

人物関係に基づくニュース映像の検索と閲覧

井手 一郎 佐藤 真一

国立情報学研究所

〒101-8430 東京都千代田区一ツ橋 2-1-2

TEL: (03)4212-2585 FAX: (03)3556-1916

{ide,satoh}@nii.ac.jp

あらまし 従来のニュース映像の検索は、付与された索引語を検索し、該当箇所を閲覧するだけのものが主流であった。我々は、より高度な映像の利用を実現するために、大量のニュース映像から抽出される構造化された知識に基づく検索と閲覧を考えている。本発表では、大量のニュース映像の話題中に出現する人物名の共起に注目することにより人物関係に関する知識を抽出し、それに基づいて映像の検索と閲覧を行なう手法を提案する。

キーワード ニュース映像, 検索, 閲覧, 人物関係, キャプション, 人物名

Retrieval and browsing of news video by human relationship

ICHIRO IDE

SHIN'ICHI SATOH

NATIONAL INSTITUTE OF INFORMATICS

2-1-2 HITOTSUBASHI, CHIYODA-KU, TOKYO 101-8430, JAPAN

TEL: +81-3-4212-2585 FAX: +81-3-3556-1916

{ide,satoh}@nii.ac.jp

Abstract Most conventional news video retrieval and browsing was based on simple keyword retrieval and browsing of the retrieved segment. We are considering a retrieval and browsing method that utilizes structured information derived from the target video contents. In this presentation, we will introduce a retrieval and browsing method based on human relationship that is derived from cooccurrence of human names in large news video corpus.

key words news video, retrieval, browsing, human relationship, caption, human name

1 はじめに

1.1 研究の背景と目的

近年、放送される映像量が増加するに伴い、放送された映像の検索や再利用への要求が高まっており、映像の閲覧や索引付けに関する研究が盛んに行われている。なかでも、資料的価値が高いことから、ニュース映像を扱うことが注目されている。

本稿では、大量に蓄積されたニュース映像の検索と閲覧を行うために、映像コーパスから抽出された情報を用いる手法を提案する。具体的には、ニュース映像コーパス中の音声の書き下しテキストから人物名の共起情報を集計し、得られた共起関係を人物関係とみなして可視化し、効果的な検索や閲覧を実現することを目指す。

1.2 用語の定義

以下のように映像に関連する用語を定義する：

- **映像**
映像とは、動画像・音声・文字放送データからなる、一まとまりのマルチメディア情報のこととする。
- **フレーム**
フレームとは、動画像の最小構成単位である静止画像のこと。NTSC放送方式では、動画像は毎秒約30フレームで構成される。
- **ショット・カット**
ショットとは、画像的に連続なフレームの集合のこと。カットとは、ショット間の不連続な点のこと。
- **シーン（話題）**
シーンとは、意味的に連続するショットの集合のこと。ニュース映像においては、シーンは話題に対応する。

これに関連して、図1に映像の物理的構成を示す。

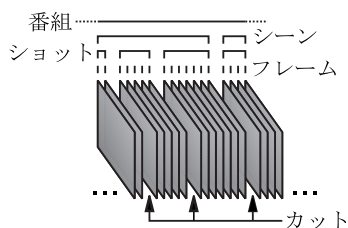


図 1: 映像の物理的構成

2 大規模映像コーパス

2.1 映像自動獲得・蓄積システム

大量のニュース映像からの情報抽出を行うために、筆者らは映像自動獲得・蓄積システムを構築している [5]。図2にその構成を、図3にその全景を示す。

このシステムでは、動画像・主音声と文字放送字幕（クローズドキャプション）をリアルタイムで自動的に獲得・蓄積するようになっている。これにより大規模な映像コーパスが自動的に構築され、特定のニュース番組¹の映像を理論的には年間180時間程度蓄積できることになる。

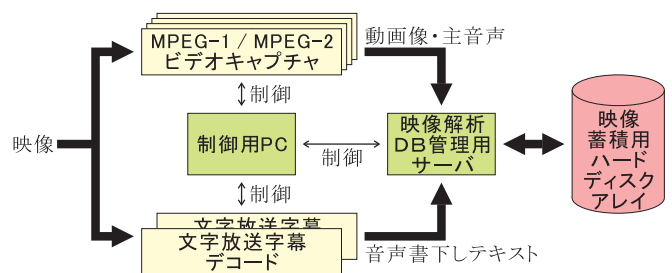


図 2: 画像自動獲得・蓄積システムの構成



図 3: 画像自動獲得・蓄積システムの全景

2.2 コーパスからの情報抽出

映像コーパスの構築と情報抽出に関する研究としては、佐藤らによるもの [7] がある。この研究では、映像の物理

¹本研究では、音声の書き下しテキストを用いるため、文字放送字幕が付与されている特定のニュース番組を利用する。文字放送字幕のニュース番組へのリアルタイムな付与については文献 [1] を参照のこと。

