

アナウンスコメントを利用したサッカー番組の メタデータ自動生成の検討

Meta-Data Generation for Football Games using Announcer's Commentary

山田 一郎†
Ichiro Yamada

佐野雅規†
Masanori Sano

住吉英樹†
Hideki Sumiyoshi

柴田正啓‡
Masahiro Shibata

1. はじめに

見たい番組や情報をいつでもすぐに取り出して視聴することができるサーバ型放送サービス[1]の準備が進められている。このサービスでは、放送局から送られる番組を説明するデータ（以後、メタデータ）を利用することにより、番組のハイライトシーンを選択して見せるダイジェスト視聴や、テレビの視聴履歴情報を参照したお薦め番組の提示が構想されている。このため、効率的なメタデータ付与が重要な課題となっている。

そこで我々は、番組とメタデータを同時に制作できる手法を開発してきた[2]。この制作手法を用いれば、取材、編集の過程を経る番組については、制作者が作成した正確なメタデータを作成することが可能である。しかし、スポーツなどの、生中継番組には同様の手法を用いることはできない。本論文では、生中継番組の一つであるサッカー番組を対象とし、アナウンサーや解説者がサッカー番組中で発した言葉であるコメント（以後、アナウンスコメント）を利用して、試合中に発生したイベントをメタデータとして自動抽出する手法を提案する。

2. サッカー番組アナウンスコメントの特徴

本研究は、音声認識で自動取得可能なアナウンスコメントからのメタデータ抽出を最終目標としている。今回、その最初のステップとして、アナウンサーと解説者が話した言葉を人手により書き起こしたテキストデータを処理対象とした。書き起こしデータには、アナウンスコメントの他に、発話者（アナウンサー、解説者、レポーターなど）、コメントの開始時間、終了時間（放送開始からのフレーム数）の情報が含まれている。

サッカー番組におけるアナウンスコメントの特徴を、放送されたサッカー番組1試合を対象として調査した。比較対象としてバラエティ番組「ためしてガッテン」のアナウンスコメント1番組分、ニュース番組1日分を利用した。文節数、異なり単語数、発話者に関する調査の結果を表1に示す。

| | 1文の文節数 | 異なり単語数 (1000語中) |
|-------|--------|--------------------|
| サッカー | 4.6文節 | 136語 |
| バラエティ | 6.1文節 | 213語 |
| ニュース | 13.3文節 | 308語 |

表1 アナウンスコメントの特徴

サッカー番組は、1文の文節数、および、異なり単語数が、他の番組と比較して少ない。このことから、サッカー

番組のアナウンスコメントは、スペースネスの問題が他より発生しにくく、統計処理に適した特徴を持つと考えることができる。

サッカーにおいて重要なイベントである「シュート」と「ファウル」を対象として、その単語がアナウンスコメントに出現した割合を調査した。結果を表2に示す。

| | 適合率 | 再現率 |
|------|---------------|---------------|
| シュート | 31/64 (48.4%) | 31/43 (72.1%) |
| ファウル | 14/19 (73.7%) | 14/42 (33.3%) |

表2 イベントがアナウンスコメントに出現した割合

調査対象とした試合のアナウンスコメント中では、「シュート」という言葉は、64回出現した。そのうち、映像中で実際にイベント「シュート」が発生していたのは31回であった（表2「適合率」）。この結果から、単純なキーワードマッチング処理だけではイベント抽出は難しいことがわかる。また、「シュート」イベントが発生していても、アナウンスコメント中には言及されていないケースも27.9%存在していた（表2「再現率」）。イベント「ファウル」については、3分の2が言及されなかった。「間接フリーキック」となるような軽微な反則については、アナウンスコメントでは、ほとんど言及されていない。この数字は、アナウンスコメントを解析してイベントを抽出する本手法における限界の再現率を示す。

3. メタデータ生成処理

本手法では、まずアナウンスコメントを解析して、各文を2種類に分類し、次に、分類結果を利用してアナウンスコメントを時間的に連続する一連のプレー（以後シーンと呼ぶ）に対応したコメント文の集合に分割し、最後に、シーン中に出現したパスやシュートなどのイベントを抽出する。以下に各処理について説明する。

3.1 アナウンスコメント分類

サッカー番組におけるアナウンスコメントは、次の2種類の文に分類することができる。

- ① 試合記述文：試合状況をリアルタイムに説明したコメント文で、ボールに絡んでいる選手を中心に言及しているもの
- ② 解説文：試合の状況には直接関係しないコメント文で、終わったばかりのプレーなどへの解説や、付加的内容に言及しているもの

アナウンスコメントを解析して、各文が上記のいずれに属するかを判定する。判定には、2つのクラスに分類するための機械学習手法であるサポートベクターマシン[3]（以後、SVMと呼ぶ）を利用する。カーネル関数として2次の多項式関数を利用し、素性は、アナウンスコメント中の以下の特徴とした。

†NHK放送技術研究所（知能情報処理）

‡NHK技術局開発センター（ニュース・番組技術）

- ① 最終形態素の動詞（または名詞）の表記
- ② 最終形態素の格
- ③ 最終形態素を修飾する人名の有無
- ④ 最終形態素を修飾する人名の格
- ⑤ 最終形態素を修飾する人名以外の表記
- ⑥ 最終形態素を修飾する人名以外の格
- ⑦ 最終形態素を修飾する動詞句の表記
- ⑧ 接続詞、間投助詞、接続助詞、終助詞の有無
- ⑨ 助詞（丁寧、完了、過去、断定、打消）の有無
- ⑩ 解説者・アナウンサーの名前の出現の有無
- ⑪ 発話者(アナウンサー、それ以外)

