

ユーザの繋がりとコメントを用いた意外性のある動画推薦システム

松原 宏和[†] 太田 学^{††}

[†] 岡山大学工学部情報工学科 〒 700-8530 岡山県岡山市北区津島中 3-1-1

^{††} 岡山大学大学院自然科学研究科 〒 700-8530 岡山県岡山市北区津島中 3-1-1

E-mail: [†], ^{††}{matsubara,ohta}@de.cs.okayama-u.ac.jp

あらまし 人気のある動画投稿閲覧サービスに YouTube やニコニコ動画などがある。これらのサービスには膨大な動画が投稿されているため、ユーザが興味を持つ動画を発見するのは非常に手間がかかる。本研究では YouTube を対象とし、閲覧動画と関連があり、かつ意外性のある動画を推薦する動画推薦システムを提案する。特にコメントに基づく動画の繋がりをを用いて動画を集め、動画の評価値等に基づいて推薦動画をランキングする点に、本研究の特徴がある。被験者実験の結果、提案手法は満足度が高く意外性のある動画を推薦できることがわかった。

キーワード 動画推薦, YouTube, ユーザの繋がり, コメント

A recommendation system for unexpected videos using user connection and comments

Hirokazu MATSUBARA[†] and Manabu OHTA^{††}

[†] Department of Information Technology, Faculty of Engineering, Okayama University
3-1-1, Tsushima-naka, Kita-ku, Okayama, 700-8530 Japan

^{††} Graduate School of Natural Science and Technology, Okayama University
3-1-1, Tsushima-naka, Kita-ku, Okayama, 700-8530 Japan

E-mail: [†], ^{††}{matsubara,ohta}@de.cs.okayama-u.ac.jp

Key words Video recommendation, YouTube, User connection, Comment

1. はじめに

近年, YouTube [1] やニコニコ動画 [2] などの動画投稿サイトによる動画共有が盛んになっている。この YouTube 等の動画共有サイトは誰もが自由に動画を投稿できるため、膨大な数の動画が存在する。例えば YouTube では、2012 年 1 月時点で 1 分当たりおよそ 60 時間分の動画が投稿されており、一日の動画視聴回数は 40 億回に及ぶ [3]。この膨大な動画の中からユーザが自分に興味のある動画を探し出すのは大変手間もかかり、困難である。そこで吉尾ら [4] は、閲覧している動画の投稿ユーザの周りのユーザに着目した。すなわち、その投稿ユーザと何らかの繋がりを持つユーザの動画を集め、閲覧動画の特徴ベクトルと集めた動画の特徴ベクトルを比較し、一定の類似度を持つ動画を推薦するシステムを提案した。彼らによると、推薦に有用なユーザの繋がりとは、ユーザが選択した動画の投稿ユーザの「お気に入り動画を投稿したユーザ」、「登録チャンネル」、「チャンネル登録者」、「友だち」によるユーザの繋がりである。よって、「お気に入り動画を投稿したユーザ」、「登録チャンネル

ル」、「チャンネル登録者」、「友だち」が投稿した動画を推薦対象としている。実験の結果、ユーザによる評価で意外性と満足度の両方の評価が高かったのが「お気に入り繋がり」であった。そこで「お気に入り繋がり」が動画推薦に有望な繋がりであるとしている。

本研究では、吉尾らが動画推薦に有望な繋がりであるとした「お気に入り繋がり」を利用しつつ、新たに動画に対するコメントを利用した意外性のある動画推薦システムを提案する。

本稿の構成は次の通りである。2 節で関連研究について述べ、3 節では動画推薦システムについて説明する。4 節で評価実験について述べ、5 節でまとめる。

2. 関連研究

2.1 動画推薦

吉尾らは、YouTube を対象にユーザの繋がりをを用いた意外性のある動画推薦システム [4] を提案した。彼らは、ユーザが選択した動画の投稿ユーザの「お気に入り」、「登録チャンネル」、「チャンネル登録者」、「友だち」の繋がりから動画を集め、関

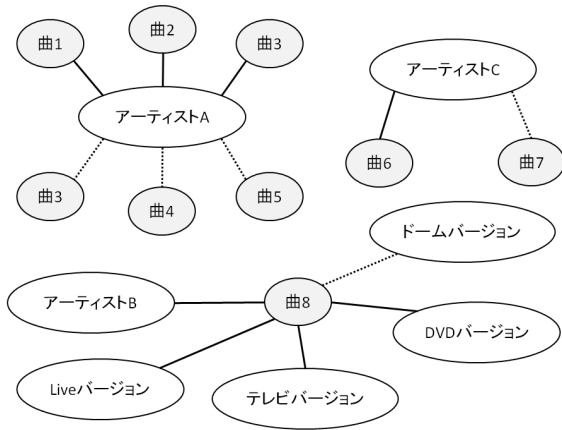


図1 利用者の嗜好を表すネットワーク

覧動画のメタデータから生成した特徴ベクトルと、ユーザの繋がりが得られた関連動画の特徴ベクトルを比較し、一定の類似度を持ち、かつ意外性のある動画を推薦している。また彼らは、元々興味のあるアイテムと類似したジャンルで、それについて知識は少ないが興味が沸くアイテムを意外性のあるアイテムと定義している。

