

再生リストを利用した Web 動画クラスタリング手法の提案

上江まり子[†] 橋本 隆子^{††} 北川 博之^{†††}

[†] 筑波大学大学院システム情報工学研究科 〒 305-8573 茨城県つくば市天王台 1-1-1

^{††} 千葉商科大商経学部 〒 272-8512 千葉県市川市国府台 1-3-1

^{†††} 筑波大学大学院システム情報工学研究科 〒 305-8573 茨城県つくば市天王台 1-1-1

E-mail: [†]kamie@kde.cs.tsukuba.ac.jp, ^{††}takako@cuc.ac.jp, ^{†††}kitagawa@cs.tsukuba.ac.jp

あらまし Web 動画は増加の一途をたどっており、Web 動画集合のクラスタリング技術が注目されている。多くの既存手法ではメタデータのテキスト情報（タイトル等）と動画本体の画像・音声情報を利用するというアプローチをとっている。しかし、動画のテキスト情報の情報量は少なく品質が悪いことから解析が困難であり、一方、画像・音声処理の技術は未だ動画の内容の類似度を測れるまでに至っていない。本研究では、動画共有サイトの提供している「再生リスト」機能に着目する。再生リストはユーザが気に入っている動画をカテゴリ化できる機能と考えることが出来、同一の再生リストに含まれている動画が類似度を持っていると考えられる。本研究では再生リスト情報が動画間の類似性情報を含むことを検証し、これを利用した Web 動画のクラスタリング手法を提案し、その効果を検証した。
キーワード web 動画, クラスタリング, ユーザ生成リスト, 動画共有サイト

Web Video Clustering using Playlist

Mariko KAMIE[†], Takako HASHIMOTO^{††}, and Hiroyuki KITAGAWA^{†††}

[†] Graduate School of Systems and Information Engineering, University of Tsukuba Tennodai1-1-1, Tsukuba-shi, Ibaraki, 305-8573 Japan

^{††} Faculty of Commerce and Economics, Chiba University of Commerce Konodai1-3-1, Ichikawa-city, Chiba, 272-8512 Japan

^{†††} Graduate School of Systems and Information Engineering, University of Tsukuba Tennodai1-1-1, Tsukuba-shi, Ibaraki, 305-8573 Japan

E-mail: [†]kamie@kde.cs.tsukuba.ac.jp, ^{††}takako@cuc.ac.jp, ^{†††}kitagawa@cs.tsukuba.ac.jp

Abstract The increase of videos on the Internet has made video clustering more important. Most existing approaches focus on visual or text information to analyze videos. However, existing approaches have problems: text information in metadata tends to be of low quality, and visual or audio information is difficult to analyze. In this paper, we propose a playlist-based video clustering method, a novel framework that can form new clusters independent of text or visual similarities. A playlist is useful because it is based on the viewers' knowledge or intuitiveness; beyond that, it is not noisy. Our experiments show good result clusters generated by playlist-based clustering and prove that it can capture relativity or proximity among videos.

Key words web video, clustering, user-generated list, video-sharing website

1. はじめに

近年、YouTube 等の動画共有サイトが普及し、動画共有サイト上に存在する Web 動画（以下、動画と記す）の数は増加の一途をたどっている。その結果、動画共有サイトではキーワード検索の結果でさえも大量の動画が得られるようになった。一方で、多大な数の動画を検索結果として得た時、「検索結果全体の概要を把握したい」、「検索結果動画集合を要約して欲し

い」というユーザの要求が生まれている。この問題に対するソリューションの一つとして、「Web 動画のクラスタリング技術」がある。

Web 動画のクラスタリングでは、「動画間の類似度計算が難しい」という課題がある。多くの既存研究では、動画の画像や音声特徴、メタデータのテキスト特徴（タイトルや説明文、等）のいずれか、あるいはそれらの組み合わせを利用して動画間の類似度を計算するというアプローチをとっている。それらの研

究では画像特徴量や音声特徴量，音声認識結果，メタデータのテキスト情報が類似する動画を類似していると判定する。しかし，画像が類似していることが必ずしも動画の内容が類似していることにつながらないという問題や，音声品質が低い動画，口調が早い動画など，音声解析が難しい動画が多いという問題，更にテキスト情報は情報量が少ない上に，自分が投稿した動画を多くの人に見てもらうために関係のない単語をテキスト情報に加えるなどといったノイズの為に解析が難しいという問題がある。

