

映像における話題の多様性抽出に基づく受動的映像視聴システムの提案

渡部 雅俊^{†1}王 元元^{†1}河合 由起子^{†1}角谷 和俊^{†2}^{†1} 京都産業大学^{†2} 兵庫県立大学

1 はじめに

近年、ビデオ・オンデマンドサービスの普及により、ユーザは観たい時にいつでも映像コンテンツを視聴できるようになった。映像コンテンツを活用する研究は活発に行われており、重要なシーンの抽出による映像の要約手法 [1] や SVM による音声の学習を用いた映像要約の研究 [2]、また、SNS を用いた映像要約の研究 [3] などがあげられる。しかしながら、映像視聴時に任意の時点で興味や疑問が生じた場合、それらに関連する話題のコンテンツ (Web ページや画像、Youtube 映像など) を検索しなければならない。そこで、本研究では、視聴映像の話題語抽出よりシーンを検出し、話題性の高低度を判定することで、詳細度に基づいて話題性の高い (あるいは低い) コンテンツを追加し、話題性の高い (あるいは低い) シーンを削除し、多様性のある映像を自動生成する手法を提案する。

具体的には、まず映像全体の話題語を抽出する。この話題語の出現頻度を用いてシーンを検出し、検出されたシーンに対する話題性を判定する。次に、各シーンに出現する話題語の検索ヒット数に基づき、話題性の高低度を算出し、生成する映像の種類に応じてランキングする。最後に、映像全体の再生時間の長さに基づき、シーンの削除と新たなシーンの追加を決定する。なお、追加シーンは、今回は Youtube 映像を用いて、関連性の高い順と映像の再生時間に基づき選択する。シーンの削除と取得した関連するシーンの挿入により 4 種類の映像コンテンツが生成される。

本論文では、話題発見に対するシーン検出および話題性に基づく受動的映像視聴システムについて提案し、被験者による実験検証を行う。

2 受動的映像視聴システム

2.1 システム概要

本研究では、字幕データを用いることで映像の話題語の抽出からシーンの検出を行い、また、検出したシーンの話題性の高さ・低さと時間制約に基づき、シーンの話題性に関連する映像の追加、話題性の高い・低いオリジナルシーンを削除することで、話題性と詳細度に合

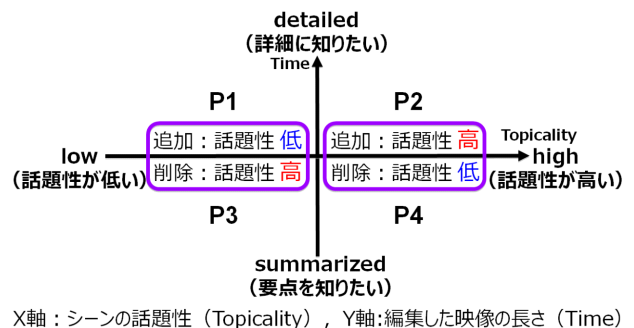


図 1: 話題性と詳細度に基づく映像コンテンツ生成

わせた 4 種類の映像コンテンツを生成する。本システムで生成される 4 種類の映像は図 1 のように分類される。P1, P3 は、シーンの話題性が低い場合に新たな映像を追加し、話題性の高いオリジナルシーンの削除を行う。逆に、P2, P4 は、シーンの話題性が高い場合に新たな映像を追加し、話題性の低いオリジナルシーンの削除を行う。また、P1, P2 は、映像の話題についてより詳細に知りたい場合とし、オリジナル映像より長い映像が生成され、逆に、P3, P4 は、映像の話題の要点だけを知りたい場合とし、オリジナル映像より短い映像が生成される。追加する新たな映像に関しては、投稿映像サイトより話題性に合わせた映像を検索し、映像の再生時間に基づき関連性の高いものを挿入する。

図 2 に本システムで生成される P2 と P4 の場合の概要を示す。映像全体の字幕データより、話題性が高い、低いシーンをそれぞれ検出する。赤い枠は話題性が高いと判定されたシーンであり、破線枠はそのシーンに含まれる話題語に関連する別の映像コンテンツを投稿映像サイトより取得し、追加する。逆に、青い枠は話題性が低いと判定されたシーンであり、該当シーンをオリジナル映像より削除する。黄色の枠は映像の編集を行わないシーンとなる。

以上より、例えば、話題性が低く、詳細を知りたい場合 (P1) は、マニア向けの長編の詳細映像が視聴でき、一方で、話題性が高く、要点を知りたい場合 (P4) は、一般受けのダイジェスト映像が視聴できる。

2.2 話題語抽出およびシーン検出手法

まず、映像コンテンツ全体の字幕データより、 tf 値が閾値以上の単語 w を抽出を行う。 tf 値は字幕に出現

A Passive Video Viewing System based on Diversity of Topics

^{†1} Masatoshi WATANABE ^{†1} Yuanyuan WANG ^{†1} Yukiko KAWAI ^{†2} Kazutoshi SUMIYA^{†1} Kyoto Sangyo University^{†2} University of Hyogo

