社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

大規模ニュース映像コーパスの意味構造解析

井手　一郎† 　孟　洋† 　片山　紀生† 　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
E-mail: †{ide,mo,katayama,satoh}@nii.ac.jp

あらまし　大規模ニュース映像コーパス内におけるトピックの連鎖（スレッド）構造を抽出する手法と，その構造に基づく閲覧インタフェースを提案する．ニュース映像コーパスは、規模が拡大するにつれ，トピック同士が相互に複雑に絡み合った，関連したトピックの集積としての本来の性質を示すようになる．このような関連性は，それ自体に重要な高次の意味情報が含まれるため，意味内容に基づく高度な閲覧インタフェースを実現するために，抽出することが重要となる．本研究では，このようなスレッド構造に基づいたインタフェースにより、利用者に提示するトピック数を必要最小限に抑えることで，大規模映像コーパスを効率的に閲覧できるようにすることを目標とする．
キーワード　マルチメディア情報検索，映像コーパス，トピック分割，トピック追跡，ユーザインタフェース

Anayzing the semantic structure of a large-scale news video corpus

Ichiro IDE†, Hiroshi MO†, Norio KATAYAMA†, and Shin'ichi SATOH†

† National Institute of Informatics　2–1–2 Hitotsubashi, Chiyoda-ku, Tokyo, 101–8430 Japan
E-mail: †{ide,mo,katayama,satoh}@nii.ac.jp

Abstract　We introduce a method to extract topic threads throughout a large-scale news video corpus as well as an interface that provides the users with the facility to browse through the corpus guided by the thread structure. Such guidance is essential to explore into the mass volume of video contents for thorough understanding of a topic developing through time, since when the size of a corpus expands, it is no longer a mere accumulation of independent stories, but a group of stories mutually related among themselves, where the relation itself provides rich semantic information.
Key words　Multimedia contents retrieval, Video corpus, Topic segmentation, Topic tracking, User interface

1. はじめに

近年の情報通信技術の発展に伴い，様々な媒体を通じて大量の映像が放送されるようになっている．放送映像は多岐にわたる人間の社会活動を記録しており，人類共通の文化的・社会的資産と考えられる．ニュース映像はその最たるものであるが，従来大規模なニュース映像コーパスにおいて，高次の意味情報を抽出する試みはほとんどなされていない．

このような問題意識のもと，我々は自動ニュース映像蓄積装置を構築し，重要なニュートピックを容易に検索し追跡するインタフェースの実現を目指している．この装置は映像として放送される動画像，音声，文字放送字幕（ディジタルデータとして提供される主音声の書下しテキスト）を自動的に取得し，現時点までに連日放送されるニュース番組を約 390 時間（MPEG-1 映像 242GB，MPEG-2 映像 1.47TB，文字放送字幕 17.4MB からなる）蓄積している．

本報告では，ニュース映像コーパス内におけるトピックの連鎖（スレッド）構造を抽出する手法と，その構造に基づく閲覧インタフェースを提案する．スレッド構造に基づき利用者に提示するトピック数を必要最小限に抑えることで，大規模映像コーパスを効率的に閲覧できるようにすることが目標である．

2. ニュース映像コーパスにおけるスレッド構造の抽出

ニュース映像コーパスは，規模が小さければ単に連日放送される映像の集積ととらえることもできるため，従来小規模の映像群を対象とた研究では，意味をもつ最小の映像単位を抽出するために，映像分割（映像内構造化）に主眼をおいたものが多かった．しかし，規模が拡大するにつれ，相互に複雑に絡み合う，関連したトピックの集積としての本来の性質を示すようになる．このような関連性は，それ自体に重要な高次の意味情報が含まれるため，意味内容に基づく高度な閲覧インタフェースを実現するために，抽出することが重要となる．図 1 に両構造の例を示す．図中左側は映像内構造を，右側は映像内構造の上

