

## 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

小山 誠\* 酒井 哲也† 福井 美佳\* 上原 龍也† 下森 大志†

\* (株) 東芝 研究開発センター 知識メディアラボラトリー

† (株) 東芝 研究開発センター マルチメディアラボラトリー

‡ (株) ニュースウォッチ

makoto3.koyama@toshiba.co.jp

あらまし 本論文では、TV 番組を話題毎に分割して得られるセグメントに対して、内容を簡潔に表す代表キーワードや代表フレーズ、代表文を付与する話題ラベリングの手法とその評価について述べる。本手法は、セグメントに対するクローズドキャプション（字幕テキスト）から、情報検索における適合フィードバック技術を応用して、対象セグメントの話題に関するキーワードやフレーズ、文を抽出する。「旅行」「タウン」「料理」ジャンルの情報番組を対象に、自動および人手により作成した文形式のラベル、キーワード・フレーズ形式のラベルに対して、39名の被験者による主観評価実験を実施した。その結果、ラベルの「わかりやすさ」と「適切さ」に関し、全体としてはキーワード・フレーズ形式のラベルが文形式のラベルより高い評価を得られることを確認した。

### Automatic Generation of Topic Labels for Efficient Video Viewing

Makoto KOYAMA\* Tetsuya SAKAI† Mika FUKUI\*

Tatsuya UEHARA† Taishi SHIMOMORI†

\*Knowledge Media Laboratory, Toshiba Corporate R&D Center

†Multimedia Laboratory, Toshiba Corporate R&D Center

‡NewsWatch, Inc

makoto3.koyama@toshiba.co.jp

**Abstract** This paper describes a method for generating keyword, phrase and sentence labels for video segments of TV programs. By using a relevance feedback algorithm in information retrieval, it selects topic keywords, phrases and sentences from closed caption text in each topical segment. 39 subjects evaluated keyword, phrase and sentence labels from TV programs about travel, town and cooking. The results show that keyword and phrase labels achieve better results than sentence labels on understandability and relevance of labels.

## 1 はじめに

TV 放送の多チャンネル化・録画機器の大容量化により、録りためた TV 番組を効率的に視聴する技術のニーズが高まりつつある[5]。映像番組を効率よく視聴できるようにする技術は多く研究されているが、ニュース放送のセグメンテーションやスポーツ番組からのハイライト抽出などが主なものである[6, 7]。

我々は、さまざまなジャンルを含む情報番組を対象に、ユーザに番組の「選択的視聴（つまみ見）」の環境を提供することを目的とするシステム Pic-A-Topic(ピカトピック)の開発を進めている[1, 2]。

Pic-A-Topic は、クローズドキャプション（字幕テキスト）を利用して、複数の話題が含まれる番組を話題毎にセグメントに分割し、各セグメントの内容を簡潔に表す代表キーワードや代表フレーズ、代表文を抽

出し、その結果を話題単位の視聴ができるようにユーザに提示する。これにより、例えば、2時間の旅行番組の中からユーザが興味を示したレストラン情報の部分のみを20分で視聴するといったことが可能になると考えられる。

本報告では、Pic-A-Topic における話題ラベリング（セグメントに対する代表キーワードや代表フレーズ、代表文の付与）について、生成したラベルの有効性を調べるための精度評価および主観評価実験を行ったので、その結果を報告する。

以下、Pic-A-Topic の構成を簡単に述べた後、話題ラベリングの方法を説明する。その後、本研究で行った評価の方法とその結果を詳述する。

## 2 Pic-A-Topic の構成

図1に Pic-A-Topic のシステム構成を示す。システムへの入力は、番組のクローズドキャプション、EPG のテキストデータであり、これらのデータに基づき番組のジャンル判定、話題分割、話題ラベリングを行い、ユーザに提示する画面を生成する。

ジャンル判定では、番組に対して、EPG で与えられているジャンル情報からシステムが独自で保持するジャンルへの割り当てを行う。ここでは、クローズドキャプション、EPG の内容に対して、人手で作成した簡単な if-then ルールを利用することにより実現している。

