

動画共有サイトにおける映像検索のための 編集度順序によるランキング法の提案と評価

入江 豪^{†,††} 日高 浩太^{†,†††} 佐藤 隆[†]
小島 明[†] 相澤 清晴^{††}

本研究では、動画共有サイトにおける消費者生成映像 (Consumer Generated Videos: CGV) 検索の新たなランキング指標として、CGV が「編集されている程度」に基づく編集度順序を導入し、これを推定する手法の提案を行う。提案手法は、編集された程度によって変化するカット点の数や、音楽区間の時間割合などの画像・音響的な特徴を利用し、評価者が主観的に判断した編集度順序を統計的に推定する。提案手法について、評価者が判断した編集度順序を、提案手法がどの程度推定可能であるかを、順位相関によって評価した。さらに、動画共有サイトでの利用を想定した効果を検証するため、約 22,000 件の CGV データベースを用いて、評価者実験を実施した。結果、提案手法が、現在の方式によるランキング結果を改善できることを示唆する結果を得た。

A Degree-of-Edit Ranking for Retrieval on Video Sharing Sites and Its Evaluation

GO IRIE^{†,††} KOTA HIDAKA^{†,†††} TAKASHI SATOU[†] AKIRA KOJIMA[†]
and KIYOHARU AIZAWA^{††}

We introduce *degree-of-edit* ranking to focus on “how much a Consumer Generated Video (CGV) is edited” as a ranking measure for CGV retrieval on video sharing sites; a method to estimate degree-of-edit ranking is proposed. In the proposed method, the degree-of-edit score of a CGV is estimated by using audio/visual features such as the number of shot boundaries and time ratio of music. We evaluate the rank correlation between degree-of-edit ranking determined by subjects and by our method. To demonstrate its performance in a practical scenario, application to video sharing sites, a user test is performed on over 22,000 CGVs in the context of CGV search. Obtained results suggest that our method significantly improves conventional CGV ranking results.

1. はじめに

『Youtube』や『ClipLife』に代表される動画共有サイトが普及し、公開される消費者生成映像 (Consumer Generated Video: CGV) の数は膨大なものとなっている。動画共有サイトで CGV を視聴するユーザは、タグやタイトルに代表される周辺テキストを利用したキーワード検索を行い、視聴したい CGV を探

しだす必要があるが、この際、より多くのユーザの興味を惹く CGV を発見しやすくするためには、検索結果に含まれる CGV の提示順序を効果的に制御するランキング指標の導入が不可欠である。

現在の一般的な動画共有サイトでは、ユーザの興味を惹く CGV をランキングの上位にあげるため、“視聴回数”や“コメント数”などの指標を用いて、キーワード検索結果をランキングできるようになっている。これらの指標によるランキングは、「多くのユーザが興味を持ちやすい CGV は、より多くのユーザによって視聴・コメントされている」という考えに基づいている。ユーザが CGV を視聴したり、コメントを付与したりするのは、その CGV に興味を持ったからである、

[†] 日本電信電話株式会社 NTT サイバーソリューション研究所
NTT Cyber Solutions Laboratories, Nippon Telegraph
and Telephone Corporation

^{††} 東京大学大学院 情報理工学系研究科
Graduate School of Information Science and Technol-
ogy, The University of Tokyo

^{†††} 慶應義塾大学大学院 政策メディア研究科
Graduate School of Media and Governance, Keio Uni-
versity

<http://www.youtube.com/>
<http://cliplife.jp/>

というのは比較的自然な考え方であることから、合理的な指標と認識されている。しかしながら、指標そのものがCGVの内容のみに依存しない外部指標であるため、いかに内容が優れ、より多くのユーザの興味を惹くようなCGVであったとしても、たまたまユーザの目に触れることがなければ、ランキングの上位にあがることのないまま埋もれてしまう、という問題も有している。したがって、何らかの観点からCGVの内容を評価し、ランキングに反映する手法の導入が必要となる。

CGVの特徴として、映像の作成経験もさまざまな、一般の消費者によって作成・公開された映像であることがあげられる。したがって、その内容に着目すれば、非常によく編集されているものから、それほど編集されていないものまで、さまざまなものが存在している。ここでいう編集とは、post-productionに含まれるものであり、例えば、音質や画質を改善する、あるいは、余分な内容を省いて主旨を明確にする、効果を付与して演出を施す、などの行為を指す。これらの編集行為は、映像の構造を整えると同時に、魅力・雰囲気高めことなどを目的として行われるため¹⁾、よりよく編集されているものほど、ユーザの興味を惹きやすいものになる傾向があると考えられる。このことから、視聴回数やコメント数などの外部指標に加え「編集されている程度」を表す内部指標を用いたランキング手法を導入することで、ユーザにとってより有益なランキングを実現できる可能性があるといえる。

