

人物紹介映像の自動編集に向けた Web コンテンツの自動選択

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

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ワード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収集する。収集した画像について、その内容を表す物体やシーンに着目し、キーワードに関する一般的な画像内容を表す特徴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。画像内容の表現には、ORB 特徴量と Visual Concept を用い、画像検索のクエリに用いたキーワードの属性によってこれらを切り替えて照合する。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

Automatic Selection of Web Contents Towards Automatic Authoring of a Video Biography

Kyoka KUNISHIRO[†], Frank NACK^{††}, Ichiro IDE[†], Yasutomo KAWANISHI[†],

Daisuke DEGUCHI^{†††}, and Hiroshi MURASE[†]

[†] Grad. School of Information Science, Nagoya Univ., Furo-cho, Chikusa-ku, Nagoya, 464-8601, Japan

^{††} Informatics Institute, Univ. of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands

^{†††} Information Strategy Office, Nagoya Univ., Furo-cho, Chikusa-ku, Nagoya, 464-8601, Japan

Abstract In this report, we propose an automatic authoring method of video biography using Web contents. A video biography is composed of three parts: childhood, profession, and personal life. The proposed method extracts 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. Focusing on the objects and scenes of the obtained images, by selecting images that share the same features that represent the keyphrases in general, The proposed method selects images that visually represent the important events of the famous person, well. We used ORB and Visual Concepts for representing image contents, and they were switched by the types of the keyphrase and matched. Finally, we produce a video biography as a slide-show of the selected images with captions.

Key words Video authoring, Personal information, Image retrieval

1. はじめに

近年、スマートフォンやパーソナルコンピュータの普及が進んでいる。このようなモバイル端末の利用目的の 1 つとして、ニュースの閲覧がある。利用者が閲覧するニュース内容のうち、最も関心が高いものはスポーツ・芸能に関するものである。こ

のようなニュースでは、著名人に関する記事が多く、彼ら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、その多くはこれまでの功績や来歴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「人物紹介映像」と呼ぶ。人物紹介映像は視聴者に対し、著名人に関する情報を視覚的に伝えられるため、テキストの

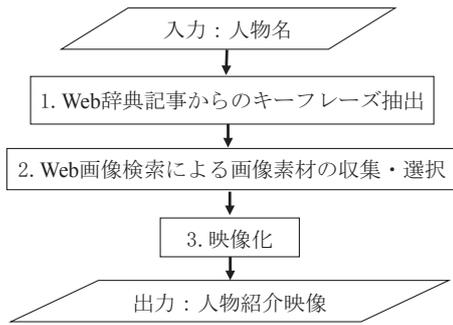


図1 提案手法による人物紹介映像の自動編集システム。

みの記事やアナウンサーによる単なる口頭説明よりも、その人物の人物像を視聴者に直感的に伝えられると考えられる。その人物について知らない視聴者でも、その人物像を知ることができ、また知っている場合でもその人物像を再認識できる。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、ニュース内容の深い理解を助けることができる。しかし、予め放映日時が分かっている場合を除き、事件や死去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。しかし、人手での素材収集や編集作業には時間を要するため、実際に放映される人物紹介の映像の内容は、近年話題になったような視聴者の記憶に新しい話題に限定されることが多い。

これに対して、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て、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手法を提案した [1]。この手法では、素材画像の選択において、Visual Concept [2] を用いてキーワードを表すモデルを生成することで、著名人とキーワードの関係を視覚的によく表すものを選択した。本報告では、この手法を更に拡張した自動編集手法を提案する。具体的には、画像選択の際に構築するモデルを、キーワードが属する分野によって変更する。

以降、まず 2. で、本報告で提案する Web コンテンツ選択手法の全体像について説明する。次に 3. で、著名人に関するキーワード抽出方法について説明する。また 4. で、著名人と抽出したキーワードの両方を視覚的によく表す画像の選択方法について説明する。さらに 5. で、Web コンテンツ選択手法の評価実験について述べる。最後に、6. で本報告をまとめる。

