

Web 画像の「らしさ」検索：語の典型的特徴を表す語集合の ソーシャルタギング情報からの取得による Web 画像検索

加藤 誠[†] 大島 裕明^{††} 小山 聡^{††} 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

あらまし 本稿では、「夏らしい画像」のように、抽象的な概念を表す語をクエリとする際の Web 画像検索の精度向上手法を提案する。現在の Web 画像検索エンジンは、具体物を表す語をクエリとした場合は比較的精度が良いが、「夏」や「平和」のような抽象的な概念を表す語をクエリとした場合には精度が悪くなってしまふ。我々は、抽象的な概念を表す語を連想させる、より具体的な物を表す語集合を取得し、それらをクエリにおいて利用することで、Web 画像検索の改善を行う。そのような語集合を取得するために、まず、抽象的な概念を表す語がタグとして付与された画像のソーシャルタギング情報を収集する。同一画像にタグとして付与された語の共起関係を利用して語のクラスタリングを行うことにより、抽象的な概念を表す語を連想させる語集合を取得する。提案手法のプロトタイプシステムとして、ソーシャルタギング情報に Flickr の情報を利用し、Web 画像検索の改善を行うシステムを作成する。

キーワード 情報検索, 画像検索, Web マイニング, フォークソノミー

“Likely” Image Search: Web Image Search using Term Sets Representing Typical Features Extracted from Social Tagging Information

Makoto KATO[†], Hiroaki OHSHIMA^{††}, Satoshi OYAMA^{††}, and Katsumi TANAKA^{††}

[†] Department of Informatics and Mathematical Science, Faculty of Engineering,
Kyoto University, Yoshida-honmachi, Sakyo, Kyoto 606-8501, Japan

^{††} Department of Social Informatics, Graduate School of Informatics, Kyoto University,
Yoshida-honmachi, Sakyo, Kyoto 606-8501, Japan

Abstract We propose a method for improving Web image search when an abstract term is submitted as a query. Current Web image search engines return enough good results for concrete term queries. However, search results for abstract term queries such as “summer” or “peace” are not good. To improve such a Web image search for an abstract term query, we obtain more concrete term sets which are associated with the abstract concept and use them as queries for Web image searches. Extracting concrete term sets for an abstract term consists of three steps: (1) collecting social tagging information of images which have the abstract term as one of the tags, (2) clustering the tags using co-occurrence relations on images, and (3) getting concrete term sets from clusters. We make a prototype system to improve Web image search using social tagging information in Flickr.

Key words Information retrieval, Image search, Web mining, Folksonomy

1. はじめに

近年、インターネットと Web 検索エンジンの普及により、文書や画像、動画等の様々なメディアをキーワードによって容易に検索することが可能となった。現在の検索エンジンの多くは、クエリと文書内の文章、クエリと画像の周辺テキスト、クエリと動画のタグ、といった 2 つのテキスト間のキーワードマッチングによって、それぞれ、文書、画像、動画を検索している。

しかし、このような方法で検索することが妥当ではないクエリも存在する。その例として、「夏」という語をクエリとした画像検索が挙げられる。現在の画像検索ではこのような抽象的なクエリによる検索は十分な結果を得ることができない。何故なら、いかに「夏らしい」画像であっても、その周囲に必ず「夏」という語が現れるとは限らないためである。また同様に、「有名な」京都のページや「心踊る」動画などはうまく検索することができない。

画像検索を行っても良い結果を得られない抽象的な語には、抽象的な概念を表す語句「夏」「平和」「京都」などや、感性的な語句「楽しい」「悲しい」「切ない」といったものがある。これらのクエリは文書検索では用いられることは多くないだろうが、画像や動画、書籍などを検索する場合には使用できることが望ましい。これらのような、内容が人間の感性に依存するメディアを検索する際には、人間の直感的なキーワードの意味を汲み取って検索すべきである。

そこで本稿では、これらのような語による Web 画像検索を改善する手段として、「らしさ」Web 画像検索を提案する。「らしさ」とは、「日本らしさ」「夏らしさ」「かわいらしさ」といったような場合に使われる語である。本システムでは、これらを表現する画像、すなわち「日本らしい」「夏らしい」「かわいらしい」画像を検索することを可能とする。

「らしさ」Web 画像検索が抽象的な語による画像検索を改善させるのは、抽象的なクエリによる画像検索で求められる画像は、その画像がクエリの特徴を含み、クエリを連想させる画像、すなわち「らしい」画像であるためである。「夏」というクエリによる画像検索の適合画像には、「夏らしい」画像、例えば「ひまわり」や「風鈴」が写った画像が期待される。

2章では、画像検索行為の分類について述べる。3章では、「らしさ」Web 画像検索システムの概要について述べる。4,5章では、「らしさ」Web 画像検索を実現するための手法について述べる。4章では、Flickr^(注1)から「らしさ」を取得する方法について、5章では、4章で得られた「らしさ」を使って、既存の画像検索エンジンにて画像を検索する手法について述べる。6章では、「らしさ」Web 画像検索の性能について、既存の Web 画像検索と比較を行う。7章では、本研究の関連研究を、8章では、本研究のまとめと今後の課題について述べる。