話題分割は、番組のクローズドキャプションデータを利用して、話題の転換を表す自然言語表現(cue phrases)に着目した方法と語彙の遷移に着目した方法を併用し、最後に結果を統合することにより行う[1, 2]。上記キューフレーズは、番組のジャンルに応じたものを利用するようになっており、番組ジャンルに応じた分割処理が可能となっている。

さらに、話題分割した各セグメントに、クローズドキャプションから話題ラベルを抽出し（話題ラベリングの処理の詳細は次章で説明する）、ユーザへの表示画面を生成する。画面例を図2に示す。

## 3 話題ラベリング

話題分割で分割されたセグメントに対するクローズドキャプションから、セグメントの代表キーワード、代表フレーズまたは代表文を抽出する。

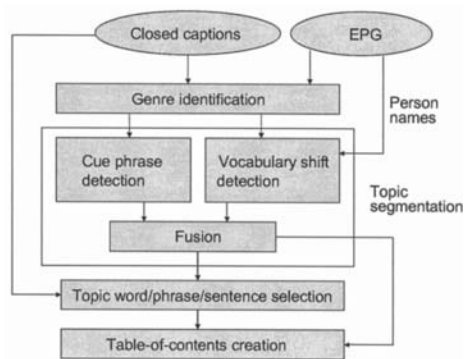


図1 Pic-A-Topic の構成



図2 Pic-A-Topic の画面例

### 3.1 代表キーワードの抽出

代表キーワードの抽出は、情報検索における確率検索モデルに基づく適合フィードバックのアルゴリズム[3]の適用により実現した。番組のクローズドキャプション中の各文を文書とみなし、ラベリングの対象セグメントに含まれる文を適合文書と見なして、以下で定義される  $ow$  (offer weight) の値の大きい順に一定文字数を超えるまで代表キーワードを選出する。

$$rw(w) = \log \frac{(r(w) + 0.5)(N - n(w) - R + r(w) + 0.5)}{(n(w) - r(w) + 0.5)(R - r(w) + 0.5)}$$

$$ow(w) = r(w) * rw(w)$$

各要素の意味は以下の通りである。

- N 番組のクローズドキャプションに含まれる文の総数
- R 対象セグメントに含まれる文の総数
- n(w) 番組のクローズドキャプションに含まれる文のうち、キーワード w を含むものの数
- r(w) 対象セグメントに含まれる文のうち、キーワード w を含むものの数

ここで、キーワードは、形態素解析または固有表現抽出[4]の処理結果で得られた形態素または固有表現を単位とする。

### 3.2 代表文、代表フレーズの抽出

代表文の抽出は、代表キーワードの重みを利用して行う。対象セグメントのクローズドキャプションの文から、スコアの大きい代表キーワードを多く含んだ文を抽出する。以下に、抽出のアルゴリズムを示す。

1. 対象セグメントに含まれるテキストを文単位に分割する。
2. 各文に対し、以下に示すスコア  $s$  を計算する。ここで、第1項は文に含まれる代表キーワードの重みの和を計算した値である。第2項における  $n$  は文のセグメント先頭からの位置（先頭文は  $n=1$ 、2番目の文は  $n=2$ ）で、 $A$  は定数であり、先頭に近くに位置する文ほどスコアが大きくなる。

$$s = \sum ow(w) + \frac{A}{n}$$

3. スコアが最大のものをひとつ代表文として採用し、この代表文に含まれる代表キーワードの重みを減少させる。（今回は重みを0点にしている）
4. 指定された数の代表文が得られるまで 2.以下の処理を繰り返す。

なお、3で代表キーワードの重みを減少させているのは、出力する文同士の内容が重複しないようにし、情報量の多い表示画面をユーザに提供するためである。

代表フレーズは以下のアルゴリズムにより抽出する。基本的な処理は代表文の抽出と同じであり、フレーズ中の代表キーワードの重みの和とセグメントの先頭からの位置に基づくスコアを用いる。

