

ユーザ参加行動に基づく蓄積映像コンテンツ 編集方式

ASO, Tasuku / 麻生, 祐

(出版者 / Publisher)

法政大学大学院情報科学研究科

(雑誌名 / Journal or Publication Title)

法政大学大学院紀要. 情報科学研究科編 / 法政大学大学院紀要. 情報科学研究科編

(巻 / Volume)

8

(開始ページ / Start Page)

69

(終了ページ / End Page)

74

(発行年 / Year)

2013-03

(URL)

<https://doi.org/10.15002/00009571>

ユーザ参加行動に基づく蓄積映像コンテンツ編集方式

A method for editing stored video contents based on consumer generated media

麻生 祐

Tasuku Aso

法政大学大学院情報科学研究科情報科学専攻

E-mail: 11t0001@cis.k.hosei.ac.jp

Abstract

Video on demand service, like YouTube, has attracted attention as a new entertainment system. Although the video on demand gives consumers opportunities to watch the video from a huge number of stored video contents, the size of contents makes it difficult for the user to select. Some systems provide key word search functions to the users, but the quality of the search results is not so high. There is another trend in Web, where user comments on the video contents are posted while watching video contents by using SNS such as Twitter or online message board. Most of these comments are related to the video contents. This paper proposes a method for editing stored video contents to generate digest and facilitate video contents retrieval. This paper generates attention graph by using user participation behavior, and analyses it to identify the scenes which should be included in the digest. After the analysis, the proposed method can set importance level for each scene, and the method generates digest from some scenes which have higher level importance. The generated digest was evaluated in terms of recall performance of the scene relative to the length ratio of the digest by experiments.

1 はじめに

近年、大容量記憶装置が小型かつ安価となり、大容量の映像コンテンツを録画して視聴する機会が増大している。また、YouTube¹やニコニコ動画²に代表されるように、Web で映像を配信するサイトも数多く存在している。これらのことは、ユーザが映像を好きな時間に視聴するオンデマンド型視聴を手助けしている。映像コンテンツの増加は、各ユーザの好みにあった視聴を提供出来る一方で、映像コンテンツの検索を困難にし、様々な検索方法が提案されている。テレビ放送では、事前に提供される番組の内容や出演者の情報といったメタデータを基に検索を行うことが出来る。一方 Web の動画配信では、キーワードを入力する検索や、事前に投稿者や視聴済みユーザが映像コンテンツに対して付与したタグを基に検索することが出来る。これらの検索は提供されたデータを基に行うため、要求とは異なる結果が得られることが度々あった。

一方で Twitter³のハッシュタグやインターネット掲示板

を利用し、映像コンテンツを視聴しつつそのコンテンツに対して内容や感想を発言する機会が増大している。特に Twitter における発言傾向は番組の盛り上がりと連動する傾向が見受けられる。ツイート・パー・セカンド(TPS)と呼ばれる1秒間にどのくらいのツイートがされたかという指標がある。日本で放送された映画のクライマックスシーンで TPS の当時の世界記録が更新された。これは視聴ユーザが最も盛り上がるシーンで他のユーザと時間を共有するために行われたためと推察できる。すなわち、盛り上がりや注目の高いシーンで、ユーザが視聴シーンに対して何かしらのリアクションを起こしていると考えられる。

本研究では、数多く存在している映像コンテンツの選択や検索を支援するための動的ダイジェスト自動生成を提案する。特にユーザの行動を利用することで、時間経過と共に変化するダイジェストを提案する。また簡潔に内容を理解することで、各ユーザの好みに適した映像コンテンツの選択が可能になる。本提案ではダイジェスト作成のために、ユーザが映像コンテンツに対して発言したデータを用いて、盛り上がっている区間やユーザから注目を集めている区間を同定する。これらの区間を繋ぎあわせ、ダイジェストを作成することを目標にする。

2 映像視聴支援

2.1 蓄積映像コンテンツ

蓄積映像コンテンツとは、大容量記憶装置に録画された映像コンテンツである。蓄積映像コンテンツはテレビのような放送型では無いため、視聴ユーザそれぞれの好きな時間帯に視聴できることが大きなメリットとなる。蓄積映像コンテンツの例として、テレビ番組の録画やオンデマンド型動画配信サイトがある。従来ではテレビ番組の録画は事前に録画予約を行ったものだけを録画していたが、近年ではテレビ放送の全ての番組を数日間録画できるホームサーバ型の録画機器が登場した。このホームサーバ型録画機器の登場により、出演者の情報や活字による番組のあらすじを基に行われていた事前録画の必要性がなくなった。一方で、録画番組は従来の選別方法に加えてすでに視聴したユーザのレビューを使用した選別が可能となった。オンデマンド型動画配信サイトではユーザのレビューに基づく人気映像コンテンツの紹介や各ユーザ自身の視聴履歴に基づいた映像コンテンツのレコメンデーションが行われている。