2. 画像検索行為の分類

現在の Web 画像検索の問題を議論する前に、入力するクエリとユーザの求める画像との関係に着目した画像検索行為の分類を行う。以下、ユーザが入力するクエリ、単語 1 語を Q とする。また、ユーザが適合画像と判断する画像集合を I し、 I の要素 i_k に写っているオブジェクト集合を O_{i_k} とする。

クエリと適合画像の関係に基づいて、画像検索を以下の 3 つに分類する。

2.1 クエリが画像に写りうる具体物である場合

クエリが画像に写りうる具体物を指す語である場合、クエリ Q と適合画像 i_k に写るオブジェクト集合 O_{i_k} のある要素との関係は「instance-of」であるか「itself」、もしくは「part-of」である。「instance-of」であるのは、例えば、 Q が「りんご」で O_{i_k} の要素に「りんご」を含むような、クラス名をクエリとする場合であり、「itself」であるのは、例えば、 Q が「田中克己」で O_{i_k} の要素に「田中克己」を含むような、インスタンス名をクエリとする場合である。最後の「part-of」については「instance-of」と「itself」とは異なった特徴を持つため後述す

表 1 クエリが具体物である画像検索の適合率

画像検索クエリ	上位 20 件の適合率	上位 100 件の適合率
moon	1.00	0.82
bird	0.95	0.68
rose	1.00	0.55
butterfly	0.80	0.70
bear	0.95	0.75
cat	1.00	0.86
dog	1.00	0.76
pigeon	1.00	0.77
ipod	1.00	0.79
headphone	1.00	0.77

る。この画像検索行為の特徴として、 Q が何であるかを視覚的に知りたい場合に用いられることが多い。また、このような画像検索を行った場合には比較的良い適合率が得られる。表 1 にそのクエリの例と画像検索における適合率を示す。なおの表 1,2,3 の画像検索は、Google^(注2)の画像検索エンジンを用いている。適合率の評価にはそれぞれの検索行為を行っているユーザを想定し、適合しているかどうかを我々が判断した。

2.2 クエリが具体物であるが画像に写すのが困難である場合

クエリが画像に写りうる具体物を指す語であるが、画像に写すことが困難である場合、クエリ Q と適合画像 i_k に写るオブジェクト集合 O_{i_k} のある要素との関係は「part-of」である。画像に写すことが困難であるとは、その具体物の全体が画像に現れることが稀であり、その具体物の一部が画像に含まれることで適合画像となる場合である。「part-of」であるのは、例えば Q が「京都」で O_{i_k} の要素に「金閣寺」「嵐山」などが含まれるような場合である。このような画像検索は主に Q が非常に大きいものである場合に用いられる。その例として、地域名や海、空などがある。「part-of」の関係は、適合画像 i_k に写るオブジェクト集合 O_{i_k} が Q のインスタンスの一部であるため、「instance-of」「itself」の関係に近いが、あるオブジェクトの一部がそのオブジェクト全体を表す、連想させるという考えからすると、後述する「associated-with」の関係に近いものである。この検索行為の適合率は、クエリが具体物である場合と比較してやや低い値となっている。表 2 にそのクエリの例と画像検索における適合率を示す。

2.3 クエリが抽象的である場合

クエリが画像に写り得ない抽象的な語である場合、クエリ Q と適合画像 i_k に写るオブジェクト集合 O_{i_k} のある要素との関係は「associated-with」である。「associated-with」であるのは、 Q が「夏」で O_{i_k} の要素に「風鈴」「ひまわり」などを含む場合である。適合画像 i_k に写るオブジェクト集合 O_{i_k} の要素は Q のインスタンスでも一部分でもなく、単にオブジェクト集合 O_{i_k} の要素が Q を連想させるという関係になっている。抽象的なクエリでの画像検索の適合率は、クエリが具対物である場合と比較すると低い値となっている。表 3 にそのクエリの例と画

(注1): Flickr, <http://www.flickr.com/>

(注2): Google, <http://www.google.co.jp/>

表 2 クエリが具体物であるが画像に写すのが困難である画像検索の適合率

画像検索クエリ	上位 20 件の適合率	上位 100 件の適合率
sea	0.95	0.45
sky	0.90	0.47
forest	0.90	0.52
lake	0.95	0.67
tokyo	0.65	0.47
kyoto	0.75	0.47
paris	0.70	0.41
kyoto-university	0.45	0.22

表 3 クエリが抽象的である画像検索の適合率

画像検索クエリ	上位 20 件の適合率	上位 100 件の適合率
summer	0.40	0.25
winter	0.90	0.54
justice	0.45	0.25
love	0.25	0.17
sad	0.95	0.37
powerful	0.35	0.10
america	0.25	0.10
japan	0.45	0.23

像検索における適合率を示す。

具体的なクエリとも抽象的なクエリとも成り得る語として表 3 の japan がある。地域名として japan を捉えた場合、クエリは具体的なものであるが、概念として捉えた場合、クエリは抽象的なものとなる。これらの違いは検索するユーザがどの画像を適合画像と判断するかによって生じ、ある語が与えられたときにその語が具体的か抽象的であるかを区別することはできない。

