

番組実況チャットを利用したテレビ番組のメタデータ自動抽出方式

宮 森 恒[†] 中 村 聡 史[†] 田 中 克 己^{†,††}

本稿では、ウェブ上の実況チャットを利用することでテレビ番組のメタデータを自動的に抽出する手法を提案する。実況チャットは、テレビ番組と同期して視聴者によって書き込まれるウェブ上の掲示板である。番組実況チャットのデータを、利用者全体あるいは特定の傾向を持つ利用者のデータ集合として扱い、チャット上に表れる特徴的な表現を認識処理することで、番組の盛り上がり場面や、視聴者の嗜好・興味に沿ったリアクションなど、視聴者視点に関連するメタデータを効率良く抽出できる。実験の結果、提案方式により、視聴者の反応の大きさや喜び、落ち込みなどの感情度数などのメタデータを効率良く抽出できることを確認した。また、提案方式で得られるメタデータを利用した番組視聴システムを実装することにより、シーンのランキング表示やダイジェスト生成など視聴者視点を反映した多様な番組視聴へ応用可能になることを確認した。

Method of Automatically Extracting Metadata of TV Programs Using Its Live Chat on the Web

HISASHI MIYAMORI,[†] SATOSHI NAKAMURA[†] and KATSUMI TANAKA^{†,††}

A method of automatically indexing broadcast content using live chat on the Web is proposed. The live chat is a Web bulletin board where the viewers post messages in sync with a TV program. The messages in the live chat written by all viewers or by viewers having similar interest or preference to the user are statistically analyzed with pattern recognition to effectively extract indices related to viewer's viewpoints such as important scenes in the program or responses by a particular viewer. Experiments indicate that the proposed method can efficiently extract indices such as the intensity of viewers' responses and degree of emotional delight or disappointment. They also indicate that a prototype TV viewing system using the extracted indices enables a new way of viewing TV content from different perspectives reflecting viewers' viewpoints such as scene ranking and digest generation.

1. はじめに

近年、HD レコーダの性能向上にともない、一般の利用者が録画できるテレビ番組のデータ量は飛躍的に増加している。現在、600 GB の容量を持つ HD レコーダが登場し、画質モードによっては 1,070 時間以上の録画が可能である。しかし、録画した映像が増大しても、人間が 1 日に視聴できる時間には物理的に限界がある。このため、膨大な映像データから興味のある部分を効率良く探したり、内容の概要を手軽に理解したり、限られた時間でハイライトだけを見るといった多様な視聴方法を提供する仕組みが重要となる。

そのような仕組みを実現するための重要な基礎技術

として、映像のメタデータ抽出があげられる。従来、動画像中の色やテキスト、カメラ操作、人物の顔などの特徴や、字幕テキスト、音の種別や大きさなどをマルチモーダルに利用した種々のメタデータ抽出手法が提案されてきた。しかし、これら従来手法が利用してきたデータは、番組の配信側（放送局）から提供される情報のみであったため、抽出されたメタデータは、基本的に、番組作成者・配信側の意図を反映しているだけであった。よって、従来手法においては、シーン探索や内容の早見、ダイジェスト視聴などの内容に、自分以外の他の視聴者の視点や反応といった要素を取り入れることはできなかった。

一方、近年ウェブ上の実況チャットの利用者数が増加している。実況チャットは、テレビ番組の視聴者がインターネット上のチャットコミュニティに集まり、テレビ番組の内容について感じたことなどやテレビ番組の内容自体のことをリアルタイムに書き込み、会話を楽しむシステムである。利用者は、テレビ番組と並行

[†] 情報通信研究機構

National Institute of Information and Communications Technology

^{††} 京都大学

Kyoto University

してチャットを楽しむことにより、視聴者との感情の共有による一体感を味わうことができる。

そこで本稿では、ウェブ上の実況チャットに基づくテレビ番組の自動メタデータ抽出手法を提案する。番組実況チャットのデータを、利用者全体あるいは特定の傾向を持つ利用者のデータ集合として扱い、チャット上に表れる特徴的な表現を認識処理することで、番組の盛り上がり場面や、視聴者の嗜好・興味に沿ったリアクションなど、視聴者視点に関連するメタデータを効率良く抽出できると考えられる。抽出したメタデータを利用することにより、視聴者視点を反映した多様な番組視聴へ応用可能と考えられる。

本稿の構成は以下のとおりである。2章では実況チャットについて述べる。3章では実況チャットを利用したメタデータ抽出の処理手順について説明し、4章で実験および考察を示す。5章で関連研究を整理する。最後に、6章でまとめを述べる。

2. 番組実況チャットとは

スポーツ試合などのテレビ番組を1人ではなく複数のグループで観戦するというスタイルは、欧米などをはじめ世界の様々な地域で普及している。イタリアではサッカーの試合のたびに多くの人がスポーツバーに集まって熱狂し、アメリカでは大リーグの試合があるたびにスポーツバーに集まって試合観戦を楽しんでいる。

この、多くの人が同じ空間に集まりテレビ番組を楽しむというスタイルの最も大きな利点の1つは、感情を実際に共有できるということであると考えられる。ともに喜び、ともに悲しみ、ともに笑い、ともに泣き、ともに怒る。こうした視聴者の感情の共有による一体感が楽しいことが、そうしたスタイルの視聴が行われるゆえんであらうと考えられる。また、テレビ番組の内容について同意を求めたいとか、他人の反応を見たいというのも1つの要素であらう。

テレビ番組ではこうした点を考慮し、いろいろな趣向を凝らしている。たとえば、バラエティ番組などで笑い声を挿入したり、テレホンショッピングにおいて客の驚きの声を挿入したりしているのは、テレビ番組の視聴者に対する感情の一提案であらう。しかし、テレビ放送で提供されるものは画一的なものであり、視聴者全員が同様に感じるには限らない。たとえば、AチームとBチームによるサッカーの試合を観戦する場合、視聴者は「Aチームの応援」「Bチームの応援」「中立」の3つの立場に大別される。この立場の違いは大きく、同じ試合を見ていてもAチームが勝って

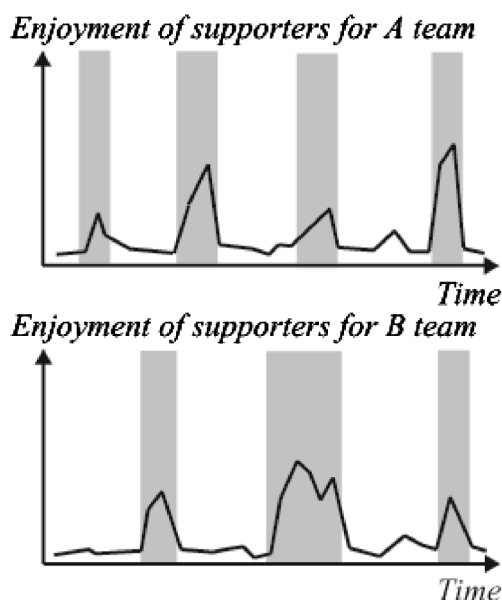


図1 異なるチームを応援する視聴者の感情の起伏の違い
Fig. 1 Rise and fall of emotions of viewers supporting different teams.

