

ユーザタグとコメントの出現確率の差異に基づく意外な動画の推薦手法

松山 卓矢[†] 北山 大輔[†]

[†] 工学院大学情報学部 〒163-8677 東京都新宿区西新宿 1-24-2
E-mail: tj113108@ns.kogakuin.ac.jp, kitayama@cc.kogakuin.ac.jp

あらまし ニコニコ動画などの動画共有サイトでは、ユーザは投稿された膨大な数の動画から効果的に動画を検索するためにランキング機能を利用することが多い。しかしながら、ランキング機能はユーザが興味のある動画を発見したり、意外性のある動画を発見するのには向かない。そこで本研究では、動画に付与されたタグの情報と動画に対して投稿されたコメントの情報を用いてユーザにとって興味があり、また意外性を感じる動画を推薦する。

キーワード 動画推薦, 動画共有サイト, CGM

1. はじめに

近年, Youtube [1] やニコニコ動画 [2] などの動画共有サイトの発展に伴ってこれらのサイトには多くの動画が存在し, ユーザが好みに合う動画を発見するには非常に手間がかかるという問題がある。ユーザは, 動画を探す手間を減らすために動画共有サイトにあるランキング機能を利用することが多いが, ランキング機能では個々のユーザの好みを反映することは出来ない。また, もともと注目されていた動画がさらに注目を受けやすいため, 同じ動画が長期間に渡って掲載されるという問題点も存在している。そのためランキング機能ではユーザの嗜好に合い, かつ意外性を感じる動画を発見することは難しい。ランキング機能とは別に「オススメの動画」という推薦サービスもあるが, ユーザの視聴履歴に存在する動画のタグに基づいて推薦されるため, ユーザが普段視聴している動画と類似した動画が推薦されやすく, ユーザにとって意外性のない推薦結果である可能性が高い。

本研究ではこのような問題を解決するため, ユーザの嗜好に合い, 普段見ている動画とは異なった反応を呼んでいる動画を意外性のある動画として推薦する手法を提案する。研究の対象はニコニコ動画とし, ニコニコ動画に投稿された動画に付与されているタグの情報と, 動画に対して投稿されたコメントの情報を利用するため, 国立情報学研究所のダウンロードサービスにより株式会社ドワンゴから提供を受けた「ニコニコ動画コメント等データ」[3] を利用した。

ユーザにとって興味があり, かつ意外性のある動画はユーザの視聴履歴に存在する動画に付与されているタグと多く一致し, コメントを比較した際に視聴履歴内に存在する同じタグを付与されている動画とは投稿されているコメントの内容が大きく違う動画であると考えられる。提案手法の流れを図1に示す。推薦候補の動画と視聴履歴内の動画それぞれでタグと単語の組み合わせを抽出し, 同じタグと単語の組み合わせ全ての出現確率の差異をカイ二乗値を用いることで算出し, その合計値が高い動画ほどユーザにとって興味があり, かつ意外性のある動画として推薦動画とする。

本論文の構成を以下に示す。2章では, 動画推薦やニコニコ

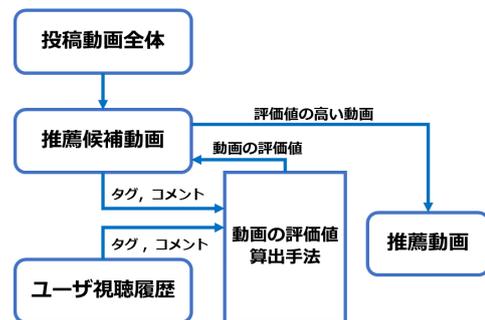


図1 提案手法の流れ

動画に関する研究について述べる。3章では, ニコニコ動画のデータについて述べる。4章では, ユーザの嗜好に合い, 意外性の有る動画の推薦手法と予備実験について述べる。5章では推薦手法による評価実験と結果についての考察を述べる。6章で本研究のまとめを述べる。

2. 関連研究

動画推薦については, 様々な研究が行われている。ニコニコ動画を対象とした動画推薦の研究として, 平沢らの研究 [4] [5] では, ニコニコ動画のログデータに注目し, 機械学習を用いて社会的に未知であるが推薦されて興味を持つ動画を推薦している。佃らの研究 [6] [7] では, コメントを用いて動画内の登場人物の活躍パターンを抽出し, ユーザが視聴している動画の活躍パターンと類似している動画を推薦している。工藤らの研究 [8] では, ユーザの視聴履歴と投稿動画間におけるタグの共起関係を用いて意外性のある動画の推薦を行っている。本研究では, コメントを用いることで意外性のある動画の推薦を行っている点で異なる。