1. 対象セグメントに含まれるテキストを文単位に分割する。
2. 各文からフレーズとなる単語列を抽出する。（各文を形態素解析した結果に対する、品詞の接続関係に基づく簡単な抽出ルールを利用する。）
3. 抽出した各フレーズに対し、以下に示すスコア  $s$  を計算する。ここで、第1項はフレーズに含まれる代表キーワードの重みの和を計算した値である。第2項における  $n$  は抽出元の文のセグメント先頭からの位置で、 $A$  は定数である。

$$s = \sum ow(w) + \frac{A}{n}$$

4. スコアが最大のものをひとつ代表フレーズとして採用し、この代表フレーズに含まれる代表キーワードの重みを減少させる。（今回は重みを0点にしている）
5. 指定された数の代表フレーズが得られるまで 3.以下の処理を繰り返す。

## 4 評価

話題ラベリングにより生成したラベルに対して行った精度評価実験、および39名の被験者による主観評価実験の方法と結果について説明する。

### 4.1 実験準備

#### ■番組データ

番組としては、情報番組の中の旅行、タウン、料理の3ジャンルから3コンテンツずつ、合計9コンテンツを用いた。各ジャンルの番組のおおよその内容は次のようになっている。ここで、旅行、タウンは1コンテンツ約1時間、料理は1コンテンツ約15分となっ

ている。

旅行：出演者が観光地を訪れ、散策、温泉、食事などを楽しむ。

タウン：街のスポット情報（場所、店、イベントなど）を順に紹介する。

料理：いくつかの料理の調理方法を順に紹介する。

### ■対象セグメント

話題ラベリングに与えるセグメントは、自動分割した場合の分割誤りの影響を受けないよう、ここでは人手による話題分割結果のセグメントを用いた。すなわち、人手によって話題毎に分割された各セグメントに対するラベリングの結果を評価する。ここで、1コンテンツあたりのセグメント数の平均は、旅行、タウン、料理でそれぞれ25.0、41.3、5.0セグメントとなっている。なお、人手による分割結果に対するシステムの分割結果の相対精度はクローズドデータに対して約8割超となっている[2]。

### ■評価ラベル

対象セグメントから表1に示す7通りの方法で生成したラベルを評価用に生成した。各ラベルの長さは、文字数が最大で30文字以下となるようにしている。

L1～L3は文形式のラベル、L4～L7はキーワードまたはフレーズの形式のラベルである。

L4は各セグメントに対するクローズドキャプションから、そのセグメントの話題を代表するフレーズを人手により抽出したものである。評価番組の9コンテンツに対して4人の正解作成者を割り当てて作成したものである（一人が1～3コンテンツの正解を作成した）。また、L1は、L4を含む文を人手により抽出したものである。

L2,L3,L5～L7はクローズドキャプションから自動抽出したものである。L2は、セグメントの先頭から一定文字数までの文を抽出したものである。

L3は3.2節の代表文抽出アルゴリズムを用いて生成した。L5、L6は代表キーワード抽出により抽出した。L5はキーワードとして形態素だけを使い、L6はキーワードとして形態素と固有表現を使った。また、L7は代表フレーズの抽出アルゴリズムを用いて生成した。ここで代表文、代表フレーズのスコアの計算には、形態素と固有表現の重みを使った。また、固有表現の重みは形態素の重みの2倍とした。

表1 評価ラベル

文	L1	人手による代表文
	L2	先頭文
	L3	代表文
キーワード・フレーズ	L4	人手による代表フレーズ
	L5	代表キーワード(形態素) 例：“いわし”、“御殿場”
	L6	代表キーワード(形態素+固有表現) 例：“いわし料理”、“御殿場市”
	L7	代表フレーズ 例：“いわし料理の専門店”

## 4.2 精度評価

### ■方法

自動生成したラベルと人手により生成したラベルとの重なりを、形態素の再現率により調べた。表1のラベルL4を正解ラベルとみなして、自動生成した各ラベルに対して再現率Rを次式により計算した。

$$R = Nsc / Nc$$

Nsc システムが出力したラベルと正解ラベルに共通して含まれる形態素の個数

Nc 正解に含まれる形態素の個数

ここで、形態素は自立語のみを対象とし、表記の同じ形態素が複数ある場合は1つとして計算した。

### ■結果

各ラベルの再現率を図3に示す。旅行、タウン、料理のジャンル毎に、3コンテンツに対する再現率の平均を出している。

まず、キーワード・フレーズ形式のラベルL5,L6,L7を比べると、L6,L7で再現率が約40～55%となり、L5に対して高くなっている。L5では形態素のみを用いているが、L6、L7では形態素と固有表現を利用しており、固有表現として地名や料理名などが得られたためである。