一方，近年複数の動画共有サイトで普及している機能として，「再生リスト」がある。再生リストはユーザが任意に作ることの出来る動画のリストで，ユーザは再生リストに登録されている動画に素早くアクセスする事ができる。一人のユーザが複数の再生リストを作成可能なので，音楽動画を集めた再生リストやスポーツに関する動画を集めた再生リストなど，再生リストを利用して頻繁にアクセスする動画をカテゴリ化することが出来る。その場合，同一再生リストに含まれる動画間にはなんらかの類似性があると考えられる。我々はこの「再生リスト」に着目し，再生リスト情報に基づいて動画間の類似度計算を行う。そして画像や音声，テキスト情報に依存しない高精度な動画クラスタリング手法を提案する。

本稿では最初に，予備実験によって同一再生リストに含まれる動画間に類似性があるかどうかを確認した後，再生リスト情報に基づく動画クラスタリング手法を提案，実験によって提案手法の効果を定量的に示す。

本稿の構成は，以下の通りである。2章で前提知識として再生リストの詳細と実験の動画検索で利用したキーワードの選択方法を説明し，3章で関連研究について述べる。4章で再生リストの動画クラスタリングにおける有効性を検証する予備実験について説明し，5章で提案手法，6章で評価実験について述べ，最後に7章でまとめと今後の課題を述べる。

2. 前提知識

2.1 再生リスト

再生リストは YouTube が提供するサービスである。ユーザは再生リストに自分が頻繁に見る動画や気に入っている動画を登録しておくことでそれらの動画に素早くアクセスすることが出来る。また，同一再生リストに登録されている動画集合は連続再生が可能である。各ユーザは複数の再生リストを作成でき，各再生リストは ID，タイトル，説明文，0~200 件の動画リストを持つ（図 1）。

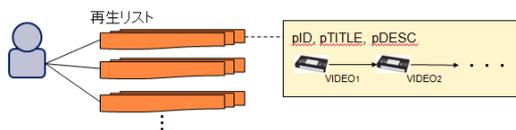


図 1 YouTube 再生リスト

YouTube 以外の複数の動画共有サイトでも，これと類似の機能を「お気に入り動画 / favorites」として提供している。こ

れらのほとんどのサイトでは，動画を「リスト」ではなく順序なし，重複ありの「多重集合」として管理する。しかし各ユーザが動画多重集合を複数作成出来る場所は再生リストと同じであり，「お気に入り動画 / favorites」機能も頻繁にアクセスする動画のカテゴリ化に利用出来る。2011 年 12 月現在で，再生リストと類似の機能を備えている動画共有サイトには次がある：ニコニコ動画 [1]，Veoh [2]，PANJEA [3]，Godtube [4]。

2.2 評価クエリの選択方法

本研究ではキーワード検索結果の動画集合に対して動画クラスタリングを行う。そのため，予備実験や評価実験ではキーワード検索を行った結果についての調査や評価を行う必要があり，その際に使用するキーワード（評価クエリ）を選択する必要がある。

様々のジャンルの動画/評価クエリに対して調査や実験を行うため，なるべく幅広いジャンルから一般的なキーワードを選出することが望ましい。本実験では様々な動画をカテゴリ化する YouTube カテゴリを利用して評価クエリを選出した。YouTube カテゴリは全 15 カテゴリからなるが，現在「非営利団体と社会活動」カテゴリは事実上使用不可となっているため，「非営利団体と社会活動」カテゴリを除く 14 カテゴリの各カテゴリからそれぞれ以下の手順で評価クエリを抽出する。

- (1) 各 YouTube カテゴリ毎，再生数の多い 100 動画を取得
- (2) 取得した動画中から出現頻度の高い 10 単語を選出
- (3) 10 単語の中から，著者がクエリにふさわしいと判断できる n 単語を選出

3. 関連研究

本章では既存のクラスタリングを始めとする Web 動画の解析に関する研究と，ユーザ生成リストを利用した研究について述べる。Web 動画の解析に関する関連研究では，「内容が類似する動画を検出する必要がある」という点で同じ課題を持つ，Web 動画のタグ付け・カテゴリ分けの研究を同類問題と見なし，Web 動画のクラスタリングに関する研究と一緒に紹介する。

3.1 Web 動画のクラスタリングに関する研究

動画コンテンツの増加に伴い，クラスタリングやタグ付け，カテゴリ分け等の動画解析に関する研究が数多く発表されている。

画像特徴量に着目して動画間の類似度計算を行っている研究としては [5]~[10] がある。S.Cheung ら [6] や Z.Huang ら [11] は，色ヒストグラム等の情報から，動画間で類似した画像フレームを見つけることで類似動画を検出している。M.J.Roach ら [5] や X.Yuan ら [7]，M.Montagnuolo ら [9]，I.Bogdan ら [10] は画像キーフレームから抽出した特徴以外にも，カメラモーションや動画中に映る物体の動き，ショットの長さ，ショットの分布やリズムなど，様々な画像特徴量を用いている。