ニコニコ動画に関する研究として, 早川らの研究 [9] では, コメントの投稿時間と動画内で表示されるタイミングをもとに, コメントのレスポンス関係を推定することで面白いコメントを抽出する手法を提案している。松原らの研究 [10] では, 動画の画像特徴量と, コメントを用いて動画の重要な場面を検出して

複数のサムネイルを抽出し、動画の要約を行っている。村上らの研究 [11] では、タグの出現頻度や共起表現 ISR(Inter-section ratio) 手法を用いてタグの階層化を行っている。

ニコニコ動画以外の動画共有サイトを対象とした動画推薦の研究として、江端らの研究 [12] では、動画に対して投稿されたコメントに TF-IDF 法を用いて特徴ベクトルを算出し、類似する動画を関連動画として推薦している。本研究では、意外性のある動画の推薦を目的としている点で異なる。

3. ニコニコ動画のデータ

本研究では、ユーザにとって興味があり、かつ意外性のある動画を推薦するために、動画に付与されたタグの情報と動画に対して投稿されたコメントの情報を用いる。ニコニコ動画に投稿された動画にはカテゴリ分けを行うためのカテゴリタグ 1 件と動画投稿者や視聴者が付与することが出来るタグ 10 件の最大計 11 件までのタグを付与することが出来る。動画に対して投稿されるコメントの情報には最大 1024 文字までのコメント内容、動画内で表示される時刻のデータや表示される時の文字の色や大きさのデータが含まれている。

今回はユーザの視聴履歴内において、ユーザが動画に付与されているタグに対してどんなコメントが大きく関係しているかを算出し、それを推薦候補の動画内でのタグとコメントの関係と比較を行うため、タグの情報とコメント本文のデータを利用した。動画に対して投稿されたコメントをそのまま比較に扱うのは難しいため、日本語形態素解析システムの一つである MeCab [13] を利用し、解析の結果から得られた単語をユーザがタグに対して持つ印象的な単語を算出するのに用いた。

4. 提案手法

4.1 概要

提案手法の流れを以下に説明する。はじめに、ユーザの視聴履歴内の動画からタグと単語の組み合わせを全て作成し、それぞれの出現確率を求めることでユーザの普段見ている動画の傾向を抽出する。その後、推薦候補の動画におけるタグと単語の組み合わせを全て作成し、出現確率を求めた後にそれらが視聴履歴内における出現確率がどれほどの差があるかをカイ二乗値を用いることで算出する。期待値に対して推薦候補の動画における出現確率、観測値に対して視聴履歴内における出現確率をそれぞれ適応することで推薦候補の動画に対してコメントの投稿数が少なかった場合にスコアが高くなりすぎないように組み合わせ毎の評価の算出を行った。組み合わせ毎に算出されたカイ二乗値を合計した数値を動画の評価値とし、評価値の高い動画を推薦する。図 2 は提案手法の流れを示す図である。

4.2 推薦候補動画における評価値の算出

推薦候補の動画と視聴履歴内の動画でユーザがタグに対して持っている印象を比較するため、推薦候補の動画でタグと単語の組み合わせを全て作成する。推薦候補の動画におけるあるタグと単語の組み合わせの出現確率と同じタグを含む視聴履歴内の動画における出現確率を比較する。出現確率を算出する流れを図 3 に示す。

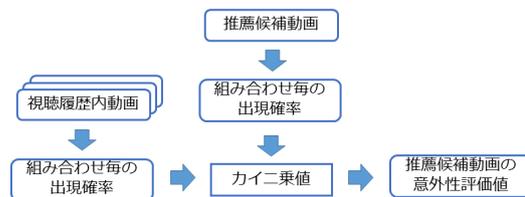


図 2 動画推薦の流れ

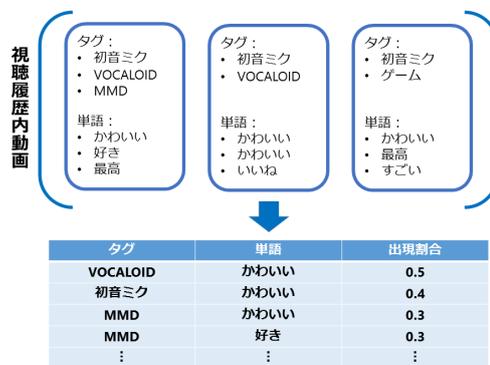


図 3 出現確率算出の流れ

「初音ミク」、「VOCALOID」、「MMD」などはタグであり、そのタグがつく動画のコメントに出現するコメントと組み合わせ、「初音ミク、かわいい」のようなペアを作る。このペアが視聴履歴内にどのくらい出現するかを確率で表す。タグと単語の組み合わせごとの評価値は出現確率を基にカイ二乗値によって算出し、その合計値を推薦候補の動画の評価値とする。動画の評価値は以下の式 (1) によって求められる。

$$score(X) = \sum_{t \in T, k \in K} \frac{(\bar{X}_{t,k} - X_{t,k})^2}{X_{t,k}} \quad (1)$$

