

アナウンスコメントを用いたサッカー番組の実映像 - リプレイ映像の対応付け

酒井 敬一[†] 高村 大也^{††} 奥村 学^{††} 山田 一郎^{†††}

[†] 東京工業大学 知能システム科学専攻 ^{††} 東京工業大学 精密工学研究所 ^{†††} NHK 放送技術研究所
sakai@lr.pi.titech.ac.jp {takamura,oku}@pi.titech.ac.jp yamada.i-hy@nhk.or.jp

1 はじめに

近年, DVD, HDDレコーダーの出現により, 保存できる映像が今までより増加しつつある. 特にスポーツ番組等, 長時間放送される番組から視聴者が希望するシーンのみを観たいという要望がある [1].

スポーツ番組では, 重要なシーンは後になって繰り返し放送される. このリプレイシーン (以降 RS とする) は視聴者が視聴を希望するシーンであると考えられる. また RS は人手により挿入されるため, RS が番組上でいつ始まるのか, どのくらいの長さであるかを知ることが容易である. したがって視聴者は RS を視聴すれば, 視聴を希望するシーンを観ることができると考えられる.

しかし, RS は重要である部分のみに絞って放送するため, その前後の映像まで放送されることが少ない. そのため RS だけでは得られない情報も存在する. また, スローモーションで放送される RS よりも実時間上でのスピーディな映像も見たい視聴者もいるだろう.

そこで, 本研究では RS に対応している実時間軸上のシーン (以降 RTS とする) を特定することにより, RS だけでは得られない情報を補完するための基礎技術を開発する. 今回はスポーツ番組の一例としてサッカーの試合中継番組を対象とする. 本来サッカー番組等のスポーツ番組の解析には, 映像認識処理や音声認識処理, そして自然言語処理など複数の技術を利用して解析し, 得られた情報を組み合わせる必要がある. 本研究では, 自然言語処理を用いてアナウンサーや解説者がサッカー番組中で発したコメント (以降アナウンスコメントとする) を解析して RS に対応する RTS の特定を行う.

しかし, RS 部分に発話されたアナウンスコメントが必ずしも映像に対して述べられているとは限らず, 映像とは関係ない内容についての言及である場合も多い. そこで提案手法では, まず RS に対して言及されたアナウンスコメントの範囲を特定する. そうすることにより RS と無関係なアナウンスコメントを取得することを防ぎ精度の向上を図る. 次に, RS に対して言及されたアナウンスコメントから得られた情報を元に, RS に対応している RTS を特定する.

RS に対しての言及であるアナウンスコメントの範囲の特定では, RS に対しての言及であるアナウンスコメントを連続した一連の固まり (chunk) と考える. そして各アナウンスコメントが chunk のどの位置にあるかを機械学習により学習し, 判定することで本問題を解く手法を提案する. また RS に対応している RTS の特定でも, RS に対応している RTS を chunk として考え, 各アナウンスコメントが chunk のどの位置にあるかを機械学習により学習し, 判定することで本問題を解く.

2 関連研究

サッカー番組のアナウンスコメントを利用して番組の解析を行う研究に山田ら [2] がある. 山田らの研究ではサッカー番組のアナウンスコメントを機械学習を用いて以下の 2 種類に分類している. 1 つはボールタッチしている選手を中心に実況しているコメントである試合記述文であり, もう 1 つは試合の流れとは直接関係しない補足的な解説である解説文である. そして分類された試合記述文からキーワードマッチングを用いて映像に映っているイベントを抽出している. しかし, このキーワードを用いてイベントを抽出する方法では, アナウンスコメントにイベントを示すキーワードが含まれていない場合も多く, 再現率は低いものとなっている. 本研究では, RTS と RS の対応付けを行うことによって情報の補完を行う. そうすることにより, アナウンスコメントにイベントを示すキーワードが含まれないことによる再現率の低下を防ぐことが期待できる.

3 提案手法

3.1 RS に対して言及されたアナウンスコメントの範囲の特定

RS に対して言及されたアナウンスコメントの範囲を特定する手法について述べる. 本問題は対象のアナウンスコメントが RS に対しての言及か否かの 2 値分類問題だと考えられる. そこで本研究では 2 値分類問題に利用される教師付き学習器である SVM を学習器として利用して, リプレイシーンに対して言及されたアナウンスコメントの範囲の特定を行う.

