

# 同一映像区間を手がかりとしたニュース映像アーカイブの トピック構造解析

社本 裕司<sup>†</sup> 高橋 友和<sup>†</sup> 井手 一郎<sup>†,‡</sup> 村瀬 洋<sup>†</sup>

<sup>†</sup>名古屋大学大学院 情報科学研究科 〒464-8603 愛知県名古屋市千種区不老町

<sup>‡</sup>国立情報学研究所 〒101-8430 東京都千代田区一ツ橋 2-1-2

E-mail: <sup>†</sup>{yshamoto,ttakahashi,ide,murase}@murase.m.is.nagoya-u.ac.jp, <sup>‡</sup>ide@nii.ac.jp

**あらまし** 近年、記憶容量の増大に伴い大量の映像資源を蓄積し、利用することが可能になり、それら映像資源の効率的な利用のための技術が求められている。我々は特に、ニュース映像のトピック構造解析に注目している。従来、トピック構造の解析は一般にテキスト情報のみによる手法が用いられているが、テキストに現れないトピック間の関連が存在する可能性は高い。そこでテキスト情報に加えてニュース映像中に出現する同一映像区間を手がかりとする。この際に問題となるのは、関連するストーリーを対応付ける際に重要ではない映像の存在である。本研究では、同一映像区間の分布情報を利用して関連性の強さに応じた重み付けを行う手法を提案し、テキスト情報のみによる手法と比較することで同一映像区間を出現パターンによって使い分けることの有効性を示した。

**キーワード** 同一映像区間, ニュース映像, トピック構造

## Analyzing topic structure of a news video archive based on near-duplicate video segments

Yuji SHAMOTO<sup>†</sup>, Tomokazu TAKAHASHI<sup>†</sup>, Ichiro IDE<sup>†,‡</sup>, and Hiroshi MURASE<sup>†</sup>

<sup>†</sup> Graduate School of Information Science, Nagoya University Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8603 Japan

<sup>‡</sup> National Institute of Informatics Hitotubashi 2-1-2, Chiyoda-ku, Tokyo, 101-8430 Japan

E-mail: <sup>†</sup>{yshamoto,ttakahashi,ide,murase}@murase.m.is.nagoya-u.ac.jp, <sup>‡</sup>ide@nii.ac.jp

**Abstract** Recently, video resources are accumulated and used in large quantities by the growth of memory. Technologies for efficient utilization of these video resources are needed. Especially, we focus on analyzing topic structure of a news video. Currently, text information is generally used for the analysis of topic structure, however, relation which do not appear in the text among topics exist. In this paper, we make use of near-duplicate video segment in addition to text information in news video for this task. In this case, the existence of videos which is not important to connect the news stories is a problem. In order to handle such cases, we put weight to near-duplicate video segments according to the strength of relation by their temporal distribution. The method is compared with a method which uses only text information, to demonstrate the effectiveness to make use of near-duplicate video segments based on the pattern of its distribution.

**Keyword** Near-duplicate video segments, News video, Topic structure

### 1. はじめに

#### 1.1. 背景と目的

近年、情報通信技術の進歩により大量の映像が放送されるようになってきている。また、記憶容量の増大に伴い、膨大な量の映像資源を蓄積し利用することが可能になっている。よって、それら映像資源の効率的な利用のために、映像の理解、検索、要約などの技術が求められている。その中で、映像が持つ意味構造を解析する研究に注目が集まっている。

映像にはスポーツ、ドラマ、CMなどのさまざまなジャンルがあるが、本研究では、特に検索や再利用の価値が高いと考えられるニュース映像に着目する。ニュース映像において、話題追跡やトピック構造の解析は、ある話題に関するニュースの検索や要約映像の生成などの効率的な利用に非常に有効であると考えられる。従来研究ではニュース映像の話題追跡や構造の解析には、テキスト情報を用いるのが一般的である[2]が、テキストに現れないようなニュースストーリー間の

関連が存在する可能性は十分にある。

そこで、本研究では、テキスト情報とともに画像情報の利用を考える。関連のあるニュースストーリー間では、特定の映像を繰り返して利用することがある。つまり、ニュース映像の中に同一の映像区間が存在する場合、その区間が属するニュースストーリー間には意味的な関連がある可能性が高い。

よって、本研究では画像情報を利用することでテキスト情報を補完して新たな関連を発見し、ニュース映像のトピック構造を解析することを目的とする。

## 1.2. 用語の整理

本研究に関連する用語を整理する。

一般に映像は以下の階層的構造により構成される。

- フレーム：映像を構成する静止画像
- ショット：画像的に連続しているフレーム群
- シーン：意味的に連続しているショット群

一方、ニュース映像固有の概念<sup>1</sup>には以下のようなものがある。

- イベント：ある日ある場所で起こった出来事
- ストーリー：あるニュース映像中で一つのイベントを扱っている単位
- トピック：直接的に関連のある複数のイベントを扱っている一連のストーリー群