文によるラベルL2,L3に関しては、一部を除き、全体的にキーワード・フレーズ形式のラベルより再現率は低くなっている。文形式の場合は30文字中に含まれる語数がキーワード・フレーズ形式の場合と比べると少なくなり、再現率も低くなっていると思われる。

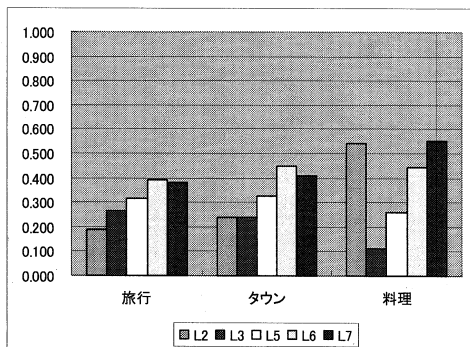


図3 再現率

料理に関しては、L2の再現率が高くなっている。この番組に関しては、話題の変わり目毎に、これから調理する料理の簡単な説明があり、セグメントの先頭付近にセグメントの内容を表すのに有用な情報が集まっていたためである。

### 4.3 ユーザ評価

#### ■方法

表1のラベルL1~L7について、ラベルの「わかりやすさ」と「適切さ」について被験者39名による評価実験を行った。

ラベルのわかりやすさは、ラベルを単独で見たときにラベルが表す内容がわかるかどうかを各被験者が、a“わかる”、b“だいたいわかる”、c“あまりわからない”、d“わからない”の4段階で評価した。

ラベルの適切さは、ラベルとラベルが表すセグメントの内容(ここではセグメントに対するクローズドキャプションデータを用いた)との一致度を各被験者がa“合っている”、b“だいたい合っている”、c“あまり合っていない”、d“合っていない”の4段階で評価した。

評価対象セグメントは、各ジャンルから2コンテンツずつ合計6コンテンツを使用し、旅行、タウン番組については番組開始から約1/3までのセグメントを抽出・利用した。容量の少ない料理番組については全セグメント利用した。評価に用いたセグメント数は、旅行:18、タウン:22、旅行:11セグメントで、各セグメントに対して2~3人の被験者が評価を与えた。

最後に、各被験者がL2,L3,L6,L7の4つのラベルに対して、それぞれを比べてわかりやすい順に順位を

付けた。

#### ■結果1(わかりやすさ)

ラベルのわかりやすさについて4段階で評価した結果から、aを1点、bを $2/3(=0.666\cdots)$ 点、cを $1/3(=0.333\cdots)$ 点、dを0点として、各セグメントに対して付けられた評価の点数を計算した。各ラベルに付けられた点数の平均値を図4に示す。

旅行、タウン番組については、人手によるラベルL4の点数が最も高い。また、人手により作成されたラベルL1(文形式)とラベルL4(フレーズ形式)とを比較すると、L4の点数が高い。また、文形式のラベルL2,L3とキーワード・フレーズ形式のラベルL4,L5,L6とを比較するとL4,L5,L6の点数が高くなっており、おおむねキーワード形式によるラベルがわかりやすいと評価されている。

旅行番組については先頭文を用いたラベルL2の点数が高くなっている。

#### ・再現率との関係

図3の再現率との関係を見ると、料理番組に関しては、再現率の高いL2,L7がわかりやすさにおいても高い点数となっており、相関の傾向がみえる。

旅行、タウン番組については、文形式のラベルがキーワード・フレーズ形式のラベルに対して再現率が低くなっているが、わかりやすさにおいても同様の傾向はみられた。ただし、再現率で見られる差よりも差は小さい。

また、形態素だけのラベルL5については、再現率はL6,L7に比べて低いが、わかりやすさではL6,L7と同等以上の点数になっている。L5において、地名、施設名などの一部だけが出力されている場合でも、およそその内容が分かる場合があったと思われる。