いるとき、Bチームが勝っているときでそれぞれ、視点が異なり、感情の起伏も違ったものになると考えられる(図1)。また、同じAチームを応援している場合でも、好きな選手の違いや、チームに対する思い入れの違いなどにより、視点の違いは生じるであらう。

一方、近年番組実況チャットの利用者数が増加している。番組実況チャットとは、テレビ番組を観ている視聴者がインターネット上のチャットコミュニティに集まり、テレビ番組の内容について感じたことなどやテレビ番組の内容自体のことをリアルタイムに書き込み、会話を楽しむシステムである。テレビ番組と並行してチャットを楽しむことにより、感情の共有および一体感を味わうことができる。これはまさに、従来の物理的な空間に人々が集まってテレビ放送を楽しんでいたものを、仮想的な空間に応用したものであるといえる。

日本の巨大掲示板サイトである2ちゃんねるでは、この番組実況チャットを大規模サポートしている。2ちゃんねるでは、日本の各チャンネル専用の掲示板が設置されており(図2)、その中に番組自体について語り合うスレッド(ある話題に関して投稿されたメッセージの集まりで、書き込まれた時間順に並べられている)が多数存在する。同一の番組であっても、視聴者ごとに興味や嗜好が異なるため、野球のそれぞれの対戦相手専用のスレッドなど様々な視点に基づくチャットコミュニティが作られている。テレビ番組放送中に

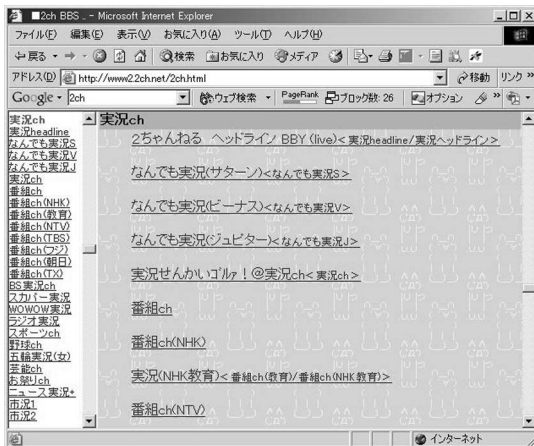


図 2 番組実況チャットの掲示板一覧の例

(http://www2.ch.net/2ch.html より引用)

Fig. 2 Example of list of bulletin boards supporting live chat.

- 80 名前:名無しさん 05/05/17 08:48:22 ID:R9D+YqFT
ものすごい体勢からTDキター————(∇)————!!!
- 81 名前:名無しさん 05/05/17 08:48:28 ID:+FNgUHKl
めちゃめちゃ点入りそうな勢いだな
- 82 名前:名無しさん 05/05/17 08:48:30 ID:BZA6/Cc
しかも1stだけで...
- 83 名前:名無しさん 05/05/17 08:48:48 ID:7iZg967n
奴はこういうときすぐムキになる。
しかしただでさえ背高いのにジャンプ投げとは反則やー
- 84 名前:名無しさん 05/05/17 08:49:15 ID:3bMzDEmi
初心者には空中戦がわかりやすくてよいかも。
- 85 名前:名無しさん 05/05/17 08:49:23 ID:R9D+YqFT
ブリッツ多かったけどバック何人残ってたかな?

図 3 番組実況チャットの書き込み内容の例

Fig. 3 Conversation in live chat.

は番組実況チャットから人が絶えることがないほど人気が高い。

番組実況チャットでは、設定による違いはあるが、書き込みの内容には時間と、書き込み者を一意に識別する ID が付加されていることが多い。この識別 ID を利用することで、どの書き込みはだれが書いたのかということを区別することができる。

3. メタデータの抽出手順

ここでは、番組実況チャットを利用した番組コンテンツのメタデータ抽出手法について説明する。

3.1 方針

番組実況チャットは、設定による違いはあるが、基本的に、書き込み時刻、書き込み者の ID、書き込み内容から構成されている (図 3)。

たとえば、このデータをもとに、ある単位時間ごと

1. キター————(∇)————!!!
2. キター————∩(∇)∩————!!!
3. TD————∩(∇)∩————!!
4. キター————∪(∇)∪————!!!
5. アツヒャツヒャ!∪(∇)∪アツヒャツヒャ!

図 4 喜びの表現に用いられるアスキーアートの例

Fig. 4 ASCII art used to express enjoyment.

1. (ノ∇)アチャー
2. ∪(∇)ノウワァン
3. (・ω・)ショボーン
4. ムー(A)
5. I I I I I I (∇) I I I I I I

図 5 落胆の表現に用いられるアスキーアートの例

Fig. 5 ASCII art used to express disappointment.

に書き込みエントリ数を計測し、その時間変化をグラフに表すと図 1 のようになる。この図より、番組実況チャットに参加している視聴者の反応の度合いを各時刻について把握することができると考えられる。

また、番組実況チャットへの書き込みでは、アスキーアートをを用いた特徴的な表現がしばしば用いられる。たとえば、スポーツ番組で盛り上がった場面などでは、図 4 のようなアスキーアートがよく利用される。

同様に、落胆した場面では、図 5 のようなアスキーアートがよく用いられる。

このような特徴的な表現の出現回数をパターンマッチングにより検出することで、書き込みを行った視聴者の喜び (盛り上がり・興味) や落胆 (悲しみ) といった感情の度合いを抽出することができると考えられる。

3.2 提案手法の説明

以上のような考えに基づき、ここでは図 6 のようなメタデータ抽出処理を行う。

まず、記録された実況チャットのデータを構文解析して、書き込み時刻、書き込み者の ID、書き込み内容を得る。

なお、書き込み時刻と、書き込み内容に対応する番組映像シーンの時刻にずれが生じる場合がしばしばある。これは、映像中のあるイベントが起きた後に、視聴者が書き込みを行うためであり、ここでは、書き込み内容のテキスト長に応じた補正秒数の算出を行い、書き込み時刻の補正をする。具体的には、以下の式によって番組中のイベント発生時刻 T_e を算出する。

$$T_e = T'_v - \delta t$$

$$\delta t = \alpha \cdot l + t_r$$

ここで、 T'_v は書き込み時刻、 l は書き込み文字数、 α は 1 文字入力に必要な時間を表す定数、 t_r はイベント発生から書き込みを開始するまでの平均時間である。

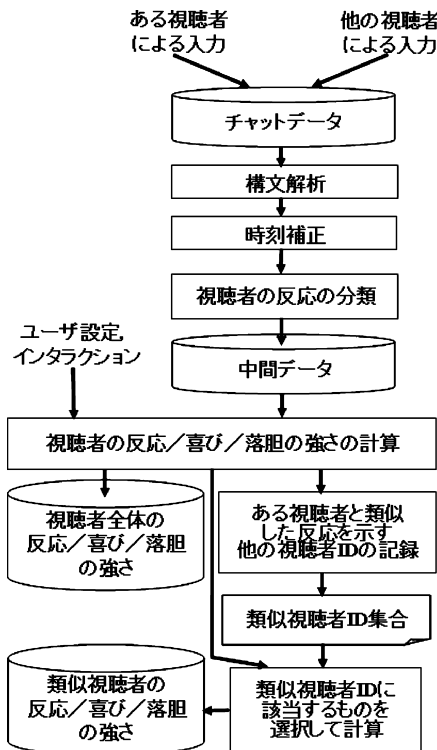


図 6 メタデータの抽出手順

Fig. 6 Block diagram of extracting metadata.