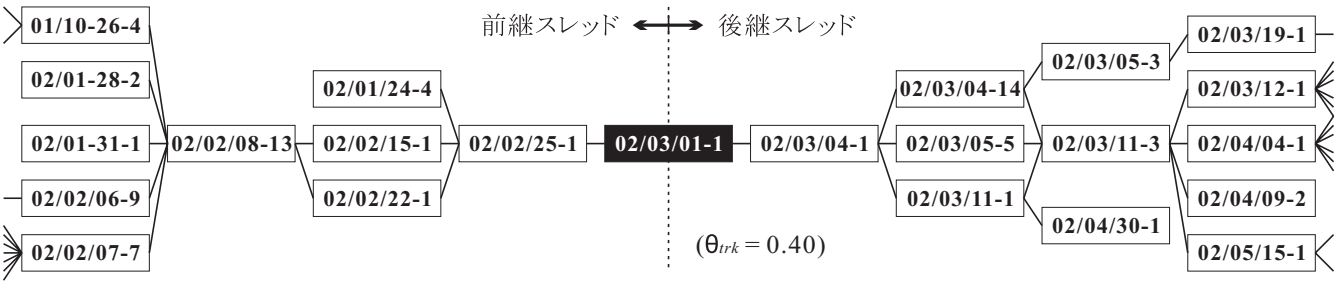

図 2　コーパスから抽出されたスレッド構造（部分）．各トピックの識別番号は次の書式になっている：「年/月/日-トピック番号」．

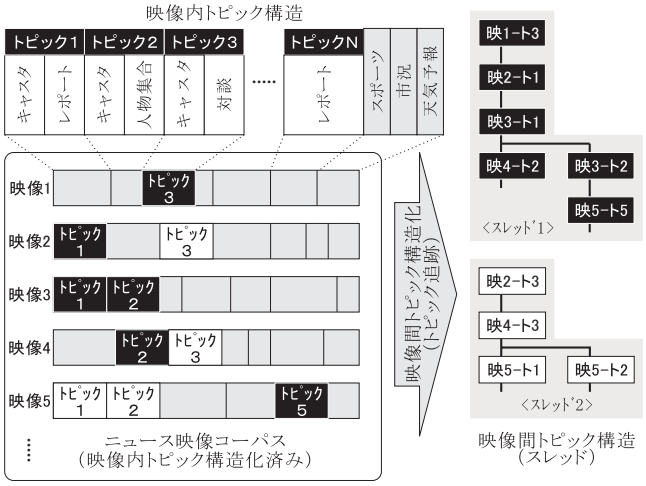

図 1　トピックに基づく映像内・映像間構造．

に立脚する映像間構造を示す．以下では，これらの構造の抽出と，関心のあるトピックを最小限の選択によりたどれるような映像間構造に基づくインタフェースを提案する．

スレッドの構築は，個別のトピック間の，1）意味的関連性，2）時間的関連性，の 2 つを満たしつつトピックを連鎖することにより行なう．ニュース映像の性質に基づき，スレッドの内容は時間とともに徐々に変遷していくため，いくつものスレッドが相互に合流・分岐を繰返し，その結果，コーパス全体は複雑に絡み合ったスレッド構造にからなることになる．図 2 に，あるトピック（この場合は 2002 年 3 月 1 日に放送された 1 番目のトピック）を起点として実際に抽出された前継・後継スレッド群を例示する．

スレッド構造は以下の 2 つの手順で抽出される：

（1）　トピック分割

（2）　トピック追跡・スレッド構築

トピック分割は，日々の映像を対象とし，文字放送字幕テキストにおける文間のキーワードベクトルの類似性に基づいて行なわれる（映像内構造化）．一方，トピック追跡・スレッド構築は，コーパス内全体にわたって，手順（1）で分割されたトピック間のキーワードベクトルの類似性に基づいて行なわれる（映像間構造化）．以下の各章で，まずこれらの手順を詳細に紹介し，次にスレッド構造に基づいて実現したユーザインタフェースを紹介する．

トピック分割・追跡は一般論としては米国 NIST により定義されている “Topic Detection and Tracking (TDT) task” [6] の一種であり，過去のワークショップや関連研究分野で様々な手法が提案され評価されてきた．一方，本研究と同規模のニュース映像コーパスを対象とした構造化，可視化，情報検索に関しては，文献 [5] や Informedia News-on-Demand プロジェクト [8] のようにいくつかのものがある．これらに対して本研究は，映像間にわたる高次の意味構造の解明を追求し，また徐々に遷移するトピックの意味内容を追跡するといった，スレッド構造の抽出手法において特徴的である．