NHK で実際に放送されたサッカー番組 6 試合のアナウンスコメントを対象とし、人手により対象とする全てのアナウンスコメントに対して「試合記述文」か「解説文」を判断した正解データを与え、実験を行った。SVM による判定処理では、学習データを 5 試合、テストデータを 1 試合として 6 通りの交差検定を行った。表 3 に結果を示す。

| | 適合率 | 再現率 |
|-------|------------------|------------------|
| 試合記述文 | 3601/4108(87.7%) | 3601/4252(84.7%) |
| 解説文 | 5442/6093(89.3%) | 5442/5949(91.5%) |

表 3 「試合記述文」「解説文」判定結果

表 3 では、適合率が、試合記述文、解説文ともに、80% を超しており、それらの調和平均を表す F 値は、試合記述文が 86.2%、解説文が 90.4% という結果が得られており、手法の有効性が確認できた。

3.2 シーン抽出

アナウンスコメントから以下の 2 項目を手掛かりとして、サッカーの試合におけるプレーの切れ目を判定する。

- ① 試合の状況には直接関係しないコメント
- ② 一定時間コメント無し

「解説文」とは、試合の状況には直接関係しないコメントであるため、①は解説文の有無により判断する。この際、プレーの途中に短時間の解説文が出現することがあるため、2 文以上連続して発話された解説文を、プレーの切れ目とする。②は、アナウンスコメントの開始時間、終了時間を利用して判断する。

この処理により、時系列に並ぶアナウンスコメント列に対して、その切れ目となるセグメント境界が挿入される。このセグメント境界間にある試合記述文集合に対応する番組映像を、時間的に連続し関連の深いプレーの集まりである一つのシーンと判定する。

3.3 イベント抽出

サッカーでは、「シュート」などの特有のイベントが出現する。これらのイベントを、あらかじめ手作業により登録し、各シーンにこの内のどのイベントが含まれているかの情報を抽出する。今回、「シュート」「ゴール」「ファウル」「イエローカード」「フリーキック」「コーナーキック」「オフサイド」の 7 つのイベントを登録した。

これらのイベントが、アナウンスコメント中に存在しても、実際にそのイベントが映像中に発生しているとは限らない。映像中で発生したイベントのみを抽出するため、「試合記述文」と「解説文」の判定結果を利用する。試合記述文は、実際に映像で発生している事を映像と同期してコメントしたものであるため、試合記述文に出現するイベ

ントは、実際に発生しているイベントと判断できる。そこで、3.2 で抽出したシーン中の試合記述文から、キーワードマッチングによりイベントを抽出した。

サッカー番組 1 試合のアナウンスコメントを対象とし、実験で得られた結果を、人手により抽出したイベント情報と比較評価した。その結果を表 4 に示す。

| | 適合率 | 再現率 |
|---------|--------------|--------------|
| シュート | 28/30(93.3%) | 28/43(65.1%) |
| ゴール | 2/2(100%) | 2/2(100%) |
| ファウル | 6/6(100%) | 6/42(14.3%) |
| イエローカード | 5/5(100%) | 5/5(100%) |
| フリーキック | 6/8(87.5%) | 6/45(13.3%) |
| コーナーキック | 3/6(50.0%) | 3/12(25.0%) |
| オフサイド | 2/3(66.7%) | 2/3(66.7%) |

表 4 イベント抽出検証結果

イベント「シュート」は、適合率 93.3%、「ゴール」「ファウル」「イエローカード」は適合率 100%、表 4 の項目を平均すると、52/60(86.7%)と高い適合率を得ることができた。しかし、イベント「ファウル」「フリーキック」「コーナーキック」の再現率は低い結果であった。7 つのイベントの再現率を平均すると、52/152(34.2%)であった。これは、実際にイベントが発生した際に、そのイベントを表す単語がアナウンスコメントに出現する割合が低いことが原因となっている。アナウンスコメントに出現する割合が、本手法の再現率の限界値となるため、もともとアナウンスコメントとして取り上げられにくいイベントは再現率が低い結果となった。

4. まとめ

本稿では、サッカー番組のアナウンスコメントからサッカー特有のイベントとそれらを捉えたシーンを抽出する手法を提案した。アナウンスコメントの分類では、F 値が試合記述文で 86.2%、解説文で 90.4%と良好な結果が得られた。またイベント抽出では、特定イベントに対するの平均適合率は 86.7%であったが、平均再現率は 34.2%と低い値であった。この問題に対処するためには、自然言語処理によるアプローチだけでなく、画像処理のアプローチなどとの併用が必要と考えられる。そこで我々は、現在、映像や音声、字幕などの複数の情報を統合して、番組に対してメタデータを付与するメタデータエディタ[4]の開発を進めている。

今回の評価では、実際にアナウンサーが発した言葉を、人手により書き起こして実験を行ったが、現在は、音声認識結果への適用も進めており、放送に付けられた字幕の利用も含めて、提案手法による、効率的なメタデータ生成の有用性が高まっていくことが期待できる。

【参考文献】

- [1] 栗岡ほか: "サーバ型放送", 映情学誌, Vo.58, No.5, pp.647-650, 2004
- [2] 住吉ほか: 階層化番組制作手法における番組情報の構造, 信学会総合大会, SMD-2, 1999
- [3] Nello Cristianini, An Introduction to Support Vector Machines and other kernel-based learning method, Cambridge University Press, 2000
- [4] 佐野ほか: メタデータエディタの開発, 映情学年次大会, 2004