トピックに基づく大規模ニュース映像コーパスの構造化

Topic-Based Structuring of a Large-Scale News Video Corpus

井手 一郎† 孟 洋† 片山 紀生† 佐藤 真一†
Ichiro Ide Hiroshi Mo Norio Katayama Shin'ichi Satoh

1. はじめに

情報通信技術の進歩に伴い，さまざまな媒体を通じて日々大量の映像が放送されている．なかでもニュース映像は，人類共通の貴重な文化的・社会的記録とみなせ，整理して蓄積し，検索や再利用に供する価値がある．しかし従来は，記憶媒体の容量や計算機の処理能力の限界から，大規模に電子的に蓄積されることは少なく，そのような映像を対象とした効率的な解析・検索手法についても，十分に検討されていない．このような背景をふまえ，我々は大規模ニュース映像コーパスを作成するために，映像（動画像・主／副音声・文字放送字幕）自動蓄積装置を構築した．本発表では，この装置により蓄積した数 100 時間規模のニュース映像群に対する意味内容に基づく知的構造化手法と，構造化された映像を効率的に検索・閲覧するためのインタフェースについて紹介する．

2. ニュース映像コーパスの構造化

2.1 ニュース映像の構造

図 1 左に示すように，1 つのニュース映像はトピックを単位として意味内容に基づいて構造化される．従来のニュース映像の知的構造化に関する研究は，このような意味的なまとまりを抽出するために，1 つの映像内の構造を解明することに主眼をおいていた．しかしニュース映像は，意味的に連続した内容を日々報じるため，特に大規模なニュース映像群を扱う際には，図 1 右に示すように，映像間の関連する話題（トピック）のつながりを解明することに大きな意味がある．このように一連のトピックのつながり（スレッド）を解明することで，単一のトピックからは得られない，関連性そのものがもつ重要な情報が得られるようになることが期待される．

このようなトピックを単位とする映像の構造化は，米国 NIST が定義する “Topic Detection and Tracking（TDT） task”[1] の一種であり，テキスト情報に基づくトピックの分割・追跡に関する様々な研究がなされている．しかし，映像を扱ううえではテキスト情報の解析のみでは不十分であり，画像情報の解析も含めて考える必要がある．そこで，本研究では図 2 に示すような統合メディア処理によるトピック分割・追跡手法を提案する．まず，映像内構造化としてトピックに分割し，次に，映像間構造化としてトピック間の関連性を評価する．その後，時系列に関連するトピックを追跡してスレッド構造を解明する．なお，本発表ではその第一段階として，テキスト情報（文字放送字幕の形で提供される主音声の書き下しテキスト）のみによる解析結果を示すが，今後画像情報の解析結果も順次導入していく．