2.2 関連研究

山本[1]らは、フォークソノミーと呼ばれるユーザが自由にタグ付けを行う分類手法を用いて、映像コンテンツに含まれる映像シーンに対してのタグアノテーション手法を提案した。タグの付与にはユーザ毎に自動生成されるタグクラウドを用いる他に自由に記述することも可能であり、それらのタグを基に検索を行うことで映像シーン単位での映像検索を可能にした。

また、Chung[2]らはフォークソノミーを用いて、ユーザが映像コンテンツに対して付与するブックマークを利用したビデオ要約自動生成を提案した。数多く付与されたブックマークを分析することにより、精度の高いビデオ要約生成を提案した。この手法の評価は自動的にビデオを要約する既存手法よりも優れていることが示されており、フォークソノミーの概念がビデオ要約に有効であることを示した。

これらの手法では、共通して、ユーザが行う行動に注目している。従来、要約作業は専門家により行われた主観的なものであったが、フォークソノミーを用いることでユーザが求めるものに近づけると考える。本研究では以前の我々の研究[3]でも注目したユーザ参加行動に焦点を当て、ダイジェスト生成に有効な手法を検討する。

3 ダイジェストの分析

3.1 ダイジェストの分類

ダイジェストとは、ある映像コンテンツの内容を要約し、簡潔に映像コンテンツの内容を理解出来るようにした短縮映像である。既存の映像ダイジェストは2種類に分類される。1つはコマーシャル型のダイジェストである。コマーシャル型とは、映像コンテンツに含まれる印象に残りやすい刺激的なシーンを主に繋ぎ合わせたダイジェストである。映像コンテンツの内容を理解するには及ばないが、映像コンテンツを魅力的に感じられるように設定されている。コマーシャル型の主な目的は番組の広告や宣伝であり、テレビ局の流す番組や映画の予告編が例として挙げられる。もう一方はニュース型である。ニュース型とは、映像コンテンツに含まれる中で内容を理解するために、あるシーンの魅力的な部分を長めに切り出し、繋ぎ合わせるダイジェストである。ニュース型は主に放送や配信が終わった後に内容を簡潔に伝える目的で利用される。例えば野球中継の結果を伝えるスポーツニュースが考えられる。

一方、ダイジェストと類似するものとして総集編がある。総集編とは主要な部分を編集し、内容をまとめたものである。総集編はこれまでの内容を振り返る上では最適なものである一方、時間が長くなる・説明をするためのナレーションを録音する必要がある等の欠点を持つ。

3.2 ダイジェスト調査

本節では、3.1節で述べたダイジェストの分類に対して実際にどのようなダイジェストが生成されているかを調査する。具体的には、コマーシャル内での映像が切り替わるまでの時間を調査し、1シーンあたりにどの程度時間がかけられているかを検討する。

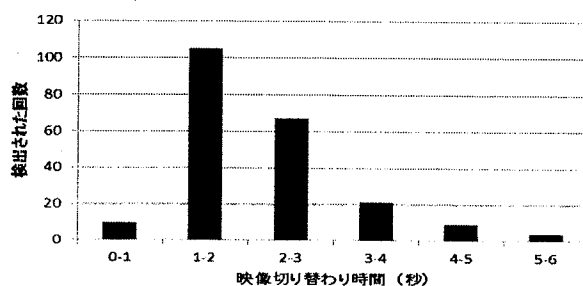


図1 映像切り替わり時間のヒストグラム

コマーシャル型の調査には実際にテレビ局で放送されたテレビ番組宣伝用のコマーシャルを用いる。分析するコマーシャルは、24番組31放送日に対して放送されたコマーシャルを基に行った。調査対象としたコマーシャル全てに含まれていたシーン数は216であった。図1に調査結果をヒストグラムで示す。調査結果では、映像切り替わりの平均時間が2.18秒、標準偏差が0.99秒となった。この結果から、映像切り替わり時間にはおよそ1-3秒程度の値が一般的には用いられていることが分かる。