本研究では、CGVが編集されている程度に基づくランキング指標として、編集度順序を導入し、これを自動推定する編集度順序推定法の提案と評価を行う。提案手法は、編集された程度によって変化するカット点の数や、音楽区間の時間割合などの画像・音響的な特徴を利用し、評価者が主観的に判断した編集度順序を統計的に推定する。まず、予備実験として、実際の動画共有サイトで流通するCGVを用い、評価者が判断した編集度順序を、提案手法がどの程度推定可能であるかを順位相関によって評価した。さらに、実際の動画共有サイトでの利用を想定した効果を検証するため、約22,000件のCGVからなるデータベースを用い、評価者実験を実施した。

2. 関連研究

2.1 映像インデクシング

映像中の構造に関する情報を検出し、これを用いて映像やシーンにインデクスを自動的に付与する手法が研究されている。これまで、カット点^{2)~4)}、カメラ

ワーク⁵⁾、音楽⁶⁾、テロップ^{7),8)}などのインデクスを検出する手法が研究されてきており、いずれもかなり高精度に検出することができるようになってきている。本研究では、これらを統合的に用いることによって、編集度順序を推定する手法を提案する。

2.2 映像検索結果のランキング

『YouTube』などの多くの動画共有サイトでは、公開日時のほか、視聴回数や、コメント数、ユーザが付与した評価値などのように、より多くのユーザに視聴・支持されたCGVが、検索結果の上位にあがるようなランキング方式が導入されている。一方で、動画共有サイトが普及し、CGVの検索が一般的に行われるようになってきたが、これに主眼をおいた研究例は数少ない。本研究では、動画共有サイトでの新たなランキング手法を提案するものであり、実際に流通するCGVを用いた評価実験によって、性能の検証を行う。

一般の映像検索におけるランキング手法として代表的なものに、画像片や映像片といった、テキストとは異なるメディア情報をクエリとして用いる類似検索型のアプローチ^{9),10)}がある。さらに、ユーザとのインタラクションを介することで、ユーザの検索意図を反映したランキング結果を提示するRelevance Feedback (RF)¹¹⁾と呼ばれる手法も研究されている。RFは、初期検索結果にリストされたいくつかの映像に対して、その映像がpositive(結果として望ましいもの)であるかnegative(望ましくないもの)であるかを、ユーザが直接評価する。例えば、画像片をクエリとした初期検索結果に対して、ユーザがpositiveであると判断した画像との類似性を利用して、検索結果を改善する手法が提案されている¹²⁾。RFでは、positive/negativeの決定の際、ユーザの操作を要求することになるが、これを自動的に類推することで、インタラクションを省略するPseudo Relevance Feedback (PRF)もある。初期検索リストのランキングに応じて、pseudo-positive(結果として望ましいと考えられる映像)とpseudo-negative(望ましくない映像)を決定し、これらとの画像特徴の類似度に応じて正、負のスコア補正を掛けてリランキングする手法が提案されている^{13),14)}。

これらのランキング手法が、クエリやユーザの行動に依存してランキングが変化する、動的な指標を利用してランキングを行うのに対して、提案手法は、クエリやユーザの行動に依存せず、CGVの内容に依存して定まる静的な尺度によってランキングする手法である。提案手法は、これらの動的な手法と競合するものではなく、むしろ併用することによって、より有益な

ランキング手法とすることもできると考えられる。

2.3 質の評価

映像の質を評価する手法に関する研究例として、音質・画質の評価手法に関するさまざまな取り組みがなされている^{15)~17)}。一方で、本稿で対象とする、編集の程度に着目した研究例はみられない。動画共有サイトでのCGV検索を目的とした場合、音質・画質に加えて、より意味内容の強い編集の程度に着目することは、ユーザの利便性を高める上で有益であると考えられる。

3. 編集度順序推定法

提案する編集度順序推定法について説明する。まず、提案手法のアプローチについて述べる。その後、提案手法で用いる特徴量と、編集度順序を推定する方法について述べる。

3.1 アプローチ

CGVの集合 V に含まれる n 個のCGV $\{v_1, v_2, \dots, v_n\} \in V$ が与えられたとき、編集度順序による順位 $\{o_V(v_1), o_V(v_2), \dots, o_V(v_n)\}$ を決定したい。ここで、 $o_V(v_i)$ は、CGV v_i の V 内の順位を表す。本稿では、 V によらず、CGV v_i の画像・音響特徴量(ベクトル) $x(v_i)$ のみに基づいて決定される編集の程度(編集度) $e(x(v_i))$ を考え、この降順によって、任意の V に対する編集度順序における順位 $o_V(v_i)$ が定まるものとすれば、本稿の問題は、関数 $e(x(v_i))$ を得ることである。ここで、 $e(x(v_i))$ を定める客観的な規則を与えることは困難であるため、本稿では、評価者により主観的に決定された編集度順序を正解事例とし、統計的に $e(x(v_i))$ を得る。