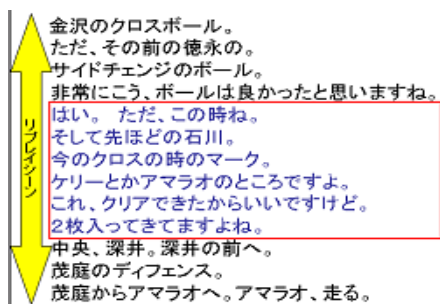


図 1: RS に対して言及されたアナウンスコメント

RS に対しての言及を開始してから終了するまでの間に RS 以外の内容について言及することが少ないことが分かっている。よって図 1 のように RS に対して言及されたアナウンスコメントを一連の固まり (chunk) と考えることができる。本研究では、機械学習によりアナウンスコメントのある文が RS に対してのアナウンスコメントのどの位置 (先頭, 中心, 終わり等) にあるかを学習し, 判定することで本問題を解く。そこで学習データに対して, 対象の 1 文が chunk のどの位置にあるのかをタグ情報として付与する必要がある。本手法では, RS 中のアナウンスコメントの位置を示すタグは IOB2 モデル [3] を使用して表現する。IOB2 モデルは各アナウンスコメントに対して B(このタグの位置は RS に関するアナウンスコメントの先頭), I(このタグの位置は RS に関するアナウンスコメントの一部), O(このタグの位置は RS に関するアナウンスコメント以外) の 3 つのタグを付与することで chunk を表現する。学習には汎用テキストチャンカーの YamCha[4]¹ を用いて, アナウンスコメントの各文に対してどのタグが付与されるかを学習し, 決定する。

使用する素性

学習時に用いる素性の説明を行う。まずテキストの特徴から得られた素性について説明を行う。その次に RS の開始, 終了してからの経過時間などテキスト以外の特徴から得られた素性についての説明をする。

・テキストから得られる素性

Yamcha の学習で用いる素性のうち, テキストの特徴から得られる素性についての説明を行う。統計処理による学習では, 単語の表記そのものを素性として利用すると学習データに出現しない単語がテストデータに頻出するスパースネスの問題が存在する。しかし, 山田ら [2] によるとサッカー番組では使われる単語の種類数は他の番組 (バラエティ番組, ニュース番組) と比べて少なく, ある特定の単語が多用されると考えられる。そのため, 単語の表記を素性として利用してもスパースネスの問題が少ないと考えられる。よって, 主に素性として単語の表記を利用する。これをふまえ, 以下に機械学習に用いた素性を示す。

- 最終文節にある動詞, 名詞, 助詞の表記
- 最終文節を修飾する文節中にある動詞, 助詞の表記
- 接続詞, 間接助詞, 接続助詞, 指示詞の有無
- フィラーの表記
- 終助詞の表記
- 時相名詞の表記

ここでの助詞とは「格助詞一般」と係助詞(「は」「も」)の計 13 種とする。また指示詞, 時相名詞の抽出には Juman²を, それ以外の形態素の抽出には ChaSen³, 最終文節の抽出には CaboCha⁴を用いた。

・テキスト以外の特徴から得られる素性
 Yamcha の学習で用いる素性のうち, テキスト以外の情報から得られる素性を以下に示す。

- アナウンスコメントの話者
- RS 中に発言されたアナウンスコメントかどうか
- RS が開始, 終了してからの経過時間
- 発話の前に沈黙があるか

調査により, RS に対する言及はアナウンスにより開始されることが多いということがわかっており, 対象のアナウンスコメントの話者が誰であるかという素性を付与した。話者がアナウンスなら 0, それ以外の人物なら 1 をとる 2 値の素性とした。また RS に対する言及を開始する前に沈黙が存在することが多いと考えられる。よって沈黙した時間が 2 秒未満あるなら 0, 2 秒以上あるならば 1 として 2 値の素性として追加した。

3.2 RS に対応する RTS の特定

次に RS に対応する RTS を特定するモデルについての説明を行なう。RS に対応する RTS 中のアナウンスコメントも RS に対して言及されたアナウンスコメントと同じく連続した一連の chunk とみなす。そして機械学習により, 対象のアナウンスコメントが RS に対応する RTS のどの位置 (先頭, 中心, 終わり等) にあるかを学習し, 判定することで本問題を解く。本タスクでも, RS に対応する RTS 中のアナウンスコメントの位置を IOB2 タグを利用して付与し, 学習・テストを行なった。

使用する素性

RS に対応する RTS の特定をする際に学習に用いる素性の説明を行う。

・RS に対して言及されたアナウンスコメントから得られる素性

RS に対しての言及から得られる素性についての説明を行う。RS に対して言及されたアナウンスコメントと RTS に対して言及されたアナウンスコメントで共通する単語が含まれていることが多いと考えられる。サッ