2.4 現在の Web 画像検索の問題点

Web 画像検索では、抽象的なクエリで検索する場合、再現率、適合率は低い値となってしまう。これは以下の 2 つの点が原因となっている。

- 抽象的なクエリは、Web 文書中の適合画像の周辺テキストにあまり出現していない
- 抽象的なクエリは、適合画像以外の画像の周辺テキストにも多く存在する

1 つ目は、例えば「夏」というクエリで検索した際に適合画像であるはずの、「海 砂浜」「ひまわり 青空」「花火 浴衣」といったオブジェクトが含まれる写真の周辺テキストには、「夏」という語はあまり出現していないことを意味している。このことは「夏」というクエリにおける再現率を下げる一因になっていると考えられる。

表 4 は、ある画像検索クエリで検索した際に、その検索結果として列挙される画像の周辺テキストに、あるキーワードを含む割合を示している。今回、周辺テキストを画像周辺の 100 語とし、検索された 500 件の画像の掲載ページの中から、検索結果として表示された画像を含んでいると判定された有効なページを選択し、その周辺テキストにキーワードを含んでいるかを

表 4 ある抽象的な語が、適合すると想定された画像の掲載ページに含まれる割合

画像検索クエリ	キーワード	キーワードを含むページの割合
sunflower	summer	0.02
sea beach	summer	0.01
pool children	summer	0.04
snow mountain	winter	0.06
white pigeon	peace	0.03
kitty	cute	0.06
geisha	japan	0.12
sky bird	freedom	0.01
america flag	freedom	0.01

Yahoo!^[注3]の画像検索 API を用いて調査した。

表 4 では、キーワードとして抽象的な語を、画像検索クエリとしてそのキーワードで画像検索を行った際に想定される適合画像が含んでいるオブジェクト名を用いた。この結果が指していることの例として、平和の象徴である「white pigeon」の画像は、「peace」で検索した際にはわずか 0.03 の割合でしか含まれていないと言ええる。以上の点から、抽象的なクエリは、その結果が含むべき適合画像の多くを取りこぼしているということになる。

2 つ目は、抽象的な語が画像の周辺で用いられたとしても、画像内容に関する言及であるとは限らないことを意味している。例えば「夏」や「平和」といった語句はその意味が広く、具体的な語に比べ、Web 文書中に遍在している語だと考えられる。そのため、画像の周辺にこれらの抽象的な語が出現したために、適合していない画像も検索結果に含まれてしまう。一方、具体物を表す語は、その語が画像の周囲に存在していれば、画像中にその語が示すオブジェクトがある可能性も高い。このことは、表 1, 3 の適合率を比較すれば明らかである。

3. 概要

3.1 システムの概要

ここで提案する「らしさ」Web 画像検索の入力と出力は以下の通りである。

- 入力：抽象的な語 X
- 出力：「X らしい」画像集合

動作としては従来の Web 画像検索システムと同様の振る舞いをする。ただし、入力されるクエリとしては抽象的な語を想定している。出力は、入力された語 X に対して、システム内で得られた「X らしさ」に分類されて表示される。図 1 に X=「夏」によるシステムの動作画面の例を示す。ユーザがクエリ「夏」による検索を行った場合、システムは内部で「夏らしさ」を表す「風鈴 縁側」「祭り 花火 浴衣」「海 浜辺」といった語集合を取得し、それぞれの語集合ごとに画像が出力される。

3.2 システムの内部動作の概要

本システムを実現するためには以下の 2 点が必要となる。

- 「らしさ」の抽出

(注3): Yahoo!, <http://www.yahoo.com/>

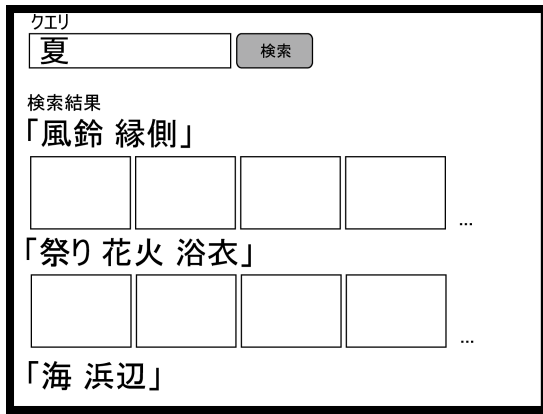


図 1 システムの動作画面

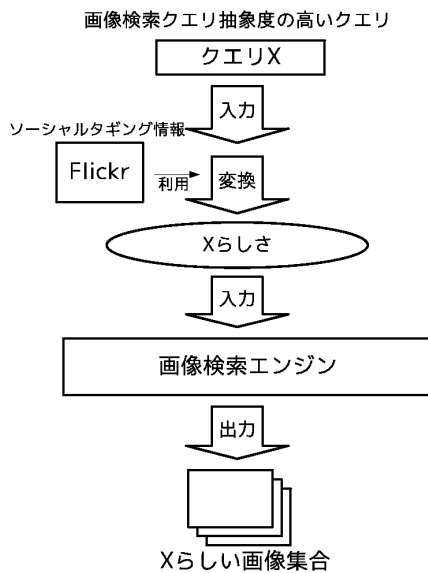


図 2 システムの内部動作図

• 「らしさ」を表す画像の検索

まず、システムに入力として語 X が与えられる。この X に対して、Flickr のソーシャルタギング情報より、「Xらしさ」を抽出する。次に、「Xらしさ」を表す画像を既存の Web 画像検索を通して検索し、その結果を選択して画面に表示する。図 2 にシステムの内部動作を図示する。

