2008－DBS－146 (36)
2008／9／22

文字放送字幕を利用した番組話題分割手法のオンエア視聴時への適用

渡辺奈夕子† 鈴木 優† 真鍋 俊彦†

† (株) 東芝 研究開発センター 知識メディアラボラトリー
〒 212–8582 神奈川県川崎市幸区小向東芝町 1
E-mail: †{nayuko.watanabe,masaru1.suzuki,toshihiko.manabe}@toshiba.co.jp

あらまし TV 番組の話題毎の選択的視聴を実現するために，これまで文字放送字幕に出現する語彙の遷移を元に自動で番組を話題分割する手法が研究されてきた．しかし従来研究は録画済み番組を前提としており，番組全体の字幕が得られていない放送中の番組に対しては適用できなかった．本論文では，手がかり表現によるスコアリングを基準とし，語彙遷移を補助的に使うことでオンエア視聴時にも適用可能な手法を提案する．料理，グルメ，旅行，タウンに関する 17 番組を用いた評価により，オンエア視聴時でも従来手法の 92%の精度で分割可能であることを確認した．また評価を通じ，途中にトーク部分が入る番組など短期間の語彙遷移による補完が有効な番組の特徴が分かった．
キーワード TV 番組のオンエア視聴, 話題分割, クローズドキャプション

Topic Segmentation of On-Air TV Shows Using Closed Caption Streams

Nayuko WATANABE†, Masaru SUZUKI†, and Toshihiko MANABE†

† Knowledge Media Laboratory, Corporate Research & Development Center, Toshiba Corp.
1 Komukai-Toshiba-cho, Saiwai-ku, Kawasaki, Kanagawa, 212–8582 Japan
E-mail: †{nayuko.watanabe,masaru1.suzuki,toshihiko.manabe}@toshiba.co.jp

Abstract For viewing topical segments of TV shows selectively, topic segmentation methods based on vocabrary shift of closed captions (CC) have been proposed. Although these methods handle recorded TV shows, they cannot be applicable for on-air shows in which only partial CC information is available. We propose a new method available for both recorded and on-air shows by using cue phrases mainly and using vocabulary shift subsidiarily. According to an experiment using 17 TV clips, we confirmed proposed method's accuracy is 92% of our previous work for recorded shows. Moreover, we found features of TV shows in which topic segmentation based on short-term vocaburaly shift is effective.
Key words on-air TV show, topic segmentation, closed caption

1. はじめに

録画機器の大容量化に伴い，大量の TV 番組を録画することができるようになった．大量の番組を効率的に見るために，番組のシーン一覧の中から見たいシーンだけを選んで視聴する「つまみ見」や，番組を視聴しつつ見たくないシーンは飛ばす「飛ばし見」のような録画番組の話題毎の選択的視聴へのニーズが高まっている．更に録画機能を搭載した TV では，TV 番組視聴と同時にその番組を録画しておき，録画済み番組と同じように一時停止や巻き戻し等が可能になった．この結果今後は，オンエア視聴している番組を一部見逃したときにその話題だけ見直す，番組の途中で内容を振り返って見るなどの機能へのニーズが高まると考えられる．

このような番組視聴を実現するために，番組の中から特定のシーンを抽出する手法や，番組をいくつかのシーンに分割する話題分割手法が研究されてきた．特に話題分割は Allan らにより提唱された “Topic Detection and Tracking (TDT) task” [1] の 1 つであり，映像情報だけでなく音声やテキストを用いた様々な手法が提案されている．歓声，拍手等の音響を自動抽出することでスポーツ番組の試合の得点シーンを抽出する手法 [2] や類似ショットの登場傾向から対話区間を取り出すことによってニュース番組やバラエティ番組をコーナーに分割する手法 [3]，フレーム内の色分布や音声の解析によりシーン分割する手法 [4] 等が提案されている．映像や音声だけでなく，文字放送字幕や EPG（Electronic Program Guide）といったテキスト情報を用いた話題分割手法も提案されている．井出らによるニュース