3.　映像内構造化：トピック分割

映像内におけるトピックの境界を以下の手順で文字放送字幕から検出する．

（1）　文字放送字幕の各文に形態素解析（日本語形態素解析システム JUMAN [4] を使用）を適用し，名詞列を抽出する．

（2）　名詞列の語義属性（一般，人物，場所・組織，時相）を解析し，文毎に属性別の出現頻度付きキーワードベクトル $(\vec{k_g}, \vec{k_p}, \vec{k_l}, \vec{k_t})$ を作成する．語義属性は，末尾の名詞を属性別名詞辞書を照合して分類する手法 [2] により解析した．

（3）　窓幅 w を設定し，各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評価する．文 i と $i+1$ の間における類似度を次のように定義する：

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

ここで，$S = \{g, p, l, t\}$ であり，i_{max} は文字放送字幕中の全文数を表す．また，以下の実験では $w = 1, 2, ..., 10$ とした．

（4）　次の関数を評価し，様々な w における $R_{S,w}(i)$ の最大値をとる：

$$R_S(i) = \max_w R_{S,w}(i)$$

予備実験において，w に関係なくほとんどのトピック境界が正しく検出できるものの，多くの過分割が発生することが観察された．ここで，各窓幅における類似度の最大値をとることで，過分割を隠し合うことを期待する．これは，図 3 に例示するように，窓幅の大小によ次のような傾向が見られることによる：

- 窓幅 w が小さい時（点線）：

大量の過分割が生じるが，短いトピック内では著しく高い類似度を示す．

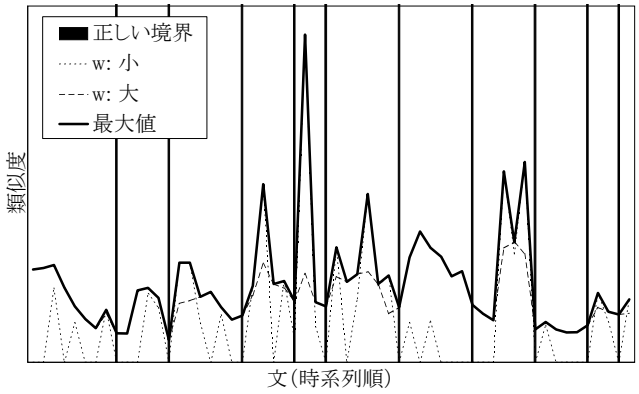

図 3　窓幅による過分割傾向.

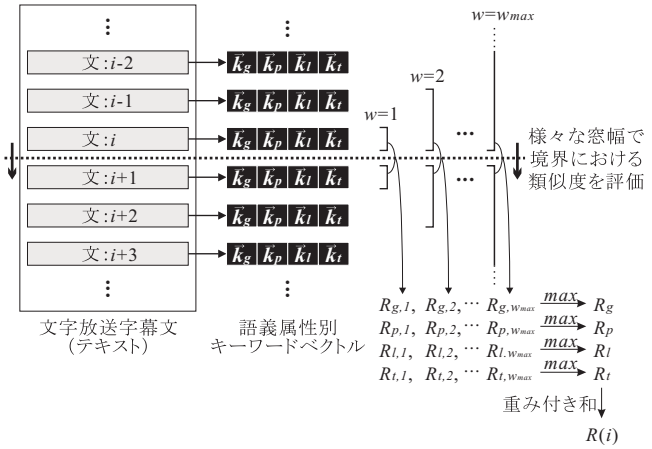

図 4　文間の類似度評価手順.

表 1　トピック分割性能.

評価条件	両端厳密	一端厳密	両端許容
再現率	30.0%	34.6%	95.4%
適合率	28.5%	32.8%	90.5%

- 窓幅 w が大きい時（破線）:

短いトピック内では著しく高い類似度は示さないが，長いトピック内では比較的高い類似度を維持する.

図中，実線は最大値をとった結果を示し，上式の効果が示されている.

（5） 次の関数を評価し，各語義属性別の類似度を重み付き和の形で統合する：

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

属性別に異なる重みを与えるのは，特にニュース映像に付随する文字放送字幕テキストにおいて，トピック分割を考える際に特定の属性が他の属性よりも重要な役割を果たすと考えたためである.