的構成や動画像の様々な画像特徴量に対する分析を行っている。しかし、本研究ではそのような映像から得られる表面的な特徴を分析対象にするのではなく、映像の意味的内容に踏み込んだ情報抽出を目指す。

ここでは、新聞記事やニュース映像の内容の理解において重要と思われる、いわゆる5W1H(When, Where, Who, What, Why, How)情報のうち、特に「誰が(Who)」,つまり登場人物に注目した情報抽出を行う。このような限定は一般の映像に対しては必ずしも有効ではないかもしれないが、ニュース映像においては、人間による人間のための人間に関する情報の提供が主眼であることを考えると、適当であると考えられる。

3 人物関係に基づく検索・閲覧

3.1 概要

以上のような考えに基づき、ニュース映像中の登場人物に注目した情報抽出を行い、その結果を利用した映像の検索と閲覧に関する手法を提案する。

従来から、ニュース映像に登場する人物に注目した索引付け手法[3]や、検索・閲覧手法[6]が提案されている。これらは、動画像中に登場する人物像と人物名とを対応付ける、いわば異メディア間の一次的な対応付け手法である。

しかし、本研究では、登場人物の共起という二次的な知識、つまり人物関係と映像とを対応付けることを考える。ここでいう人物関係とは、図4に例示する、テレビ雑誌などに掲載されることがある、ドラマにおける登場人物関係図に示される登場人物間の関係のようなものである。このように人物関係を可視化することにより、特定の人物とその周囲の人物関係をたどりながら映像内容を追跡することができ、より話題内容に則した検索・閲覧が可能になる。

具体的には以下の手順で人物関係の抽出を行い、映像との対応付けを行う。

1. 音声書き下しテキストから人物名を検出
2. 動画像から話題境界を検出
3. 話題内の人物名の共起から人物関係を抽出

以下の各節で各手順について詳細を述べる。

3.2 人物名の検出

テキストから人物名を抽出する手法は、固有表現抽出問題(Named entity task)[8, 9]の一部として、研究さ

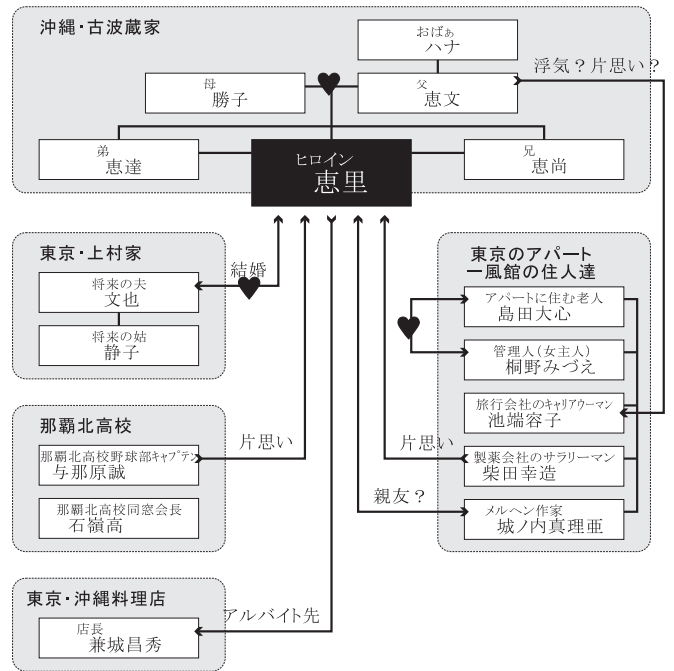


図 4: ドラマにおける登場人物関係図の例

れている。本研究では、筆者らが開発した、名詞句の末尾の名詞に着目して句の語義属性を解析する比較的単純な手法[2]を採用した。この手法は、日本語において一般に、名詞句の末尾の名詞により句全体の語義が決定される傾向が強い、という性質を利用している。以下にその例を示す。