3.3 Flickr について

知識の取得元として利用する Flickr は、ユーザが画像を投稿し、その画像に対して投稿者や閲覧者がタグとしてキーワードを付加できるサービスである。知識抽出のデータ元として Flickr を採用した理由は二つある。1 つ目は、本稿では画像検索を扱うため、画像に即した知識が必要であるからである。2 つ目は、利用者が写真に対して自由にタグづけを行える点にある。タグをつける理由として、投稿者が写真の情報を閲覧者に知らせるためと、画像を検索する側に対しての配慮が考えられる。Flickr のタグと写真の関係は、タグは必ずその写真の情報であること、タグが写真中に現れるオブジェクトを Web 文書中画像よりもよく指していること、の 2 点が Web 文書中の画像とその周辺テキストとの関係とは異なり特徴的な点である。以

上の点から、知識抽出の取得元として Flickr を採用している。

4. 特徴語集合の抽出

4.1 「らしさ」の定義

日本語における「らしさ」という言葉は場合によって様々な意味を持つが、今回取り扱う「らしさ」とは、形容詞性接尾辞の「らしい」が名詞化したものである。例えば、「夏らしさ」「京都らしさ」のように、名詞を伴って用いられる場合や、「可愛らしさ」のように形容詞を伴う例などがある。「らしい」の言語学的な意味は以下の通りである。

「そのものが十分にその特徴や性格を備えている」という話し手の感じ方 [1] [2]

本稿では、この言語学定義を前半部の「十分にその特徴や性格を備えている」という部分と、後半部の「話し手の感じ方」という部分の 2 つの側面に分けることにことにより、「らしさ」を捉える。

- ある X が与えられた時に Y が「Xらしさ」であるとは、
- (1) Y は X の特徴を持っている
 - (2) Y は X を連想させる
- という 2 つの条件を満たすことである。

(1) では、「Xらしさ」は X の特徴を持っていることを意味している。例えば、「夏らしさ」であれば、それが「夏」の特徴を持っていないとてはならない。(2) では、「Xらしさ」には人に対して X を感じさせなくてはならないことを意味している。例えば、「夏らしさ」であれば、それが「夏」を連想させなくてはならない。

また、これら 2 つの条件が共に成立しなくては「らしさ」であるとは言えない。例として、「夏」に対して「春夏秋冬」という語集合は 2 つ目の条件を満たすが、1 つ目の条件は満たさない。よって、「春夏秋冬」は「夏らしさ」であるとは言えない。対極の例として、「夏」の特徴として「水泳」という語が得られても、そのことが「夏」を連想させなくては「夏らしさ」であるとは言えない。この例の場合、「水泳」は「夏」以外のことも多く連想させてしまう。例えば、「オリンピック」や「海」「プール」や「水着」なども連想させる。連想させるもの全体の内「夏」が占める割合を考えると、「水泳」が「夏」を連想させる度合いは低いと考えられる。つまり、「水泳」という語は「夏らしさ」としては妥当ではない。

「らしい」のより正確な言語学的意味は上記に述べたこと以上のものを含むため、ある特定の名詞、形容詞の後に続けて「らしい」を用いることが適切でない場合が多々あり、また、ニュアンス的にもずれる場合がある。本稿では「らしさ」のこれらの言語学的な問題には深く立ち入らず、以降「らしさ」とは、上述の 2 点を満たすものだけを指す。

4.2 特徴語集合の定義

前節では、「らしさ」を性質の面から定義したが、その実体についての定義は与えていなかった。ここで、「らしさ」を表すものである「特徴語集合」を以下のように定義する。

ある語 X の特徴語集合 L とは、「 X らしさ」の条件を満たす任意の語集合である。

例として、「夏」の特徴語集合として、「海 浜辺」「風鈴 縁側」「花火 浴衣 祭り」などが挙げられる。これらは、「夏らしさ」の条件、すなわち、夏の特徴を持っており、夏を連想させるという2点を満たしている。

4.3 ソーシャルタギング情報からの特徴語集合の抽出

特徴語集合を取得に Flickr のタグデータを利用する。Flickr のタグデータを用いることについての理由は前述した通りである。Flickr から得られるデータは、写真と複数のタグの組み合わせとなっており、それらは一対多の関係になっている。特徴語集合の抽出は、与えられた X の特徴を持った語集合の取得をした上で、その語集合の中から X を、ある一定の水準以上連想させるものだけを取り出すことによって行う。

4.3.1 特徴の抽出

第1段階として、与えられた X の特徴を持つ特徴クラスタ集合 C_X の取得を行う。特徴を得るためには、 X というタグがつけられている写真集合に現れるいずれかのタグを選択するという方法が考えられる。しかし、画像における特徴が1語のみでしか与えられない場合、十分にその特徴を表すことができず、画像が特定できない場合が多々ある。例えば、「海」1語においても、「海 やしの木 太陽」や「海 暗雲 漁船」「海 日の出」といった組み合わせのように、様々な種類の画像が考えられるため、1語だけでは X らしい画像を特定できない可能性が考えられる。そのため、以下のように X というタグがつけられた写真集合をクラスタリングすることによって、その特徴をクラスタによって表現する。