彼らの実験から意外性があり満足度の高い動画は「お気に入り繋がり」から得られる動画に多いことが分かっている。「お気に入り繋がり」とは、ある動画投稿者のお気に入り動画の投稿者という繋がりのことである。そこで本研究でも、推薦に有用な繋がりとして「お気に入り繋がり」を利用する。ただし本研究の動画推薦ではコメントによるユーザの繋がりも利用する点が異なる。

動画検索サイトにおける動画推薦では、利用者が「お気に入り」フォルダに新たに登録したいと思う動画を推薦する方法が提案されている[5]。利用者のお気に入りフォルダに登録されている動画のtitleから利用者の嗜好の情報となり得る単語をキーワードとして抽出し、その関連を分析することにより利用者の嗜好を明らかにし、新たな動画を推薦する。具体的には、お気に入りフォルダの動画titleから抽出したキーワードを節点とし、複数のキーワードを同時に含む動画titleがあれば、それらのキーワードに対応する節点を実線で結びネットワークを構成する。そして1,780,481件の動画データが登録されている動画マスタの中に、お気に入りフォルダの動画titleから抽出したキーワードを含む動画titleが存在すれば、動画マスタの動画titleから新しいキーワードを抽出し、お気に入りフォルダの動画titleから抽出したキーワードの節点と破線で結び。この利用者の嗜好を表したネットワークを図1に示す。このようにしてできたネットワークの各節点から出ている全ての線に対して、実線の数が多い節点から出る破線の先のキーワードに対応する動画を推薦する。例えば、図1であれば「曲8のドームバージョン」を推薦する。しかし、お気に入りには十分な数の動画データを登録していない利用者には推薦できない。

2.2 意外性のある推薦

ユーザ利用履歴を用いた意外性推薦システムが提案されてい

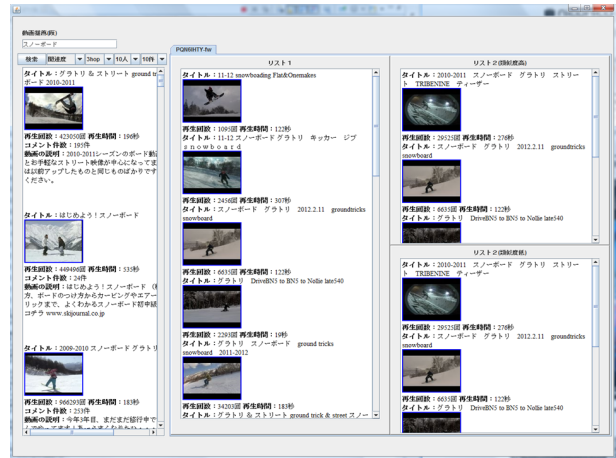


図2 動画推薦システムのGUI

る[6]。これは、推薦対象の映画情報からキーワードを抽出した後、概念辞書を利用し、多数のキーワードを概念データに変換した特徴集合と、ユーザのクリック履歴情報により生成したユーザプロファイルを利用し、「推薦結果に発見性を含み、ユーザ状況が一致するもの」と定義した意外性のある情報を提供する情報推薦システムである。発見性はユーザが知らない情報を見つけたことによる驚きの要素で、ユーザ状況はユーザの趣味に合った情報が提供される喜びの要素としている。ユーザが今求めている情報を得るためにクリック履歴を利用することで、最新のユーザ状況を推薦に反映できるとしている。本研究ではコメントによるユーザの繋がりをを用いるが、ユーザ状況は考慮しない。

3. 提案する動画推薦システム

3.1 推薦システムの概要

本研究では吉尾らの動画推薦システムのGUIを利用する。その実行画面を図2に示す。このGUIの左上のテキストボックスに検索語を入力すると、その下にYouTubeの検索結果が表示される。このYouTubeの検索結果の中から一つ動画を選ぶと、システムはユーザの繋がりをを用いて動画を集める。このユーザの繋がりととは、3.2.3項で示すコメントとお気に入りを組み合わせたユーザの繋がりのことである。このユーザの繋がりを利用して3.3節で示す方法で動画を収集する。そしてユーザの繋がりによって集められた動画をランキングし、ランキングの上位20件を図2のGUIの右側に表示する。GUIの“リスト1”は本研究の推薦アルゴリズムで推薦された動画20件、“リスト2”は比較のための吉尾らの提案した推薦アルゴリズムで推薦された動画20件である。

3.2 ユーザの繋がり

ここでは推薦する動画を集めるために用いたユーザの繋がりについて説明する。推薦で用いるユーザの繋がりにには、コメントによるユーザの繋がりとお気に入りによるユーザの繋がりの2種類がある。そこで、まずコメントによるユーザの繋がりについて説明し、次にお気に入りによるユーザの繋がりについて説明する。そしてコメントによるユーザの繋がりとお気に入り



