	0.504	0.491	0.534	0.487	0.489	0.511	0.485	0.473	0.508	0.485	0.534
16	0.563	0.514	0.581	0.559	0.478	0.546	0.507	0.556	0.537	0.538	0.517	0.513	0.445	0.465	0.538	0.472	0.501	0.477	0.501	0.478
17	0.563	0.507	0.556	0.540	0.531	0.511	0.531	0.502	0.526	0.540	0.509	0.499	0.535	0.515	0.505	0.452	0.475	0.484	0.514	0.495
18	0.547	0.567	0.556	0.550	0.529	0.505	0.513	0.491	0.502	0.500	0.541	0.441	0.452	0.459	0.505	0.493	0.486	0.505	0.546	0.484
19	0.568	0.542	0.558	0.552	0.535	0.477	0.454	0.527	0.536	0.493	0.513	0.511	0.442	0.489	0.546	0.503	0.452	0.467	0.492	0.553
20	0.560	0.512	0.565	0.555	0.539	0.515	0.506	0.516	0.501	0.500	0.534	0.521	0.486	0.509	0.504	0.532	0.481	0.477	0.484	0.501

図7 各秒数 n と各区間数 m における平均予測精度

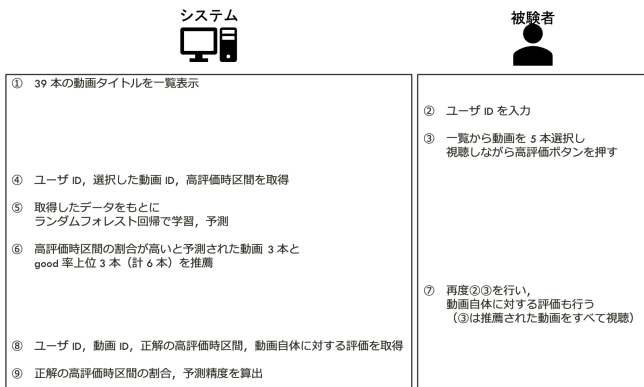


図8 評価実験の概要

動画 ID, 高評価時区間をもとに, 39 本のうち, 被験者が視聴した 5 本をランダムフォレスト回帰で学習し, 残りの 34 本の高評価を予測する. その後, 高評価時区間の割合が高いと予測された上位 3 本の動画と, good 率上位 3 本の動画を推薦する. 被験者は再度ユーザ ID を入力し, 推薦された計 6 本の動画をすべて視聴しながら高評価ボタンを押す. ここで得られる高評価時区間が正解データとなる. また, 推薦された各動画自体に対して「面白かった」, 「面白くなかった」という 2 択での評価も行う. 被験者が視聴を終了すると, ユーザ ID, 動画 ID, 正解の高評価時区間, 動画自体に対する評価をシステムが取得する. 最後に, 正解の高評価時区間の割合, 予測精度を算出する. 予測精度は予備実験と同様, 予測結果と取得した正解の高評価時区間の誤差が 0.25 未満である割合とする.

## 5.2 動画推薦の評価

図 8 の③で各被験者が視聴した動画を表 3 に示す. 被験者 F は誤って動画を 6 本視聴してしまったため, 6 本目の「野球どんでん下手になっていく奴」は, 被験者 F の学習データおよび推

薦対象から除外した.

提案手法を用いた推薦について, すべての被験者の予測値が 0.75 を下回ったため, 予測高評価時区間は得られなかった. それに伴い, 予測高評価時区間の割合が 0 となり, 予測高評価時区間の割合による推薦は行うことができなかった. 代わりに, 予測高評価時区間の平均を算出し, その上位 3 本の動画を提案手法における推薦動画とした.

各被験者の結果を表 4 に示す. 評価は各動画自体に対する評価であり, 1 が「面白かった」, 0 が「面白くなかった」である. また, 高評価割合は, 被験者が実際につけた正解の高評価時区間の割合である. 「面白かった」と評価された動画の本数を比較すると, 提案手法を用いた推薦が 13 本, good 率を用いた推薦が 9 本であり, 提案手法の有用性を示すことができたと思う.