またテキストや画像情報など，複数情報源から抽出した特徴量をいかに統合するかに着目した研究として [12]~[14] がある。D.Zhang ら [12] や J.Shao ら [13] はテキスト情報や音声情報，その他あらゆる情報源から得た特徴量，それぞれの長所を生

かせるような動画のクラスタリングを行うことを目的として、どちらもグラフベースのクラスタリング手法を提案している。A.Hindleら[14]は複数特徴量を使用した動画クラスタリングにも既存のクラスタリング手法を当てはめるため、種類の異なる特徴量を適切に融合させる方法を提案している。

教師付き学習に基づく動画の解析を目的とし、正解データの自動作成に着目した研究として[15]~[17]がある。H.Aradyeら[15]は人手による正解データの作成を必要とせずに特徴量の学習を行い、動画のタグ付け手法を提案している。この手法では人手で作られたタグ付けの正解データの代わりに、動画のテキスト情報を正解タグとして利用している。また特徴量は画像と音声情報から算出している。K.Filippovaら[17]は[15]で作成したタグ情報を正解データと見なし、更に動画のテキスト情報による学習を行っている。これにより、実際のタグ付け作業ではテキスト情報からのタグ付けが可能になり、計算コストが非常に低くなる。一方、W.Yangら[16]は[15]のタグ作成の際に視聴履歴情報を利用することでより精度のよいタグ推定をおこなっている。

また近年、画像やテキスト、音声特徴量以外の新たな情報源として、ウェブ上の情報を用いた動画解析手法が提案されている[18],[19]。K.Jensら[18]は動画のテキスト情報から動画の内容をWikipediaエントリと結びつけ、Wikipediaエントリに付随しているWikipediaカテゴリを基に、動画のカテゴリ化を行うことを提案している。Z.Chenら[19]は動画のテキスト情報からクエリを作成し、ウェブでキーワード検索することで動画に関連するイベントを検出し、タグ付けを行うことを提案している。

異なる情報源から作成した特徴量の合成方法やある情報源からのデータを有効に活用する方法等、様々な視点から動画解析を行う研究が多く発表されている。しかしどの研究も画像や音声、メタデータテキスト情報を利用しており、それらの情報源の持つ難点が依然問題として残る。Wikipediaや検索エンジンなど、新しい情報源の活用した手法についても提案されている[18],[19]が、どちらの研究もやはり動画のテキスト情報に依存しており、「テキスト情報の情報量と信頼性が低い」という問題は解決されない。本研究では、新しい情報源として「再生リスト」に着目する。

3.2 ユーザ生成リストを利用した研究

近年、ユーザ自らがその知見や嗜好を反映したリスト(以下、ユーザ生成リストと記す)を作成出来るサービスが、さまざまなWebサービスで提供されている。

例として、はてなブックマーク[20]やdelicious[21]が提供するソーシャルブックマーク、Twitter[22]が提供するユーザリストの作成機能がある。ソーシャルブックマークとは、自分が頻繁にアクセスするWebサイト等のURLにタグを付けて管理(ブックマーク)し、それをインターネット上で公開・共有するサービスである。例えば、音楽関連のWebサイトには「音楽」、仕事関連のWebサイトには「仕事」とタグを付けておくことで、頻繁にアクセスするWebサイトを整理する事が可能となる。一方、Twitterの提供するユーザリストの作成機

能とは、例えば、自分の学校の友達のTwitterアカウントには「学校」、芸能人のTwitterアカウントには「芸能人」とタグを付けることで、自分の注目するTwitterユーザを管理/整理するための方法である。タグの名前を指定するとそのタグが付けられているユーザー一覧やそのユーザのツイート一覧を見ることが出来る。

ユーザ生成リストの普及に伴い、ユーザ生成リストを利用した研究がいくつか行われている。B.Krauseら[23]はソーシャルブックマーク情報を利用し、スパマーの検出を、T.Takahashi, M.Harvey, D.Carmelら[24]~[26]はウェブページの推薦を行っている。X.Liら[27]はソーシャルブックマークの情報から、社会で注目されている話題の抽出を行っている。

Twitterのユーザリストを利用した研究としては、S.Wuら[28]の研究やY.Yamaguchi[29]らの研究がある。S.Wuら[28]はTwitterのリスト情報を用いて、マイクロブログ上でのユーザの交流関係や情報の流れを調査、Twitterに流れているURLの情報源が限られていることや同業者間での交流が多いことを明らかにしている。Y.Yamaguchi[29]らはtwitterのユーザ生成リストを利用し、タグの推薦を行っている。

ソーシャルブックマークやTwitterのユーザ生成リストを利用した研究は多々提案されているが、YouTubeのユーザ生成リストである再生リストを利用した研究は未だ行われていない。サービスや対象メディアが違えば、それぞれのユーザ生成リストは異なった特徴を持つと思われる。ソーシャルブックマークやTwitterのユーザ生成リストは基本的にテキストベースのウェブページやツイートに対して行われるものであり、動画コンテンツの内容理解に関してユーザ生成リストを利用する研究は著者の知る範囲では本研究が初めてである。

