

Web コンテンツを用いた人物紹介映像の自動編集に向けて

國代 京花[†] ナックフランク^{††} 井手 一郎[†] 川西 康友[†]
出口 大輔^{†††,†} 村瀬 洋[†]

[†] 名古屋大学大学院情報科学研究科 〒464-8601 愛知県名古屋市千種区不老町
^{††} アムステルダム大学 情報学研究所 〒1098 XH オランダ アムステルダム市 サイエンスパーク 904
^{†††} 名古屋大学情報戦略室 〒464-8601 愛知県名古屋市千種区不老町
E-mail: [†]kunishirok@murase.m.is.nagoya-u.ac.jp, ^{††}nack@uva.nl,
^{†††}{ide,kawanishi,murase}@is.nagoya-u.ac.jp, ^{††††}ddeguchi@nagoya-u.jp

あらまし 本報告では Web コンテンツを用いて、著名な人物の紹介映像を自動で生成する手法を提案する。紹介映像は、幼少期、功績、私生活の3部で編成する。各部に関する重要な事象を表すキーフレーズをオンライン百科事典 Wikipedia から抽出し、人物に関連する重要な事象に対応する画像を画像検索エンジンを用いて取得する。取得した画像について、その内容を表す Visual Concept に着目し、キーフレーズに関する一般的な Visual Concept と一致度が高いものを選ぶことで、人物に関する重要な事象を視覚的によく表現した画像を選択する。選択された画像にキーフレーズをテロップとして付与することで、映像に用いる最終的な画像とする。最後に、テロップ付きの画像をスライドショーの形式で映像化する。

キーワード 映像編集, 人物情報, 画像検索

Towards Automatic Authoring of a Video Biography using Web Contents

Kyoka KUNISHIRO[†], Frank NACK^{††}, Ichiro IDE[†], Yasutomo KAWANISHI[†],
Daisuke DEGUCHI^{†††,†}, and Hiroshi MURASE[†]

[†] Graduate School of Information Science, Nagoya University
Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan
^{††} Informatics Institute, Univ. of Amsterdam
Science Park 904, 1098 XH Amsterdam, The Netherlands
^{†††} Information Strategy Office, Nagoya University
Furo-cho, Chikusa-ku, Nagoya-shi, Aichi, 464-8601 Japan
E-mail: [†]kunishirok@murase.m.is.nagoya-u.ac.jp, ^{††}nack@uva.nl,
^{†††}{ide,kawanishi,murase}@is.nagoya-u.ac.jp, ^{††††}ddeguchi@nagoya-u.jp

Abstract In this report, we propose an image selection method for automatic authoring of a video biography using Web contents. The video biography is composed of three parts; childhood, profession, and personal life. We extract keyphrases on important events for each of these parts from the online encyclopedia Wikipedia, and then retrieve images from an image search engine that correspond with the important events of the person in focus. Focusing on the Visual Concepts of the obtained images, by selecting images that share the same Visual Concepts that represent the keyphrases in general, we select images that visually represent the important events of the person in focus, well. We super-impose the extracted keyphrases as captions to the selected images. Finally, we make a slideshow movie showing the selected images one after another.

Key words Video authoring, Personal information, Image retrieval

1. はじめに

訃報や事件、著名人の来日などの日々のニュースにおいて、著名な人物はそのニュースの中で生い立ちや経歴を紹介されることが多い。著名人について紹介する場合、その多くは人物のこれまでの功績や様子をまとめた映像を伴って伝えられる。本研究ではそのような映像を「紹介映像」と呼ぶ。紹介映像は視聴者に対し、著名人に関する情報を視聴覚的に伝えられるため、著名人について知らない人でもどのような人物であるか知ることができる。このような紹介をすることで、視聴者にその著名人に関心を持たせ、ニュース内容への深い理解を助けられる。しかし、予め放映日時が分かっている場合を除き、事件や訃報などの突発的事象に関するニュースのために製作する紹介映像は速報性が求められることが多い。人手での素材収集や編集作業に時間を要するため、生成される紹介映像の内容は近年話題になったような視聴者の記憶に新しい話題で構成されることが多い。