## 2. 関連研究

ニュース映像のトピック構造解析に関する研究は、テキスト情報に基づく手法が一般的であった。しかし、最近では画像情報を用いた研究が増加してきている。小川らはテキスト情報だけでは困難な、多言語ニュース映像間の同一ニュースイベントの検索を、映像間に出現する同一映像区間を手がかりにした[3]。Wuらは、映像のキーフレームに対してクラスタリングを行うことで類似しているフレームを抽出し、それを手がかりにトピック構造の解析を行う手法を提案した[4]。Duyguluらは、類似したキーフレーム中の出現するロゴ情報や同一映像区間が、ストーリーが同一のトピックに属するかどうかを判定するのに有効であることを検証した[5]。これらの手法は、後で詳しく述べるように強い意味的関連性を示さない映像の存在を考慮していないという問題点がある。

## 3. 同一映像区間を手がかりとしたトピック構造解析

### 3.1. ストーリー分割

はじめにニュース映像をストーリーごとに分割する。本研究では井手らの手法[2]を用いてストーリー分割を行った。処理の大まかな流れは以下のものである。

1. 映像に付随するテキスト各文ごとに名詞列を抽出
2. 文の境界を基準として前後  $w$  文を各々結合したキーワードベクトル間の類似度を求める
3.  $w$  を一定の範囲内で変化させたときの類似度の最大値があるしきい値  $\theta_{seg}$  以下の時にその境界をストーリーの境界とする

### 3.2. 同一映像区間の検出

次に、ニュース映像中に出現する全ての同一映像区間を検出する。本研究では野田らの手法[1]を用いて同一映像区間の検出を行った。この手法は時間方向と空間方向の2段階次元圧縮を行うことによって単純な照合と比べて1000倍以上に検出を高速化することに成功している。

### 3.3. 同一映像区間に基づく関連性評価

3.2節で検出された同一映像区間を手がかりとしてストーリー間の関連性を評価する。この際、どのような同一映像区間も同等に扱えないという問題点が挙げられる。例えば、図1のような特定の機能を持つ建物などの資料映像は一般的にその機能と関連して長期間にわたって利用される傾向があり、ストーリー間の直接の関連を示すとは限らない。一方で、図2のような事件・事故が起きた瞬間の映像(以下事件映像)などは、短期間に集中して利用される傾向があり、ストーリー間の直接の関連を示している可能性が高い。



図1 資料映像の例



図2 事件映像の例

そこで本研究では同一映像区間の分布状況に着目する。山岸らは、同一映像区間の分布状況を閲覧するブラウザの実装を行った[6]。

先に述べた傾向を示す特徴として同一映像区間対  $i, j$  に対して出現密度  $d_{i, j}$  を(1)式のように定義する。

$$d_{i, j} = \frac{\text{ある期間で } i, j \text{ と同一の映像が出現するストーリー数}}{\text{ある期間の日数}} \quad (1)$$

ここである期間は同一映像区間対  $i, j$  のうち時系列的に遅いほうからさかのぼって  $i, j$  と同一の映像が最初に出現するまでとする。これは、事故等の映像は最初に出現した直後に集中して利用されることを考慮している。また、同一映像区間対に出現密度を定義したのは、同じ映像でも出現時期によって重要度が異なるためである。

出現密度の求め方の例を図3に示す。

<sup>1</sup> 米国NIST が実施したTDT(Topic Detection and Tracking) ワークショップにおける定義

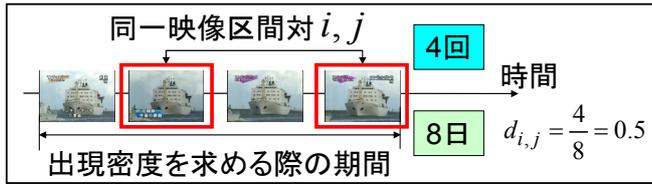


図3 出現密度の求め方の例

また、同一映像区間に基づくストーリー  $k, l$  間の類似度  $R_{image}(k, l)$  を式(2)のように定義する。

$$R_{image}(k, l) = \begin{cases} \max(d_{i_1, j_1}, d_{i_2, j_2}, \dots) & (k \text{ と } l \text{ の間に同一映像区間対} \\ & (i_1, j_1), (i_2, j_2), \dots \text{ が存在する}) \\ 0 & (k \text{ と } l \text{ の間に同一映像区間対 が存在しない}) \end{cases} \quad (2)$$

### 3.4. テキスト情報に基づく関連性評価

一方テキスト情報に基づく関連性評価は、井手らの手法[2]を用いた。この手法は文字放送字幕テキストから各ストーリーごとに名詞ベクトルを抽出し、抽出されたベクトルのコサイン距離を類似度としている。