4. 予備実験：再生リストの有効性検証

最初に、再生リスト情報の動画クラスタリングにおける有効性を検証するため、「同一再生リストに登録されている動画間に類似度があるか」を検証する予備実験を行った。動画間の類似性は「同一再生リストに登録されている動画集合になんらかの共通するトピックがあるか」というトピックの観点から観察した。

実験は $n = 1$ として2.2節の方法で選択した14個の評価クエリを基に、次の手順で行う：

- (1) 各評価クエリを基に再生リストを検索。
- (2) 再生リスト検索結果のランキングトップから最下位まで、均等間隔で選択した20件の再生リストを取得。
- (3) 各再生リストに保存されている動画の内、1番目の動画から最後の動画まで、20件の動画を均等間隔で選択し閲覧。
- (4) 各再生リストを評価。

再生リストの評価は「トピック統一度」として、動画集合に共通するトピックがあるかどうかを次の3段階で評価した：「 \cup 」：全動画に共通するトピックがある、「 \cap 」：ほとんどの動画に共通するトピックがある、「 \times 」：共通するトピックがない。更に全動画、あるいはほとんどの動画に共通するトピックがある場合は、その再生リストにトピックを表すラベル付けを行った。

各カテゴリから選択した 14 個の評価クエリは次のようになった。なお () 内はカテゴリ名を示す。ピカチュウ (映画とアニメ), JR(自動車と乗り物), kpop(音楽), ねこ (ペットと動物), 野球 (スポーツ), 温泉 (旅行とイベント), ポケモン (ゲーム), 初音 (コメディ), NG 集 (ブログと人), 地震 (ニュースと政治), マジック (エンターテインメント), lesson(ハウツーとスタイル), 料理 (教育), 原発 (科学と技術)。

この 14 個の評価クエリを用いて再生リストを検索したところ、全体で 280 個の再生リストが得られた。各再生リストについて、そのトピック統一度を評価した結果を表 1 に示す。約 85% (237 個) の再生リストにおいて、全動画に共通するなんらかのトピックが見られた。

表 1 トピック統一度

再生リスト数			x
237	10	33	

作成したラベルの例として「料理」という評価クエリに対して作成したラベルを表示する：料理 (3), 食べ物 (1), レシピ (2), 汁もの (1), 日本料理 (2), とんかつ (1), 中華料理レシピ (1), 松本人志 (1), 音楽 (1), 邦楽 (4) () 内の数字はそのラベルが付けられた再生リストの数を示している。料理という評価クエリに対しては 10 種類 17 個のラベルを作成した。ほとんどの再生リストが料理に関するものであり、一部、再生リスト全体としては料理に関係のない再生リストも混じていたことが分かる。

予備実験により、ほとんどの再生リストがなんらかのトピックを持って作成されており、同一再生リストに登録されている多くの動画間になんらかの類似性があることが分かった。

5. 提案手法

本研究では、画像や音声、テキスト情報に依存しない高精度な検索結果動画集合のクラスタリングを目的とする。我々は再生リスト情報に着目し、これを利用したクラスタリング手法を提案する。対象動画共有サイトは YouTube とし、ユーザーからの検索クエリを入力とする。

提案手法は次の 3 ステップからなる：1. 動画情報の取得 (再生リストの情報等), 2. 動画間の類似度計算 (再生リスト情報を利用), 3. 動画クラスタリング。順に各ステップについて説明する。

5.1 動画情報の取得

本ステップでは各評価クエリに対して、次の情報を取得する：

- クラスタリング対象となる動画集合
 - 動画集合の各動画について「動画 ID」、「テキスト情報」「リスト情報」の 3 種の情報
- 「テキスト情報」とは動画のタイトル, 説明文, タグ, 最近のコメント最大 50 件を指し, 「リスト情報」とは動画が登録されている再生リストの ID を指す。

これらの情報を取得するには、図 2 のように、動画のキーワード検索によってクラスタリング対象動画集合を取得し、そ

の後、各動画のテキスト情報とリスト情報を順に取得する、というプロセスが単純である。しかし現在、YouTube は各動画が登録されている再生リストの情報を提供していない。そこで本研究では YouTube が提供している「再生リストのキーワード検索」機能で代用し、類似の機能を実現した。