## 6 考察

### 6.1 概要

評価実験により得られたデータに, 「視聴した動画自体に対する評価」, 「視聴した動画における高評価の時系列データ」があるため, これらを用いて以下の観点で考察を行う.

- 高評価時区間の割合と動画自体に対する評価の関係
- 学習データ量と予測精度の関係
- 中間時区間の長さとの予測精度の関係

なお, それぞれのデータは提案手法および good 率手法で推薦された動画であるが, 上記の分析では, どちらの推薦手法で推薦された動画であるかは無関係であるため, 特に区別せず扱う.

### 6.2 高評価時区間の割合と動画自体に対する評価の考察

高評価時区間の割合と動画自体に対する評価の関係を明らかにするため, 各被験者の動画自体に対する評価と正解の高評価時区間の割合の相関係数を算出した. その結果を表 5 に示す.

表 2 対象動画と各動画の good 率

動画タイトル	good 率
マラソン大会でスタートできないパンクボーイって奴	0.035
ヤバい先輩に呼び出された奴	0.034
バックダンサーの〈シェイ〉って奴	0.032
漁師って、良い奴	0.032
できる子、できない子の奴	0.032
野球どんどん下手になっていく奴	0.031
まだ泣いてないのに背中さする奴	0.030
泣いてる他人を励ましに行く奴	0.030
軽い知り合いに、ミツパチに襲われてるとこ見られる奴	0.030
車で猛スピードでクラッシュして、ハズがる奴	0.030
友達作るの下手な奴〜アホ田とキショ田〜	0.028
みんな「自分は普通の人」と思っていることを解説する奴	0.028
一瞬考えただけなのに、ノーコメントと思われた奴	0.027
軽い知り合いに、「ホイップ多めで」と注文してるとこ見られる奴	0.027
最後の思い出作りで試合出る奴	0.027
知らん間に SNS でバズった奴	0.026
友達の家あがって「くっさ〜」って言う奴	0.025
ガードせず「やめてー！」って言う奴	0.025
自分の話に大笑いする奴	0.025
器用に生きてて幸せになれる奴	0.025
軽い知り合いに、脱出ゲーム参加してるとこ見られる奴	0.025
めっちゃしゃべりな奴	0.024
大人になるということを、実感する奴	0.024
説得するとき、言葉も拳も使う奴	0.024
4年ぶりに会ったらめっちゃ変わった奴	0.024
自分のおもしろい話を台無しにする奴	0.024
スカウト早過ぎる奴	0.024
関係者のデブな奴	0.023
めっちゃ失礼なスカウトする奴	0.023
口調だけは綾小路きみまろの奴	0.023
答え全然違う奴	0.023
ついにスカウトされた奴	0.022
いつも遅刻する友達を覚悟して待つ奴	0.022
意見却下されてるのに前向きな奴	0.022
途中、理解しがたい NG 出す奴	0.021
バイト先の先輩にネタを予想される芸人な奴	0.020
部屋探しの条件、無茶な奴	0.018
年収 5 千万円の道を捨てて、ボクサーなる奴	0.016
ラーメン屋のくせに、ぼったくろうとする奴	0.014

被験者 B は、すべての動画に対して「面白かった」と評価していたため、相関係数は算出できず、平均値の算出にも含まれていない。被験者 B を除くすべての被験者については、いずれも正の相関があった。この結果から、ユーザが実際につける正解の高評価時区間の割合が高い動画は、そのユーザにとって面白い動画であるという傾向にあることがわかる。このことから、予測精度が高ければ、そのユーザにとって面白い動画をよりの確に推薦することができると思われる。

