

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[†] 藤田 悟[‡]
Tasuku Aso Satoru Fujita

1. はじめに

大容量記憶装置が小型かつ安価となることで、大きなデータ容量が必要な映像コンテンツを録画し視聴する機会が増大している。また、Web で映像を配信するサイトも数多く存在し、映像コンテンツの選択が煩雑化している。

一方で SNS 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。

本研究では煩雑化している映像コンテンツの選択を支援するために、内容を簡潔に理解できるダイジェストの生成を提案する。映像コンテンツ視聴時に多数のユーザが行なった発言などの行動を蓄積データとして、映像コンテンツに対する時系列の注目度を生成する。注目度のパターンを抽出し、分析を行うことでダイジェストの生成を行う。蓄積データは時間経過と共に変化していくため、生成される注目度も常に変動する。そして、注目度に動的に適応したダイジェスト生成を行うことができる。

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、映像コンテンツに対する多種多様なタグアノテーション手法を提案し、映像シーン検索を容易にした。

また、Chung[2]らはフォークソノミーを用いてビデオ要約を作成する手法を提案した。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。本研究でもユーザの行動に注目し、ダイジェスト生成に有効な手法を検討する。

3. 提案手法

3.1 提案手法の概要

ダイジェストを作成する際、注目度が高いシーンのみを切り出すと、盛り上がり続けたシーンにおいて全てが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。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、全体にはコメント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シーンとしてダイジェストにしたいと考えた。そこで本研究では、ダイジェストシーンを「前後のシーンと比較して際立って注目度の高い 5-10 秒程度のシーン」と定義し、相対的な注目度の高まりを求めるために、長期的な注目度の傾向を示す長期移動平均を求め、これと発言数の差分で短中期的な注目度の高まりを求めることにした。

一方、観測される素データは 1 秒毎の注目度データであるが、そのままのデータでは注目度の上下動が激しく、適切な区間を切り出すことが出来ない。そこで素データについても中期移動平均を求め、データの平滑化を試みる。

以上、長期移動平均と中期移動平均を用い、中期移動平均が長期移動平均を上回る区間をダイジェスト候補として抽出する方法を提案する。この考え方は、株式データのゴールデンクロスとデッドクロスの関係と相似である。

3.2 時系列データ分析

ピークの検出には移動平均を用い、株式の分析方法であるゴールデンクロスとデッドクロスを利用する。株式において移動平均を行う場合は前日から一定区間までのデータで平均をとるが、本研究では前後の一定区間までのデータで平均をとる。これはダイジェストに入れるべき点は前後の盛り上がりがあるシーンと考えたためである。この一定区間は映像コンテンツの長さによって変更する必要があると考える。一方、本研究においてゴールデンクロスと中期移動平均の値が長期移動平均の値を上回る点、デッドクロスと中期移動平均の値が長期移動平均の値を下回る点と定義する。ゴールデンクロスからデッドクロスまでがシーンの盛り上がりが高い点と考えられる、これをダイジェストに入れるべきピークとして検出する。このピークには重要度をつける。あるシーンの重要度 w_s は、中期移動平均におけるヒストグラムの t 時間における発言数を HM_t 、長期移動平均におけるヒストグラムの t 時間における発言数を HL_t 、ゴールデンクロスの時刻を GC 、デッドクロスの時刻を DC としたとき、以下の式のように定義する。

$$w_s = \sum_{t=GC}^{DC} (HM_t - HL_t) \quad (1)$$

(1)式は、ゴールデンクロスからデッドクロスまでの中期移動平均と長期移動平均がなす積分値を意味する。

3.3 ピーク選定

ゴールデンクロスを用いて分析を行う場合、多くのゴールデンクロスの検出が予想される。このため、中期移動平均が長期移動平均を多少超えている場合は、ダイジェストに含めないことが望ましい。そこで(1)式の重要度を利用し、それを指標として小さなピークを除去する方法を検討する。ダイジェストに含むべきピークは重要度の高い上位数点であると考えられる。そこで、重要度の高い上位数点のみをダイジェストに含むべきデータとして取ることにより、小さなピークの除去を行う。

