

情報統合機能を実装したメタデータエディタ

A Metadata Editor with Information Integration

佐野 雅規† 住吉 英樹† 八木 伸行†
Masanori Sano Hideki Sumiyoshi Nobuyuki Yagi

1. まえがき

近年メタデータの重要性が高まるなか、我々ではできるだけ効率的にメタデータを制作する手法の研究を進めており、これまでにメタデータ制作のためのフレームワークを提案した[1]。また、このフレームワークに基づいたシステムを構築し、2005年5月のNHK放送技術研究所の一般公開で展示した[2]。本稿では、このシステムの一部である映像/音声/自然言語など多様な情報を統合処理し、目的のメタデータを制作するメタデータエディタについて報告する。

これまでにも複数の情報を扱う、マルチモーダル解析によるイベント検出の重要性が指摘されている[3]が、我々は今回開発したエディタに時間方向の処理も含めたマルチモーダル統合処理機能を実装し、サッカー番組で検証を行った。マルチモーダル統合により、良好にイベントが検出できることを確認したので報告する。

2. フレームワーク内でのエディタの位置づけ

メタデータ制作フレームワーク[1]では、番組のメタデータを蓄積/管理するメタデータサーバを中心に、抽出/編集/利用の様々なモジュールを接続して制作環境を形成する。

図1は、このフレームワークに基づいて構築した制作システム[2]の全体図である。抽出モジュールは、それぞれが対象とするデータ(映像/音声/テキスト)から、メタデータに有用な情報を抽出する。ここで抽出モジュールからの出力をプリミティブメタデータ(PM)と呼ぶ。編集モジュールに相当するメタデータエディタは、これら抽出された複数のプリミティブメタデータを読み込み、その情報を統合して確度の高いメタデータ(統合メタデータ)を制作する。この場合、プリミティブメタデータの種類や統合の方法はいくつも考えられ、汎用性をどう確保してエディタに実装するかは問題となる。本稿では、この点を考慮して開発した統合処理の手法について説明する。

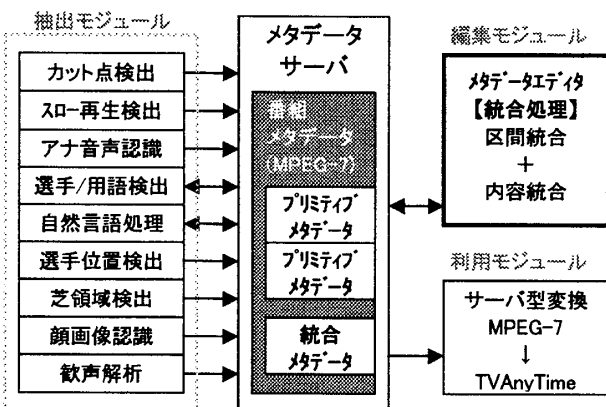


図1 メタデータ制作システムの全体図

† NHK放送技術研究所(知能情報処理)

3. 統合処理の方法

3.1 区間統合と内容統合

メタデータ制作フレームワークでは、MPEG-7 Part 9 CDPをベースにしたメタデータの表現モデルを規定している。プリミティブメタデータと統合メタデータの表現は同じであり、区間(シーン)を示す時間情報と、その区間の内容情報(意味的記述)の2つが主な要素である。このことからエディタにおける統合処理は、注目するプリミティブメタデータの時刻情報と内容情報を使って、統合メタデータの時刻情報と内容情報を作り出す処理と言える。そこで、時刻情報をつくりだす区間統合処理と、内容情報をつくりだす内容統合処理の2つに分けて実装することとした。

この2つの統合処理について、図2を使って説明する。横軸は時間を表し、各プリミティブメタデータA,B,C,Dが図のように存在すると仮定する。この例における区間統合処理は、プリミティブメタデータA,Bの時刻情報を用い、オーバーラップする区間を統合メタデータTの区間として処理している。内容統合処理は、作成された各統合メタデータTx(x=1,2)についてそれぞれ行われる。図では各統合メタデータTxの直後に存在するプリミティブメタデータCと、Txの区間内に存在するDの内容情報を用い、それらを連結させて各統合メタデータTxの内容情報を作成している。この例では、区間/内容統合とも1回の処理で完結しているが、実際はいくつでも処理を重ねることができ、また利用するプリミティブメタデータも複数でなく、1種類でも良いこととした。

3.2 統合処理のための基本操作関数

図2の統合例では、区間統合処理において、時刻がオーバーラップする区間を出力するという操作があり、内容統合処理では、2つの内容情報を連結するという操作があった。このような区間統合と内容統合の具体的な処理は、いくつかの基本操作関数として外部に用意しておき、エディタ上で組み合わせて目的の統合処理を実現するよう設計した。表1に用意した基本操作関数を示す。特に、区間統合に関しては、これらの関数を組み合わせることで、複数プリミティブメタデータの複雑な時間関係を表現することができ

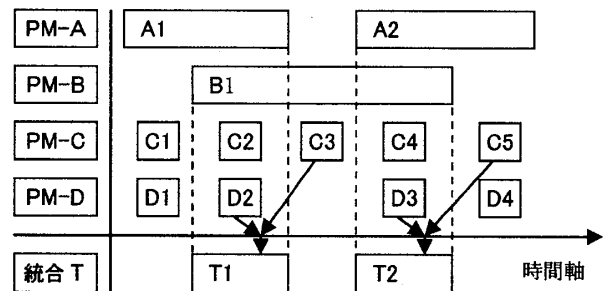


図2 統合処理の例

る。これは時間軸を考慮したイベントテンプレートに相当する。

4. 統合処理の検証実験とその結果

2004年12月11日放送のサッカー天皇杯を対象に検証実験を行った。表1に示す基本操作関数①~⑧を組み合わせ、ダイジェスト用メタデータ、シュート、コーナーキック、得点シーンなどについてそれぞれ区間統合と内容統合のルールを作成した。以下、3つの例を説明する。

a) ダイジェストシーン抽出 (図3参照)