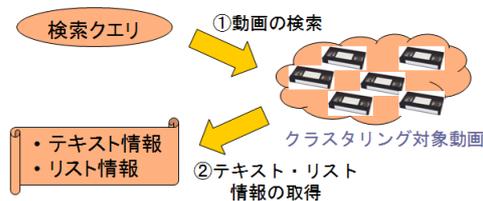


図 2 単純な動画情報の取得方法

再生リストのキーワード検索では、検索キーワードが「再生リストのタイトルや説明文」、あるいは「再生リストに登録されているいずれかの動画のタイトルや説明文」とマッチする再生リストを検索していると見られる。したがって、動画のタイトルをクエリとして再生リストを検索すると、その動画が登録されている再生リストを取得出来る可能性があると考えられる。これを利用して図 3 のような情報収集のためのクロールシステムを開発した。

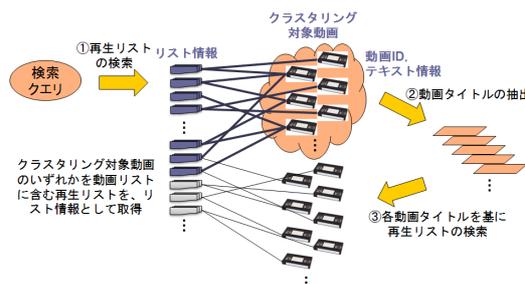


図 3 開発したクロールシステム

システムはユーザーが入力したクエリを入力とし、次のように動作する：1. クエリを利用して再生リストのキーワード検索を行う。本研究では、この際に取得した再生リストに含まれる動画集合をクラスタリング対象動画集合とし、この動画集合から動画 ID とテキスト情報を取得する。2. クラスタリング対象動画集合からタイトルを抽出。3. 抽出したタイトル集合の各タイトルをクエリとして再生リストの再検索を行う。4. 最初の再生リストの検索で取得した再生リストに加え、再生リストの再検索で取得した再生リストの内、クラスタリング対象動画集合のいずれかの動画を含む再生リストをリスト情報として取得する。

5.2 動画間の類似度計算

取得した情報を基に動画間の類似度を計算する。本研究では、テキスト類似度, リスト類似度, 合成類似度の 3 種類の動画類似度を計算する。

テキスト類似度は動画のテキスト情報 (タイトル, 説明文, タグ, コメント) を基に、最もシンプルと言える文章処理の技術を用いて計算した。最初に、テキスト情報から抽出した単語

を元に、tf-idf で重み付した単語頻度ベクトルを作成、LSI によって次元圧縮を行い「テキスト特徴量」を作成した。LSI の次元圧縮はフロベニウスノルムに基づく累積寄与率^(注1) α が 0.8 となるまで行った。テキスト類似度はテキスト特徴量のコサイン類似度で定義した。

リスト類似度は動画のリスト情報に基づいて算出する。リスト情報から図 4 のような、「動画-再生リスト」の登録関係行列を作成しその行列に対してノイズ除去と計算コスト軽減のため、LSI に基づく次元圧縮を行う。これをリスト特徴量と定義する。リスト類似度はテキスト類似度と同様に、リスト特徴量のコサイン類似度で定義する。

		再生リスト		
		L1	L2	L3
動画	V1	1		
	V2	1		1
	V3	1	1	

「動画-再生リスト」登録関係行列

図 4 「動画-再生リスト」登録関係行列

合成類似度 ms はテキスト類似度 ts とリスト類似度 ls の重み付き平均で定義する。

$$ms = (1 - \theta) * ts + \theta * ls \quad (1)$$

$\theta = [0, 1]$ は合成割合のパラメータとし、 θ が大きい程リスト類似度の割合が大きくなる。

5.3 動画クラスタリング

作成した類似度を基に動画集合のクラスタリングを行う。クラスタリングには比較的高速で妥当な精度が出るとされている Bisecting K-means [30] を利用する。Bisecting K-means は簡単に言うと、所望のクラスタ数が得られるまで、kmeans による 2 クラスタへの分割を繰り返すクラスタリング手法である。分割は、分割前後で大きな変化がなくなるまで分割する為、分割前後の cohesion の差が 1.0 になるまで分割を行った。cohesion は全要素のセントロイドまでの距離の和とする。

6. 評価実験

提案手法の有効性を検証するため、以下に述べる 3 つの評価実験を行った：

(1) リスト特徴量の次元圧縮効果を「クラスタリング精度」と「時間コスト」の 2 つの観点から検証する、「リスト特徴量の次元圧縮の有効性検証」。

(2) テキスト、リスト、合成類似度間で最も動画クラスタリングに向いている類似度を検証するための「最適な類似度種類の検証」。

(3) 合成類似度において、テキストとリスト類似度の割合の変化によるクラスタリング精度の変化を検証する「合成割合の変化による精度の推移」。

(注1)：徳永 健伸, 辻井 潤一: “情報検索と言語処理 (言語と計算)”, 東京大学出版会 (1999-11)

