

画像検索におけるファセットの自動抽出と多様性計算

金 応教[†] 山本 岳洋^{††} 田中 克己^{††}

[†] 京都大学工学部情報学科 〒 606-8501 京都府京都市左京区吉田本町

^{††} 京都大学大学院情報学研究科社会情報学専攻 〒 606-8501 京都府京都市左京区吉田本町

E-mail: [†]{kim,tyamamot,tanaka}@dl.kuis.kyoto-u.ac.jp

あらまし 本論文では、画像検索における、画像ファセットの抽出手法と画像のファセットの多様性計算手法を提案する。現在、画像検索システムでは、ユーザの検索意図を検索質問にフィードバックする手法として、画像のコンテンツや画像に付与されたタグや周辺テキストを元にした適合性フィードバックが用いられている。これにより、ユーザの検索要求に合致した画像を容易に見つけることができる。しかし、このような適合性フィードバックでは、“この画像の多様な観点に基づいて他の画像を探したい”という検索要求に応えることは困難である。そこで、本研究では、1つの画像から様々なファセットを抽出する手法を提案する。また、抽出されたファセットに基づき、画像のファセット多様性を計算する手法を提案する。

キーワード 画像検索, ファセット, ファセット多様性

1. はじめに

ウェブの発展にともない、ウェブ上に存在する画像の数も増加している。例えば、カメラつき携帯デバイスの普及により、人々は容易に写真をウェブ上に投稿することができる。近年では、Flickr^(注1)やInstagram^(注2)などに代表される、写真共有サービスなども広く用いられるようになっており、我々は膨大な数の画像にウェブを通してアクセスすることができる。

膨大な画像集合から求める画像を検索するため、Yahoo!, Google, Bing といった検索エンジンでは、テキストクエリに基づく画像検索機能をユーザに提供している。また、Flickr では画像につけられたタグを利用して、画像を検索することができる。さらに、一部の検索エンジンでは、テキストクエリに基づく検索だけでなく、ユーザが選択した画像に基づき、類似画像を検索するという、コンテンツベースの適合性フィードバックに基づく画像検索も行うことができる。こうした画像検索機能を用いることで、ユーザはある特定の観点に基づく検索や、“この画像と類似した画像が欲しい”といった情報要求を満たすことができる。

しかし、現在の画像検索システムでは、“この画像の多様な観点に基づいて他の画像を探したい”という情報要求に対応することができない。例えば、日本への観光を計画しており、観光地をいくつか探しているユーザを想定する。そのユーザはまず、“日本 ∧ 観光地”というクエリで画像検索を行い、その検索結果の画像集合を閲覧し興味を引く画像を探すと考えられる。その中で、“厳島神社の大鳥居”の画像に興味を引かれ、それに関する観光地の画像を探そうと考えた場合、以下の様に、多くの

の情報要求が考えられる。

- (1) “厳島神社の大鳥居”の画像をもっと他に見てみたい
- (2) 厳島神社だけでなく“広島”にある他の観光地の画像を見たい
- (3) 鳥居がきれいなので、他の“鳥居”に関する観光地の画像を見たい
- (4) 厳島神社だけでなく、他の“世界遺産”に関する観光地の画像を見たい

現在のコンテンツベースの適合性フィードバックでは、1件目の情報要求にしか応えることができない。また、他の“広島”や“鳥居”といった観点が、ユーザにとって知らなかったり気づいていない場合はテキストクエリを入力することができない。上記のような、ある画像に関するさまざまな情報要求に基づく検索を可能にするためには、ユーザが選択した画像に関するファセットを提示することが重要であると考えられる。ファセットとは、階層的な分類構造であり、たとえば、“広島”という概念の下に“厳島神社”や“平和記念公園”といった階層的な構造をユーザに提示することができれば、上記で述べたような情報要求を満たすことができると考えられる。このような、ファセットを提示することが有効な画像検索として、他にも以下のような例が挙げられる。

- 花壇に植える花を探す場合、気に入った花を選択し、それに対する色、種、開花時期などのファセットを提示する
- パーティーの準備で料理を探す場合、良さそうなものを選択し、それに対する食べ物の種類、材料、調理方法などのファセットを提示する

このようにユーザが選択した画像に対してファセットを提示することは、Web 画像検索において有効ではないかと考えられる。

そこで、本稿では、画像に対するファセットを抽出し、有効なファセットをユーザに提示する手法を提案する(図1)。ファセッ

(注1): <http://www.flickr.com/>

(注2): <http://instagram.com/>

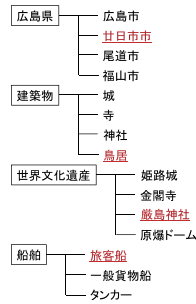


図 1 ファセット提示のイメージ図

トとは物事を見る側面を意味するもので、1つの画像でも様々なファセットを持つことになる。提案手法では、写真投稿サイト上に投稿された画像についているタグに着目し、ファセットを抽出する。しかし、タグにも画像の特徴とは関係のないものが多く存在し、すべてのタグがファセットに属する語として相応しいとは限らない。そこで、まず画像の特徴を表すような語を抽出した上で、ファセットの抽出を行う。その際、画像の特徴を表すタグはファセットというより、ファセットの一例であるファセット値に近いものであるので、得られる語をファセット値としてそれが属するファセットの抽出を行う。

さらに、ファセットを抽出するだけでなく、抽出されたファセットの多様性に基づく画像ランキング手法についても提案する。様々なファセットを持つ画像は、様々な目的に対して使用できる画像であると考えられる。検索結果の上位の画像が観点をほとんど持っていないとすると、ユーザが選択した上位画像の観点をユーザに提示したところであまり効果がない。そこで、抽出された観点を有効活用するために、画像に関するファセットの多様性の高い画像を上位にランキングすることで、ユーザの再検索を支援できると考えられる。