そこで我々は、充実した内容の紹介映像を自動で生成することを目的とし、映像に用いるための素材を Web 上のコンテンツから自動的に収集する方法を提案した [1]。現在、多くの人が Web 上の様々な媒体から情報を得ることができる。文字媒体ならば Wikipedia^(注1) のようなオンライン百科事典から、映像媒体や音声媒体ならば Youtube^(注2) のような映像共有サービスや画像検索エンジンから、必要な情報を容易に得ることができる。これらの Web コンテンツは、一般ユーザが盛んに投稿することで莫大な量の情報が日々追加、更新されているため、著名な人物が関わった出来事について、現在から過去に遡って広範な情報が得られる。しかし、大量の情報からある人物に関係する出来事を網羅的に抽出・選択して編集するのは大変な作業である。特に、編集者がその著名人について詳しく知らない場合は抽出すべき内容を調べるところから始めなければならず、負担が大きい。そのため、我々の提案する手法によって充実した内容の映像を自動で生成し、編集者の負担を軽減することを目指す。

2. 関連手法

オンライン百科事典 Wikipedia は、人物に限らず様々な事象に関する詳細な情報を抽出する情報源として注目されている。また、Zhang らはその内容を構造化したデータベース DBpedia [2]^(注3) を利用した属性抽出を行った [3]。DBpedia を用いれば、名前や出生地、職業などの人物に関する基本的な情報を属性として抽出することができる。Jiang らは属性とそれに関連の深いフレーズ群を、その人物に関する重要な属性情報の直前に現れる目印として利用し、各目印には抽出するルールを詳細に設定している [4]。

また、抽出したテキスト情報に関する Web 上の画像の取得

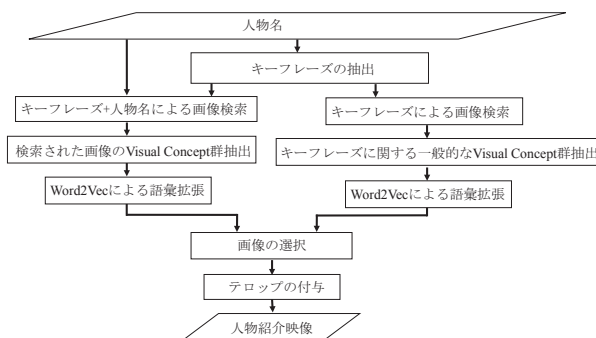


図1 人物紹介映像生成の処理手順

には、Google 画像検索^(注4) や Yahoo!画像検索^(注5)、Bing 画像検索^(注6) などの画像検索エンジンを用いた収集が行われる。一般に画像検索で画像を収集する場合、上位の画像ほどクエリと類似していると考えられる [5]。しかし、上位の画像でもクエリと画像の内容が一致しない画像が現れることもある。そのため、Zhang らは Visual Word [6] を用いた再ランキングを行っており、高い精度を示している [7]。ここで用いられているクエリは直接画像内の物体の名称になっているために、画像内の物体が直接クエリ名を表さない場合の結果は保証されていない。Jing らはクエリによって検索された画像に対して PageRank アルゴリズムを取り入れることで画像間の視覚的な類似度を算出し、絵画などのクエリに対しても、高い精度のランキング結果を得ている [8]。

3. 人物紹介映像の自動生成のための提案手法

2. で紹介した研究では、属性情報の抽出が主な目的だった。しかし、本研究では著名人の属性の他に、一生のうちに生じた重要なイベントを満遍なく抽出することが目的の1つである。単純な DBpedia の利用では著名人に関連のある特徴的な事象を満遍なく抽出できないため、DBpedia からの属性情報抽出の他にも、Wikipedia の見出しを利用した固有名詞の抽出を行う。また、抽出した人物情報に即した画像を得るため、本研究では、Visual Concept [9] を用いて Jing らの考えに類似した方法を採用する。