X は推薦候補の動画、 T は X に付与されたタグで、 t はそのうちのひとつ、 K は X に投稿されているコメント内の単語で、 t はそのうちのひとつ、 X は推薦候補の動画におけるタグと単語の組み合わせが t, k の出現数、 \bar{X} は視聴履歴内におけるタグと単語の組み合わせが t, k の出現確率に推薦候補内のタグと単語の組み合わせ数を掛けて割合を合わせたものである。例としてタグが $\{A, B, C\}$ 、単語が $\{a: 5 \text{ 回}, b: 2 \text{ 回}, c: 3 \text{ 回}\}$ という推薦候補の動画があるとすると、この場合タグと単語の組み合わせのパターンは $\{A, a\}, \{A, b\}, \{A, c\}, \{B, a\}, \{B, b\}, \{B, c\}, \{C, a\}, \{C, b\}, \{C, c\}$ の計 9 個となる。また $\{A, a\}$ の出現数は 5 個で、全体の組み合わせの個数はタグの種類 3 個と単語数 10 個を掛けて 30 個となる。A, a の組み合わせが視聴履歴内において出現確率が 0.25 としたときカイ二乗値は以下の式 (2) のようになる。

$$\frac{(0.25 \times 30 - 5)^2}{5} = 1.25 \quad (2)$$

計 9 個の組み合わせそれぞれのカイ二乗値を求め、それを全て合計した値が推薦候補の動画の評価値となる。動画の評価値算出の例を図 4 に示す。

例とした推薦候補の動画には「初音ミク」、「VOCALOID」

表 4 ゲーム動画を対象としたカイ二乗値上位 10 件

tag	word	score
初音ミク	好き	0.08454
初音ミク	すぎ	0.06611
初音ミク	さん	0.01272
初音ミク	絵	0.01207
初音ミク	声	0.01151
初音ミク	最高	0.01109
初音ミク	歌詞	0.00905
初音ミク	さ	0.00663
初音ミク	サビ	0.00483
初音ミク	ちゃん	0.00282

表 5 予備実験の結果

	含むタグ	タグの一致数	評価値
A	MMD	3	0.4121
B	第 10 回 MMD 杯本選	3	0.5583
C	作ってみた	2	0.2934
D	ゲーム	1	0.3030

は全く含まれない単語のスコアが高くなり、全体としても高い評価値が算出された。これは今回対象とした動画がカップ麺を調理する際の待ち時間に視聴する動画であるという趣旨があったためと考えられる。この事から提案手法を用いることでタグの一致数が同じ動画間においてより意外性のある動画を推薦することが可能であると考えられる。次に動画 C と動画 D を比較する。動画 C は一致するタグが 2 つあり、タグと単語の組み合わせ数が多くなるため、タグが 1 つしか一致していない動画 D よりスコアが高くなると予想したが、結果は動画 D の評価値が高くなった。理由としては、動画 C では初音ミクに対して注目しているコメントが多く、視聴履歴内に存在する動画と傾向が似たためスコアが伸びなかったと考えられる。それと比べ動画 D では、ゲームの内容に対して注目するコメントが多く、カイ二乗値の上位 10 件には入らなかったがリズムゲームに関する単語である「クリア」、「パーフェクト」、「コンボ」などの単語のスコアが高くなり、結果全体の評価値が高く算出されたと考えられる。この事から提案手法を用いることでタグの一致数が少ない動画においても、視聴履歴内の動画と傾向を比較することでタグに一致数がより多い動画と同等の評価をすることができると考えられる。これらより提案手法を用いることでユーザの興味に合い、より意外性のある動画に対して高い評価値を与えられることを確認した。

5. 評価実験

提案手法を用いることでユーザにとって興味があり、かつ意外性のある動画が推薦出来ることを客観的に評価するため、指定した視聴履歴を用いた推薦動画の評価実験と、被験者の視聴履歴を用いた推薦動画の評価実験の 2 つを行った。