P : 全写真集合

T : 全タグ集合

T_{p_i} : $p_i \in P$ に付加されているタグ集合

X をタグとして含む P の部分集合 $P_X = \{p_i | X \in T_{p_i}\}$ に含まれる全ての $p_i (i = 1, 2, 3, \dots)$ に対して、

$$V_{p_i} = (tf_{p_i, t_1} \cdot idf_{t_1}, tf_{p_i, t_2} \cdot idf_{t_2}, \dots, tf_{p_i, t_n} \cdot idf_{t_n})$$

を求める。ただし、

t_j : T の要素

df_{t_j} : タグ t_j を含む写真数

N : 全写真数 (= 2,000,000,000)

$$idf_{t_j} = \log \frac{N}{df_{t_j}}$$

$$tf_{p_i, t_j} = \begin{cases} 1 & (p_i \text{ が } t_j \text{ をタグとして含む場合}) \\ 0 & (otherwise) \end{cases}$$

(1)

である。

以上のように、各 p_i に特徴ベクトルを定義した後、 P_X に含まれる全ての p_i に対して階層的クラスタリングを行う。

表5 $X = \text{"summer"}$ で得られた重心ベクトル

語	値	語	値	語	値
minimal	10.97	mother	9.90	surf	9.05
mar	9.24	children	8.27	ocean	7.73
sea	7.42	fun	7.19	water	6.92
blue	7.24	sky	7.05	spray	10.73
water	6.93	beach	6.50	California	6.56
beach	6.50	vacation	6.41	sea	7.41

語	値	語	値	語	値
food	7.17	nature	6.60	grass	8.73
chubby	11.82	vacation	6.41	green	7.34
estate	9.94	blue	7.24	Canada	7.02
woman	8.24	travel	6.20	warm	10.57
beach	6.50	moraine	13.27		
people	6.99				

2つの写真 p_i, p_j の類似度 $Sim_P(p_i, p_j)$ を以下のように定義する。

$$Sim_P(p_i, p_j) = \frac{V_{p_i} \cdot V_{p_j}}{|V_{p_i}| |V_{p_j}|}$$

また2つのクラスタ c_n, c_m 間の類似度 $Sim_C(c_m, c_n)$ を以下のように定義する。

$$Sim_C(c_m, c_n) = \min_{x \in c_m, y \in c_n} Sim(x, y)$$

階層的クラスタリングの手順は以下の通りである。

- (1) 写真集合 P_X の要素 p_i を、1つの要素しか持たないクラスタ c_i の要素とする
- (2) クラスタの数が1であれば終了する
- (3) 全てのクラスタの中で最も類似度の高い2つのクラスタを取り出すすなわち、 $Sim_C(c_i, c_j)$ が最大となる c_i, c_j の組を見つける
- (4) c_i, c_j を結合し1つのクラスタとし(2)に戻る

階層的クラスタリングを行った後、根から探索を行い、予め設定した閾値を下回る類似度で結合された節の1つ下の深さにある全ての節を X の特徴クラスタ集合 C_X として採用する。

表5に $X = \text{"summer"}$ で得られたクラスタの重心ベクトルの例を示す。

4.3.2 アソシエーションルールを用いた特徴クラスタの評価

階層的クラスタリングにより得られたものは、 X の特徴ではあるが、 X を連想させるものであるとは限らない。したがって、特徴クラスタ集合を X を連想させるかという基準でそれぞれ評価し、設定した閾値を越えたものだけを採用するという方法が考えられる。今回、評価値として、 X の特徴クラスタが X を連想させるというアソシエーションルールを適用した際の「確信度」と「サポート」という2つの値を採用する。

評価するにあたり、各特徴クラスタ c_i の重心ベクトル m_i を算出する。 m_i の中から値の大きい順に、その次元に対応するタグを k 個取り出し、その冪集合を特徴語集合候補 T_i とする。 T_i の全ての要素 a に対して、

$$a \Rightarrow X$$

表 6 X="summer" の特徴語集合の例

特徴語集合	サポート	確信度
estate woman	91.39	0.54
vacation nature travel blue	1035.84	0.48
sky beach fun	2316.08	0.27
grass warm	1093.79	0.23
smile boy pool	73.11	0.22
August portrait	1488.63	0.2
light landscape	16483.46	0.16
sun holidays	4354.11	0.14
water sea blue	16388.21	0.11
beach girl	25058.74	0.11