α および t_r は通常経験的にあらかじめ与えられる。

次に、アスキーアートによる特徴的な表現を検出することで、視聴者による各書き込み内容が、喜び、落胆、その他のいずれの感情に分類されるかを判定する。実際には、書き込み内容に出現する「ナイス」、「すげー」、「凄い」、「すげえ」、「キター」といった喜びを表す単語や、「あーあ」、「アホ」、「ださださ」、「だせえ」、「なんだかなあ」といった落胆を表す単語を、感情を表現する典型的なフレーズと考え、アスキーアートや典型的なフレーズのパターンを perl の正規表現でヒューリスティックに記述し、それらを組み合わせたパターンマッチングにより喜び/落胆を表すテキストを検出した。各書き込み内容において、検出されたテキスト数の最も大きかった感情（喜び/落胆）を分類結果とし、いずれの感情を表すテキストも検出されなかった場合は、その他に分類した。

図 7 に中間データの例を示す。Data タグ中の st 属性は時刻補正されたイベント発生時刻を表し、id 属性は書き込み者の ID を、class 属性は書き込み内容の分類（ここでは、喜び=1、落胆=-1、その他=0）を表す。Data タグの要素は書き込み内容に対応する。

次に、各単位時間ごとの視聴者の反応/喜び/落胆の

```
<?xml version="1.0" encoding="Shift_JIS" ?>
<InterimData>
  <VideoFile>050124nfl.wmv</VideoFile>
  ...
  <Data st="3590" id="ROJGrftU" class="1">喜ぶ</Data>
  <Data st="3591" id="WBK6JfKS" class="0">うわ</Data>
  <Data st="3593" id="hRXOKHVI" class="0">とめた</Data>
  <Data st="3593" id="7IfKoKl6" class="-1">止められた</Data>
  <Data st="3694" id="DPZQjvAx" class="0">いつもの展開だ</Data>
  ...
</InterimData>
```

図 7 中間データの例

Fig. 7 Example of interim data.

大きさ度数を計算する。視聴者の反応は、各単位時間ごとの書き込みエントリ数を計測することにより計算され、喜び/落胆の大きさ度数は、中間データの class 属性値がそれぞれ喜び/落胆に一致した各単位時間ごとのエントリ数を計測することにより計算される。ここでは、書き込みエントリ数をそのまま度数とする方法、および、その番組内で最大となる書き込みエントリ数で正規化した数値を度数とする方法の 2 種類で記録した。また、単位時間はユーザ設定やインタラクションによって必要に応じて適宜設定できる。

また、視聴者全体の情報を利用するだけでなく、自分と嗜好が似通った書き込みを行う視聴者（類似視聴者）を束ねて集団として扱い、これらの情報をメタデータ抽出に反映させる。これは、視聴者の嗜好・興味に沿ったダイジェストなどを生成する際に有効であると考えられる。

同一単位時間内に視聴者 A の反応と似た反応（喜び/落胆を含む）を示す、A 以外の他の視聴者 B の書き込み者 ID を、類似視聴者 ID として記録する。類似度 S の算出方法は様々な方法が考えられるが、本稿では次のように行った。

step1. 時刻 $T = 0$ において $S = 0$

step2. 時刻 T から単位時間 δT の時区間における視聴者 A の反応 R_A （ここでは、喜び=1、落胆=-1、その他=0 のいずれか）を次のように求める。同時区間における視聴者 A の書き込み内容の class 属性 c の頻度を $N_A(c)$ とする。0（その他）以外の c について、
If ($\max(N_A(c)) = 0$) then $\{ R_A = 0 \}$
else $\{ R_A = \operatorname{argmax}(N_A(c)) \}$
ここで、 $\max(f(x))$ は $f(x)$ の最大を返す関数、 $\operatorname{argmax}(f(x))$ は $f(x)$ を最大とする x を返す関数である。視聴者 B の反応 R_B も同様に求める。

step3. 視聴者 A と視聴者 B の反応が

- 同じ反応である場合（ここでは、 $(R_A, R_B) = (1, 1)$ or $(-1, -1)$ の場合）、 $S = S + 1$
- 互いに逆の反応と見なせる場合（ここでは、 $(R_A, R_B) = (1, -1)$ or $(-1, 1)$ の場合）、

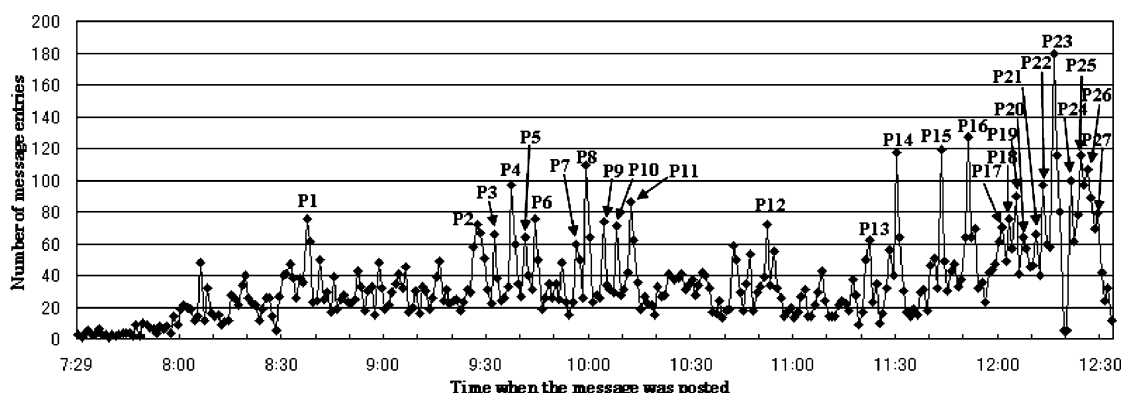


図 9 生成されたメタデータと視聴者の反応との比較 (抜粋)

Fig. 9 Comparison of extracted metadata with responses by viewers (excerpt).

```

<?xml version="1.0" encoding="Shift_JIS" ?>
- <Chat>
  <VideoFile>050124nfl.wmv</VideoFile>
  <SimilarViewersID type="numOfId">3</SimilarViewersID>
  <SimilarViewersID type="Id">WBK6JfK8</SimilarViewersID>
  <SimilarViewersID type="Id">JUqD6rFF</SimilarViewersID>
  <SimilarViewersID type="Id">DPZQjvAX</SimilarViewersID>
  :
  <Intensity st="3580" dur="5" response="2" happiness="0" depression="0" />
  <Intensity st="3585" dur="5" response="3" happiness="0" depression="0" />
  <Intensity st="3590" dur="5" response="5" happiness="1" depression="1" />
  <Intensity st="3595" dur="5" response="10" happiness="3" depression="0" />
  <Intensity st="3600" dur="5" response="5" happiness="0" depression="0" />
  :
</Chat>