4 ダイジェスト自動生成手法の提案

4.1 ユーザ参加行動の応用

本研究ではユーザの参加行動を応用したダイジェストの自動生成手法を提案する。ユーザの参加行動とは、映像コンテンツを視聴しながらユーザが残す行動と定義する。ユーザ参加行動の例として、映像を視聴しながらその内容に関連した発言を残すことが考えられる。この発言にも様々な種類が考えられるが、大きく3種類に分類して考える。1つは内容に関連する発言である。これは映像コンテンツに関連した感想やリアクションを書き込むことで、同じ映像コンテンツを視聴しているユーザと感覚を共有することが目的である。2つめは発言に触発された発言である。他のユーザの発言にリアクションすることで、コミュニケーションを図っている。3つめは映像コンテンツに全く関連しない発言である。記号で作るアスキーアートなどで視聴ユーザを盛り上げることが目的である。発言として1つめと2つめは相当数あると考えられるが、3つめは非常に少ないと考えられる。すなわち、ユーザが残す発言は映像コンテンツに関連するものであると仮定する。

映像コンテンツに関連する発言を集めヒストグラムとしたものを注目度と定義する。注目度は時刻と参加行動数で構成される。ある時刻 t における参加行動数を $act(t)$ で表す。評価に用いる一例として、図2に注目度をグラフ化した例を示す。このデータは再生時間1404秒、総参加行動数が148184回であった。図2において、横軸は映像コンテンツの経過時間 t を表し、縦軸は t における参加行動数を示している。図2より、注目度は映像コンテンツの経過時間に応じて大きく変化していることが分かる。これは映像コンテンツに含まれる場面(以降、シーン)に応じて視聴ユーザがコメントを残すためである。すなわち、700秒付近でこの映像コンテンツの最も盛り上がったシーンが存在したと予想出来る。

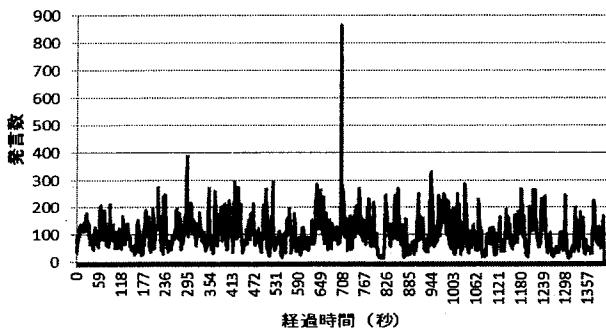


図2 注目度の一例

4.2 提案手法概要

本研究ではダイジェストの自動生成を提案する。ダイジェストの自動生成では 4.1 節で述べたユーザ参加行動を利用する。ユーザ参加行動は時間経過で変化する特性があるため、本研究で提案するダイジェストも同様の性質を持つ。すなわち、時間が経過するとダイジェストの内容も変化する動的ダイジェストを提案する。

ある映像コンテンツに対するダイジェストは各シーンの集合である。映像コンテンツを Video, ダイジェストを Digest, 映像コンテンツに含まれるシーンを Scene とすると、以下の集合式で表せる。

$$Digest = \{scene \mid scene \in Video\} \quad (1)$$

本研究ではユーザの参加行動を応用したダイジェストの自動生成としてピーク検出法とゴールデンドロス法の 2 つの手法を提案する。以降、それぞれの手法について述べる。

4.3 ピーク検出法

注目度において発言が突出した箇所をピークと定義する。ピークはある時刻 t に対して、 $act(t-1) < act(t)$ かつ $act(t) > act(t+1)$ を満たす点とする。ピーク検出法はこのピークをつなぎあわせダイジェストとする手法である。しかしながら、注目度には数多くのピークが存在する。そこで参加行動数を用いてピークの選別を行う。すなわち、参加行動数が多いものほど注目度の高い時刻とみなし、希望ダイジェスト時間に近づくようにピークを選別する。ある時刻 t_1 から t_2 までの区間を (t_1, t_2) と表現すると、ピーク検出法におけるダイジェストは以下のように定義出来る。

$$Digest = \{ (t_1, t_2) \mid (t_1, t_2) = period(t) \\ \wedge act(t-1) < act(t) \\ \wedge act(t) > act(t+1) \} \quad (2)$$