### 6.3 学習データ量と予測精度の考察

提案手法を用いた推薦について、各被験者が図 8 の③、⑦で視聴した計 11 本分のデータを用いて、11 分割交差検証で評価を行った。各被験者の各学習データ量における予測精度および再現率を表 6 に示す。ここで、再現率とは、被験者が実際につけた高評価時区間を正解としたときに、それを正しく予測できた割合である。学習データが 5 本の場合の予測精度は、表 4 のものと同一である。学習データが 5 本の場合と 10 本の場合の平均予測精度を比較すると、5 本の場合が約 0.430、10 本の場合が約 0.625 であり、学習データが多い方が予測精度が高いことがわかる。また、学習データが 5 本の場合、一度も予測値

表 3 各被験者の視聴動画

被験者	動画タイトル
A	ヤバい先輩に呼び出された奴
A	めっちゃ失礼なスカウトする奴
A	友達作るの下手な奴〜アホ田とキショ田〜
A	途中、理解しがたい NG 出す奴
A	自分のおもしろい話を台無しにする奴
B	野球どんどん下手になっていく奴
B	年収 5 千万円の道を捨てて、ボクサーなる奴
B	めっちゃ失礼なスカウトする奴
B	部屋探しの条件、無茶な奴
B	説得するとき、言葉も拳も使う奴
C	ヤバい先輩に呼び出された奴
C	いつも遅刻する友達を覚悟して待つ奴
C	めっちゃしゃべりな奴
C	ガードせず「やめてー！」って言う奴
C	軽い知り合いに、「ホイップ多めで」と注文してるとこ見られる奴
D	ヤバい先輩に呼び出された奴
D	みんな「自分は普通の人」と思っていることを解説する奴
D	口調だけは綾小路きみまろの奴
D	意見却下されてるのに前向きな奴
D	ついにスカウトされた奴
E	一瞬考えただけなのに、ノーコメントと思われた奴
E	最後の思い出作りで試合出る奴
E	ラーメン屋のくせに、ぼったくろうとする奴
E	部屋探しの条件、無茶な奴
E	軽い知り合いに、脱出ゲーム参加してるとこ見られる奴
F	自分の話に大笑いする奴
F	知らん間に SNS でバズった奴
F	ヤバい先輩に呼び出された奴
F	口調だけは綾小路きみまろの奴
F	ラーメン屋のくせに、ぼったくろうとする奴
F	(野球どんどん下手になっていく奴)

が 0.75 を上回っていないため、再現率も 0 となった。それに対し、学習データが 10 本の場合については、平均値は約 0.378 と低めだが、最大値は 1 であることから、動画によっては完全に高評価時区間を予測することができているということがわかる。以上のことから、ユーザの視聴履歴が増加すれば、より高精度な予測を行うことができ、より適切な推薦を行うことができると考えられる。

### 6.4 中間時区間の長さとの予測精度の考察

中間時区間の秒数を 1 秒、3 秒、5 秒とした場合における予測精度の比較を行った。この比較は、6.3 節とは独立に、学習データを被験者が図 8 の③で視聴した 5 本として、中間時区間の長さのみを変更して実行した。各被験者の各中間時区間の秒数における予測精度を表 7 に示す。結果、中間時区間の秒数を 3 秒とした場合の平均予測精度が、約 0.530 で最も高かった。このことから、中間時区間の秒数は 3 秒が適切であると考えられる。

## 7 まとめ

本稿では「ユーザが動画内で特に良いと感じた時区間」を高評価時区間と定義し、高評価時区間と動画の音の特徴量を用いた動画推薦を行った。本稿では、以下の 2 つの研究課題を設定した。