(a) 動画メタデータ



(b) コメントデータ

図3 YouTubeの動画エン트리

によるユーザの繋がりを組み合わせたユーザの繋がりについて説明する．

3.2.1 コメントによるユーザの繋がり

コメントによるユーザの繋がりを得るために、YouTube Data API [7] を用いて動画メタデータ (図3) とユーザのアクティビティ (図4) を取得する．動画メタデータとは、動画のタイトルや動画の説明文、再生回数、評価値、コメント内容、コメント投稿者等を指す．一方ユーザのアクティビティとは、ユーザがどんな動画をアップロードしたか、どの動画にコメントや評価をしたか等である．本研究では、「コメント投稿者」「ユーザがどの動画にコメントしたか」の情報を利用してユーザの繋がりを作成する．

コメントによるユーザの繋がりとは、動画にコメントを投稿しているユーザを利用する繋がりである．このコメントによるユーザの繋がりを図5に示す．図5のCommenterはコメント投稿者を表し、Videoはコメント投稿者によってコメントが投稿された動画を表す．具体的には、ある動画V1にコメントを投稿しているコメント投稿者C1, C2, C3がいる．そしてコメ



図4 ユーザのアクティビティ

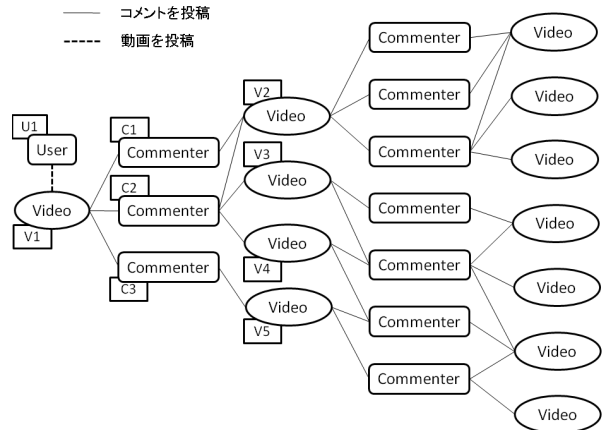


図5 コメントによるユーザの繋がり

ント投稿者C1は動画V2、コメント投稿者C2は動画V2, V3, V4、コメント投稿者C3は動画V5へもコメントを投稿している．そしてさらに、動画V2, V3, V4, V5にコメントを投稿しているコメント投稿者がC1, C2, C3以外にも存在する．これがコメントによるユーザの繋がりである．

3.2.2 お気に入りによるユーザの繋がり

お気に入りによるユーザの繋がりとは、ユーザのお気に入り動画とその動画の投稿ユーザの繋がりである．お気に入りによるユーザの繋がりを図6に示す．図6のUserとは、動画の投稿ユーザを表し、Fav Videoはお気に入り動画、Uploaded Videoはユーザが投稿した動画である．具体的には、ある動画の投稿ユーザがお気に入り登録した動画の投稿ユーザ、そしてお気に入り登録された動画の投稿ユーザのお気に入り動画の投稿ユーザという繋がりがお気に入りによるユーザの繋がりである．

お気に入りによるユーザの繋がりとは、吉尾らが動画推薦システムで利用した繋がりである [4]．吉尾らの提案システムでは、お気に入りによるユーザの繋がり以外の他の繋がりも利用していたが、お気に入りによるユーザの繋がりが一番推薦に有効であるとしているため本研究ではこれを使用する．お気に入りによるユーザの繋がりを作成するために、YouTube Data API を用いてチャンネル情報とユーザのアクティビティを取得する．

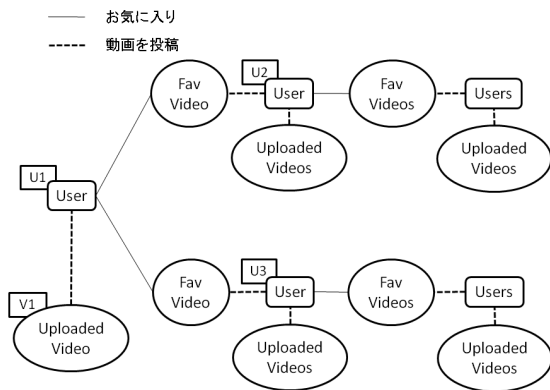


図6 お気に入りによるユーザの繋がり

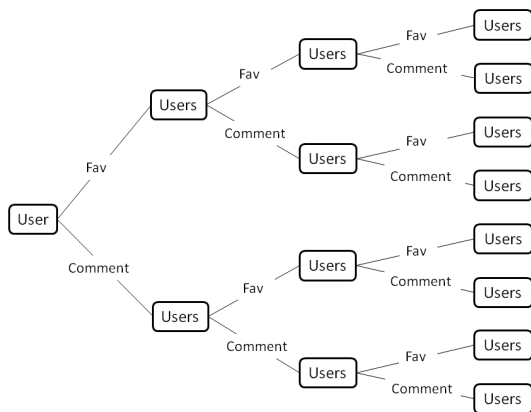


図7 お気に入りとコメントを組み合わせたユーザの繋がり

チャンネル情報には「お気に入り動画」「再生リスト」「友だち」等がある。お気に入りによるユーザの繋がりでは「お気に入り動画」「ユーザがどんな動画をアップロードしたか」これらの情報を利用する。