以降、本稿の構成は以下の通りである。2章では関連研究を挙げ、本研究との違いについて述べる。3章では本研究で扱う、ファセットおよびファセット多様性について定義する。4章でファセット抽出手法について説明し、5章でファセット多様性の計算手法について述べる。6章で実験及び実験結果について述べ、得られた結果を考察する。最後に、本稿のまとめと今後の課題を述べる。

2. 関連研究

2.1 ファセットによる画像の分類

画像検索にファセットを適用した研究としては、Yee らの研究や川野らの研究がある。Yee らは与えられた画像をあらかじめ決められたファセットに割り当ててことで [1]、また、川野らはクエリに応じてファセットを動的に抽出して割り当ててことで [2] [3]、複数のファセットから画像を検索できるシステムをそれぞれ提案している。これらの研究では、ある決められた画像集合を分類するためのファセットを作成しているのに対し、本研究では、1つの画像に基づいて、再検索をするためのファセットを発見するという点で、彼らの研究とは大きく異なる。本研究で提案するファセット抽出手法を利用することで、多様

な観点による画像検索を支援できるのではと考えている。

2.2 検索結果多様化

近年、情報検索の分野において、検索結果多様化に関する研究が盛んに行われている [4] [5]。それらの研究の多くは、ウェブ文書を対象としたものが多いが、画像検索においても検索結果多様化が取り組まれている。たとえば、van Leuken らは画像特徴量を用いた画像検索結果の多様化手法を提案している [6]。こうした、検索結果多様化に関する研究は、ウェブページ集合や画像集合といった、集合の多様性を考えているのに対し、本研究では、ある特定の画像が属するファセットの多様性を考えているという点で、彼らの研究とは異なる。しかし、画像検索結果の多様化のために、本研究で提案するファセットおよびファセットの多様性を利用することも可能であると考えられる。

3. ファセットとファセット多様性

3.1 ファセット

一般的に物事を見る側面を意味するファセットは、画像検索においては、画像が持つ観点を意味する。言い換えると、画像集合を分類する基準でもあり、例えば、“西洋の絵画全体”の画像集合があるとすると、“時代”、“地域”、“主義”、“技法”、“素材”などがファセットとなりうる。ファセットはファセット名とファセット値を持つ、階層的な分類構造である。ファセット名はファセットを表す語、即ち画像が持つ観点の名前である。ファセット値はファセットに属する一例、即ち画像が分類される項目である。例えば、“西洋の絵画全体”の画像集合では、ファセット名“技法”のファセットにはファセット値“油彩”、“水彩”などがあり、ファセット名“主義”のファセットにはファセット値“古典主義”、“印象主義”、“キュビズム”などがある。また、ある画像は複数のファセットに属することもある。例えば、フィンセント・ファン・ゴッホの“星月夜”の画像は、ファセット名“技法”のファセット値“油彩”、ファセット名“主義”のファセット値“ポスト印象主義”に分類される。

このように、ファセットはファセット値の集合で構成されることができる。本稿におけるファセットの定義を以下に表す。

[定義 1] 画像が持つ観点を表す同位の語からなる語集合 $F_i = \{f_i^{(1)}, f_i^{(2)}, \dots, f_i^{(m)}\}$ をファセットという。ここで、各語 $f_i^{(j)}$ をファセット値といい、語集合 F_i を表す語 $\text{Name}(F_i)$ をファセット名という。

また、ファセット集合を以下のように定義する。

[定義 2] ファセットによって構成される集合 $\mathcal{F} = \{F_1, F_2, \dots, F_n\}$ をファセット集合という。

3.2 ファセット多様性

本研究の目的は、様々な目的の再検索に利用できる画像を発見することである。これは、多様な観点を持っている画像を発見することで、実現できると考えられる。本研究では、多様な観点を持つ画像として、以下に示す条件を考えた。

- (1) オブジェクトが多く写っている
- (2) 各オブジェクトに関連する単語が多く思い浮かぶ
- (3) 思い浮かぶ単語の間は関連性がないものが多い

本研究で提案するファセット多様性とは、上記の条件を考慮した、画像に対するファセット集合がいかに様々なファセットで構成されているのかを意味する概念である。ファセット集合で、ファセットの間に上下関係が存在するものや、ファセットの間に共通する部分が多く存在するもの、そもそもファセットの数が少ないもの、重要ではないファセットだけを持っているものなどは、多様な観点を持っているとは言えない。

本稿で定義するファセット多様性を以下に示す。

[定義3] ファセット多様性とは、ある画像に属するファセット集合が、多様な観点を持っていることをいう。ファセット集合が、より多様な観点を持っているとき、ファセット多様性が高いという。

4. ファセット抽出

本章では、画像検索におけるファセットの抽出手法について述べる。まず、画像検索のための効果的なファセットについて述べる。その後、提案手法の概要を説明した後、手法の詳細を述べる。なお、本稿では、写真共有サービスの1つであるFlickrにアップロードされた画像を対象とする。Flickrでは、画像に対してユーザが自由にタグを付与することができる。

4.1 画像検索のための効果的なファセット

多様な観点を持っている画像からは、多数のファセットが抽出できる。しかし抽出されたファセットの中には、画像とは関連性が低いファセットも存在する。そのようなファセットは画像をよく表しているとは言えない。また、画像とある程度の関連性のあるファセットに対しても、ファセットによって画像に対する重要度が異なる。

以上のことを考慮するために、本研究では画像検索における効果的なファセット値と効果的なファセット名の条件を提案する。これらの条件を数値で表したものを、本稿ではファセットのスコアと呼ぶことにし、その計算方法の詳細は4.5節で述べる。