2. 提案手法

図 1 に、本報告で提案する人物紹介映像の自動編集システムの処理手順を示す。人物紹介映像は、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 部で編成する。始めに、Web 辞典であるオンライン百科事典 Wikipedia^(注1) とその内容を構造化したデータベース DBpedia [3]^(注2) から、著名人に関する特徴的な事象を表現したキーワードを抽出する。次に、このキーワードをもとに画像検索をする。一般に画像検索で

(注1) : <https://en.wikipedia.org/wiki>

(注2) : <http://dbpedia.org/sparql>

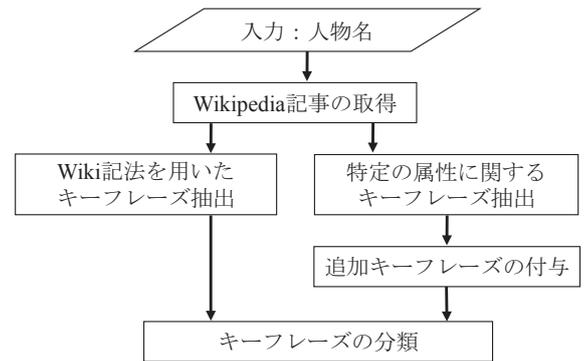


図2 Wikipedia 記事からのキーワード抽出の処理手順。

表1 キーワード抽出のための Wiki 記法

Wiki での表現	Wiki 記法
リンク	<code>\[\[(.+?)\]\]</code>
斜体	<code>''(.+?)''</code>
太字	<code>'''(.+?)'''</code>

画像を収集する場合、上位の画像ほどクエリと一致していると考えられる [4] が、クエリと視覚的な内容が一致しない画像が現れることもある。そこで、キーワードを表す一般的なモデルを構築することで画像の視覚的な内容を分析し、著名人に関係する画像のうちキーワードの内容と最も一致するものを選択する。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。本報告では、主に Web 辞典からのキーワード抽出方法と Web 画像検索による画像素材の収集・選択方法について紹介する。

3. Web 辞典からのキーワード抽出

3.1 DBpedia からのキーワード抽出

DBpedia は Wikipedia の情報を構造化したものであり、SPARQL と呼ばれる問い合わせ言語を用いることで構造化情報にアクセスできる。本報告では、これから生年月日、職業、出生地に関する定型的な基本情報のみを抽出する。しかし、功績や私生活に関する情報の分類や構造化は人物ごとに異なり、定型的でないため、Wikipedia から直接抽出する。

3.2 Wikipedia からのキーワード抽出

Wikipedia 記事から、著名人に関する特筆すべきキーワードを抽出したい。図 2 に、Wikipedia 記事からのキーワード抽出手順を載せる。各処理の詳細を以下に述べる。

3.2.1 Wiki 記法によるキーワード抽出

Wikipedia 記事の編集者は、Wiki 記法に従い、重要と考えられる語句を太字や斜体にして強調する。また、記事中の語句のうち、別の記事の見出しとなっているものについては、その語句にリンクを張ることができる。このような語句は、誰もが知っているような事象であることが多い。そこで、提案手法では著名人に関する Wikipedia 記事内にある強調された語句とリンクのアンカテキストを、キーワード候補として抽出する。

キーワードを抽出する範囲は記事冒頭の文章（最初の文章から次の見出しまで）とする。この文章は記事全体の概要であ

表 2 キーフレーズ抽出のための属性別テンプレート

属性	テンプレート
趣味	“* love *”, “* favorite *”, “* like *”, “* prefer *”
副業	“* business *”, “* legacy *”, “* other venture *”, “* launch *”
社会活動	“* activism *”, “* philanthropy *”, “* charity *”, “* causes *”, “* donate *”, “* supporter *”, “* right *”, “* foundation *”, “* goodwill *”
配偶者	“\ spouse = *”
論文	“\ trans-title = *”