3.2.3 コメントとお気に入りによるユーザの繋がり

3.2.1 項と 3.2.2 項で示したそれぞれのユーザの繋がりを組み合わせて提案システムでは動画を収集する。コメントによるユーザの繋がりとお気に入りによるユーザの繋がりを組み合わせた繋がり図7のようになる。図7の Fav はお気に入りによるユーザの繋がり、Comment はコメントによるユーザの繋がりである。また図7の「User」は、図5と図6の U1 に対応し、図7の「User」と「Comment」で繋がる「Users」とは、図5の「Commenter」の C1, C2, C3 である。図7の「User」と「Fav」で繋がる「Users」は、図6の「User」の U2, U3 のことである。

3.3 推薦動画の収集法

推薦に有望と考える繋がりをもとに動画を収集する。ユーザの繋がりを広げると、集まる動画の数はホップ数が増えるにつれて爆発的に増えていく。しかしそれらの動画の中には、推薦すべきでない動画も多く含まれる。そこで本節では、推薦のためにユーザの繋がりを選別して動画を収集する方法を示す。

推薦に有望な繋がりを見つけるためにユーザが選んだ動画とユーザの繋がりから収集された動画の類似度を次のように算出する。すなわち、動画のメタデータである「動画のタイトル」「動画の説明文」から特徴語を抽出して、各動画の特徴語ベクトルを生成し、ユーザの閲覧動画とユーザの繋がりから収集した動画の類似度を計算する。特徴語の抽出には、形態素解析器 ChaSen [8] および専門用語（キーワード）自動抽出用 Perl モジュール「TermExtract」[9] を利用する。TermExtract により抽出された特徴語を基に、特徴語ベクトルを termmi [10] を用いて生成する。ここで termmi とはテキストファイルの専門用語を複数ファイル間で比較するシステムである。形態素解析による単語への分割、TermExtract による専門用語の抽出、それらの文章中における重要度の計算という三つのステップにより、複数ファイル間での専門用語の重要度を比較する。

また termmi は重要度比較以外に、ベクトル空間法を利用したテキストファイル間の類似度を比較することも可能である。本研究ではこれを用いて閲覧動画と収集した動画の類似度を計算する。類似度は、閲覧動画と集めた動画が類似していれば 1 に、類似していなければ 0 に近い値となる。

推薦に有望な繋がりであるか、不適切な繋がりであるかの判断は、先ほど述べた termmi により計算される動画間の類似度に基づいて行う。閲覧動画の特徴語ベクトルとユーザの繋がりから集めた動画の特徴語ベクトルの類似度を算出し、算出した類似度が 0 であれば、その動画は閲覧動画とは関係が無いと判断し、その動画から繋がる動画も閲覧動画との関連が小さいとみなして、その先の動画を収集しない。一方、類似度が 0 より大きい動画は閲覧動画と何らかの関連があるので、この繋がりをもとに動画を収集する。

3.4 動画の推薦アルゴリズム

3.3 節で示したユーザの繋がりにより集めた動画をランキングし、ユーザに推薦する方法を説明する。吉尾らは閲覧動画との類似度が、ある一定の値の動画を推薦して、閲覧動画とあまり似すぎていないことを意外性とした。しかし、彼らの推薦動画では、意外性は YouTube の関連動画より高いが、満足度が YouTube の関連動画より劣るという問題があった。そこで本研究では閲覧動画と一定の関連があり、かつユーザが満足できる動画を推薦することを目的とする。

3.4.1 ランキングに利用する動画のメタデータ

ユーザの繋がりをもとに集めた動画を、ランキングしユーザに提示する。そのランキングに利用する動画のメタデータとして、YouTube Data API を用いて取得できる「評価値」「コメント件数」「再生回数」を使用する。YouTube のメタデータには「カテゴリ」もあるが、本研究では類似度により動画の関連度を測っており、カテゴリの横断も許しているため考慮しない。

3.4.2 動画のランキング手法

3.3 節で説明した方法で、ユーザの繋がりに基づいて集めた動画は、すでに閲覧動画と何らかの関連がある。そのためこれらは閲覧動画と全く無関係の動画ではなく、閲覧動画とよく似た動画や一部似ている動画が含まれていると考えられる。そこで、これらを類似度ではなく「評価値」「コメント率」に基づ