¹<http://chasen.org/~taku/software/yamcha/>

²<http://nlp.kuee.kyoto-u/nl-resource/juman.html>

³<http://chasen.naist.jp/hiki/Chasen/>

⁴<http://chasen.naist.jp/~taku/software/cabocha/>

カー中継では、RS と RTS では表現が変わることが多い。しかし、選手名やプレイ名ならば RS と RTS では表現が変わることが少ないと考えられる。したがって RS、RTS 中のアナウンスコメントに共通する名詞、未知語があればその表記を素性として用いた。

・ RTS 自身が持つ特徴から得られる素性

学習に用いた素性のうち RTS 自身が持つ特徴から得られる素性についての説明を行う。具体的に機械学習に用いた素性を以下に示す。

- 最終文節にある名詞、未知語の表記
- 最終文節を修飾する文節中にある名詞、未知語の表記
- 最後の形態素の表記
- 終助詞の表記
- 最後の形態素が選手名か、プレイ名か
- 同じ名前が連続して出現するか
- 1 発話に 3 人以上違う人名が存在するか
- (選手名)の(プレイ名)という表現があるか
- 1 発話中の文の数
- アナウンスコメントの話者
- 発話に聞き取れない部分があるか

RTS のように盛り上がるシーンでは「一気に前線に、そこにエメルソン。」「山田のシュートだ。平野」等のように選手名で文が終了したり、1 文に複数の選手名が存在することが多い。よって、そのような表現がアナウンスコメント中に出現するかという情報も素性として追加した。

ここでプレイ名というのは、RS に対応する RTS のイベントに関連する単語名である。RS に映るイベントに関連する単語はだいたい決まっているので、そういった単語名は素性として有効だと考えられる。そこで人手により、RS で起こるイベントに関連すると思われる単語名を簡単に収集しリストを作成した。ここで「(選手名)の(プレイ名)」という表現は「山田のシュート」や「ジャンのクリア」という表現の有無という 2 値の素性として定義した。

また「発話に聞き取れない部分があるか」という素性について説明する。RS に対応している RTS 中の発話は短く、アナウンサーは早口で説明する必要があるので不明瞭な発話が起りやすい。書き起こしの際、そういった不明瞭な発話がある部分は * で示される。そこでアナウンスコメント中に発話が不明瞭な部分があるなら 1、無いならば 0 をとる 2 値の素性を定義した。

4 実験

4.1 RS に対して言及されたアナウンスコメントの範囲の特定

前節で述べた「RS に対して言及されたアナウンスコメントの範囲」を特定する提案手法の性能を実験により示す。

実験対象とするデータは実際にテレビ番組上で放送されたサッカー番組である。それには、試合中継だけではなくハーフタイムや試合前、試合終了後の放送も含まれる。基本的にサッカー番組においてはアナウンサーと解説者の 2 人が出演して実況と解説を行なう。提案手法で処理するコーパスはこのアナウンサーと解説者の発話を書き起こしたテキストである。正例クラスとして RS に対する言及である文が 808 文、負例クラスとして RS 以外の内容に対する言及を 1911 文を利用して実験を行った。

学習・テストには SVM の結果に基づいてテキストのチャンキングを行う Yamcha を用いて実験を行った。実験では、学習データを 10 試合、テストデータを 1 試合として 1:1 分割の交差検定を行った。学習に使用したカーネル関数は 2 次多項式であり、ソフトマージンパラメータの値は 1.0 とした。また判定する文の後ろ 2 文、前 1 文、そして直前にシステムが付与した 2 文のタグも素性として学習に利用した。評価尺度は、精度、再現率、F 値を用いる。

表 1: RS に対して言及されたアナウンスコメントの範囲の特定の実験結果

実験モデル	再現率	精度	F 値
ベースライン	79.0	53.5	62.6
提案手法	69.4	71.7	70.6

実験の結果を表 1 に示す。ベースラインは RS の間に発話されたアナウンスコメントは全て RS に対する言及であるとした時の結果である。提案手法による実験結果はベースラインより F 値で 8.0 ポイントほど良い。

4.2 RS に対応する RTS の特定

RS に対応する RTS を特定する提案手法の性能を実験により示す。RS の開始の 272 秒前までのアナウンスコメント全 6105 文から 248 文の「RS に対応する RTS」を特定する実験を行った。

判定処理には SVM によりテキストのチャンキングを行う Yamcha を用いて実験を行なった。学習データは 10 試合、テストデータは 1 試合として 1:1 分割の交差検定を行った。学習に使用したカーネル関数は 2 次多項式であり、ソフトマージンパラメータの値は 1.0 とした。また判定する文の後ろ 2 文、前 6 文、そしてシステムが付与した直前の 2 文のタグも素性として学習に利用した。尚、タグを打つ向きは後ろ向き（過去から未来）にタグ付けを行った。評価尺度は、精度、再現率、F 値を用いる。実験の結果を表 2 に示す。ベースライン 1 は選手名、プレイ名のどちらかが RS、RTS 中のコメントに共通して出現している場合、その RTS は RS と対応関係にあるとしたものである。ベースライン 2 は選手名、プレイ名の両方が RS、RTS 中のコ