利用する特徴量 x は、編集度 $e(x(v_i))$ に関連するものであり、編集の程度に応じて変化しやすいものであることが望ましい。本稿では、CGVを含む映像一般に実施される、下記の4つのpost-productionを編集行為として定義する。

- (1) 映像区間の除去・結合、
- (2) 音質・画質の改善、
- (3) 視覚的効果の付与(テロップの挿入)、
- (4) 聴覚的効果の付与(音楽・BGMの挿入)。

これらの行為を行う場合、いくつかの特徴量に変化が起こる。例えば、(1)を実行するためには、図1(a)に示すように、カットを実行しなければならず、実行後の映像にはカット点が生じる。複数の映像をつなぎ合わせる場合にも同様にカット点が生じ、また、映像の時間長も長くなることが多い。あるいは、(4)のように音楽を挿入した場合には、図1(b)に示すように、音響信号に変化が起こり、音楽信号の時間割合が増加す

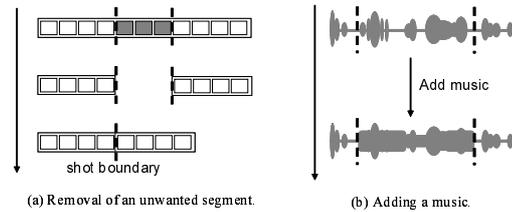


図1 編集行為による変化の例
Fig. 1 Examples of changes with an editing

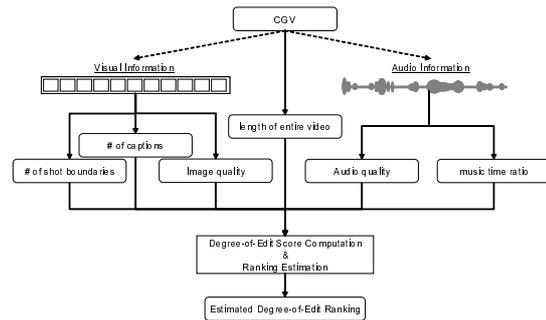


図2 編集度順序推定法の概要
Fig. 2 Proposed degree-of-edit ranking estimation method

る。このような考えに基づき、提案手法では、下記の6種類の特徴量を用いる。

- 映像の時間長： L
- カット回数： NC
- 音質： AQ
- 画質： IQ
- テロップの出現回数： NT
- 音楽区間の時間割合： RM

図2に、提案する編集度順序推定法の概要を示す。提案手法は、まず、CGV v_i の画像信号と音響信号から、前述の6種類の特徴量 $x(v_i) = \{L(v_i), NC(v_i), AQ(v_i), IQ(v_i), NT(v_i), RM(v_i)\}$ を抽出する。その後、これらの特徴量 $x(v_i)$ から、編集度順序を定める編集度 $e(x(v_i))$ を推定し、これに基づくランキングを行う。

3.2 特徴量抽出

前節で述べた6種類の特徴量のうち、容易に取得可能な映像時間長 L を除いた5種類の特徴量を抽出する方法について説明する。

3.2.1 カット回数

カット回数 NC を取得するためには、画像信号からカット点を検出し、映像全体に対して何回カットが検出されたかをカウントすればよい。カット検出手法には、多くの先行研究が存在する^{2),3)}が、本稿では、

高速かつ高精度にカット点を検出できる手法の一つである谷口らの方法を用いる⁴⁾。検出されたカットの数を、 NC とする。

3.2.2 音質・画質

音質や画質の尺度として、ビットレートや信号成分の大きさ(パワー)とノイズ成分の大きさの割合を表す SNR がある。後者については、CGV のように、ノイズ成分の周波数特性が予め想定しにくいような元信号では、厳密に求めることが難しい。そこで、提案手法では、音質 AQ と画質 IQ を次のように算出する。音質 AQ は、音響信号から 60Hz~400Hz の間で $F0$ が推定された音響信号区間を信号成分、それ以外の区間を雑音成分とみなし、前者の RMS 値 \bar{r}^s と、後者の RMS 値 \bar{r}^n によって、次式で計算する。

$$AQ = \log \frac{\bar{r}^s}{\bar{r}^n}. \quad (1)$$

$F0$ の抽出法は RAPT¹⁸⁾ を用いた。画質 IQ は (平均) ビットレート b と画像横サイズ w 、画像縦サイズ h を用い、次式によって計算する。

$$IQ = \frac{b}{w \times h}. \quad (2)$$

3.2.3 テロップの出現回数

テロップ出現回数 NT を取得するには、画像信号からテロップの出現する区間を検出し、映像中に何回テロップが出現しているかをカウントすればよい。テロップ検出手法としては、MPEG のマクロブロックを利用する手法⁷⁾ や、輝度分布から計算するエッジ密度を用いるもの⁸⁾ などが提案されている。提案手法では、CGV が必ずしも MPEG フォーマットでないことを踏まえ、桑野らによって提案された手法⁸⁾ を用いる。この手法では、テロップが出現してから消失するまでを一区間として検出することができる。この区間を一回の出現とし、出現回数 NT を求める。