をトピック毎に分割する手法 [5] は，文字放送字幕中に出現する名詞列の分割候補点前後の類似度を解析している．また新田らは，文字放送字幕に対しベイジアンネットワークを利用することにより，スポーツ番組のシーン分割を行っている [6]．更に新田らは映像によるシーン分割と文字放送字幕によるシーン分割を個別に行いその結果を対応付けることでそれぞれのシーンの意味内容の自動獲得を行っている．

ニュースの大規模データベースのような対象においては，話題分割の結果は効率的な番組視聴だけでなく情報編集，整理等にも役立つ．井出らは大規模ニュース映像をトピック分割し，トピック間のスレッド構造を閲覧，編纂するインタフェース mediaWalker を提案している [7]．

我々はさまざまなジャンルを含む TV 番組を対象に，録画済み番組の選択的視聴をユーザに提供することを目的とする文字放送字幕を利用した話題分割システム Pic-A-Topic の開発を進めている [8]．Pic-A-Topic ではニュース以外の様々なジャンルに適用するために文字放送字幕中の話題の転換を示す手がかり表現と，番組中での語彙の遷移に注目し，それぞれに基づく分割点候補を生成した後，正規化，統合して分割する．

しかし従来の Pic-A-Topic では番組全体の文字放送字幕がある状態で分割を行う必要があるため，途中までの字幕しかないオンエア視聴時には分割することができない．そこで本論文では従来の Pic-A-Topic の処理手順を再構成し，分割の手がかり表現によるスコアリングを基準とし，語彙遷移によるスコアリングによって補助的に分割の制御をすることで，現在放送中の番組を話題分割する手法を提案する．またこの提案手法と従来手法の比較と，手がかり表現，語彙遷移がどれだけ分割精度に影響を与えているかの評価を行う．一方放送中の番組を分割するには，今放送している時点の前後でどれくらい内容に違いがあるのかを調べる必要があり現在放送している点が分割点になるかどうかがわかるのは実際には分割点よりもいくらか時間が経過した後にならざるを得ない．本論文ではさらに，この判断のためにどのくらいの窓幅（分割点らしさのスコアを算出するために必要とする時間幅）が必要かの調査を行う．

以下では，まず従来の Pic-A-Topic の概要とその課題を説明し，その手法を放送中の番組の話題分割に適用するための提案手法について述べる．次に従来手法と提案手法の比較評価，提案手法を適用する際に問題となるスコアリングの窓幅とその分割精度についての評価と結果についての考察を行う．

2. 従来の番組話題分割とその課題

我々はさまざまなジャンルを含む TV 番組を対象に，録画済み番組の選択的視聴をユーザに提供することを目的とするシステム Pic-A-Topic の開発を進めている．Pic-A-Topic は文字放送字幕（クローズドキャプション）を利用し，複数の話題が含まれる番組を話題毎に分割する．ここでは従来開発されてきた録画番組向けの Pic-A-Topic と，オンエア視聴時での分割への適用における課題について述べる．

2.1 従来の Pic-A-Topic

図 1 に Pic-A-Topic のシステム構成を示す．システムは文字放送字幕及び EPG データを入力として受け取る．最初に分割点の候補を全文字放送字幕の各文のタイムスタンプとし，次に，各分割点候補の分割点らしさのスコアを「手がかり表現」と「語彙の遷移」に基づいて算出する．最後に分割点ごとにスコアを統合し，最終的な分割点を得る．

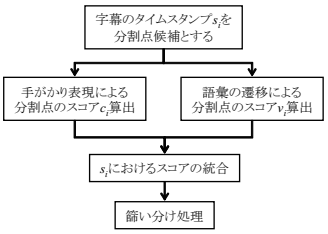

図 1　従来の Pic-A-Topic 構成図

2.1.1 手がかり表現

手がかり表現とは,「まずは」「次に」のような話題の転換を表す自然言語表現である．接続詞のみでなく，ランキング番組であれば各ランクを示す「第（数字表現）位」というパターンのように，どのような点で分割するかによりさまざまな手がかり表現が考えられる．

Pic-A-Topic では番組のジャンル毎に想定される手がかり表現を人手で作成し，それぞれの手がかり表現にスコアを付与している．i 番目に表示される字幕のタイムスタンプ s_i を分割点候補とする．s_i から始まる文字放送字幕が手がかり表現にマッチすれば，手がかり表現による分割点らしさのスコアに加算する．全分割点候補におけるスコアの最大値を用いて正規化し，最終的な手がかり表現によるスコア c_i が算出される．

