

ライブ番組向けダイジェスト視聴サービスのためのリアルタイムメタデータ生成技術

スポーツ中継など、ライブ配信される番組を対象として番組配信中にシーン検索やダイジェスト視聴等の高度視聴サービスを提供するために必要なリアルタイムメタデータ生成技術を紹介します。メタデータを利用することで、家庭のTVやモバイル端末等、さまざまな視聴環境に適したライブ番組のダイジェスト視聴サービスが可能になります。

くわの ひでたか こんや ゆうこ
 桑野 秀豪 / 紺家 裕子
 やまだ ともかず かわぞえ かつひこ
 山田 智一 / 川添 雄彦

NTTサイバーソリューション研究所

メタデータを利用したライブ番組のダイジェスト視聴サービス

放送通信連携サービスにおけるメタデータを利用した新しいTV視聴方法の1つに番組のダイジェスト視聴があります。番組中の複数の注目シーンを連続的に再生視聴するダイジェスト視聴は、好きなときに、好きな番組の内容を短時間に把握でき、時間に余裕のない現代社会の生活スタイルにマッチした視聴方法です。家庭でのTV視聴時はもちろん、特に、時間に制約のあ

る外出先等でのモバイル環境においては、よりユーザーズにマッチした番組視聴方法と考えられます。

ダイジェスト視聴を実現するためには、番組中のシーン内容に関するメタデータを用意する必要があります。NTTサイバーソリューション研究所では、これまでに映像・音声認識、言語処理といったメディア解析技術を利用して半自動的にメタデータを生成する技術の研究開発に取り組んできました⁽¹⁾。ここでは、制作済みの番組映像を対象としたメタデータ生成の作業

フローを紹介しました。本稿では、メタデータを利用した番組視聴サービスの適用範囲の拡大のため、新たに、ライブ放送される番組を対象として、番組放送中にリアルタイムにメタデータを生成する技術、作業フローを紹介します。リアルタイムメタデータ生成技術は、メディア流通システムの6種類の放送通信連携サービス⁽²⁾、すべてのサービスの発信側において適用可能な技術です。