4.4 ゴールデンドロス法

ダイジェストを作成する際、注目度が高いシーンだけを切り出すと、番組のクライマックスの前後だけが集中的に切りだされてしまい、ダイジェストとしての意味を成さない。そこで我々は相対的に注目度が高い部分区間を同定し、ダイジェストにしたいと考えた。逆に、全体には参加行動数が少なく盛り上がり欠けるシーン群の中でも一部区間だけ注目度が高まっているシーンは、映像コンテンツの流れの中で重要な意味の区間とみなし、

ダイジェストにしたいと考えた。そこでゴールデンドロス法ではダイジェストに含めるシーンを「前後のシーンと比較して際立って注目度の高い 5-10 程度のシーン」と定義し、それぞれのシーンをつなぎあわせることで一つのダイジェストとする。

相対的な注目度の高まりを同定するために株取引のテクニカル指標として用いられるゴールデンドロスと呼ばれる手法を使用する。ゴールデンドロス検出のために、移動平均を用いる。移動平均とは、時系列データに対して平滑化を行う手法である。本研究では移動平均の中でも単純移動平均を用いる。単純移動平均はある地点からの一定範囲を平均し、それをある地点での値とする手法である。我々はダイジェストにはある地点における前後の注目度が影響すると考えた。そこで今回用いる移動平均における一定範囲では、前後に同様の長さを持つものと定める。

ゴールデンドロスを検出するために、長期移動平均と短期移動平均を用いる。長期移動平均とは時系列データにおける長期的なノイズを除去する。長期移動平均では映像コンテンツにおける長期的な傾向が分かる。短期移動平均は短期的な傾向が分かる。短期移動平均は、ノイズがデータに含まれる微小変動のノイズを除去するために用いる。ゴールデンドロスとは、短期移動平均の値が長期移動平均の値を上回るポイントのことである。ゴールデンドロスは今後値の上昇が見込めるポイントのため、株取引では買いシグナルとなる。本研究において、ゴールデンドロスは蓄積映像コンテンツにおける注目度に対して今後参加行動数の上昇が見込めるポイントとして認識する。一方でゴールデンドロスとペアで発生するものとしてデッドクロスがある。デッドクロスとは短期移動平均の値が長期移動平均の値を下回るポイントのことである。デッドクロスは今後参加行動数の下降が見込まれるポイントのため、株取引では売りシグナルとなる。本研究において、デッドクロスは今後の注目度下降が予想されるポイントとして認識する。ある時刻 t における短期移動平均の値を $MAS(t)$ 、同様に長期移動平均の値を $MAL(t)$ 、ゴールデンドロスとなる時刻を t_{GC} 、デッドクロスを t_{DC} とすると、満たすべき条件を以下のように定義する。

$$\begin{aligned} MAS(t_{GC} - 1) < MAL(t_{GC} - 1) \\ \cap MAS(t_{GC}) \geq MAL(t_{GC}) \end{aligned} \quad (3)$$

$$\begin{aligned} MAS(t_{DC} - 1) > MAL(t_{DC} - 1) \\ \cap MAS(t_{DC}) \leq MAL(t_{DC}) \end{aligned} \quad (4)$$

ゴールデンドロス法を用いてダイジェスト生成を行う場合、多くのゴールデンドロスが検出されるためダイジェストが冗長なものになりがちである。このため、短期移動平均が長期移動平均を多少超えている場合は、そのシーンをダイジェストに含めないことが望ましい。そこで短期移動平均と長期移動平均がなす積分値を求め、それを指標としてゴールデンドロスを選定する方法を検討する。ダイジェストに含むべきシーンは積分値の高い上位数点であると考えられる。この積分値をシーンの重要度と定義し、検出されたシーンの選別を行う。ある地点 t_1 からある地点 t_2 までの重要度を $w(t_1, t_2)$ として表す

とき、ゴールデングロス法における重要度は以下のように定義する。

$$w(t_{GC}, t_{DC}) = \sum_{t=t_{GC}}^{t_{DC}} (MAS(t) - MAL(t)) \quad (5)$$

以上から、ゴールデングロス法におけるダイジェストを以下のように定義する。

$$Digest = \{ (t_{GC}, t_{DC}) \mid w(t_{GC}, t_{DC}) > threshold \} \quad (6)$$

5 実験

5.1 ユーザ参加行動遅延調査

ユーザ参加行動は映像コンテンツを視聴した後に行うため、ユーザ参加行動が映像コンテンツに関連付けられる時間は遅延していると考えられる。そこで本節ではユーザ参加行動はどの程度の遅延が生じているかを実験する。実験は以下の手順で行う。

