

E-011

クローズドキャプションを対象とした番組シーン分割手法の検討

A Study on Scene Segmentation of TV Programs Using Closed Captions

三浦 菊佳† 山田 一郎† 小早川 健† 住吉 英樹† 柴田 正啓† 八木 伸行†
Kikuka Miura Ichiro Yamada Takeshi Kobayakawa Hideki Sumiyoshi Masahiro Shibata Nobuyuki Yagi

1. はじめに

近年インターネットの普及と高速通信の実現とともに、動画を利用したサービスが増加している。放送局においても、NHK オンデマンド、アーカイブスなど過去に放送された番組を利用した動画サービス展開が期待されている。番組がシーン単位で分割されていると、たとえば動物の映っているシーンのみを集めた映像百科事典のように、様々な形態の二次活用が考えられる。番組を内容に応じたシーンに自動分割するためには、映像や音声を解析する手法も提案されているが、本稿ではクローズドキャプション（聴覚障害者用番組字幕）を対象とし、言語情報からシーンの切れ目を特定する手法を検討する。

テキストから話題変化点を検出する従来手法として、語彙的結束度による手法 [1]、手がかり語を利用した手法 [2]、ベイズ学習を利用した手法 [3][4]、などがあげられるが、番組 1 本通して一貫したトピックの下、少しずつ場面が展開するような番組には、語の分布の変化が乏しく適切ではない。また、話し言葉は省略が多いため、一般的な接続詞などの手がかり語を抽出するだけでは十分とはいえない。さらに、シーンを分割するといっても、目的や人によってその分割点の基準は様々であり、正確な教師ありデータを作成することが難しいため、機械学習も導入しにくい。

レギュラー番組（毎週もしくは毎日同じ時間帯に放送される番組）は、毎回扱われるトピックは異なるが、番組特有の共通した場面転換が多い。そのような転換箇所を特定できれば、シーン分割の手がかりになると考えられる。本稿では、シーン単位で映像を再利用するニーズの高い情報番組を対象とし、意味単位のシーン分割に向けて、話題転換の手がかりとなる語を教師なしで抽出する手法を提案する。

2. 時間別単語分布調査

レギュラーの情報番組では、毎回扱うトピックは異なっているが、各コーナーの転換パターンは似ていることが多いと考えられる。例えば、料理番組「きょうの料理」では料理内容に関わらず、冒頭に挨拶や先生紹介、料理の簡単な説明があり、終わり間近に再度材料と作り方の説明がなされるといったケースが多い。これらの場面では特有の単語が出現すると考えられる。

同一番組のどの回にも、ある時間帯のみに偏って出現する単語は、話題転換の手がかりになりやすいと仮定する。ここで、NHK のレギュラー番組「ためしてガッテン」30 番組（1 番組 43 分）を対象とし、一定時間で 30 番組共通の窓枠を設けて単語の分布を調査した。5 分間を窓幅とし 1 分ずつずらし、窓に対応するクローズドキャ

プション中の各単語 w について、以下の指標 w_{cost} を算出した。

$$w_{cost} = \frac{f}{f_{all}} \cdot \frac{df}{D} \quad (1)$$

f : 窓枠中に出現する単語 w の頻度

f_{all} : 全番組中で出現する単語 w の頻度

df : 単語 w が窓枠中に出現する番組数

D : 番組数 ($D=30$)

w_{cost} の値が高い単語が、時間帯に特有な単語といえる。ここで、横軸に窓枠の時間位置、縦軸に w_{cost} が 0.20 以上の単語種類数を表したグラフを図 1 に示す。番組のはじめ、終わりでは毎回同じ単語が偏って使われている傾向がわかる。それ

