

公開動画リストを利用した動画検索手法における 公開動画リスト選定の改善に関する一考察

近藤 悠大[†] 西 友規[‡] 山口 実靖^{†‡}

[†]工学院大学 工学部 情報通信工学科

[‡]工学院大学大学院工学研究科電気・電子工学専攻

1. はじめに

インターネットを活用した動画共有サービスが普及し、動画共有サイトにおいて多くの動画が共有されている。しかし、動画共有サイトで提供されている動画検索機能では、検索語と関連した動画を得られないことがある。よって、動画共有サイトにおける動画検索機能の精度の向上は重要な課題の一つであると考えられる。

公開動画リストを用いる既存の検索手法は、多くの場合において高い精度で検索を行うことが可能である。しかし、検索語「原爆」など一部の検索語において精度が低くなる問題がある[1]。

そこで、本研究では、動画共有サイトで提供されている「タグ」と「公開動画リスト」と「Web 百科事典 (Wikipedia)」に着目し、検索語の表記揺れを修正することで、既存手法における公開動画リスト選定を改善し動画検索精度を向上させる手法を提案する。

2. Web コミュニティ抽出手法

Web 空間の中で共通の話題を有する Web ページ群を“Web コミュニティ”と呼ぶことができる。Web 空間から Web コミュニティを抽出する手法に関しては多くの研究成果が得られている[2,3,4]。

これらの手法では、特定の話題を持つ Web ページを Center とし、Center の集合を共通の話題を有する Web コミュニティと考える。また、Center の Web ページ群に対し多数のリンクを出している Web ページを Fan としている。この Center の集合と Fan の集合を用いて二部グラフを作成し、Center 集合からの Fan 集合の作成および Fan 集合からの Center 集合の作成を繰り返すことで、Center 集合の改善を行っていく。

3. 動画コミュニティ抽出手法

福井らは、動画共有サイト内の「共通の話題

を持つ動画の集合」を動画コミュニティとし、2章の Web コミュニティ抽出手法を用いて動画コミュニティを抽出する手法を提案している[1]。本手法では、Web コミュニティ抽出手法における Center, Fan, Fan から Center へのリンクを、動画共有サイトにおける動画、動画リスト、動画リストによる動画の登録に対応させることにより Web コミュニティ抽出手法を動画共有サイトに適用している。

公開動画リストを Fan, 公開動画リストに登録されている動画を Center とし、図1に示す手順で Center 動画集合から Fan 動画リストの抽出、Fan 動画リスト集合から Center 動画の抽出を繰り返し、動画コミュニティを抽出する。

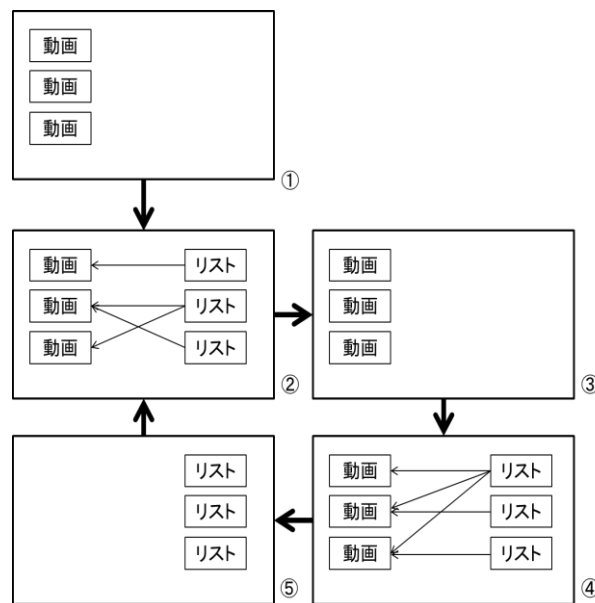


図1 動画コミュニティ抽出手法

4. WCTI 手法, WCTIZ 手法

動画コミュニティと TF-IDF を用いた WCTI 手法[1]は、2章の Web コミュニティ抽出手法におけるリンク先ページ、リンク元ページ、リンクを、動画共有サイトにおける動画、動画リスト、動画リストへの動画の登録で置き換え、Web コミュニティ抽出手法を適用する。また、TF-IDF における文書を動画リストに、単語を動画のタグに、文書内の全単語を動画リスト内の全動画

A Study on improvement of Selection of Public Video Lists in Video search using Public Video Lists

Kondo Yudai[†], Yuki Nishi[‡], Saneyasu Yamaguchi^{†‡}

[†]Department of Information and Communications Engineering, Kogakuin University

[‡]Electrical Engineering and Electronics, Kogakuin University Graduate School

の全タグに置き換え、TF-IDF を動画共有サイトに適用し、動画コミュニティ抽出を行う。そして、TF-IDF を用いて検索語と動画リストの関連度を評価している。

WCTI 手法と Zipf の法則を用いる WCTIZ 手法 [1] は、動画リストにおけるタグの出現頻度と、出現頻度の順位の関係を調査し、その動画リストの話題一貫性の強弱を推定し、一貫性が強い動画リストのみを用いる。WCTI 手法における Center 集合からの Fan の抽出の際に、タグの出現頻度順位と順位×出現頻度の関係のグラフが右上がりになる動画リストを除外し、抽出を行う。それ以外は、WCTI 手法と同一の手順を用いて動画コミュニティ抽出を行う。