野球中継を例として、メディア流通システム上に実装したライブ番組のダ

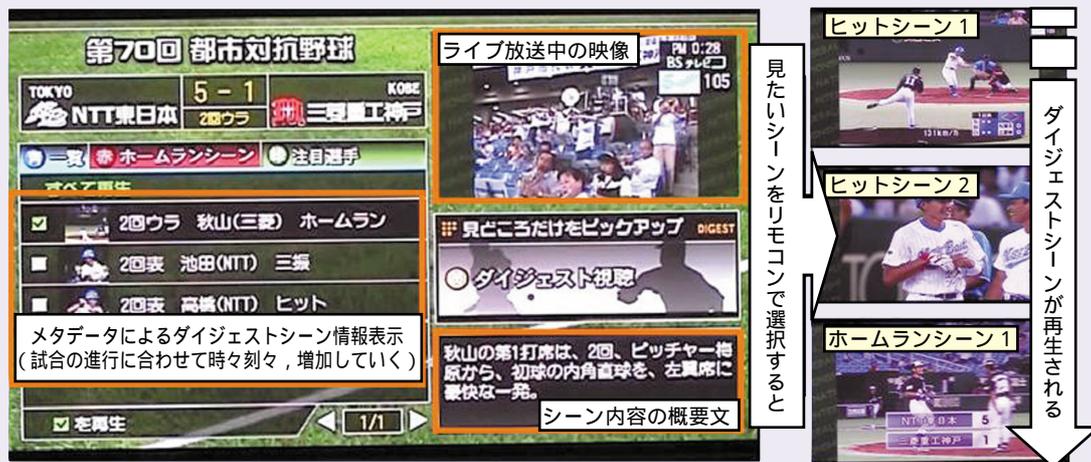


図1 メタデータを利用したライブ野球中継のダイジェスト視聴サービスのTV画面例

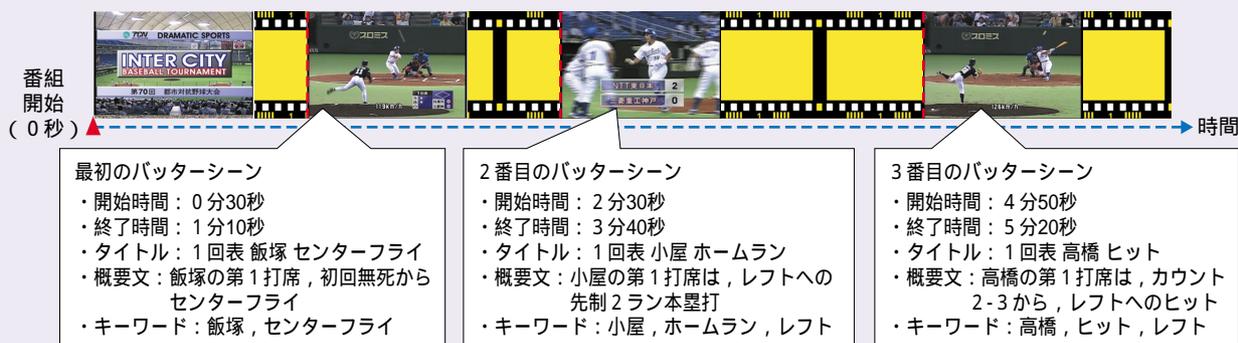


図2 野球中継のダイジェスト視聴サービスに必要なメタデータの例

イジェスト視聴サービスのデモ画面例を図1に示します。TV画面には、番組の進行に合わせて、ダイジェスト視聴用のシーンに関するメタデータが時々刻々増えて表示されていくようなイメージになります。視聴者は見たいシーンに関する情報を選択するとそのシーンのリプレイ映像が視聴できます。野球中継での好きな選手のシーンなどの他、ニュース番組中の注目トピック等、ライブ番組中の注目シーンを見逃してしまった場合やもう1度見たい場合に、好きなタイミングで簡単に該当シーンをリプレイ視聴することが可能になり、TV視聴時間を有効に活用できます。

次にこのような新しいITV視聴サービスを実現するための基盤技術となるリアルタイムメタデータ生成技術を紹介します。

ライブ番組向けのメタデータ生成の課題

ライブ番組の例として野球中継番組のダイジェスト視聴に必要なメタデータ項目を図2に示します。番組中の個々のシーン区間の開始時間、終了時間、およびシーン内容に関するタイトル、概要文、キーワードがメタデータ

項目になります。これらの項目はTV Anytime Forum⁽³⁾においてもメタデータの国際標準仕様として規定されています。

ライブ番組のダイジェスト視聴サービスを実現するためには、メタデータを番組進行中にリアルタイムにつくり出し、すぐに視聴者端末に向けて送り出す必要があります。メタデータの生成作業をすべて人手で行うと膨大な手間が必要となり、作業の疲労感も大きくなります。特に、シーン内容に関するタイトル、概要文、キーワード等のテキスト情報については、番組台本のような既存テキスト情報が存在すれば、それを有効活用して効率的にメタデータとして生成できますが、スポーツ番組のように事前に筋書きの読めない番組については、実際にシーンが終わった段階で、初めてシーンのタイトル等の表現内容を検討し、書き起こす必要があります。この作業を手作業だけで進めていくと、疲労感が大きいだけでなく、作業時間に余裕がなくなり、情報の入力ミスも生じやすくなるのが想定されます。そのため、ライブ番組の進行に遅れることなく、かつ、作業者の疲労感も少なくすむメタデータ生成の作業フローを確立することが重要

になります。

音声認識をベースとしたメタデータ生成の作業フロー

野球中継のライブ放送を対象として、メタデータを番組進行に遅れることなく高速に生成していくための作業フローを策定しました。策定にあたっては、メディア解析技術、ユーザインタフェース技術を最大限活用し、手作業の分量が少なくすみ、かつ作業者は野球中継の状況からなるべく目を離さずにメタデータ生成の作業が効率的に進めていけることを意識しました。この考え方に基づき、私たちは特に音声認識技術を積極的に活用した作業フローを策定しました(図3)。また策定した作業フローを実現するためのシステムとして「ライブメタデータ生成システム」を開発しました。システムの画面例を図4に示します。