ることが多く、冒頭文章中の強調された語句やリンクのアンカテキストは、著名人について特に重要な情報が厳選されていると考えられる。Wiki 記法では、記事を構成する各章の大きな見出し語はページのソースに “==” の記述で表されるため、これを目印にして冒頭文章の抽出範囲を決めることができる。

冒頭文章からキーフレーズを抽出するために、本研究では表 1 に示すような正規表現のテンプレートを用いる。“.” は任意の 1 文字，“+” は 1 回以上の繰り返し，“+?” は最短マッチを表している。ただし，“\” はエスケープ文字である。具体的なキーフレーズの抽出は、表 1 の各テンプレートと一致したそれぞれの語句を抽出することで行う。

3.2.2 特定の属性に関するキーフレーズ抽出

冒頭文章から得たキーフレーズは重要なものが厳選されているが、その内容は著名人の主な職業に関するものに偏る傾向がある [5]。冒頭文章中に私生活に関するキーフレーズが少ない場合も考えて、冒頭文章外から私生活に関するキーフレーズを抽出することを考える。そこで、表 2 のようなテンプレートを用意し、これを Wikipedia 本文中の文章に適用する。表 2 にある属性の種類は、実際に放送番組の映像を調査した際に、映像内で紹介されていた内容である。表 2 の各テンプレートの “*” はワイルドカードであり、このテンプレートを用いてキーフレーズを含む文を検出する。

趣味や副業、慈善活動などに関するキーフレーズを抽出するため、まずは冒頭文章外の文章を 1 文ごとに区切り、表 2 の各テンプレートと照合する。照合した結果、テンプレートを含む文を検出できれば、その文に形態素解析を施す。その結果、名詞（固有名詞を含む）と判断された語句を、その属性に関するキーフレーズとして抽出する。

それに対し、作品・著作物の標題や配偶者に関するキーフレーズについては、各テンプレートが Wiki 記法を含んでいるため、それらを使って照合した後に、一致した文に対して 3.2.1 項の手法と同様に強調表現やリンクのアンカテキストをキーフレーズとして抽出する。

3.2.3 追加キーフレーズ付与

3.2.2 項におけるテンプレート照合により、特定の属性に関するキーフレーズ抽出を行った。このとき、Wikipedia 記事中の小見出しが表 2 の各テンプレートと完全に一致する場合には、小見出しがそのままキーフレーズとして抽出されてしまうことがある。しかし、これはキーフレーズとして有用でないため、

表 3 キーフレーズ分類のためのテンプレート例

分野	テンプレート
功績	“* career *”, “* award *”, “* honors *”, “* public *”, “* prize *”, “* work *”
私生活	“* personal life *”, “* marriage *”, “* love *”, “* fashion *”, “* partner *”, “* lifestyle *”, “* domestics *”, “* activism *”, “* philanthropy *”, “* supporter *”

より具体的な内容を表すキーフレーズを付与する必要がある。これを、追加キーフレーズと呼ぶ。

追加キーフレーズは、キーフレーズとして抽出された小見出しが付いていた文章中から抽出する。その各文に対して、表 1 を用いて強調表現やリンクのアンカテキストを抽出した後、これらを追加キーフレーズ候補とする。次に表 2 のうち、見出し語と完全に一致したテンプレートが属する属性の、全てのテンプレートを用意する。これらを用いて追加キーフレーズ候補と照合し、一致したものを追加キーフレーズとして付与する。

このとき、表 2 の各属性のテンプレートだけでは種類が少ないため、追加キーフレーズ候補のどれとも一致しない可能性がある。そこで、Word2Vec [6] を用いて各属性のテンプレートの共起語を求め、テンプレートを拡張する。これを拡張テンプレートと呼ぶ。この拡張テンプレートを用いて追加キーフレーズ候補と照合し、一致したものを追加キーフレーズとして抽出する。