提案手法の処理手順を図1に示す。紹介映像は著名人の人生を要約したものであり、幼少期、功績、私生活の3部で構成する。始めに著名人に関する特徴的な事象を表現したキーワードを Wikipedia と DBpedia から抽出する。次に、このキーワードをもとに画像検索を行う。このとき、画像検索結果の画像とキーワードが表す対象の視覚的な内容が一致するように、Visual Concept を用いて画像の視覚的な内容を分析する。キーワードの内容と最も一致するものを、紹介映像に用いる画像として選択する。最後に、取得した画像に対応するキーワードをテロップとして付与し、スライドショーの形式で映像化する。

(注1) : <https://en.wikipedia.org/>

(注2) : <https://www.youtube.com/>

(注3) : <http://wiki.dbpedia.org/>

(注4) : <https://images.google.com/>

(注5) : <http://search.yahoo.co.jp/image>

(注6) : <http://www.bing.com/images>

3.1 キーフレーズの抽出

3.1.1 定型情報の抽出

Wikipedia ページには infobox という、そのページタイトルに関する基本的な定型情報がテンプレートを用いて掲載されている。DBpedia の情報は、これらの情報が元になっている。そのため、ノイズが少ない情報を抽出しやすく、これから生年月日、職業、出生地など定型的な情報を SPARQL^(注3) を用いて抽出する。しかし、元の情報となる infobox のテンプレートは功績や私生活に関する情報の分類や構造化は行っていない。そのため、それらについては Wikipedia から直接抽出する。

3.1.2 非定型情報の抽出

Wikipedia は閲覧する人物が自由に書き換え可能な情報媒体である。そのため、厳密に決まった文章の書き方は存在せず、各人物について記述される情報は編集者の見方や知識に依存する。このような自由文から必要な情報だけを抽出するため、本研究ではキーフレーズとして取得したい語句に現れる特徴に着目し、それらを目印とした文字テンプレートマッチングを行う。

まず、キーフレーズとなりそうな語句を、Wikipedia の文章全体からキーフレーズ抽出用のテンプレートを使って抽出する。Wikipedia では、執筆者が強調したい語句はアポストロフィなどを用いた Wiki 固有の記法により強調される。また文章中から張られるリンクのアンカテキストは固有名詞や重要な語句であることが多い。そこでこれらの表記があれば、それに対応する語句をキーフレーズとして抽出する。一般に Wikipedia 冒頭の概要では、その人物の特に重要な特徴が簡潔にまとめられているため、概要内から抽出できるキーフレーズを主に利用する。

次に、概要には功績と私生活の内容が混在して記述されているため、抽出したキーフレーズがどの内容を表しているのか判別する必要がある。Wikipedia はページの冒頭部分に見出し一覧が表示されており、閲覧者はページ内の大まかな内容を把握することができる。そのため、それらの見出しには功績や私生活をそのまま表すようなキーワードが含まれやすい。そこで、予め功績と私生活に関連ある語句を含むキーワードを文字列テンプレートとしてそれぞれ用意しておき、それらと一致した見出しが Wikipedia 記事中にあれば、その見出し以下の文章に含まれるキーフレーズは功績と私生活のいずれかの分野に関係するものとして判別する。

功績と私生活の2つの分野へ分類のために使用する文字列テンプレートの例の一部を表1に示す。功績として抽出するキーフレーズは、著名人の主な職業に関係するものである。私生活として抽出すべきキーフレーズは、結婚相手や嗜好品、副業、社会活動等に関係するものである。

Wikipedia の見出しのうち、細かい見出しはテンプレートと一致しないものもある。その場合は、直前の項目と関連があると仮定し、直前の見出しが分類された分野にキーフレーズを分類する。表1のテンプレートによって分類されたキーフレーズと、概要文章内のキーフレーズとを比較し、一致するものを取得することで概要文章内のキーフレーズを功績と私生活に分類する。