4.1.1 効果的なファセット値

画像から得られるファセット値の中には、画像の特徴を表すのに適切ではないものも存在する。例えば、画像の投稿者名や写真を撮ったカメラの名前などがタグとして使われていることがある。また、家族旅行で訪ねた観光地の写真の場合、周辺テキストやタグに“家族”というコンテキストを表す語が存在することもある。それらの情報は画像そのものの観点を表しているとは言えない。

また、画像をよく表しているファセット値であっても、ユーザの検索目的とは全く関係がない場合がある。例えば、“花”というクエリで画像検索を行い、“薔薇”と“車”が写っている画像が得られたとする。“車”という語はクエリ“花”との関連性が低いので、ユーザの検索目的に合うものとは言えない。

このような、画像そのものとは関連性の低いメタ情報やコンテキストによるファセット値と、ユーザの検索目的に合わないファセット値を取り除くため、ファセット値の中から画像と検索目的に関連する効果的なファセット値を発見する必要がある。本稿で提案する、効果的なファセット値の条件を以下に示す。

クエリとの適合性 効果的なファセット値は、クエリとの適合性が高い。クエリと全く関連性がないようなファセット値は、画像検索を行うユーザの目的に合わない可能性が高い。

類似画像集合内での典型性 効果的なファセット値は、そのファセット値が得られた画像の類似画像集合内での典型性が高い。言い換えれば、端的に類似画像集合を表しているものが効果的なファセット値である。類似画像集合をよく表しているということは、元の画像もよく表しているとも言える。メタ情報やコンテキストによるファセット値は、この典型性が低くなると考えられる。

共用性 効果的なファセット値には、多くのユーザが使用できるものが望ましい。多くの画像から抽出できるファセット値でも、1つのユーザだけが使う語は主観的すぎるもので、画像の観点を表しているとは言えない。

4.1.2 効果的なファセット名

効果的なファセット値を含むファセットを抽出できたとしても、それが各画像の特徴を表すのに適切なファセットであるとは言えない。また、同じファセット値でも複数のファセットに属することがある。そのため、単にファセット値の良さを計算するだけでなく、ファセットとしての良さも考慮してファセットの重要度を求める必要がある。本研究では、このファセットの重要度を、ファセットを一言で表すファセット名から求める。本稿では、画像検索における効果的なファセット名の条件として、以下を提案する。

クエリとの適度な重なり 画像から得られたファセットを再検索に用いる場合、元のクエリの画像集合と関連し、かつ、元のクエリの画像集合には存在しない画像を検索可能なファセットが再検索に望ましいと考えられる。例えば、あるファセット名で検索したとしても、元のクエリで得られる画像集合やその部分集合しか得られないようなファセットは、再検索には適していないと考えられる。また、元のクエリと全く関連しない画像集合ばかりが得られてしまうようなファセットも、ユーザの検索目的に合うとは言えず、抽出すべきファセットとして相応しくないであろう。つまり、クエリが対象とする画像集合と、ファセット名が対象とする画像集合が適度に一致しているようなファセット名が、再検索に相応しいと考えられる。

4.2 ファセット抽出の概要

まず、本手法の問題設定を述べる。入力となるクエリを q 、 q によって得られる検索結果の画像集合を $\text{Ans}(q)$ とし、ある1つの画像を p ($p \in \text{Ans}(q)$) とすると、クエリ q と画像 p からその画像の観点を表すファセット集合 $\mathcal{F} = \{F_1, F_2, \dots, F_n\}$ を生成し、各ファセットのスコア $S(F_i)$ を得ることが本手法の目的である。

図2に提案手法の概要を示す。図2に示すように、本手法は大きく分けて、ファセット値の抽出、ファセット名の取得、ファセットのスコアリングの3つのフェーズに分けられる。以降、それぞれの手法について説明する。

4.3 類似画像集合を用いたファセット値の抽出

画像の観点を表すファセットを抽出するため、まずファセット値を抽出する。本手法では、画像に付与されているタグを用

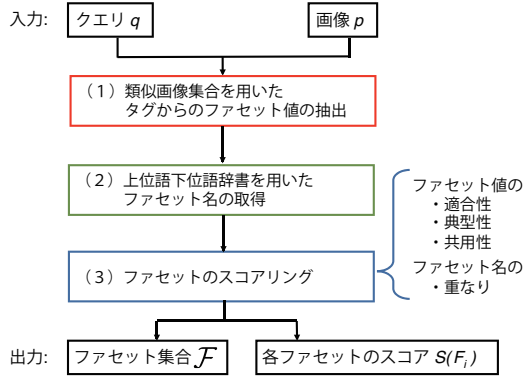


図2 ファセット抽出の流れ

いてファセット値を抽出する。この時、ある画像には豊富な数のタグが付与されていることもあれば、ほとんどタグが付与されていないタグも存在する。画像に付与されたタグが少ない場合、その画像を表す十分な数のファセットを抽出することができなと考えられる。しかし、ある画像とよく似た画像は、同様のファセットに属する可能性が非常に高いと考えられる。そこで、対象画像の類似画像集合に付与されているタグもファセット値の候補とすることで、より多くのタグをファセット値の候補とする。

いま、画像 p がクエリ q によって得られる検索結果の画像集合 $\text{Ans}(q)$ 内の要素だとすると、画像特徴量を用いることで、画像 p の類似画像集合 $\text{Ans}_{\text{similar}}(p) \subset \text{Ans}(q)$ を得ることができる。本稿では、画像 p および、類似画像集合 $\text{Ans}_{\text{similar}}(p)$ で用いられている全てのタグを対象とし、ファセット値の候補とする。以降、得られたタグの語集合を T で表す。

なお、本稿では、画像特徴量として、画像の RGB ヒストグラム、HSV ヒストグラム [7]、ILSVRC2010 の画像コレクション [8] から構築した、Bag of Visual Words [9] 特徴量を用い、 $\text{Ans}(q)$ 内の類似画像を $\text{Ans}_{\text{similar}}(p)$ として用いる。