2.1.2 語彙の遷移

分割点候補の前後にある窓幅を想定し，分割点候補よりも前の時間の窓（左側の窓）に出現するタームと後の時間の窓（右側の窓）に出現するタームとを比較して語彙の遷移度を算出する．これを語彙の遷移による分割点候補のスコアとする．ここで言うタームとは，人物，地名などの固有表現と形態素の双方である．通常のテキスト情報での話題分割 [9] と比較してタイムスタンプが文字放送字幕の大きな特徴となっているため，窓幅は文書量ではなく時間幅で与える．

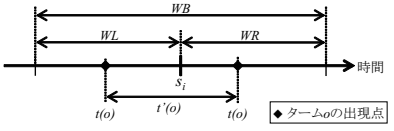

図 2　語彙の遷移度算出における窓の概念

図 2 に語彙遷移度を測る処理の模式図を示す．左側の窓を WL，右側の窓を WR，両方の窓を合わせた窓を WB としたとき，分割点候補 s_i におけるスコア v_i は，以下のように算出される．

窓 W 中に出現するターム o の集合を $O(W)$ とすると，

$O(WB)$ は，以下の 3 種類に分割できる．

- $O_{WL} = O(WB) \setminus O(WR)$
- $O_{WR} = O(WB) \setminus O(WL)$
- $O_{WB} = O(WL) \cap O(WR)$

O_{WL} は s_i よりも前 (WL) にのみ出現するターム，O_{WR} は s_i よりも後 (WR) にのみ出現するターム，O_{WB} は s_i の前後両方に出現するタームの集合である．この 3 種類のタームについて以下の式を用いることで，s_i におけるターム o の新規性 $n_i(o)$ を算出する．

$$n_i(o) = \begin{cases} \alpha \cdot dw(o) \left\{ 0.5 - 0.5cos\pi \left(\frac{s_i - t(o)}{WL\ の幅} + 1 \right) \right\} & o \in O_{WL} \\ \beta \cdot dw(o) \left\{ 0.5 - 0.5cos\pi \left(\frac{t(o) - s_i}{WR\ の幅} + 1 \right) \right\} & o \in O_{WR} \\ -\gamma \left\{ 0.5 - 0.5cos\pi \left(\frac{t'(o)}{WB\ の幅} + 1 \right) \right\} & o \in O_{WB} \end{cases}$$

$t(o)$ はターム o が出現するタイムスタンプ，$t'(o)$ は WR と WL における 2 つの $t(o)$ の差である．窓の中に複数回ターム o が出現する場合は，s_i に一番近い o のタイムスタンプを用いる．また，α，β，γ はいずれも正の定数，$dw(o)$ は番組ジャンルに応じて各タームに付けられる重みである．新規性 $n_i(o)$ は，ターム o が片側の窓にしか出現しない場合はその出現時刻が s_i に近いほど高く，両側の窓に出現する場合はその出現時刻の差が小さいほど低い．

分割点候補 s_i の新規性は，各ターム o の s_i における新規性 $n_i(o)$ の和で求められる．s_i における新規性を全分割点候補の新規性の最大値と最小値によって正規化したものを，s_i における語彙遷移のスコア v_i とする．

有効な手がかり表現や語彙遷移の新規性算出における適切な $dw(o)$ は番組によって大きく異なると考えられる．例えば旅行番組においては，場所の移動を表すような手がかり表現が有効であり，また語彙遷移においても場所の固有表現が重要である．そこで Pic-A-Topic ではいくつかの種類の手がかり表現辞書を作成しておき，分割時には番組を構成を表すジャンルに分類し手がかり辞書をそのジャンルにおいて有効な辞書へ切り替え，$dw(o)$ を重くするべき固有表現の種類を切り替えている．構成を表すジャンルは，現在は EPG に付与されているジャンルからルールベースでマッピングされる．