3.2.4 音楽区間の時間割合

音楽区間の時間割合 RM は、映像時間長 L に対する音楽の含まれる区間の時間長 LM の比によって求める。提案手法では、Minami らにより提案された手法を用いて音楽区間を検出する⁶⁾。この手法は、音楽のスペクトルが、時間方向に安定したピークを有することを利用し、音楽の開始点 t_s 、終了点 t_e を検出する。検出された音楽区間毎に、 $\Delta t = t_e - t_s$ を計算し、その累積和 $LM = \sum \Delta t$ と映像時間長 L との比

$$RM = \frac{LM}{L} \quad (3)$$

によって、 RM を求める。

3.3 編集度と編集度順序の推定

提案手法では、評価者が主観的に判断した編集度順序を正解事例とした、統計的な推定を行う。本稿では、学習モデルとして、線形回帰 (Linear Regression: LR)、サポートベクター回帰 (Support Vector Regression: SVR)、および、Ranking SVM (RSVM)¹⁹⁾ の3つのモデルを利用した場合について検討を行う。これらを学習するための学習用データは、4.2.1 に後述する方法によって得る ground truth データを用いる。ground truth データは、評価者が判断した、CGV の編集の程度を7段階で表したものであり、編集度順序を定める正解データとなる。各学習モデルは、CGV v_i から抽出した特徴量ベクトル $x(v_i)$ を正規化した $\bar{x}(v_i) \in R^6$ から、編集度順序を定める編集度 $e(\bar{x}(v_i)) \in R$ を計算する関数を学習する。最終的に、出力として得られる編集度 $e(\bar{x}(v_i))$ に基づいて、初期検索結果の集合 V に含まれる CGV を降順にランキングし、編集度順序の推定結果とする。

4. 評価実験

提案手法の有効性を評価するために、実際に流通する CGV を用いて実施した評価実験について述べる。まず、予備実験として、評価者が主観的に判断した編集度順序を、提案手法がどの程度推定可能であるかを評価するため、これらのランキング間の順位相関を評価する。その後、実際の動画共有サイトでの利用を想定し、約 22,000 件規模の CGV データベースを用いた評価者実験を実施することによって、提案手法の効果を検証する。

4.1 データの取得

本稿では、動画共有サイト『ClipLife』に公開されている CGV を対象とする。『ClipLife』は、商用(放送)映像の不法アップロードに厳密に対処しており、消費者が作成した CGV の割合が高いという特徴を持つ。『ClipLife』から、CGV 27,564 ファイルを収集した。『ClipLife』では、動画がユニークに所属するカテゴリが設定されている。表 1 に、各カテゴリ名と、取得した CGV の内訳を示す。本実験では、収集した CGV 27,564 ファイルから、「その他」として分類した「携帯」、「テスト」、および、未所属の3つの例外的なカテゴリに含まれる 4,769 ファイルを除く、22,795 ファイルを、評価用映像として用いる。

動画共有サイトにおける映像検索のための編集度順序によるランキング法の提案と評価

表 1 取得した CGV のカテゴリと内訳

Table 1 The number of CGVs belongs to each category

カテゴリ名	取得 CGV 数
動物・ペット	6488
観光・旅行	3017
趣味	2513
スポーツ	1974
家族・結婚	1557
映像作品・アニメ	1333
乗り物	1327
テクノロジー	1306
音楽	733
教育・語学	567
ニュース	485
ヘルスケア	483
グルメ	345
ユーモア・ジョーク	306
ビジネス	256
ファッション・美容	105
その他	4769

4.2 予備実験

評価者が主観的に判断した編集度順序を、提案手法がどの程度推定可能であるかを評価するために、収集した CGV 中から、ランダムに抽出した 100 ファイルを用いて、順位相関を評価する。

4.2.1 ground truth の取得

100 ファイルの CGV それぞれに対して、編集度順序の正解を定める ground truth データを得る。評価者 5 名 (A, B, C, D, E) に、100 ファイルの CGV をランダムな順序で視聴させ、そのそれぞれに対して「編集されている程度」を 7 段階 (1:全く編集されていない~4:普通に編集されている~7:非常によく編集されている) の評価値で評価させた。