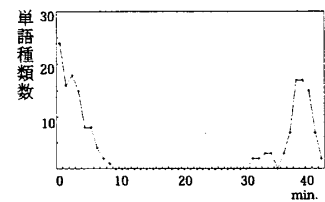


図 1. 出現時間に偏りがある単語分布

に対し中間部は、 w_{cost} 値が低いものが多く、時間窓枠で偏りを持つ単語が集中する傾向がなかった。これは中間部では各回の内容次第でシーンの長さが異なるので、単に時間で区切った単語の分布では抽出できないためと考えられる。

3. 提案手法

前章の問題を解決するため、時間軸によらず番組に共通して話題転換を表す手がかり語を取り出す手法を提案する。本手法では、自立語と付属語の組み合わせに特徴があると仮定し、文節の分布の偏りを考える。

図 2 にその概要図を示す。文節の分布が似ているセグメントを番組ごとに集め、そのセグメント系列から話題転換の手がかりとなる語を抽出する。

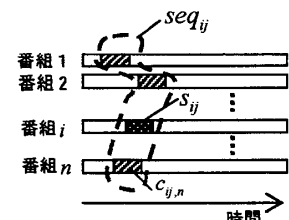


図 2. セグメントアラインメント

3.1 セグメントアラインメント

まず、クローズドキャプションを一定時間 t の長さでセグメントに分割する。 n 回分の入力番組中、番組 i の j 番目のセグメント s_{ij} のベクトルを以下のように算出する。

$$s_{ij} = (x_1, x_2, \dots, x_k, \dots, x_K) \\ x_k = 1 - f_k \quad (2)$$

K : セグメント s_{ij} 中の文節種類数

f_k : セグメント s_{ij} 中の文節 $w_{s_{ij},k}$ に対する Fisher の正確確率検定による有意確率

文節の偏りと頻度を考慮するために、Fisher の正確確率検定を用いる。 f_k の算出法については 3.2 章で詳しく説明する。セグメント s_{ij} に対して、最もコサイン距離の近い

† NHK 放送技術研究所 (人間・情報)

ベクトルを持つセグメントを残りの $n-1$ 番組から 1 つずつ選ぶ。すなわち、番組 l から選ばれたセグメント $c_{ij,l}$ は以下の式を満たす。

$$c_{ij,l} = \underset{s_{lm}}{\operatorname{argmax}} \frac{s_{ij} \cdot s_{lm}}{|s_{ij}| |s_{lm}|} \quad (3)$$

セグメント s_{ij} に対応するセグメント系列 $seq_{ij} = \{c_{ij,1}, c_{ij,2}, \dots, c_{ij,j}, \dots, c_{ij,n}\}$ が得られる。

3.2 Fisher の正確確率検定

Fisher の正確確率検定は、2 つの変数の間に統計学的に有意な差があるかを判定する検定手法で、近似せずにすべての可能な事象について列挙し、直接有意確率を計算する。枠と文節 w の頻度の組み合わせを考える場合、表 1 に示すような 2×2 分割表を作成する。各要素の値はそれぞれの条件下における各文節の頻度を表す。この事例が出現する確率 p は、以下の式で与えられる。

$$p = \frac{a+b}{a+b+c+d} \frac{c+d}{a+c} \frac{c}{c+d} = \frac{(a+b)!(c+d)!(a+c)!(b+d)!}{(a+b+c+d)!a!b!c!d!} \quad (4)$$

片側検定により、有意確率は以下に表される。

$$f = \sum_{a \geq a_0} p(w \text{ と枠内の共起頻度} = \alpha) \quad (5)$$

表 1. Fisher の 2×2 分割表

	枠	枠外	計
w	a	b	$a+b$
w 以外	c	d	$c+d$
計	$a+c$	$b+d$	$a+b+c+d$

3.3 文節のスコア算出

セグメント s_{ij} に対応するセグメント系列 seq_{ij} 内に出現する文節 $w_{seq_{ij},k}$ のスコア $score_{ij,k}$ を以下の式で算出する。この値が高いほど、手がかり語になりやすいと考える。