いてランキングしユーザに提示する。

まず動画に対するユーザの評価として一番わかりやすい「評価値」を用いる。評価値が低い動画を提示してもユーザは満足しないと考えるので、ユーザの繋がりによって集めた動画から評価値が低い動画を推薦対象から除外し、評価値が一定の値以上の動画を残す。ただしマイナーまたは新しい動画には評価値が付いていないこともあり得るため、評価値が0の動画は推薦対象に残しておく。

次に「コメント件数」「再生回数」を利用する。先ほどの「評価値」による選別で残った動画は、ある一定以上の評価を受けている動画と評価値が0の動画である。そこで「コメント件数」を「再生回数」で割った「コメント率」を定義する。この「コメント率」の大きい上位20件をユーザに提示する。コメント率の大きい動画とは、他の動画に比べて再生回数の割にコメント件数が多いことを意味している。つまり、ある一定以上の評価値があるとき、再生回数が少なくても、コメント件数が多いということは、その動画のコメントには動画に対して好意的なコメントが多く投稿されていると推測できる。また評価値が0で、あまり再生されていないくても、コメントが多ければ、その動画を推薦する。一方、有名で多くのユーザが見ている動画でも、コメントが少なければ提案手法では推薦しない。

4. 実験

3節で説明した動画推薦システムにより推薦された動画が適切なものであるか評価するための実験を行った。まず3.3節で説明した動画の収集法が妥当であるかを確かめる動画収集実験について述べる。次に3.4.2項で説明した動画のランキングに使用する評価値の閾値を定めるために行った実験について述べる。最後に、推薦動画を被験者が評価した実験について説明する。

4.1 動画収集実験

ユーザの繋がりはホップ数が増えるにつれて大変多くなる。そのためユーザの繋がりを利用して集める動画も膨大になり、その全ての動画と閲覧動画の類似度を比較しながら収集するかどうかを決定するのは現実的でない。そこで、3.3節で提案したように類似度により収集に利用するユーザの繋がりを制限したが、この方法が妥当かどうかを実験により確かめる。

まず、ユーザが検索クエリ「野球」を用いて動画を検索し、その検索結果から動画を一つ選択する。これを閲覧動画とする。動画推薦システムは、閲覧動画からユーザの繋がりをたどり1ホップ目の動画を集める。1ホップ目で集まった動画と閲覧動画の類似度を計算し、その類似度が0の動画と0より大きい動画に分ける。このときの動画の件数は表1の左の列のようになる。

次に類似度が0の動画510件からユーザの繋がりをたどり2ホップ目の動画を集める。1ホップ目の類似度が0より大きい動画32件からも、ユーザの繋がりをたどり2ホップ目の動画を集める。それらの結果を表1の右の2列に示す。表1より、集まる動画の合計では1ホップ目の類似度が0の繋がりに集めた方が多いが、類似度が0のユーザの繋がりに集めた動画

表1 収集される動画

	1 ホップ	2 ホップ	
		1 ホップの類似度=0	1 ホップの類似度>0
類似度=0	510	30,185	2,455
類似度>0	32	530	91
合計	542	30,715	2,556

表2 類似度=0と類似度>0の動画に含まれる動画の件数

	満足	不満足	合計	関連有り	関連無し	合計
類似度=0	127	383	510	198	312	510
類似度>0	25	7	32	17	15	32

より、1ホップ目の類似度が0より大きいユーザの繋がりに集めた動画の方に、類似度が0より大きい動画が含まれる割合が高いことが分かる。よって閲覧動画との類似度が0の動画からは、類似度が0の動画が集まりやすく、類似度が0より大きい動画からは類似度が0より大きい動画が集まりやすいといえる。ここで1ホップ目の類似度が0より大きい動画32件と類似度が0の510件の動画の中に、本稿の第一著者が満足する動画と、関連を認める動画がどれだけ含まれているか調べた。これを表2に示す。表2より類似度が0の動画には閲覧動画と関連のないものが多く、満足とはいえない動画も多いことがわかる。また類似度が0より大きい動画には閲覧動画と何らかの関連のある動画の方が多く含まれる。つまり類似度が0の動画は推薦しても意味がないものが多いことがわかるため、ユーザの繋がりをたどる途中で類似度により集める動画を制限しながら、ユーザの繋がりを広げていくことは有効である。本研究では、類似度が0より大きい動画からのユーザの繋がりを有望な繋がりとして動画を収集する。

4.2 動画の評価値とユーザの満足度に関する実験

3.4.2項で説明したように、動画のランキングには動画のメタデータである「評価値」を用いる。YouTubeの「評価値」は最高5、最低1の5段階である。この「評価値」がどの値以上であればユーザが満足する動画であるかを調べた。

まず、吉尾らの実験結果より、YouTubeの関連動画の満足度が高いことが分かっている。そこで、表3に示す各選択動画に対するYouTubeの関連動画25件の評価値を調べ、各動画を本稿の第一著者が視聴し満足度を「大きい」「小さい」「無い」の3段階で評価した。表3の4つの動画の関連動画100件の中に、満足度の「大きい」動画は46件あり、この46件の平均評価値は4.36だった。そこで本研究では、この値以上の評価値であればユーザが満足する可能性が高いとみなす。