人手でトピック境界を与えた訓練事例（39 日分，合計 384 の境界をもつ文字放送字幕テキストからなる）に対して重回帰分析を施すことで，次の重みを得た：

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

最後に，$R(i)$ がある閾値 θ_{seg} を下回る場合に，文 i と $i+1$ の間にトピックの境界を検出する.

図 4 に手順（2）から（5）の処理を図示する.

（6） 過分割された断片を再結合するため，分割された各トピックに対してキーワードベクトル $\vec{K_S}$ を作り，次の関数を用いて隣接するトピック i と $j(=i+1)$ の類似度を評価する：

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

a_S としては，式 1 を用いた.

$R(i,j)$ がある閾値 θ_{cat} を上回ればトピック i と j を結合し，再結合が起きなくなるまで再帰的に処理を繰り返す.

まず，適当な閾値を決めるために，手順（4）で用いたのと同じ訓練事例に対して以上の手順を適用し，その結果，再現率と適合率が共に良くなるような閾値として，$\theta_{seg} = 0.28, \theta_{cat} = 0.08)$ を得た．この閾値を用いて，2001 年 3 月 16 日から 2003 年 7 月 31 日までのコーパス全体（のべ 774 日，97,591 文からなる）に以上の手順を適用したところ，2 文以上からなるトピックが 11,089 件抽出された．ここで，1 文からなるトピック（のべ 24,785 件存在した）は，過分割による断片であることが多く，それらはトピックとして扱うほどの情報を含まないため，除外した.

人手でトピック境界を与えた評価事例（14 日分，合計 130 トピックからなり，上で用いた訓練事例とは異なるもの）を用いて，分割の結果得られたトピックの抽出性能を評価したところ，表 1 に示しような結果を得た.「厳密」な評価条件では，境界が完全に一致している場合に正解とし,「許容」する評価条件では，境界が前後 1 文以内で一致している場合に正解とした．トピックの境界周辺の文は短かく，核となる情報を含まないことが多いため，本研究の目的のためには，1 文程度の誤差は許容できる．そこで，提案手法により現実的なトピック分割性能が得られたと考える.

4.　映像間構造化：トピック追跡・スレッド構築

次に追跡のために，式 2 を用いて，分割された全てのトピック相互間の類似度を評価する．将来は，式中の重み a_S を動的に調整することで利用者の意図を反映した追跡を実現する予定だが，現時点では式 1 で定めた値をそのまま用いた．トピック i と j の類似度 $R(i,j)$ がある閾値 θ_{trk} を上回れば，これらのトピックは強く関連しているとみなす.

このようにして強く関連すると判断されたトピック対を連鎖することでスレッドを構築する．スレッド構築の目的は，強く関連したトピックを時系列に連鎖し，利用者が関心のあるトピックの顛末を理解できるような道筋を示すことである．利用者が大規模な映像コーパスに分け入る際には，沢山の「似た」トピックの中から関心のあるトピックを追跡することが大きな負担になるため，道を選ぶ際の選択肢の数を最低限に抑えることが重要である．そのため，以下で提案するスレッド構築手法は，いずれ必然的にたどり着くようなトピックや結果的に選ば