- ・手順1 映像コンテンツを視聴
- ・手順2 注目度の高い時刻を選択
- ・手順3 注目度の上昇時刻（正解時刻）を同定
- ・手順4 正解時刻に関連する参加行動の時刻を同定

今回の実験では、映像コンテンツに関連した発言をユーザ参加行動として用いている。厳密には発言がどの時刻に対して行われたものかを同定することは、発言したユーザだけが分かることである。そこで今回は映像コンテンツの注目度が高い部分でも特に映像のセリフや動きに関連するものを選択し、比較的関連が強いものに対して調査を行うこととした。実験対象として6つの映像コンテンツを用意し、6つの映像コンテンツから合計100件のユーザ参加行動について調査した。調査の結果を図3に示す。調査の結果、100件のユーザ参加行動の平均遅延時間が3.84秒、標準偏差が2.13秒となった。この結果より、約4秒の遅延時間を考慮することが望ましいと考える。

5.2 映像ダイジェスト評価

5.2.1 評価方法

【実験手法1】

実験として、交差検定の考えを参考にした実験を行う。この実験手法を実験手法1とする。実験手順を以下に示す。

- ・手順1 注目度からn回の参加行動を残したユーザを抽出
- ・手順2 手順1で抽出したユーザを2分割(それぞれをa, bとする)

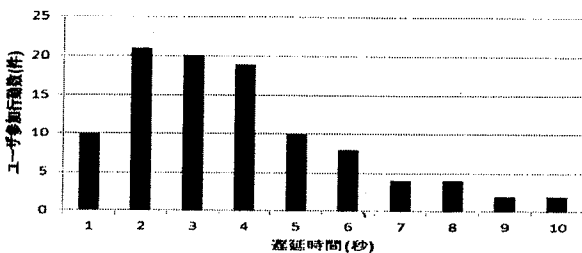


図3 参加行動遅延時間ヒストグラム

- ・手順3 手順1で抽出したユーザ以外を2分割(それぞれをc, dとする)
- ・手順4 aとc, bとdを足しあわせて注目度を生成する(それぞれをa-c, b-dとする)
- ・手順5 注目度a-cに対してbを正解ファイルとし、注目度b-dに対してdを正解ファイルとして評価を行う

今回の実験では、手順1で示したn回を5回の参加行動数とした。また、ファイルを2分割したことにより、実験結果として2つの結果が観測できる。このそれぞれの実験を指す場合に便宜的に【実験手法1-1】、【実験手法1-2】と記す。

2つの実験の評価の基準として、再現率を参考にした指標を提案する。通常の再現率は正解ファイルをどの程度再現しているかを示す。正解ファイルの正解時間データ集合を $Cor = \{t_1, \dots, t_n\}$ として表し、集合の長さを $size()$ とする。このとき正解時間データ集合 Cor に対し、生成されたダイジェスト $Digest$ に含まれる区間の集合 $Cor(Digest)$ を以下の集合式で定義する。

$$Cor(Digest) = \{ t \mid t_1 \leq t \leq t_2 \text{ for } \forall (t_1, t_2) \in Digest \} \quad (7)$$

生成されたダイジェストの再現率 $recall(Digest)$ は、ある映像コンテンツの開始時間を $Start$ 、終了時間を End とすると以下のように表せる。(7)式を用いて以下のように表せる。

$$recall(Digest) = size(Cor(Digest)) / size(Cor) \quad (8)$$

一方でダイジェストは映像コンテンツの一部を切り出すため、(8)式について分子が小さくなり分母が大きくなる傾向にある。そこでダイジェストの時間割合を考慮させる。映像の長さを取得する関数を $length()$ とするとき、ある映像コンテンツにおけるダイジェストの時間割合 $ratio(Digest)$ は、以下の式で表される。

$$ratio(Digest) = length(Digest) / length(Video) \quad (9)$$

ある映像コンテンツに対して、ダイジェストの時間割合を考慮させることで相対的な参加行動の密度を求めることが出来る。この密度は、ある映像コンテンツに対してダイジェストの占める時間帯における再現率を意味する。この密度を相対参加行動密度と定義する。相対参加行動密度 ρ は(8), (9)式を用いて以下のように定義出来る。