```

図 8 自動抽出されたメタデータの例

Fig. 8 Example of automatically extracted metadata.

$$S = S - 1$$

step4. If ($T < \text{終了時刻}$) then $\{T = T + \delta T, \text{goto step1}\}$

step5. A 以外のすべての視聴者の中で、最も高い S の値を示す上位 m 名 (m はあらかじめ与えられた人数) を類似視聴者と判断する。

類似視聴者 ID は、ID の集合としてグループ管理され、その視聴者による書き込み入力があり新たな類似視聴者 ID が検出された場合は、適宜グループに追加される。その視聴者による書き込み入力のない番組 (未視聴番組など) の場合は、これまでに累積した類似視聴者 ID を持つ書き込み内容のみを対象として、反応/喜び/落胆の強さを計算する。

以上のような処理の結果、以下のような数値の組からなるメタデータが生成される (図 8)。

- (1) 視聴者 (書き込み者) 全体について
 - 計測時刻, 計測単位時間
 - 反応/喜び/落胆の強さ
- (2) 類似視聴者 (書き込み者) について
 - 類似視聴者 ID の集合
 - 計測時刻, 計測単位時間

● 反応/喜び/落胆の強さ

図 8 において、Intensity タグは視聴者全体あるいは類似視聴者が示す反応の強さを表し、属性 st は計測時刻、属性 dur は計測単位時間、属性 response, happiness, depression はそれぞれ計測期間における反応/喜び/落胆の強さを表す。また、VideoFile タグは対象番組の映像ファイル名を表し、SimilarViewersID タグは類似視聴者 ID の集合を指定する際に用いられる。SimilarViewersID タグの属性 type の値 numOfId は類似視聴者の総数を、値 Id は各類似視聴者の ID をそれぞれ表す。

4. 実験および考察

4.1 実験 1

生成されたメタデータが、どの程度番組の盛り上がりなどを表現できているのかを調べる実験を行った。本実験では、番組実況チャットに書き込みをした人を「視聴者」と呼び、番組例としてアメリカンフットボールの中継試合を用いることとする。

まず、それぞれ違うチームを応援するアメフト好きの被験者 3 名に、番組中、各自にとって重要な映像区間を記録してもらう。具体的には、コメントしたいと思うところ、盛り上がったところ、落胆したところの映像区間を記録する。この記録結果と、実際の番組実況チャットから抽出した視聴者全体の反応の大きさを示すメタデータとを比較した結果を図 9 に示す。被験者の 3 名が記録した映像区間はインターセプトやタッチダウンといったシーンであり、応援チームに関係なく重要視されることが多かった。また、3 名が記録した映像区間のうちの多くは、番組実況チャットから抽出されたメタデータのピークともかなり一致していることを確認した (図 9)。たとえば、図 9 におい

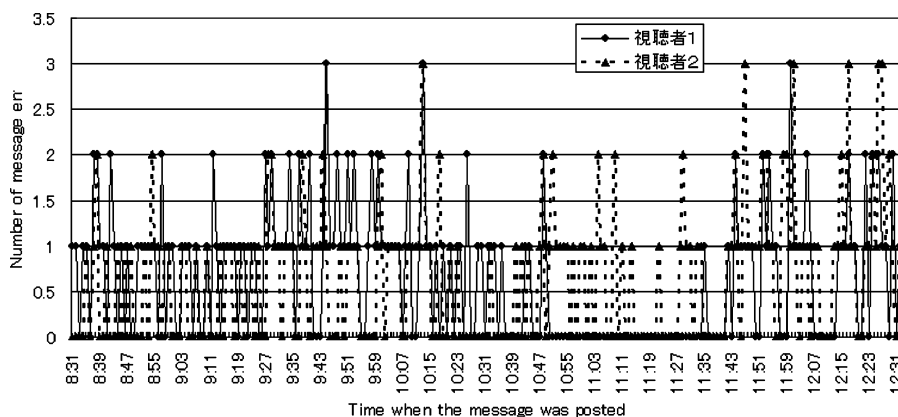


図 10 異なる視聴者の反応の比較 (抜粋)

Fig. 10 Comparison of responses from different viewers (excerpt).

て、P1～P27 は抽出メタデータのある一定以上の大きさのピークを表す。ここで、P2 はジャンブル、P4 はフィールドゴールのブロック、P6、P8、P9、P13、P14、P16、P19、P23 はタッチダウン、P15 はインターセプトに対応しており、これらのプレーはいずれもアメフトの試合では重要な場面である。これらの場面は被験者の 3 名が共通して記録した映像区間となっており、これより、抽出メタデータのピークと被験者の記録した映像区間が、重要な場面において特に一致していることが分かる。なお、利用した番組実況チャットのデータには、前記 3 名の被験者の書き込みは存在していない。

また、異なるチームを応援する視聴者の反応の大きさを示すメタデータどうしを比較した例を図 10 に示す。ここで異なるチームを応援する視聴者は、実際の番組実況チャット中から選択した 2 名であり、前記アメフト好きの 3 名の中の 2 名ではない。図中、視聴者 1 と視聴者 2 は 12:00 前後でともに同程度の強い反応を示しているが、視聴者 1 は 9:40 前後で、視聴者 2 は 10:10 前後で、それぞれ異なる程度の反応を示している。今後、より詳細な評価を行う必要はあるが、視聴者個別の反応を示すメタデータが、視聴者の嗜好の違いをある程度反映した形で番組内の盛り上がりを表現できていることが確認できる。

4.2 実験 2

また、喜び/落ち込み表現の検出精度について調べた。ここで、正解データは、実際の番組実況チャット全体のデータ（視聴者数：360 名、視聴時間：約 3 時間 50 分）を、チャットをよく利用する第三者に見てもらい、各エントリが喜び、落胆、どちらでもない、のいずれに分類されるかを主観的に判断してもらうことで作成した。また、アスキーアートや典型的なフレー

表 1 喜び/落胆表現の検出結果

Table 1 Results of detecting enjoyment and disappointment expressions.

	適合率	再現率	F-measure
喜び	0.989 (376/380)	0.900 (376/418)	0.942
落胆	0.812 (125/154)	0.839 (125/149)	0.825

ズのパターンを表す正規表現は、喜びについて 19 種類、落胆について 23 種類用意した。

表 1 に検出結果を示す。喜びの表現では、再現率が 0.900 となり、正解データの中から 10% の漏れがある可能性があるが、適合率が 0.989 とほぼ 99% に達しているため、検出結果の誤りが結果全体に及ぼす影響はきわめて小さいと考えることができ、結果的に、非常に良好な検出結果が得られたと考えられる。これより、喜びの表現の検出は、今回の実験で用いた比較的単純なパターンマッチングによる手法でも十分に高い精度で検出できることが分かる。検出誤りの例としては、普通の人と違った意味で解釈してアスキーアートが使われた場合、喜び表現や悲しみ表現など複数のアスキーアートから構成されていた場合に見られた。また、検出もれは、アスキーアートが使われない言葉のみによる表現の場合に見られた。落ち込み表現の検出精度が比較的低くなったのは、落ち込み表現が、喜び表現よりも複雑な感情に基づき直接的でない表現が多かったためであると考えられる。今後は前後の文脈などを考慮して精度を上げるためのパターンマッチングの改善を図る必要がある。今回の実験で検出できた感情は、比較的その感情レベルが大きなものであると考えられる。驚きや怒り、恐れなど感情の種類が多くなった場合の抽出手法については今後の課題である。



図 11 反応の大きさの時間順によるビュー

Fig. 11 Example of chronological views of viewers' responses.