最後に、必要に応じて付加情報の付与を行う。上記のキーフ

表1 キーフレーズ分類のための文字列テンプレート例

分野	文字列テンプレート
功績	“career”, “award”, “honors”, “public”, “prize”, “work”, “credit”
私生活	“personal life”, “marriage”, “philanthropy”, “love”, “activism”

表2 提案手法によって取得したキーフレーズ例

(a) Elizabeth Taylor

分野	取得したキーフレーズ
出生地	“England”
功績	“Academy Award for Best Actress”, “American Film Institute”, “National Velvet (film)”, “Who’s Afraid of Virginia Woolf? (film)”
私生活	“Richard Burton”, “Conrad Hilton, Jr”, “Fragrance and jewelry brands”, “hiv/aids activism+American Foundation for AIDS Research”

(b) Albert Einstein

分野	取得したキーフレーズ
出生地	“Ulm”
功績	“Introduction to quantum mechanics”, “University of Bern”, “Nobel Prize in Physics”, “On the Quantum Mechanics of Radiation(paper title)”
私生活	“Mileva Marč”, “Elsa Löwentha”, “Love of music+violin+playing”, “assisting zionist causes + World Zionist Organization”

レーズの抽出法では、しばしば「carrier」や「activism」などの見出しがそのままキーフレーズとして取得され、不十分な場合がある。このようなキーフレーズを見出しキーフレーズと呼ぶ。見出しキーフレーズは、そのままでは適切な画像を収集するのに使えないため、付加情報となるキーフレーズを追加し、その特定性を高める。付加情報となるキーフレーズは、その見出しが示す文章内から抽出したキーフレーズ群から、新たに文字列テンプレートマッチングを行い選択する。

まず表1で紹介したような、各分野に関連の強いキーワードを用いて文字テンプレートマッチングを行う。ここでテンプレートと一致するキーフレーズが、見出しキーフレーズの示す文章内から抽出したキーフレーズ群になれば、さらに Word2Vec [10] を用いたテンプレートマッチングを行う。まず、見出しキーフレーズの取得の際に使用した表1の文字列テンプレートの共起語を、Word2Vec を用いて取得する。それらを拡張テンプレートとし、見出しキーフレーズの示す文章内から抽出したキーフレーズ群とマッチングを行う。もし一致するキーフレーズがあれば、それを付加情報として取得する。これにより、後の画像選択の際により精度の高い画像収集を可能にする。

表2に Elizabeth Taylor と Albert Einstein の2名に関して取得できたキーフレーズの一部を紹介する。私生活分野内のキーフレーズで“+”を含むものはより詳細な情報を含んだキーフレーズであり、“+”以下のキーフレーズがより詳細な情報を表している。

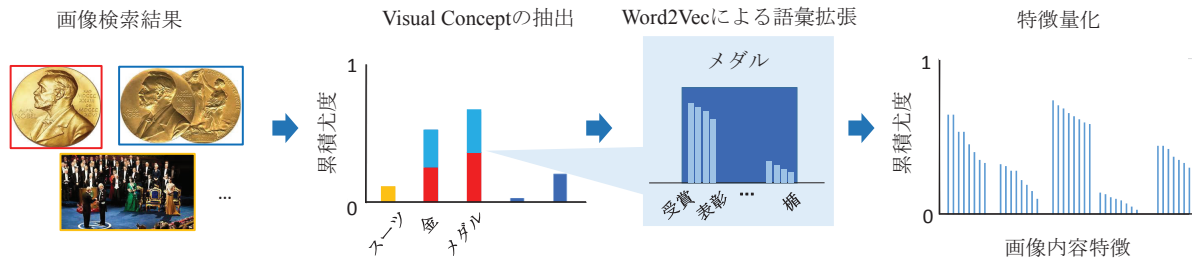


図 2 画像内容特徴の抽出

3.2 画像の選択

キーワードをクエリとした画像検索で得られる画像の中から、その視覚的な内容がキーワードの内容と一致したものを選択したい。そこで本研究では画像検索をキーワードのみをクエリとしたものと、人物名とキーワードをクエリとしたものの2回に分けて行い、それぞれ得られた画像から特徴を抽出することで一般的なキーワードとの視覚的内容の一致を図る。