取得した ground truth データの性質を調査するため、評価者毎に、評価値による CGV のランキングを行った後、それぞれのランキング間の順位相関 (Spearman 順位相関) を計算した。結果を表 2 に示す。全ての評価者間で 0.8 を超える高い相関を持つことが示されており、その平均値は 0.895、標準偏差は 0.031 であった。このことから、編集度順序は、主観的差異による個人差の影響が小さく、客観性の高い指標であるといえ、本稿で用いる ground truth データが、一定の妥当性を有するものであると考えられる。

4.2.2 予備実験結果

本実験では、ある評価者 1 名の評価データを試験データ、残り 4 名の評価データを学習データとして用いることとし、起こりうる 5 通りの場合について評価を行った。学習データについて、RSVM の学習の際には、4 名の評価値の中央値を用いた順位を用い、LR, SVR の 2 つの回帰モデルに対しては、尺度の整合性

表 2 評価者間の編集度順序の順位相関

Table 2 Rank correlations of degree-of-edit rankings between two subjects

評価者 ID	B	C	D	E
A	0.877	0.910	0.930	0.906
B	-	0.889	0.852	0.837
C	-	-	0.922	0.918
D	-	-	-	0.908

表 3 提案手法による編集度順序の順位相関

Table 3 Rank correlations of degree-of-edit rankings

	LR	SVR	RSVM
A	0.753	0.740	0.827
B	0.768	0.795	0.870
C	0.743	0.761	0.831
D	0.740	0.719	0.794
E	0.729	0.720	0.785
Ave.	0.747	0.747	0.822

を取るため、便宜的に評価値を間隔尺度と見做して平均値を用いた。SVR, RSVM には、ガウスカネルを用いた。結果を、表 3 に示す。いずれのモデルを用いた場合においても、それぞれの評価者との順位相関は 0.7~0.8 程度と高いことがわかる。また、RSVM を用いた場合が最も高い順位相関となっており、その平均は 0.822 となった。

4.3 実験条件

実際の動画共有サイトでの利用を想定した効果を検証する。本実験では 4.1 において収集した「その他」を除く 16 種のカテゴリに所属する CGV 22,795 ファイルからなる CGV データベースを用いる。

本実験では、下記の 4 種類のランキング方式について比較を行う。

- 公開日時 (DATE): 最近公開されたものほど上位、
- 視聴回数 (VIEW): 視聴回数の多いものほど上位、
- 編集度順序 (EDIT): 推定された編集度の高いものほど上位、
- 視聴回数 + 編集度順序 (VIEW+EDIT): 視聴回数が多く、推定された編集度の高いものほど上位。

ここで、編集度順序については、4.2.1 で得た評価値の中央値を用いて学習した RSVM を用いた。VIEW+EDIT については、おのおのの CGV の視聴回数と推定された編集度を、22,795 ファイルに対して、それぞれ 0.0~1.0 となるよう線形正規化し、和をとることによってスコアを計算してランキングした。これら 4 種のランキング方式を適用して、上記 16 種のカテゴリ毎に CGV をランキングし、それぞれ上位 10 件の検索結果を表示したページ (4 方式 × 16 カテゴリ、計 64 ページ) を準備した。評価者は、予備実験

とは異なる 14 名 (男性 6 名, 女性 8 名) とし, 各評価者は以下の手順で評価作業を実施するものとした.

- (1) 16 種のカテゴリから, 最も興味のあるカテゴリを 1 つ選択する. また, $i = 1$ とする.
- (2) $i \leq 4$ ならば, (3) に進む. そうでなければ, (7) へ進む.
- (3) 選択したカテゴリについて, 4 種のランキング方式による検索結果のうち, i 番目を閲覧する.
- (4) 検索結果に含まれる CGV の視聴を行い, 各 CGV に対して [質問 1] [質問 2] に回答する.
- (5) 全ての CGV の視聴を終えた後, 検索結果に対する設問 [質問 3] に回答する.
- (6) $i \leftarrow i + 1$ とし, (2) に戻る.
- (7) 4 種のランキング方式全ての回答が終了した後, [質問 4] に回答する.

[質問 1] ~ [質問 4] の内容は, 以下の通りである.

[質問 1] (Interestingness) 検索結果に含まれる各 CGV を視聴し, 各 CGV に対する興味の度合いを 7 段階 (1: 低い ~ 7: 高い) で評価する.

[質問 2] (Quality) 検索結果に含まれる各 CGV を視聴し, 各 CGV の品質の高さを 7 段階 (1: 低い ~ 7: 高い) で評価する.

[質問 3] (Satisfaction) 検索結果に満足した度合いを 7 段階 (1: 低い ~ 7: 高い) で評価する.