3.2.4 キーフレーズの分野分類

3.2.1 項で抽出したキーフレーズは、冒頭文章から抽出した。冒頭文章には功績と私生活の記述が混在しているため、各キーフレーズをどちらかに分類する必要がある。そこで、Wikipedia 記事中の小見出しに対してテンプレートマッチングを適用する。このとき、予め功績と私生活に関連する語句を含む表現をテンプレートとしてそれぞれ用意しておき、それらと一致した小見出しがあれば、その直下の文章中に存在するキーフレーズは功績と私生活のいずれかに関係するものとして分類する。表 3 に、分野分類に用いるテンプレートの一部を示す。最後に、分類された各キーフレーズと、冒頭文章内から抽出されたキーフレーズとを比較し、その一致に基づいて功績か私生活に分類する。

4. Web 画像検索を用いた画像の収集・選択

キーフレーズ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フレーズの内容と一致したものを選択したい。このとき、そのキーフレーズに関する画像モデルを構築するために、まず「キーフレーズ」のみをクエリとした画像検索を行う。この結果得られた上位 N 枚の画像を、「一般的なキーフレーズ画像群」と呼ぶ。この N 枚の画像群からモデルを構築することで、「キーフレーズを表す一般的な画像内容」を表現する。次に、「人物名+キーフレーズ」をクエリとした画像検索を行い、その結果得られた上位 M 枚の画像を、「個人に特化したキーフレーズ画像群」と呼ぶ。この M 枚の画像群から「キーフレーズを表す個人に特化した画像内容」を表現する。その後、個人に特化したキーフレーズ画像群から一般的なキー

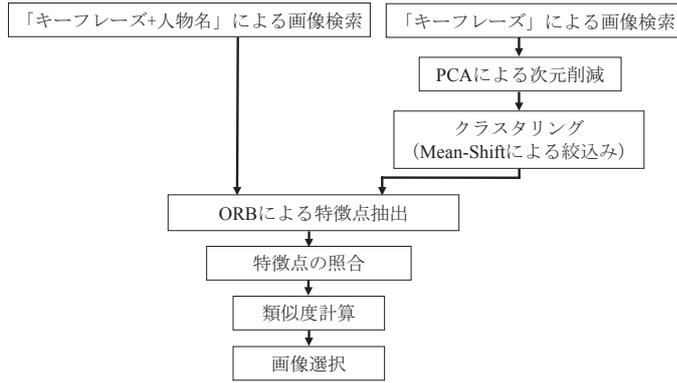


図3 局所特徴量による画像選択手順.

フレーズ画像群との類似度が最も高い1枚を選択する.

このとき、画像検索のクエリに用いたキーワードが属する分野によって、画像内容の表現方法及び照合方法を切り替える.

4.1 局所特徴量を用いた画像特徴の表現による画像選択

クエリに用いたキーワードが功績に関する場合、局所特徴量を用いて著名人とキーワードの関係を表す画像を選択する. 図3に、局所特徴量を用いた画像選択手順を示す.

4.1.1 画像検索結果の雑音除去

一般的なキーワード画像群には、画像検索結果のランキング上位のものを用いているため、キーワードとは関係ない画像がランキング上位に出現することもある. そこで、一般的なキーワード画像群を、より視覚的に類似するものだけに絞り込むために、Mean-shift法[7]を用いたクラスタリングを施す.

まず、クラスタリングの前に、画像群中の全画像に主成分分析(PCA)[8]を施す. これにより、画像群を表す主成分を抽出し、後のクラスタリングでその特徴に基づいて分類することができる. 本研究では、入力画像を双1次補間法によって50画素四方に拡張した後、2,500次元のベクトルとする. これをあらかじめ設定した累積寄与率 E になるまで次元を削減する.

クラスタリングの後、各クラスターのうち要素数が最大であるものに属する \hat{N} 枚の画像群から、キーワードを表す一般的な画像特徴を抽出する.