3.2.1 画像内容特徴の抽出

まずキーワードのみをクエリとして画像検索を行い、得られた画像から Visual Concept 群を抽出する。得られた Visual Concept 群は、そのキーワードに関する一般的な Visual Concept 群とみなす。次に、著名人の名前とキーワードの両方をクエリとして画像検索を行い、検索された画像から Visual Concept 群を抽出する。これを先ほど得た一般的な Visual Concept 群と比較し、最も一致する画像を検索結果から抽出する。このようにして、著名人に関するそのキーワードが表す事象の典型的な画像を選択する。しかし、画像検索から得られた画像にはクエリに関係ない画像も含まれていることがある。そこでノイズとなる余分な特徴を排斥するために、特徴として利用する Visual Concept 群は閾値処理を行った後に、累積尤度が上位のものだけを利用する。

ここで Visual Concept 群の語彙のみでは画像内容を表す特徴として不十分であると考えたため、Word2Vec を用いて、各 Visual Concept の名称と共起確率が高い語句を用いて語彙を拡張することで特徴量を増やす。まず、取得した各 Visual Concept をクエリとして、Word2Vec を用いて関連する共起語彙のうち尤度が閾値以上のものを取得する。次に、クエリとなった Visual Concept の尤度とそれに対応する共起語彙の尤度を乗算してヒストグラムの形式で並べる。最終的にこのように拡張された特徴量を画像内容特徴とし、比較の際に用いる。画像内容特徴の抽出の様子を図 2 に載せる。

3.2.2 画像の比較

最後に一般的なキーワードを表す画像から生成した画像内容特徴と、それぞれの画像が持つ画像内容特徴を比較する。それぞれのヒストグラムの差分をとり、最も少ないものを一般的なキーワードを表す画像と視覚的内容が似ている画像として選択する。

表 3 実験結果

選択手法	比較手法	提案手法
平均選択率	0.243	0.476



(a) Elizabeth Taylor の “hiv/aids activism”



(b) Albert Einstein の “Love of music”

図 3 取得した画像の結果

(左) 詳細な情報の付与なし (右) 詳細な情報の付与あり

4. 実験

画像選択手法の有効性を検証するため、15 人の著名人に対し、その人に関連あるキーワードをよく表す画像を選択する実験を行った。20 代の男女含めた 9 名を被験者として選び、各被験者は人物名とキーワードで検索した結果の上位 4 枚から、最もキーワードを表していると考えられる画像を 1 枚選択した。比較手法、提案手法ともに、被験者に与えられたものと同じ 4 枚の画像から 1 枚だけ選び、それが被験者が選んだ画像と一致する割合を調べた。評価指標となる手法の選択率は、各手法が選択した画像と被験者によって選ばれたものが一致した数を、総設問数で除したものとす。

本実験の比較手法は、著名人の人物名とそれぞれに関連するキーワードの両方をクエリとして画像検索を行った結果から第 1 位の画像を利用したものである。また、提案手法は画像特徴の抽出において上位 10 位までの Visual Concept を用い、Word2Vec は尤度が 0.5 以上の単語のみ利用し、類似度が最も高い 1 枚を選出する。表 1 に実験結果として各手法の選択率を載せる。提案手法は比較手法より 23.3% の精度向上が見られた。これにより、提案手法はより人間の判断に近い画像の選択を行えることが示唆された。