4.4 上位語下位語辞書を用いたファセット名の取得

得られたファセット値 $t_k \in T$ を用いて、語 t_k をファセット値として持つようなファセット F_i ($t_k \in F_i$) を求める。本研究では、ファセット値からファセットを求めるために、上位語下位語辞書を利用する。辞書には、ALAGIN フォーラムから提供されている上位語階層データ^(注3)を用いる。このデータは、Wikipedia で記事の見出し語やカテゴリ名となっている名詞句をその上位語、下位語関係に基づいて階層化したものであり、これを用いることで、ある語の上位語や下位語を求めることができる。

まず、ファセット値 t_k の上位語を辞書で調べて得られる語のすべてをファセット名 $\text{Name}(F_i)$ とする。次に、ファセット名 $\text{Name}(F_i)$ とファセット値 t_k を用いて、ファセット F_i を構成する。すべてのファセット値 t_k に対して上記の手法を実行すると、ファセット集合 $F_0 = \{F_1, F_2, \dots, F_m\}$ が得られる。最後に、画像集合 $\text{Ans}(\text{Name}(F_i))$ が得られないファセットを取

り除くことで、最終的にファセット集合 $F \subseteq F_0$ を得る。

4.5 ファセットのスコアリング

これまでの過程で、ファセット F_i を抽出する際に使われたファセット値 t_k とそのファセット名 $\text{Name}(F_i)$ 、画像集合 $\text{Ans}(\text{Name}(F_i))$ が存在するファセットがなすファセット集合 F を得ることができる。しかし、得られたファセットの中には画像と関連の無いものや、画像検索において有益でないもの、ユーザの検索目的に合わないものも多く存在する。

本節では、4.1 節で提案した効果的なファセットの定義に基づき、得られたファセット F_i に対して、画像のファセットとしての重要度 $S(F_i)$ を求める手法を提案する。手法では、ファセット F_i に対して、ファセットを抽出する際に使われたファセット値の重要性和ファセット名の重要性的の両者を考慮し、 $S(F_i)$ を求める。なお、本節で用いる集合の類似度 Sim_{set} には、Dice 係数を用いる。

4.5.1 ファセット値に基づくスコア

4.1.1 節で提案した効果的なファセット値の条件に基づいてスコアを定義するため、クエリとの適合性、類似画像集合内での典型性、共用性を表す各スコアを考える。本稿で提案する、ファセット値の各条件によるスコアを以下に示す。

クエリとの適合性 ファセット値とクエリの適合性は、ファセット値とクエリがどの程度類似しているのかを表す。ファセット値 f によって得られる画像集合 $\text{Ans}(f)$ とクエリ q によって得られる画像集合 $\text{Ans}(q)$ の類似度を用いると、その適合性を定量的に表すことができる。ファセット値 f に対するクエリ q との適合性のスコア $S_{\text{rel}}(f)$ を以下に示す。

$$S_{\text{rel}}(f) = \text{Sim}_{\text{set}}(\text{Ans}(f), \text{Ans}(q))$$

類似画像集合内での典型性 ファセット値の類似画像集合内での典型性は、ある画像が持つファセット値がどれほどその画像の類似画像集合をよく表しているのかを指す。本稿ではタグをファセット値として用いるため、類似画像集合内でタグ(ファセット値) f が用いられている画像の比率をもって典型性を表す。画像 p の類似画像集合を $\text{Ans}_{\text{similar}}(p)$ とした場合、ファセット値 f に対する典型性のスコア $S_{\text{typ}}(f)$ を以下に示す。

$$S_{\text{typ}}(f) = \frac{|\{p_f | p_f \in \text{Ans}_{\text{similar}}(p), p_f \text{ はタグ } f \text{ を持つ}\}|}{|\text{Ans}_{\text{similar}}(p)|}$$

共用性 ファセット値の共用性は、そのファセット値がどれほど多くのユーザから用いられているのかを表す。一般的な画像検索の場合は、検索結果に現れる各画像の情報源が異なり、画像の投稿者を表す ID が存在しないため、それらの投稿者を区別するのが難しい。この場合、画像の元ページのドメインなどを用いて投稿者を表すことになる。しかし本稿では Flickr を用いるため、画像の投稿者の ID を容易に得ることができる。そのため、ファセット値 f を持つ画像の投稿者の ID を用いて、ファセット値を持つ画像の集合に対する投稿者の数をもって共用性を表す。ファセット値 f によって得られる画像集合を $\text{Ans}(f)$ 、画像 p の投稿者を $u(p)$ とした場合、ファセット値 f に対する共用性のスコア $S_{\text{com}}(f)$ を以下に示す。

(注3): <https://alaginrc.nict.go.jp/>

$$S_{\text{com}}(f) = \frac{|\{u(p_f) | p_f \in \text{Ans}(f)\}|}{|\text{Ans}(f)|}$$

上記の三つの式をパラメータを用いてまとめたものを、ファセット値 f によるファセットのスコアとして計算する．このスコア $S_{\text{value}}(f)$ を以下に示す．

$$S_{\text{value}}(f) = \lambda_{\text{rel}} S_{\text{rel}}(f) + \lambda_{\text{typ}} S_{\text{typ}}(f) + \lambda_{\text{com}} S_{\text{com}}(f)$$

(ただし, $\lambda_{\text{rel}} + \lambda_{\text{typ}} + \lambda_{\text{com}} = 1$)

4.5.2 ファセット名に基づくスコア

4.1 節で述べたように、ファセットのスコアはファセット値とファセット名の両者を考慮して計算する．ファセット名に基づくスコアは、4.1.2 節で提案した、クエリとファセット名との適度な重なりに基づいて計算する．