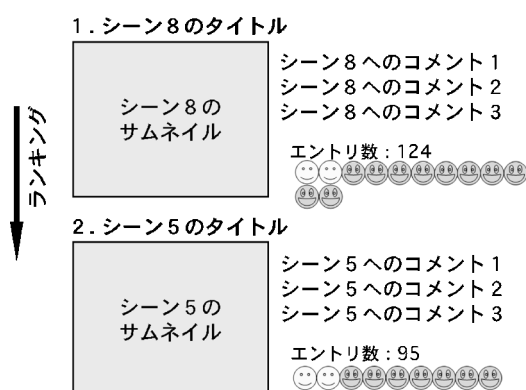


図 12 反応の大きさのランキングによるビュー

Fig. 12 Example of ranking views of viewers' responses.

以上より、抽出メタデータが視聴者の嗜好の違いを効果的に反映していることを確認した。

4.3 実装システム 1

さて、提案手法で得られるメタデータは、実際の書き込みデータと併用することで、録画番組を様々な視点や見方によって視聴するシステムに応用できる。表示インターフェースとしては、種々の方式が考えられるが、ここでは、ストーリーボード形式のインターフェースを用いて説明する。

まず、様々なシーンに対する時間順およびランキング順によるビューが可能になる。ここでシーンとは、適当な分割手段を介して得られる番組の部分映像という意味である。以下の図 11~14 では、各区分=1 分として分割された映像をシーンとして説明する。

たとえば、各区分に含まれる書き込みエントリの総数から視聴者全体の反応をシーンの出現順に従って表示すると図 11 のようになる。シーンごとに、サムネイルと書き込みコメント、書き込みエントリ数の数値、および、エントリ数の大きさを表した顔アイコンが表示され、それらが縦方向に時間順に配置されている。このビューは、シーンの時間経過に沿った流れを把握する際に役立つ。

また、顔アイコンの種類は、視聴者が示す喜びや落胆、それ以外の感情の違いを表現している。これにより、エントリ全体の中で各感情を持った視聴者がどの程度存在するか大まかに把握することができる。

一方、各区分の書き込みエントリ総数でソートすることにより、視聴者全体の反応が大きかった順に各シーンを並べ替えて表示することができる(図 12)。このビューは、一般的に反応の大きかったシーンをいくつかいつまんで視聴する際に役立つ。特に、このランキングによるビューは、番組実況チャットから視



図 13 視聴者全体の視点を表すビュー

Fig. 13 Example of views based on total viewers' responses.

聴者の反応を得ることではじめて可能になる表示であり、これまでにはない新しい表示法の1つであると考えられる。

また、視聴者全体あるいは類似視聴者 ID に基づくビューが可能である。たとえば、各シーンに対する任意の視聴者のコメントを表示すると図 13 のようになる(すべてのコメントを表示することはあまり現実的でないので、通常はいくつかのコメントを適宜選択することにより表示する)。図中、時間/ランキングは、矢印の向きに時間順あるいはランキング順にシーンが並んでいることを表している。このビューは、視聴者全体の反応を把握する際に有効となる。

一方、類似視聴者 ID に対応する視聴者のコメントのみを表示すると図 14 のようになる。このビューは、自分と同じ嗜好を持つ視聴者の反応に注目して視聴する際に特に有効である。

これらのビューは、それぞれいずれも時間順あるい

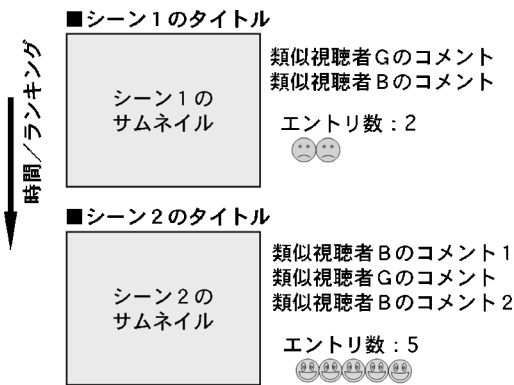


図 14 類似視聴者の視点を表すビュー

Fig. 14 Example of views based on similar viewers' responses.

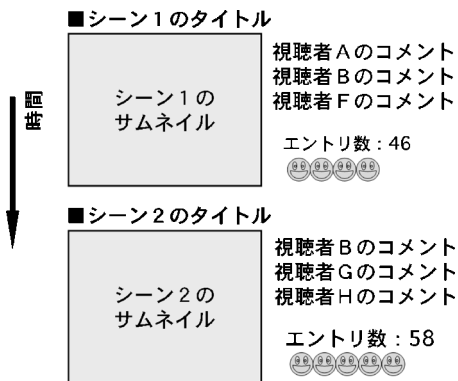


図 15 喜びの反応のみに限定したビュー

Fig. 15 Example of views based on viewer's enjoyment.

はランキング順による表示、また視聴者の喜び・落胆の度合いに基づくビューが可能であり、利用者は状況に応じて様々な見方を選択することが可能となる。たとえば、喜びの反応のみに絞ったビューを生成することが可能である（図 15）。このビューは、視聴者の喜びの感情が表れた部分だけを把握する際に役立つ。

以上で示した全体/類似視聴者のビュー、あるいは、反応/喜び/落胆の度合いに基づくビューについても、番組実況チャットから視聴者の反応を得られることではじめて可能になる表示であり、これまでにはない新しい表示法の 1 つになると考えられる。

さらに、これらのビューを応用すると様々なインタラクションや提示を利用者に提供できると考えられる。たとえば、番組実況チャットのメタデータや書き込みデータと並行して、スポーツの試合などにおけるスコア情報を利用することにより、特定のチームが点を入れた際にだけ喜びの感情度数が高い視聴者や、チームに関係なく点が入った際に喜びの感情度数が高い視聴

者を抽出することができる。これにより、たとえば、A チームを応援している視聴者の視点、中立の立場で視聴している視聴者の視点を表すビューを提供することができると考えられる。

また、特定の選手名が含まれており、その書き込み内容の感情度数が高いものを抽出することができる。この情報を利用することで、ある選手について強い思い入れを持つ人を推定し、熱狂的なファンの視点に基づくビューを提供することができると考えられる。