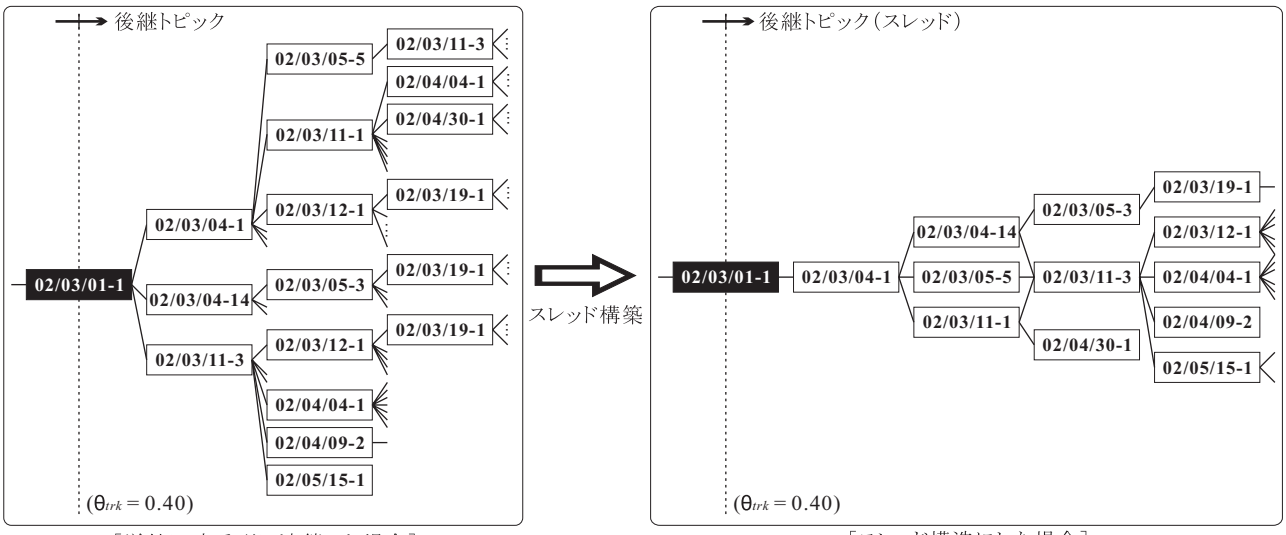

図 5　スレッド構築前後のトピック構造の変化．ここでは図 2 の右半分と同じ実例を示している．

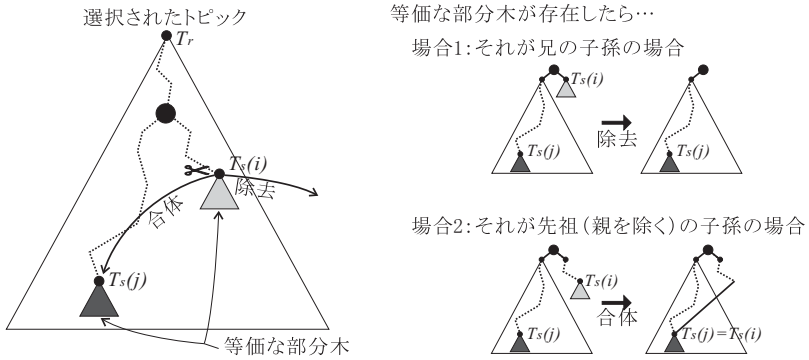

図 6　トピックスレッド構築機構.

れることがないかもしれないトピックをなるべく手前では隠すために，後続のトピックと関連するトピックは可能な限り後回しにするという条件を満たすように設計されている．図 5 に強く関連するトピック間の構造を，スレッド構築前後に分けて例示する．スレッド構築前は各トピックにおいて強く関連するものの候補として，多数の分岐（選択肢）が見られるが，スレッド構築により比較的少数に抑えられることが分かる．

あるトピック（T_r）から派生するスレッド構造を以下のアルゴリズムで構築する：

（1） 次の条件を満たしながら，T_r を根とする木を展開する：

（a） 子供は親と強く関連し，必ず時系列的に親よりも新しいトピックになる．

（b） 兄弟は必ず若い方（右側）が時系列的に新しいトピックになる．

（2） 次に，T_r を根とする木の中の全ての部分木 $T_s(i)$ に対して，左側に等価な部分木 $T_s(j)$ が存在するとき，次の操作を施す：

（a） $T_s(j)$ が $T_s(i)$ の兄の子孫であるときは，$T_s(i)$ を除去する．

（b） $T_s(j)$ が $T_s(i)$ の先祖（親を除く）の子孫であるときは，$T_s(i)$ を $T_s(j)$ と合体する．

ここで，(a) において，合体する代わりに除去しているのは，経路上にトピックが存在せず，経由することに意味がない短絡路をスレッド構造内に作るのを防ぐためである．以上の，除去・合体の仕組みを図 6 に示す．

なお，ここでは後継スレッド構造の構築の際のアルゴリズムを示したが，前継構造の場合は，「新しい」を「古い」に置き換えて実現する．

上記のアルゴリズムでは計算時間がかかるため，実装上は以下のような工夫をしている：

（1） 手順（1）で木を展開している最中に同一トピックを検出すると，随時手順（2）を実行して枝刈りをしておく．

（2） 木の展開をある深さ N_{trk} で中断する．

工夫（2）の結果，得られるスレッド構造は近似解となるが，N_{trk} をある程度深く設定しておけば，次章で紹介する追跡インタフェースのように，1 段階先の分岐しか利用しない場合は十分であると考える．