という関係がアソシエーションルールによる評価対象となる．アソシエーションルールにおける確信度 $Conf(a, X)$, サポート $Sup(a)$ の定義は以下の通りである．

$$Conf(a, X) = \frac{df(a \cup X)}{df(a)}$$

$$Sup(a) = df(a)$$

ただし, $df(a)$ は a の全てのタグ集合を含む写真数である．

ただし, この手法の場合, Flickr 特有の問題が発生する． $Sup(a)$ の値が小さいとき, すなわち a の要素を含む写真の枚数が少ない際に, 少数のユーザが大量の写真を同一のタグで投稿したために $Sup(a), Conf(a, X)$ の値とも不当に高い値を獲得してしまうことがある．そこで, ユーザ 1 人の重みを等しくするために df の代わりに $uf(User Frequency)$ を用いる．すなわち,

$$Conf(a, X) = \frac{uf(a \cup X)}{uf(a)}$$

$$Sup(a) = uf(a)$$

ただし, $uf(a)$ は a の全てのタグ集合を含む写真を投稿したユニークな投稿者数となる．

以上のようにして, 各特徴語集合候補 T_i の要素を評価し, そのサポートが min_sup 以上, また確信度が min_conf 以上のものが c_i の特徴語集合として採用される．また, 各特徴クラスタ c_i の評価値はそのクラスタから得られた特徴語集合の値とする．特徴語集合候補 T_i の中から最も良く X を連想させると判断された語を特徴語集合として用いるのは, X の特徴として得られた語集合から冗長性を排除し, 特徴語集合を最適な大きさにするためである．

表 5 は $X=summer$ とし, P_X の要素数を 200, $min_sup = 50, min_conf = 0.1$ としたときに得られる特徴語集合である．

ただし, サポートと確信度を計算する際の uf の値は, 実装上の問題から最初に得られた画像 500 件中の, ユニークなユーザ数を用いて全体におけるユニークな投稿者数を近似的に求めた値を用いている．

5. 特徴語集合を用いた画像検索の改善

得られた特徴語集合及び特徴クラスタを用いた Web 画像検索の改善方法について述べる．単純に得られた特徴語集合を

表 7 複数語クエリの上位 20 件の適合率と上位 100 件の適合率

複数語クエリ	上位 20 件の適合率	上位 100 件の適合率
shrine kimono	0.60	0.33
sea sun	0.55	0.18
orange apple	0.50	0.21
girl beach	0.55	0.35
sun beach sky	0.30	0.58

AND で繋ぎ, クエリとただけでは十分な結果を得ることができない．これは, 現在の Web 画像検索エンジンでは複数語のクエリを用いた場合, 十分な結果を得ることができないためである．そこでまず, Web 画像の周辺テキストと Flickr のタグの相違について議論した上で, Flickr で得た知識を元に Web 画像検索を行う手法について述べる．

5.1 Web 画像検索における複数語クエリ

簡易的な実験による複数語クエリとその適合率の関係は表 7 の通りである．なお画像検索には Yahoo! の画像検索 API を用いている．表 1, 2, 3 と比較すると分かるように, その適合率は総じて低い値となっている．このことの原因については, また後ほど述べる．

5.2 Web 画像の周辺テキストの特徴

Web 文書中出现する画像の周辺テキストの特徴の 1 つとして, 「画像内のオブジェクトを指す語句は, Flickr 画像のタグと比較した場合, あまり周辺テキスト中出现しない」ということが挙げられる．Web 文書中の画像の周辺テキストは必ずしも画像に言及しているわけではなく, また, 画像中のオブジェクトについて 1 つずつ解説していることは稀である．一方, Flickr のタグのほとんどが画像に関するものであり, 画像のオブジェクトに対する語が多く現れる．これらの Flickr と Web 画像の差異が, 複数語クエリによる画像検索の適合率が低下する原因の 1 つになっていると考えられる．

5.3 特徴クラスタを用いた Web 画像検索

複数語のクエリは現在の Web 画像検索の適合率を低下させてしまう．そのため, 得られた特徴語集合を直接クエリとして用いることは適当ではない．そこで, 本稿では, 特徴語集合のみだけでなく, 特徴語集合を求めた際に得られた特徴クラスタを用い, 各語に重みをつけた検索を行うことによって画像検索を行う．具体的には, クラスタに対応した特徴ベクトルを算出し, それと画像の周辺に現れる語によって作られた特徴ベクトルとの類似度を元に画像集合をソートすることによって, 上位により適合する画像を配置することにより適合率の向上を図る．

特徴クラスタを用いた Web 画像検索は以下の通りである．

- (1) 特徴語集合を AND でつないだクエリを用いて, 既存の画像検索エンジンで画像を検索する
- (2) 画像の周辺テキストを用いて特徴ベクトルを作成する
- (3) 特徴語集合に対応する特徴クラスタの重心ベクトルと特徴ベクトルの類似度を計算する
- (4) 画像をベクトル間の類似度に基づいてソートする

特徴語集合をクエリとして画像検索を行い, 得られた画像集合 I の各要素 i_k に対して以下のように特徴ベクトルを定義する．

表 8 既存の画像検索エンジンと本システムによって得られた画像上位 20 件の適合率

抽象的なクエリ	既存	本システム	本システム (ソートあり)
summer	0.25	0.60	0.60
winter	0.85	0.65	0.70
justice	0.65	0.15	0.10
love	0.35	0.45	0.35
sad	0.60	0.70	0.40
powerful	0.10	0.50	0.50
america	0.60	0.30	0.45
japan	0.70	0.50	0.55

表 9 「summer」の特徴語集合

特徴語集合	サポート	確信度
estate, woman	91.63	0.54
fun, pool	8682.68	0.24
sky, fun	11631.20	0.15
ocean, portrait	4046.87	0.10
girl, beach	25080	0.11

$$V_{i_k} = (tf_{i_k,1} \times idf_1, tf_{i_k,2} \times idf_2, \dots, tf_{i_k,n} \times idf_n)$$

ただし,

$tf_{i_k,l}$: 画像 i_k の周辺テキストに現れる語 $term_l$ の数

$$idf_l = \log \frac{N}{df_l}$$

df_l : 周辺テキストに語 $term_l$ を含む画像の総数

N : I の要素数

検索に用いた特徴語集合の取得元の特徴クラスタ c_i と画像 i_k の類似度には, c_i の重心ベクトル M_{c_i} と i_k の特徴ベクトル V_{i_k} とのコサイン類似度を用いる.