5.1 指定した視聴履歴を用いた評価実験

指定した視聴履歴を用いた評価実験では、9 人の被験者に視聴履歴として設定した「VOCALOID」カテゴリの動画 10 件

表 6 提案手法による動画の評価

	1 件目	2 件目	3 件目	4 件目	5 件目
嗜好に合う	3.22	3.56	3.78	3.33	4.22
意外性がある	4.44	3.33	2.89	2.22	3.22
合計値	7.67	6.89	6.67	5.56	7.44

表 7 比較手法による動画の評価

	1 件目	2 件目	3 件目	4 件目	5 件目
嗜好に合う	3.56	3.56	3.78	4.00	3.56
意外性がある	3.33	3.89	3.11	2.89	3.67
合計値	6.89	7.44	6.89	6.89	7.22

表 8 指定した視聴履歴による実験結果

	比較手法	提案手法
嗜好に合う	3.69	3.62
意外性がある	3.38	3.22

を予め視聴してもらい、その後に提案手法によって推薦された動画 5 件、比較手法によって推薦された動画 5 件の合わせて 10 件をランダムな順番に視聴し、各動画に対して評価を付けてもらった。動画の評価は視聴履歴の動画に対して嗜好に合う、意外性があるの 2 つの指標に 5 段階の評価を付け、より良い結果であれば 5、悪い結果であれば 1 を付けてもらった。比較手法は「VOCALOID」カテゴリ内において、視聴数、コメント数、マイリスト数それぞれのランキング作成したものであり、順位を合計した値が低い動画を推薦動画とした。提案手法、比較手法それぞれの推薦動画は 2016 年 7 月の 1ヶ月間にニコニコ動画に投稿された動画約 16 万件から抽出した。実験の結果を集計し、平均値を取った結果、提案手法による推薦動画の評価は表 6 の結果となり、比較手法による推薦動画の評価は表 7 の結果となった。また、2 つの手法の評価平均は表 8 のようになった。

実験の結果、提案手法は比較手法に対して嗜好および意外性共に下回る結果となったが、嗜好に対しての評価と意外性に対しての評価の 2 つとも平均値が 3 を上回っていることから、ユーザにとって興味があり、かつ意外性のある動画を推薦することが出来ていると考えられる。また、提案手法による推薦動画と提案手法による推薦動画の計 10 件の動画の内、最も評価の合計値が高い動画は提案手法の 1 件目の動画の 7.67 であるため、提案手法によって最も評価の高い動画を発見していることが確認できる。しかし、提案手法は最も評価の低い動画も推薦している。理由としては、提案手法では様々なカテゴリから動画の推薦を行っており、別のカテゴリから推薦された動画に対して被験者は良い結果であると感じなかったためだと考えられる。

5.2 被験者の視聴履歴を用いた評価実験

被験者の視聴履歴を用いた評価実験では、3 人の被験者に予め過去に視聴したことのある動画 10 件を提示してもらい、その視聴履歴から提案手法によって推薦された動画 5 件、比較手法によって推薦された動画 5 件の合わせて 10 件をランダムな順番に視聴し、各動画に対して評価を付けてもらった。動画の評価は視聴履歴の動画に対して嗜好に合う、意外性があるの 2

表 9 被験者の視聴履歴による実験結果

	比較手法	提案手法
嗜好に合う	2.87	2.87
意外性がある	3.33(2.9)	3.53(3.83)

つの指標に5段階の評価を付け、より良い結果であれば5、悪い結果であれば1を付けてもらった。比較手法は被験者の視聴履歴中に一番多くの動画存在するカテゴリ内において、視聴数、コメント数、マイリスト数それぞれのランキング作成したものであり、順位を合計した値が低い動画を推薦動画とした。提案手法、比較手法それぞれの推薦動画は2016年5月の1ヶ月間にニコニコ動画に投稿された動画約16万件から抽出した。実験の結果を集計し、平均値を取った結果評価平均は表9のようになった。また、意外性があるの括弧内の数値は嗜好にあうの評価が3以上の動画に対しての意外性があるかの評価となっている。

実験の結果、比較手法と提案手法では嗜好についての評価が同程度だが、意外性についての評価は提案手法が上回る結果となった。また、嗜好に合うの評価が3以上の動画を対象に意外性についての評価を見ると、比較手法では評価が落ち、提案手法では評価が上がっていることが確認できた。ユーザにとって嗜好に合う動画は、ユーザが内容についてよく理解している動画であると考えられる。そのため、意外性についての評価が落ちることが予想されるが、提案手法を用いた際には意外性の評価が落ちていなかったため、提案手法を用いることで被験者に対しての嗜好と意外性が両立することが出来たと考えられる。

