

要約済異種メディアを利用した一般向け映像要約規則の抽出

Extracting of Typical Video Summarization Rules using Heterogeneous Summarized Media

平林 衛 稲葉 大樹 川口 克則 大平 茂輝 白井 克彦
Mamoru HIRABAYASHI Taiki INABA Katsunori KAWAGUCHI Shigeki OHIRA Katsuhiko SHIRAI

早稲田大学理工学部
School of Science and Engineering, Waseda University

In this paper, we propose a method to extract typical video summarization rules using heterogeneous summarized media, such as digest videos, news videos, newspapers, and magazines. The extracted rules are available to make a same type of video summary.

1. はじめに

近年、我々の身近に存在する動画の急増に伴い、映像の要点をまとめた要約映像の自動作成が必要となっている。

機械による要約映像の自動作成は、映像の各イベントに重要度を割り振り、ある閾値以上のシーンを繋ぎ合わせるという方法が最も主流である。しかし、従来の研究は映像とイベントの対応付け(自動インデクシング)に重点を置いており、どのイベントにインデクスを振るか、どこからどこまでが一つのイベントか、そのイベントはどれ程重要であるかについて、確立した指標は見当たらない。

本研究では、ダイジェスト番組やニュース映像、新聞、雑誌などの人手によって作られた要約済異種メディアから、一般性のある要約規則の抽出を試みた。抽出した要約規則を任意の映像に適用することで、同型の要約映像が作成可能となる見通しを得た。

2. 要約規則抽出手法

要約の対象をサッカー中継映像に限定し、要約済異種メディアとして同TV放送局により作成された20分要約映像、5分要約映像、及びニュースサイトの速報記事を使用する。本研究で抽出する情報は、以下の通りである。

1. 重要イベント
2. イベント重要度

要約を行うには、要約対象となるデータの内容記述がなされていることが望ましい。本研究で対象とするサッカー映像においては、試合におけるプレイの流れを把握することがそれに当たる。サッカー映像における重要なイベントの抽出ならびに各イベントの重要度を決定する方法として、ここでは対象の試合に関するニュースサイトの速報記事をみなし要約として利用する方法を提案する。

2.1 TF・IDF法

TF・IDF法[1]は、文書内に含まれる索引語の重み付けとして一般的に利用されている手法である。出現頻度の高い索引語がその文書の特徴付けているというTF法、その文書で見られないような索引語こそが重要であるというIDF法の2つの特徴を兼ね備えている。

文書 d 中に出現する索引語 t の重み W_t^d は、索引語頻度と逆文書頻度の積として式(1)で表される。

$$W_t^d = \frac{tf(t, d)}{\sum_{s \in d} tf(s, d)} \cdot \left\{ \log \frac{N}{df(t)} + 1 \right\} \quad (1)$$

2.2 重要イベント

「茶釜」[2]を用いて速報記事の形態素解析を行い、品詞情報から不要語を削除する。(1)式のTF・IDF法により、残った各単語の重要度を調べる。 t が抽出単語、 d が各試合の速報記事、 N が対象試合数、 $df(t)$ が抽出単語 t が出現する速報記事数である。

重要度の高い単語を重要イベントと定義し、インデクスとして利用する。

2.3 イベント重要度

サッカーに関する単語とそれ以外の単語の頻度情報が同じ場合は、重要度に違いが出ないため単語の頻度情報のみでイベント重要度を定めることは難しい。そこで提案するのが、式(2)(3)のように、各要約時間に相当する要約テキストにおける重みと、90分フル映像の要約テキストにおける重みとの比をとり、この値が大きいものほど重要度が高いと考える方法である。これにより余分なイベントが削減され、必要なイベントの重要度が補正される。

$$\text{重要度 (20分)} = \frac{20 \text{分要約テキストにおける重み}}{90 \text{分要約テキストにおける重み}} \quad (2)$$

$$\text{重要度 (5分)} = \frac{5 \text{分要約テキストにおける重み}}{90 \text{分要約テキストにおける重み}} \quad (3)$$

具体的には、図1のように90分速報記事から20分要約映像、5分要約映像に対応するシーンテキストを抽出し、それぞれに含まれるイベント語の重みをTF・IDF法で計算する。さらに、各重み付けを90分速報記事全体を用いた場合のTF・IDF値で割ることによって、対象要約時間における最終的なイベント重要度とする。

3. 抽出結果

3.1 重要イベント

図2は、各試合ごとに算出したTF・IDF値の平均をカテゴリごとにソートしたものである。国際試合13試合より抽出される全784単語のうち、サッカーに関するイベントとして171単語を、プレイ・ポジション・身体・状況・時間情報の5カテゴリに分類して分析した。

これより、試合のプレイ内容の記述に直接関係のない時間情報の重みが大きくなっていることが分かる。プレイ主体の要約を作成する場合、このままでは重要度として利用できないため、前述の補正処理を加える。