6. 実験と考察

抽象的な語に対して, 既存の画像検索エンジンを用いた場合と, 本稿で提案するシステムを用いた場合の結果を比較することにより, 本システムの有為性について検証する. 抽象的な語として, 表 3 で用いられた 8 語を用いる.

X を各抽象的な語とし, P_X の要素数を 250, $min_sup = 50$, $min_conf = 0.1$ としたときに得られる特徴語集合上位 5 件を用いて Yahoo! の画像検索 API により画像を取得する. 各画像の周辺テキスト 50 語を用いて特徴ベクトルを作成し, 特徴クラスタの重心ベクトルとのコサイン類似度によって値の大きい順にソートしたものの上位 4 件, すなわち 5 つの特徴語集合に対し 4 件の画像, 計 20 枚の画像を本システムで得られる画像とする. またこれに加え, ソートせずに各特徴語集合による検索結果の上位 4 件ずつ取得した 20 件も比較の対象とする. 一方, 既存の画像検索エンジンの結果として同様に Yahoo! の画像検索 API を用い, 上位 20 件の画像を取得する.

表 8 に適合率を比較したものを示す. また, 表 9, 10, 11, 12, 13 に本システムで用いられた特徴語集合を例として示す.

summer, powerful には本システムよっての改善が見られるものの, winter, justice, america, japan においては適合率が損なわれてしまっている. 適合率を低下させてしまう原因とし

表 10 「winter」の特徴語集合

特徴語集合	サポート	確信度
ice, cold	35516.43	0.54
cold, path	462.62	0.51
ice, icy	2868.12	0.50
frost	44023.67	0.43
cold	167924.09	0.42

表 11 「justice」の特徴語集合

特徴語集合	サポート	確信度
law, court	189.43	0.18
protest	123478.52	0.004
freedom	48816.08	0.003
divination	69.18	0.002
punishment	1519.06	0.001

表 12 「love」の特徴語集合

特徴語集合	サポート	確信度
bride, emotion	60.83	0.40
kiss	79147.37	0.23
heart	78342.43	0.21
sad, hate	109.76	0.20
romantic	20152.77	0.19

表 13 「japan」の特徴語集合

特徴語集合	サポート	確信度
tokyo, asia	4075.2	0.94
tokyo	362781.37	0.63
shrine, girl	62.83	0.55
pagoda	17847.02	0.14
maple, autumn	5863.36	0.10

て, 確信度が低く特徴語集合として妥当でない語が採用されてしまっているという点が挙げられる. これは表 11 にあるような justice を入力とした例に現れている. また, 表 10 にある winter の例のように, 確信度が高く winter の特徴語集合として妥当だと思われる語であるが, 良い適合率が得られない例もある. このことについては, 本システムで得られる特徴語集合全体に共通することであるが, 語集合が抽象的な語で構成される場合, 十分な語の具体化が行われず, システムの最後の段階において抽象的な語をクエリとした画像検索を行ってしまっているため, 良い結果が得られていないと考えられる. 再現率においては, 複数のクエリに拡張し画像を検索する本手法の方が優れていることは明らかである.

得られた特徴語集合のみに着目した場合, 概ね良好な結果が得られている. しかし, 表 13 にある japan の例における「tokyo asia」ように, 単純に共起度が高いために確信度が高くなってしまっている語がいくつか含まれている. この場合, 「tokyo asia」が japan を連想させるとは言い難い. そのため, アソシエーションルールだけでは語と語の連想関係を評価しきれない部分があると考えられる.

特徴語集合だけでなく特徴クラスタを用いた画像検索においても, その結果はクエリによって異なっている. 精度が向上しない原因として, 画像の周辺テキストが多く得られない場合に十分な特徴ベクトルが得られず, 適切なソートが行われないこ

とが挙げられる。また今回、特徴クラスタの重心ベクトルと画像の周辺テキストから作られる特徴ベクトルのコサイン類似度を用いたが、重心ベクトルの値は写真集合のクラスタリングのためのものであるため、画像検索において用いるのは妥当でないと考えられる。