迷惑な書き込みをする人がある場合は、書き込み者 ID やそのような迷惑な書き込み内容の特徴を学習した識別器を用いることで、迷惑な特定の人や書き込み内容を検出し適宜表示内容から除くことも可能になる。

図 16、図 17、図 18 に提案方式によるビューを実装した例を一部示す^{*}。

図中、見出しは番組の各トピックの先頭を示す字幕データであり、これに対応するシーンがサムネイルとして表示されている。ここで、字幕データとは、テレビ番組に付随して送信されるクロズドキャプションと呼ばれるテキストデータのことで、特別なデコーダを使うことにより、アナウンサの読み上げ原稿などのテキストデータを時刻データとともに実時間で取得することができる。ここでは、この字幕データを用いて番組映像のトピックへの分割を行った。具体的には、字幕データ中の単語分布に基づく統計的手法により分割を行った¹⁾。この手法は、学習に基づく手法と異なり、トレーニングデータを必要とせず、任意のドメインで利用できるという利点がある。映像のカット検出などと組み合わせる分割方法も考えられるが、本実装では、字幕データのトピック分割のみによって映像を分割した。

各トピックの映像区間は、同様の分割を行うことでサブトピック単位の映像区間にさらに分割される。サムネイル右に表示されたリンクのアンカーテキストは各サブトピックの先頭を示す字幕データに対応しており、このリンクをクリックすることにより、対応する映像がその先頭から全画面で再生されるようになっていく。

さらに、サブトピック単位の字幕データが表示されたリンクの下部に、該当するトピック区間における視聴者の反応や感情度数が表示されており、感情を表す顔アイコンとともに実況チャットの書き込み内容が選択的に表示されている。ここで、反応や感情の強さは、

^{*} 図 16～18 は、2004 年 10 月 19 日に NHK BS1 で放映された番組「NFL・アメリカンフットボール」に基づき処理された例である。

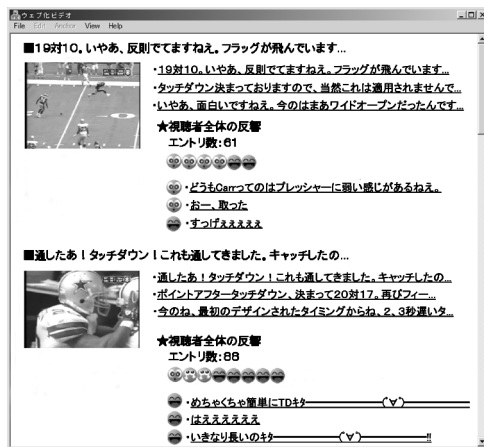


図 16 視聴者全体の反応を表す時間順ビュー☆

Fig. 16 Example of chronological views of total viewers' responses.

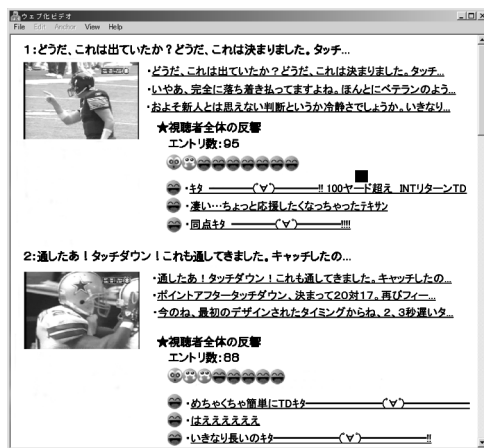


図 17 視聴者全体の反応を表すランギングビュー☆

Fig. 17 Example of ranking views of total viewers' responses.

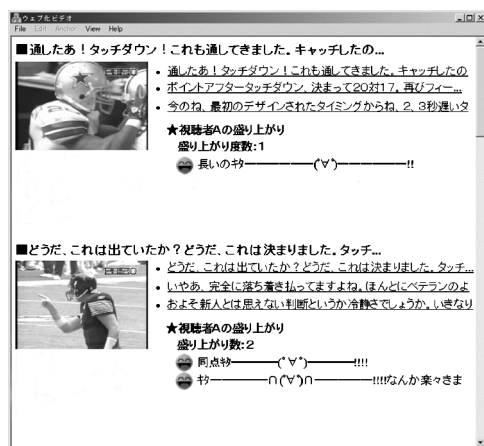


図 18 喜びの度合いに基づく時間順ビュー☆

Fig. 18 Example of views based on viewers' enjoyment.

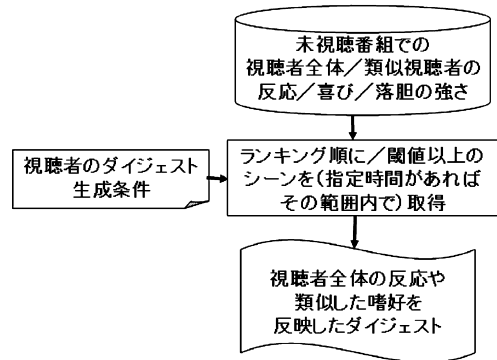


図 19 視聴者の反応を利用したダイジェスト生成手順

Fig. 19 Digest generation using viewers' responses.

各トピック区間に含まれる視聴者の反応や感情の強さの和として表示されており、書き込み内容は反応や感情の強さの最も強い区間から適当に選択され表示されている。

4.4 実装システム 2

本稿では、ビューの表示インタフェースとして、ストーリーボード形式をベースに説明を進めてきた。しかし、ダイジェストを視聴する場合などは、通常のテレビ画面形式のインタフェースも有効である。たとえば、図 19 の手順で、視聴者全体の反応、あるいは、各利用者に対してその利用者と類似した興味・嗜好を反映したダイジェストを生成することができる。視聴者のダイジェスト生成条件としては、ダイジェスト全体の時間、重要シーンと判断する閾値（視聴者反応の強さなど）、反応/喜び/落胆の指定といった条件が考えられる。

たとえば、本稿では、ダイジェスト全体の時間を指定せず、視聴者反応から生成するという条件を与えた場合、時間 t_d 秒間あたりに含まれる視聴者反応の強さの和が、与えられた閾値 θ_d を超えている場合に、対応する映像区間がダイジェストの 1 映像区間として選択されるという方法で生成している。

このダイジェスト生成を実装した例を図 20～22 に示す☆☆。図 20 は、インターセプトなど試合中の重要なプレーの例である。一方、図 21 や 図 22 は、珍しいグッズの紹介やチアガールの登場シーンなど、試合内容には直接関係がない部分映像だが、視聴者の反応が大きいために、ダイジェストとして選択された例である。このような例は、従来の字幕や音響、映像などのマルチモーダルな解析手法では得られないシーンで

☆☆ 図 20～22 は、2004 年 10 月 19 日に NHK BS1 で放映された番組「NFL・アメリカンフットボール」に基づき処理された例である。



図 20 ダイジェスト生成の例 (1)☆☆

Fig. 20 Example of digest generation (1).



図 21 ダイジェスト生成の例 (2)☆☆

Fig. 21 Example of digest generation (2).



図 22 ダイジェスト生成の例 (3)☆☆

Fig. 22 Example of digest generation (3).