小泉首相	=	小泉+首相 _{人物}	→	人物
小泉町	=	小泉+町 _{場所・組織}	→	場所・組織

このような語義決定を行うために、名詞句の末尾に存在して人物を表すような名詞を「RWCテキストコーパス」[10]と「京都大学テキストコーパス」[11]から一定の条件に従って語を抽出し、さらに「『分類語彙表』形式による語彙分類表(増補版)」[13]を用いて語彙の拡張を行い、辞書を作成した[2]。

以上の手法を用いて、以下の手順で人物名の検出を行う。

1. テキストを形態素解析
日本語形態素解析システム JUMAN[12]を利用
2. 名詞列(名詞句のうち名詞が連続するもの)を検出
3. 固有名詞で始まるものを選別
4. 末尾の名詞に着目して語義属性を解析

5. 属性が人物であると判断されたものを検出

3.3 話題境界の検出

話題境界の検出は、ニュース映像の画像的構成を利用して行う。一般に、ニュース映像は図5に示すような構成をとり、話題の冒頭にはキャストが登場する「スタジオショット」が出現する。ここでは、この性質を利用して、スタジオショットを画像的に検出することにより、話題境界の検出を行う。

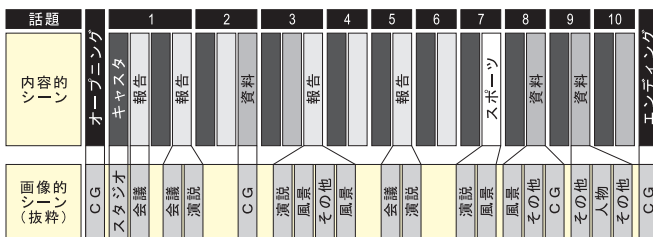


図5: ニュース映像の構成の例

スタジオショットの特徴は、以下の通りである。

- 大きめの顔領域（キャストの顔）が一つ存在
- 背景領域（スタジオ）の色彩的特徴の変動は少ない

このような特徴を考え、以下の手順でスタジオショットの検出を行う。

1. 大きめの顔領域が一つ存在するショット（人物ショット）を検出
2. 人物ショットの代表フレーム（ショットの先頭フレームを採用）の色ヒストグラムを特徴量ベクトルとみなし、特徴量ベクトル空間中でもっとも稠密なクラスターを形成するものをスタジオショットと判断

しかし、このようにして検出したスタジオショットは、話題の途中に出現することもある。そこで、話題の冒頭のスタジオショットのもう一つの特徴として、以下の特徴を加味する。

- タイトル字幕が出現

タイトル字幕の出現を検出するために、さらに以下の手順を付け加える。

3. フレーム全体のエッジ検出を行い、エッジ強度の変動を検出

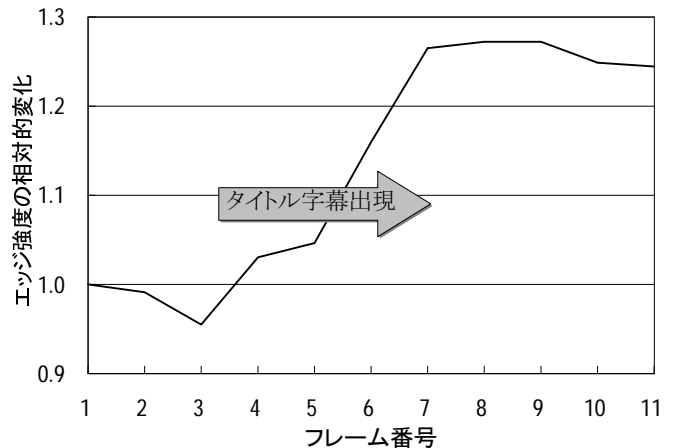


図6: タイトル字幕出現に伴うエッジ強度の相対的变化

これは、タイトル字幕の出現前後で、エッジ強度が図6に示すように相対的に大きく変化することを利用して。

以上の手順に従い、75分間（計468ショット）のニュース映像に対して検出実験を行った[3]。この実験で、顔領域検出には修正HSI表色系[4]を利用した肌色領域抽出とテンプレートマッチングを利用した手法を用いた。また、カット検出は正しく行われたものとした。その結果、表1に示すように良好な結果が得られた。なお、スタジオショット（冒頭）における検出洩れは、話題の冒頭にもかかわらず、タイトル字幕が出現しなかった場合である。