[†] 法政大学大学院情報科学研究科

Graduate School of Computer and Information Sciences, Hosei University

[‡] 法政大学情報科学部

Faculty of Computer and Information Sciences, Hosei University

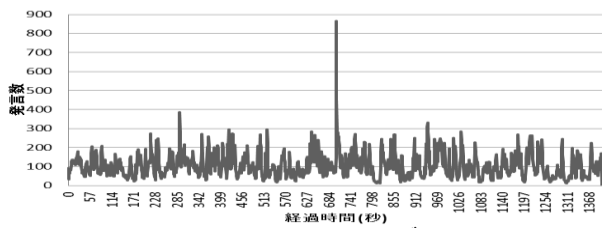


図 1 発言ヒストグラム

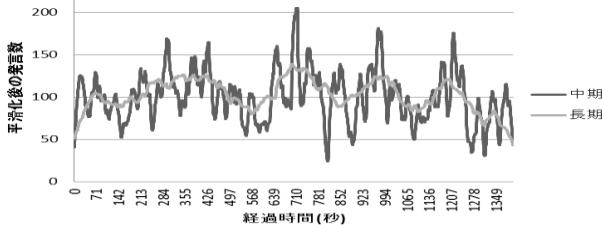


図 2 総発言数による平滑化されたヒストグラム

4. 実験・評価

4.1 発言数ヒストグラム

評価に用いるデータとして、図 1 に発言数をヒストグラムとしたグラフを示す。このデータは再生時間 1404 秒、総発言数が 148184 回のデータを使用して分析を行う。図 1 において、横軸は動画の再生時間 t を表し、縦軸は t における発言数を示している。また、図 2 に図 1 で示した発言数ヒストグラムに対し、移動平均で平滑化したグラフを示す。今回の移動平均では短期移動平均の区間を前後 10 秒間、長期移動平均の区間を前後 60 秒間と設定した。図 2 において、検出されたゴールデクロスは 34 ヶ所であった。しかしグラフからも読み取れるように、すぐにデッドクロスに至るポイントも見受けられるため、3.2 節で述べた手法を用いてこのような小さなピークを除去する。

4.2 ピーク妥当性評価

検出されたピークがダイジェストに相応しいシーンであるかを主観評価する。評価はピーク上位 10 シーンに対して行なった。結果、10 シーンそれぞれ内容的にも問題なく、それぞれのシーンでは発言数の平均値を超える発言が多いことからダイジェストに相応しいシーンだと判断する。しかし、単に積分値の大きなシーンを検出したため各シーンが長めに設定される傾向があり、ピークをさらに意味的なシーンに分割する必要があると考える。

4.3 発言サンプル数検討

本研究で対象とする映像コンテンツの発言は、時間経過と共にデータ数が溜まっていく特性がある。全ての発言に対して分析を行う場合、処理の重さが懸念される。そこで本節では、発言のサンプル数の検討評価を行う。図 3 に発言された日時が最新の 1000 件で作成されたヒストグラム

表 1 検出されたピーク上位 5 シーン

| | 総発言 | 10000 件 | 1000 件 |
|-----|-------|---------|--------|
| 1 位 | シーン 1 | シーン 2 | シーン 5 |
| 2 位 | シーン 2 | — | — |
| 3 位 | シーン 3 | シーン 1 | — |
| 4 位 | シーン 4 | シーン 5 | — |
| 5 位 | シーン 5 | シーン 7 | — |

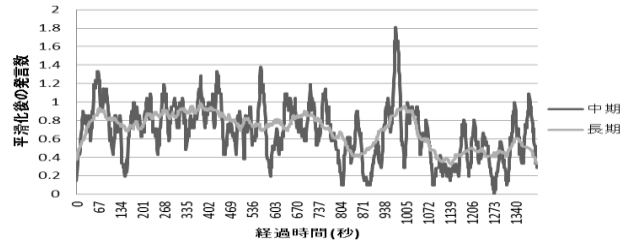


図 3 発言数 1000 件の平滑化されたヒストグラム

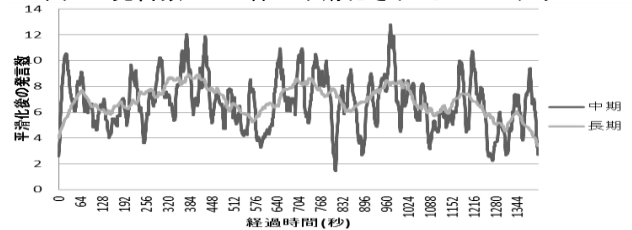


図 4 発言数 10000 件の平滑化されたヒストグラム

の中期移動平均と長期移動平均で平滑化したグラフを示す。図 2 と比較すると長期移動平均が大きく異なっていることが分かる。また、図 4 には発言された日時が最新の 10000 件で作成されたヒストグラムの中期移動平均と長期移動平均で平滑化されたグラフを示す。図 2 と比較すると長期移動平均が同時刻でほぼ同様の傾向を示している。また、ピークについても評価を行う。表 1 に検出されたピークの上位 5 シーンを示す。基準とする総発言数で検出されたピークを、上位 10 シーンに順番にシーン番号を割り当てる。10000 件と 1000 件に記載されているシーンはこのシーンのいずれかを指し、記載が無い箇所は該当データが無かったことを意味する。上位 10 シーンで再現率を比較したところ、発言数が 10000 件のデータでは 10 シーン中 8 シーンが検出され 80%、1000 件のデータでは 10 シーン中 4 シーンが検出され 40% となった。また、表 1 の結果から、再現率が高くともピークの順位は異なることが分かる。ピークの順位は、ダイジェスト生成時に順に重要な点として認識されるため、ダイジェストに含めるピークの数を変更すると発言数を減らした場合に同様の結果が得られない可能性がある。しかし、本研究では動的に変化するダイジェストを想定しているため、大きな問題は無いと考えられる。以上の結果から、総発言数の 15 分の 1 程度の発言数で、ピーク検出には十分な精度が期待できる。

5. まとめ

本研究では映像コンテンツ視聴時にユーザが残す内容についての発言を基にしたダイジェスト生成手法を提案した。移動平均を用いて短期的なノイズと長期的なノイズを除去し、株式のテクニカル分析指標であるゴールデクロスを用いてダイジェストに含有すべきピークの検出を行なった。今後の課題として、様々なジャンルの映像コンテンツに対して実験や、実際にダイジェストを再生するシステムの作成が挙げられる。

参考文献

- [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 卷 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).