

Web ページ補完のための共有動画に付与されたユーザコメントを用いた シーン抽出手法

若宮 翔子† 北山 大輔‡ 角谷 和俊‡

† 兵庫県立大学大学院環境人間学研究所

‡ 兵庫県立大学環境人間学部

1 はじめに

Web ページを閲覧しているときに、Web ページのテキストが示す具体的な動作や状態を知りたい場合があるが、テキストによりこれらが具体的に表現されることは少ない。たとえば、フィギュアスケートの演技を解説している Web ページの「見事なトリプルアクセルを披露した」というテキストには、トリプルアクセルがどのような動作であるのか、見事と形容されるトリプルアクセルはどのような状態であるのかといった具体的な動作や状態に関する表現が含まれていない。このような場合、Web ページのテキストを動画シーンによって補完することは有効である [1] [2] [?].

そこで本研究では、Web ページのテキストのみでは表現することが困難な動作や状態を共有動画のシーンによって補完する手法を提案する。共有動画はニコニコ動画*のような動画共有サイトに投稿されている動画であり、ユーザは共有動画を視聴しながらコメントを付与することができる。すなわち、共有動画には映像に出現する対象やその動作あるいは状態を形容するキーワードを多く含むユーザコメントが付与される。そしてこのようなユーザコメントを利用することで、Web ページのテキストのみでは表すことが難しい動作や状態を補完するシーンを抽出することが可能となる。

2 本研究のアプローチ

本稿では、Web ページのテキストに対する補完シーンを共有動画から抽出する手法を提案する。しかし表 1 から分かるように、Web ページのテキストには対象や動作を説明するキーワードが、ユーザコメントには映像に出現する対象やその動作あるいは状態を形容するキーワード、ユーザの感情を表すキーワードやコミュニティに特化した語 [3] が用いられることが多い。たとえば、前述のフィギュアスケートに関する Web ページでは「トリプルアクセル」の前後に「跳ぶ」や「着氷

表 1: Web ページのテキストとユーザコメントの性質

| | Webページの テキスト | 共有動画の ユーザコメント |
|---------|-----------------|--|
| 文書・文法構造 | あり | なし |
| 文脈 | あり | 不明 |
| キーワード | 名詞、動詞が中心 | 名詞、動詞、形容詞 副詞、感動詞 オノマトペ コミュニティに特化した語 |

する」などトリプルアクセルを直接説明するキーワードが用いられるのに対し、共有動画のユーザコメントでは「高い」、「フワフワ」などトリプルアクセルを形容するキーワードが用いられる。

このような Web ページのテキストとユーザコメントにおけるキーワード間のギャップを解消するために、本手法では Web ページ側のキーワードをユーザコメント側のキーワードに変換する。このとき補完対象となる Web ページのテキスト領域に応じて、補完シーンを抽出するために必要なキーワードも変わるため、それを考慮したキーワード変換が必要である。たとえば、Web ページのテキスト領域から「跳ぶ」というキーワードを抽出した場合、「跳ぶ」という状態が分かるシーンを抽出するために必要なキーワード（「高い」や「フワフワ」など）へ変換する。一方「跳ぶ」、「着氷する」というキーワードを抽出した場合には「跳ぶ」、「着氷する」という一連の動作が分かるシーンを抽出するために必要なキーワード（「ジャンプ」、「決まる」など）へ変換する。そして変換したキーワードから生成したクエリを用いて動画シーンを検索することで、単に Web ページ側のキーワードやその共起語を用いて生成したクエリ（「トリプルアクセルへ跳ぶ」や「トリプルアクセルへ着氷する」）では抽出できない、ユーザコメント側のキーワードのみで形容されているシーンの抽出が可能となる。

3 補完シーン検索

本手法ではまず Web ページを閲覧中にユーザが選択した領域内のテキストから、句読点の前の動詞と、そ

Scene Extraction Method to Complement Web Page using User Comments Attached to Shared Video