[質問 4] (Result-Difference) 4 つの検索結果にどの程度差があったかを 7 段階 (1: 差がない ~ 7: 差がある) で評価する.

4.4 結果

評価者 14 名により, 興味のあるカテゴリとして選択された頻度は, 動物・ペット: 3, 観光・旅行: 1, スポーツ: 2, 映像作品・アニメ: 1, 乗り物: 1, 音楽: 1, グルメ: 2, ユーモア・ジョーク: 2, ファッション・美容: 1 であった. 各ランキング手法の性能に関する質問である [質問 1] ~ [質問 3] に対する回答結果のヒストグラムを図 3 に示す. また, 評価項目毎に, 各ランキング方式間の差が有意であるか否かを確認するために行った Steel-Dwass 検定の結果を表 4 に示す.

まず, Interestingness について, その中央値は, DATE: 4, VIEW: 4, EDIT: 4, VIEW+EDIT: 4 であり, 方式間の差は見られなかった. しかしながら, 図 3(a) をみると, 評価値の傾向として, DATE は 2~5, VIEW は 3~6, EDIT と VIEW+EDIT は 4~6 に高い頻度で分布しており, 差がみられた. 特に, 評価値 6, 7 の高い値では, DATE に比べ VIEW, EDIT, VIEW+EDIT の 3 種が相対的に高い頻度で分布している. 表 4(a) に示すように, DATE と EDIT および

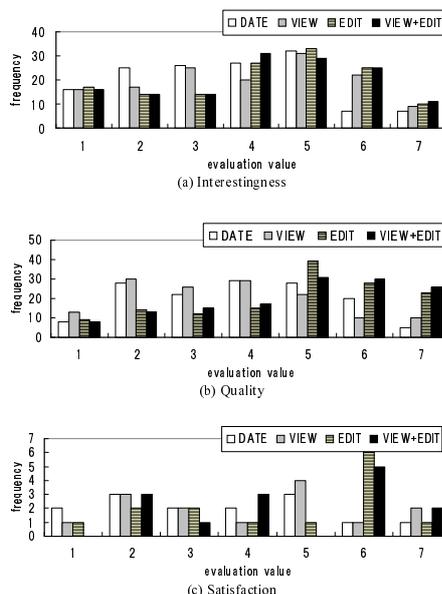


図 3 評価値の分布

Fig. 3 Distribution of evaluation results

VIEW+EDIT の間に有意水準 5% の有意差があった.

次に, Quality に着目すると, 中央値は, DATE: 4, VIEW: 4, EDIT: 5, VIEW+EDIT: 5 であり, EDIT, VIEW+EDIT の 2 種が相対的に高かった. 図 3(b) の分布も, EDIT, VIEW+EDIT の 2 種が, DATE, VIEW よりも高い評価値を得ている傾向にある. また, 表 4(b) によれば, EDIT と DATE および VIEW, EDIT+VIEW と DATE および VIEW の間に, それぞれ有意水準 1% の有意差が認められた.

最後に, Satisfaction についての結果を述べる. 中央値は, DATE: 3.5, VIEW: 4.5, EDIT: 5.5, VIEW+EDIT: 5 であった. VIEW, EDIT, VIEW+EDIT が, DATE に比べて高いことがわかる. しかしながら, 図 3(c) に示すヒストグラムをみると, 評価値 6 の点で EDIT, VIEW+EDIT の頻度が高いものの, 明らかな分布の違いは確認できず, 表 4(c) の結果からも, 有意傾向は認められなかった.

以上の結果から, EDIT によるランキングについて, 以下のことが確認できたといえる.

- (1) DATE よりも Interestingness の高い CGV を上位にランキングする,
- (2) DATE, VIEW よりも Quality の高い CGV をより上位にランキングする,
- (3) 結果に対する Satisfaction は, DATE, VIEW と同程度である.

動画共有サイトにおける映像検索のための編集度順序によるランキング法の提案と評価

表 4 Steel-Dwass 検定の結果 (t, p)

Table 4 Results of Steel-Dwass test (t, p)

(a) Interestingness			
	VIEW	EDIT	VIEW+EDIT
DATE	(1.89, .23)	(2.84, .02)	(2.86, .02)
VIEW	-	(0.89, .81)	(0.92, .79)
EDIT	-	-	(0.01, >.99)
(b) Quality			
	VIEW	EDIT	VIEW+EDIT
DATE	(1.37, .52)	(4.19, <.01)	(4.36, <.01)
VIEW	-	(5.13, <.01)	(5.25, <.01)
EDIT	-	-	(0.28, .99)
(c) Satisfaction			
	VIEW	EDIT	VIEW+EDIT
DATE	(0.70, .90)	(1.35, .53)	(1.49, .44)
VIEW	-	(0.70, .90)	(0.82, .85)
EDIT	-	-	(0.26, .99)