4.1.2 特徴点抽出と類似度の算出

画像特徴として、まずORBにより特徴点を抽出する. その後、キーワードを表す一般的な画像特徴と個人に特化した画像特徴を照合する. 検出された特徴点のうち、照合で対応づいたものの割合を類似度とする. クラスタリングによる絞り込みを行った後の一般的なキーワード画像 \hat{x}_i ($i \in \hat{N}$)から検出された特徴点の数 $P_{\hat{x}_i}$ と、個人に特化したキーワード画像 x_j ($j \in M$)から検出された特徴点の数 P_{x_j} の類似度の算出方法を以下に示す.

$$S(x_j, \hat{x}_i) = \frac{2M_{(x_j, \hat{x}_i)}}{P_{x_j} + P_{\hat{x}_i}} \quad (1)$$

ここで、 $M_{(x_j, \hat{x}_i)}$ は、一般的なキーワード画像 \hat{x}_i と、個人に特化したキーワード画像 x_j の2枚において、対応づいた特徴点の数を表す. 類似度 $S(x_j, \hat{x}_i)$ は、1対の画像間の類似度なので、1枚の画像 x_j と一般的なキーワード画像群全体

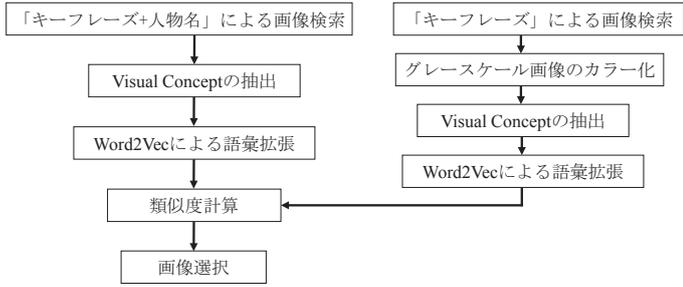


図4 Visual Conceptによる画像選択手順.

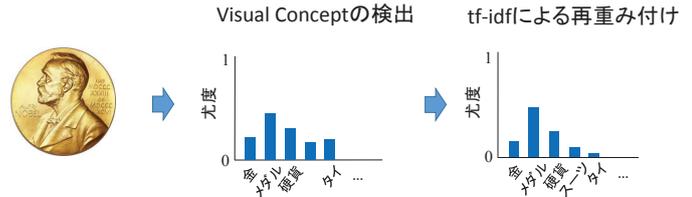


図5 Visual Concept 尤度の重み付けの様子

$\{\hat{x}_i | i \in \hat{N}\}$ の類似度は、 $i = 1, 2, \dots, \hat{N}$ として、 $S(x_j, \hat{x}_i)$ の平均をとることで得られる. これを M 回繰り返すことで、個人に特化した画像群の全ての画像とキーワードとの類似度を算出する. 最終的にこの類似度が最も高い画像を、人物紹介映像に用いる視覚素材として選択する.

4.2 Visual Conceptを用いた画像内容の記述による画像選択

クエリに用いたキーワードが私生活に関する場合、Visual Conceptによる一般物体認識結果により画像内容を記述することで著名人とキーワードの関係を表す画像を選択する. 図4に、Visual Conceptによる画像選択手順を示す.

4.2.1 グレースケール画像の着色

画像検索結果にはグレースケール画像が混入することがあるが、Visual Conceptを精度よく検出するためには、色情報が重要である. そこで事前処理としてIizukaらが提案したAIC(Automatic Image Colorization)[9]を用いて着色を行う.