- RQ1 動画の音の特徴量を用いた高評価時区間の予測は可能か
- RQ2 高評価時区間の予測に基づく動画推薦は効果的か

表 4 各被験者の結果

被験者	提案手法を用いた推薦				good 率を用いた推薦			
	動画タイトル	評価	高評価割合	予測精度	動画タイトル	評価	高評価割合	予測精度
A	まだ泣いてないのに背中さする奴	1	0.054	0.098	バックダンサーの〈シェイ〉って奴	0	0.016	0.464
	年収 5 千万円の道を捨てて、ボクサーなる奴	1	0.031	0.096	漁師って、良い奴	0	0.000	0.522
	マラソン大会でスタートできないパンキンボーイって奴	1	0.058	0.505	できる子、できない子の奴	0	0.004	0.452
	平均値			0.048	0.233			0.007
最大値			0.058	0.505			0.016	0.522
最小値			0.031	0.096			0.000	0.452
B	4 年ぶりに会ったらめっちゃ変わった奴	1	0.286	0.569	マラソン大会でスタートできないパンキンボーイって奴	1	0.458	0.426
	まだ泣いてないのに背中さする奴	1	0.352	0.445	ヤバイ先輩に呼び出された奴	1	0.436	0.525
	自分のおもしろい話を台無しにする奴	1	0.315	0.501	バックダンサーの〈シェイ〉って奴	1	0.224	0.429
	平均値			0.318	0.505			0.372
最大値			0.352	0.569			0.458	0.525
最小値			0.286	0.445			0.224	0.426
C	年収 5 千万円の道を捨てて、ボクサーなる奴	0	0.148	0.522	マラソン大会でスタートできないパンキンボーイって奴	1	0.205	0.636
	みんな「自分は普通の人」と思っていることを解説する奴	1	0.199	0.411	バックダンサーの〈シェイ〉って奴	0	0.188	0.464
	まだ泣いてないのに背中さする奴	1	0.182	0.435	漁師って、良い奴	1	0.169	0.493
	平均値			0.176	0.456			0.188
最大値			0.199	0.522			0.205	0.636
最小値			0.148	0.411			0.169	0.464
D	まだ泣いてないのに背中さする奴	1	0.173	0.586	マラソン大会でスタートできないパンキンボーイって奴	1	0.299	0.572
	年収 5 千万円の道を捨てて、ボクサーなる奴	1	0.120	0.703	バックダンサーの〈シェイ〉って奴	0	0.136	0.478
	器用に生きてて幸せになれる奴	0	0.114	0.665	漁師って、良い奴	0	0.073	0.565
	平均値			0.136	0.652			0.169
最大値			0.173	0.703			0.299	0.572
最小値			0.114	0.586			0.073	0.478
E	まだ泣いてないのに背中さする奴	1	0.249	0.548	マラソン大会でスタートできないパンキンボーイって奴	1	0.219	0.614
	いつも遅刻する友達を覚悟して待つ奴	0	0.045	0.520	ヤバイ先輩に呼び出された奴	1	0.323	0.593
	年収 5 千万円の道を捨てて、ボクサーなる奴	1	0.203	0.689	バックダンサーの〈シェイ〉って奴	0	0.084	0.518
	平均値			0.166	0.585			0.209
最大値			0.249	0.689			0.323	0.614
最小値			0.045	0.520			0.084	0.518
F	みんな「自分は普通の人」と思っていることを解説する奴	0	0.043	0.096	マラソン大会でスタートできないパンキンボーイって奴	0	0.000	0.873
	途中、理解しがたい NG 出す奴	0	0.028	0.110	バックダンサーの〈シェイ〉って奴	1	0.048	0.560
	年収 5 千万円の道を捨てて、ボクサーなる奴	1	0.031	0.242	漁師って、良い奴	0	0.000	0.622
	平均値			0.034	0.149			0.016
最大値			0.043	0.242			0.048	0.873
最小値			0.028	0.096			0.000	0.560
合計		13				9		
平均値			0.146	0.430			0.160	0.545
最大値			0.352	0.703			0.458	0.873
最小値			0.028	0.096			0.000	0.426

表 5 各被験者の評価と高評価割合の相関係数

被験者	相関係数
A	0.905
B	-
C	0.512
D	0.624
E	0.911
F	0.542
平均値	0.699

高評価時区間と動画の音の特徴量を用いることで、再生回数や good 数など、動画自体のメタデータでは扱えない、動画の部分的な好みに基づく推薦が可能であると考えた。