5. 提案手法

本研究では、検索語の表記揺れを統一することで、既存手法において抽出される Fan 選定を改善する方法を提案する。

提案手法では、Web 百科事典 (Wikipedia) において同一の言葉として扱われる際に転送されるリダイレクトを基準に、リダイレクトされる前の言葉とされた後の言葉を同一の単語として扱うことにより、検索語の表記揺れを統一し、Fan 選定の改善を図る。

6. 性能評価

提案手法が有効かどうか確認するために、評価実験を行った。動画共有サイトで提供されている検索機能、Web 検索エンジン、既存手法、提案手法のそれぞれによる検索結果の比較を行った。動画共有サイトの検索結果は、キーワード検索およびタグ検索の結果を再生回数順あるいは動画リスト登録数順に並び替えて上位 50 件を結果としたもの、の 4 通りを用いた。Web 検索エンジンは検索範囲に当該動画共有サイトのみを指定し単語検索を行った上位 50 件を検索結果とした。また、既存手法および提案手法の初期 Center 動画の集合としては、動画共有サイトにより提供されているタグ検索機能の結果を動画リスト登録数順に並び替えた上位 10 件を用いた。抽出には、2013 年 4 月 1 日から 2014 年 1 月 2 日にニコニコ動画より収集した 1,756,422 件の動画と、181,335 件の動画リストを用いて行った。

検索結果の評価は各動画を閲覧し主観により行い、次の 3 段階 (A,B,C) の評価に分類した。

A 評価[+2 点]: 検索語と関係の深い動画

B 評価[+1 点]: 検索語と関係があるが関係が深くない動画

C 評価[±0 点]: 検索語と関係の無い動画

既存手法では初期 Center を「原爆」とし、提案手法では初期 Center を「原爆」、「原子爆弾」、「原爆と原子爆弾」とした。ただし、「原爆と原子爆弾」の初期 Center は、両方のタグ検索の結果を一つにした上でリスト登録回数順に並び替えた上位 10 件を用いた。検索語を「原爆」としたときの各検索手法の評価結果を表 1 に示す。提案手法では Web 百科事典 (Wikipedia) においてリダイレクトされている「原爆」と「原子爆弾」を統一し一つの単語として抽出を行った。

表 1 より、提案手法は最も A 評価が多く、他の検索手法より検索精度が高いことが確認された。

表 1 評価結果

	A[+2点]	B[+1点]	C[±0点]	合計点
キーワード+再生数	20	11	19	51
キーワード+動画リスト登録数	19	12	19	50
タグ+再生数	20	15	15	55
タグ+動画リスト登録数	20	13	17	53
Web検索エンジン	14	23	12	51
動画コミュニティ抽出手法	0	0	50	0
WCTI手法	20	6	24	46
WCTIZ手法	20	8	22	48
WCTI手法を基にした提案手法(初期Centerを原爆のみとした)	30	5	15	65
WCTIZ手法を基にした提案手法(初期Centerを原爆のみとした)	27	7	16	61
WCTI手法を基にした提案手法(初期Centerを原子爆弾のみとした)	30	6	14	66
WCTIZ手法を基にした提案手法(初期Centerを原子爆弾のみとした)	28	8	14	64
WCTI手法を基にした提案手法(初期Centerを原爆と原子爆弾両方とした)	27	1	22	55
WCTIZ手法を基にした提案手法(初期Centerを原爆と原子爆弾両方とした)	24	9	17	57

7. おわりに

本研究では、検索語の表記揺れを修正し、検索語と同一の単語として扱われている言葉を統一することにより、動画リストの選定を改善する手法を提案した。

また評価実験により、提案手法を用いることで動画コミュニティ抽出の精度が向上することがわかり、提案手法の有効性が確認された。

今後は検索語以外も含むすべてのタグの表現揺れの修正について考察していく予定である。

謝辞

本研究は JSPS 科研費 24300034, 25280022 の助成を受けたものである

参考文献

- [1] 西友規, 山口実靖, 小林亜樹, “動画リストの主題の偏りを利用した動画コミュニティ抽出”, 信学技報, vol. 113, no. 150, DE2013-28, pp. 163-168, 2013
- [2] Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, Andrew Tomkins, “Trawling the Web for emerging cyber communities,” In Proc. of the 8th international conference on World Wide Web, pp. 1481 - 1493, 1999.
- [3] P. Krishna Reddy, Masaru Kitsuregawa, “An approach to relate the web communities through bipartitegraphs,” Proc. of the 2nd International Conference on Web Information Systems Engineering, 2001.
- [4] Jon M. Kleinberg, “Authoritative sources in a hyperlinked environment,” Journal of the ACM (JACM), Volume 46 Issue 5, pp. 604 - 632, 1999.
- [5] 福井紀彦, 山口実靖, “動画共有サイトにおけるコミュニティ抽出”情報処理学会 2012 年全国大会 6N-7