4.3 被験者実験

被験者に推薦動画を評価させることで、提案手法の有効性を評価した。また比較対象としてGoogleの提供するYouTubeの「関連動画」と吉尾らが提案した動画推薦システムの推薦動画を利用した。

被験者は、推薦されたこれらの動画に対して、「意外性」「関連度」「満足度」の3項目について「大きい(2)」「小さい(1)」「無い(0)」の3段階で評価する。ここで、閲覧動画と被験者が見た動画が、無関係または関係はあるがそれについて知識が

表3 ユーザの満足度調査に利用した動画

検索語	選択動画のタイトル	関連動画の平均評価値
スノーボード	グラトリ & ストリート ground trick & street スノーボード 2010-2011	4.15
AKB	【PV】Everyday、カチューシャ / AKB48[公式]	3.79
サッカー	サッカー 日本 VS フランス 香川の劇的決勝弾！	4.58
面白動画	ネットで拾った面白画像特集 良い子は見てはいけませんっ!!	4.23

少なければ意外性があると判断する．関連度は、閲覧動画と被験者が見た推薦動画の関連の度合いを被験者の主観で判断したものである．満足度は、被験者がその推薦動画に満足できたかどうかで判断する．

4.3.1 推薦動画リスト

被験者が評価する推薦動画リストは合計で 14 ある．具体的には、3.2 節で説明した三つのユーザの繋がりによって集めた動画に対してそれぞれ次に示す四つのランキング方法で推薦したリストと、吉尾らの手法の推薦動画のリスト、Google の提供する YouTube の関連動画のリストである．

一つ目のランキング方法は、3.4.2 項で説明したように、「評価値」と「コメント率」を用いた方法で、「評価値」の閾値に 4.2 節で示した満足度の大きい動画の平均評価値 4.36 を用いて、「評価値」がこの値以下の動画は推薦対象から除外し、「コメント率」の上位 10 件をランキングするものである．ただし、「評価値」が 0 の動画はマイナーまたは新しい動画であるとし、推薦対象から除外せずコメント率に基づいてランキングに反映する．二つ目のランキング方法は、「評価値」に閾値を設けず「評価値」の上位 10 件をランキングするものである．三つ目のランキング方法は、「評価値」を考慮せず「コメント率」の上位 10 件をランキングするものである．四つ目のランキング方法は、動画のメタデータである「動画を評価したユーザの人数」が少ない上位 10 件をランキングするものである．この方法では、まだ評価の付いていない新しい動画とあまり評価のついていないマイナーな動画が上位になる．これら 14 の推薦動画リストを表 4 にまとめる．それぞれのリストには最大 10 件の推薦動画が含まれる．

4.3.2 推薦動画の評価

岡山大学工学部情報工学科の四年生 3 名と同大学院自然科学研究科の大学院生 3 名からなる被験者が、提案手法の推薦動画と吉尾らの手法の推薦動画、YouTube の「関連動画」を評価した「関連動画」は Google 提供の YouTube Data API で取得した．

まず被験者らが閲覧した動画のタイトルと、動画を選ぶために利用した検索語を表 5 にまとめる．被験者が検索結果から動画の一つ選ぶと、4.3.1 項で説明した動画リストが推薦される．被験者は、提案手法の推薦動画と吉尾らの推薦動画、YouTube の「関連動画」を無作為並べた動画リストを評価した．動画を集めるのに利用したユーザの繋がりは 10 ホップまでとした．

被験者による評価結果を図 8 に示す．図 8 の縦軸は 3 段階評価の平均で、横軸は左から表 4 で示した各推薦動画リストを表

表4 推薦動画リスト

リスト名	リストの種類
Rec1-com-rate	コメントとお気に入りによるユーザの繋がりを 用いて集めた動画のランキングに評価値とコメント 率を用いた推薦動画
Rec1-com	コメントとお気に入りによるユーザの繋がりを 用いて集めた動画のランキングにコメント率のみを 用いた推薦動画
Rec1-rate	コメントとお気に入りによるユーザの繋がりを 用いて集めた動画のランキングに評価値のみを用い た推薦動画
Rec1-#raters	コメントとお気に入りによるユーザの繋がりを 用いて集めた動画のランキングに動画を評価したユ ーザの人数のみを用いた推薦動画
Rec2-com-rate	コメントによるユーザの繋がりを 用いて集めた動画のランキングに評価値とコメント率を用いた推 薦動画
Rec2-com	コメントによるユーザの繋がりを 用いて集めた動画のランキングにコメント率のみを用いた 推薦動画
Rec2-rate	コメントによるユーザの繋がりを 用いて集めた動画のランキングに評価値のみを用いた推薦動画
Rec2-#raters	コメントによるユーザの繋がりを 用いて集めた動画のランキングに動画を評価したユーザの人数の みを用いた推薦動画
Rec3-com-rate	お気に入りによるユーザの繋がりを 用いて集めた動画のランキングに評価値とコメント率を用いた 推薦動画
Rec3-com	お気に入りによるユーザの繋がりを 用いて集めた動画のランキングにコメント率のみを用いた推薦 動画
Rec3-rate	お気に入りによるユーザの繋がりを 用いて集めた動画のランキングに評価値のみを用いた推薦動画
Rec3-#raters	お気に入りによるユーザの繋がりを 用いて集めた動画のランキングに動画を評価したユーザの人数 のみを用いた推薦動画
Rec-Yoshio	吉尾らの手法の推薦動画
YouTube	Google の提供する YouTube の「関連動画」