連絡先: 平林 衛, 早稲田大学, hirabayashi@ruri.waseda.jp

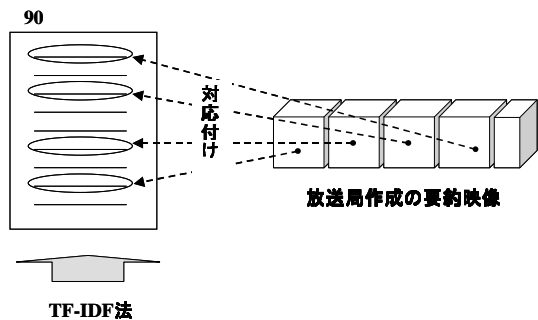


図 1: 90 分速報記事と放送局作成要約映像との対応付け

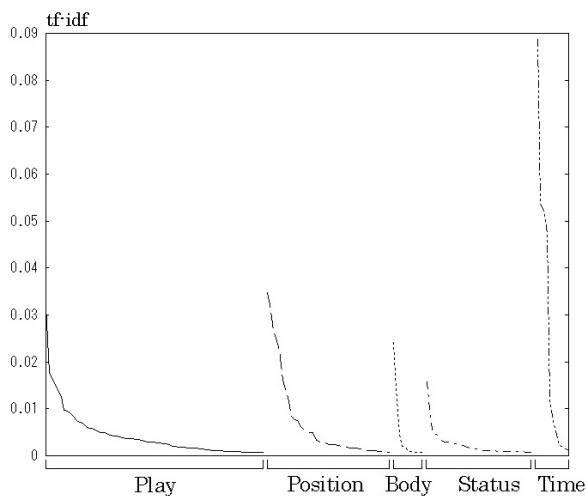


図 2: 各イベントの $TF \cdot IDF$ 値

3.2 イベント重要度

図 3 は、要約映像に相当する要約テキストとの比較によりイベント重要度を補正したものである。

これより、プレイ内容に直接関係のない時間情報等のイベントの重みが相対的に下げられていることが分かる。また、要約時間が 20 分から 5 分に短縮されると、各イベント重みの差がより大きくなっている。

5 分要約テキストを用いて補正した場合、プレイに属するイベントの上位には、ゴール、レッドカード、クロス、シュート、パス、FK といった、ニュース等の要約映像で中心となるイベントが集まっており、本手法の妥当性が確認された。

4. 考察

図 3 で示されるイベント重要度を利用して 2 試合のサッカー映像について要約映像を作成し、実際に TV で放送された要約映像との比較評価を行った。要約映像に含まれる同一シーンの割合と、不一致シーンにおいて本研究で対象外とした選手/観客アップ映像の占める割合についてまとめた結果を表 1 に示す。

アップ映像については、重要シーンの近傍に出現する確率が高いことからルールベースでの要約処理が可能と考えられるため、残り 2 割の不一致シーンに対する要約精度向上が今後の課題である。

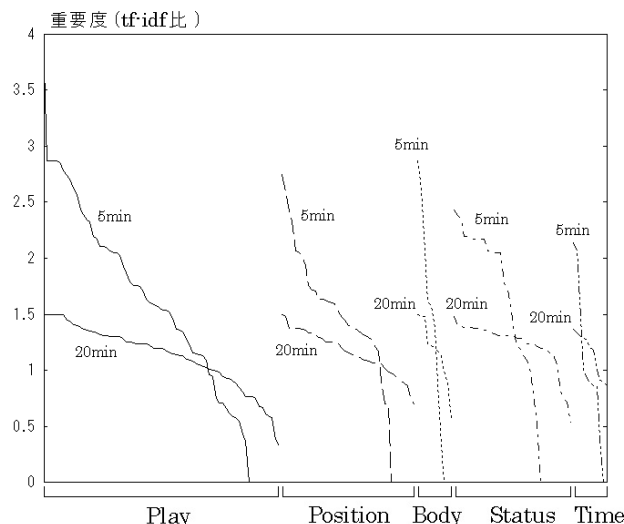


図 3: 補正後のイベント重要度

表 1: 5 分映像要約結果

試合 ID	同一シーン率 (%)	選手/観客アップ映像率 (%)
1	66.4	18.0
2	61.2	10.8
平均	63.8	14.4

本研究では、20 分要約映像、5 分要約映像を利用して要約規則の抽出を行った。得られた規則は、任意のサッカー映像に適用する場合、20 分、5 分要約映像の作成にのみ有効である。20 分、5 分要約映像の長さは一般的な要約映像の長さであり、一般向けの要約映像を作成する際には、この 2 種類の媒体があれば作成可能であると言える。また別時間の媒体があれば、同手法において重要度を算出することも可能である。

本手法は、テキストから要約映像作成規則を抽出しているところが特徴である。したがって、同じような媒体が存在すれば、サッカー映像以外の要約にも適用させることが可能である。他スポーツ映像はもちろんのこと、事件速報等のニュース映像もその対象と考えられる。

5. 今後の展開

今後は、任意の要約時間への対応、すなわち個人性を考慮した要約映像作成のための重要度に関する検討が必要である。また、大量のデータにおける精度向上を目指し、動画像処理・音声処理によるインデックス情報を組み合わせた要約手法の検討を行う予定である。

参考文献

- [1] 徳永健伸, "言語と計算 5 情報検索と言語処理", 東京大学出版会, 1999-11.
- [2] 日本語形態素解析システム「茶釜」,
<http://chasen.aist-nara.ac.jp/index.html>