表1: スタジオショット検出実験の結果

ショット	正検出	誤検出	検出洩れ	再現率	適合率
スタジオ	53	0	0	100%	100%
川（冒頭）	37	0	6	86%	100%

3.4 人物関係の抽出

人物関係の抽出は、3.2で検出した人物名の共起を、3.3で検出した話題内において数え、全話題に対して集計することにより行う。このようにして抽出した人物関係は、図4に示したドラマにおける登場人物相関に示されるような人物関係の意味付け（「親友」「片思い」など）を行うまでにはいたらないが、どのような話題で共起しているかによって関係を大雑把に推定することはできるだろう。

3.5 検索・閲覧インタフェース

以上の手順に従って抽出されたニュース映像中の登場人物の人物関係を手がかりに映像を検索・閲覧するイン

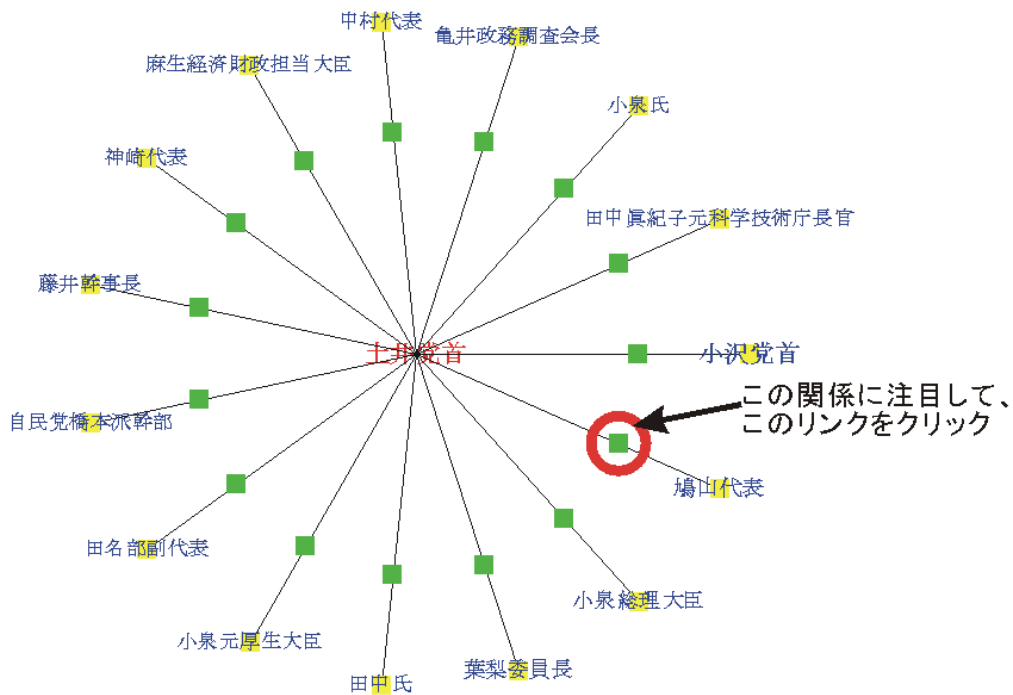


図 7: 人物関係の可視化：人物関係閲覧インターフェース

タフェースを実装し、提供する。図 7 に人物関係を可視化した人物関係の閲覧インターフェースを示す。ここでは、中央に存在する人物に注目して、円周上に配置した人物との一階層の人物関係を可視化している。人物関係をたどるためのインターフェースとして、円周上に配置した人物名の中央にある四角形をクリックすることにより、注目する人物を切り替えることができる。また、関係の強さ（同一話題内での共起頻度の多さ）に比例して人物名を大きく表示し、右回りに順に並べてある。

次に、図 7 の人物関係のうち、特定のものに注目して映像を閲覧するためのインターフェースを図 8 に示す。この映像閲覧インターフェースは、図 7 の人物関係閲覧インターフェースの人物間のリンク（中心の人物と円周上の人物を接続するスポーク状の線分）上にある四角形（この場合は図 7 中の○で囲まれた四角形）をクリックすることにより表示される。このインターフェースにより、注目する人物関係が得られた映像を閲覧できるだけでなく、それがどのような人物関係であったかを動画像および文字放送字幕の情報から得ることができる。表示されている代表画像は、各々の話題において円周上の人物名（この例の場合は「鳩山代表」）が文字放送字幕中に出現した時刻のフレームである。