クエリとファセット名との画像集合の適度な重なりは、クエリ q による画像集合 $\text{Ans}(q)$ とファセット名 $\text{Name}(F)$ による画像集合 $\text{Ans}(\text{Name}(F))$ の共通部分が、適切な大きさを持っているのかを表す．集合がどれほど共通しているのかは集合の類似度を用いれば良いが、4.1.2 節で述べたように、二つの画像集合が全く異なるものでも、完全に一致するものでも効果的なファセット名とは言えず、適度な数の共通画像を持っているものが効果的なファセット名になる．そのため、集合の類似度をそのまま用いるのではなく、類似度が適度な中程度の値に近いき高い値となるような関数を考える必要がある．

ここで、類似度を引数として受け取る関数 Moder を考える．引数が 0.0 と 1.0 のときに最小値、0.5 のときに最大値になる関数が、上記の条件に合致する関数の一つである．この性質を満たす関数には様々なものがあるが、本稿では正規分布関数に注目し、平均が 0.5、最大値が 1.0 になるよう正規分布関数を修正したものを関数 Moder として用いる．関数 Moder を以下に示す．

$$\text{Moder}(x) = \exp\left(-\frac{(x - 0.5)^2}{0.1}\right)$$

この関数を用いて、クエリ q とファセット名との適度な重なりを式で表したものが、ファセット名 $\text{Name}(F)$ によるスコア $S_{\text{name}}(\text{Name}(F))$ である．ファセット名に基づくスコア $S_{\text{name}}(\text{Name}(F))$ を以下に示す．

$$S_{\text{name}}(\text{Name}(F)) = \text{Moder}(\text{Sim}_{\text{set}}(\text{Ans}(\text{Name}(F)), \text{Ans}(q)))$$

4.5.3 ファセットのスコア

ファセット F のスコアを、ファセットを抽出する際に使われたファセット値 f によるスコア S_{value} の式と、ファセット名 $\text{Name}(F)$ によるスコア S_{name} の式を、パラメータ λ を用いて求める．ファセットのスコアを以下の式で表す．

$$S(F) = \lambda S_{\text{value}}(f) + (1 - \lambda) S_{\text{name}}(\text{Name}(F)) \quad (1)$$

5. ファセット多様性の計算

3.2 節で提案したように、多様な目的の再検索に利用できる画像はファセット多様性が高い．多様な目的の再検索に利用で

きるということは、様々な観点を持っていると言い換えられる．ある画像が多数のファセットを持っており、かつ、持っているファセットが類似していない場合、その画像は多様な観点を持っていると言える．

画像がどれほど多様な観点を持っているかを計算することは、その画像のファセット集合がどれほど類似しているのかを計算することである．本稿では、ファセット間の類似性を計算するため、ファセットによって得られる画像集合の画像特徴量と集合の類似性を用いる．例えば、画像特徴量を利用することで、あるファセットと別のファセットが、画像的に異なる画像を検索することが可能なのかを判断することができる．また、ファセットから得られる画像集合間の集合的な類似を比較することで、互いのファセットの意味的な類似性を計算することができる．また、画像特徴量と集合の類似性は、それぞれ異なる観点からの類似性を評価しており、ファセット間の類似性を評価するためには両者を考慮する必要があると考えられる．

以降、本章では画像特徴量と画像集合の類似に基づいた、ファセット多様性の計算手法を説明する．

5.1 画像特徴量に基づくファセット距離

ファセット多様性を計算するために、ファセット間の違いを表すファセット距離を考える．まず、画像特徴量を用いたファセット距離の計算手法を説明する．本稿では、画像特徴量を用いてファセットの類似度を計算し、その類似度をファセット距離の計算に用いる．

画像特徴量をベクトルとして用いる、ある 2 つのファセット F_i と F_j 間の、画像特徴量に基づくファセット類似度の式を以下に示す．

$$\text{FSim}_{\text{feat}}(F_i, F_j) = \text{Sim}_{\text{vec}}(\mathbf{f}(F_i), \mathbf{f}(F_j))$$

上記の式の $\mathbf{f}(F_i)$ は、ファセット名 $\text{Name}(F_i)$ による画像集合の画像に対する特徴量ベクトルの平均を表す． Sim_{vec} は特徴ベクトルの類似度を意味し、本研究ではコサイン類似度を用いる．

上記の式を用いて、画像特徴量によるファセット距離の式を以下のように表す．

$$\text{FDist}_{\text{feat}}(F_i, F_j) = 1 - \text{FSim}_{\text{feat}}(F_i, F_j)$$

5.2 画像集合の類似に基づくファセット距離

次に、画像集合間の集合的類似に基づいたファセット距離の計算手法について説明する．5.1 節と同様に、画像集合の類似を用いて類似度を計算し、それをファセット距離の計算に用いる．

ある 2 つのファセット F_i と F_j 間の、画像集合の類似に基づくファセット類似度の式を以下に示す．

$$\text{FSim}_{\text{set}}(F_i, F_j) = \text{Sim}_{\text{set}}(\text{Ans}(\text{Name}(F_i)), \text{Ans}(\text{Name}(F_j)))$$

上記の式は、画像特徴量に基づく類似度と非常に似ているが、ベクトルの類似度 Sim_{vec} の代わりに集合の類似度 Sim_{set} を Dice 係数で計算するという点で異なる．

上記の式を用いて、画像集合の類似によるファセット距離の式を以下に示す．

$$\text{FDist}_{\text{set}}(F_i, F_j) = 1 - \text{FSim}_{\text{set}}(F_i, F_j)$$

表 1 実験に使用したクエリの一覧

クエリ	“ケーキ”, “食べ物”, “春”, “京都 寺”, “鳥居”, “厳島神社”, “祭り”, “チューリップ”, “スキー”
-----	--

5.3 ファセット集合の多様性

ファセット多様性が高いファセット集合は、類似していないファセットを多く持つ必要がある。逆に言えば、ファセットの数が少ない場合や、ファセットが類似している場合は、ファセット多様性が低い。