す．図 8 からわかるように、意外性に関しては、Rec3 の四つの推薦動画リストが一番高い．しかし各リストには被験者 6 名の推薦動画をあわせると、最大 60 件の動画が含まれるが、Rec3 の各リストでは被験者 6 名の推薦動画はあわせて 7 件しか推薦されず、十分なデータが得られなかった．お気に入りによるユーザの繋がりのみを使って動画を収集する Rec3 で動画があまり推薦されなかったのは、被験者が選んだ閲覧動画の投稿者がお気に入り動画を公開しておらず、ユーザの繋がりが作成できなかったからである．Rec3 を除くと、意外性は吉尾らの推薦動画リストが一番高く、次に提案手法の Rec2-com の推薦動画リストが高い．一方満足度に関しては、YouTube の「関連動画」が一番高く、次に提案手法の Rec1-rate の推薦動画リストが高い．また四つのランキング方法で推薦動画リストを作成したが、「評価値」のみを利用してランキングした推薦動画リスト

表 5 被験者らが選択した動画

利用した検索語	選択動画のタイトル
manu ginobili	Manu Ginobili 2010-2011 MVP Highlights - HD
ペン回し	ペン回し講座 No.1 ノーマル
Mr.children	Mr.Children 常套句の PV
イチロー	イチロー 8 打数 7 安打 4 盗塁 & 決勝タイムリーの 神モード突入
面白動画	ネットで拾った面白画像特集 良い子は見ては いけませんっ!!
スノーボード	グラトリ & ストリート ground trick street スノーボード 2010-2011

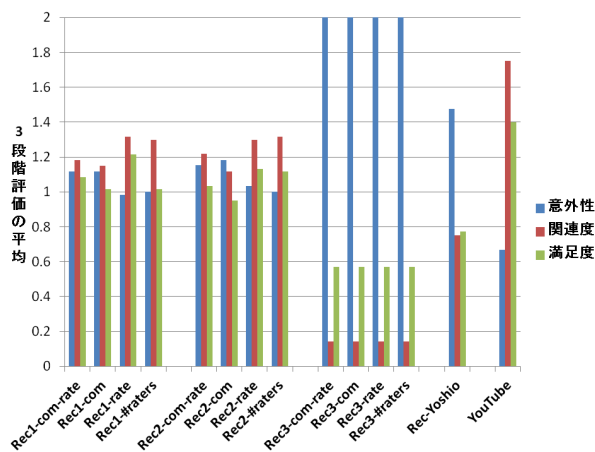


図 8 推薦リストの評価

が一番満足度が高くなり、「コメント率」「評価値」を共に利用した推薦動画リストは意外性は高いが満足度は低い傾向にある。コメントとお気に入りによるユーザの繋がりをを用いて動画を収集する Rec1 とコメントによるユーザの繋がりをを用いて動画を収集する Rec2 の推薦動画には大きな違いが見られなかった。

4.4 考 察

提案手法の推薦動画は吉尾らの手法の推薦動画より満足度が高いことが分かった。Google の提供する YouTube の「関連動画」と比べると提案手法の推薦動画の満足度はやや低かった。また提案手法の推薦動画の意外性は、吉尾らの手法の推薦動画の意外性に比べて低い。Google の提供する YouTube の「関連動画」の意外性より高いことがわかった。つまり YouTube の「関連動画」より意外性があり、吉尾らの手法の推薦動画より満足度の高い動画が推薦される。しかし、提案手法の推薦動画は YouTube の「関連動画」より満足度は低く、吉尾らの手法の推薦動画より意外性は低い。よって両者の中間の性質を持つ推薦手法となっている。提案手法の推薦動画の意外性が高い理由の一つには、3.3 節で説明した動画収集において、わずかでも閲覧動画と類似点があればそのような動画を全て収集していることがあげられる。また満足度が高いのは、集めた動画に対して評価値が一定以上の動画やコメント率の高い動画を推薦しているからである。

Rec3 を除く被験者 6 名分の推薦動画リスト (Rec1 と Rec2 と Rec-Yoshio と YouTube) 中に、被験者が関連度も満足度も大き

表 6 推薦動画の分類 (関連度と満足度の関係)