作業フローとしては、まず、ホームランやヒットなどのシーン終了時点で、メタデータ生成の作業者がシーン内容を説明する概要文を音声認識しやすい条件下(周囲の雑音が少ない環境)で発話します。概要文はシーン内容を説明する文章であり、TV画面の限られた表示スペースに表示される文字

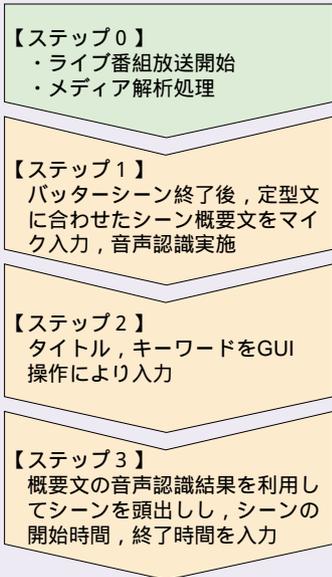


図3 ライブ番組向けのメタデータ生成作業フロー



図4 ライブメタデータ生成システムの画面例

情報として利用されるため、文字数を考慮して設定する必要があります。そのため、定型フォーマットに合わせた文章として発話することとしました(図5)。定型文としての音声認識処理は、認識結果に対する制約条件を設けることができることから、自由発話の音声認識処理に比べ高い認識率を得ることができ、シーン概要文の定型化が想定可能な番組のメタデータ生成に有効な方法です。実際のシステム操作としては、作業者が発話音声システムに対し、マイク入力すると、音声認識された結果が図4右側の音声認識結果ブラウザに表示されます。音声認識結果ブラウザ上のテキストを必要に応じて修正し、図4左下のメタデータエディタに設定することで、シーンの概要文が完成します。

シーン内容のタイトル、キーワードもメタデータエディタ上で入力します。タイトルは選手名やホームラン、ヒットなどの打席結果をラジオボタンなどの選択式GUI(Graphical User

Interface)を使って入力します。キーワードに関しては、メタデータエディタ上の要約ボタンを押すと、言語処理が実行され、タイトルや概要文としてすでに設定されているテキストから自動的に抽出される単語が得られ、これがキーワードとして設定されます。

シーンの開始時間、終了時間の設定には、前述の音声認識結果ブラウザのほか、図4中央のキー画像ブラウザ、および図4左上の再生モニタを利用します。音声認識結果ブラウザ上の音声

認識結果のテキストやキー画像ブラウザ上に表示される映像認識結果の代表画像を選択すると、選択されたそれらに対応する映像中のシーン画像が再生モニタ上に表示されます。前述のように概要文の音声認識はメタデータ入力対象のシーンの終了直後に実施されますので、音声認識結果ブラウザ上のテキストを選択すると再生モニタ上で対応するシーンの頭出しができます。その後、キー画像ブラウザ上の代表画像や再生モニタ上の±10秒、±1秒

【シーン概要文のフォーマット】
(バッター名)(打席数)(インニング)(アウトカウント)(ボールカウント)(対戦ピッチャー名)(結果)
発話ルール: この順で発話する。全項目が必須ではない。

【シーン概要文のサンプル】
「池葉の第一打席は、左翼席への先制2ラン本塁打」
「高橋の第一打席は、カウント2-3から、三遊間を抜けるヒット」
「秋山の第一打席は2回、ピッチャー梅原から、初球の内角ストレートをレフトスタンドに豪快な一発」

図5 野球番組のバッターシーンに対する定型概要文の例

などのボタンで時間を微調整することでメタデータ入力対象のシーンの開始時間、終了時間が設定できます。

以上の作業フローによる、ライブ番組のダイジェスト視聴用のメタデータ生成作業では、タイトルなどのテキスト情報の入力を1からキーボードなどをタイプしていく必要はなく、野球中継の状況を見ながら、シーン内容に関して概要文を発話したり、簡単なGUI操作だけで作成することが可能となります。またシーンの開始時間、終了時間もメディア解析処理の結果を利用して効率的に設定することができます。

