

## 実世界シナリオに基づく映像の再編纂

[正会員] 井手 一郎<sup>†</sup>, 村瀬 洋<sup>†</sup>

キーワード 放送映像, 映像アーカイブ, 編集, 要約, ニュース, スポーツ, 料理

## 1 ま え が き

近年, 記憶媒体の大容量化に伴い, 大量の放送映像を電子的に蓄積したアーカイブを構築し, コンピュータ上で検索・閲覧する技術に関する研究が盛んである。しかし, 映像はテキストや画像と比べて一覧性が悪いいため, 検索結果を確認するだけでも長時間を要するという問題がある。

そこで筆者らは, 単純に検索結果を一覧表示するのではなく, 映像コンテンツに関連する何らかの実世界情報(ウェブを含む, さまざまなソーシャルメディアから得られる情報)をいわば「シナリオ」とみなして, アーカイブ中の映像を再編纂して提示する方法について検討している。本稿では, ニュース, スポーツ, 料理の3分野における筆者らの取組みについて紹介する。

## 2 ニュースコンテンツ

ニュースコンテンツに関しては, 2種類の実世界シナリオを想定した映像再編纂方法を紹介する。

## 2.1 Wikipedia記事のマルチメディア化

- ・映像: ニュース映像
- ・シナリオ: 時事関係のWikipedia記事

オンライン百科事典Wikipedia<sup>\*1</sup>は, 不特定多数の利用者が随時加筆修正を行う枠組みにより実現された, 実世界のさまざまな事象に関する即時性が高い情報源である。しかし, テキスト中心の記述であり, 画像・映像が付加されていることは少ない。そこで筆者らは, 時事関係のWikipedia記事中の経緯に関する時系列的記述に対して, ニュース映像の該当するストーリー<sup>\*2</sup>を自動的に対応付ける手法を開発するとともに, 閲覧インタフェース「Videopedia」を実装した。以下でその手法について簡単に紹介する。

\*1 <http://ja.wikipedia.org/wiki/>

\*2 一つのイベントについて扱った, ニュース映像の意味的な最小単位。ここでは, 文献2)の手法により, 自動的に分割されていると想定する。

<sup>†</sup>名古屋大学 大学院情報科学研究科

"Video Recompile Based on Real-world Scenarios" by Ichiro Ide and Hiroshi Murase (Graduate School of Information Science, Nagoya University, Nagoya)



図1 「Videopedia」インタフェース

## [Wikipedia記事とニュースストーリーの対応づけ]

まず, Wikipedia記事中出现する日付表現を抽出し, 放送日により, 対応づけるストーリー候補を絞り込む。次に, Wikipedia記事とストーリーの文字放送字幕<sup>\*3</sup>中の名詞の出現頻度ベクトルを比較することで, 両者の話題の類似度を測り, 閾値以上のものを対応づける。

## [トピックスレッド構造に基づく対応づけの補間]

ニュース番組とWikipedia記事では, 記述の粒度が異なるため, 以上の処理だけでは充分な対応づけが得られない。そこで, 筆者らが別途提案した, トピックの変遷を追跡した「トピックスレッド構造」<sup>2)</sup>を利用して, トピックスレッド上の対応づいたストーリー間を補間するように, 洩れたストーリーを映像に対応づける。

以上の手法を, 三つの時事関連Wikipedia記事に適用した結果について人手で確認したところ, 適合率82.1%, 再現率75.8%の精度で正しく対応づけられていた。

このようにして得られた対応づけに基づいて, 時事関係のWikipedia記事中の, 経緯に関する時系列的記述と該当する映像を同期して, 閲覧するためのインタフェース「Videopedia」を図1に示す。

\*3 主に聴覚障害者のために, 音声を手動的に書き下したテキストで, 映像とともに放送されている。



なお、本研究の詳細については、文献1)を参照されたい。

## 2.2 特定人物に注目した要約映像の制作

- ・映像：ニュース映像
- ・シナリオ：内閣支持率\*4

放送局などにおいて、社会的に広く知られた人物の引退・死去などにあたり、その来歴を説明する映像を制作することがある。引退・死亡という出来事は往々にして突発的に発生するため、アーカイブ中の膨大な量の映像から、限られた時間の中で必要な映像素材を収集して、短い映像に編集し直す必要がある。