7. 関連研究

ある語を、その語を連想させる語集合で表現するという研究は、稲川ら [3] によって行われている。本研究では、ある概念を表現する画像を特定するために、語集合に変換しており、抽象的な概念を複数の語集合で表現することによって具体化を図るという側面もある。また、画像検索のクエリとして抽象的な語を用いる研究も行われている。その中でも印象語や感性に着目した研究がある [4] [5]。本稿では、これらの印象語なども対象として含んでおり、より直感的な画像の検索を実現する、改善するという点においては類似するところが大きい。複数語クエリによる画像検索の再現率の向上を、クエリ緩和によって行う研究は桑原ら [6] によって行われている。再現率の向上は本研究の目的の一つでもあり、画像検索と Web 検索の両方の結果を用いたクエリ緩和手法は、本研究の複数語クエリによる画像検索をする段階で有効に働くのではないかと考えられる。Flickr に限らず、フォークソノミーの情報を利用した研究も盛んに行われている [7] [8]。このような情報の特性として、Web 上の文書とは異なり、ページの製作者に限らず、利用する各々のユーザが情報を Web 上に反映でき、また、利用するユーザが検索しやすいようにそれぞれのユーザがタグづけなどを行うため、知識抽出の情報源としては有用である。その他にも、画像検索におけるインターフェースにおける研究も行われている。今回は画像を特徴語集合ごとに分類して提示するシステムであるため、画像のクラスタリングを行って提示する手法と類似している [9]。Liu ら [10] による画像検索のインターフェースの比較実験によれば、Liu らの提案手法に加えて、クラスタごとの画像提示は、既存の画像検索エンジンのようにリスト形式で画像を提示するよりも閲覧性に優れている。

8. まとめ

本稿では、抽象的な語による Web 画像検索の再現率、適合率を改善する、Web 画像の「らしさ」検索を実現する手法について述べた。「らしさ」の言語的な意味合いから、 X の「らしさ」を表現する X の特徴語集合を、 X の特徴かつ X を連想させる語集合であると定義した。システムのプロトタイプとして、Flickr の画像をタグの共起関係に基づいてクラスタリングし、各クラスタ中の数語の幕集合に対してアソシエーションルールを適用することで評価を行い、サポートが一定値以上のものの中で確信度が最大の語集合を特徴語集合とした。また、特徴語集合に加え、特徴クラスタを用いることにより、特徴語集合による複数語の画像検索の適合率向上を図った。今回「らしさ」検索の結果例として数件のクエリに対して本研究手法を適用したが、従来の画像検索エンジンに対する大きな優位性は見られなかった。しかし、「らしさ」の抽出のみに着目した場合、本研

究における「らしさ」の条件を満たすものの抽出には概ね成功しているため、画像検索の段階においての改善の余地があると考えられる。

今後の課題としては、前述の通り画像検索段階における改善が考えられる。また、結果例として挙げた各種画像検索の精度については、サンプル数を十分にした上で問題を明確にする予定である。同様に、本システムの評価実験についても妥当なクエリ数で結果を検証する必要がある。

本研究では、主に抽象的な語による画像検索の精度向上に主眼を置いたが、1 章でも述べた通り、画像検索以外、特に動画や書籍においても抽象的な語の意図を汲み取って検索できることが望ましい。今回は画像投稿サイトである Flickr を用いたが、他のソーシャルタグ情報を利用すれば画像以外への「らしさ」検索の適応が可能であると考えられる。

謝辞

本研究の一部は、京都大学グローバル COE プログラム「知識循環社会のための情報学教育研究拠点」、および、文部科学省科学研究費補助金特定領域研究「情報爆発時代に向けた新しい IT 基盤技術の研究」、計画研究「情報爆発時代に対応するコンテンツ融合と操作環境融合に関する研究」(研究代表者: 田中克己, A01-00-02, 課題番号 18049041), ならびに、計画研究「情報爆発に対応する新 IT 基盤研究支援プラットフォームの構築」(研究代表者: 安達淳, Y00-01, 課題番号: 18049073), および、文部科学省研究委託事業「知的資産の電子的な保存・活用を支援するソフトウェア技術基盤の構築」、異メディア・アーカイブの横断的検索・統合ソフトウェア開発(研究者代表: 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] 山下喜代: “形容詞性接尾辞「ばい・らしい・くさい」について”, 講座日本語教育.
- [2] 小島聡子: “接尾語「ばい」の変化”, 明海日本語, 8, (2003).
- [3] 稲川雅之, 大島裕明, 小山聡, 田中克己: “オブジェクトを特定する語集合の Web からの抽出”, データベースと Web 情報システムに関するシンポジウム (DBWeb2007).
- [4] 板橋美子, 佐々木史織, 吉田尚史, 清木康: “時代・文化に応じた色・印象知識による文化財画像メタデータ自動抽出システムの実現方式”, 日本データベース学会 Letters, 5, 4, pp. 21–24.
- [5] 北川高嗣, 中西崇文, 清木康: “静止画像メディアデータを対象としたメタデータ自動抽出方式の実現とその意味的画像検索への適用”, 情報処理学会論文誌, 43, .
- [6] 桑原昭裕, 角谷和俊, 田中克己: “質問緩和法によるクロスメディア・メタサーチ”, DBSJ Letters, 3, 1.
- [7] P. Schmitz: “Inducing ontology from flickr tags”, Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland, May (2006).
- [8] 山家雄介, 中村聡史: “ソーシャルブックマークの特性を利用した Web 検索のランキング精度の向上”, DBSJ Letters, 6, 1.
- [9] D. Cai, X. He, Z. Li, W. Ma and J. Wen: “Hierarchical clustering of WWW image search results using visual, textual and link information”, Proceedings of the 12th annual ACM international conference on Multimedia, pp. 952–959 (2004).
- [10] H. Liu, X. Xie, X. Tang, Z. Li and W. Ma: “Effective browsing of web image search results”, Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval, pp. 84–90 (2004).