#### ■結果2(適切さ)

ラベルの適切さについて、4段階で評価した結果から、aを1点、bを $2/3(=0.666\cdots)$ 点、cを $1/3(=0.333\cdots)$ 点、dを0点として、各セグメントに対して付けられた評価の点数を計算した。各ラベルに付けられた点数の平均値を図5に示す。

旅行番組については、文形式のラベルに対してキーワード・フレーズ形式のラベルの点数がややたかくなっている。

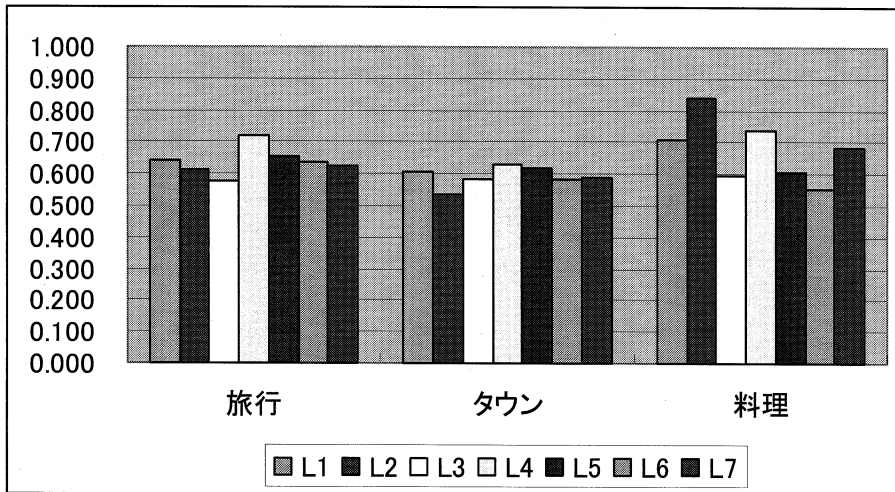


図4 ラベルのわかりやすさ

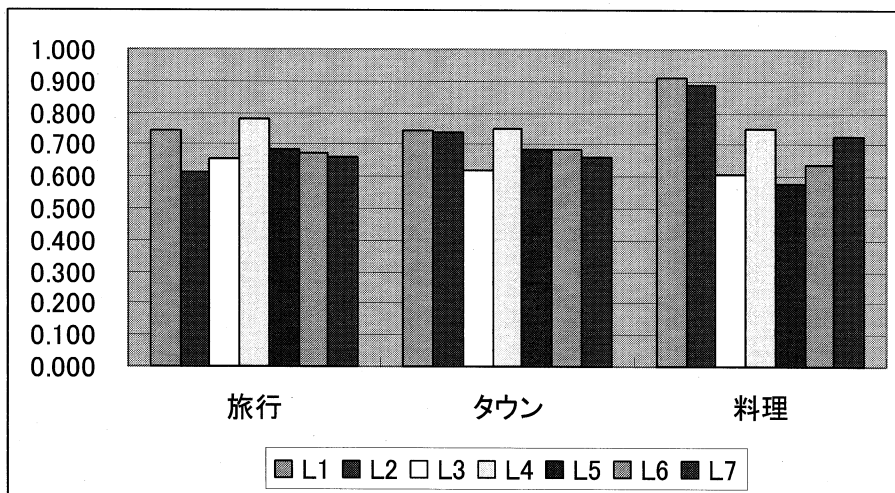


図5 ラベルの適切さ

タウン番組, 料理番組については先頭文を用いたラベルL2の点数が高くなっている

・再現率との関係

料理番組については, わかりやすさの場合と同様, 再現率の高かった L2,L7 の点数が高く, 相関の傾向がみられる。

旅行, タウン番組については, 文形式のラベルがキーワード・フレーズ形式のラベルに対して再現率が低くなっている。適切さにおいてもおおよそ同様の傾向がみられるが, タウンにおいては先頭文によるラベル

L2 の点数が高くなっている。先頭文に関しては, 再現率が低い場合でも, 適切さにおいて高く評価される場合があることがわかる。

また, 形態素だけのラベルL5はL6,L7に対して再現率は低くなっているが, 適切さではL6,L7と同等以上の点数が得られており, L5において地名, 施設名の一部だけが出力されている場合でも, 主観評価ではあっていると評価されたと考えられる。