メタデータ生成の効率化実験

野球中継のすべてのバッターシーンを対象にメタデータを生成する作業について、ライブメタデータ生成システムを用いる場合とすべて手作業で行う場合の作業効率の比較実験を行いました。

実験は複数の作業員によって行われ、それぞれの作業モデルごとの作業効率を評価しました。

実験の結果、すべて手作業で行う場合は、メタデータの入力作業に時間がかかり、番組の進行ペースに追いつかない結果となりました。具体的には、10人くらいのバッターシーンが終わった段階で5人分くらいのメタデータしか入力できない状況でした。これに対し、ライブメタデータ生成システムを利用すると、おおよそ次のバッターシーンが終わる段階で、前のバッターシーンに関するメタデータを入力し終えることができ、番組進行のペースに遅れることなく、メタデータを生成していくことができました。

また作業員からのアンケートによると、作業の疲労感に関して、総じて手作業だけのほうが「疲れた」「大変だ

った」との意見が得られ、ライブメタデータ生成システムを利用した場合は「大変ではなかった」との意見が得られました。特に、概要文の入力作業に関しては、ライブメタデータ生成システムを利用した場合は、野球中継の状況から目を離さず、音声を入力するだけで実施できるため、手作業に比べ、非常に効率的に実施できたとの意見が得られました。さらに、ライブメタデータ生成システムを利用すると、音声認識結果を手がかりにシーンの頭出しが簡単にできるため、シーンの開始時間、終了時間を探す作業が非常に簡単に行えたとの意見が多く得られました。再生モニタの±1秒、±10秒といったジャンプボタンも非常に有効であったとの意見もありました。

メタデータ生成技術の今後の展開

本稿ではライブ番組中の見逃したシーン、あるいはもう1度見たいシーンを放送中の好きなタイミングで簡単にダイジェスト視聴するサービスのイメージ、およびサービスに必要なメタデータの生成技術を紹介しました。野球中継の場合、特にシーンの概要文を「定型文に対する音声認識処理」という高精度の認識処理が実施できる方法を利用することで、番組の進行に遅れることなく、メタデータを生成することができることを確認しました。

放送通信連携サービス向けのメタデータ生成は参考文献(1)で紹介した制作済みの番組向けや本稿で紹介したライブ番組向けでは作業フローが異なります。それ以外にドラマ、バラエティといった番組のジャンルの違い、メディア解析技術等、自動化のための技術の到達レベル、また番組台本などの再利用できる情報が事前にどこまでそろっているか等の条件によっても最適

なメタデータ生成の作業フローは変わり得る性質を持つものです。

今後は、メタデータ生成の自動化技術の高度化の検討のほか、番組ジャンル等の条件に合わせて最適な作業フローが効率的に策定できるような作業フロー策定モデルを検討し、作業フロー策定の段階も含めてトータルでメタデータ生成の一層のコスト経済化を図れるような研究開発を進めていきたいと思っております。

参考文献

- (1) 桑野・松尾・川添：“映像・音声認識、言語処理の適用による経済化メタデータ生成技術,” NTT技術ジャーナル, Vol.16, No.5, pp.22-25, 2004.
- (2) 川添・端山：“メディア流通システム,” NTT技術ジャーナル, Vol.17, No.6, pp.8-9, 2005.
- (3) <http://www.tv-anytime.org/>



(後列左から) 山田 智一 / 川添 雄彦
(前列左から) 紺家 裕子 / 桑野 秀豪

固定・携帯連携サービス上のメタデータを利用した新しいTV視聴サービスとサービスに必要なメタデータ生成のより一層のコスト削減を実現するための研究開発に取り組んでいきます。

問い合わせ先

NTTサイバーソリューション研究所
第一推進プロジェクト
TEL 046-859-2508
FAX 046-855-3495
E-mail kuwano.hidetaka@lab.ntt.co.jp