本章ではまず、6.1 節でクラスタリング精度の人手による測定方法について説明し、6.2 節で実験のために取得したデータセットについて説明、その後 6.3 節～6.5 節で各実験の方法と結果について説明する。なお、実験では人手によるクラスタリング結果の評価が多く必要になった。被験者の負担を減らす為、2 番目の「最適な類似度種類の検証」以外の実験ではクラスタリング結果の評価は著者が一人で行った。

6.1 クラスタリング精度の測定方法

クラスタリング結果がどれだけ人の感覚にマッチしているかを検証する、「クラスタリング精度の測定方法」について説明する。

評価者にはクラスタリング結果として、10 クラスタを提示し、クラスタ毎にクラスタを代表する 10 動画を提示した (図 5)。評価者は全ての動画を閲覧した後に、クラスタリング結果



図 5 被験者に提示した画面の例

の妥当性を評価する。クラスタリング結果が妥当であるとは、「動画集合がその内容毎に適切にグルーピングされている」ことであるとした。クラスタリング結果の妥当性は 4 段階評価 (4: 妥当だと思う / 3: まあまあ妥当だと思う / 2: あまり妥当だと思わない / 1: 妥当だと思わない) とし、その評価値を付けた理由も記入して頂いた。

提示する 10 クラスタは、10 動画以上を含むクラスタをクラスタに含まれる動画数順に並べ、一番動画数が少ないクラスタから一番動画数が多いクラスタまで、均等間隔で選択したものとした。クラスタを代表する 10 動画はクラスタのセントロイドに近い 10 動画とした。

6.2 データセット

$n = 3$ として 2.2 節の手順に従い、各 YouTube カテゴリ毎に 3 クエリ、合計で 42 個のキーワードを評価クエリとして選択した。表 2 に選択した評価クエリを示す。また、開発したクローリングシステムを利用し、2011 年 9 月 10 日～2011 年 11 月 25 日の間、実験用データのクローリングを行った。表 3 に各評価クエリに対して取得した動画集合の平均動画数と各特徴量の次元圧縮前の平均次元数を示す。また、表 4 に累積寄与率 $\alpha = 0.8$ $\rho = 0.6$ で次元削減した結果得られた特徴量の平均次元数を示す。表から、特徴量の次元削減を行わない場合リスト特徴量の方が次元数が大きく、次元削減を行うと、リスト特徴

表 2 評価クエリ一覧 (YouTube カテゴリ毎に 3 つ)

YouTube カテゴリ	評価クエリ
映画とアニメ	アニメ, ピカチュウ, naruto
自動車と乗り物	車, JR, ドリフト
音楽	AKB, kpop, PV
ペットと動物	ねこ, スコティッシュフォールド, モアレ
スポーツ	ルービックキューブ, 野球, ブラインドサッカー
旅行とイベント	旅行, 東京, 温泉
ゲーム	ポケモン, playstation, 遊戯王
コメディ	おもしろ, Elvira, NG 集
ブログと人	マギボン, ブログ, 初音
ニュースと政治	TBS, 尖閣諸島, 地震
エンターテインメント	マジック, キャンペーン, CM
教育	minxy, こども, lesson
ハウツーとスタイル	クッキング, How, 折り紙
科学と技術	iPhone, robot, 原発

表 3 クラスタリング対象動画集合の平均ビデオ数, 特徴量の平均次元数

ビデオ数	テキスト特徴量	リスト特徴量
3060	2230	3679

表 4 クラスタリング対象動画集合の平均ビデオ数, 特徴量の平均次元数

テキスト特徴量		リスト特徴量	
$\alpha = 0.8$	$\alpha = 0.6$	$\alpha = 0.8$	$\alpha = 0.6$
214	90	90	52

量の方が次元数がテキスト特徴量の半分程度まで小さくなるのが分かる。これは再生リストに含まれる動画の相関が高いためと考えられる。

表 5 に各特徴量を使用したクラスタリング結果の平均クラスタ数を示す。

表 5 特徴量毎の平均クラスタ数

	$\alpha = 1.0$	$\alpha = 0.8$	$\alpha = 0.6$
テキスト特徴量	268	213	146
リスト特徴量	67	60	51

テキスト特徴量に基づくクラスタリングのクラスタ数が, リスト特徴量に基づくものよりも大きくなっていることが分かる。これは, リスト特徴量では早い段階である程度クラスタリング結果がまとまり, それ以上の分割では cohesion に大きな差が出来なくなったためである。