$$\rho = recall(Digest) / ratio(Digest) \quad (10)$$

ある映像コンテンツにおける時刻 t に対して、常に相対参加行動密度を最大値で取り続ける値を理論限界値と定義する。ユーザ参加行動を参加行動数の降順に並び替えたときの時刻を $t_0^a, t_1^a, \dots, t_n^a$ とすると、ある時刻 t に対する長さ1のダイジェストの理論限界値は以下のように定義出来る。なお、理論限界値は便宜的に理論値と表記する。

$$theoretical_1 = \frac{\sum_{i=0}^{l-1} act(t_i^a) / \sum_{t=0}^{End} act(t)}{l / length(Video)} \quad (11)$$

5.2.2 ダイジェスト評価

提案手法で生成されたダイジェストについて評価する。

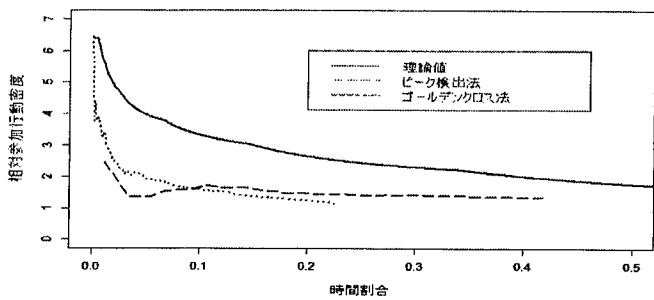


図4 映像Aにおける実験手法1-1の結果

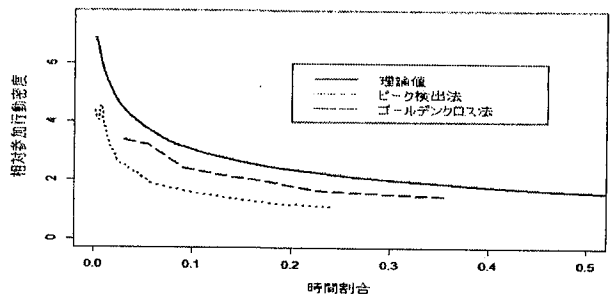


図7 映像Bにおける実験手法1-2の結果

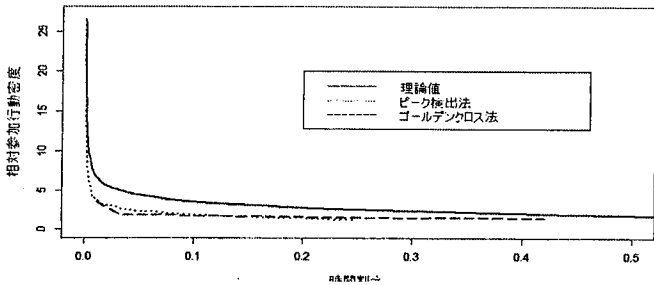


図5 映像Aにおける実験手法1-2の結果

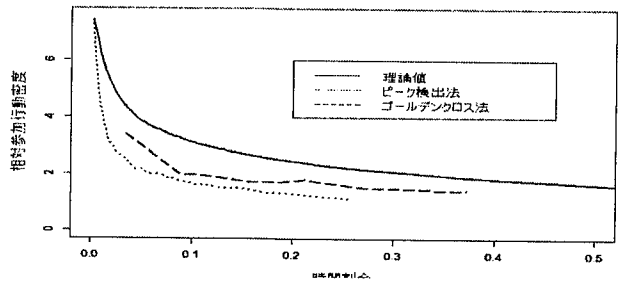


図8 映像Cにおける実験手法1-1の結果

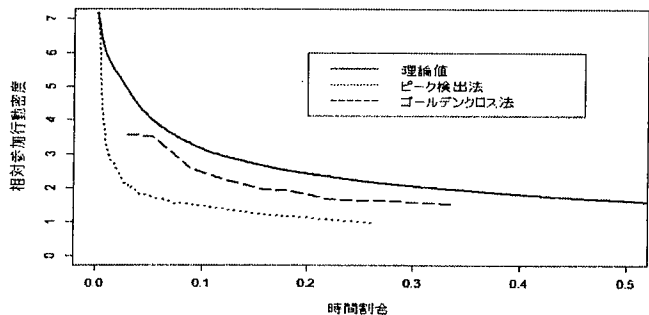


図6 映像Bにおける実験手法1-1の結果

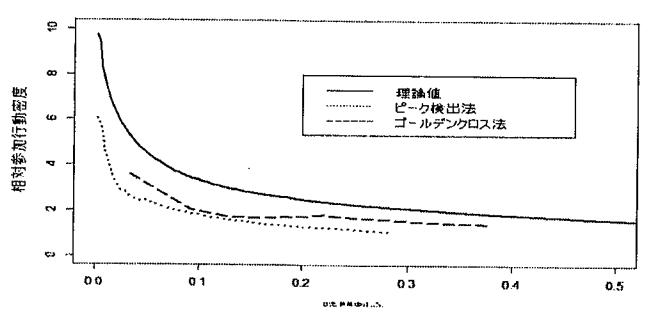


図9 映像Cにおける実験手法1-2の結果

提案したピーク検出法とゴールデングロス法について、その理論値と共に示したグラフを示す。

映像Aについて実験手法1-1で評価した結果を図4、実験手法1-2で評価した結果を図5に示す。図4から、シーン数が増加すると共にゴールデングロス法がピーク検出法よりも理論値に近い値を示すことが分かる。一方図5ではピーク検出法がゴールデングロス法よりも良い値を示している。

映像Bについて実験手法1-1で評価した結果を図6、実験手法1-2で評価した結果を図7に示す。図6では安定的にゴールデングロス法が理論値に近い値を取っている。図7について、ピーク検出法で検出したピーク上位数点の相対参加行動密度が、図6に示した実験手法1-1のピーク検出法で検出したピーク上位数点の相対参加行動密度の半分程度の値しか計測されなかった。これは最も参加行動数が大きかった時刻 t 点ではピークの条件を満たしていなかったためこのような結果になったと考えられる。また、実験結果を検証したところ、最もピークの点が高かった点の前後1秒地点に参加行動が集中していた。よって正解ファイルを分割して生成したときに偶発的に起こったものと考えられる。