4.5 考 察

4.5.1 予備実験結果について

予備実験の結果, 特に RSVM を用いた場合, 評価者による編集度順序と高い順位相関が得られることがわかった。しかしながら, 評価者間同士の順位相関平均 0.895 に対して, 0.07 程度低い値となった。この原因として, 評価者が評価値を判断する際に, 判断材料とする画像・音響的な特徴と, 提案手法で用いた 6 種の特徴の間に乖離があることが考えられる。

評価者から回答を得た後, 何を基準に評価値を判断したかをたずねたところ, 概ね, カット点の数や音楽の有無など, 提案手法で利用した特徴量に対応する回答が多かった。一方で, 「カメラ動作が安定しているほど, 良質なシーンを選定している」という観点で, 編集の程度が高いと感じた」といった, 提案手法が考慮していないカメラ動作に対する言及があった。この言及は, どちらかといえば「編集の良し悪し」に関わるものであるが, カメラ動作自体は映像文法などの観点からも重要とされており²⁰⁾, これまでにも, CGV の高品質化を目的として, これを安定化する手法が提案されている²¹⁾。したがって, 主観との一致性を向上する, あるいは, さらなる Quality の改善を図るという観点では導入する価値があるものと考えられる。

また, 今回利用した特徴量のうち, IQ が実際の評価者が感じる画質とは乖離していた可能性も指摘された。この原因の一つとして, 動画共有サイトでは, 動画投稿時にビットレートや画像サイズが変更される場合があることが考えられる。本実験で対象とした『ClipLife』においても, このような変更により, 必ずしも画質を正確に反映しないものが含まれていた。今後は, より正確に, 信号レベルで画質を評価する手法

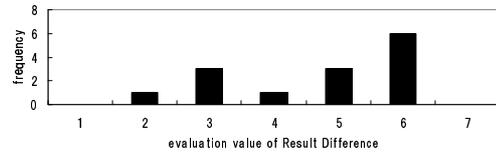


図 4 Result-Difference の分布
Fig. 4 Distribution of Result-Difference

の導入が必要であると考えられる。

4.5.2 ランキング結果の差異について

[質問 4] として, 方式の違いによるランキング結果の差異を質問した。この結果を, 図 4 に示す。評価値 5, 6 などの, 違いがあると回答した評価者の割合が相対的に高い頻度で分布していることがわかる。全ての評価者に, どのランキング方式が類似していると感じたかを尋ねたところ, EDIT と VIEW+EDIT が類似していると回答したものが最も多く 9 名, 続いて, VIEW と EDIT と VIEW+EDIT が 2 名, DATE と VIEW, DATE と EDIT, VIEW と VIEW+EDIT がそれぞれ 1 名ずつであった。

また, DATE, VIEW, EDIT の 3 種について, カテゴリ毎のランキング結果の順位相関を求めたところ, その平均値は, DATE と VIEW 間で-0.447, DATE と EDIT 間で 0.029, VIEW と EDIT 間で-0.045 となった。EDIT と他の 2 種のランキング結果との間に相関がないことがわかる。DATE と VIEW 間にみられる負相関は, 公開されてからある程度時間が経過しないと, 視聴回数が増加しないことによる。

以上のことから, EDIT は, DATE や VIEW とは主観的にも客観的にも異なるランキング結果を提示できるといえる。

4.5.3 評価項目間の相互相関について

実験では, 編集度順序が, Quality の高い CGV を上位に挙げやすいランキング方式であることを確認した。ここで, Interestingness, Quality, Satisfaction の 3 つの評価項目間の順位相関を計算したところ, Interestingness-Quality 間で 0.664, Interestingness-Satisfaction 間で 0.514, Quality-Satisfaction 間で 0.483 となった。Interestingness-Satisfaction 間と比較すれば, 相対的には弱いものの, Quality と Interestingness, Satisfaction の間にも少なからず相関があることが確認できる。この結果は, ユーザの興味や満足度に対して, Quality が寄与していることを示唆している。同時に, Quality を得やすい編集度順序の導入により, ユーザの興味や満足度を向上できることを示唆するものであると考えられる。

5. ま と め

本研究では、CGV の検索結果をランキングするための新たな指標として編集度順序を導入し、その推定手法を提案した。主観的に評価した編集度順序と、提案手法が推定した編集度順序との順位相関を評価し、提案手法が主観による編集度順序を精度よく推定できることを確認した。さらに、約 22,000 件規模の CGV データベースを用いた評価者実験によって、提案手法が現在の動画共有サイトにおけるランキング結果を改善できることを示唆する結果を得た。