6.3 リスト特徴量の次元圧縮の有効性検証

リスト特徴量の次元圧縮の有効性検証では, 下記二点を検証する。

- (1) ノイズ除去が有効に働いているか。
- (2) 類似度計算にかかる時間コストはどの程度軽減されるか。

検証は次元圧縮前後の「クラスタリング精度」, 「動画の類似度計算の実行時間」を比較することで行う。類似度計算の実行時

間の比較では, 作成した動画集合の特徴量からコサイン類似度 (余弦尺度) を求めるのに要した時間の平均を比較する。次元圧縮する前のリスト特徴量と累積寄与率 $\alpha = 0.8, 0.6$ で次元圧縮したリスト特徴量の 3 種類の特徴量間で「クラスタリング精度」と「類似度計算の実行時間」を比較する。

クラスタリング精度は図 6 のようになった。 $\alpha = 1.0$ は次元

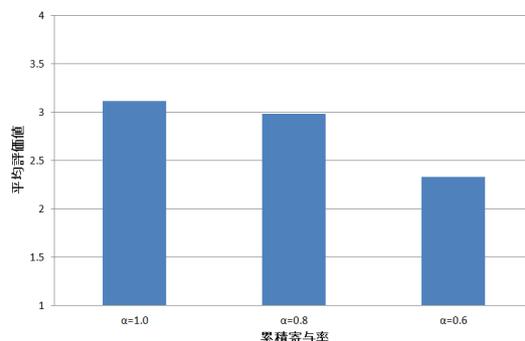


図 6 次元圧縮の有効性検証

圧縮する前の特徴量に基づくクラスタリング結果を示している。 $\alpha = 0.6$ で大きくクラスタリング精度が低下しているが, 次元圧縮前と $\alpha = 0.8$ で次元圧縮した特徴量間ではクラスタリング精度に大きな違いがなかった。

類似度計算の実行時間は表 6 のようになった。 $\alpha = 0.8$ で計

表 6 類似度計算の平均実行時間 (秒)

	$\alpha = 1.0$	$\alpha = 0.8$	$\alpha = 0.6$
実行時間	170.3628	21.66968	13.6432

算時間が 8 分の 1 に, $\alpha = 0.6$ で更に, その 2 分の 1 になっていることが分かる。

実験結果全体としては, リスト特徴量の次元圧縮によってクラスタリング精度向上にはならなかったが, 時間面で大きな速度向上になっていることが分かった。

6.4 最適な類似度種類の検証

テキスト類似度, リスト類似度, 合成類似度の 3 類似度の内, 動画クラスタリングに最適な類似度の種類を検証するため, 被験者 8 名によるクラスタリング精度の測定を行った。被験者には, 1 人 4 クエリ 3 種類のクラスタリング結果, 全体として 12 個のクラスタリング結果を評価して頂いた。評価するクエリは 6.2 で説明した評価クエリの内, 被験者が興味のある 4 クエリを選択して頂いた。リスト類似度に基づくクラスタリング結果では次元圧縮前のリスト特徴量を用いて計算したものを評価し, 合成特徴量は合成比率 $\theta = 0.875$ で合成した特徴量に基づくクラスタリング結果を評価した ($\theta = [0, 1]$ で, θ の値が大きい程リスト類似度の割合が高い)。

図 7 に結果を示す。リスト類似度に基づくクラスタリング結果が最も高い精度を示した。

評価値の理由としては, 良い評価値を持っているクラスタリング結果に対しては, 「全体的にクラスタとしてまとまっている」, 「初めの方のクラスタは, 歌 (MV, ライブなど) の映像

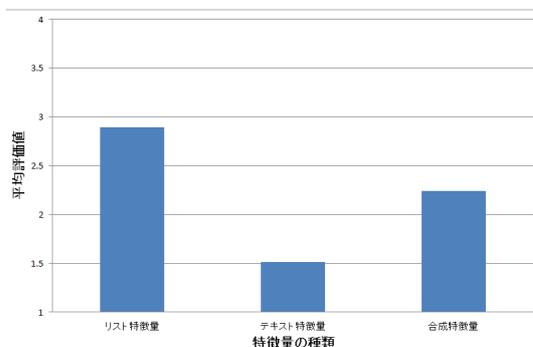


図7 最適な類似度種類検証

や、特典映像などでわかっていた」、「ゲームの映像、アニメの映像、音楽などが比較的よく分類されているように思う。」などというコメントが得られた。逆に、悪い評価値となったクラスタリング結果に対しては、「音楽の映像、議会の映像、動物の映像などが全クラスタにわたって入り乱れており、まとまりを感じられなかった」、「いくつかのクラスタのグルーピングが不完全」、「各クラスタにほとんど統一性がないように感じた。BGMや改造動画が複数のクラスタに点在していた。」というようなコメントを頂いた。「特典動画のクラスタがある」などといった、各被験者の専門性を反映したコメントがあったことから、各被験者にクエリを選択してもらい、各被験者が多少なりとは興味があり、知識があるクエリに関する動画集合のクラスタリング結果を評価してもらったことの有効性も確認できた。