$$score_{ij,k} = \frac{df_{seq_{ij},k}(1-f_{ij,k})}{D} \quad (6)$$

$df_{seq_{ij},k}$: 文節 $w_{seq_{ij},k}$ が seq_{ij} 内に出現した番組数
 D : 全番組数

$f_{ij,k}$: 文節 $w_{seq_{ij},k}$ に対する Fisher の正確確率検定による有意確率

4. 実験および考察

前章で述べた手法の有効性を確認するために、2 章で用いたものと同じ NHK の情報番組「ためしてガッテン」30 番組を対象とし、話題転換の手がかりとなる語を抽出する実験を行った。セグメント長は $t=2$ 分とし、1 分単位でセグメント境界をずらし、全てのセグメントについて

処理を行った。また、句点は字幕作成上の都合で省略される場合があるため排除し、全ての文節について処理を行った。構文解析には Cabocha を使用した。セグメント 1 つに対して 1 セグメント系列と、そこに含まれる各文節のスコアが算出されるため、表 2 に 2000 年 12 月 6 日放送分の番組をセグメント基準にした抽出語の一例を示す。

アラインメントの過程では、番組開始、および終了付近では、それぞれ近い時間帯のセグメント同士が選ばれており、2 章の検証と同じくこれらの時間帯では毎回共通する語が集中したといえる。表 2 の例のほか、冒頭部分からは「『ためして』」「ガッテン』の」「時間です」「今日は」、終了部分からは「『小論文』」「『実習コーナー』です」「ガッテンでしょうか」といった表現が多く抽出された。

それに対し、中間部のアラインメントでは番組によってそれぞれ時間帯が離れているセグメントが抽出されていた。ここでは、「はいテーマです」「ご覧」「下さい」といった決まり文句や、「さて」「まずは」「では」のように一般的に話題を変えるときに接続詞が抽出された。一方、「この」「その」などの連体詞や「ある」といった全般的に高い頻度のもも抽出されており、候補の文節を制限する必要がある。また、今回はセグメント長を 2 分に固定したが、適切な長さを推定する必要がある。

5. まとめ

本稿では、レギュラー番組において各回に共通して偏った分布で出現する表現をシーン分割の手がかり語として自動抽出する手法を提案した。文節の指標に Fisher の正確確率検定を用い、セグメントアラインメントを行うことで出現時刻が異なる手がかり語も抽出できた。今後、セグメント長や語の単位を見直すとともに、他番組にも適用させ、シーン分割に応用させていく予定である。

参考文献

- [1]Hearst, "Multi-paragraph segmentation of expository text", 32nd Annual Meeting of the Association for Computational Linguistics, pp.9-16, 1994.
- [2]望月源, 本田岳夫, 奥村学, 「複数の表層的手がかりを統合したテキストセグメンテーション」自然言語処理, vol.6, No.3, pp.43-58, 1999.
- [3]持橋大地, 菊井玄一郎, 「Gibbs Sampling による確率的テキスト分割と複数観測への拡張」NLP-2006, pp.212-215, 2006.
- [4]越仲孝文, 奥村明俊, 磯谷亮輔, 「HMM の変分ベイズ学習によるテキストセグメンテーション及びその映像インデキシングへの応用」, 電子情報通信学会論文誌 D, Vol.J89-D, No.9, pp.2113-2122, 2006.

表 2. 手がかり語抽出例

(a) 冒頭部 00:03:00 から 2 分		(b) 中間部 00:26:00 から 2 分		(c) 終了部 00:39:00 から 2 分	
score	抽出語	score	抽出語	score	抽出語
0.97838	では	0.70000	はいテーマです	0.63333	ありがとうございました
0.56667	問題	0.66667	下さい	0.60000	それでは
0.56667	(拍手) こんばんは	0.63333	ご覧	0.56667	ガッテン!
0.53333	「基礎講座」から	0.63326	さて	0.53772	では
0.53333	ゲストの	0.49732	でも	0.46667	本日の