そこで筆者らは、ニュース映像に頻出する人物である「内閣総理大臣」に注目し、辞任へ至った過程を説明するような要約映像の制作を支援する方法を考えた。具体的には、支持率の変化やトピックの重要度を手がかりにして、編集に必要と思われる素材映像を、大規模ニュースアーカイブから自動的に収集する方法を開発した。以下でその手法について簡単に紹介する。

### 〔処理対象映像の抽出〕

まず、要約映像を制作したい内閣総理大臣の在任期間中に、その名前が主語になっている文を含むすべてのストーリーを、処理対象映像として抽出しておく。

### 〔定型的イベントに関する映像の検出〕

就任直後や辞任直前の期間については、施政方針演説や辞任会見など定型的に生じるイベントについて、時期とキーワードからなるテンプレートを用意して、検出する。

### 〔支持率変化に基づく重大イベントに関する映像の抽出〕

まず、月例支持率に以下の現象が生じた期間に、重大イベントが発生したと考え、イベント発生期間とする。

- (1) 支持率の大幅な増減
- (2) 支持率と不支持率の逆転
- (3) 圧倒的な高支持率または低支持率

図2に、支持率グラフと、検出されたイベント発生期間の例を示す。

次に、イベント発生期間中のストーリーから、イベントに関する情報を含むと思われるものを推定する。具体的には、期間中の各ストーリーについて、以下の条件を満たすものをイベント候補として抽出する。

- (1) トピックスレッド構造の始点または終点
- (2) トピックスレッド構造上で前後にストーリーが密集

これらの条件は、当該ストーリーがトピックの時間的変遷の中で、その発生または決着に関するイベントを含むか、大きな話題になっていることから、重大イベントを含む可能性が高いと考えて設定した。

最後に、各イベント候補ストーリーのテキストに対して、評判分析を適用する。その後、編集者の意図が、肯定的な映像を制作したいか、否定的な映像を制作したいかによって、肯定度または否定度順に整列しておく。

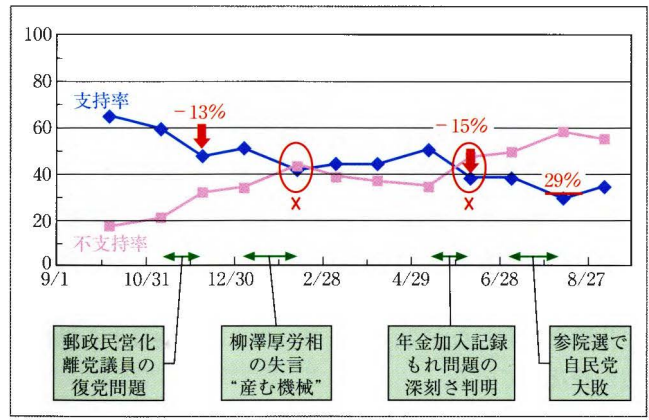


図2 内閣支持率とイベント期間検出の例 (安倍首相)



図3 制作された要約映像の例 (安倍首相)

### 〔要約映像の編集〕

最終的に制作する映像長と、イベント発生期間数に依りて、各イベントに対する採用映像数・映像長を割り振る。この際、各イベント発生期間について、整列したストーリーの上位から順に、指定された映像数だけ選択する。

最終的な映像は、支持率グラフにイベント発生時の支持率変化を示したCG (図2のような画像) を挿入し、その後に変化を裏付けられる映像素材から切出された映像断片を結合することによって得る。切出しにあたっては、当該総理大臣の名前が発話された瞬間を始点とすることで自動化も可能であるが、実用レベルの完成度を求めるならば、これより先は人手に委ねた方が良い場合もある。