今後の課題として、より有益なランキング結果の提示に向け、手法の精度改善、および、RF のように、ユーザからのレスポンスを利用してリランキングする手法などとの組み合わせを検討することが考えられる。また、実際の動画共有サイトへの実用化を目指すことも重要な課題である。

謝辞 東京大学大学院 講師 山崎俊彦氏、NTT 加藤洋一氏、谷口行信氏に多大なご協力をいただきました。また、実験にご協力頂いた皆様に感謝申し上げます。

参 考 文 献

- 1) Anderson, G.H.: *Video Editing and Post-Production: A Professional Guide 3rd Edition*, Knowledge Industry Publications, White Plains, NY (1993).
- 2) Zhang, H.-J., Kankanhalli, A. and Smoilar, S.W.: Automatic Partitioning of Full-Motion Video, *Multimedia Systems*, Vol. 7, No. 5, pp. 771-785 (1997).
- 3) 長坂晃朗, 田中護: カラービデオ映像における自動索引付け法と物体検索法, 情処論, Vol. 33, No. 4, pp. 543-550 (1992).
- 4) 谷口行信, 外村佳伸, 浜田洋: 映像ショット切換え検出法とその映像アクセスインタフェースへの応用, 信学論, Vol. J79-D-II, No. 4, pp. 538-546 (1996).
- 5) Taniguchi, Y., Akutsu, A. and Tonomura, Y.: Panorama Excerpts: Extracting and Packing Panoramas for Video Browsing, *Proc. ACM Multimedia*, pp. 427-436 (1997).
- 6) Minami, K., Akutsu, A., Hamada, H. and Tonomura, Y.: Video Handling with Music and Speech Detection, *IEEE Multimedia*, Vol. 5, No. 3, pp. 17-25 (1998).
- 7) 佐藤隆ほか: MPEG 符号化映像からの高速テロップ領域検出法, 信学論, Vol. J81-D-II, No. 8, pp. 1847-1855 (1998).
- 8) 桑野秀豪, 倉掛正治, 小高和己: 映像データ検索のためのテロップ文字抽出法, 信学技報 PRMU, Vol. 96, No. 385, pp. 39-46 (1996).
- 9) Dimitrova, N. and Abdel-Mottaleb, M.: Content-based Video Retrieval by Example Video Clip, *Proc. SPIE*, Vol. 3022, pp. 59-70 (1997).
- 10) Chen, L. and Chua, T.-S.: A Matching and Tiling Approach to Content-based Video Retrieval, *IEEE Int. Conf. Multimedia & Expo.*, pp. 417-420, (2000).
- 11) Rui, Y., Huang, S.H., Ortega, M. and Mehrotra, S.: Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval, *IEEE Trans. Circuits and Systems and Systems or Video Technology*, Vol. 8, No. 5, pp. 644-655 (1998).
- 12) Doulamis, A.D., Avrithis, Y.S., Doulamis, N.D. and Kollias, S.D.: Interactive Content-Based Retrieval in Video Databases Using Fuzzy Classification and Relevance Feedback, *Proc. of IEEE Inter. Conf. on Multimedia, Comp. & Syst.*, Vol. 2, pp. 954-958 (1999).
- 13) Yan, R., Hauptmann, A. and Jin, R.: Multimedia Search with Pseudo-Relevance Feedback, *LNCS*, Vol. 2728/2003, pp. 649-654 (2003).
- 14) Hsu, W.H., Kennedy, L.S. and Chang, S.-F.: Video Search Reranking via Information Bottleneck Principle, *Proc. ACM Multimedia*, pp. 35-44 (2006).
- 15) 宮地悟史, 浜田高宏, 松本修一: 人間の視覚特性を考慮した画像品質評価システムの開発, 信学論, Vol. J81-D-II, No. 6, pp. 1084-1094, (1998).
- 16) 林孝典ほか: マルチビットレート映像配信サービスにおける QoS 制御方式の検討, 信学技報 IE, Vol. 104, No. 313, pp. 55-60 (2004).
- 17) Wang, Z., Bovik, A.C., Sheikh, H.R. and Simoncelli, E.P.: Image Quality Assessment: From Error Measurement to Structural Similarity, *IEEE T. Image Processing*, Vol. 13, No. 4, pp. 600-612, (2004).
- 18) Talkin, D.: A Robust Algorithm for Pitch Tracking, *Speech Coding and Synthesis*, (1995).
- 19) Joachims, T.: Optimizing Search Engines using Clickthrough Data, *Proc. ACM SIGKDD*, pp. 133-142 (2002).
- 20) Arijon, D.: *Grammar of the Film language*, Silman-James Press (1991).
- 21) Gleicher, M.L. and Liu, F.: Re-Cinematography: Improving the Camera Dynamics of Casual Video, *Proc. ACM Multimedia*, pp. 27-36 (2007).