5. トピックスレッドを反映した映像閲覧インタフェース

得られたスレッド構造に基づき、図 7 に示すような映像閲覧インタフェース “Topic Browser” を実装した。インタフェースの左側は、“Topic Finder” と “Topic Tracker” の 2 つのイン

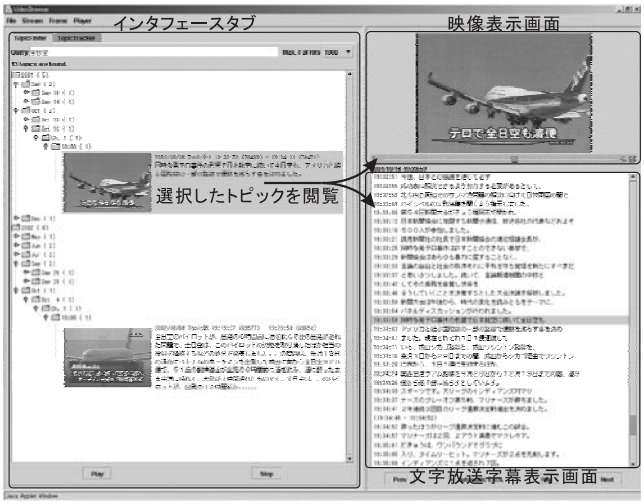

図 7　“Topic Browser” インタフェース.

図 8　“Topic Finder” インタフェース.「ビンラディン」を検索した結果.

タフェースのタブを切り替えられるようになっている。一方、右側は、共通の映像閲覧インタフェースとして、左側のインタフェースで選択したトピックに対応する映像および連動して文字放送字幕を再生できるようになっている。

“Topic Finder”（図 8）は本インタフェースの入口である。利用者が検索語句を入力すると、その語句が含まれるトピックの一覧を時系列順に列挙する。各トピックは放送日時などと併せて、サムネイル画像（トピックに対応する動画像の先頭フレーム）と文字放送字幕の一部（先頭数百文字）により表現される。利用者はこれらのトピック中から関心のあるものを選び、以降の追跡過程に入るための始点として指定する。

図 9　“Topic Tracker” インタフェース.［前継トピック］スレッド 1：同時多発テロの発生.［後継トピック］スレッド 1：攻撃に関する捜査；スレッド 2：外務省による安全情報の発出；スレッド 3：エジプトにおけるラマダン入り.

次に、“Topic Tracker”（図 9）はトピックをインタラクティブにたどるためのインタフェースである。始点となるトピックは “Topic Finder” を通じて選ばなければならないものの、それ以降の追跡過程は基本的にこのインタフェース内で行なわれる。このインタフェースでは、選択されたトピックから派生するスレッドが時系列的に過去に向かうもの（前継）と未来に向かうもの（後継）に分けて表示される。各スレッドは、選択されたトピックの子供のトピックが代表として文字放送字幕付きで大きく表示され、その下にスレッド内で以降に続くトピックの一部が小さなサムネイルとして列挙されている。代表として表示されるトピックは、スレッド構造における分岐点であるため、これらの中から次のトピックを選ぶことにより、利用者が関心のあるスレッドを絞り込んで追跡していくことになる。前継スレッドと後継スレッドを選んで追跡できるため、関心のあるトピックが成立する過程を知りたい場合、逆にその後の展開を知りたい場合、というように目的に合わせた追跡ができるようになっている。

このように次々にスレッドを選択していくことで追跡が進んでいく。また、選択する度にスレッド構造をその場で構築するようになっているため、将来は利用者の選択履歴に基づき追跡意図を推定することで、動的にスレッド構造を最適化して提示することを考えている。図 9 に、図 8 の例で選択したトピックから派生するスレッドの一覧を示す。各スレッドが同時多発テロ事件から派生し、それに関連するトピックではあるものの、