予備実験では、 $n$  秒前までの高評価の秒数  $n$ 、テンポおよび音の大きさの指数平滑移動平均の区間数  $m$  の 2 つの値を決定するため、秒数  $n$  を  $1 \leq n \leq 20$ 、区間数  $m$  を  $1 \leq m \leq 20$  とし、すべての組み合わせで、ランダムフォレスト回帰による 10 分割交差検証を行った。予測結果と著者の 1 人が記録した高評価時区間の誤差が 0.25 未満である割合を算出し、予測精度とした。結果として、 $n=9$ 、 $m=2$  における平均予測精度が約 0.583

で最も高かったため、この組み合わせを採用した。

評価実験では、提案手法を用いた動画推薦、good 率を用いた動画推薦の 2 手法を比較した。提案手法を用いた推薦について、すべての被験者の予測値が 0.75 を下回ったため、予測高評価時区間が得られず、予測高評価時区間の割合による推薦は行うことができなかった。代わりに、予測高評価時区間の平均を算出し、その上位 3 本の動画を提案手法における推薦動画とした。「面白かった」と評価された動画の本数を比較した結果、提案手法を用いた推薦が 13 本、good 率を用いた推薦が 9 本であり、提案手法の有用性を示すことができたと考える。また、各被験者の動画自体に対する評価と正解の高評価時区間の割合の相関係数から、高評価時区間の割合が高い動画は、面白い動画であるという傾向にあることがわかった。このことは、RQ2 に関して効果的であることを支持すると考える。さらに、学習データを増やした結果、予測精度が向上し、動画によっては完全に高評価時区間を予測することができていた。これらのことから、ユーザが動画を視聴すればするほど、より高精度かつ適切な推薦を行うことができると考えられる。このことは、RQ1 を支持すると考える。

表 6 各被験者の各学習データ量における予測精度および再現率

被験者	各値	5 本		10 本	
		予測精度	再現率	予測精度	再現率
A	平均値	0.233	0.000	0.953	0.827
	最大値	0.505	0.000	0.988	1.000
	最小値	0.096	0.000	0.894	0.000
B	平均値	0.505	0.000	0.529	0.475
	最大値	0.569	0.000	0.685	0.924
	最小値	0.445	0.000	0.359	0.000
C	平均値	0.456	0.000	0.431	0.524
	最大値	0.522	0.000	0.800	1.000
	最小値	0.411	0.000	0.247	0.000
D	平均値	0.652	0.000	0.393	0.160
	最大値	0.703	0.000	0.944	0.620
	最小値	0.586	0.000	0.060	0.000
E	平均値	0.585	0.000	0.524	0.285
	最大値	0.689	0.000	0.822	0.750
	最小値	0.520	0.000	0.241	0.000
F	平均値	0.149	0.000	0.922	0.000
	最大値	0.242	0.000	1.000	0.000
	最小値	0.096	0.000	0.555	0.000
平均値		0.430	0.000	0.625	0.378
最大値		0.703	0.000	1.000	1.000
最小値		0.096	0.000	0.060	0.000

表 7 各被験者の各中間時区間の秒数における予測精度

被験者	各値	1 秒	3 秒	5 秒
A	平均値	0.356	0.530	0.472
	最大値	0.522	0.665	0.677
	最小値	0.096	0.246	0.211
B	平均値	0.482	0.392	0.363
	最大値	0.569	0.489	0.419
	最小値	0.426	0.173	0.230
C	平均値	0.494	0.533	0.528
	最大値	0.636	0.675	0.758
	最小値	0.411	0.443	0.414
D	平均値	0.595	0.577	0.584
	最大値	0.703	0.823	0.773
	最小値	0.478	0.474	0.474
E	平均値	0.580	0.519	0.449
	最大値	0.689	0.593	0.508
	最小値	0.518	0.468	0.370
F	平均値	0.417	0.628	0.640
	最大値	0.873	0.916	0.809
	最小値	0.096	0.467	0.500
平均値		0.487	0.530	0.506
最大値		0.873	0.916	0.809
最小値		0.096	0.173	0.211