今後の課題として、安定して評価値の高い動画を推薦するためにユーザの視聴履歴に多く存在するタグなどを重視した動画の評価値算出手法が今後必要であると考えている。また、今回提案した手法ではコメントがネガティブな印象を持つものであるかを考慮していないため、全体的に不評な動画や、タイトルや説明文と内容が全く異なる釣り動画などでも評価値が高くなる可能性がある。そのため、単語の意味を考慮した評価値の算出を今後行う必要があると考えている。

6. まとめ

本研究では、ニコニコ動画に投稿された動画に対して付与されているタグの情報と投稿されているコメントの情報を用いて、ユーザにとって興味があり、かつ意外性のある動画の推薦手法を提案した。提案手法ではユーザが動画に付与されているタグに対して持っている印象をタグと単語の組み合わせによって推定し、推薦候補動画における組み合わせと視聴履歴内における組み合わせの出現確率をカイ二乗値を用いて推薦候補の動画の評価値とした。実験結果から提案手法を用いることでユーザにとって興味があり、かつ意外性のある動画を推薦することが出来ていることと、ユーザに取っての嗜好と意外性の両立が出来ていることが確認できた。しかし、様々なカテゴリから推薦を行うため評価の揺れ幅が大きく、視聴履歴に多く存在するタグを重視した評価値の算出手法が今後必要であると考えている。また、今回の手法ではコメントがネガティブな印象を持つ

ものであるかを判断していないため、評価値が高いのでユーザが推薦されて嬉しい動画であるとは限らない。そのため、単語の意味を考慮した上で評価値を算出していく手法が今後必要であると考えている。本研究では、意外な動画をユーザが普段見ている動画とトピックが同じだが、コメントの傾向が異なる動画とした。そのため、ユーザの嗜好をタグ、意外性をコメントと分担して動画の推薦を行った。しかし、タグが異なることで意外性を感じたり、コメントの傾向が似ているためユーザの嗜好に合うということも考えられるため、新しい意外な動画の定義を提案していくことも重要であると考えている。

謝 辞

本研究の一部は、平成28年度科研費若手研究(B)(課題番号:15K16091)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] <https://www.youtube.com/>.
- [2] <http://www.nicovideo.jp/>.
- [3] <http://www.nii.ac.jp/dsc/idr/nico/nico.html>.
- [4] 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を用いた新たな面白い動画の発見に関する研究. 日本社会情報学会全国大会研究発表論文集, Vol. 26, pp. 225-228, 2011.
- [5] 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, 2013.
- [6] 洗根恒, 聡史中村, 岳洋山本, 克己田中. 映像に付与されたコメントを用いた登場人物が注目されるシーンの推定. 情報処理学会論文誌, Vol. 52, No. 12, pp. 3471-3482, 2011.
- [7] 佃洗根, 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. *WISS2011*, p. B15, 2012.
- [8] 工藤真之, 北山大輔. ユーザ視聴履歴と投稿動画におけるタグの興味関係に基づく動画推薦手法. *DEIM Forum 2015 B8-6*, 2015.
- [9] 卓弥早川, 嘉徳土方, 正吾西田. 実時間と動画時間から面白い動画コメントを抽出する手法の提案. 情報処理学会論文誌, Vol. 56, No. 9, pp. 1929-1942, 2015.
- [10] 松原宏和, 新妻弘崇, 太田学. 画像特徴量とコメントを用いたニコニコ動画の指示的要約サムネイルの生成手法. 研究報告システムソフトウェアとオペレーティング・システム(OS), Vol. 2014, No. 18, pp. 1-9, 2014.
- [11] 直至村上, 栄典伊東. 動画投稿サイトで付与された動画タグの階層化. 研究報告数理モデル化と問題解決(MPS), Vol. 2010, No. 17, pp. 1-6, 2010.
- [12] 佑介江端, 秀憲川村, 恵二鈴木. ユーザコメントのtf-idf法による分析を用いたインタラクティブな関連動画の提示(社会システムと情報技術研究ウィーク). 電子情報通信学会技術研究報告. AI, 人工知能と知識処理, Vol. 109, No. 439, pp. 7-10, 2010.
- [13] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. Applying conditional random fields to japanese morphological analysis. In *In Proc. of EMNLP*, pp. 230-237, 2004.