2.1.3 スコアの統合と篩い分け

上述したように算出された，手がかり表現によるスコア c_i と，語彙の遷移によるスコア v_i との重み付きの和が，分割点候補のスコアとなる．後処理として各分割点候補 s_i において，その前 T_{sL} 秒，後 T_{sR} 内の他の分割点候補に s_i よりもスコアが高いものがあれば，s_i のスコアを 0 とする．この操作を篩い分けと呼ぶ．これにより，間隔が $T_{sL} + T_{sR}$ 秒以上あいたスコアの高い点のみが分割点候補として残る．このスコアが高い順に N 個，もしくはある閾値 R 以上になっている点が分割点となる．

2.2 オンエア視聴時への適用における課題

TV 番組のオンエア視聴時には，見逃してしまった話題だけを視聴する，現在の番組のこれまでの話題の中でもう一度見直したいものを視聴する，等のニーズがある．また，番組の途中から視聴し始める場合には区切りの良いところから追っかけ再生をしたい，というニーズがある．このような場合に，巻き戻しで試行錯誤しつつ再生位置を探すのは非常に煩わしい．そのため番組の途中で分割結果を参照し，各話題の中身を確認して視聴したい部分を選択して話題の先頭から再生する機能は有効であると考えられる．

従来の Pic-A-Topic では録画済の番組を対象としていた．このため番組全体の文字放送字幕を利用できることが前提となっており，手がかり表現による分割点候補のスコアの正規化や語彙遷移のスコアの正規化時には全ての分割点候補のスコアが必要となる．

しかしオンエア視聴時には，視聴時点までの文字放送字幕しか利用することができない．このため番組全体の文字放送字幕を利用した正規化処理が不可能である．また，最終的な分割点を選択する際に，スコアの上位 N 件という選び方はできなくなる．番組によってスコアの分布は異なるため閾値との比較で分割点を選択すると，番組により分割点が多くなったり，少なくなったりしてしまう可能性がある．本研究の目的はこのような部分的な文字放送字幕のみでの高精度な話題分割である．

3. 話題分割のオンエア視聴時への適用

3.1 提案手法

2.2 で述べた課題を解決するため，従来の Pic-A-Topic の処理手順をオンエア視聴時にも適用できるように閾値を用いたルールベースの手順に変更した．提案手法では，**(1)** 語彙遷移による分割よりも効果が大きい手がかり表現による分割結果を優先し，**(2)** 手がかり表現による過分割を抑制し，また **(3)** 手がかり表現による分割が失敗して 1 つの話題の長さが長くなりすぎないよう考慮している．

(1) については，手がかり表現における分割点候補のスコアが閾値を越えるかどうかで分割の判断を行う．これは従来手法における手がかり表現での分割点と語彙の遷移での分割点を比較したところ，手がかり表現の方がより正解の分割点を再現できたからである．評価対象の 17 番組について再現率を比較すると，手がかり表現での分割点 (従来手法の手がかり表現でのスコア上位 n 件を採用したもの) が 46.0%，語彙の遷移での分割点 (従来手法の語彙遷移でのスコア上位 n 件を採用したもの) が 37.7%となり，手がかり表現の方が正解の分割点にヒットしやすいことがわかる．ここで n は，正解データの平均話題長で全て分割されたとしたときの対象番組での分割数である．また手がかり表現による分割点は，そこから再生した際にユーザに分割の理由が伝わりやすいという利点もある．

(2) 手がかり表現には「そして」等の一般的な接続詞も含まれ，手がかり表現が出現したからと言ってその点全てで分割を行うと過分割になってしまうという問題がある．そこで基本は手がかり表現による分割点を使用しつつ，語彙の遷移度によって過分割を抑制する．

(3) 現在手がかり表現は人手で辞書を作成しているため，辞書に登録されていない手がかり表現が出現すると精度が劣化する．このため手がかり表現が全く出現しない場合せず時間が経

過した場合は，語彙の遷移度によって分割を行う．

分割点の候補は，従来の Pic-A-Topic と同じく文字放送字幕の各行の開始タイムスタンプ s_i である．本手法においては，手がかり表現による分割を行うかどうかの閾値 C_1, C_2 $(C_1 > C_2)$，語彙遷移による分割を行うかどうかの閾値 V_1, V_2 $(V_1 > V_2)$ を使用する．現在注目している文字放送字幕開始のタイムスタンプ s_i が分割点になるかを判断するには，まず s_i における手がかり表現での分割点候補のスコア c_i と語彙の遷移による分割点候補のスコア v_i を算出する．そして，以下の条件に合致しないものは分割点の候補から外す．各条件は上述した **(1)**–**(3)** のアイディアと対応し，上から順番にチェックが行われる．