図3に、提案手法により選択されたストーリーから、最終的に人手で切出した区間を結合して制作した要約映像の例を示す。

なお、本研究の詳細については、文献4)を参照されたい。

\*4 「政治意識月例調査」<sup>3)</sup>中の月ごとの内閣支持率。



### 3 スポーツコンテンツ：twitterのつぶやきに基づくスポーツ映像の要約

- ・映像：スポーツ中継映像
- ・シナリオ：twitterの投稿数・投稿者の属性

近年、マイクロブログサービスtwitter<sup>\*5</sup>による手軽なつぶやきの投稿が流行している。特に、スポーツの試合などは、間接・直接に視聴する多数の利用者が、リアルタイムに体験を共有しながらtwitterでつぶやきを投稿する視聴形態が普及している。

そこで筆者らは、大量のtwitterのつぶやきを収集し、属性別投稿数に基づいて、応援チーム別に要約映像を制作する方法を開発できるようにした。以下でその手法について簡単に紹介する。

#### 〔投稿者の属性推定〕

試合時間中に投稿されたすべてのつぶやきの投稿者ごとに、一連のつぶやきから応援しているチーム（＝投稿者の属性）を推定する。投稿者の属性を推定するために、応援チームごとに出現しやすいキーワードを事前に学習しておく。このために、応援チームに関するタグが付与されている投稿について、SO-PMI (Semantic Orientation Using Pointwise Mutual Information) と呼ばれる手法<sup>5)</sup>を適用して、属性評価辞書を作成した。これにより、属性を示すタグが付与されていない個々のつぶやきについても、属性を推定することができるようになった。

#### 〔属性別投稿頻度に基づくチーム別盛り上がり推定〕

次に、属性別の投稿頻度から、チーム毎の盛り上がり状況を分析する(図4)。これに基づいて、試合の盛り上がりを集めた要約映像を制作する。図5に、実装した要約映像閲覧インタフェース「twiSpo」を示す。このインタフェースにより、視聴者の属性を切替えたうえで、同じ属性の投稿者が多数つぶやきを投稿したハイライトシーンからなる要約映像を制作して提示できるようになった。

なお、本研究の詳細については、文献6)を参照されたい。

### 4 料理コンテンツ：テキストレシピのマルチメディア化

- ・映像：料理番組映像
- ・シナリオ：調理レシピ

近年、投稿型調理レシピサイトが爆発的に普及し、日々大量のレシピが投稿されている。これらのサイトは一般ユーザーが気軽に調理レシピを投稿できる反面、撮影や編集の手間がかかるため、手順に対応する画像や映像が付与されることは依然、少ない。