映像Cについて実験手法1-1で評価した結果を図8、実

験手法1-2で評価した結果を図9に示す。図8、図9共にゴールデングロス法がピーク検出法よりも理論値に近い値を示すことが分かった。

図4-図9について、全体的にピーク検出法よりもゴールデングロス法が理論値に近い値を示す傾向にあり、ダイジェスト生成には向いていると考えられる。ピーク検出法はその性質上、最大でもピークとして検出される時間が映像コンテンツ再生時間の半分となる。また、連続して注目度が上昇を続ける区間において、注目度が高いものでもピークとして検出されない。これらの要因が作用してピーク検出法が理論値の半分程度の値になると考えられる。

5.2.3 ダイジェストシーンのばらつきの評価

本節では、5.2.2節で有効性を示したゴールデングロス法で検出したシーン群が、映像コンテンツ全体に対してどのような割合で検出されているかを評価する。ダイジェストとは内容が簡潔に理解できるものであり、そのためには映像全体からバランスよく検出されることが望ましいためである。そこで映像コンテンツを均等に3分割し、便宜的に映像コンテンツの前半・中盤・後半を作成する。これを映像コンテンツの時間群と定義する。そし

て検出したシーンが時間軍のどこに属するかを決定する。評価に使用する 3 つの映像コンテンツのデータを表 1 に示す。映像 A・映像 B はそれぞれバラエティの映像である。映像 C は野球を題材にしたスポーツ映像である。

今回の実験では、ゴールデンクロス法で使用する移動平均を 5.1 節の実験結果を踏まえて変更する。すなわち、ユーザ参加行動では平均的に約 4 秒の遅延が発生するため、4.3 節で述べた移動平均の定義を「移動平均における一定範囲では、前後に同様の長さを持つものから 4 秒の幅を後退させたものと定める」と変更する。

移動平均のパラメータとして短期移動平均を 10 秒、長期移動平均を 60 秒とした場合、ゴールデンクロスで検出したシーンの個数を表 2 に示す。また、全ての検出されたシーンに対して、時間群毎に検出されたシーンの個数を表 3 に示す。表 3 の結果より、評価した 3 つの映像コンテンツに対して全ての映像コンテンツにおいて時間群からバランス良くシーンが検出可能なことを確認出来た。