### 3.5. ストーリー間の対応付け

3.3節で求めた同一映像区間に基づく類似度と3.4節で求めたテキスト情報に基づく類似度を組み合わせることで関連のあるストーリー間の対応付けを行う。

本研究では、以下の2つの条件を満たす場合に2つのストーリー間に関連があるとして対応付ける。

$$\cdot R_{text}(k, l) \geq \theta_1 \quad (3)$$

$$\cdot wR_{image}(k, l) + (1-w)R_{text}(k, l) \geq \theta_2 \quad (4)$$

### 3.6. トピックスレッド構造の構築

「トピックスレッド」とは、時系列順に連鎖している一連のストーリーのことであり、複数にトピックスレッドが組み合わせられたものをトピックスレッド構造と呼ぶ。本研究では、井手らの手法[2]を用いてトピックスレッド構造の構築を行った。

## 4. 実験結果と考察

### 4.1. 実験方法

2003年5月から2003年7月に実際に放送されたNHKのニュース番組News7に対してストーリー間の対応付けとトピックスレッド構造の構築を行った。ストーリー間の対応付けの際は検出された同一映像区間のうち人手で正解と判断したもののみを利用した。

実験で用いたパラメータを以下に示す。

- テキスト情報のみのしきい値  $\theta_1 = 0.1$
- 重み付け和のしきい値  $\theta_2 = 0.2$
- 組み合わせる際の比率を示すパラメータ  $w = 0.5$

比較手法として従来手法であるテキスト情報のみによる手法でも同様の実験を行った。

また、評価のための被験者実験を行った。

被験者実験では同一映像区間が存在する45組のニュースストーリーに対して以下の手順で実験を行った。

1. 同一映像区間を提示

2. 1の同一映像区間を含む2つのニュースストーリーを提示
3. 2つのニュース間に関連があるかどうかを判断  
1組のニュース映像に対して5人の被験者に評価させ、3人以上が関連有と評価したものを対応付けの正解とした。

### 4.2. 実験結果

被験者に評価させた45組のニュースストーリーに対して、提案手法と従来手法による対応付けの結果を表1に示す。

表1 ストーリー間対応付けの実験結果

	提案手法	従来手法
適合率	90% (35/39)	89% (17/19)
再現率	95% (35/37)	46% (17/37)

同一映像区間を手がかりとすることで対応付けの精度が向上したことが確認された。

次に、出現密度と被験者による評価との関係を図4に示す。

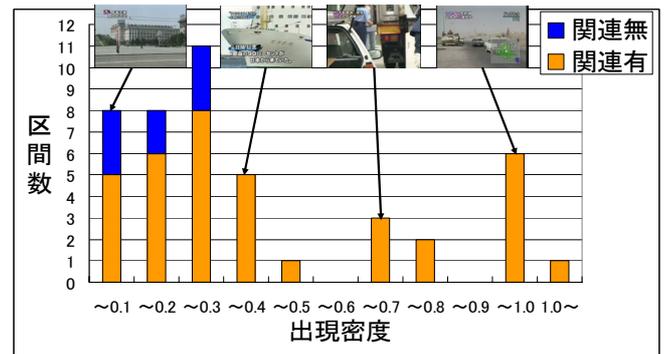


図4 出現密度と被験者による評価との関係

被験者によって関連なしと評価されたストーリー間に存在する同一映像区間対は出現密度が低いことが確認された。

同一映像区間の利用によりトピックスレッド構造が変化した例とそのなかに存在した同一映像区間を図5に示す。

ここで「ダスキン特別背任事件」と「埼玉県知事長女政治資金規制法違反」という2つのイベント間には関連がある。従来手法ではこの2つのイベントはつながっていない。しかし、同一映像区間を用いることで2つのイベントがつながるようになり、提案手法によるトピックスレッド構造に意味があることが確認された。

### 4.3. 考察

出現密度と被験者による評価との関係については、関連ありと評価されたストーリー間に存在する同一映像区間対の出現密度は一樣に分布していた。これは時間的に間隔をおいて進行するイベントの存在が原因であると考えられる。このようなイベントの場合には、資