4.2.2 Visual Conceptへの再重み付け

検出したVisual Conceptの各尤度に対し、tf-idf[10]を用いて重み付けを行う. 図5に、tf-idfを用いてVisual Conceptの各尤度に重み付けを行う様子載せる. 本来、tf(Term Frequency)は文章中のある単語の出現回数、idf(Inverse Document Frequency)はその単語を含む文章数の逆数である. 本研究では画像から抽出されるVisual Conceptに対して適用するため、tfとidfを式(2)のように定義した.

$$tf_i = v_i \quad (2)$$

$$idf_i = \log \left(\frac{K}{|\{d | i \in d, d \in K\}|} \right) \quad (3)$$

ここで、 v_i はある画像から検出されたVisual Concept i の尤度、 K はVisual Conceptを抽出する画像の総枚数、 d は i を上位 U 位以内に検出した画像枚数を表す. 最終的なVisual Concept i への重み ω_i は、 tf_i 、 idf_i の値を乗じたものとなる. ここで、

Visual Concept の重みとなる idf 値を算出するためのデータセットは、GoogLeNet モデルの評価用テストデータ 100,000 枚を用いた。

4.2.3 Visual Concept の語彙拡張

検出した上位 U 位の Visual Concept の語彙のみでは画像内容を表現しきれないため、Word2Vec [6] 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。具体的には、まず上位 U 位の各 Visual Concept を対象単語としたときの周辺単語を Word2Vec を用いて抽出し、共起語とする。各共起語の共起確率と対象単語の Visual Concept の尤度を乗算し、ヒストグラムに追加する。この Word2Vec による拡張が行われた後の画像内容のヒストグラムを、画像内容記述子と呼ぶ。図 6 に、Word2Vec による拡張を行い、画像内容記述子を生成する様子を示す。

4.2.4 Visual Concept を用いた画像内容記述と類似度の算出

個人に特化した画像内容記述子の記述は、個人に特化したキーフレーズ画像群からその都度行う。1 枚の画像の画像内容記述子を記述する場合、尤度が上位 U 位の Visual Concept について、その尤度を用いて各 Visual Concept の尤度ヒストグラムを生成する。一方、キーフレーズを表す一般的な画像内容記述子の記述は、一般的なキーフレーズ画像群からまとめて行う。このような画像群から画像内容記述子を記述する場合は、先ほどと同様に、各画像から作成した Visual Concept の尤度ヒストグラムを全て加算する。最後に正規化を施し、画像群の尤度ヒストグラムを生成する。尤度ヒストグラムを生成後、4.2.3 項にて述べた方法で語彙を拡張し、最終的な画像内容記述子とする。

最後に、以下の式を用いてキーフレーズを表す一般的な画像内容記述子 \mathbf{H} と個人に特化した画像内容記述子 $\mathbf{H}_j (j \in M)$ を比較し、類似度を算出する。

$$S(\mathbf{H}, \mathbf{H}_j) = 1 - \frac{(\mathbf{H} - \mathbf{H}_j)^2}{m} \quad (4)$$

ここで、 m は個人に特化した画像群における $(\mathbf{H} - \mathbf{H}_j)^2$ の最大値である。この類似度が最も高い画像を、人物紹介映像に用いる視覚素材として選択する。

5. 評価実験

5.1 キーフレーズ抽出手法に関する評価実験

提案した Wikipedia 記事からのキーフレーズ抽出手法の評価を行い、その有用性を検証した。なお、DBpedia から抽出した基本情報については信用して評価しない。

評価のために、29 名の著名人のデータセットを構築した。このデータセットに含まれる人物の職業を、専門家、芸術家、スポーツ選手、公人の 4 つに分類した。各分類の基準は、その著名人の主な職業がスポーツに関係するものならスポーツ選手、絵画や音楽、執筆に関係するものなら芸術家、首相や軍人などの公職に関係するものなら公人、起業や学問、料理などに関するものなら専門家とした。

提案したキーフレーズ抽出の評価を行うために、各人物の人

表 4 職業ごとの平均誤取得率と平均誤分類率

職業	平均誤取得率	平均誤分類率
専門職	14.0%	2.6%
公人	20.6%	4.3%
芸術家	10.1%	3.4%
スポーツ選手	17.0%	4.2%
平均	14.6%	3.4%