提案手法が有効だったキーワードは、「アカデミー賞」や「ユニセフ親善大使」など、よく知られているものであった。こ

これらのキーワードは、「トロフィー」や「ドレス」などといった一般的な特徴が抽出しやすいものであったために精度が高かったと考えられる。

一方、提案手法が有効でなかったキーワードは、海外の著名人に関するキーワードであまり認知されていないキーワードであった。この場合、被験者はキーワードから連想される一般的な視覚特徴よりも、画像内にあるテキスト情報からキーワードに関連があるものか判断する傾向にあった。これより、被験者はキーワードの内容がわからない場合は、視覚的な特徴だけでは正しく画像選択が行えなかったと考えられる。

次に、見出しキーワードへの付加情報に対する有効性を検証する。図3に表2で取得できたキーワードのうち、「+」以下のキーワードの有無で取得できる画像の変化を載せる。図3の上側はElizabeth Taylorの「hiv/aids activism」に関する結果であり、下側はAlbert Einsteinの「Love of music」に関する結果である。この結果は、詳細な情報の付加によって取得できる画像が、よりキーワードに一致したものになる可能性を示唆している。

5. むすび

本研究ではWeb上のコンテンツを用いた人物紹介映像の自動生成手法について述べた。提案手法では画像を選択する際にVisual Conceptを用いたが、語彙にないものは検出できない。今後は語彙にないVisual Conceptに対しても、より適した画像を選択する手法へと拡張したい。また、Visual Conceptのみでは表現しきれない画像構造を捉えることで、選択画像の更なる精度向上へと努めたい。

謝辞 実験に協力して下さった方々に感謝いたします。本研究の一部は、科学研究費補助金及び国立情報学研究所との共同研究による。

文 献

- [1] 國代 京花, F. Nack, 井手 一郎, 川西 康友, 出口 大輔, 村瀬 洋, “Web コンテンツを用いた訃報映像の自動編集へ向けた画像選択方法の検討,” 信学技報, MVE2015-63, Mar. 2016.
- [2] C. Bizer, “DBpedia —A crystallization point for the Web of data,” Web Semantics: Science, Services and Agents on the World Wide Web, vol. 7, no. 3, pp.154-165, Sept. 2009.
- [3] K. Zhang, M. Wang, X. Cong, F. Huang, H. Xue, L. Li, and Z. Gao, “Personal attributes extraction based on the combination of trigger words, dictionary and rules,” in Proc. 3rd CIPS-SIGHAN Joint Conf. on Chinese Language Processing, pp.114-119, Oct. 2014.
- [4] I. Russo, T. Caselli, and M. Monachini, “Extracting and visualising biographical events from Wikipedia,” in Proc. 1st Conf. on Biographical Data in a Digital World 2015 CEUR Workshop Procs., vol.1399, pp.111-115, April 2015.
- [5] Y.-G. Jiang, C.-W. Ngo, and J. Yang, “Towards optimal bag-of-features for object categorization and semantic video retrieval,” in Proc. 6th ACM Int. Conf. on Image and Video Retrieval, pp.494-501, July 2007.
- [6] Csurka, G., Bray, C., Dance, C. and Fan, L. “Visual categorization with bags of keypoints”, in Proc. of ECCV Workshop on Statistical Learning in Computer Vision pp. 59-74, May 2004.
- [7] S. Zhang, Q. Tian, G. Hua, Q. Huang, and S. Li, “Descriptive visual words and visual phrases for image applications,” in Proc. 17th ACM Int. Conf. on Multimedia, pp.75-84, Oct. 2009.
- [8] Y. Jing and S. Baluja, “Visualrank: Applying PageRank to large-scale image search,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.30, no.11, pp.1870-1890, Nov. 2008.
- [9] S.K. Divvala, A. Farhadi, and C. Guestrin, “Learning everything about anything: Webly-supervised visual concept learning,” in Proc. 2014 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, pp.3270-3277, June 2014.
- [10] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” arXiv Pre-print, arXiv:1301.3981, Sept. 2013.