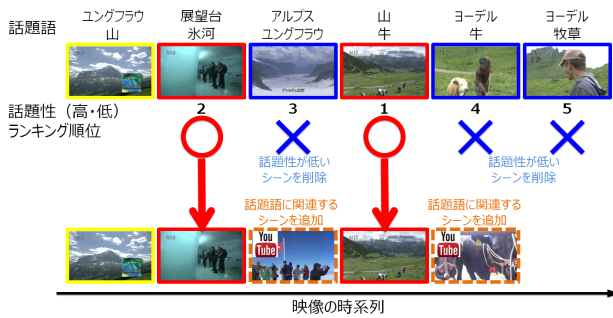


図 2: システム概要図

するそれぞれの単語の出現頻度である。抽出した単語のうち、字幕に出現する m 回以上の単語 w を話題語 k として抽出する。映像の時系列に沿って、抽出された話題語 k が n 回出現した場合、シーンとして検出を行う。

2.3 シーンの追加・削除判定

2.2 節で検出したシーンに対して、オリジナル映像に別映像からシーンの追加、オリジナル映像からシーンの削除を判定する。まず、検出したシーン の話題語 k をクエリとして、投稿映像サイトより各シーンごとに検索を行い、検索ヒット数を取得する。次に、検索ヒット数から追加、削除の判定を行うための閾値を算出する。閾値 T は以下の式を用いて算出する。

$$T = \frac{|S(k)|}{N}$$

$|S(k)|$: 映像全体の話題語により検索された結果数
 N : 検出したシーンの総数

この閾値 T と各シーンごとの検索ヒット数 $|S(k_i)|$ の比較を以下の判定式で行う。

- P1, P3 の場合: if $|S(k_i)| \leq T$ then 追加 else 削除
- P2, P4 の場合: if $|S(k_i)| \geq T$ then 追加 else 削除

3 評価実験

構築した受動的映像視聴システムの実験を行った。被験者は大学生 6 名、「シリーズ世界遺産 100」の以下の映像から、P1~P4 の 4 種類の映像を生成した。

- オリジナル映像 O1: スイスアルプス ユングフラウとアレッチェ〜スイス〜 (再生時間: 5 分 31 秒)
- オリジナル映像 O2: ヴァッハウ渓谷の文化的景観〜オーストリア〜 (再生時間: 5 分 30 秒)

提案手法により $n=3$ で検出されたシーン数は O1 が 9, O2 が 11 で、シーンの平均再生時間は約 33 秒であった。また、詳細映像 (P1, P2) の平均再生時間は 7 分 22 秒で、ダイジェスト映像 (P3, P4) の平均再生時間は 2 分 41 秒であった。

P1~P4 の 4 種類の映像に対して下記の設問項目を 5 段階評価によるアンケートを実施した。

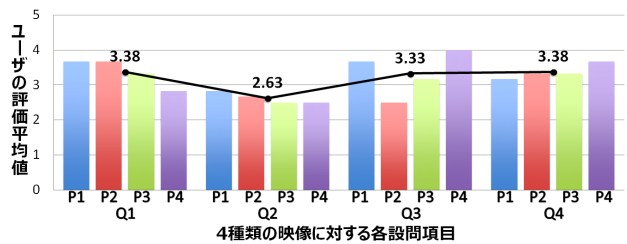


図 3: 各設問項目の評価結果

- Q1: 映像の内容が理解できた
- Q2: シーンの切り替わりの違和感がなかった
- Q3: 映像の内容に関係ないシーンがなかった
- Q4: P1, P2 の場合、映像は長いと感じた
P3, P4 の場合、映像は短いと感じた

各設問項目の評価平均値を図 3 に示す。Q1 と Q4 は全体的に高い評価を得られた。しかしながら、Q2 は全体的に低い評価となった。また、Q3 は特に P2 の評価が低くなった。以上より、短い映像を生成するため、シーンの削除の妥当性と切り替わりの違和感についての検討が必要である。なお、Q1~Q4 の全評価平均値は 3.12 となり、本手法により生成した 4 種類の映像の有用性が確認できた。

4 おわりに

本研究では、映像の字幕データを用いることで、映像の話題発見に対するシーン検出および話題性に基づく受動的映像視聴システムの構築し、シーンの削除と追加により生成した 4 種類の映像コンテンツが有効な結果であることを確認した。今後、新たに挿入するコンテンツのタイプ (Web ページや画像、地図など) を検討する。また、音声認識による音声情報と字幕の両方から話題語の抽出を検討する。

謝辞

本研究の一部は、総務省戦略的情報通信研究開発推進事業 (SCOPE) および JSPS 科研費 26280042 の助成を受けたものである。ここに記して謝意を表す。

参考文献

- [1] S. Kawamura, T. Fukusato, T. Hirai, and S. Morishima. Efficient video viewing system for racquet sports with automatic summarization focusing on rally scenes. Proc. of ACM SIGGRAPH 2014, p. 62, 2014.
- [2] C. Liu, Q. Huang, S. Jiang, L. Xing, Q. Ye, and W. Gao. A framework for flexible summarization of racquet sports video using multiple modalities. Computer Vision and Image Understanding, 113(3): 415–424, 2009.
- [3] 小林尊志, 野田雅文, 出口大輔, 高橋友和, 井手一郎, 村瀬洋. Twitter の実況書き込みを利用したスポーツ映像の要約. 電子情報通信学会技術研究報告. MVE, マルチメディア・仮想環境基礎, 110(457): 165–169, 2011.