図 8: 映像閲覧インターフェース：図 7 中の○で囲まれたリンクをクリックして表示される結果を例示

4 実験

前章で述べた手順および手法に基づき、実際のニュース映像から人物関係を抽出し、検索・閲覧インタフェースを作成した。実験には、平成13年3月16日から5月31日の期間内の53日分の特定のニュース映像（毎日30分）を用いた。この映像群に含まれる文字放送字幕は合計2,807文、話題数は合計560であった。また、今回の実験では、話題境界の検出はほぼ正確に自動化できることをふまえ、3.3で述べた手法を用いずに、人手で切り分けた結果を利用した。その結果、のべ2,799件の共起、異なり数292件の人物関係が得られた。

得られた人物関係を見ると、表記が異なる同一人物の共起も人物関係とみなして抽出されていた。これには単純に同義語による異表記（「小泉首相」と「小泉総理大臣」など）のほか、同義語ではないが意味的に同義である異表記（「小泉氏」と「小泉元厚相」など）場合も含まれる。このような異表記の名寄せを行うことは膨大な背景知識を必要とするため困難であるが、これも人物関係の一種（等価（is-a）な関係）とみなすことにより、検索の際の問合せ語の表記的・意味的多様性を吸収することができると思われる。

5 おわりに

本稿では、大規模ニュース映像コーパスから、話題中の共起を手がかりとして人物関係を抽出し、それを索引とした映像検索・閲覧手法を提案し、実装した結果を紹介した。

今後は、より大規模な映像コーパスを対象にし、さらに画像中の人物の顔の利用方法などの検討を進め、より映像内容に則した検索・閲覧手法の実現を目指す。また、共起が生じた話題の種類（政治、事件など）を判別することにより、人物関係の大雑把な意味付け（人物間のリンクの意味）と可視化についても検討していきたい。

参考文献

- [1] 安藤彰男, 今井 亨, 小林彰夫, 本間真一, 後藤 淳, 清山信正, 三島 剛, 小早川健, 佐藤庄衛, 尾上和穂, 世木寛之, 今井 篤, 松井 淳, 中村 章, 田中英輝, 都木 徹, 宮坂栄一, 磯野春雄: “音声認識を利用した放送用ニュース字幕製作システム”, 信学論, vol.J84-D-II, no.6, pp.877-887 (June 2001).
- [2] 井手一郎, 田中英彦: “末尾の名詞に着目したテレビニュース字幕の語義解析”, 情処学論, vol.39, no.8,

pp.2543-2546 (Aug. 1998).

- [3] 井手一郎, 山本晃司, 浜田玲子, 田中英彦: “ショット分類に基づく映像への自動的索引付け手法”, 信学論, vol.J82-D-II, no.10, pp.1543-1551 (Oct. 1999).
- [4] 松橋 聡, 藤本研司, 中村 納, 南 敏: “顔領域抽出に有効な修正 HSV 表示系の提案”, テレビ誌, vol.49, no.6, pp.787-797 (June 1995)
- [5] 佐藤真一, 于 川, 孟 洋, 片山紀生, 浜田 喬: “映像インデクシングの実証的研究のための映像データベースシステムの試作”, 2001年信学総大SD-5-9, vol. 情報システム2, pp.367-368 (March 2001).
- [6] S. Satoh, Y. Nakamura, and T. Kanade: “Name-It: Naming and detecting faces in news videos”, *IEEE Multimedia*, vol.6, no.1, pp.22-35 (March 1999).
- [7] 佐藤 隆, 児島治彦, 阿久津明人, 外村佳伸: “映像コーパスの構築と分析”, 信学論, vol.J82-D-II, no.10, pp.1552-1560 (Oct. 1999).
- [8] B. M. Sundheim: “Named entity task definition, version 2.1”, *Proc. Sixth Message Understanding Conf. (MUC-6)*, pp.317-332 (Nov. 1995).
- [9] “付録H: NE定義”, IREX ワークショップ予稿集, pp.264-273 (Sep. 1999).
- [10] 技術研究組合新情報処理開発機構 (RWCP): “RWCテキストデータベース第1版”.
- [11] 京都大学大学院情報学研究科知能情報学専攻言語メディア研究室: “京都大学テキストコーパス第2.0版”, <http://www-lab25.kuee.kyoto-u.ac.jp/nl-resource/corpus.html> より入手.
- [12] 京都大学大学院情報学研究科知能情報学専攻言語メディア研究室: “日本語形態素解析システムJUMAN第3.6版”, <http://www-lab25.kuee.kyoto-u.ac.jp/nl-resource/juman.html> より入手.
- [13] 国立国語研究所: “「分類語彙表」形式による語彙分類表 (増補版) [増補モニタ用電子化データ]”.