今後行うべきことについて、以下で述べる。まず、被験者の高評価と、テンポや音の大きさなどの音データの関係について確認する必要がある。また、別の機械学習を用いることや、特徴量の追加等についても検討する必要があると考える。次に、本稿では対象動画のジャンルをコント動画に限定しており、本数もかなり少なめであったため、今後はジャンルの幅を広げ、

本数も増やしていきたいと考えている。最後に、本稿では高評価時区間を取得するにあたって高評価ボタンを用いたが、将来的には、視聴中のユーザの表情による高評価時区間の取得を行いたいと考えている。

## 謝 辞

本研究の一部は、2021 年度科研費基盤研究 (C)(課題番号: 21K12147) によるものです。ここに記して謝意を表すものとします。

## 文 献

- [1] 太田祥一郎, 河田隼季, 牟田将史, 益子宗, 星野准一. Anireco: ユーザの嗜好を反映させたアニメ作品推薦システム. 研究報告ヒューマンコンピュータインタラクション (HCI), Vol. 2017-HCI-172, No. 10, pp. 1-8, feb 2017.
- [2] 小川祐樹, 諏訪博彦, 山本仁志, 岡田勇, 太田敏澄. 動的なトピック分類に基づく novelty を考慮した推薦アルゴリズムの提案. 情報処理学会論文誌, Vol. 50, No. 6, pp. 1636-1648, jun 2009.
- [3] 平澤真大, 小川祐樹, 諏訪博彦, 太田敏澄. ニコニコ動画のログデータに基づくソーシャルノベルティのある動画の発見手法の提案. 情報処理学会論文誌, Vol. 54, No. 1, pp. 214-222, jan 2013.
- [4] 中村聡史, 田中克己. 印象に基づく動画検索. 研究報告ヒューマンコンピュータインタラクション (HCI), Vol. 2009, No. 5(2009-HCI-131), pp. 77-84, jan 2009.
- [5] 佃洗撰, 中村聡史, 田中克己. 視聴者のコメントに基づく動画検索および推薦システムの提案. WISS 2011, dec 2011.
- [6] 小西敦郎, 細部博史. ユーザの興味とコメント分布によるニコニコ動画の分析. 第 13 回データ工学と情報マネジメントに関するフォーラム, mar 2021.
- [7] 土屋駿貴, 中村聡史, 松下光範. セリフの重要度を考慮した漫才動画再検索のための検索行動の調査. 研究報告グループウェアとネットワークサービス (GN), Vol. 2019-GN-106, No. 44, pp. 1-8, jan 2019.
- [8] 本井佑衣, 岡本雅史. 漫才対話の「テンポの良さ」を支える発話リズムの同期・変調パターン. 社会言語科学会第 43 回研究大会, pp. 170-173, mar 2019.
- [9] 橋原雅広, 松村耕平, 野間春生. 登場人物に注目した動画内のシーン探索支援システム. WISS 2016, dec 2016.
- [10] 坂本将暢, 大谷尚. 発話のリズムに着目した発話分析の手法の一提案: 質的研究におけるデータの解釈とコード化の支援のために. 日本教育工学会論文誌, Vol. 30, No. 1, pp. 37-49, may 2006.
- [11] Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., and Duchesnay E. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, Vol. 12, pp. 2825-2830, 2011.
- [12] Brian McFee, Alexandros Metsai, Matt McVicar, Stefan Balke, Carl Thomé, Colin Raffel, Frank Zalkow, Ayoub Malek, Dana, Kyungyun Lee, Oriol Nieto, Dan Ellis, Jack Mason, Eric Battenberg, Scott Seyfarth, Ryuichi Yamamoto, viktorandreevichmorozov, Keunwoo Choi, Josh Moore, Rachel Bittner, Shunsuke Hidaka, Ziyao Wei, nullmightybofo, Darío Hereñú, Fabian-Robert Stöter, Pius Friesch, Adam Weiss, Matt Vollrath, Taewoon Kim, and Thassilo. *librosa/librosa: 0.8.1rc2*, may 2021.