個別には別の流れのトピックに分岐していることが分かる。

“Topic Finder” インタフェースは既存のキーワードによるトピック検索の域を出ないものの、“Topic Tracker” との組み合わせにより、利用者の関心に沿って検索を絞り込む機能を提供することから、大規模ニュース映像コーパスに適した検索インタフェースと考える。また、関心に沿って検索を絞り込む一方で、徐々に変化するトピックを追跡する過程は一種の問合せ拡張ともみなせる。このような一見相反する特徴を備えることにより、提案手法による追跡過程は、利用者の関心に沿って、無駄なくかつ最大限の情報を含むスレッドを提示するように設計されている。さらに、スレッドの変遷・分岐・合流を意識した閲覧ができるため、トピックの総合的な理解にも貢献することを期待している。著者らは試行を通じて，以上に挙げた特徴が関心のあるトピックを追跡するのに有効であることを確認した。

文献 [1] でも類似した試みがなされているが、特定の意味属性に特化した閲覧インタフェースであることや、徐々に遷移する内容をまとまりとしてとらえようとはしていない点で本研究と異なる。

6. おわりに

本報告では、大規模ニュース映像コーパス中を網羅する意味構造を明らかにするために、トピックのスレッド構造を抽出する手法を提案した。まず、文字放送字幕テキストを用いたトピック分割・追跡・スレッド構築手法を紹介した。次に、スレッド構造に基づき、映像コーパス内を閲覧するインタフェースを紹介した。詳細な評価は今後の課題であるが、試行を通じて、関心のあるトピックを絞り込みながら追跡することの有効性を確認した。

今後は、分割・追跡・スレッド構築に画像特徴を導入することで、キーワード不足などテキストだけでは情報が不足する部分を補うことを目指す [3]。具体的には、トピック分割では、キャスタショット検出による映像内構造化など、既存研究で効果が認められている手法を採用する。また、トピック追跡・スレッド構築では、ニュース映像において、配信源や撮影場所の制約から、類似トピックにおいて全く同じ映像が繰り返し放映される傾向を利用し、同一映像セグメント検出 [7] の技術を用いた映像間の類似度を採用する。

インタフェースにおいては、各スレッドを特徴付けるようなサムネイルやキーワードを提示することで、より的確に進むべき方向を選択できるようにすることを考えている。また、検索に用いた語句の語義属性や選択履歴に応じて、式 2 の重みを動的に調整するなどして、利用者の意図や関心に沿ってスレッド構造そのものを適応的に変えることも考えている。

7. 謝　　辞

本研究の一部は科学研究費補助金若手研究（B）課題番号 15700116 および，特定領域研究（C）（2）公募研究課題番号 15017285 の補助を受けて実施した.

文　　献

[1] M. G. Christel, A. G. Hauptmann, H. D. Wactler, and T. D. Ng. Collages as dynamic summaries for news video. In *Proc. 10th ACM Intl. Conf. on Multimedia*, pp.561–569, Dec. 2002.

[2] I. Ide, R. Hamada, S. Sakai, and H. Tanaka. Semantic analysis of television news captions referring to suffixes. In *Proc. 4th Intl. Workshop on Information Retrieval with Asian Languages*, pp.37–42, Nov. 1999.

[3] I. Ide, H. Mo, N. Katayama, and S. Satoh. Topic-based structuring of a very large-scale news video corpus. In *AAAI 2003 Spring Symposium on Intelligent Multimedia Knowledge Management*, Mar. 2003.

[4] Kyoto Univ. Japanese morphological analysis system JUMAN version 3.61, May 1999.

[5] A. Merlino, D. Morey, and M. Maybury. Broadcast news navigation using story segmentation. In *Proc. 5th ACM Intl. Conf. on Multimedia*, pp.381–391, Nov. 1997.

[6] National Institute of Standards and Technology. The 2002 topic detection and tracking (TDT2002) task definition and evaluation plan, May 2002.

[7] S. Satoh. News video analysis based on identical shot detection. In *Proc. 2002 IEEE Intl. Conf. on Multimedia and Expo*, vol.1, pp.69–72, Aug. 2002.

[8] H. D. Wactlar, M. G. Christel, Y. Gong, and A. G. Hauptmann. Lessons learned from building a Terabyte digital video library. *IEEE Computer*, vol.32, no.2, pp.66–73, Feb. 1999.