物名を用いてキーフレーズを抽出した。得られたキーフレーズに対して、次の評価指標を用いて、提案手法の評価を行った。以下に、本実験で用いた評価指標を示す。

(1) 誤取得率：取得した全てのキーフレーズのうち、著名人の人物像を表すのに関係しない名詞を誤って取得した割合

(2) 誤分類率：取得した全てのキーフレーズのうち、功績と私生活を表すキーフレーズの分類を誤ったキーフレーズの割合

キーフレーズが人物像を表すのに関係しない名詞か否かの判別は、当該人物の Wikipedia 記事を読み、主観的に判断した。

表 4 にデータセット中の各人物について、抽出したキーフレーズに関する誤取得率と誤分類率の全体平均を載せる。平均誤取得率は 14.6%、平均誤分類率は 3.4%であった。最終的に人手で誤ったキーフレーズを除くことを考えると、この程度の誤りは大きな手間にはならないため、自動でキーフレーズを抽出するシステムとして十分な精度と考える。表 4 のうち、公人やスポーツ選手などの誤分類率が高いことから、主な職業が一般生活に近い場合はキーフレーズとして固有名詞が現れにくいのではないかと考えられる。一方で専門家や芸術家などは、キーフレーズとして抽出されたものが研究内容や作品タイトルなどの固有名詞であることが多かったため、低い誤取得率につながったと考えられる。

5.2 画像選択手法に関する評価実験

提案した画像選択手法において、キーフレーズの分野によってモデルを変えたことの有効性と、それぞれの特徴を適切に利用できたか評価した。本評価のために、被験者実験によりデータセットを構築した。データセットには、5.1 節で用いた 29 名の各人物と関係するキーフレーズを目視で設定し、32 個のクエリを生成した。このクエリを bing 画像検索^(注3)に入力し、検索結果の上位 4 枚の画像を個人に特化したキーフレーズ画像群のデータセットとした。データセットの各画像群に対し、被験者 12 名による真値付けを行った。被験者は、クエリに対し、上位 4 枚の中から最もクエリをよく表していると思う画像を 1 枚選択した。これにより、個人に特化したキーフレーズ画像群の 4 枚の画像それぞれに、被験者ごとの評価が付与された。

このようにして構築したデータセットに対して、提案手法と以下の比較手法を用いて、人手による選択の再現性を評価した。実験方法を以下に述べる。データセット中の個人に特化したキーフレーズ画像群中の 4 枚の画像から、提案手法を用いて 1 枚選択する。これを全ての画像群に対して行い、平均一致率

(注3) : <https://www.bing.com/>

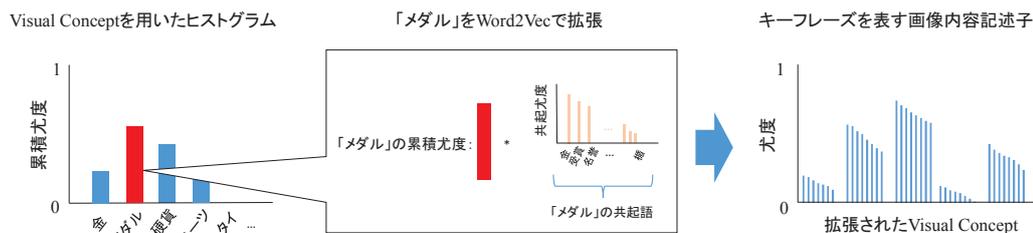


図 6 Word2Vec による画像内容記述の拡張の様子。

表 5 異なる画像の表現方法を用いて画像選択を行った場合の平均一致率

手法	平均一致率
提案手法	38.8%
比較手法 1 (画像検索第 1 位)	26.6%
比較手法 2 (SIFT)	25.4%
比較手法 3 (AKAZE)	27.8%
比較手法 4 (ORB のみ)	36.8%
比較手法 5 (Visual Concept のみ)	28.7%