†国立情報学研究所
National Institute of Informatics, Japan

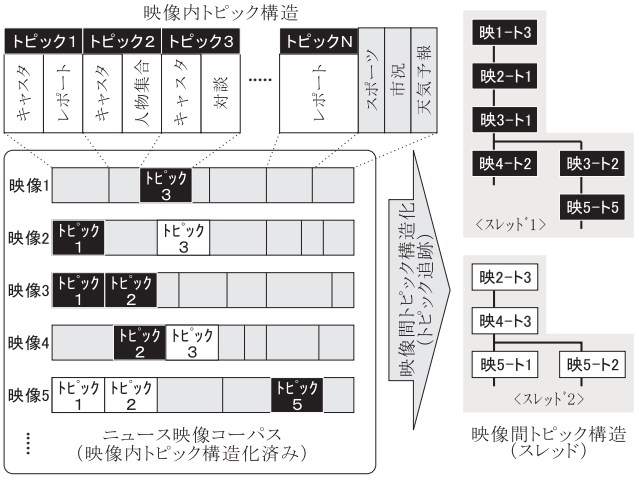

図 1: トピックを単位とするニュース映像の構造

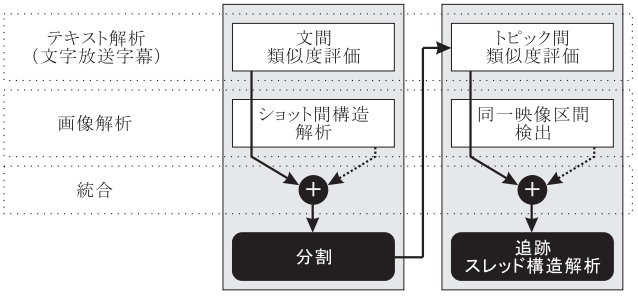

図 2: 統合メディア処理によるトピック構造の解析手法

2.2 映像内の構造化：トピック分割

以下の手順でトピックを分割する．

1. 文字放送字幕を形態素解析（日本語形態素解析システム JUMAN[2] を使用）して名詞列を抽出
2. 名詞列の語義属性（一般，人物，場所・組織，時相）を解析し [3]，文単位に属性別の出現頻度付きキーワードベクトル $\vec{k_g}, \vec{k_p}, \vec{k_l}, \vec{k_t}$ を作成
3. 各文の境界で，前後 w 文のキーワードベクトルの類似度を評価
   文 i と文 $i+1$ の境界における類似度を次のように定義する：

$$R_{S,w}(i) = \frac{\sum_{m=i-w+1}^{i} \vec{k_S}(m) \cdot \sum_{n=i+1}^{i+w} \vec{k_S}(n)}{\left|\sum_{m=i-w+1}^{i} \vec{k_S}(m)\right| \left|\sum_{n=i+1}^{i+w} \vec{k_S}(n)\right|}$$
$$(i = w, w+1, ..., i_{max} - w)$$

ここで，$S = \{g, p, l, t\}$ で， i_{max} はある 1 本の映像中の文字放送字幕テキストの総文数である．また以下では，$w = 1, 2, ..., 10$ とした．

4. 各文の境界で，次式により最終的な類似度を評価

$$R(i) = \sum_{S=\{g,p,l,t\}} a_S \max_w R_{S,w}(i)$$

まず，窓幅の大小によって以下の傾向が観察されるため，両者の短所を補うべく，各窓幅で評価した類似度の最大値を求める．これらの傾向は，同一トピック内でも文単位ではトピックを特徴付けるような語の出現のばらつきが大きいことによる．

- $w =$ 小：長いトピックを過分割する傾向
- $w =$ 大：短いトピックを過分割する傾向

次に，属性別の類似度の加重和を以下のように求め，閾値 θ_{seg} を越えた箇所をトピック境界と判定する．これは，特にニュース映像において，属性によりキーワードの重要性が異なることを想定したためである．ここで重みの値は，39 日分の訓練用文字放送字幕テキストに対して $\max_w R_{S,w}(i)$ を求め，人手で与えたトピック境界と照合し，重回帰分析により次のように定めた：

$$(a_g, a_p, a_l, a_t) = (0.23, 0.21, 0.48, 0.08) \quad (1)$$

5. トピック毎にキーワードベクトル $\vec{K_S}$ を作り，隣接するトピック間の類似度を下式により評価し，閾値 θ_{cat} を越える場合に再結合

$$R(i,j) = \sum_{S=\{g,p,l,t\}} a_S \frac{\vec{K_S}(i) \cdot \vec{K_S}(j)}{\left|\vec{K_S}(i)\right| \left|\vec{K_S}(j)\right|} \quad (2)$$

これは，過分割トピックを減らすための措置であり，再結合が生じなくなるまで，再帰的に繰り返す．

以上の手順を，手順 4. で用いたのと同じ訓練用データに適用し，検出結果の F 値が最大となるように，閾値を $\theta_{seg} = 0.28$，$\theta_{cat} = 0.08$ とした．

以上の条件で 2001 年 3 月 16 日より 2003 年 4 月 25 日までの正味 684 日分（全 85,123 文）のトピック分割を行なった結果，2 文以上からなるトピックが 9,592 件抽出された．なお，1 文からなるもの（15,110 件存在）は過分割の結果の雑音であることが多いために除外した．

表 1: トピック抽出性能

	両端一致	片端許容	両端許容
再現率	30.0%	34.6%	95.4%
適合率	28.5%	32.8%	90.5%

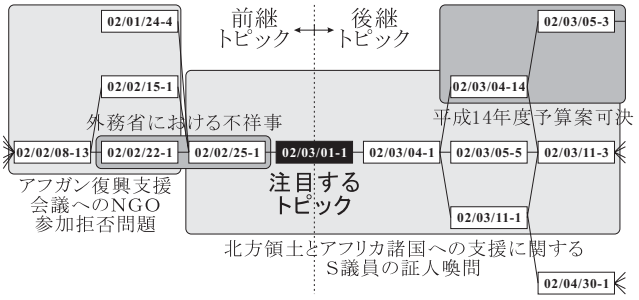

図 3: 解明されたスレッド構造の例

一方，14 日分の評価用文字放送字幕文（全 983 文，130 トピック）に同様に上記手法を適用し，人手で与えた正解と照合してトピック単位の抽出性能を評価した．表 1 にその結果を示す．表中,「両端一致」は始端と終端を正しく検出したもの,「一端許容」はいずれか一端において,「両端許容」は両端において正解と 1 文以内のずれを許した場合の性能を示す．本研究で目指すトピックの大局的な流れの解明のためには，両端 1 文程度の誤差は許容できるため，良好な抽出性能が得られたことになる．