（1） c_i が C_1 よりも高い

（2） c_i が C_2 よりも高く，v_i が V_1 よりも高い

（3） 一定時間 T_b 秒分割点が無く，v_i が V_2 よりも高い

このように閾値を用いるため，手がかり表現によるスコアリング時の正規化処理は不要となる．このチェック後に，分割点候補の前 T_{sL} 秒，後 T_{sR} 秒内の他の分割点候補のスコアを見て篩い分けを行う．現状では $T_{sL} = WL, T_{sR} = WR$ と固定している．

3.2 窓幅と遅延

2.1 の従来の番組話題分割で述べた通り，語彙遷移によるスコアを算出するためには分割点候補の周りに幅を持った窓 (WL, WR) を想定し，その中で出現するタームを考慮する必要がある．また，後処理として行う篩い分けのときに $T_{sL} + T_{sR}$ の窓を想定する必要がある．これらの窓の幅 (長さ) が，オンエア視聴時の現在視聴時刻からの処理遅延となる (図 3)．

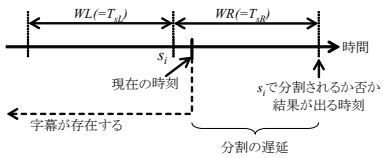

図 3 分割の遅延に影響を与える窓幅

分割精度を向上させるために最適な窓の幅は番組によって違うと思われるが，窓幅が大きいほど分割の遅延が大きくなるため，ユーザ側から見るとこの遅延はより小さい方が良い．そこで窓幅によって分割精度がどのように変わるかについて，4. の評価で述べる．

3.1 で述べた閾値 C_1, C_2, V_1, V_2 は番組によって最適な値が違う．現状では値を固定しているが，本来であれば手がかり表現の辞書や語彙遷移のタームの重みのようにジャンル等によって調整が必要であると思われる．

4. 評　　価

オンエア視聴時の精度評価，及び前述の窓の大きさ (時間幅) がどの程度分割精度に影響を与えるかの調査を行った．

4.1 実 験 設 定

4.1.1 対象番組と正解データ

話題分割の正解データは 17 番組について，各番組 2 人が作成した (表 1)．正解作成者は番組を視聴しながら，話題の切れ目だと感じた時点でタイムスタンプを記録する．なお，正解作成時に最終的な分割点数の上限・下限は指定していない．番組は従来 Pic-A-Topic [8] が主な対象としてきた料理・旅行番組を選択した．シリーズとは，同番組の異なる回を用いたことを表している．

シリーズ A は基本的には料理番組であるが，その合間に出演者たちによるある程度の長さのトークが行われる．B のジャンルは料理であるが，内容としては主に様々な料理を食べ歩き様々なスポットも訪れる内容となっており，料理に関係ないシーンや移動シーン等が含まれ，様々な人物が登場する．それに対して C は純粋にいくつかの料理の調理を行う内容となっている．A, C は調理後，料理の試食を行う．D, E, G は複数のグループが登場しそれぞれが違う場所を訪れて様々な観光場所やを紹介する旅行番組であるが，F は番組を通して 1 人の出演者が比較的狭い範囲の中で，その町の様子や店の紹介を行う内容である．C や F に共通して，番組の最後にその放送で紹介された料理や店のまとめを行っている．

番組 ID	シリーズ	ジャンル	番組長 (分)	平均分割数
1	A	料理	30	14.0
2	A	料理	30	15.0
3	A	料理	30	14.0
4	B	グルメ	54	9.0
5	B	グルメ	54	14.5
6	B	グルメ	54	8.0
7	C	料理	25	6.0
8	C	料理	25	4.0
9	C	料理	25	5.5
10	D	旅行	87	27.5
11	E	旅行	54	20.0
12	E	旅行	54	21.0
13	E	旅行	62	19.0
14	F	タウン	60	14.5
15	F	タウン	60	16.0
16	F	タウン	60	25.5
17	G	旅行	120	39.5