あり、提案手法は比較的簡単な手法であっても、視聴者の視点を反映した映像を高い精度で抽出できることを示している。

なお、本稿で紹介した2つの実装例はC++で開発され、Windows 2000 SP4, CPU Pentium4 2.8GHz, メインメモリ 512 MB, グラフィックスコントローラ NVIDIA Quadro2 EX の PC 上で動作している。

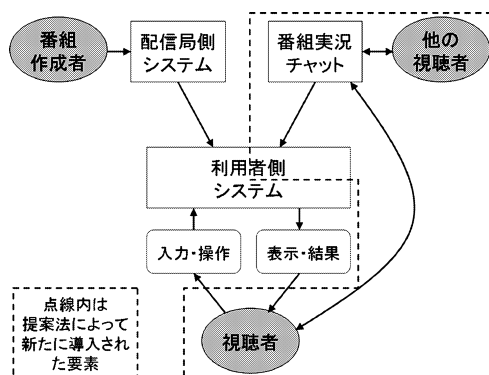


図 23 提案方式の概念

Fig. 23 Concept of our method.

4.5 考 察

提案手法は、スポーツだけでなく、様々なジャンルの番組に適用することが可能である。しかしながら、番組実況チャットは視聴者が自由にメッセージをやりとりする場であるため、しばしば番組内容とは関係のない話題で多くのメッセージがやりとりされる場合がある。本稿で用いたスポーツの実況番組の場合、比較的攻守がはっきりしている場合が多く、少なくとも試合が進行している間は番組内容に沿った盛り上がりを検出できる。しかし、ドラマなどの場合は、登場人物の好き嫌いによって、特定の俳優のゴシップで盛り上がるなど、番組の話の流れとは関係のない話題で盛り上がりが検出されることがある。より詳細な評価は今後の課題だが、番組のジャンルによって、盛り上がりの検出結果には定性的な違いが出てくることが考えられる。

番組実況チャットを利用して抽出されるメタデータの最大の利点は、メタデータに視聴者の反応といった新たな要素を含めることができた点である。図 23 は、提案法により従来の番組視聴の形態には存在していなかった視聴者の反応という新たな要素が導入されたことを示している。

これにより、

- 視聴者による反応を利用したランキングによる提示
- 視聴者の反応の大きさや喜び・落胆の度合いを共有した視聴
- 自分と同じ（あるいは異なる）興味を示す他の視聴者の価値観の発見・共有

といった様々な付加価値を持つ視聴が実現できるようになると考えられる。

視聴者の反応を利用したランキングは、1つの番組だけでなく、他の番組やある一定期間に放送された番

組をまとめた複数の番組に対して行うことができる。提案手法で得られるメタデータは、たとえば、この1週間での視聴者の反応や喜びの感情度数が大きかったシーンのランキングといった提示に応用することができると思われる。

また、提案手法で得られるメタデータは、リアルタイム視聴および録画後の視聴のいずれの視聴形態にも適用可能である。テレビの実況中継を見ながら実況チャットを生で見る使い方や、録画テレビ映像と録画実況チャットを、後から様々なビューを設定して楽しむという使い方が考えられる。また、テレビ(HDレコーダ)だけでなく、PCや携帯電話といった様々な機器を用いた利用形態が考えられる。

さて、本稿では、番組実況チャットの基本的なデータである書き込み時刻、書き込み者のID、書き込み内容に基づいたメタデータ抽出を提案した。これに加えて、番組実況チャットに参加している視聴者プロフィールが利用できると、さらに多様な番組視聴が可能になると予想される。たとえば、男性・女性による視点、年齢による世代別の視点、関東や関西といった地域別の視点、日本、韓国、中国、米英仏独といった国別の視点などを反映させたビューなどが実現できると考えられる。こうした利用スタイルが受け入れられると、いわゆる2ちゃんねるなどの掲示板だけでなくいろいろな番組実況チャットコミュニティができていくのではないかと期待される。

本稿では、番組実況チャットのみを利用したが、性別、年齢、地域、国籍などを記述した視聴者プロフィールも含めた形で視聴者情報を積極的に流通・管理・利用可能な仕組みができることになると、より多様な番組視聴が可能になると考えられる。なお、この件については、プライバシー情報の保護など多くの問題がある。

5. 関連研究

従来、映像のメタデータ抽出の研究は活発に行われてきた。たとえば、色を用いた例として、色ヒストグラムや特徴的な色の集合、隣接する特徴的な色対などを利用した手法が提案されている²⁾。これにより、問合せで指定された画像中の物体が、動画中に存在するかどうかの判定などが可能になる。また、動きを用いた例としては、動きベクトルを用いた手法が提案されている³⁾。ブロックマッチング法などで得られた動きベクトルを、ズームやパンなどを特徴づけるパラメータでモデル化し分類することにより、各カメラ操作の有無や強さの判定が可能になる。

さらに、マルチモーダルなメタデータ抽出の例とし

て、シーン分割、カメラ操作検出、顔検出、映像キャプション検出などを利用して、シーンの重要度を判定・選択する手法⁴⁾や、テキスト解析で得られたキーセンテンスと画像解析で得られたキー画像を、動的プログラミングを使って関連付ける手法⁵⁾が提案されている。マルチモーダルな情報に加えて、特定分野のドメイン知識を利用することにより、画像中の人物動作などの、より詳細な内容に関するメタデータ抽出が可能であることも示されている^{6)~8)}。

しかし、これら従来手法が利用してきたデータは、番組の配信側(放送局)から提供される情報のみであったため、抽出されたメタデータは、基本的に、番組作成者・配信側の意図に基づく画一的なものであった。よって、従来手法においては、シーン探索や内容の早見、ダイジェスト視聴などの内容に、自分以外の他の視聴者の視点や反応といった要素を取り入れることはできなかった。提案手法では、番組実況チャットを利用することで、放送番組に対する視聴者の視点や反応を反映したメタデータ抽出が可能となり、これを利用することで番組を様々な見方で視聴できるようになると考えられる。

チャットからテレビ番組のメタデータを生成する試みとして文献 9)、10) が報告されている。

文献 9) では、チャットの書き込み内容を形態素解析して得られる固有名詞や形容詞など特定の品詞の単語が、視聴者の感想を反映したメタデータとして利用できることを示している。しかし、十数名程度の少人数の被験者によるチャットに基づいた結果となっており、多くの視聴者が参加する実際のチャットにおける有効性が十分に示されているとはいえない。

また、文献 10) では、ドラマ番組に関する2ちゃんねるの実況チャットを対象として、書き込み内容からあるインターバルで連続的に出現する人物名などのキーワードを検出することで、視聴者の嗜好をある程度反映したグラフ構造が生成できることを示している。しかし、ドラマに出現する人物名などのキーワード候補リストをあらかじめ用意することを前提としており、検出される視聴者の反応が、ドメインに特化したキーワード候補リストに限定されるという問題点がある。

提案手法では、2ちゃんねるの実況チャットを対象とし、汎用的な特徴(視聴者の書き込み数や視聴者ID、喜びと落胆という比較的大まかな感情を表す典型的なフレーズ、および、アスキーアート)に着目することで、ドメインに特化されにくい、より適用範囲の広いメタデータ抽出法を実現していると考えられる。