2.3 映像間の構造化：トピック追跡

次に，分割した各トピックに対し，他のトピックとの関連性を式（2）で定義した類似度 $R(i,j)$ で評価し，閾値 θ_{trk} を越えると関連するトピックとみなす．ここで属性別の重みは，式（1）と同じに定めた．次に，以下の手順を再帰的に繰返すことで，関連するトピック間を階層的に構造化し，スレッド構造を解明する．

1. あるトピック（親）に関連する（$R(i,j) > \theta_{trk}$ を満たす）全てのトピック（子）をリンクする

2. ある子トピックに対して，兄弟やその子孫で関連するものがあれば，時系列的に最も近いものの子としてリンクし直す

なお，将来的に後述のインタフェースを通じて利用者の行動に基づく関連度の動的調整を考えており，その際には実時間処理が必要となることから，ここでは注目する話題を基準として前方・後方に連なる一方向の関連性のみを評価することで計算量の削減をはかる．図 3 に，実際に解明された 2002 年 3 月 1 日の 1 番目のトピックを起点としたスレッド構造と話題の推移を示す．

3. トピック検索・追跡インタフェース

以上のようにして分割されたトピックと，解明されたスレッド構造を反映した “Topic Browser” インタフェー

図 4: “Topic Finder” インタフェース

図 5: “Topic Tracker” インタフェース

スを構築した．このインタフェースは，クエリに対してトピック単位に検索する “Topic Finder”（図 4）と，指定したトピックに連なるスレッドを利用者が選択しながらたどる “Topic Tracker”（図 5）からなる．いずれも，選択したトピックの映像及びそれと同期した文字放送字幕を画面右側で再生して閲覧できる．

前者は，クエリと一致する語を含むトピックの一覧を時系列順に表示する．後者は，“Topic Finder” で一覧から選択したトピック，あるいは “Topic Tracker” で 1 試行前に選択したスレッドの先頭トピックを基準とした前継・後継スレッドの先頭トピックの詳細と各スレッド中の一連の前継・後継トピックの一部の先頭フレームを表示する．このようにスレッドの分岐点のみを提示することで，大量の関連トピックを効率的に絞り込んでいく．一方，トピック追跡の過程で，当初のクエリを拡張しているとも考えられる．このようなインタフェースにより，利用者の意図に最大限適合したトピックスレッドをたどれるようになることが期待される．

詳細な性能評価は将来の課題であるが，著者らが使用したかぎりにおいては，興味があるトピックに関する一連の顛末をたどって内容を把握するのに大変有効であった．

4. おわりに

本発表では，大規模ニュース映像コーパスに潜在する意味内容（トピック）に基づく構造を解明する手法を提案し，それを反映して実装した検索・閲覧インタフェースを紹介した．

今後は，各要素技術の性能向上のほか，図 2 に示した画像情報からの手がかりから得られる情報（ショット間構造解析および同一映像区間検出 [4]）も導入し，テキスト情報のみによる解析を補強するような統合メディア処理による構造化を実現する．また，“Topic Tracker” インタフェースにおいて，利用者の行動履歴に基づきトピック間の類似度を修正することで，スレッド構造を検索意図に沿って動的に再構築し，さまざまな検索要求に柔軟に対応できるようにする．

参考文献

[1] National Institute of Standards and Technology: “The 2002 topic detection and tracking (TDT2002) task definition and evaluation plan” （May 2002）.

[2] 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 3.61 版” （May 1999）.

[3] 井手一郎，浜田玲子，坂井修一，田中英彦：“テレビニュース字幕の語義属性解析のための辞書作成”，信学論（D-II），vol.J85-D-II, no.7, pp.1201–1210（July 2002）.

[4] Shin’ichi Satoh: “News video analysis based on identical shot detection”, *Proc. 2002 IEEE Intl. Conf. on Multimedia and Expo*, vol.1, pp.69–72（Aug. 2002）.