そこで筆者らは、料理番組映像やウェブ上の調理映像から、特定の素材に対して特定の調理作業を施している映像断片を大量に収集、索引付けしたデータベースを作成して

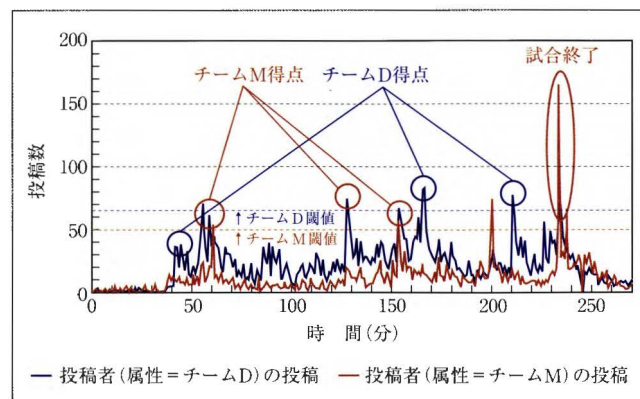


図4 twitterのつぶやきによる属性別ハイライト検出の例

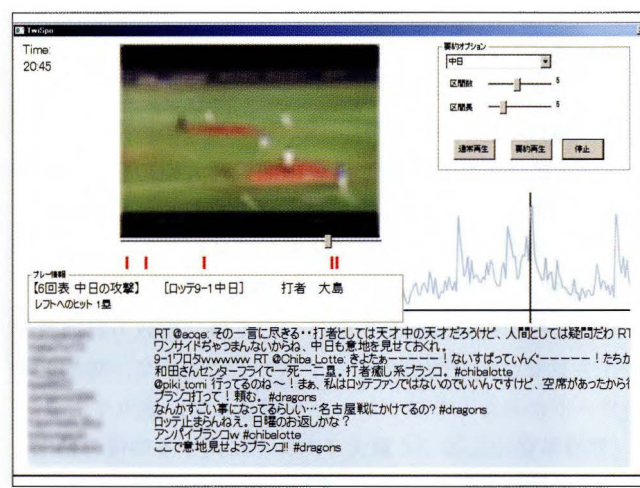


図5 twitterのつぶやきに基づくスポーツ映像要約閲覧インタフェース「twiSpo」

おき、それらを自動的にテキストレシピに対応付けすることで、任意の調理レシピに映像を付与する枠組みを考えた。以下でその手法について簡単に紹介する。

#### 〔調理動作映像データベースの構築〕

多くの料理番組映像には、音声を書き下した文字放送字幕が付随して提供される。これについて、食材を表す名詞と調理動作を表す動詞に関する簡単な係り受け解析を行うことで、対応する映像区間に「食材－調理動作」の組を索引付けする。

しかしこれだけでは、映像中で実際に当該調理動作を行っている保証がない。また、当該調理動作だけではなく、前後に冗長な映像が含まれることが多い。そこで、図6に示すように、映像区間中の動作特徴(特徴空間中の軌跡の形状、および、画像中の繰り返し動作領域の分布)を分析、分類したうえで、調理動作の種類との整合性を確認し、整合する場合のみ、動作区間のみ切出した映像断片をデータベースに格納する。

\*5 <http://twitter.com/>



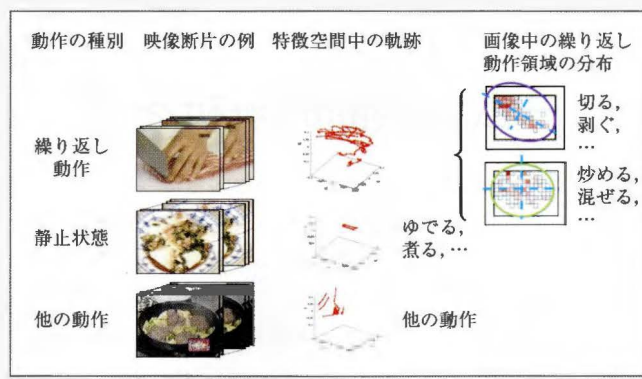


図6 動作特徴に基づく調理動作の分類

### [テキストレシピへの映像付与]

既存のテキストレシピの構造を解析し、調理手順に関する記述から、「食材-調理動作」の組を抽出する。次に、調理動作映像データベースを参照することで、この組と対応するすべての映像断片を検索する。また、組と一致する映像断片がなかった場合は、改めて調理動作のみ一致するものを検索する。

実際には、大量の料理番組映像に本手法を適用し、さまざまな食材と調理動作の組合せを網羅したデータベースを構築する必要があるが、筆者らは、人手で必要な映像を収集、索引付けして作成した小規模なデータベースを用いた、マルチメディア化レシピインタフェース「Video CookKing」(図7)を試作して、提案手法の効果を確認した。

なお、本研究の詳細については、文献7)を参照されたい。

## 5 む す び

本稿では、映像コンテンツに関連するさまざまな実世界情報をいわば「シナリオ」とみなして、アーカイブ中の映像を再編集して提示する方法について、ニュース、スポーツ、料理の3分野における取組みについて紹介した。

実用レベルのコンテンツを制作するためには、各手法の精度を高めるとともに、より多様な素材を収集して用いる必要があるため、今後は、放送映像だけでなく、ウェブ上の画像・映像を含めた枠組みを考えていきたい。