また、膨大な映像データから興味のある部分を効率

良く探したり、内容の概要を手軽に理解したり、限られた時間でハイライトだけを見るといった多様な視聴方法についての研究も従来活発に行われてきた。

たとえば、映像コーパスから興味のあるショットを探索するために視覚的要素を利用したインタフェースが提案されている¹¹⁾。映像セグメントを表す複数のキーフレームから構成されるストーリーボードと視覚的特徴によるフィルタリング機能を組み合わせることで効率的な探索が可能だとしている。

また、番組概要の把握やシーン探索を効率良く行うための表示インタフェースとして TV2Web¹²⁾ や Video Manga¹³⁾ が提案されている。TV2Web では、映像セグメントのサムネイルと字幕テキストから構成されるストーリーボードを異なる詳細度レベル（セグメント、シーン、ショット）で構造化し、これら構造データをズームメタファによってスムーズにつなぎながら（切り替えながら）表示することで効率的な概要把握やシーン探索が可能だとしている。Video Manga では、クラスタリングされた映像セグメントが長くて稀なものの重要度を高く計算し、その値に応じて表示するキーフレームの大きさを制御する。異なるサイズのキーフレームをマンガのような表示形態に配置しサマリとして表示することで、より重要な場面を視覚的に把握しやすいインタフェースを実現している。

限られた時間でハイライトを視聴するために、テレビ番組のためのダイジェスト視聴システムが提案されている¹⁴⁾。番組映像中に配信側で用意された汎用的な番組索引と検索用ルールが含まれることを前提とし、これらを利用して利用者の嗜好を考慮したハイライトをダイジェスト視聴することができる。

これらの手法は、いずれもキーフレームや重要な映像セグメントの時間順による提示を基本としていた。一方、提案手法は、このような時間順によるビューだけでなく、他の視聴者による喜びや反応を反映した番組シーンのランキングビューといった提示を可能としている。このような見せ方は従来手法には存在しておらず、提案手法によって初めて可能になるものであると考えられる。

6. ま と め

本稿では、番組実況チャットに基づく視聴者反応を利用したテレビ番組のメタデータ抽出手法を提案した。番組実況チャットのデータを、利用者全体あるいは特定の傾向を持つ利用者のデータ集合として扱い、アスキーアートなどの特徴的な表現を認識処理することで、番組の盛り上がり場面や、視聴者の嗜好・興味に

沿ったリアクションなどを効率良く抽出することができる。これら抽出メタデータを利用することで、シーンのランキング表示やダイジェスト視聴など様々な視聴が可能となった。これらはいずれも、視聴者による反応を利用することで初めて可能になったものであり、従来手法にはない新しい付加価値をもたらすメリットがある。

今後は、精度を上げるためのパターンマッチングの改良を行う必要がある。また、性別、年齢、地域、国籍などを記述した視聴者プロフィールを積極的に流通・管理・利用可能なサイトの構築などを行うことによって、提案手法がより多様な番組視聴に適用可能であることを実証することができると期待される。

参 考 文 献

- 1) Utiyama, M. and Isahara, H.: A Statistical Model for Domain-Independent Text Segmentation, *ACL/EACL-2001*, pp.491-498 (2001).
- 2) 長坂晃朗, 田中 譲: カラービデオ映像における自動索引付け法と物体探索法, 情報処理学会論文誌, Vol.33, No.4, pp.543-550 (1992).
- 3) Akutsu, A., Tonomura, Y., Hashimoto, H. and Ohba, Y.: Video indexing using motion vectors, *SPIE Proc. VCIP '92*, pp.522-530 (1992).
- 4) Smith, M. and Kanade, T.: Video Skimming and Characterization through the Combination of Image and Language Understanding Techniques, *CVPR* (1997).
- 5) Nakamura, Y. and Kanade, T.: Semantic analysis for video contents extraction — spotting by association in news video, *ACM Multimedia*, pp.393-401 (1997).
- 6) Intille, S.S. and Bobick, A.F.: Closed-world tracking, *Proc. 5th International Conference on Computer Vision*, pp.672-678 (1995).
- 7) Saur, D.D., Tan, Y.-P., Kulkarni, S.R. and Ramadge, P.J.: Automated analysis and annotation of basketball video, *Storage and Retrieval for Image and Video Databases V*, SPIE-3022, pp.167-187 (1997).
- 8) Miyamori, H.: Automatic annotation of tennis action for content-based retrieval by integrated audio and visual information, *CIVR2003*, LNCS2728, pp.331-341 (2003).
- 9) 大黒泰平, 加藤友規, 土居清之, 亀山 渉: 視聴者入力情報からの時系列キーワード抽出システムの試作, 情報処理学会第 67 回全国大会, pp.3-5, 3-6 (2005).
- 10) 上原 宏, 吉田健一: インターネット上の対話文に基づくドラマ番組の構造化—注目状態グラフによる視聴者による視聴者コミュニティの嗜好パ

ターン認識, 信学技報パターン認識・メディア理解, Vol.104, No.369, *PRMU2004-87*, pp.25-30 (2004).

- 11) Christel, M.G. and Huang, C.: Enhanced access to digital video through visually rich interfaces, *ICME*, MD-L5.1 (2003).
- 12) Sumiya, K., Munisamy, M. and Tanaka, K.: TV2Web: Generating and browsing web with multiple LOD from video streams and their metadata, *ICKS2004*, pp.158-167 (2004).
- 13) Uchihashi, S., Foote, J., Girgensohn, A. and Boreczky, J.: Video Manga: Generating semantically meaningful video summaries, *Proc. ACM Multimedia 99* (1999).
- 14) 橋本隆子, 白田由香利, 真野博子, 飯沢篤志: TV受信端末におけるダイジェスト視聴システム, 情報処理学会論文誌: データベース, Vol.41, No.SIG3 (TOD6), pp.71-84 (2000).

(平成 17 年 6 月 20 日受付)

(平成 17 年 10 月 14 日採録)

(担当編集委員 池田 哲夫)



宮森 恒 (正会員)

独立行政法人情報通信研究機構 (NICT) メディアインタラクショングループ研究員。1997 年早稲田大学大学院後期博士課程修了。1996～1997 年同大学理工学部助手。工学博士。主に、マルチメディアコンテンツ処理に関する研究に従事。電子情報通信学会, 映像情報メディア学会, 日本データベース学会各会員。



中村 聡史

独立行政法人情報通信研究機構 (NICT) メディアインタラクショングループ専攻研究員。2004 年大阪大学大学院後期博士課程修了。工学博士。主に、ヒューマンインタフェースに関する研究に従事。日本データベース学会会員。



田中 克己 (正会員)

京都大学大学院情報学研究科社会情報学専攻教授。1976 年京都大学大学院前期博士課程修了。工学博士。主にデータベース, マルチメディアコンテンツ処理の研究に従事。IEEE Computer Society, ACM, 人工知能学会, 日本ソフトウェア科学会, 日本データベース学会各会員。