本稿では、ファセット集合のファセット多様性を数値化したものを、ファセット集合 $\mathcal{F}$ の多様度 $\text{Div}(\mathcal{F})$ と定義する。その多様度は、ファセット距離 $\text{FDist}_{\text{feat}}$ と $\text{FDist}_{\text{set}}$ の式を用いて求めるが、重要度の高いファセットの比較を重視するため、ファセットのスコアの式 (1) を重みとして用いる。

また、ファセットの組み合わせの数でファセットの比較を行うため、そのまま比較結果の和をとると、ファセット間の類似度より、ファセット数の方が多様度に与える影響が大きくなってしまう。ファセット間の距離とファセット数の両者を考慮するため、比較結果の和の値をファセット数で割ることで多様度の値を補正する。

以上より、ファセット集合の多様度を以下の式に従い求める。

$$\text{Div}(\mathcal{F}) = \frac{1}{|\mathcal{F}|} \sum_{F_i, F_j \in \mathcal{F}} \min(S(F_i), S(F_j)) \times \{\text{FDist}_{\text{feat}}(F_i, F_j) + \text{FDist}_{\text{set}}(F_i, F_j)\} \quad (2)$$

6. 実 験

本稿では、画像検索結果の各画像に対してファセットの抽出とファセットのスコアリングを行い、得られたファセット集合と各ファセットのスコアを用いてファセット集合の多様度を求め、検索結果の画像のランキングに利用する手法を提案している。

そこで、抽出したファセットから得られたスコアの妥当性と、多様度を用いたランキングの有用性を検証するため、本手法を用いるシステムのプロトタイプを Python を用いて実装し、2 種類の実験を行った。

6.1 実験設定

今回の実験では、画像とその周辺情報を取得するため、Flickr が提供している API^(注4)を用いて各クエリに対する画像検索結果と各画像のタグ情報を取得した。表 1 に実験で用いた 9 個のクエリを示す。各クエリから取得する画像は 1,000 件とし、その 1,000 件の画像の中から等間隔で 20 件選び、各画像に対してファセットの抽出とファセット多様性の計算を行った。

画像に対するファセットの抽出では、クエリから得られた 1,000 件の画像との距離を画像特徴量を用いて求め、距離が小さい 10 件を類似画像集合とし、類似画像集合から抽出されるタグのすべてをその画像のファセット値とした。その後、上位語下位語辞書を用いて各ファセット値に対する上位語をすべて

取得し、形態素解析器 MeCab^(注5)を用いて上位語から名詞句と推定される語を抽出した。上位語と形態素解析で抽出した語のすべてを、ファセット値から抽出されたファセット名として、そのファセット値とファセット名の組をファセットとして抽出した。

ファセットのスコアリングでは、ファセット値 f とファセット名 $\text{Name}(f)$ から得られる画像集合として、それぞれをクエリとして Flickr から得られる画像集合を用いた。4.5 節で定義したスコアの計算式で使われるパラメータとしては、 $\lambda_{\text{rel}} = 0.6, \lambda_{\text{typ}} = 0.2, \lambda_{\text{com}} = 0.2, \lambda = 0.4$ を用いた。

6.2 ファセットのスコアの妥当性評価

まず、4 章で提案した手法によって画像から抽出されるスコアの高いファセットが、本当にその画像を表す妥当なものであるかを評価した。本稿で抽出したファセットはファセット名とファセット値の組で表されるため、ファセット名とファセット値が画像を良く表しているのかを調べることでファセットの妥当性を評価した。

提案手法を評価するため、画像に対するファセット集合を、式 (1) で示したスコア S でランキングし、上位 10 件のファセットを評価の対象とした。比較手法として、まずファセット値によるスコア S_{value} のみでランキングしたファセット集合から選んだ上位 10 件のファセットを用いた。本稿のファセット値はタグから抽出したものであり、ファセット値によるスコアのみを考慮した場合は、画像を表すタグの良さのみを考慮したことになる。また、別の比較対象として、ファセット名によるスコア S_{name} のみでランキングしたファセット集合から選んだ上位 10 件のファセットを用いた。

$S, S_{\text{value}}, S_{\text{name}}$ の各ランキングから得られた計 30 個のファセットを画像とともに評価者に提示し、それぞれのファセットが画像を表している用語として適切なかを適切、不適切の 2 値で回答してもらった。また、30 個のファセットの中から重複するファセットは取り除いて評価者に提示した。

評価者の回答から、画像に対するファセットの適合率と nDCG の平均を求めた。適合率は正解を割合を表し、nDCG は理想のランキングに対するランキングの良さを表す指標で、その適合率と nDCG を用いて各ランキングの場合に対する平均を求めて評価に利用した。本評価では各クエリに対して 3 件の画像を選び、その各画像に対しての適合率と nDCG を求め、計 27 件の画像に対して、得られたファセットの適合率と nDCG の平均を求めた。また、得られたファセットが本研究の目的に適しているのかを分析するため、各ランキングの場合に対する上位 10 件のファセット中のファセット名の種類数の平均とファセット値の種類数の平均も求めた。

各スコアに基づくランキングで得られた結果を表 2 に示す。この表を見ると、適合率と nDCG が最も高い値を持つ場合は、ファセット値によるスコアのみを考慮したランキングを用いた場合である。しかし、この場合に上位にランキングされるファセットはファセット値が同じものばかりであり、表から分かる

(注4): <http://www.flickr.com/services/api/>

(注5): <http://mecab.sourceforge.net/>

表 2 ファセットのスコアに対する評価結果

ファセットの ランキング手法	ファセットの適合性		ファセット名の 種類数の平均	ファセット値の 種類数の平均
	適合率	nDCG@10		
S_{value}	0.60	0.77	9.41	2.33
S_{name}	0.51	0.60	5.11	6.44
S	0.58	0.72	6.93	4.55