を用いて評価した。平均一致率とは、各被験者による選択画像と、提案手法による選択画像が一致した割合を平均したものである。このとき、被験者実験によって評価された各クエリで最も多くの被験者に評価された画像を 1 枚選び、真値としたときの平均一致率は 65.9%であった。

比較手法は、キーフレーズの分野によらず、画像選択を行う際に、SIFT, AKAZE, ORB の各特徴量による画像特徴の表現、または Visual Concept による画像内容の記述を用いた。さらに、各クエリに対して画像検索エンジンが出力したランキング 1 位の画像を常に選択する方法も比較手法とした。

表 5 に評価結果を示す。これより、提案手法が 38.8 ポイントと高評価であることから、人手で選ぶ画像に最も近いものを選択できる手法であることが示された。また、用いた特徴量別に見た場合、提案手法で用いた ORB 特徴量と Visual Concept の平均一致率がそれぞれ提案手法に次いで高かったことから、それぞれが画像選択のための特徴表現として有用であったことが伺える。さらに、それらと提案手法を比較した場合に、提案手法のようにキーフレーズの分野に応じて用いる特徴を切り替えることの有用性が示唆された。最後に、提案手法において、功績について ORB 特徴量を用いて正しく選択できた例を図 7、私生活について Visual Concept を用いて正しく選択できた例を図 8 に、それぞれ示す。

6. むすび

本報告では Web 上のコンテンツを用いた人物紹介映像の自動生成におけるキーフレーズ抽出手法と画像選択方法について提案した。提案した画像選択手法では、用いた Visual Concept の語彙にないものは検出できなかった。今後は、Visual Concept のモデルを拡張することで、より高い一致率を得られるようにすることが必要である。

謝辞 本研究の一部は、科学研究費補助金及び国立情報学研



図 7 ORB 特徴量を用いて選択された功績に関する画像。



図 8 Visual Concept を用いて選択された私生活に関する画像。

究所との共同研究による。

文 献

- [1] 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, “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,” HCG シンポジウム 2016 論文集, pp.493–497, Dec. 2016.
- [2] S. K. Divvala, A. Farhadi, and C. Guestrin, “Learning everything about anything: Webly-supervised Visual Concept learning,” Proc. 2014 IEEE Conf. on Computer Vision and Pattern Recognition, pp.3270–3277, June 2014.
- [3] 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.
- [4] Y.-G. Jiang, C.-W. Ngo, and J. Yang, “Towards optimal bag-of-features for object categorization and semantic video retrieval,” Proc. 6th ACM Int. Conf. on Image and Video Retrieval pp.494–501, July 2007.
- [5] 村上晴美, 小西利宗, 浦芳 伸, “Web 上の人物の概要文の作成,” 第 30 回人工知能学全大論文集, no.3P1-10in2, 2p., June 2016.
- [6] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” Advances in Neural Information Processing Systems 26, pp.3111–3119, Sept. 2013.
- [7] D. Comaniciu and P. Meer, “Mean shift analysis and applications,” Proc. 7th IEEE Int. Conf. on Computer Vision, vol.2, pp.1197–1203, Sept. 1999.
- [8] N. P. Chandrasiri, R. Suzuki, N. Watanabe, N. Yoshida, H. Yoshida, H. Yamada, and H. Harashima, “Is average face recognized as the average,” Proc. 9th Int. Symposium on Signal Processing and its Applications, pp.1–4, Feb. 2007.
- [9] S. Iizuka, E. Simo-Serra, and H. Ishikawa. “Let there be color!: Joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification,” ACM Trans. Graphics, vol.35, no.4, pp.110, July 2016.
- [10] G. Salton and M. J. McGill, “On the specification of term values in automatic indexing,” J. of Documentation, vol.29, no.4, pp.351–372, Dec. 1993.