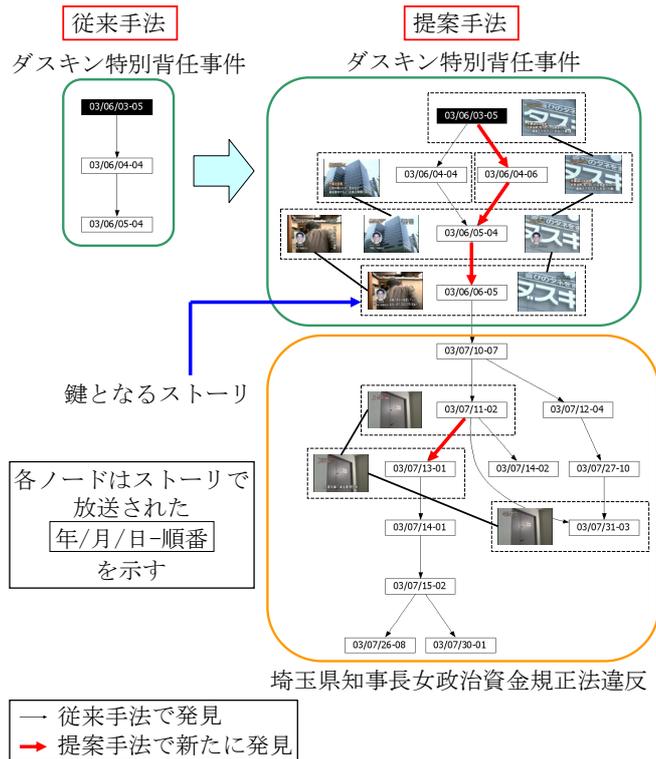


図5 提案手法によるトピックスレッド構造(一例)

料映像でない場合でも分散して出現すると考えられる。

実験では徐々に進行するイベントの映像と資料映像の差が分布上に現れなかったことから、必ずしも3ヶ月では1つのイベントが収束しないことがあることを示している。よって、数年規模での実験を行うことにより、徐々に進行するイベントの映像と資料映像を区別することが可能であると考えられる。

トピックスレッド構造の構築については、この例では関連のある2つのイベント間には同一映像区間が存在しなかった。しかし、同一映像区間の利用により、2つのイベントをつなぐ鍵となるストーリー(03/06/06-05)との関連が発見されたため2つのイベントが繋がったと考えられる。実際、鍵となるストーリーの内容は「ダスキンの特別背任事件で埼玉県知事の長女が経営する会社に金が支払われていた」という内容であった。

このように1つのイベントを扱っているストーリーを数多く発見することによって、関連のあるイベントが繋がる可能性がある。また、誤った関連が増えると、まったく関連のないイベントとつながってしまう可能性も考えられる。よって、ストーリー間の関連を数多くかつ正確に発見することが重要である。そのために、よりよい特徴や画像情報とテキスト情報の連携手法を検討していく。

## 5. むすび

本研究ではニュース映像に出現する同一映像区間を手がかりとしてニュースストーリー間の関連性を評価

し、トピックスレッド構造を構築する手法を提案した。

また、ストーリー間の関連性を評価する場合に、すべての同一映像区間を同等に扱うのは適切ではないので、出現密度による重み付けを導入した。

その結果、出現密度がストーリー間の関連性の強弱をある程度示しており、ストーリー間対応付けの精度が向上することが確認された。

今後の課題としては以下のような点が挙げられる。

- 重み付けに適した新たな特徴量の検討
- より大規模なニュース映像への適用
- 画像情報とテキスト情報の新たな連携手法の検討

## 謝辞

日頃より熱心に御討論頂く名古屋大学村瀬研究室の諸氏に感謝する。また、本研究で使用した映像を提供して頂いた国立情報学研究所の佐藤真一研究室に感謝する。本研究の一部は日本学術振興会科学研究費補助金、21世紀COEプログラム「社会情報基盤のための音声・映像の知的統合」による。

## 文 献

- [1] 野田和広, 高橋友和, 井手一郎, 目加田慶人, 村瀬 洋: “適応的特徴選択を用いた長時間放送映像からの高速な繰り返し区間検出”, 電子情報通信学会技術研究報告, PRMU2005-289, (Mar 2006)
- [2] 井手一郎, 孟 洋, 片山紀生, 佐藤真一: “大規模ニュース映像コーパスの意味構造解析”, 電子情報通信学会技術研究報告, PRMU2003-97 (September 2003)
- [3] 小川 晃, 野田和広, 高橋友和, 井手一郎, 村瀬洋: “同一区間を手がかりとした同一ニュースイベントの言語横断検索”, 画像の認識・理解シンポジウム(MIRU 2006), IS-3-45, pp.1278-1283 (July 2006)
- [4] Xiao Wu, Chong-Wah Ngo, and Qing Li: “Threading and Autodocumenting News Videos”, IEEE Signal Processing Magazine, Vol.23, No.2, pp.59-68 (March 2006)
- [5] P. Duygulu, J.-Y. Pan, and D. A. Forsyth: “Towards Auto-Documentary: Tracking the Evolution of New Stories” Proc. 12th ACM International Conference on Multimedia, pp.820-827 (October 2004)
- [6] 山岸史典, 佐藤真一: “同一映像断片に基づくニュース映像ブラウザの実装”, 電子情報通信学会技術研究報告, PRMU2003-51, (July 2003)