区間統合処理は、歓声が沸いた区間に前10秒、後5秒を付け加える時間指定による伸縮操作①を設定した。内容統合については、統合メタデータの区間に少しでも重なる言語解析結果(誰が何をしたという情報)と、後60秒までの区間に存在する顔認識結果の論理積⑥⑧を設定した。

この統合処理では、特筆すべき何かが起こったシーンが抽出できており、ダイジェスト候補としてふさわしい結果となった。内容記述については、言語解析と顔認識の情報を用い、両者から同じ情報が出力されていれば確からしいと判断し、情報を自動的に出力した。また、言語解析では発生したプレーだけが抽出され、顔認識からそのプレーに関係した選手の情報が出力されている場合もあった。開発したエディタは、このような候補の情報も同時に表示しており、クリック1つで統合メタデータに反映できるように設計してある。

表1 実装した基本操作関数

区間統合用	内容統合用
① 時間指定による伸縮操作	⑥ 論理演算 (1対1)
② 論理演算による操作	⑦ 新規内容追加
③ 重複関係による操作	⑧ MPEG-7記述子によるフィルタ
④ キーワードによる操作	
⑤ 間隔指定による統合操作	

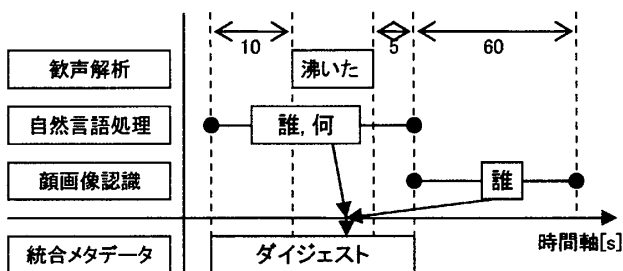


図3 ダイジェスト用メタデータ生成

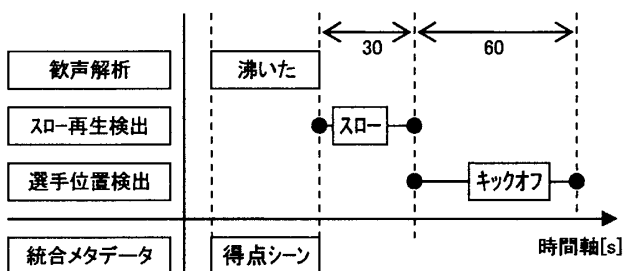


図4 得点シーン抽出

b) 得点シーン抽出 (図4参照)

この統合処理では、歓声が沸いた後30秒以内にスロ再生があって、かつ30~90秒の間にキックオフらしき選手位置が検出された場合、その歓声が沸いた部分は得点シーンであるというルールを基本操作関数②③を組み合わせで作成した。内容統合については、新規内容追加で得点シーンを付与する操作⑦⑧を設定をした。

実験素材には得点シーンが1つ存在し、その部分だけが間違いなく検出された。これは各プリミティブメタデータの時間方向の存在関係を利用した検出といえる。また、音声認識や言語処理によらず得点シーンが特定できていることにも意味がある。

c) シュートシーン抽出

区間統合処理として、まずアナウンサーのコメント中に「シュート」が存在する区間として、選手/用語検出の結果からキーワード操作④により抽出する。この結果の中で隣り合う区間が1秒以内の場合は、間隔指定⑤の操作によりそれらを1つの区間にまとめた。次に、その区間がスロ再生でないものだけを②③の操作で選択し、更に直後5秒以内に選手のアップ(顔認識結果)が存在する区間を③の操作により残した。これによりシュートというコメントでは16ヶ所あった区間が、正しく精査され4ヶ所となった。

一方、アナウンスコメントを自然言語解析した結果がシュートシーンである区間は5ヶ所あった。統合メタデータの区間としては、この5ヶ所と前者の結果4ヶ所の論理和②を設定した。内容統合にはシュートシーンを付与する操作⑦⑧を設定した。

この実験では、最終的に6ヶ所のシュートシーンが抽出された。これは実際のシュートシーンと完全に一致しており、情報を組み合わせることで情報補完ができた例である。

5. まとめ

本報告では、複数の情報を統合してメタデータを作成するメタデータエディタと、その試用実験について述べた。本エディタの統合処理は、小さな処理単位を基本操作関数として外部に用意し、これらを組み合わせるものであり、柔軟性・拡張性が高いことを確認した。また、検証実験から、用意した基本操作関数の組み合わせで、目的とした統合メタデータが生成可能であることを確認した。今回は1試合だけで実験を行ったので、多くの試合で検証する必要がある。また、統合ルールについてもデータを見ながら人手で作成したので、今後は機械学習などを用いて自動的にルールを生成する研究を進めていく。

参考文献

[1] 佐野ほか, “メタデータエディタの試作:メタデータ制作活用プラットフォームの提案,” 信学技報, PRMU2004-146, Dec. 2004
 [2] 住吉ほか, “メタデータ制作・活用システムの試作,” 2005年映情学年大, 11-3, 2005
 [3] 馬場口登, “長時間映像メディアのセマンティック処理,” 信学総大, DT-3-4, 2005