6.5 合成割合の変化による精度の推移

合成割合の変化による精度の推移では、 $\theta = 0.125, 0.25, 0.5, 0.75, 0.875$ と θ の値を変化させ、クラスタリング精度を測定した。実験結果を図8に示す。リスト類似度の割

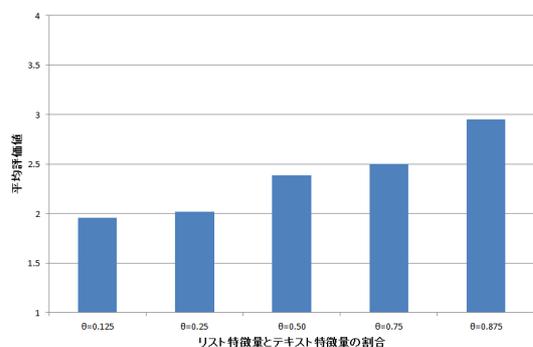


図8 合成割合の変化による精度の推移

合が高い程クラスタリング精度が高いことが分かる。リスト情報がテキスト情報に比べて動画クラスタリングに向いていること、2特徴量間の合成方法を再度検討する必要があることが示唆できる。

7. まとめと今後の課題

再生リスト情報の動画クラスタリングにおける有効性を検証し、再生リスト情報に基づく動画の特徴量を提案した。実験により、テキスト情報に比べて再生リスト情報が動画クラスタリ

ングに適していることを示した。今後の課題として、被験者数を増やした実験と評価を行うこと、動画のタグ付けへの応用を考えている。

8. 謝 辞

本研究は科学研究費補助金・基盤研究 A(# 21240005) の助成を受けたものである。

文 献

- [1] Nico nico douga. <http://www.nicovideo.jp/>.
- [2] Veoh. <http://www.veoh.com/>.
- [3] Panjea. <http://www.panjea.com/>.
- [4] Godtube. <http://www.godtube.com/>.
- [5] M.J.Roach et al. *Video genre classification using dynamics*, 2001.
- [6] S.Ching et al. *Efficient video similarity measurement with video signature*. IEEE Transactions on Circuits and Systems for Video Technology, 2002.
- [7] X.Yuan et al. *Automatic video genre categorization using hierarchical svm*. Science And Technology, 2006.
- [8] S.Liu et al. *Mining similarities for clustering web video clips*. CSSE '08.
- [9] M.Montagnuolo et al. *Parallel neural networks for multi-modal video genre classification*. Multimedia Tools Appl., 2009.
- [10] L.Bogdan et al. *Content-based video description for automatic video genre categorization*. In MMM, 2012.
- [11] Z.Huang et al. *Mining near-duplicate graph for cluster-based reranking of web video search results*. ACM Trans. Inf. Syst., 2010.
- [12] D.Zhang et al. *Semantic video clustering across sources using bipartite spectral clustering*. In ICME, 2004.
- [13] J.Shao et al. *Topic discovery of web video using star-structured k-partite graph*. In ACM Multimedia, 2010.
- [14] A.Hindle et al. *Clustering web video search results based on integration of multiple features*. World Wide Web, 2011.
- [15] H.Aradye et al. *Video2text: Learning to annotate video content*. ICDMW '09.
- [16] W.Yang et al. *Discriminative tag learning on youtube videos with latent sub-tags*. 2011.
- [17] K.Filippova et al. *Improved video categorization from text metadata and user comments*. SIGIR '11.
- [18] K.Jens et al. *Videoclef 2008: Asr classification based on wikipedia categories*. 2008.
- [19] Z.Chen et al. *Context-oriented web video tag recommendation*. WWW '10.
- [20] Hatena bookmark. <http://b.hatena.ne.jp/>.
- [21] delicious. <http://delicious.com/>.
- [22] Twitter. <http://www.twitter.com/>.
- [23] B.Krause et al. *The anti-social tagger: detecting spam in social bookmarking systems*. AIRWeb '08.
- [24] T.Takahashi et al. *A ranking method for web search using social bookmarks*. DASFAA '09, 2009.
- [25] M.Harvey et al. *Improving social bookmark search using personalised latent variable language models*. WSDM '11.
- [26] D.Carmel et al. *Social bookmark weighting for search and recommendation*. The VLDB Journal.
- [27] X.Li et al. *Tag-based social interest discovery*. WWW '08.
- [28] S.Wu et al. *Who says what to whom on twitter*. WWW '11.
- [29] Yuto Yamaguchi, Toshiyuki Amagasa, and Hiroyuki Kitagawa. *Tag-based user topic discovery using twitter lists*. In ASONAM, pp. 13–20, 2011.
- [30] S.Michael et al. *A comparison of document clustering techniques*, 2000.