表 2: RS に対応する RTS の特定の実験結果

実験モデル	再現率	精度	F 値
ベースライン 1	16.9	46.2	24.8
ベースライン 2	50.8	22.2	30.9
提案手法	58.5	83.4	68.4

メントに共通して出現している場合にその RS と RTS は対応関係にあるとしたものである。ベースラインよりも提案手法の性能が大きく上回り、単純な選手名やプレイ名のキーワードマッチによる方法より良い結果が出る事が分かる。

4.3 1つのRSに対して必ず1つのRTSを特定するように制約を課した実験

前節の実験では、システムは1つのRSに対し、RSに対応するRTSを1つも判定しない場合がある。普通RSには対応するRTSが1つは存在するはずである。そこで1つのRSに対して必ず1つのRTSを特定するように制約を課して実験を行う。

もしシステムがRSに対応するRTSを1つも判定しなかった場合、Yamchaで算出されたBタグのスコアのうち最も高かった1文をRSに対応するRTSとした。その様な制約を課して実験を行った結果を表3に示す。

表 3: 必ず1つのRTSを判定した実験結果

実験モデル	再現率	精度	F 値
制約を加えず実験	58.5	83.4	68.4
制約を加えて実験	62.1	77.8	69.1

その結果、制約を加える前と比較してF値が0.7上昇し、良い結果が得られた。

4.4 RSから得られる情報を変えた実験

4.2節の実験では、RSに対して言及されたアナウンスコメントを手で判定し実験を行った。そこで節4.1の実験で特定したRSに対する言及を利用して実験を行う。各実験の設定を以下に示し、表4に各実験結果を示す。

Rerefertag

4.2節での実験結果と同じく、RSに対して言及されたアナウンスコメント範囲を手で判定し実験を行った。

RSresult

4.2節の実験で特定したRSに対して言及されたアナウンスコメントを利用する。

Replayresult

RSが放送されている間に発話された全てのアナウンスコメントを利用する。

表4より、RS中の全てのアナウンスコメントを用いてRTSの特定をするよりも、RSに対しての言及であ

表 4: RSから得られる情報を変えた実験結果

実験モデル	再現率	精度	F 値
Rerefertag	62.1	77.8	69.1
RSresult	58.5	82.4	68.4
Replayresult	57.7	74.9	65.1

るアナウンスコメントの範囲を特定し、その情報を手がかりにRTSの特定を行う方が、F値にして2.7ポイントほど良い結果になった。

よって、まず「RSに対しての言及であるアナウンスコメント」の範囲を特定し、次にRSに対応するRTSを特定する2段階の特定を行う提案手法を用いることにより優れた結果が得られることがわかった。

5 おわりに

本研究では、サッカー番組のRSに対して言及しているアナウンスコメントを機械学習を用いて特定する手法を提案した。提案手法では、RSに対して言及しているアナウンスコメントを一連のchunkとみなし、対象のアナウンスコメントがそのchunkのどの位置になるかを学習することによって、F値70.6%が得られた。

また、特定できたRSに対して言及されたアナウンスコメントから得られた情報を使用して、RSに対応するRTSの特定を行った。提案手法では、同じくRTSに対して言及しているアナウンスコメントを一連のchunkとみなし、RTSの特定を行っている。結果としてF値68.4%が得られた。

以上から、まず「RSに対して言及されたアナウンスコメント」の範囲を特定し、次にRSに対応するRTSの特定を行なう2段階の手法を用いることにより優れた結果が得られることがわかった。今後の課題としては、これまでの自然言語処理を用いた手法に加え、画像処理や音声処理を併用することが挙げられる。

参考文献

- [1] <http://prius.hitachi.co.jp/prius/pc/2005may/iitokomi/index.html>
- [2] 山田一郎 佐野雅規 住吉英樹 柴田雅弘 八木伸行, サッカーと解説者のコメントを利用したサッカー番組のセグメントメタデータ自動作成. 電子情報通信学会文誌 VolJ89-D-2, pp.2328-2337, 2006.
- [3] 工藤拓 松本祐治, Support Vector Machine を用いた Chunk 同定. 情報処理学会自然言語処理研究会 SIGNAL-140, 2000.
- [4] <http://www.chasen.org/~taku/software/Yamcha>