表 1 正解データの概要

4.1.2 評 価 尺 度

この正解セットに対し，2.1 で述べた従来手法と提案手法の双方において語彙遷移の窓の大きさ (WL) を変化させて，その精度を比較した．ただし，語彙の遷移を算出するための左右の窓の大きさは同じとしている $(WL = WR)$．従来手法においては，各番組の分割点数を n とし，スコアの上位 n 件を最終分割結果としている．このため従来手法では窓幅の大きさに関わらず一定の分割数となるが，提案手法では過分割が起きる可能性がある．本評価では，各番組に対する分割点数 n を正解データにおける平均話題長 (221 秒) に基づき決定した．

精度は番組によっては人手でも分割がぶれてしまい，1 つの正解を定めるのが困難なため，正解の分割精度に対する相対精度で算出する．これは正解作成者 α による分割を基準とした時の正解作成者 β による分割の精度とシステムによる分割の精度を比較したものと，β による分割を基準とした時の α による分割の精度とシステムによる分割の精度を比較したものの平均で

ある．

まず，分割点セット x を正解としたときの分割点セット y の絶対精度を算出するとする．y の中のある分割点 b に対し，$[b-5$ 秒$,b+5$ 秒$]$ の範囲の中に x の分割点 a が入っているとき，b は正解 a と合致するとみなす．ただし b の他にも a と合致する分割点が y の中にある場合は，一番 a に時間が近いもののみを合致するとみなし，それ以外は合致しないとする．このようにして y 中で x 中の分割点に合致するものを抽出した分割点のセットを y'_x とする．分割点セット x 中の分割点の数を $N(x)$ とすると，x に対する y の絶対精度は以下のように計算できる．

$$
\begin{aligned}
precision_x(y) &= N(y'_x)/N(y) \\
recall_x(y) &= N(y'_x)/N(x) \\
F\ 値_x(y) &= \frac{2 * precision_x(y) * recall_x(y)}{precision_x(y) + recall_x(y)}
\end{aligned}
$$

ある番組に対する 2 つの正解 (α, β) のうち α を正解のベースとしたとき，α に対するシステムによる分割 S の相対精度は，

$$
相対\ F\ 値_\alpha(S) = \frac{F\ 値_\alpha(S)}{F\ 値_\alpha(\beta)}
$$

で算出される．最終的に S の相対精度は，複数の正解に対する相対 F 値 (S) の平均を用いる．

4.2 結　　果

調査対象番組の 17 番組について，語彙遷移の窓幅 (WL, WR) を 30 秒とした際の手がかり表現 (c) のみ，語彙遷移 (v) のみ，従来手法，提案手法での分割精度を調べた．この結果を図 4 に示す．横軸は表 1 に示した各番組を表している．また，各手法の平均を図 5 に示す．手がかり表現 (c) のみ，語彙遷移 (v) のみ，従来手法は 4.1.2 で述べたスコアの上位 n 件を採用している．

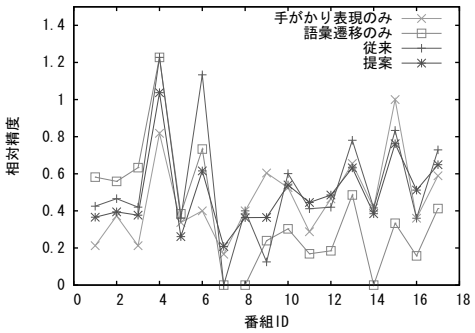

図 4　手がかり表現と語彙遷移の分割精度への影響

旅行番組については従来版と提案手法で同程度の精度が出ている．これらの番組については手がかり表現による分割精度が従来版と同程度の精度を出しているのに対し，語彙遷移があまり有効でないことが多い．また全体的にも，語彙遷移の平均精度が 0.38，手がかり表現が 0.46 であり手がかり表現の方が有効である (図 5)．

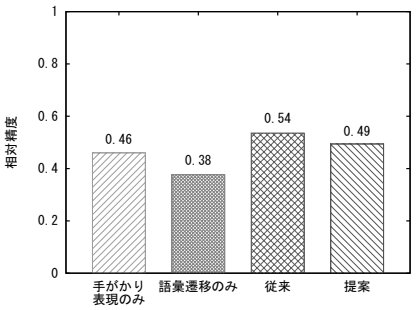

図 5　手がかり表現と語彙遷移の分割精度 (平均)