表 3 各クエリに対するスコア S の評価結果

クエリ	ファセットの適合性		ファセット名の 種類数の平均	ファセット値の 種類数の平均
	適合率	nDCG@10		
厳島神社	1.00	1.00	6.00	4.00
春	0.77	0.84	8.67	3.00
ケーキ	0.60	0.82	7.00	5.00
鳥居	0.60	0.66	5.33	6.67
スキー	0.60	0.62	8.67	4.40
京都 寺	0.57	0.70	3.33	5.40
チューリップ	0.53	0.71	10.00	3.00
祭り	0.27	0.71	6.33	4.50
食べ物	0.27	0.46	7.00	5.00

ように、上位の 10 件の中に約 2 種類のファセット値しか出現しない。つまり、ファセット値のみに基づくランキングでは、画像に関連する観点が少数しか得られないことになるため、本研究の目的である、様々な観pointsの提示には適していないと考えられる。

また、ファセット名によるスコアのみを考慮したランキングの場合は、ファセット値の種類数はファセット値によるスコアの場合より多くなるが、逆にファセット名の種類数が減ってしまい、これも本研究の目的に適していないと考えられる。さらに、適合率と nDCG の値も他の手法と比べて大きく減少しており、画像を表すファセットが上位にランキングされづらくなっていることが分かる。

一方、本稿で提案したファセット値とファセット名の両方を考慮したスコアでランキングを行った場合は、ファセット値によるスコアの場合と適合率と nDCG の差はほとんどないが、提示されるファセット名とファセット値の種類数が上記の 2 つの場合の間であることが分かる。したがって、本稿で提案したスコアを用いると、様々な観点を提示するという目的に反することなく、画像に関連した適切なファセットを提示することができるのではないかとと思われる。

しかし、本稿で提案したスコアでランキングを行った場合も、適合率の値は 0.58、nDCG の値は 0.72 であり、これはファセット抽出の精度が 6、7 割程度の低い値であることを意味する。表 3 は、スコア S によるランキングの場合の各クエリに対するファセットの評価を表しているが、全体の精度が低い理由はこの表 3 を見て分かるように、抽象的なクエリに対する精度が低いためである。抽象的でも“春”のように日本人なら誰でも“桜”を連想するクエリならファセットの精度が高いが、はっきりしたイメージが存在しない“食べ物”のような抽象的なクエリはファセットの精度が悪くなっていることが分かる。

精度が低い理由としては、まず類似画像集合の精度が考えられる。ファセット値の抽出を行うため、本手法ではまず類似画像集合を求めているが、抽象的なクエリで得られる画像集合が

表 4 多様性を用いた画像ランキングに対する評価結果

ランキング手法	多様性に関する		クエリに対する画像の適合性
	手法の平均得点	適合率	
Flickr の適合性	0.89	0.81	0.82
タグの数	1.00	0.78	0.78
多様度	1.56	0.96	0.97

らは、画像特徴量が類似した画像があまり存在しない場合がある。本手法では距離が小さい上位 10 件を類似画像として選択しているので、その 10 件の中に類似画像とはいえない画像が含まれると、画像とは関連性のないタグがファセット値として抽出されてしまう。そのため、抽象的なクエリに対してはファセットの精度が低くなったのではないかと考えられる。

また、上位語下位語辞書の精度も理由として考えられる。本手法では、各ファセット値に対応するファセット名を辞書を用いて取得しているが、辞書の精度が悪いと、抽出されるファセット名が画像と関係のないものばかりになってしまう。実際、今回の実験では、画像とは関連性のない上位語が多く辞書から抽出され、ファセット抽出の精度低下の原因となっていた同様に、クエリ間のファセット値の種類数の差より、ファセット名の種類数の差が大きい理由も、画像と関連性のない上位語が辞書から大量に得られてしまうことが 1 つの要因ではないかと考えられる。

6.3 多様性に基づくランキングの有用性評価

5 章で提案した手法によって求められたファセット集合の多様性に基づくランキングが、多様な目的の再検索に有用かどうかを評価した。

多様性を用いたランキングの有用性を評価するため、各クエリに対する上位の 3 件の画像を用いた。この 3 件の画像がそれぞれ多様な観点を持っていると評価されると、多様性に基づくランキングが多様な目的の再検索のために有用であると言える。

比較手法として、まず Flickr が提供している、クエリとの適合性に基づくランキングから得られる画像集合から選択した上位 3 件の画像を用いた。また、別の比較対象として、画像に付与されたタグの数に基づいてランキングした画像集合から選択した上位 3 件の画像を用いた。これは、タグも画像の観点を表すと考えられるので、タグの数が多い画像ほど多様な目的の再検索に使えるかどうかを評価するためである。

実験では、3 種類の各ランキング手法から得られた計 9 件の画像を重複を除いて評価者に提示し、まず、それぞれの画像がクエリに適合しているかを適合、不適合の 2 値で評価してもらった。その後、適合であると判定された画像の中から多様な観点を持っているとと思われる画像を選択してもらい、選択された画像が得られたランキング手法の得点を 1 点加算した。重複した画像が多様な観点を持っている画像として評価者に選択された場合は、その画像が得られる全てのランキング手法に 1 点を加算した。また、評価者には基本的には 3 件の画像を選択してもらったが、重複を除いた画像の総数が 6 件以下の場合は選択数を 2 件、画像の総数が 3 件以下の場合は選択数を 1 件にして 1 回の選択に対する総得点を調整した。

評価者の回答から得られた得点の平均とクエリに対する画像の適合率と nDCG の平均を、表 4 に示す。この表を見ると、平均得点が最も高いのは、提案手法である、多様度を用いたランキングであることが分かる。つまり、単なる適合性やタグ数より、本稿で提案した多様度に基づくランキングの方が、多様な目的の再検索のために有用であると考えられる。