ダイジェストは短縮映像であるため、実際にはゴールデンクロス法により検出されたシーン全てを含むことはダイジェストが冗長になりがちとなるため、ダイジェストに含めることは望ましくない。そこでシーンの重要度上位 10 点に対して同様の調査を行う。重要度上位 10 点に対して、時間群ごとに検出されたシーンの個数を表 4 に示す。表 4 の結果より、上位 10 点でもおよそバランス良く検出出来たと考える。映像 B においては、前半部分に単調なシーンが連続したため、ユーザの参加行動が分散し、結果として検出個数が若干低めになったものと考えられる。

表 1 実験に用いる映像コンテンツデータ

| | 映像 A | 映像 B | 映像 C |
|--------|---------|--------|--------|
| 時間 (秒) | 1,404 | 585 | 408 |
| 参加行動数 | 148,185 | 99,801 | 34,721 |

表 2 検出されたシーンの個数

| | 映像 A | 映像 B | 映像 C |
|------|------|------|------|
| 検出個数 | 47 | 16 | 12 |

表 3 検出されたシーンの分布

| | 映像 A | 映像 B | 映像 C |
|----|------|------|------|
| 前半 | 16 | 6 | 5 |
| 中盤 | 15 | 5 | 3 |
| 後半 | 16 | 5 | 4 |

表 4 上位 10 点における検出されたシーンの分布

| | 映像 A | 映像 B | 映像 C |
|----|------|------|------|
| 前半 | 3 | 2 | 4 |
| 中盤 | 3 | 4 | 3 |
| 後半 | 4 | 4 | 3 |

6 考察

本研究では、数多く存在している映像コンテンツの選択や検索を支援するために、ユーザ参加行動を用いた動的ダイジェスト自動生成を提案した。

ダイジェストの自動生成には様々なものがすでに存在している。特に音声や画像の視聴覚特徴に依存して作成するダイジェスト自動生成が知られている。一方でこれらのダイジェストでは、ジャンルが制限される等の問題点が存在した。これら既存手法と本提案手法の大きな違いはユーザ参加行動を用いている点である。ユーザ参加行動を用いる利点として、様々なジャンルの映像コンテンツに対応出来ることとダイジェストが時間経過と共に変化する事が挙げられる。それぞれの映像コンテンツを視聴したユーザのデータを用いるため、様々なジャンルでも同様の手法で適用することが可能である。また、ユーザ参加行動には時間経過と共にその数が増加する特性が存在するため、特定の時期によるダイジェストが作成可能である。すなわち、ある時期には盛り上がりつつあるポイントでも時間経過とともに注目度が下がり、違う時期では盛り上がり小さいことが考えられる。本提案手法ではこの差を吸収し、時期ごとによって内容が変化する動的ダイジェストの特性を持たせることに成功した。

本提案手法には様々な応用例が考えられる。ユーザ毎に参加行動情報を取得すれば、参加行動が似ているユーザの情報から各ユーザにパーソナライズされたダイジェストが作成出来る。

7 おわりに

本研究ではユーザ参加行動を利用し、蓄積された映像コンテンツに対しての動的ダイジェスト生成方法としてピーク検出法とゴールデンクロス法の 2 つの生成方法を提案した。生成されたダイジェストに対し、交差検定の考えを利用した実験を行い、ゴールデンクロス法の有効性を示した。また実験に用いる指標として相対参加行動密度を提案し、ダイジェストに含むべきシーンの再現性を示した。ユーザそれぞれの特性に注目し、似ているユーザからダイジェストを生成するユーザ依存型で視聴しユーザ毎に異なるダイジェストが生成されるパーソナルダイジェストが今後の課題として考えられる。

文献

- [1]山本 大介, 増田 智樹, 大平 茂輝, 長尾 確, “タグクラウド共有に基づく協調的アノテーション”, 人工知能学会論文誌, 25 巻 2 号 B (2010).
- [2]Min Gyo Chung, Taehyung (George) Wang, Phillip C.-Y. Sheu, “Video summarization based on collaborative temporal tags”, Emerald Online Information Review, Vol.35, No.4 (2011).
- [3] 麻生祐, 藤田悟, “ユーザ参加行動に基づく蓄積映像コンテンツ編集方式”, FIT2012 第 11 回情報科学技術フォーラム, k-020(2012).

ⁱ YouTube. <http://www.youtube.com/>

ⁱⁱ ニコニコ動画. <http://www.nicovideo.jp/>

ⁱⁱⁱ Twitter. <https://twitter.com/>