図 4 によると全体的に手がかり表現による分割 (0.46) と提案手法による分割 (0.49) は同程度の精度であるが，手がかり表現のみに頼る分割は辞書が古くなると精度が低下する可能性が高く，実用的でない．それに対して提案手法では手がかり表現がある番組で無効であっても語彙遷移で分割が補完できるという利点がある．

シリーズ A や B については，語彙遷移も有効であることが図 4 からわかる．これらは，手がかり表現による直接的なシーンの切り替わりだけでなく，シーン毎に登場人物が変わる，料理部分とトーク部分が分かれている等の間接的なシーンの切り替わりが多いという特徴がある．

また，語彙遷移の窓幅 (WL, WR) を 1 秒，5 秒，10 秒，20 秒，30 秒，40 秒，50 秒，60 秒と変化させた際の分割精度の挙動を図 6 に示す．

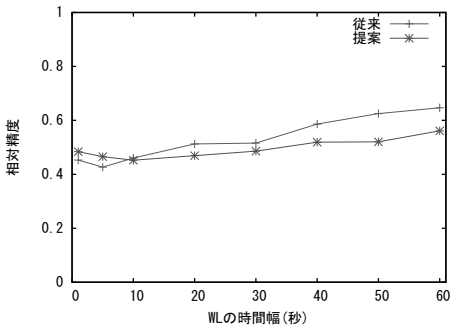

図 6　語彙遷移の窓幅の分割精度への影響

従来手法，提案手法共に窓幅が大きいほど相対精度は向上する傾向にある．ただし向上は従来手法の方が大きく，提案手法ではあまり大きく無い．

4.3 考　　察

手がかり表現・語彙遷移の有効性　図 4 において番組 7 や 8，14 で語彙遷移の分割精度が悪い理由としては，これらの番組に見られる番組最後のまとめ (その番組内の内容を振り返る部分) が考えられる．まとめの中には番組中の様々な語句が短

い間に出現し，その途中で分割されてしまうことが多い．またCのような料理番組においては，1つの材料がいくつもの料理に出現することがあり，これも語彙遷移による分割が失敗する原因となり得る．提案手法は手がかり表現による分割を基準とするため，このような場合においても手がかり表現による分割による補完が行われていることがわかる．

逆に語彙遷移の分割が成功し手がかり表現の分割があまり良くない番組6のような場合においても，語彙遷移の分割を補助的に使用することで分割点が補完されている．手がかり表現の分割が失敗し語彙遷移の分割が成功する例としては結果で述べた通り，シリーズAやBのように手がかり表現が有効な直接的なシーンの切り替わりは無いが間接的なシーンの切り替わりのある番組で発生する．

タウンや旅行番組などでは提案手法でも従来手法とあまり変わらない，十分な精度が得られることが結果からわかる．全ての文字放送字幕が揃っていなくても分割ができるところを，従来では余分に正規化処理を行っていたと言える．

窓幅の影響　従来手法，提案手法において窓幅の大きさの違いによる精度の違いは見られたものの，従来手法に比べると提案手法における精度の変動は少ない．このため，窓幅を大きくとって精度向上を目指すよりも辞書の拡充や番組のジャンルによるパラメータの切替等により精度向上を目指す方が，ユーザにとって有益であると思われる．

辞書のメンテナンス　提案手法においては手がかり表現での分割精度が重要であり，質の良い手がかり表現の辞書が豊富に用意されるほど精度は向上すると考えられる．しかし実際は，語句表現が増える等の状況の変化が起こるため，手がかり表現の抽出量は時間と共に減少していくことが予想される．このため，手がかり表現辞書を新しい状況に応じてメンテナンスしていくことが重要である．

評価尺度の妥当性　なお今回は評価尺度としてF値の相対精度を用いているが，人による分割結果に揺れがあるため，重要なのは適合率であると考えられる．更に，映像のみを視聴してユーザが判断した分割点のみを正解とするよりは，実際にアプリケーション上で分割結果を見た上で，ユーザが「許容できる」と判断したものを正解とした方が，分割結果の有効性をより正確に示すと言える．このようなユーザの主観評価は今後の課題となっている．