タグ数でランキングした場合の平均得点が多様度でランキングした場合より低い理由は、タグが多い画像は、類似しているタグが多くついている場合が多いためである。タグ数によるランキングが適合性によるランキングより平均得点が高い結果となっているのは、タグの数が多いと類似しているタグも多いが、類似していないタグも多く含まれるため、結果的に多くの観点を含んだ画像が上位にランキングされやすくなるためであると考えられる。

また、表 4 から分かるように、適合率と nDCG の値が最も低いのは、タグ数を用いたランキングの場合である。タグが多く付与されている画像には画像と関係のない無駄なタグも多い場合がある。画像にタグを付けるユーザの中には、画像と関連性の低いタグをつけたり、画像を検索対象とするためわざと画像と関連性の全く無いタグをつけたりするユーザもあり、そうしたタグによって、クエリと関連しない画像が上位にランキングされてしまうと考えられる。一方、多様度を用いたランキングの場合は適合率と nDCG の値も最も高い。Flickr の適合性の場合よりもこれらの値が高いのは、Flickr の適合性によるランキングがあまり高い精度ではないことも考えられるが、多様度の計算式 (2) に重みとして、クエリとの適合性を考慮している、ファセットのスコアを用いているのが大きな理由であると考えられる。以上より、本稿で提案した多様度の用いたランキング手法では、クエリとの適合性が高く、多様な目的の再検索にも利用できる画像を上位にランキング可能ではないかと考えられる。

しかし、多様度に基づくランキングの場合の平均得点が 1.56 ということは、理想の平均得点である 3.00 より低い値である。平均得点の値が低い理由としては、まず、ファセット値をタグから抽出していることが考えられる。比較手法としてタグの数に基づいたランキング手法を用いているが、多様度を求めたファセット集合も本来はタグから得られたものなので、由来が同じである 2 つの手法を比較してもあまり大きな差は生じないのではないかと考えられる。

また、抽出したファセットが本当の意味で画像から得られたものではないということも、平均得点が低い理由ではないかと考えられる。画像の出現するオブジェクトを正確に抽出し、そのオブジェクトを表す語をファセット値として用いれば理想的であるが、システムが自動的に画像からオブジェクトを判別することや、そのオブジェクトを表す語を発見することは困難な課題である。より精度良くファセットを抽出するためには、周辺テキストやメタ情報、タグ、画像特徴量だけでなく、そうした、画像に出現するオブジェクトそのものを理解するような技術を組み合わせる必要があると考えられる。

7. ま と め

本稿では、画像検索において多様な観点からの検索を支援するために、1 つの画像からファセットを抽出する手法を提案した。提案手法では、まずクエリと画像からその画像の類似画像集合を求めてファセット値を抽出し、そのファセット値からファセット名を取得する。その後、各ファセットに対するスコア付けを行うことで、画像に関連したファセットを重要視する。また、抽出したファセットから得られるファセット集合と各ファセットのスコアを利用し、ファセット多様性の計算手法としてファセット集合の多様度を提案した。それからファセットのスコアの妥当性を評価し、ファセット値とファセット名を考慮したスコアを用いると、0.58 の適合率と 0.72 の nDCG で妥当なファセットが得られることを示した。また、多様度に基づくランキングの有用性の評価し、クエリとの適合性とファセット多様性の側面で、タグの数に基づくランキングの手法より優れた結果が得られることを示した。

本稿で提案した手法は、1 章で提案したファセット提示のイメージとは違い、ファセットに含まれる残りのファセット値を抽出していない。言語パターンを用いることで、そのような残りのファセット値を抽出できるのではないかと考えられる。今後は、手法の評価結果から考えられた課題を解決しながら、ファセットに基づく画像再検索システムの構築に取り組む予定である。

謝辞

本研究の一部は、文科省科研費基盤 (A) 「ウェブ検索の意図検出と多元的検索意図指標にもとづく検索方式の研究」(24240013 , 研究代表者：田中克己) によるものです。ここに記して謝意を表します。

文 献

- [1] Yee, K., Swearingen, K., Li, K. and Hearst, M.: Faceted metadata for image search and browsing, *Proc. of CHI*, pp. 401–408 (2003).
- [2] 川野悠, 大島裕明, 田中克己: クエリに応じたファセットの動的抽出による Web 画像検索結果の提示, *DEIM2010*, pp. A1–5 (2010).
- [3] Kawano, Y., Ohshima, H. and Tanaka, K.: On-the-Fly generation of facets as navigation signs for web objects, *Proc. of DASFAA*, pp. 382–396 (2012).
- [4] Agrawal, R., Gollapudi, S., Halverson, A. and Ieong, S.: Diversifying search results, *Proc. of WSDM*, pp. 5–14 (2009).
- [5] Sakai, T., Dou, Z., Yamamoto, T., Liu, Y., Zhang, M. and Song, R.: Overview of the NTCIR-10 INTENT-2 task, *Proc. of NTCIR-10* (2013).
- [6] van Leuken, R. H., Garcia, L., Olivares, X. and van Zwol, R.: Visual diversification of image search results, *Proc. of WWW*, pp. 341–350 (2009).
- [7] Makadia, A., Pavlovic, V. and Kumar, S.: A new baseline for image annotation, *Proc. of ECCV*, pp. 316–329 (2008).
- [8] Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. and Fei-Fei, L.: ImageNet: A Large-Scale Hierarchical Image Database, *Proc. of CVPR* (2009).
- [9] Csurka, G., Dance, C., Fan, L., Willamowski, J. and Bray, C.: Visual categorization with bags of keypoints, *Proc. of ECCV*, Vol. 1, p. 22 (2004).