リスト名	関連度 2 かつ 満足度 2	関連度 1 かつ 満足度 2	関連度 0 かつ 満足度 2
Rec1-com-rate	20	1	1
Rec1-com	18	0	1
Rec1-rate	21	3	0
Rec1-#raters	17	1	0
Rec2-com-rate	19	0	0
Rec2-com	14	2	0
Rec2-rate	18	3	0
Rec2-#raters	16	4	0
Rec-Yoshio	5	0	0
YouTube	29	1	1

表 7 推薦動画の分類 (意外性と満足度の関係)

リスト名	意外性 2 かつ 満足度 2	意外性 1 かつ 満足度 2	意外性 0 かつ 満足度 2
Rec1-com-rate	4	3	15
Rec1-com	3	3	13
Rec1-rate	4	4	16
Rec1-#raters	0	5	13
Rec2-com-rate	2	6	11
Rec2-com	3	3	10
Rec2-rate	3	6	12
Rec2-#raters	2	7	11
Rec-Yoshio	0	1	4
YouTube	3	7	21

い (3 段階評価で 2) とした動画と、関連度は小さい (3 段階評価で 1) が満足度が大きい (3 段階評価で 2) とした動画と、関連度は無い (3 段階評価で 0) が満足度が大きい (3 段階評価で 2) とした動画の件数を集計した (表 6)。同様に被験者が意外性も満足度も大きい (3 段階評価で 2) とした動画と、意外性は小さい (3 段階評価で 1) が満足度が大きい (3 段階評価で 2) とした動画と、意外性は無い (3 段階評価で 0) が満足度が大きい (3 段階評価で 2) とした動画の件数を集計した (表 7)。表 6 より、被験者が関連度 0 かつ満足度 2 とした動画はほぼ無いため、ユーザは全く関連の無い動画では満足しないことが分かる。また図 8 と表 6 より、Rec1 や Rec2 といった提案手法の推薦動画には、閲覧動画と何らかの関連のある動画が集まっていることが分かる。よって 3.3 節で説明した閲覧動画との類似度を利用した方法は閲覧動画と何らかの関連のある動画を集めるのに有効であるといえる。さらに表 7 より、提案手法の推薦動画リストの Rec1-com-rate と Rec1-rate に意外性と満足度が共に高い (3 段階評価で 2) 動画が多く含まれることが分かる。

5. ま と め

本稿では、コメントによるユーザの繋がりと、お気に入り動画によるユーザの繋がりを利用して、動画を収集して推薦するシステムを提案した。コメント投稿ユーザと動画投稿ユーザを利用してユーザの繋がりを広げて推薦に利用する動画を集め、またその際閲覧動画との類似度を利用して、推薦に有望な繋が

りだけを選択的にたどる方法を提案した。

動画収集実験で、閲覧動画との類似度が0より大きい動画からは、閲覧動画と何らかの関連のある動画が集まりやすいことがわかった。そこで閲覧動画との類似度が0より大きい動画からの繋がりを利用して、被験者実験では推薦動画の候補を収集した。被験者実験では、吉尾らの提案した手法の推薦動画よりも、Google の提供する YouTube の「関連動画」や本研究で提案した手法の推薦動画の満足度が高いことが分かった。一方意外性については吉尾らの提案手法の推薦動画が一番高かった。しかし本研究の提案手法の推薦動画は、意外性と満足度が共に高い動画が多く含まれることが分かった。よって提案手法は、YouTube と比べると満足度はやや低くなるものの、吉尾らの手法と YouTube の両者の中間の性質を持つといえる。

今後の課題として、3.3 節で説明した動画の類似度以外に基づく動画の選別法の検討が挙げられる。そのために意外性と満足度の両者の関連をさらに詳しく調査する予定である。

文 献

- [1] YouTube: <http://www.youtube.com/>
- [2] ニコニコ動画: <http://www.nicovideo.jp/>
- [3] The Official Youtube Japan Blog: <http://youtubejpblog.blogspot.jp/2012/01/60-40.html>
- [4] 吉尾透, 太田学, “ユーザの繋がりを利用した意外性のある動画推薦システム”, DEIM Forum 2012, B4-1, 2012.
- [5] 長谷川喬一, 山川栄樹, 壇寛成 “動画検索サイトにおけるレコメンデーション手法の一提案”, 日本オペレーションズ・リサーチ学会春季研究発表会アブストラクト集 2010, 136-137, 2010.
- [6] 吉見哲哉, 辻秀一, “ユーザ利用履歴を用いた意外性推薦システムの提案”, 電子情報通信学会技術研究報告. AI, 人工知能と知識処理 110(105), 37-41, 2010.
- [7] YouTube の API とツール: <http://code.google.com/intl/ja/apis/youtube/overview.html>
- [8] 形態素解析システム茶筌: <http://chasen.naist.jp/hiki/ChaSen/>
- [9] 専門用語 (キーワード) 自動抽出用 Perl モジュール “TermExtract” の解説: <http://gensen.dl.itc.u-tokyo.ac.jp/termextract.html>
- [10] Windows 用テキストマイニングツール “Termmi” の解説: <http://gensen.dl.itc.u-tokyo.ac.jp/termmi.html>