・「わかりやすさ」との関係

わかりやすさと適切さについては, わかりやすさに

対して適切さの点数が、全体的にやや大きくなっている。わかりやすさについては、旅行、タウン、料理の各番組の全ラベルの平均は、0.64, 0.59, 0.68となっているが、一方、適切さではそれぞれ 0.69, 0.70, 0.73 となり、3 番組とも適切さの点が高くなっている。これは、わかりやすさにおいて、低い評価を付けた場合も、実際に内容を見た場合に適切であった場合があったためと考えられる。

### ■結果 3 (ラベル間の比較)

前述のわかりやすさと適切さの評価を行った後、各被験者は、ラベル L2,L3,L6,L7 を見比べてわかりやすい順に 1 位から 4 位まで順位を付けた。各ラベルに付けられた順位の平均を図 6 示す。

3 ジャンルとも L7 が各番組で最も良い順位となっている。料理番組に関しては L2 と L7 の平均順位は同じ値になっている。

L6 と L7 はそれぞれ単独でわかりやすさ、適切さを評価した場合は大きな差は見られなかったが、比較評価した場合は明らかな差が出ており、キーワードとフレーズとを比べた場合は、より長いフレーズの方がラベル単位の情報量が多く、わかりやすいと評価されたと思われる。

## 5 おわりに

効率的な番組視聴を支援するための話題ラベリングの方法とその評価実験について報告した。自動生成および人手により作成した 7 種類のラベルに対し、ラベルの「わかりやすさ」と「適切さ」について被験者 39 名による主観評価を実施した結果、全体としてはキーワード・フレーズ形式のラベルの結果が良くなることを確認した。また、ジャンルによっては先頭文を用いた文形式のラベルの結果が良くなることを確認した。

今回はラベリング単独での評価を実施したが、話題分割、ユーザーインターフェースを含めたシステム全体としてのユーザビリティの評価は今後の課題である。また、番組に対する字幕情報としては、オープンキャプション(テロップ)もある。クローズドキャプションからは得られない情報がオープンキャプションに含まれる場合もあり(その逆もある)、こうしたオープンキャプションの利用も課題として挙げられる。

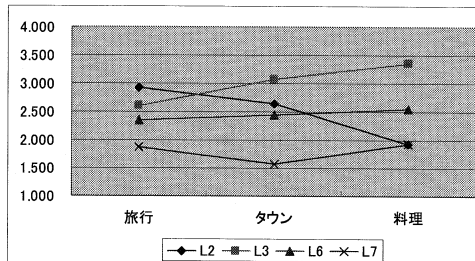


図 6 ラベルの比較評価

## 参考文献

- [1] Sakai, T., Uehara, T., Sumita, K. and Shimomori, T.: Pic-A-Topic: Gathering Information Efficiently from Recorded TV Shows on Travel. AIRS 2006, Lecture Notes in Computer Science 4182, pp. 374-389, Springer-Verlag (2006)
- [2] Sakai, T. et al.: Pic-A-Topic: Efficient Viewing of Informative TV Contents on Travel, Cooking, Food and More, RIAO 2007 Proceedings, to appear.
- [3] Robertson, S. E. and Sparck Jones, K.: Simple, Proven Approaches to Text Retrieval, University of Cambridge Computer Laboratory, TR356 (1997)
- [4] 市村ほか: 固有表現抽出と回答タイプ体系が質問応答システムの性能に与える影響. 電子情報通信学会論文誌 Vol.J88-DII, No.6 (2005)
- [5] Smeaton, A., Kraaij, W. and Over, P.: The TREC Video Retrieval Evaluation (TRECVID): A Case Study and Status Report, RIAO 2004 Proceedings (2004)
- [6] Hauptmann, A. G. and Witbrock, M. J.: Story Segmentation and Detection of Commercials in Broadcast News Video. Advances in Digital Libraries '98 (1998)
- [7] Rui, Y., Gupta, A. and Acero, A.: Automatically Extracting Highlights for TV Baseball Programs. ACM Multimedia 2000 Proceedings (2000)