5. おわりに

本論文では従来のPic-A-Topicの特性を踏まえ，手がかり表現を基準とし語彙遷移の影響を少なくしたルールベースの処理によって番組のオンエア視聴時に話題分割を行う手法を提案した．従来手法と提案手法の精度評価及び手がかり表現と語彙遷移の影響調査，スコアを算出するために必要とする時間幅が分割精度にどのような影響を与えるかの評価実験を行った．評価結果を通じて手がかり表現が有効な番組については録画，正規化処理をせずとも話題分割が可能であることがわかった．また手がかり表現があまり有効でない場合においても，シーン毎に人物が変わるなど間接的なシーンの切り替わりが存在する番組においては語彙遷移が有効であることがわかった．

今回はオンエア視聴時の話題分割のみに注目したが，番組の話題分割をする際には話題へのラベリング (話題の要約) が重要である録画番組の話題に対するラベリングについては研究がなされている [10] が，オンエア視聴時に同様にラベリングをする方法を検討する必要がある．

コンピュータの演算速度が上がってきていることにより，従来では困難であったリアルタイムの映像処理が実現されつつある．これまでも文字放送字幕の分割と映像の分割とを組み合わせる手法が提案されている [6] [11]．今回提案した手法においても映像による番組の分割手法を組み合わせることで，更に精度の高い話題分割結果が得られると考えられる．

現在は手がかり表現を人手で作成しているが，新しい番組が増えていくと新たに手がかり表現が必要になることもある．そのため，今後はシリーズ番組の字幕情報から手がかり表現を自動で抽出する，汎用的に使える手がかり表現をWeb文書等から抽出する等，手がかり表現の辞書を自動で作成する機構に取り組む予定である．

文　献

[1] J. Allan, J. Carbonell, G. Doddington, J. Yamron and Y. Yang: “Topic detection and tracking pilot study: Final report”, In Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop, pp. 194–218 (1998).

[2] Y. Rui, A. Gupta and A. Acero: “Automatically extracting highlights for tv baseball programs”, MULTIMEDIA ’00: Proceedings of the eighth ACM international conference on Multimedia, New York, NY, USA, ACM, pp. 105–115 (2000).

[3] H. Aoki: “High-speed topic organizer of tv shows using video dialog detection”, Syst. Comput. Japan, **Vol.37**, No.6, pp. 44–54 (2006).

[4] K. Hoashi, M. Sugano, M. Naito, K. Matsumoto and F. Sugaya: “Video story segmentation based on generic low-level features”, IEICE Transactions on Information and Systems, D, **Vol.J89-D**, No.10, pp. 2305–2314 (2006).

[5] 井手, 孟, 片山, 佐藤：“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. パターン認識・メディア理解, **Vol.103**, No.296, pp. 13–18, PRMU2003–97 (2003).

[6] 新田, 馬場口：“放送型スポーツ映像の意味内容獲得のためのストーリー分割法”, 電子情報通信学会論文誌. D-II, **Vol.86**, No.8, pp. 1222–1233 (2003).

[7] 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”, 人工知能学会論文誌, **Vol.23**, No.5, pp. 282–292 (2008).

[8] T. Sakai, T. Uehara, T. Shimomori, M. Koyama and M. Fukui: “Pic-a-topic: Efficient viewing of informative tv contents on travel, cooking, food and more”, RIAO (2007).

[9] S. E. Robertson and K. S. Jones: “Simple, proven approaches to text retrieval”, Technical Report TR356, University of Cambridge Computer Laboratory (1997).

[10] 小山, 酒井, 福井, 上原, 下森：“効率的な番組視聴を支援するための話題ラベルの生成とその評価”, 情報処理学会研究報告. DD, **Vol.2007**, No.34, pp. 17–23, IPSJ–FI07086003 (2007).

[11] N. Babaguchi, Y. Kawai and T. Kitahashi: “Event based indexing of broadcasted sports video by intermodal collaboration”, Multimedia, IEEE Transactions on, **Vol.4**, No.1, pp. 68–75 (2002).