†Shoko WAKAMIYA ‡Daisuke KITAYAMA ‡Kazutoshi SUMIYA

†Graduate School of Human Science and Environment, University of Hyogo

‡School of Human Science and Environment, University of Hyogo

*ニコニコ動画 : <http://www.nicovideo.jp/>

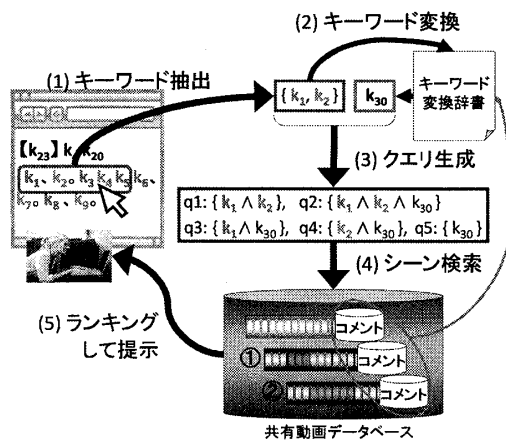


図 1: 補完シーン検索の概念図

表 2: 抽出した動詞と名詞に基づくキーワード変換

| | 動詞の数が多 | 動詞の数が少ない |
|----------|--------------------|-------------------|
| 名詞の種類が多様 | 動詞もしくは副詞のキーワードに変換 | 動詞もしくは副詞のキーワードに変換 |
| 名詞の種類が一様 | 名詞もしくは形容詞のキーワードに変換 | 動詞もしくは副詞のキーワードに変換 |

の動詞と係り受け*の関係がある名詞をキーワードとして抽出する (図 1 の (1))。そして抽出したキーワードを分析することで、どのようなキーワードに変換するかを決定し、ユーザコメントのキーワードの共起関係に基づき作成した辞書を用いて変換する (図 1 の (2))。具体的には、図 1 の (1) で抽出する動詞の数が少ないとき、選択領域の状態が分かるシーンが補完シーンとして適切であると推定し、動詞のキーワードもしくは動詞を修飾する働きがある副詞のキーワードへ変換する。一方、動詞の数が多く、名詞の種類が多様であれば、それぞれの名詞によって表される対象よりも一連の動作を見ることが出来るシーンが補完シーンとして適切であると推定し、動詞のキーワードもしくは動詞を修飾する副詞のキーワードへ変換する。動詞の数が多く、名詞の種類が一様である場合は、動詞によって表される個々の動作よりも名詞によって表される対象に関するシーンが補完シーンとして適切であると推定し、名詞のキーワードもしくは名詞を修飾する形容詞のキーワードへ変換する (表 2)。

次に Web ページから抽出したキーワードと変換したキーワードを用いて、動画シーン検索のためのクエリを生成する (図 1 の (3))。Web ページのテキストのキーワードと共有動画のユーザコメントのキーワードは、同じキーワードが用いられていない、あるいは同じキーワードが用いられていても共起語が異なる場合

があるという仮説に基づき、Web ページから抽出した名詞 k_n ・動詞 k_v と、変換されたキーワード k_t を組み合わせてクエリを生成する。 $k_n \wedge k_t$ や $k_v \wedge k_t$ というクエリを生成することで、Web ページのテキストとユーザコメントに共通するキーワードが一部欠けている場合にもシーンを検索できる。また、 k_t というクエリを生成すると、共通のキーワードが用いられていない場合でも、Web ページのテキストと共有動画シーンを対応付けて検索することができる (図 1 の (4))。検索結果は、選択領域内のテキストから抽出したキーワードとユーザコメントのキーワードの順序を比較し、一致するシーンを Web ページ文脈と近い順序を持つ動画シーンとして上位にランキングして提示する (図 1 の (5))。

4 おわりに

本稿では、Web ページのテキストにおいて具体的に表現されない動作や状態を補完するために、共有動画から補完シーンを抽出する手法を提案した。今後は提案したアルゴリズムに関する予備実験を行い、プロトタイプシステムを開発する予定である。

謝辞

本研究の一部は、平成 21 年度科研費基盤研究 (B)(2) 「ユーザの潜在的意図を用いたレス・コンシャス情報検索基盤の構築」(課題番号: 20300039) および平成 21 年度特別研究員奨励費 (21.197) によるものです。ここに記して謝意を表すものとします。

参考文献

- [1] 吹野 直紀, 馬 強, 角谷 和俊, 田中 克己. ニュース記事を利用したサッカー要約映像の生成. 第 14 回データ工学ワークショップ (DEWS2003), 2003 年
- [2] 馬強, 田中克己. 話題構造に基づく放送と Web コンテンツの統合のための検索機構. 情報処理学会論文誌 データベース (TOD), vol.45 No.SIG10, 2004 年
- [3] H. Miyamori, S. Nakamura, and K. Tanaka. Generation of views of tv content using tv viewers' perspectives expressed in live chats on the web. Proc. of the 13th annual ACM international conference on Multimedia (ACM Multimedia2005), Nov. 2005, pp. 853-861.

*CaboCha : <http://chasen.org/taku/software/cabocha/>