最後に、以上に紹介した研究の大部分は、科研費特定領域研究「情報爆発IT基盤」、若手研究(B)、基盤研究(B)、日本学術振興会優秀若手研究者海外派遣事業、国立情報学研究所およびアムステルダム大学との共同研究による。また、各研究に携わった研究室のスタッフおよび学生諸氏に感謝する。

(2011年12月19日受付)

## [文 献]

- 1) T. Okuoka, T. Takahashi, D. Deguchi, I. Ide and H. Murase: "Labeling news topic threads with Wikipedia entries", Proc. 11th IEEE Int. Symposium on Multimedia, pp.501-504 (Dec. 2009)

【田舎風野菜スープ】

エネルギー: 70kcal 調理時間: 20分 講師: 成川 朝

タグ: じゃがいも たまねぎ にんじん キャベツ スープ パーコン 食べるスープ 野菜スープ

元レシピ: [http://www.kyounoryouri.jp/recipe/4162\\_田舎風野菜スープ.html](http://www.kyounoryouri.jp/recipe/4162_田舎風野菜スープ.html)

【材料】 (2人分)

- ・キャベツ 1枚
- ・にんじん 1/4本
- ・たまねぎ 1/2コ
- ・じゃがいも 1コ
- ・パーコン(薄切り) 1枚
- ・ハーブ\* 一つまみ (サラダ油・塩・こしょう)
- \*タイム、バジルなど

【作り方】

1. キャベツは縦切りにする。にんじんは2mm厚さのいちょう切りにする。たまねぎは横切りにする。じゃがいもは半分に切ってから、2mm厚さに切り、水にさらして1~2度水を切り替える。水けをきる。パーコンは5mm幅に切る。
2. 耐熱性容器(ボウルでもよい)に、1の材料を入れ、サラダ油小さじ2を加えてざっと混ぜ、ふたをして電子レンジ(500W)で約3分間加熱する。
3. 2に塩小さじ1、こしょう少々、ハーブ、水カップ2+1/2を加え、電子レンジで約12分間加熱する。
4. 取り出して器に盛る。

動画での解説

調理動作: 「薄切りにする」

・食材: たまねぎ / たまねぎ2 / たまねぎ3

図7 マルチメディア化レシピ閲覧インタフェース「Video CookKing」

- 2) 井手, 木下, 高橋, 孟, 片山, 佐藤, 村瀬: “大量ニュース映像を対象とした時系列意味構造に基づく情報編纂手法の提案”, 人工知能学論, 23, 5, pp.282-292 (Sep. 2008)
- 3) NHK放送文化研究所: “放送研究と調査 (月報)”, NHK出版, ISSN: 0288-0008
- 4) F. Nack and I. Ide: "Why did the Prime Minister resign? -Generation of event explanation from large news repositories-", Proc. 19th ACM Int. Multimedia Conf., pp.313-322 (Nov. 2011)
- 5) P.D. Turney: "Thumbs up? Thumbs down? Semantic orientation applied to unsupervised classification of reviews", Proc. 40th Annual Meeting of the Association for Computational Linguistics, pp.417-424 (July 2002)
- 6) 小林, 野田, 出口, 高橋, 井手, 村瀬: “Twitterの実況書込みを利用したスポーツ映像の要約”, 信学技報, MVE2010-162 (Mar. 2011)
- 7) 道満, カイ, 高橋, 井手, 村瀬: “マルチメディア料理レシピ作成のための料理レシピテキストと料理番組映像との対応付け”, 信学論(A), J94-A, 7, pp.540-543 (July 2011)



**井手 一郎** 2000年, 東京大学大学院工学系研究科博士課程修了。同年, 国立情報学研究所助手。2004年, 名古屋大学大学院情報科学研究科助教授。2007年, 同准教授となり, 現在に至る。博士(工学), 正会員。



**村瀬 洋** 1980年, 名古屋大学大学院工学研究科修士課程修了。同年, 日本電信電話公社(現NTT)入社。2003年, 名古屋大学大学院情報科学研究科教授となり, 現在に至る。博士(工学)。